

MICROBIOLOGY MONOGRAPHS

ALEXANDER STEINBÜCHEL

Series Editor

Edward Schwartz

Editor

Microbial Megaplasmids



Springer

Microbiology Monographs

Volume 11

Series Editor: Alexander Steinbüchel

Münster, Germany

Microbiology Monographs

Volumes published in the series

Inclusions in Prokaryotes

Volume Editor: Jessup M. Shively
Vol. 1, 2006

Complex Intracellular Structures in Prokaryotes

Volume Editor: Jessup M. Shively
Vol. 2, 2006

Magnetoreception and Magnetosomes in Bacteria

Volume Editor: Dirk Schüler
Vol. 3, 2007

Predatory Prokaryotes – Biology, Ecology and Evolution

Volume Editor: Edouard Jurkevitch
Vol. 4, 2007

Amino Acid Biosynthesis – Pathways, Regulation and Metabolic Engineering

Volume Editor: Volker F. Wendisch
Vol. 5, 2007

Molecular Microbiology of Heavy Metals

Volume Editors: Dietrich H. Nies and Simon Silver
Vol. 6, 2007

Microbial Linear Plasmids

Volume Editors: Friedhelm Meinhardt and
Roland Klassen
Vol. 7, 2007

Prokaryotic Symbionts in Plants

Volume Editor: Katharina Pawlowski
Vol. 8, 2009

Hydrogenosomes and Mitosomes: Mitochondria of Anaerobic Eukaryotes

Volume Editor: Jan Tachezy
Vol. 9, 2008

Uncultivated Microorganisms

Volume Editor: Slava S. Epstein
Vol. 10, 2009

Edward Schwartz
Editor

Microbial Megaplastids

 Springer

Editor

Dr. Edward Schwartz
Institut fuer Biologie, Mikrobiologie
Humboldt-Universität zu Berlin
Chausseestr. 117
10115 Berlin
Germany
edward.schwartz@rz.hu-berlin.de

Series Editor

Professor Dr. Alexander Steinbüchel
Institut für Molekulare Mikrobiologie und Biotechnologie
Westfälische Wilhelms-Universität
Corrensstraße 3
48149 Münster
Germany
steinbu@uni-muenster.de

Microbiology Monographs ISSN 1862-5576 e-ISSN 1862-5584
ISBN 978-3-540-85466-1 e-ISBN 978-3-540-85467-8
DOI: 10.1007/978-3-540-85467-8

Library of Congress Control Number: 2008940579

© 2009 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: WMX Design GmbH, Heidelberg, Germany

Printed on acid-free paper

9 8 7 6 5 4 3 2 1 0

springer.com

Preface

The term “megaplasmid” was first used by Rosenberg and coworkers in 1981 as a designation for plasmids “...with a molecular weight clearly greater than 300×10^6 ” (Rosenberg et al. 1981). Using gentle lysis techniques to minimize shearing of DNA molecules, these authors prepared lysates of *Sinorhizobium meliloti* cells and examined their plasmid content by means of agarose gel electrophoresis. They found a high-molecular weight species in several different strains of *S. meliloti* and showed that it carried determinants for both symbiosis and nitrogen fixation. As was customary at the time, the size of the DNA element was given as apparent molecular weight (100×10^6 is equivalent to about 150 kb). The high-molecular weight plasmid discovered in *S. meliloti* by Rosenberg and coworkers has been intensively studied in the past two decades, especially the species present in *S. meliloti* strain 1021, which is called pSymA.

The new term for “very large plasmids” was arbitrary in more than one respect. Clearly Rosenberg and his coauthors coined the term simply to indicate plasmids in a size range that had hitherto not been detected. There was no particular reason for the threshold of 300×10^6 (~ 450 kb) and no grounds to assume that these DNA species had biological properties distinguishing them from smaller plasmids. Inherent in the term “megaplasmid” were the tacit assumptions that bacteria have only a single chromosome, i.e. a replicon carrying genetic information that is essential for the cell under all growth conditions, and that, in bacteria harboring multiple replicons, the chromosome is always the largest replicon. Hence, in the original article cited above, the question of essentiality was not raised.

Since the discovery of giant plasmids in *S. meliloti*, our knowledge of bacterial genomes has progressed remarkably. On the one hand, many new plasmids in the size range >100 kb have been discovered in various bacteria and archaea and some have been shown to encode capabilities of immense ecological and industrial significance. More importantly, the advent of whole genome sequencing has led to fundamental changes in our thinking. It is now clear that diverse bacteria, such as *Rhodobacter sphaeroides* 2.4.1, *Brucella suis* and *Burkholderia pseudomallei* to name a few examples, possess more than one chromosome (Bentley and Parkhill 2004). This necessarily casts doubts on the classification of replicons as chromosomes or megaplasmids based on size alone. The case of *S. meliloti*, host of the prototypical megaplasmids, is interesting in this context. Curing experiments established

that the 1.3 Mb replicon pSymA is in fact nonessential and thus is a true megaplasmid. The status of the 1.6 Mb replicon pSymB was for long a matter of debate. Most but not all of pSymB can be deleted without significantly affecting viability. pSymB carries singular copies of a few essential housekeeping genes – a gene for $\text{tRNA}_{\text{CCG}}^{\text{Arg}}$ and the *minCDE* genes – making it essential for the growth of *S. meliloti* (Wong et al. 2002). This example illustrates that it is a very fine line separating chromosomes and megaplasms, as a single recombinative event can result in the transfer of an essential chromosomal gene to a plasmid. Indeed, it has been proposed that secondary chromosomes originated from megaplasms via the acquisition of essential housekeeping genes. Many secondary chromosomes have plasmid-type replication origins. Furthermore, analysis of the ParA partitioning proteins reveals that the sequences of ParA proteins deduced for secondary chromosomes cluster with plasmid ParA sequences, suggesting that they are more closely related (MacLellan et al. 2004).

From the above discussion it should be clear that the class of extrachromosomal elements that have been designated as megaplasms has no sharp boundaries, but rather is an arbitrary segment or slice of the spectrum of microbial replicons. What, then, is the justification for publishing a collection of articles devoted to “megaplasmid biology”?

The giant replicons surveyed in this book contribute to the biology of their hosts in a special way that sets them apart from smaller plasmids. They provide complex biochemical pathways or whole repertoires of gene products that in some cases enable the organism to prevail in special niches, such as H_2 -rich soil microcosms or the hostile environment of a mammalian host. Paradoxically, many of the megaplasms described in the following chapters are nonessential, but carry genetic determinants that define the unique biology of the organisms they inhabit.

For purely pragmatic reasons, a lower threshold of 100 kb was adopted for this selection of megaplasms. Notwithstanding this limitation it was not possible to include all of the interesting and well-studied megaplasms in this book. The reader will find that the scope of the various chapters is not uniform: Some chapters survey a lifetime of work, while others present a story that is just beginning to unfold. In our opinion they are all fascinating. To all the authors who in spite of their numerous other duties found time to review the work in their field in a novel vein, we express our gratitude. We are grateful to Springer and especially to Jutta Lindenborn of the editorial staff for her support and counsel.

Berlin and Münster
September 2008

Edward Schwartz
Alexander Steinbüchel

References

- Bentley SD, Parkhill J (2004) Comparative genomic structure of prokaryotes. *Annu Rev Genet* 38:771–791
- MacLellan SR, Sibley CD, Finan TM (2004) Second chromosomes and megaplasms in bacteria. In: Funnell BE, Phillips GJ (eds) *Plasmid biology*. ASM Press, Washington DC, pp 529–542
- Rosenberg C, Boistard P, Dénarié J, Casse-Delbart F (1981) Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol Gen Genet* 184:326–333
- Wong K, Finan TM, Golding GB (2002) Dinucleotide compositional analysis of *Sinorhizobium meliloti* using the genome signature: distinguishing chromosomes and plasmids. *Funct Integr Genomics* 2:274–281

Contents

Part I Megaplastids in Archaea

- Haloarchaeal Megaplastids** 3
Shiladitya DasSarma, Melinda Capes, and Priya DasSarma

Part II Megaplastids and Biodegradation

- Megaplastids and the Degradation of Aromatic Compounds by Soil Bacteria**..... 33
Eve Vedler

- Catabolic Plasmids Involved in the Degradation of Polycyclic Aromatic Hydrocarbons and Heteroaromatic Compounds**..... 55
Hideaki Nojiri, Masahiro Sota, and Masaki Shintani

Part III Megaplastids in Plant Symbionts and Pathogens

- Sinorhizobium meliloti* Megaplastids and Symbiosis in *S. meliloti*** 91
Frédérique Barloy-Hubler and Mohamed Jebbar

- The Megaplastid pNGR234a of *Rhizobium* sp. Strain NGR234**..... 119
Patrick Mavingui

- Ti and Ri Plasmids** 133
Katsunori Suzuki, Katsuyuki Tanaka, Shinji Yamamoto, Kazuya Kiyokawa, Kazuki Moriguchi, and Kazuo Yoshida

Part IV Megaplastids in Human Pathogens

- The Virulence Plasmids of *Shigella flexneri*** 151
Charles J. Dorman

The Three Bears and Virulence-Associated Plasmids in the Genus <i>Yersinia</i>.....	171
Sarah L. Howard, Philippa C.R. Strong, and Brendan W. Wren	
Virulence Megaplastmids in <i>Bacillus anthracis</i> and Their Relatives in the <i>Bacillus cereus</i> Group	187
Agnès Fouet and Marie Moya	
Part V Megaplastmids and Special Metabolic Functions	
Megaplastmids in <i>Cupriavidus</i> Genus and Metal Resistance.....	209
Max Mergeay, Sébastien Monchy, Paul Janssen, Rob Van Houdt, and Natalie Leys	
Megaplastmids of Aerobic Hydrogenotrophic and Carboxidotrophic Bacteria	239
Edward Schwartz	
The Megaplastmid pAO1 of <i>Arthrobacter Nicotinovorans</i> and Nicotine Catabolism	271
Petra Ganas, Gabor L. Igloi, and Roderich Brandsch	
The pMUM Megaplastmid of <i>Mycobacterium ulcerans</i> and Closely Related Mycobacteria: A Blueprint for the Synthesis of Mycolactones.....	283
Sacha J. Pidot, Nicholas J. Tobias, and Tim Stinear	
Megaplastmid pKB1 of the Rubber-Degrading Bacterium <i>Gordonia westfalica</i> Strain Kb1	297
Daniel Bröker and Alexander Steinbüchel	
Megaplastmid pMP118 of <i>Lactobacillus salivarius</i>	311
Paul W. O'Toole and Emma J. Raftis	
Index.....	327

Haloarchaeal Megaplasמידs

Shiladitya DasSarma (✉), Melinda Capes, and Priya DasSarma

Contents

1	Introduction.....	4
2	<i>Halobacterium</i> sp. NRC-1.....	6
2.1	Restriction Mapping pNRC100 and Inversion Isomers.....	8
2.2	Gas Vesicle Genes and Mutants.....	9
2.3	Mapping an Origin of Replication.....	10
2.4	pNRC100 Megaplasמיד Genes.....	11
2.5	pNRC200 Megaplasמיד Genes.....	12
2.6	Additional Genetic Studies on pNRC100 and pNRC200.....	15
3	<i>Haloferax volcanii</i>	16
3.1	pHV4 Megaplasמיד.....	16
3.2	pHV3 Megaplasמיד.....	18
3.3	pHV1.....	19
3.4	pHV Replication Origins.....	19
4	<i>Haloarcula marismortui</i>	20
4.1	pNG700 Megaplasמיד.....	20
4.2	Chromosome II.....	21
4.3	pNG600 Megaplasמיד.....	22
4.4	pNG500 Megaplasמיד.....	22
4.5	Other <i>H. marismortui</i> Plasmids.....	22
5	<i>Natronomonas pharaonis</i>	24
6	<i>Haloquadratum walsbyi</i>	24
7	Evolution of Haloarchaeal Megaplasמידs.....	25
	References.....	27

Abstract Many salt-loving archaea, or haloarchaea, commonly harbor one or more megaplasמידs in their genomes. The haloarchaeal model organism, *Halobacterium* sp. NRC-1, contains two related replicons, pNRC100 (191 kb in size) and pNRC200 (365 kb), both of which code for buoyant gas vesicles as well as other genes that are

S. DasSarma, M. Capes, and P. DasSarma
University of Maryland Biotechnology Institute, Center of Marine Biotechnology,
Baltimore, MD 21202, USA

S. DasSarma and M. Capes
Graduate Program in Life Sciences, University of Maryland, Baltimore, MD 21201, USA
e-mail: dassarma@umbi.umd.edu

important or essential for the host. Two other haloarchaea, *Haloarcula marismortui* and *Haloferax volcanii*, with larger genomes have a more complex complement of extrachromosomal replicons, eight in the former, including four megaplasmids, and four in the latter, including two megaplasmids and one very large plasmid. Two other sequenced haloarchaea, *Natronomonas pharaonis* and *Haloquadratum walsbyi*, have fewer and smaller extrachromosomal replicons, with a single megaplasmid in the former and only a relatively small plasmid in the latter. We review the current state of knowledge on these sequenced megaplasmids, including the eukaryotic type replication protein genes present in most, IS elements populating many, and unusual (e.g., gas vesicle), important (cytochrome oxidase and thioredoxin/thioredoxin reductase), and even essential (transcription and replication factors and an aminoacyl-tRNA synthetase) genes on several of these large replicons. The role of megaplasmids, some of which may qualify as small chromosomes (minichromosomes), in the evolution of haloarchaeal genome architecture is discussed.

1 Introduction

The advent of high-throughput mapping and sequencing technologies developed in the 1980s and 1990s and subsequent post-genomic methodologies of the early twenty-first century have allowed detailed understanding of prokaryotic genomes, including members of the third domain of life, the archaea (Fraser et al. 2004). Archaeal genomes range in size from 0.5–5.75 Mbp and are organized, like most bacterial genomes, with a single main circular chromosome, and on occasion, with accessory plasmids or extrachromosomal elements. Among the halophilic or salt-loving Archaea (also called haloarchaea), which are geographically widely distributed in salty environments around the world, large plasmids are very common. Nine of these haloarchaeal megaplasmids have been completely sequenced and several have been classified as minichromosomes due to the presence of important or essential genes (DasSarma 2004). The best studied example, pNRC100, in the model *Halobacterium* sp. NRC-1, is distinguished by an abundance of transposable IS elements that promote DNA rearrangements and cause frequent phenotypic variability (Fig. 1) (DasSarma and Arora 1997; Ng et al. 1998).

Currently, of the five complete haloarchaeal genome sequences (Tables 1 and 2), four harbor megaplasmids larger than 100 kb. The genome of the first sequenced haloarchaeon, *Halobacterium* sp. NRC-1, 2.57 Mbp in size, contains two large, related megaplasmids, pNRC100 and pNRC200, specifying buoyant gas vesicles (Ng et al. 1998; Ng et al. 2000; DasSarma 2004). Because of the presence of many important and a few essential genes, these replicons were proposed to be essential minichromosomes. *Halobacterium* is widely distributed in nature, for example, as a major component of Great Salt Lake in the western USA (Baxter et al. 2005), and is one of the most intensively studied archaea to date (DasSarma et al. 2006). Two halophiles from the Dead Sea, *Haloferax volcanii* and *Haloarcula marismortui*, have bigger, more complex genomes with larger numbers of megaplasmids (Charlebois et al. 1991; Baliga et al. 2004). *H. marismortui* has a 4.3 Mbp genome

with four megaplasms ranging in size from 411 to 133 kb (pNG700, pNG600, and pNG500, and one which was designated as Chromosome II). *H. volcanii* contains a slightly smaller genome, 4.1 Mbp, with two megaplasms, pHV4 (636 kb) and pHV3 (438 kb), which are the largest characterized among haloarchaea to date. In addition, *H. volcanii* contains a third large plasmid, pHV1 (85 kb), which is

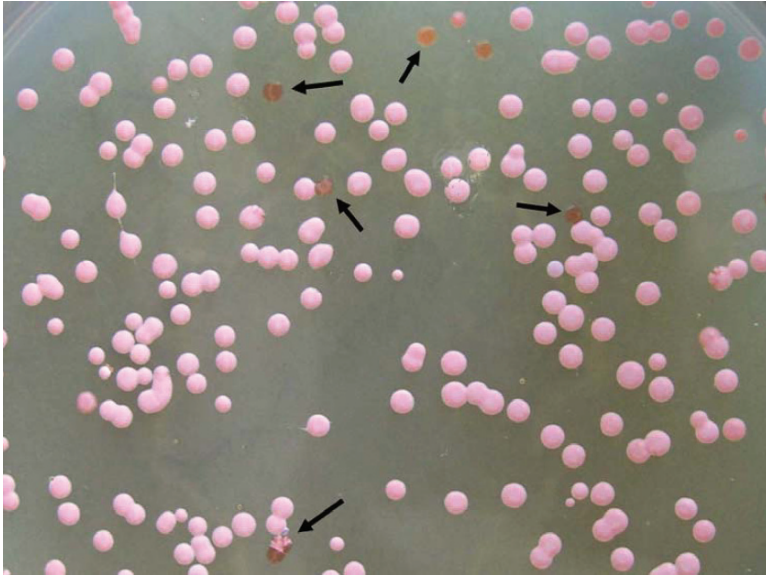


Fig. 1 Colonies of *Halobacterium* sp. NRC-1 on a salt agar plate showing wild-type (pink, opaque) gas vesicle-containing colonies, and with *arrows*, mutant (orange, translucent, or sectorized) gas vesicle mutants. Mutants result from high-frequency DNA rearrangements in megaplasms pNRC100 and pNRC200

Table 1 Sequenced haloarchaeal organisms

Organism	Physiology and ecology	Genome	Reference
<i>Halobacterium</i> sp. NRC-1 ATCC 700922, JCM 11081	Extreme halophile, facultative anaerobe, and phototroph	2.6 Mbp, 2 megaplasms	DasSarma (2004)
<i>Haloferax volcanii</i> DS2 ATCC 29605, DSM 3757, JCM 8879	Moderate halophile, prototroph, from Dead Sea mud	4.1 Mbp, 2 megaplasms	Charlebois et al. (1991)
<i>Haloarcula marismortui</i> ATCC 43049, DSM 3752; JCM 8966	Extreme halophile, metabolically versa- tile, from Dead Sea	4.3 Mbp, 4 megaplasms	Baliga et al. (2004)
<i>Natronomonas pharaonis</i> strain Gabara ATCC 35678, DSM 2160, JCM 8858	Alkaliphilic extreme halophile from soda lake	2.8 Mbp, 1 megaplasmid	Falb et al. (2005)
<i>Haloquadratum walsbyi</i> HBSQ001, DSM16790, JCM 12895	Square-shaped extreme halophile from solar saltern	3.2 Mbp, no megaplasmid	Bolhuis et al. (2006)

Table 2 Haloarchaeal replicons in sequenced genomes

Organism	Replicon name	Length (bp)	G + C content (%)	NCBI gene count	NCBI RefSeq accession
<i>Halobacterium</i> <i>sp. NRC-1</i>	Chromosome	2,014,239	67.90	2,127	NC_002607
	pNRC200 ^a	365,425	59.22	371	NC_002608
	pNRC100 ^a	191,346	57.91	176	NC_001869
<i>Haloferax</i> <i>volcanii</i>	Chromosome	2,847,757	66.64	2,960	ND
	pHV4 ^a	635,786	61.67	638	ND
	pHV3 ^a	437,906	65.56	362	ND
	pHV1	85,092	55.50	89	ND
	pHV2	6,359	56.06	6	ND
<i>Haloarcula</i> <i>marismortui</i>	Chromosome	3,131,724	62.36	3,186	NC_006396
	pNG700 ^a	410,554	59.12	363	NC_006395
	chromosome II ^a	288,050	57.23	285	NC_006397
	pNG600 ^a	155,300	58.33	167	NC_006394
	pNG500 ^a	132,678	54.48	131	NC_006393
	pNG400	50,060	57.35	51	NC_006392
	pNG300	39,521	60.02	40	NC_006391
	pNG200	33,452	55.63	42	NC_006390
	pNG100	33,303	54.25	36	NC_006389
	<i>Natronomonas</i> <i>pharaonis</i>	Chromosome	2,595,221	63.40	2,726
pL131 ^a		130,989	57.22	132	NC_007427
pL23		23,486	60.59	36	NC_007428
<i>Haloquadratum</i> <i>walsbyi</i>	Chromosome	3,132,494	47.90	2,875	NC_008212
	pL47	46,867	47.67	36	NC_008213

ND not deposited, ^aMegaplasmid replicons

slightly smaller than the minimum size to be classified as a megaplasmid, and a miniplasmid, pHV2, which is only 6 kb. The alkaliphilic haloarchaeon, *Natronomonas pharaonis*, from the soda lakes of the Sinai Peninsula has a 2.8 Mbp genome with a single 131 kb megaplasmid, pL131, and a smaller plasmid, pL23 (23 kb), that is also found integrated in the chromosome (Falb et al. 2005). The square-shaped haloarchaeon from crystallizer ponds, *Haloquadratum walsbyi*, is the only haloarchaeal genome devoid of megaplasmids, although it does harbor a large (47 kb) plasmid, pL47 (Bolhuis et al. 2006).

In this chapter, our current knowledge on each of the sequenced haloarchaeal mega- and large plasmids is reviewed from the perspective of gene content, and where available, genome structure and replication characteristics. Certain of these studies have led to deeper insights into the evolution of megaplasmids and chromosomes resident within an individual species.

2 *Halobacterium* sp. NRC-1

Early genomic studies of several related *Halobacterium* species indicated that they had a novel composition based on the finding of two components, a major G + C-rich (67%) fraction and a relatively A + T-rich (58% G + C) satellite (Joshi et al.

1963; Moore and McCarthy 1969). The satellite DNAs were subsequently shown to correspond mainly to large and heterogeneous extrachromosomal replicons containing many IS elements (Pfeifer et al. 1981; DasSarma 1989). For a typical *Halobacterium* isolate, strain NRC-1, extensive restriction mapping showed the presence of two megaplasms, pNRC100 and pNRC200 (Ng et al. 1991; Ng and DasSarma 1991). Additionally, several minor circular DNAs related to the pNRC megaplasms with lower copy numbers were also found (Ng et al. 1994). Surprisingly, the pNRC100 and pNRC200 replicons were found to be identical over an extended region and to contain large inverted repeats, recombination between which resulted in the formation of inversion isomers (Figs. 2 and 3a). The presence of deletion derivatives and minor circular DNAs could also be ascribed to recombinational activities of IS elements resident on these pNRC plasmids.

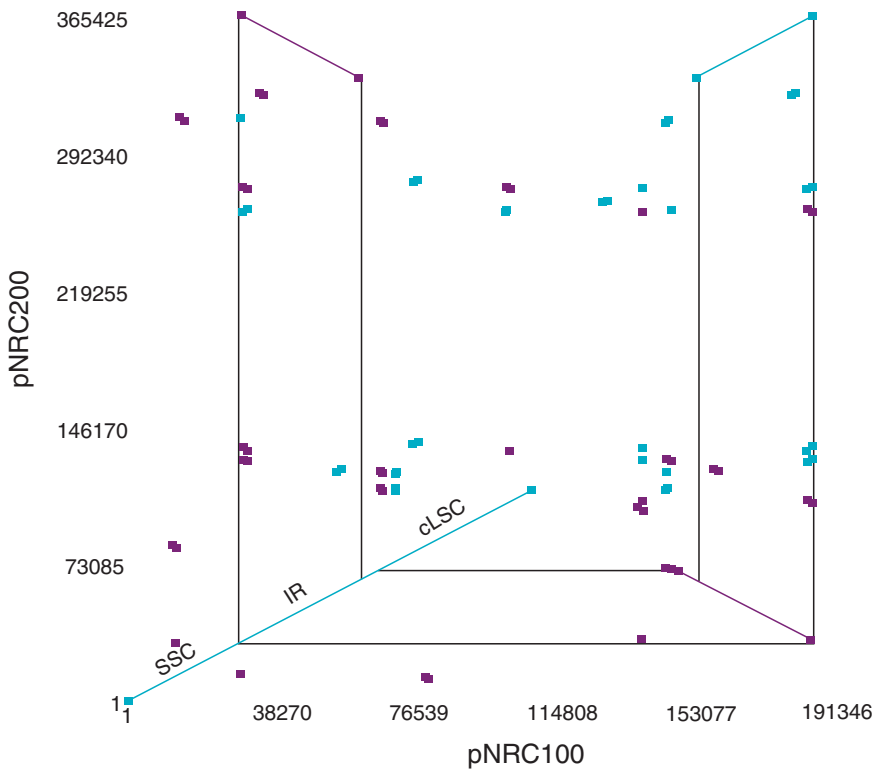


Fig. 2 Alignment of pNRC100 and pNRC200 using the TIGR genome alignment tool. Extended regions of homology are indicated by *diagonal lines* and short regions of homology are shown by *blocks*. The small single copy regions shared by the two megaplasms are labeled SSC and the common regions of the large single copy regions are labeled cLSC. One copy of the inverted repeats is labeled IR. The horizontal and vertical lines are used for alignment of the IRs, which are slightly larger on pNRC100 than on pNRC200

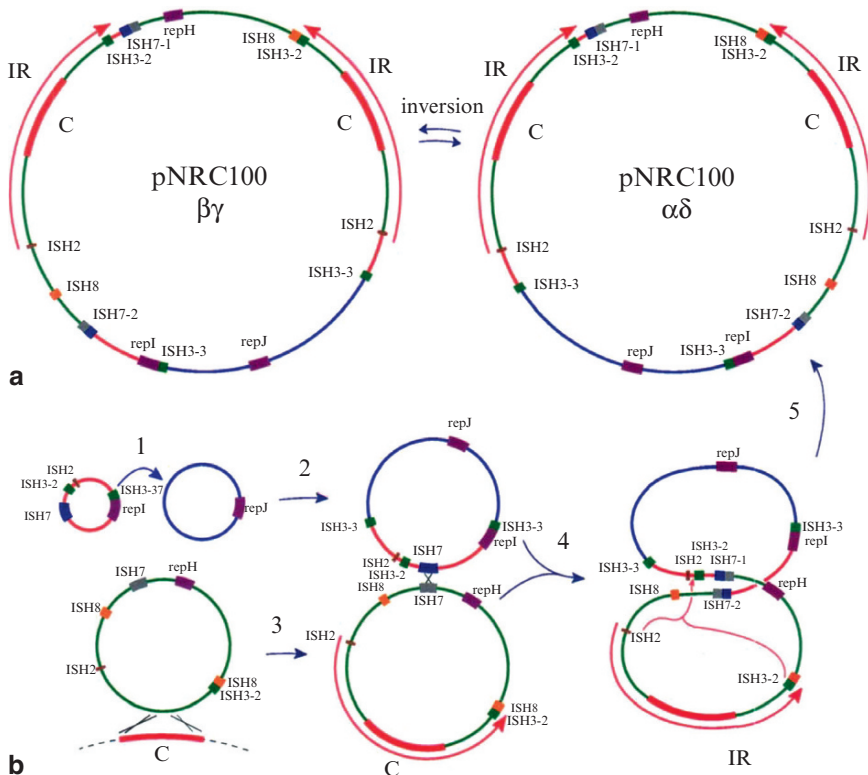


Fig. 3 (a) Inversion isomerization and (b) possible mechanism of evolution of pNRC100. Three replicase genes and a subset of IS elements are shown in colored boxes as is the chromosome-like region, labeled “C.” The large inverted repeats are shown with *pink arrows*. Recombinational steps are shown with *numbered arrows*, and discussed in the text. Adapted from Ng et al. 1998

2.1 Restriction Mapping pNRC100 and Inversion Isomers

Physical mapping of pNRC100 using pulsed-field gel electrophoresis showed the presence of a 35,000–38,000 bp inverted repeat (IR) sequence (Figs. 2 and 3) (Ng and DasSarma 1991). Inversion isomers of pNRC100 were demonstrated by Southern hybridization analysis using rare-cutting restriction enzymes (*Afl*III and *Sfi*I), which cut asymmetrically within the intervening small single-copy and large single-copy regions, respectively, but not within the large IRs (Fig. 3a). In this regard, the pNRC100 structure resembled some chloroplast and mitochondrial genomes which also contain large IRs (Gray 1999). Subsequent sequencing analysis of pNRC100 showed the presence of important genes, such as *trxA*B coding for thioredoxin and thioredoxin reductase, involved in protein reduction, and *cydAB* specifying cytochrome oxidase, important in aerobic respiration (Ng et al. 1998). However, no rRNA genes were located in the pNRC IRs or elsewhere on the replicons, as is common in

organellar genomes (Gray 1999). No inversion isomers were observed for a deletion derivative of pNRC100 lacking one copy of the IRs, indicating that both copies are required for inversion to occur (Ng et al. 1994). These early studies also established the identities and approximate positions of over a dozen IS elements in pNRC100, using Southern hybridization and limited nucleotide sequence analysis across the IS element-target site junctions. Four classes of IS elements were initially found: ISH2, a 0.5-kb element; ISH3, a heterologous family of 1.4-kb elements; ISH4, a 1.0-kb element; and ISH8, a 1.4-kb element, representing a relatively small fraction of all the elements in this megaplasmiid. The large IRs of pNRC100 terminated at an ISH2 element at one end and an ISH3 element at the other end.

2.2 Gas Vesicle Genes and Mutants

One of the most interesting early findings was that *Halobacterium* sp. NRC-1 plasmid pNRC100 codes for synthesis of buoyant gas-filled vesicles used for cell flotation (DasSarma et al. 1987). This finding was based on the phenotypic changes observed in *Halobacterium* colonies containing or lacking gas vesicles; since gas vesicles refract light, they impart an opaque appearance to colonies while mutant colonies lacking gas vesicles form translucent colonies (DasSarma et al. 1988) (Fig. 1). In some *Halobacterium* strains, the colony opacity was highly variable, with loss of gas vesicle formation at frequencies of 1%. In early studies, two genes, *gvpA* and *C*, were identified on pNRC100 through cross hybridization of the *gvpA* gene of the cyanobacterium *Calothrix* sp. and sequencing analysis (Jones et al. 1989). Analysis of gas vesicle-deficient mutants showed the presence of DNA rearrangements associated with this megaplasmiid, including insertions, deletions, inversions, and more complex rearrangements. The rearrangements were mostly IS element-mediated, and when they occurred in a region of pNRC100 around *gvpAC*, mutants lacking gas vesicles resulted. Further sequencing and analysis of mutants in the *gvpAC* gene region identified a gene cluster of 13 genes, *gvpMLKJIHGFEDACN*, involved in gas vesicle synthesis (Jones et al. 1991; Halladay et al. 1993). The gene cluster was found to be organized in two divergent operons, with *gvpDEFGHIJKLM* oriented leftward and *gvpACN* operons oriented rightward (Yang and DasSarma 1990). A transcriptionally unlinked fourteenth gene, *gvpO*, was subsequently found.

The requirement of the entire *gvp* gene cluster, *gvpMLKJIHGFEDACN* and possibly *O*, on pNRC100, for gas vesicle synthesis and cell flotation was demonstrated by genetic transformation (Halladay et al. 1992). The *gvp* gene cluster was cloned and reconstructed on a *Halobacterium*-*E. coli* shuttle plasmid, and transformation of gas vesicle-deficient mutants lacking the entire gas vesicle gene region resulted in restoration of the cells' ability to float. To study the functions of the *gvp* genes in gas vesicle formation, the gene cluster was subjected to linker scanning mutagenesis (DasSarma et al. 1994). A kanamycin resistance (*kappa*) cassette was introduced into each *gvp* gene (except for *gvpA*) on the shuttle plasmid pFL2. Transformation of the *Halobacterium* sp. NRC-1 mutant, SD109, which had the entire *gvp* gene

cluster deleted, with pFL2 and mutated derivatives of pFL2 showed that while the unmutated gene cluster successfully programmed gas vesicle formation, derivatives with insertion of the *kappa* cassette in any of the *gvp* genes, except *gvpM*, did not lead to production of normal gas vesicles. Insertions in *gvpL*, *K*, *J*, *I*, and *F* resulted in a complete block in gas vesicle synthesis, while insertions in *gvpH*, *G*, *E*, *D*, *C*, and *N* resulted in greatly reduced gas vesicle synthesis. In most cases, the block in gas vesicle synthesis did not result from polar effects, since similar results were obtained for derivatives of the insertion mutants in which most of the internal portion of the *kappa* cassette was deleted and only small (15–54bp) insertions remained. The only exceptions were for *gvpH* and *gvpD*, where deletion of the internal portion of the *kappa* insertions resulted in phenotypic reversion. Electron microscopic analysis of the *kappa* mutants revealed that interruptions of *gvpC* and *gvpN* cause the formation of smaller gas vesicles than in the wild type (DasSarma et al. 1994). These results indicated that the *gvpA*, *C*, and *N* genes encode structural proteins, with either *gvpC* or *gvpN* necessary for late stages of vesicle formation, and *gvpL*, *K*, *J*, *I*, *H*, *G*, and *F* involved in early steps in the assembly of gas vesicles.

Immunoblotting analysis was used to probe for the presence of gas vesicle proteins corresponding to the major GvpA and GvpC proteins and five additional minor *gvp* gene products (Halladay et al. 1993; Shukla and DasSarma 2004). Polyclonal antisera raised in rabbits against LacZ-GvpA, -GvpC, -GvpF, -GvpJ, and -GvpM fusion proteins and against synthetic peptides from GvpG and L were used as probes. Immunoblotting analysis was performed on cell lysates of wild-type *Halobacterium* sp. strain NRC-1, gas vesicle-deficient mutants, and purified gas vesicles, after purification of LacZ fusion antibodies on protein A and β -galactosidase affinity columns. The presence of seven gas vesicle proteins (GvpA, GvpC, GvpF, GvpG, GvpJ, GvpL, and GvpM) was demonstrated, indicating that the corresponding pNRC100 genes were indeed important for gas vesicles formation. Two of the minor gas vesicle proteins were found to be similar to GvpA (GvpJ and GvpM), and two proteins contained predicted coiled-coil domains (GvpF and GvpL). GvpL exhibited a multiplet ladder on SDS-polyacrylamide gels, indicative of oligomerization and self-assembly (Shukla and DasSarma 2004).

Interestingly, two copies of a *gvp* gene cluster similar to those in *Halobacterium* sp. NRC-1 were described on the *H. walsbyi* chromosome (Bolhuis et al. 2006), and the gas vesicles were found to localize around the perimeter of these unusual square cells. The clusters each consisted of 12 divergently transcribed genes, *gvpACNO* and *gvpFGHIJKLM*, separated by 3.6 kb containing several nonfunctional IS element fragments and a ParA domain protein gene. The missing genes *gvpD* and *E* are likely to be regulators and their absence in this organism was a surprising finding.

2.3 Mapping an Origin of Replication

A replication origin of pNRC100 was identified by assaying for autonomous replication ability of miniplasmids containing cloned fragments of pNRC100, using the mevinolin resistance selectable marker of *H. volcanii* (Ng and DasSarma 1993).

Initially, the autonomous replication ability of the 19-kb *Hind*III-C fragment of pNRC100 was demonstrated, followed by the definition of the minimal replication origin of approximately 3.9 kb by subcloning of successively smaller regions and assaying for plasmid replication in either *Halobacterium* or *H. volcanii*. The same replication origin was also recovered after transformation of *H. volcanii* with a library of partial *Sau*3AI fragments of pNRC100. The nucleotide sequence of the minimal replication origin contained a gene, *repH*, and a highly A + T-rich region located upstream to it. Later, after the genome sequence was determined, it became clear that this pNRC100 origin region was also present on pNRC200 (Ng et al. 2000).

The *repH* gene predicted a large acidic protein similar to predicted gene products of *H. volcanii* miniplasmid pHV2 and *Halobacterium* phage FH, suggesting that each RepH-like protein is involved in DNA replication (Ng and DasSarma 1993). The replication region of pNRC100 on a minireplicon was analyzed by linker scanning mutagenesis, which showed the requirement of *repH* for replication. Restoration of the *repH* reading frame of one replication-defective gene by introduction of a second small insertion resulted in reversion to replication proficiency. The replication ability of the minireplicon was lost when the entire A + T-rich region, about 550 bp long, was deleted but not when small insertions or deletions were introduced into this region. The presence of only 52 bp of the A + T-rich segment was sufficient to permit replication. The *repH*-containing minireplicon was lost at high frequency from cells grown without mevinolin selection, suggesting that any plasmid partitioning locus of pNRC100 was absent in the minimal replication origin region (Ng et al. 1993; DasSarma and Arora 1997).

2.4 pNRC100 Megaplasמיד Genes

As a pilot project to the *Halobacterium* sp. NRC-1 genome sequencing, the pNRC100 replicon was sequenced by a combination of random shotgun sequencing of libraries made from purified covalently closed circular DNA and directed sequencing of the mapped *Hind*III fragments (Ng et al. 1998). The sequence showed pNRC100 to be a 191,346-bp circle with 58% G + C and 29 transposable IS elements known at that time. Complete shotgun sequencing subsequently revealed a 2,571,010-bp genome in *Halobacterium* NRC-1, including a 2,014,239-bp 68% G + C-rich chromosome, and two megaplasמידs, pNRC100, and the 365,425 bp pNRC200 (59% G + C) (Table 2) (Ng et al. 2000; DasSarma 2004). Interestingly, pNRC100 and pNRC200 contained a 145,428 bp region of 100% identity, including nearly the entire large IRs, the small single copy region, and a portion of the large single copy regions (Figs. 2 and 3a). The unique regions of the large single copy region contained 45,918 bp for pNRC100 and 219,997 bp for pNRC200. The presence of large inverted repeats in pNRC200 suggested that this megaplasמיד also undergoes frequent inversion isomerization like the smaller pNRC100 replicon, although this prediction was not tested experimentally.

The gene prediction program, Glimmer, was used to predict 197 genes in pNRC100 and 384 genes in pNRC200 (Ng et al. 2000; DasSarma 2004). About 40 genes in pNRC100 and pNRC200 coded for housekeeping proteins, some of which are unique and likely to be essential or important for cell viability. These findings suggested that these replicons should be classified as minichromosomes rather than megaplasmids (Ng et al. 1998, 2000). Among the most interesting putative gene products encoded by pNRC100 (Table 3) were the following: (1) Four genes, *tbpA*, *B* (on the IRs), *C*, and *D*, code for transcription factors similar to the eukaryotic and archaeal general transcription factor TBP (TATA binding protein; Thomm 1996). The multiplicity of TBPs suggested by this finding was shown, with one exception, to be a general characteristic of haloarchaea. (2) Four genes code for proteins similar to replication proteins, with one, *orc9*, coding a member of the Orc/Cdc6 family of eukaryotic origin-binding proteins (Gavin et al. 1995), while the other three code members of the RepH family (Ng and DasSarma 1993). (3) Five genes are similar to the ParA and Soj family proteins, which may be involved in plasmid and chromosomal partitioning, for orderly segregation of replicated molecules into daughter cells (Ghosh et al. 2006). (4) Two genes (*cydA* and *B*), present on the IRs, code for proteins similar to subunits I and II of *E. coli* cytochrome *d* oxidase, a terminal electron acceptor in the electron transport chain (Miller and Gennis 1983). (5) Two genes (*trxA* and *B*), also on the IRs, code for thioredoxin and thioredoxin reductase, which function in protein reduction, including turnover of the ribonucleotide reductase enzyme used for conversion of ribonucleotides to deoxyribonucleotides for DNA synthesis (Arner and Holmgren 2000). (6) Another gene, *htlD*, codes a protein similar to protein kinases of two-component regulatory systems, including both components of the FixLJ-like system involved in oxygen regulation (David et al. 1988; Lois et al. 1993). (7) Four genes code proteins similar to arsenic-resistance proteins and regulators, including the catalytic subunit of the arsenite pump and arsenate reductase (Rosen 2002). (8) Two genes code for proteins similar to type III restriction-modification systems. (9) Finally, as described earlier, over a dozen *gvp* genes code for proteins involved in gas vesicle formation (Halladay et al. 1992, 1993; DasSarma et al. 1994; DasSarma and Arora 1997; Shukla and DasSarma 2004).

2.5 *pNRC200 Megaplasmid Genes*

pNRC200 was sequenced as part of the *Halobacterium* sp. NRC-1 whole genome via random shotgun sequencing, with selected regions verified using polymerase chain reaction fragments and by primer walking (Tables 2 and 3). Since pNRC200 contained a 145,428-bp region of 100% identity with pNRC100 (including the entire small single copy region and nearly all of the IRs), the larger megaplasmid contained most of the genes found on the smaller. They included the *repH* gene and origin region, the gas vesicle gene cluster, the *htlD* gene, and on the large IRs, *tbpB*, *cydA* and *B*, *trxA* and *B*. The arsenic resistance gene region, which is present on the large unique single copy region of pNRC100, was absent in pNRC200.

Table 3 Gene content in large extrachromosomal haloarchaeal replicons

	pNRC100 ^a	pNRC200 ^a	pHV1	pHV3	pHV4	pNG100	pNG200	pNG300	pNG400	pNG500	pNG600	chr2	pNG700	pL23	pL131	pL47
DNA replication and transactions																
Origin recognition	1	6	2	1	4	1	1	1	1	2	2	2	1	1	1	1
Plasmid replicases	3	2	-	-	-	-	1	1	1	-	-	-	-	-	-	-
Replicative helicases	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-
Primase	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
B-type DNA polymerase	-	1	-	-	1	-	-	-	-	-	1	-	-	-	-	-
RPA-family ssDNA binding	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Transposases	8	13	14	3	36	-	-	1	3	22	-	8	2	-	10	-
Site-specific recombinases ^b	2	2	-	-	3	1	-	-	-	-	-	1	1	1	-	2
Res/Mod enzymes and exonucleases	2	3	-	-	2	-	-	-	-	-	2	1	-	-	2	3
Dam methylase	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-
UvrABC repair	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-
Partitioning and segregation	5	8	1	2	2	-	2	1	1	2	1	3	1	-	1	1
Conjugation	-	-	1	-	-	-	-	-	-	3	2	-	-	-	-	-
Other DNA-binding proteins ^c	-	-	-	6	11	-	1	1	1	3	1	2	2	-	2	-
Other helicases	1	1	1	-	6	1	1	-	1	-	1	-	-	-	-	1
CRISPR-associated proteins/repeats	-	-	-	-	7	-	-	-	4	-	-	-	-	-	-	-
Transcription and sensing/regulation																
TATA-binding proteins	5	5	-	1	-	-	-	-	-	-	-	-	-	-	-	-
Transcription initiation factor IIB TFB	2	2	-	1	-	-	-	-	-	-	2	1	-	-	-	-
Transcriptional regulators	2	2	1	19	22	1	3	1	3	7	9	6	16	-	6	-
Regulatory systems ^d	1	4	1	6	6	-	-	-	-	-	-	11	6	1	4	-
Ribonucleases	-	-	-	1	2	-	-	-	-	-	1	1	2	-	-	-
Protein synthesis and stability																
tRNAs	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-
rRNAs	-	-	-	-	-	-	-	-	-	-	-	3	1	-	-	-

(continued)

Table 3 (continued)

	pNRC200 ^a	pHV1	pHV3	pHV4	pNG100	pNG200	pNG300	pNG400	pNG500	pNG600	chr2	pNG700	pL23	pL131	pL47
AA tRNA synthetases and charging ^e	1	-	-	1	-	-	-	-	-	-	1	-	-	-	-
Translation initiation factors	-	-	3	1	-	-	-	-	-	-	-	1	-	-	-
Ribosomal proteins	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-
Chaperones and stress proteins	-	-	1	7	-	-	-	-	1	1	3	7	-	3	-
Metabolism	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cytochrome oxidases	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-
Other cytochromes	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-
Nitrate metabolism	-	1	2	3	-	-	-	-	-	1	-	1	-	-	-
Thioredoxin and trx reductase	4	4	-	-	-	-	-	-	-	2	1	3	-	-	-
Cobalamin biosynthesis	-	-	-	16	-	-	-	-	-	-	-	-	-	-	-
Heavy metal transporters	4	1	1	1	-	-	-	-	9	1	1	-	-	2	-
Other transporters	2	27	4	57	-	-	-	-	3	2	21	41	1	17	-
Drug resistance	-	1	-	2	3	-	-	-	-	-	2	1	1	-	-
Superoxide dismutase	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
Gas vesicle	14	26	-	-	-	-	-	-	-	-	-	-	-	-	-
FtsZ/tubulin GTPase	-	1	-	1	-	-	-	-	-	-	-	1	-	-	-
Flagellar components	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cobalamin biosynthesis	-	-	-	16	-	-	-	-	-	-	1	-	-	-	-

^aEach copy of genes repeated in the IRs has been counted

^bResolvases and integrases

^cIncludes zinc-finger/HTH/ribbon/helix/winged

^dIncludes sensors and kinases, but not transcriptional regulators

^eIncludes aminoacyl-tRNA synthetases and Asn/Glu-tRNA^{Asn/Gln} amidotransferase

pNRC200 was also found to contain additional genes not present on pNRC100 within its 219,997 bp unique region in the large single copy region. Among the most interesting additional genes in this region of pNRC200 are the following: (1) A fifth TBP transcription factor gene, *tbpF*; (2) Two TFB (transcription factor IB) genes, *tfbC* and *E*; (3) Five additional *orc/cdc6* genes (*orc1*, 2, 3, 4, and 5); (4) A eukaryotic family B DNA polymerase gene, *polB2*, which is located upstream of *orc4*, an arrangement similar to that found on haloarchaeal megaplasmids in both *H. volcanii* and *H. marismortui*; (5) The only arginyl-tRNA synthetase gene, *argS*, in the genome, and therefore predicted to be critical for protein synthesis; (6) Two additional Soj family partitioning protein genes, *sojC2* and *E*; (7) A FtsZ gene, *ftsZ5*, coding a probable cell division GTPase; (8) An additional restriction system protein gene; (9) An additional nearly complete gas vesicle protein gene cluster, missing only *gvpM*; and (10) The *arcRABC* gene cluster, coding the arginine deiminase pathway for anaerobic arginine fermentation and its regulation.

2.6 Additional Genetic Studies on pNRC100 and pNRC200

In addition to genetic studies of *gvp* genes and the *repH* gene region of pNRC100, the *ura3*-based gene knockout method was applied to a number of additional pNRC100 and pNRC200 genes. Genes homologous to arsenic resistance were targeted on pNRC100 (*arsADRC* and *arsR2M*) (Wang et al. 2004). Deletion of the entire *arsADRC* gene cluster resulted in increased sensitivity to arsenite and antimonite but not arsenate. Knockout of the *arsM* gene produced sensitivity to arsenite, suggesting a second novel mechanism of arsenic resistance, involving a likely arsenite(III)-methyltransferase. These results taken together indicated that pNRC100 encodes two different arsenite and antimonite extrusion systems. The identity of the ArsB pump itself was not established, however, since deletion of a chromosomal gene predicted to code this function did not confer an arsenic-sensitive phenotype. The arsenite and antimonite resistance elements were also shown to be regulated, with resistance to arsenic in the wild type inducible by exposure to a sublethal concentration of the toxic metal ions. *arsA*, *arsD*, *arsR*, *arsM*, *arsC*, and surprisingly, the *arsB*-like gene, but not *arsR2*, were inducible by arsenite and antimonite (Wang et al. 2004). These results are interesting in light of the fact that heavy metals like arsenic are frequently concentrated in hypersaline environments (Oremland et al. 2004).

A genetic study of DNA replication in *Halobacterium* sp. NRC-1 targeted the 10 *orc* gene alleles, eight of which were coded by pNRC100 and 200 (Berquist et al. 2007). Seven of the ten *orc/cdc6* genes in the genome are found on the pNRC replicons, but only one extrachromosomal gene, *orc2*, on pNRC200 was found to be essential. The *orc2* gene was classified as essential on the basis of statistical analysis and complementable in trans by demonstrating the ability to isolate genomic knockouts in the presence of a plasmid copy of the gene. Of the three replicative-type DNA polymerase genes in the genome, only the pNRC200-encoded

B family polymerase, *polB2*, was nonessential (Berquist et al. 2007). Of the multigene families in *Halobacterium*, the expansion of the *orc* gene family in NRC-1 is exceptional, being the largest number of alleles for any gene, with the exception of some transposases.

Another genetic study of transcription factor genes targeted the six *tbp* and seven *tfb* genes in the *Halobacterium* sp. NRC-1 genome (Coker and DasSarma 2007). Of the five *tbp* genes located on the pNRC replicons, two (*tbpA* and *B*) were found to be essential, while three (*tbpC*, *D*, and *F*) were nonessential. For the two *tfb* genes on pNRC200, one (*tfbE*) was essential and one was non-essential (*tfbC*) (Coker and DasSarma 2007). The results reported in another study were slightly different, although no documentation was provided in the second report (Facciotti et al. 2007).

3 *Haloferax volcanii*

The *Haloferax volcanii* genome was first mapped using an overlapping cosmid library, which provided the overall genome structure (Charlebois et al. 1991), and its complete genome sequence has recently been reported on a genome database (Schneider et al. 2006; J. Eisen et al., unpublished). Compared to *Halobacterium* sp. NRC-1, *H. volcanii* contains a relatively large genome of 4,012,900 bp, with three large plasmids (85–636 kb) and one miniplasmid (only 6,359 bp) (Table 2). The two megaplasmids, pHV4 and pHV3, have G + C composition very similar to the chromosome (62–67%), while the two smaller plasmids have significantly lower G + C content. The large plasmid, pHV1, has a G + C composition of 55.5%. The *H. volcanii* miniplasmid, pHV2, has been extensively used as a cloning vehicle (Charlebois et al. 1987; Lam and Doolittle 1989) and to facilitate genetic studies of this organism (Allers and Mevarech 2005; Berquist et al. 2006). Although preliminary whole-genome analysis of genes in *H. volcanii* was originally conducted by hybridization analysis, the locations of relatively few genes were established (Cohen et al. 1992). For the present discussion, we used derived protein sequences from the genes predicted on the megaplasmids available on the UCSC database (April 2007 release; Schneider et al. 2006) to search for homologs in the NCBI nonredundant database and assign their functions (Table 3, and DasSarma et al. unpublished). In addition to these investigations, a single genetic study aimed at the replication regions of *H. volcanii* megaplasmids has also been reported (Norais et al. 2007).

3.1 *pHV4 Megaplasmid*

The *H. volcanii* plasmid, pHV4, is 635,786 bp in size, which makes it the largest megaplasmid known among haloarchaea, containing over 600 genes. Interestingly, it has four *orc/cdc6* genes, two of which are in the Orc5 family found on pNRC200

and two are in the Orc8 family (Table 4). pHV4 also encodes two ATPase-family partitioning proteins, SojE and ParA family homologs and one DNA polymerase B family protein that is similar to the nonessential homolog coded by pNRC200. Interestingly, like pNRC200, a divergently transcribed *orc/cdc6* gene is present upstream to the pHV4 *polB* gene.

Table 4 Distribution of haloarchaeal Orc/Cdc6 family proteins

Organism and replicon	Haloarchaeal Orc5	Archaeal Orc6	Archaeal Orc7	Haloarchaeal Orc8	Haloarchaeal Orc9
<i>Halobacterium</i> sp. NRC-1 chromosome	–	Orc6	Orc7	Orc8 and 10	–
<i>Halobacterium</i> sp. NRC-1 pNRC100	–	–	–	–	Orc9
<i>Halobacterium</i> sp. NRC-1 pNRC200	Orc2, 3, 4, 5	–	–	–	Orc1
<i>Haloferax volcanii</i> chromosome	Orc5 and 11	Orc9	Orc1	Orc2 and 4	Orc14
<i>Haloferax volcanii</i> pHV1	Orc8	–	–	Orc10 ^a	–
<i>Haloferax volcanii</i> pHV3	Orc6	–	–	–	–
<i>Haloferax volcanii</i> pHV4	Orc7 and 13	–	–	Orc3 ^b	Orc12
<i>Haloarcula marismortui</i> chromosome I	Cdc6i	Cdc6c and j	Cdc6d	Cdc6f, g, h	Cdc6e
<i>Haloarcula marismortui</i> pNG100	Cdc6q	–	–	–	–
<i>Haloarcula marismortui</i> pNG200	–	–	–	–	–
<i>Haloarcula marismortui</i> pNG300	–	–	–	–	Cdc6p
<i>Haloarcula marismortui</i> pNG400	–	–	–	–	–
<i>Haloarcula marismortui</i> pNG500	Cdc6n	–	–	Cdc6o	–
<i>Haloarcula marismortui</i> pNG600	Cdc6m	–	–	–	Cdc6l
<i>Haloarcula marismortui</i> chromosome II	–	–	–	Cdc6a and b	–
<i>Haloarcula marismortui</i> pNG700	–	–	–	Cdc6k	–
<i>Natronomonas pharaonis</i> chromosome	–	Cdc6-2	Cdc6-1	Cdc6-3	Cdc6-4
<i>Natronomonas pharaonis</i> pL23	–	–	–	–	–
<i>Natronomonas pharaonis</i> pL131	Cdc6-5	–	–	–	–
<i>Haloquadratum walsbyi</i> chromosome	Cdc6-4, 5, 6, 7	Cdc6-2 and 9	Cdc6-1	–	Cdc6-3 and 8
<i>Haloquadratum walsbyi</i> pL47	–	–	–	–	–

^a Orc10 may also be coded on pHV1

^b Orc3 may instead be encoded on the chromosome

Among genes likely to be essential, plasmid pHV4 codes for a tryptophanyl-tRNA synthetase in *H. volcanii*. This is one of two tryptophanyl-tRNA synthetase genes in the *H. volcanii* genome, and the only aminoacyl-tRNA synthetase coded by *H. volcanii* megaplasmids. The only other such example among haloarchaeal replicons is an arginyl-tRNA synthetase on the *Halobacterium* megaplasmid, pNRC200, shown to be a case of a recent lateral gene transfer (Berquist et al, 2005). In addition, a translation initiation factor is coded by pHV4, similar to others coded on pHV3 and the *H. marismortui* megaplasmid pNG700. Interestingly, however, all of the ribosomal proteins in *H. volcanii* are coded on the chromosome.

The pHV4 replicon contains the largest number of transposases (36) of any haloarchaeal plasmid, with the exception of the *Halobacterium* pNRC200 replicon. pHV4 also contains three integrases, suggestive of integrated prophages or cryptic phages, and CRISPR sequences, involved with acquired resistance against viruses in prokaryotes (Kunin et al. 2007) or possibly in cell division (Mojica et al. 2005). Four CRISPR-associated sequences likely involved in the propagation and functioning of CRISPRs were also found on this megaplasmid. The only other haloarchaeal megaplasmid containing CRISPRs was the *H. marismortui* pNG400 replicon. A cell division protein gene encoding a FtsZ GTPase is present on pHV4, as was also reported on pNRC200.

A conjugation-related protein was found encoded on pHV4 and is interesting in light of conjugative genetic transfer reported for this organism (Tchelet and Mevarech 1994). Several restriction-modification enzyme genes are also present on pHV4, including a type II restriction enzyme and two DNA modification methylases. Genes coding two Dam methylases and an excinuclease, as well as an UvrC subunit and two RNaseH proteins are present. Surprisingly, pHV4 contains six helicases (compared to two on each of the pNRC plasmids), including two DEAD/H box helicases, associated in many processes involving RNA (Tanner and Linder 2001). Seven stress protein genes are present, including genes coding Hsp4, universal stress response proteins, and a superoxide dismutase (product of the *sod* gene). Interestingly, this is the only example of a *sod* gene on a haloarchaeal megaplasmid.

Among over 50 transporters coded on pHV4 is an ion transport protein similar to ArsB, a putative arsenite transport protein on the chromosome of *Halobacterium* sp. NRC-1. Although transcriptional studies showed that the *arsB* gene of NRC-1 is regulated by arsenic, its knockout did not result in arsenic sensitivity, suggesting an alternate function (Wang et al. 2004). A similar gene is also present on the chromosome of *N. pharaonis*. Genes for three drug resistance proteins and a homolog of the multiple antibiotic resistance regulator, MarR, were also identified on pHV4.

3.2 *pHV3 Megaplasmid*

The second largest megaplasmid in *H. volcanii*, pHV3, 437,906 bp, also the second largest among all the characterized haloarchaeal megaplasmids, codes for over 350 proteins. In contrast to pHV4 and pNRC200, which are rich in IS elements, pHV3

contains only three transposase genes associated with these elements (two ISH3 and one ISH1). It does not code any DNA polymerases, but does contain a single *orc/cdc6* gene similar to the Orc5 family found on pNRC200, and a gene likely involved in partitioning (*sojC2*). There are also single *thp* and *tfb* genes, the only examples of these transcription factor genes found on any of the *H. volcanii* megaplasms. However, the essentiality of these genes has not been tested. Three proteins involved in translation are coded on pHV3, translation initiation factors IF-2 g-subunit and SUI1 (which contains an initiation domain), and a putative initiation inhibitor.

Among other interesting and potentially important proteins coded by pHV3 are a chaperonin-like protein, endoribonuclease (L-PSP), two DNA repair proteins (UvrC and Rad50), and two restriction modification proteins. Two drug-related protein genes are found on this plasmid, a Major Facilitator Superfamily permease and another putative drug resistance protein. Also, there are two nitrate reductase subunits (catalytic and electron transfer subunits), suggesting their involvement in anaerobic nitrate reduction reported for *H. volcanii* (Wanner and Soppa 1999), a cytochrome b_6 , a Rieske iron-sulphur protein, and two of the four subunits of the cytochrome b_6f complex (plastoquinol-plastocyanin reductase) encoded on pHV3. Finally, a large cluster of cobalamin biosynthesis genes is present on this plasmid. In the other halophiles, the cobalamin biosynthesis genes appear only to reside on the chromosome.

3.3 *pHV1*

The large plasmid of *H. volcanii*, pHV1 is 85,092 bp and encodes 89 predicted proteins, including 46 hypothetical proteins. Two *orc* genes are found on this large replicon, coding an Orc5 family homolog and an Orc8 family homolog (found on the chromosome in NRC-1). It is rich in transposable elements, with many transposase genes (including three coded by ISH4, two by ISH8, one by ISH4, and five by ISH3 elements). There is only one partitioning gene on this plasmid and it codes for the ParA ATPase domain. Other potentially important genes include those coding a helicase, bacterial conjugation protein homolog, nitrate reductase, protein kinase, a putative nickel transport protein, and four transporters predicted to be involved in peptide transport.

3.4 *pHV Replication Origins*

A study of autonomous replication ability of *orc/cdc6* gene regions in *H. volcanii* suggested that some of the genes were located near replication origins. Autonomously replicating sequences were isolated with each of the large plasmids, upstream of *orc10* on pHV1, *orc6* on pHV3, and *orc3* on pHV4 (Norais et al. 2007).

Hybridization analysis of the autonomously replicating sequence associated with pHV1 suggested that this region is also found on pHV4, indicating an additional level of complexity. Similarly, sequence analysis of the *orc3* gene region showed it to be present on pHV4, but hybridization analysis suggested that it may instead be associated with the chromosome. These complications may reflect either strain-specific differences or rearrangements that have occurred recently during subculturing. Despite the organizational complexity of these autonomously replicating sequences, replication initiation point mapping suggested that they are indeed functional origins in vivo. These results supported the idea that the *orc/cdc6* genes mentioned earlier and their upstream regions function in plasmid replication initiation.

4 *Haloarcula marismortui*

In contrast to the *Halobacterium* sp. NRC-1 and *H. volcanii* genomes, characterization of which included extensive genomic mapping, studies of the relatively large, 4,274,642 bp genome of *Haloarcula marismortui* have been conducted solely by whole genome shotgun sequencing (Table 2) (Baliga et al. 2004). The *H. marismortui* genome is reported to be organized into nine replicons; however, neither pulsed-field nor restriction mapping information is available to corroborate these findings. Two of the replicons were classified as chromosomes, including the largest replicon (3,131,724 bp) designated chromosome I with a 62% G + C content, and the third largest replicon (288,050 bp), chromosome II, with G + C content of 57%. Of the three replicons reported to be megaplasmids, pNG700 (410,554 bp) is bigger than chromosome II, with a G + C content of 59%, while pNG600 (155,300 bp and 58% G + C content) and pNG500 (132,678 bp and 54% G + C content) are smaller. The non-megaplasmid replicons of *H. marismortui* are still quite large, pNG400 being 50,060 bp and with a G + C content of 57%, pNG300 being 39,521 bp, with a G + C content of 60%, and both pNG200 and pNG100 approximately 33,000 bp, with 56% and 54% G + C content, respectively. Gene content analysis of the replicons of *H. marismortui* suggested that along with the main chromosome, pNG600 and pNG700, as well as chromosome II, may be essential replicons (Table 3). However, in the single genetic test published thus far, essentiality was not corroborated for chromosome II (Tu et al. 2005).

4.1 *pNG700 Megaplasmid*

pNRC700 is chromosome-like, based on its large size and relatively low G + C content. Moreover, this megaplasmid carries only two transposases and a single phage integrase, suggesting a more stable replicon compared to pHV4 and pNRC100/200. Phylogenetic analysis of archaeal Orc/Cdc6 homologs places the protein encoded by the single *orc/cdc6* gene on pNG700 in a clade with *Halobacterium* sp. NRC-1 Orc10 family, the chromosomally encoded essential *orc/*

cdc6 gene (Berquist and DasSarma 2008). A gene annotated as a translational inhibitor coded on pNG700 belongs to a family of proteins that inhibits mRNA translation via endoribonuclease activity. Genes for a 5S rRNA and the archaeal specific small subunit S17e protein are also present, as are thioredoxin, and two copies of thioredoxin reductase genes. There are two copies of endoribonucleases, an endonuclease III, and a RNaseH gene, both potentially key nucleases. A gene encoding the NosL protein, a nitrous oxide reductase accessory protein, is found on this megaplasmid as well. Structural studies of the copper (I) binding site suggest NosL functions as a metallochaperone and may aid in copper delivery during biosynthesis of the catalytic site of nitrous oxide reductase under copper limiting conditions (Taubner et al. 2006). The only copies of several genes essential for folate metabolism, 5,10-methylenetetrahydrofolate reductase, formyltetrahydrofolate synthetase, and formimidoyletetrahydrofolate cyclodeaminase, are also found on pNG700. There are multiple genes involved in stress response, including the thermosome small subunit, and the only PpiC-type peptidyl-prolyl *cis-trans* isomerase, assisting in protein folding through chaperone-like activity (Ou et al. 2001).

4.2 Chromosome II

Although chromosome II is much smaller than pNG700, it was designated as a minichromosome based on the presence of an rRNA operon (*rrnB*) encoding the 5S, 16S, and 23S ribosomal RNA (Baliga et al. 2004). A subsequent attempt at deletion of this operon has been successful, showing this operon and possibly the replicon itself to be nonessential (Tu et al. 2005). However, a tRNA (*trn49*) is also coded on chromosome II, a tRNA^{Lys}, recognizing the AAA codon, as is *gatA*, an amidase for pretranslational amino acid modification to form Asn- and Gln-tRNA (Namgoong et al. 2007). The main chromosome also harbors a *gatA* gene, forming a possible operon with *gatC* (similar to *Halobacterium* sp. NRC-1). Since the heterotrimeric GatCAB is needed for transamidation and more specifically the A and C subunits need to be coexpressed for full amidotransferase activity (Curnow et al. 1997), the chromosome II *gatA* gene may not be essential to cells. The replicon also harbors two copies of the *orc/cdc6* gene, both phylogenetically similar to the essential Orc10 family of *Halobacterium* sp. NRC-1 (Berquist et al. 2007). Although these genes are presumably important in initiation of DNA replication in *Halobacterium* sp. NRC-1, no parallel studies are available for *H. marismortui*. Of the nine different transcription TFB factor genes, a lone copy, *tfbC*, is on chromosome II. This gene may or may not be essential as studies of the multiple copies of *tfb* genes in *Halobacterium* sp. NRC-1 have shown that not all are essential (Coker and DasSarma 2007; Facciotti et al. 2007).

Several enzymes, succinate-semialdehyde dehydrogenase, acetyl-CoA acetyltransferase, and GMP synthase, important for metabolism are only encoded on chromosome II. There are also multiple genes found on chromosome II that are involved in cell division and chromosome partitioning, including several *parA*-like genes.

Interestingly, genes involved in the synthesis of penicillins, penicillin acylase and amidase, are also present. Chromosome II also has some characteristics more consistent with a plasmid designation, for example, there are eight transposase genes.

4.3 *pNG600 Megaplasmid*

Like the larger replicons, pNG600 also encodes several important and possibly essential genes. It harbors the only *H. marismortui* gene encoding a DEAD/H family helicase, as well as a family B DNA polymerase, with another family B DNA polymerase found on the chromosome. This is similar to *Halobacterium* sp. NRC-1, where two separate family B DNA polymerases are also found, one on the chromosome (*polB1*) and another (*polB2*) on pNRC200. Furthermore, these genes have high amino acid identities, 46% comparing the chromosomally encoded and essential protein and 81% for the nonessential megaplasmid protein (Berquist et al. 2007). However, similar studies are needed to determine which of the genes are essential in *H. marismortui*. The *polB2* gene is encoded just downstream of one of the two *orc/cdc6* genes on pNG600. Unlike the chromosome II and pNG700 replicons, this *orc/cdc6* gene codes a protein that is in the *Halobacterium* sp. NRC-1 Orc5 family (all of which are encoded on pNRC200), while the second *orc/cdc6* gene found on this replicon codes a protein that is similar to Orc9 family proteins (Table 4) (Berquist and DasSarma 2008). Two transcription factor genes, *tfbE* and *tfbG*, a tRNA^{Glu} gene, a partitioning protein gene, and a large number of transcriptional regulatory genes are found on this megaplasmid. Both the large and small subunits of exonuclease VII along with an Mrr-like restriction system protein are present. Two genes for bacterial-type conjugation are present; however, there is no drug resistance gene on the replicon. Surprisingly, there are no transposases, despite the relatively low G + C content. A single putative stress response gene is present, in contrast to the large number found on pNG700. A gene encoding thioredoxin as well as a DsbA-like thioredoxin are present, but unlike in pNRC100/200, there are no thioredoxin reductase genes. The lone copy of aconitase, an enzyme in the citric acid cycle is encoded on pNG600. Over a dozen genes are present for processing of a variety of heavy metals (cadmium, copper, mercury, and zinc), including multiple transport proteins, suggesting that the organism has to contend with toxic concentrations of these heavy metals in its natural habitat.

4.4 *pNG500 Megaplasmid*

The smallest of the *H. marismortui* megaplasmids, pNG500, has the largest number of transposases (22) and a relatively low G + C content (54.5%). Only

51% of the replicon is coding and approximately 20% of this region codes transposases. The remainder of the coding region harbors multiple transcriptional regulators, two of which are Spo0A activation inhibitors. Spo0A can either activate or repress transcription to initiate sporulation, and Spo0A activation inhibitors such as Soj are thought to link chromosome partitioning status and sporulation (Hilbert and Piggot 2004). However, so far there have been no reports of sporulation among the haloarchaea. Also coded on this replicon are two Orc/Cdc6 homologs, one in the Orc5 family on pNG600, and one similar to the Orc8 family, falling in the same clade as the Orc proteins encoded by chromosome II and pNG700. A single stress response gene, as in pNG600, and three genes involved in conjugation, including a relaxase and a gene encoding TraD, are found on this replicon. In bacteria, TraD is proposed to be the coupling protein that links the relaxosome to the transferosome for F-plasmid conjugative transfer (Sastre et al. 1998). However, unlike *H. volcanii*, conjugation has not been reported for *H. marismortui*.

4.5 Other *H. marismortui* Plasmids

The four remaining replicons of *H. marismortui* are not classified as megaplasmiids but rather large plasmids based on their size, ranging from 33 to 50 kb. Each codes for interesting gene products, although none appears to be essential to the cell. pNG400 encodes three transposases, three transcriptional regulators, a ParA chromosome partitioning protein, and a DNA binding protein. Unlike the larger replicons in *H. marismortui*, pNG400 has a pNRC-like *repH* gene for DNA replication. pNG400 also encodes four CRISPR-associated proteins, as does pHV4, as well as a nearby putative helicase.

Interestingly, pNG300 is the only replicon that, like pNRC100 and pNRC200, has both an *orc/cdc6* homolog, coding a protein similar to the Orc9 family of proteins encoded on pNRC100 and pNRC200, as well as a *repH* gene for replication. Unlike the other *H. marismortui* plasmids, pNG300 also encodes a replicative helicase, Mcm, in addition to the two found encoded on the main chromosome. This replicon also encodes a single transposase, a plasmid partitioning protein, a transcriptional regulator, and a DNA binding protein.

The 33.5 kb replicon pNG200 contains 42 genes. Among them is a *repH*-like gene, three different transcriptional regulatory genes, two chromosome partitioning genes, a helicase gene, and a gene encoding a putative DNA binding protein. Uncharacteristic of a smaller plasmid, pNG200 also encodes a large ribosomal subunit L36-like protein and lacks any transposases. The smallest replicon of *H. marismortui*, pNG100, contains less than 40 genes, including an Orc/Cdc6 protein (similar to the Orc5 family), a phage integrase, a transcriptional regulator, and a helicase. There is also a gene encoding a flagellin A-like protein, although no flagella have been reported for *H. marismortui*.

5 *Natronomonas pharaonis*

The genome of the alkaliphilic haloarchaeon, *Natronomonas pharaonis*, is more similar in size (2.8 Mbp) and arrangement (two plasmids) (Tables 2 and 3) to *Halobacterium* sp. NRC-1 than to the more complex *H. marismortui* and *H. volcanii* genomes (Falb et al. 2005). The larger plasmid, pL131, is 130,989 bp and encodes 125 putative proteins, of which 53 are hypothetical and 12% have been shown to be present by proteomic analysis (Konstantinidis et al. 2007). There are many IS elements, at least 10 based on the number of transposases, including one each of ISH4, ISH6, and ISH9, as well as five copies of another element similar to IS1341 (DasSarma et al. unpublished) distributed throughout the plasmid. Other genes of interest include those coding an Orc/Cdc6 homolog and a ParA domain protein, as well as two DNA-binding proteins and six transcriptional regulators. As in the *Halobacterium* sp. NRC-1 plasmid pNRC100, genes coding a probable arsenite/heavy metal transporter and a probable arsenate reductase are present. Two Mrr restriction-modification proteins are coded by pL131 as are three stress proteins, including two thermosome subunits. There are 17 transporter as well as two signal transducer genes distributed over the plasmid. Genes for the two subunits of 3-oxoacid CoA-transferase, involved in transferring a CoA moiety (Jacob et al. 1997), and both subunits of *n*-methylhydantoinase involved in *N*-methylhydantoinase (ATP-hydrolyzing) activity are also present.

The smaller *N. pharaonis* plasmid, pL23, is only 23 kb and encodes just 36 predicted proteins, of which 22 are hypothetical. Only a single protein from this plasmid was identified by proteomic analysis of this organism (Konstantinidis et al. 2007). A copy of the plasmid was also found integrated in the chromosome and was found to contain a 13 kb insert (Falb et al. 2005). Among the identified proteins, there is one integrase, one drug and one metal resistance transporter, and a regulator. As in the case of the *H. walsbyi* plasmid, there is also a gene coding for a probable cell surface glycoprotein as well as a viral or phage gene, which is homologous to viral protease (Bahr and Darai 2001) and a gene coding a protein with an integrin-like domain. Interestingly, pL23 contains a spermidine synthase, which is in contrast to *Halobacterium* sp. NRC-1, where spermidine uptake genes are located on the chromosome. Spermidine synthase catalyzes the transfer of aminopropyl groups from decarboxylated SAM to putrescine to form spermidine, which is important for outer membrane permeability in response to pH stress in gram-negative bacteria (Kozbial and Mushegian 2005).

6 *Haloquadratum walsbyi*

Among the sequenced haloarchaea, *H. walsbyi* is the only one without any megaplasmid (Table 2). The genome of this rather unusual square-shaped, flat organism also has the lowest G + C content (48%) and lowest coding density (Bolhuis et al.

2006). The large plasmid, pL47, is 46,867 bp and encodes 36 putative proteins, 24 of which are hypothetical (Table 3). The remaining 12 code for three phage-like proteins (encoding two primases encoded next to one another and one integrase), a partitioning ParA protein ATPase, three restriction enzyme-related proteins, and one each of a superfamily II helicase and a resolvase. In addition, a cell surface glycoprotein, a ribonuclease H1, and a SCP-like extracellular protein potentially involved in signaling are coded by pL47.

7 Evolution of Haloarchaeal Megaplasmiids

The abundance of megaplasmiids in haloarchaeal genomes is remarkable in light of their relative rarity among archaea generally. Four of the five sequenced haloarchaeal genomes harbor megaplasmiids while none of the nearly 50 other nonhalophilic archaea sequenced thus far harbor them. Studies of the megaplasmiids in *Halobacterium* sp. NRC-1 and similar plasmids in other *Halobacterium* strains (Pfeifer and Ghahraman 1993; Oesterhelt, personal communication) suggest that many haloarchaeal megaplasmiids are highly dynamic and rapidly evolving. In all cases examined, plasmid dynamics are the result of recombinational activities promoted by an abundance of IS elements and other repeats within these replicons (DasSarma and Arora 1997; Ng et al. 1998; DasSarma 2004).

The mechanism of evolution of haloarchaeal megaplasmiids has been proposed only for *Halobacterium* sp. NRC-1 (Fig. 3b), involving a series of fusions of smaller replicons, each carrying a *repH*-like gene. The replicon fusions may have resulted from the intermolecular transposition of IS elements (e.g., depicted in steps 1 and 2, in Fig. 3b) and/or by site-specific or homologous recombination between IS elements (step 4). Additional events proposed in pNRC100 evolution were recombinational exchanges leading to acquisition of chromosomal genes (step 3), and duplication of the acquired chromosomal region (step 5) to generate the large IRs via unequal crossing over. Recombination between the IRs was demonstrated experimentally to show the occurrence of inversion isomers (Ng et al. 1991) and recombination between IS elements were also detected to form minor circular DNAs (Ng and DasSarma 1991; Ng et al. 1994). Similar recombinational events likely led to replacement of a portion of the large single copy region of pNRC100 with additional genes to form pNRC200.

The acquisition of many important genes on megaplasmiids like pNRC100 and especially pNRC200, including the *argS* gene coding arginyl-tRNA synthetase likely necessary for protein synthesis (Ng et al. 1998) and the origin recognition complex *orc2* gene, experimentally shown to be essential (Berquist et al. 2007), indicates that these replicons have achieved a status as essential replicons in the cell. Presence of both an integrated copy and an autonomous version of a plasmid have also been observed in the case of *N. pharaonis* plasmid pL23 (Falb et al. 2005), providing direct evidence for the occurrence of replicon fusions in haloarchaeal genomes. A similar event has recently been hypothesized in the genomes of

some thermophilic archaeal species as well (Robinson and Bell 2007). Plasmid integration and imprecise excision from the chromosome would allow the transfer of genes between replicons, either from chromosomes to plasmids or vice versa, and would lead to scrambling of the genome over time.

Not surprisingly, a number of important genes have been found on most if not all of the sequenced megaplasmids, and at least one, the 288 kb replicon of *H. marismortui*, which contains an rRNA operon, led the authors to designate it as chromosome II (Baliga et al. 2004). However, since *H. marismortui* contains several more rRNA operons on the large chromosome, and the chromosome II ribosomal RNA genes were shown to be nonessential (Tu et al. 2005), chromosomal designation for this replicon is in doubt. Instead, this replicon may have recently acquired the rRNA operon through integration and excision from the chromosome. Such gene exchanges may be quite common among haloarchaeal plasmids, and perhaps many if not most of the resident haloarchaeal plasmids have integrated into the chromosome and excised during their evolutionary history, leading to diverse gene content. Therefore, our current view from genomic mapping and sequencing may be limited to snapshots of the dynamic genomic architecture in haloarchaea.

The reasons for the finding of such complex and dynamic genome architecture among haloarchaea are intriguing. The recombinational activities of IS elements and other repeated elements resulting in the capture of critical genes would promote the maintenance of multiple replicons. The lack of incompatibility between related replicons may result from the presence of multiple origins of replication of different compatibility groups, including some eukaryotic types. Although megaplasmid maintenance and compatibility have not been studied in any detail in haloarchaea, examples of multiple replicase genes and origin recognition complex genes, as well as ParA/Soj genes among the sequenced megaplasmids suggests that this is indeed the case. The duplication of a portion of a megaplasmid, as has occurred for pNRC100 and pNRC200, would serve to stabilize the region within the repeats. It is also possible that megaplasmids may assist in capturing new genes from the environment and contribute to lateral gene transfers, but studies published thus far do not provide many examples of such events. The single clear example is the bacterial type *argS* gene in *Halobacterium* sp. NRC-1, which was likely laterally acquired on pNRC200 after the divergence of NRC-1 from other haloarchaea (Berquist et al. 2005). It is conceivable that additional detailed studies of the recently sequenced haloarchaeal megaplasmids may uncover additional cases of laterally acquired genes.

The existence of megaplasmid and minichromosome replicons harboring multiple essential genes and with the capability to recombine with a large main chromosome and other resident replicons suggests a highly dynamic state for haloarchaeal genomes. It is possible that the haloarchaeal genomic condition may be described as one of a competitive dynamic equilibrium between essential replicons in the genome, maintained through recombination between replicons mediated by regions of DNA homology. It is conceivable that such a genomic condition may be a reflection of the dynamic hypersaline environments facing many haloarchaea (DasSarma et al. 2006), in contrast with many other archaea that inhabit relatively unchanging,

albeit extreme environments. If so, it is likely that the heterogeneity of megaplasmiids and minichromosomes among haloarchaea is a testament to its successful adaptation to a highly dynamic environment. Given these findings for the haloarchaea, competition between replicons may indeed be a general phenomenon in evolution, and is likely to play an important role in shaping the genomes of other species, and in the long-term, contributing to the evolution of new chromosomes and megaplasmiids.

Acknowledgments Research in the authors' laboratory is supported by National Science Foundation grant MCB-0450695 and National Aeronautics and Space Administration grant NNX08AT70G. M. Capes was supported by a Meyerhoff Graduate Fellowship.

References

- Allers T, Mevarech M (2005) Archaeal genetics – the third way. *Nat Rev Genet* 6:58–73
- Arner ES, Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. *Eur J Biochem* 267:6102–6109
- Bahr U, Darai G (2001) Analysis and characterization of the complete genome of tupaia (tree shrew) herpesvirus. *J Virol* 75:4854–4870
- Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, Deutsch EW, Shannon P, Chiu Y, Weng RS, Gan RR, Hung P, Date SV, Marcotte E, Hood L, Ng WV (2004) Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* 14:2221–2234
- Baxter BK, Litchfield CD, Sowers K, Griffith JD, DasSarma PA, DasSarma S (2005) Microbial diversity of Great Salt Lake. In: Gunde-Cimerman N, Oren A, Plemenitaš A (eds) Adaptation to life at high salt concentrations in archaea, bacteria, and eukarya, vol 9. Springer, Dordrecht, Netherlands, pp 9–26
- Berquist B, DasSarma S (2008) DNA replication in archaea. In: Blum P (ed) Archaea: new models for prokaryotic biology, Horizon Press, Norwich, UK, pp 95–120
- Berquist BR, Soneja J, DasSarma S (2005) Comparative genomic survey of information transfer systems in two diverse extremely halophilic archaea, *Halobacterium* sp. strain NRC_1 and *Haloarcula marismortui*. In: Gunde-Cimerman N, Oren A, Plemenitaš A (eds) Adaptation to life at high salt concentrations in archaea, bacteria, and eukarya, vol 9. Springer, Dordrecht, Netherlands, pp 148–182
- Berquist BR, Müller JA, DasSarma S (2006) Genetic systems for halophilic archaea. In: Oren A, Rainey F (eds) Extremophiles, vol 35. Elsevier, Amsterdam, pp 649–680
- Berquist BR, DasSarma P, DasSarma S (2007) Essential and non-essential DNA replication genes in the model halophilic Archaeon, *Halobacterium* sp. NRC-1. *BMC Genet* 8:31
- Bolhuis H, Palm P, Wende A, Falb M, Rampp M, Rodriguez-Valera F, Pfeiffer F, Oesterhelt D (2006) The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* 7:169
- Charlebois RL, Lam WL, Cline SW, Doolittle WF (1987) Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaeobacterium. *Proc Natl Acad Sci USA* 84:8530–8534
- Charlebois RL, Schalkwyk LC, Hofman JD, Doolittle WF (1991) Detailed physical map and set of overlapping clones covering the genome of the archaeobacterium *Haloferax volcanii* DS2. *J Mol Biol* 222:509–524
- Cohen A, Lam WL, Charlebois RL, Doolittle WF, Schalkwyk LC (1992) Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea. *Proc Natl Acad Sci USA* 89:1602–1606

- Coker JA, DasSarma S (2007) Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet* 8:61
- Curnow AW, Hong K, Yuan R, Kim S, Martins O, Winkler W, Henkin TM, Soll D (1997) Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci USA* 94:11819–11826
- DasSarma S (1989) Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. *Can J Microbiol* 35:65–72
- DasSarma S (2004) Genome sequence of an extremely halophilic archaeon. In: Fraser CM, Read T, Nelson KE (eds) *Microbial genomes*, Humana, Totowa, NJ, pp 383–399
- DasSarma S, Arora P (1997) Genetic analysis of the gas vesicle gene cluster in haloarchaea. *FEMS Microbiol Lett* 153:1–10
- DasSarma S, Damerval T, Jones JG, Tandeau de Marsac N (1987) A plasmid-encoded gas vesicle protein gene in a halophilic archaebacterium. *Mol Microbiol* 1:365–370
- DasSarma S, Halladay JT, Jones JG, Donovan JW, Giannasca PJ, Tandeau de Marsac NT (1988) High-frequency mutations in a plasmid-encoded gas vesicle gene in *Halobacterium halobium*. *Proc Natl Acad Sci USA* 85:6861–6865
- DasSarma S, Arora P, Lin F, Molinari E, Yin LR-S (1994) Wild-type gas vesicle formation requires at least ten genes in the *gvp* gene cluster of *Halobacterium halobium* plasmid pNRC100. *J Bacteriol* 176:7646–7652
- DasSarma S, Berquist BR, Coker JA, DasSarma P, Müller JA (2006) Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1. *Saline Systems* 2:3
- David M, Daveran ML, Batut J, Dedieu A, Domergue O, Ghai J, Hertig C, Boistard P, Kahn D (1988) Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. *Cell* 54:671–683
- Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, Bonneau R, Shannon P, Srivastava A, Donohoe SM, Hood LE, Baliga NS (2007) General transcription factor specified global gene regulation in archaea. *Proc Natl Acad Sci USA* 104:4630–4635
- Falb M, Pfeiffer F, Palm P, Rodewald K, Hickmann V, Tittor J, Oesterheld D (2005) Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res* 15:1336–1343
- Fraser CM, Read T, Nelson KE (eds) (2004) *Microbial genomes*. Humana, Totowa, NJ
- Gavin KA, Hidaka M, Stillman, B (1995) Conserved initiator proteins in eukaryotes. *Science* 270:1667–1671
- Ghosh TK, Mickelson DJ, Fink J, Solberg JC, Inglefield JR, Hook D, Gupta SK, Gibson S, Alkan SS (2006) Toll-like receptor (TLR) 2–9 agonists-induced cytokines and chemokines: I. Comparison with T cell receptor-induced responses. *Cell Immunol* 243:48–57
- Gray MW (1999) Evolution of organellar genomes. *Curr Opin Genet Dev* 9:678–687
- Halladay JT, Ng WL, DasSarma S (1992) Genetic transformation of a halophilic archaebacterium with a gas vesicle gene cluster restores its ability to float. *Gene* 119:131–136
- Halladay JT, Jones JG, Lin F, MacDonald AB, DasSarma S (1993) The rightward gas vesicle operon in *Halobacterium* plasmid pNRC100: identification of the *gvpA* and *gvpC* gene products by use of antibody probes and genetic analysis of the region downstream of *gvpC*. *J Bacteriol* 175:684–692
- Hilbert DW, Piggot PJ (2004) Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol Mol Biol Rev* 68:234–262
- Jacob U, Mack M, Clausen T, Huber R, Buckel W, Messerschmidt A (1997) Glutaconate CoA-transferase from *Acidaminococcus fermentans*: the crystal structure reveals homology with other CoA-transferases. *Structure* 5:415–426
- Jones JG, Hackett NR, Halladay JT, Scothorn DJ, Yang C-F, Ng W-L, DasSarma S (1989) Analysis of insertion mutants reveals two new genes in the pNRC100 gas vesicle gene cluster of *Halobacterium halobium*. *Nucleic Acids Res* 17:7785–7793
- Jones JG, Young DC, DasSarma, S (1991) Structure and organization of the gas vesicle gene cluster on the *Halobacterium halobium* plasmid pNRC100. *Gene* 102:117–122
- Joshi JG, Guild WR, Handler P (1963) The presence of two species of DNA in some halobacteria. *J Mol Biol* 6:34–38

- Konstantinidis K, Tebbe A, Klein C, Scheffer B, Aivaliotis M, Bisle B, Falb M, Pfeiffer F, Siedler F, Oesterheld D (2007) Genome-wide proteomics of *Natronomonas pharaonis*. *J Proteome Res* 6:185–193
- Kozbial PZ, Mushegian AR (2005) Natural history of *S*-adenosylmethionine-binding proteins. *BMC Struct Biol* 5:19
- Kunin V, Sorek R, Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 8:R61
- Lam WL, Doolittle WF (1989) Shuttle vectors for the archaeobacterium *Halobacterium volcanii*. *Proc Natl Acad Sci USA* 86:5478–5482
- Lois AF, Ditta GS, Helinski DR (1993) The oxygen sensor FixL of *Rhizobium meliloti* is a membrane protein containing four possible transmembrane segments. *J Bacteriol* 175:1103–1109
- Miller MJ, Gennis RB (1983) The purification and characterization of the cytochrome d terminal oxidase complex of the *Escherichia coli* aerobic respiratory chain. *J Biol Chem* 258:9159–9165
- Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60:174–182
- Moore RL, McCarthy BJ (1969) Characterization of the deoxyribonucleic acid of various strains of halophilic bacteria. *J Bacteriol* 99:248–254
- Namgoong S, Sheppard K, Sherrer RL, Soll D (2007) Co-evolution of the archaeal tRNA-dependent amidotransferase GatCAB with tRNA (Asn). *FEBS Lett* 581:309–314
- Ng W-L, DasSarma S (1991) Physical and genetic mapping of the unstable gas vesicle plasmid in *Halobacterium halobium* NRC-1. In: Rodriguez-Valera F (ed) Genetic and applied aspects of halophilic microorganisms, vol 201. Plenum, New York, pp 305–311
- Ng W-L, DasSarma S (1993) Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. *J Bacteriol* 175:4584–4596
- Ng W-L, Kothakota S, DasSarma S (1991) Structure of the gas vesicle plasmid in *Halobacterium halobium* inversion isomers, inverted repeats, and insertion sequence. *J Bacteriol* 173:1958–1964
- Ng W-L, Arora P, DasSarma S (1994) Large deletions in class III gas vesicle-deficient mutants of *Halobacterium halobium*. *Syst Appl Microbiol* 16:560–568
- Ng WV, Ciufo SA, Smith TM, Bumgarner RE, Baskin D, Faust J, Hall B, Loretz C, Seto J, Slagel J, Hood L, DasSarma S (1998) Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? *Genome Res* 8:1131–1141
- Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H, Isenbarger TA, Peck RF, Pohlschroder M, Spudich JL, Jung KW, Alam M, Freitas T, Hou S, Daniels CJ, Dennis PP, Omer AD, Ebhardt H, Lowe TM, Liang P, Riley M, Hood L, DasSarma S (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA* 97:12176–12181
- Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T (2007) Genetic and physical mapping of DNA replication origins in *Haloflex volcanii*. *PLoS Genet* 3:e77
- Oremland RS, Stolz JF, Hollibaugh JT (2004) The microbial arsenic cycle in Mono Lake, California. *FEMS Microbiol Ecol* 48:15–27
- Ou WB, Luo W, Park YD, Zhou HM (2001) Chaperone-like activity of peptidyl-prolyl cis-trans isomerase during creatine kinase refolding. *Protein Sci* 10:2346–2353
- Pfeifer F, Ghahraman P (1993) Plasmid pHH1 of *Halobacterium salinarium*: characterization of the replicon region, the gas vesicle gene cluster and insertion elements. *Mol Gen Genet* 238:193–200
- Pfeifer F, Weidinger G, Goebel W (1981) Genetic variability in *Halobacterium halobium*. *J Bacteriol* 145:375–381

- Robinson NP, Bell SD (2007) Extrachromosomal element capture and the evolution of multiple replication origins in archaeal chromosomes. *Proc Natl Acad Sci USA* 104:5806–5811
- Rosen BP (2002) Biochemistry of arsenic detoxification. *FEBS Lett* 529:86–92
- Sastre JI, Cabezón E, de la Cruz F (1998) The carboxyl terminus of protein TraD adds specificity and efficiency to F-plasmid conjugative transfer. *J Bacteriol* 180:6039–6042
- Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe, TM (2006) The UCSC archaeal genome browser. *Nucleic Acids Res* 34:D407–D410
- Shukla HD, DasSarma S (2004) Complexity of gas vesicle biogenesis in *Halobacterium* sp. strain NRC-1: identification of five new proteins. *J Bacteriol* 186:3182–3186
- Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* 8:251–262
- Taubner LM, McGuirl MA, Dooley DM, Copie V (2006) Structural studies of Apo NosL, an accessory protein of the nitrous oxide reductase system: insights from structural homology with MerB, a mercury resistance protein. *Biochemistry* 45:12240–12252
- Tchelet R, Mevarech M (1994) Interspecies genetic transfer in halophilic Archaeobacteria. *Syst Appl Microbiol* 16:578–581
- Thomm M (1996) Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev* 18:159–171
- Tu D, Blaha G, Moore PB, Steitz TA (2005) Gene replacement in *Haloarcula marismortui*: construction of a strain with two of its three chromosomal rRNA operons deleted. *Extremophiles* 9:427–435
- Wang G, Kennedy SP, Fasiludeen S, Rensing C, DasSarma S (2004) Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system. *J Bacteriol* 186:3187–3194
- Wanner C, Soppa J (1999) Genetic identification of three ABC transporters as essential elements for nitrate respiration in *Haloferax volcanii*. *Genetics* 152:1417–1428
- Yang CF, DasSarma S (1990) Transcriptional induction of purple membrane and gas vesicle synthesis in the archaeobacterium *Halobacterium halobium* is blocked by a DNA gyrase inhibitor. *J Bacteriol* 172:4118–4121

Megaplasmiids and the Degradation of Aromatic Compounds by Soil Bacteria

Eve Vedler

Contents

1	Introduction.....	34
2	Aromatic Catabolic Traits Conferred by Megaplasmiids in Gram Negative Bacteria.....	39
2.1	Degradation of Chloroaromatics Mediated by Megaplasmiids	42
2.2	Degradation of Phenolic Compounds Mediated by the IncP-2 Plasmid pVII150 and the IncP-1 Plasmid pIJB1.....	43
3	Aromatic Catabolic Traits Conferred by Megaplasmiids in Gram Positive Bacteria	48
	References.....	49

Abstract Microbial biodegradation pathways are often, either fully or partially, encoded on mobile genetic elements, including catabolic plasmids. As these plasmids frequently contain beside catabolic genes/operons the full set of determinants necessary for conjugative transfer, they are relatively large, from approximately 50 kb up to more than 1 Mb. This chapter briefly describes 37 catabolic megaplasmiids (≥ 100 kb), involved in the degradation of aromatic compounds by soil bacteria, with an emphasis on the degradative traits conferred by these plasmids. Eighteen of the discussed plasmids have been completely sequenced. Eighteen plasmids have *Pseudomonas* as the host strain. The representatives of this genus are predominant among the biodegradative bacteria discovered to date, indicating that catabolic plasmids play a crucial role in biodegradative capabilities of soil bacteria. As more sequences from bacterial genomes and catabolic plasmids become available, it may eventually be possible to determine how all these different genetic arrangements have evolved.

E. Vedler

Department of Genetics, Institute of Molecular and Cell Biology, Tartu University,
23 Riia Street, Tartu 51010, Estonia
e-mail: evedler@ebc.ee

1 Introduction

Biodegradative capabilities of microorganisms are overwhelming. Microbial biodegradation is the cornerstone of all biogeochemical cycles. The potential of soil and water bacteria to degrade various kinds of environmental pollutants, including man-made chemicals (xenobiotics) which have only recently been introduced into nature, is very large. New metabolic pathways are continuously evolving, which enable these microorganisms to use compounds that they have not encountered before. In February 2008, the Biocatalysis/Biodegradation Database of the University of Minnesota (UM-BBD, <http://umbbd.msi.umn.edu/>) contained 171 microbial biodegradation pathways, 1,185 biodegradation reactions, 1,106 biodegradable compounds, 758 enzymes, and 440 microorganism entries; this is most certainly only the “peak of an iceberg,” studied so far. Such an amazing armory of microbial functions depends on the diversity and flexibility of microbial genomes.

To survive, bacteria must rapidly adapt to changing environmental stimuli such as transient nutrient resources, and also exposure to xenobiotics, toxic chemicals, or unusually high amounts of natural chemical compounds (oil spills, etc.). As the result of this adaptation, bacteria acquire novel catabolic abilities. Mobile genetic elements (MGEs), acquired via horizontal (lateral) gene transfer (HGT), play a major role in such adaptation. MGEs – transferable (mega)plasmids, (conjugative) transposons, integrons, genomic islands, or phages – are able to move within and/or between genomes, allowing “evolution in quantum leaps” (Hacker and Carniel 2001, Nojiri et al. 2004, Osborn and Boltner 2002). MGEs provide a location where catabolic and anabolic genes can be assembled from the prokaryotic horizontal gene pool to provide the response to environmental stresses (Thomas and Smalla 2000). Beside biodegradation genes/operons, MGEs also carry determinants for resistance to antibiotics, heavy metals and radiation, symbiosis and virulence, bacteriocin production, and increased mutation frequencies.

Soil is a chemically complex environment, in part due to the wide range of compounds produced by plants, and further complexity is introduced by decomposition and soil biogenesis. Many compounds of plant origin are chemically related, although they are usually present at concentrations that individually do not support bacterial growth. Thus, the alternative, potentially successful competitive strategy for soil bacteria is to use the diverse compounds simultaneously, rather than rapidly respond to transient nutrient resources. This strategy may underlie selective pressure for large genomes having multiple chromosomes and/or (mega)plasmids, which encode numerous paralogous catabolic genes. Two examples are *Burkholderia xenovorans* LB400 and *Rhodococcus* sp. RHA1 (see further). As these two organisms are phylogenetically very diverse, this suggests the ancient origin of this catabolic capacity (McLeod et al. 2006).

Catabolic plasmids contain genes encoding the enzymes required for the degradation and utilization of chemical compounds. Typically, these plasmids have low copy numbers and they are relatively large (from 50 kb up to megaplasmids of more than 1 Mb). Although the catabolic plasmids described to date are mostly circular,

linear catabolic plasmids have been isolated from gram positive bacteria. The first plasmids involved in the degradation of xenobiotic compounds were described more than 30 years ago. At the present time, the list of catabolic plasmids is increasing continuously due to massive sequencing.

Comprehensive overviews of catabolic plasmids (Dennis 2005, Nojiri et al. 2004) and degradation pathways of aromatic compounds (Williams and Sayers 1994, Harwood and Parales 1996) can be found in the scientific literature. Based on these papers and the collected up-to-date information, for example, from the Plasmid Genome Database (<http://www.genomics.ceh.ac.uk/plasmiddb/>), this chapter will focus on catabolic megaplastids ≥ 100 kb, involved in the degradation of aromatic compounds by soil bacteria. In **Table 1** there is a list of 37 such plasmids. Two large plasmids, pIJB1 and pVII150, have been included because of their importance: the first one is slightly below the size criteria (99,448 bp) and the size of the latter one has not been determined exactly. These two plasmids will be discussed in more detail in a separate chapter. Eighteen plasmids listed in **Table 1** have been completely sequenced. Eight plasmids in this list are linear, they all have gram positive host strains – *Rhodococcus* (7 plasmids) and *Arthrobacter* (1 plasmid). The bacterial strains and catabolic plasmids involved in the degradation of polycyclic aromatic hydrocarbons and heterocyclic aromatic compounds is the subject of another chapter of this book (see Nojiri et al. 2008) and therefore will not be discussed in detail here.

The large size of catabolic plasmids is attributable to the fact that they often contain the full set of plasmid transfer genes as well as the collection of catabolic genes/operons required for (complete) degradation of chemical compound(s), associated with numerous transposons and/or IS elements or their remnants. Although in many cases the respective data are not available, only six of the plasmids listed in **Table 1** have actually been shown to be nontransferable. Catabolic modules are typically inserted into the plasmid backbone at nonessential sites, that is, where the backbone can be disrupted without cost to the plasmid's normal functions. Like plasmid backbone gene organization, catabolic pathway gene organization should improve (clustering of genes that code for degradation pathways) until gene expression is optimized in terms of maximal degradation capability and regulatory flexibility; the most well-conserved catabolic pathways are those that are most tightly clustered (Thomas 2000). As more sequences from bacterial genomes and catabolic plasmids are made available, it should be possible to trace at least some stages of the evolution of the different genetic arrangements. Plasmid genome sequence comparisons, especially those of the promiscuous IncP plasmids, indicate that catabolic plasmids have complex genetic histories resulting from transposition and recombination events (Dennis 2005).

Only eight plasmids in **Table 1** have been assigned to an incompatibility group, namely IncP-1, -2, -7, and -9, and six of them have *Pseudomonas* as the host strain. It has been pointed out (Top et al. 2000) and it is also apparent in **Table 1** that genes for biodegradation of naturally occurring compounds (phenol, toluene, xylene, naphthalene) seem to be located on IncP-2, -7, or -9 plasmids, while genes for degrading xenobiotic compounds (2,4-D, atrazine) are frequently found on IncP-1

Table 1 Overview of catabolic megaplasmids involved in the biodegradation of aromatic compounds

Plasmid	Substrate(s) degraded by the host strain	Host strain	Size (kb)	Inc group	Conjugation	References	GenBank accession numbers
pLJB1	2,4-D	<i>Burkholderia cepacia</i> 2a	99,5 (cs) ^a	P-1	Na (+)	Poh et al. (2002), Xia et al. (1998)	DQ065837
pTV1	2,4-D	<i>Variovorax paradoxus</i> TV1	200	Na	Na	Vallaey et al. (1998)	AB028643
pSAH	2-Aminobenzenesulfonate	<i>Alcaligenes</i> sp. O-1	172	P-9	+	Jahnke et al. (1990)	AF109074
pRA500	3,5-Xylenol, <i>p</i> -cresol	<i>Pseudomonas putida</i> NCIB9869	500	Na	+	Hopper and Kemp (1980)	AF101076, U96339
pAC25	3-Chlorobenzoate	<i>Pseudomonas putida</i> AC858	117	P-1	+	Chatterjee et al. (1981)	–
pAC27	3-Chlorobenzoate	<i>Pseudomonas putida</i> AC867	110	Na	+	Chatterjee and Chakrabarty (1982, 1984), Frantz and Chakrabarty (1987), Ghosal et al. (1985), Ghosal and You (1988)	M16964
pCIT1	Aniline	<i>Pseudomonas</i> sp. CIT1	100	Na	Na	Anson and Mackinnon (1984)	–
pHMT112	Benzene	<i>Pseudomonas putida</i> ML2	112	Na	+	Fong et al. (2000), Tan and Fong (1993), Tan and Mason (1990)	AF148496, AF176355
pWW174	Benzene	<i>Acinetobacter calcoaceticus</i> RJE174	200	Na	+	Winstanley et al. (1987)	–
pWW100	Biphenyl, benzoate	<i>Pseudomonas</i> sp. CB406	200	Na	–	Lloyd-Jones et al. (1994)	AY040093
pNL1	Biphenyl, benzoate, <i>m</i> -xylene, salicylate, naphthalene	<i>Novosphingobium aromaticivorans</i> DSM 12444	184 (cs)	Na	+	Romine et al. (1999)	NC_009426
pNL2	Biphenyl, benzoate, <i>m</i> -xylene, salicylate, naphthalene	<i>Novosphingobium aromaticivorans</i> DSM 12444	487 (cs)	Na	Na	Romine et al. (1999)	NC_009427
pLP6 (linear)	Biphenyl/PCB-s	<i>Rhodococcus globerulus</i> P6	650	Na	Na	Kosono et al. (1997)	–

pSP6 (linear)	Biphenyl/PCB-s	<i>Rhodococcus globervulius</i> P6	360	Na	Na	Kosono et al. (1997)	-
pTA421 (linear)	Biphenyl/PCB-s	<i>Rhodococcus erythropolis</i> TA421	500	Na	Na	Kosono et al. (1997)	AB014348
pRHL1 (linear)	Biphenyl/PCB-s, ethylbenzene	<i>Rhodococcus</i> sp. RHA1	1123 (cs)	Na	Na	McLeod et al. (2006)	NC_008269
pRHL2 (linear)	Biphenyl/PCB-s, ethylbenzene	<i>Rhodococcus</i> sp. RHA1	443 (cs)	Na	+	McLeod et al. (2006)	NC_008270
pRHL3 (linear)	Biphenyl/PCB-s, ethylbenzene	<i>Rhodococcus</i> sp. RHA1	332 (cs)	Na	Na	McLeod et al. (2006)	NC_008271
Mega-plasmid	Biphenyl/PCB-s, phenylacetate, anthranilate, salicylate, benzoate	<i>Burkholderia xenovorans</i> LB400	1471 (cs)	Na	Na	Chain et al. (2006)	NC_007953
pCAR1	Carbazole, dioxins	<i>Pseudomonas resinovorans</i> CA10	199 (cs)	P-7	+	Maeda et al. (2003)	NC_004444
pCAR3	Carbazole	<i>Sphingomonas</i> sp. KA1	255 (cs)	Na	-	Shimani et al. (2007)	NC_008308
pP51	Chlorobenzene	<i>Pseudomonas</i> sp. P51	110	Na	-	van der Meer et al. (1991b)	U15298, M57629
pBD2 (linear)	Isopropylbenzene	<i>Rhodococcus erythropolis</i> BD2	210 (cs)	Na	+	Stecker et al. (2003)	NC_005073
pRE4	Isopropylbenzene, toluene, ethylbenzene, phenol, benzoate	<i>Pseudomonas putida</i> RE204	105	Na	Na	Eaton and Timmis (1986)	AF006691
pND6-1	Naphthalene	<i>Pseudomonas</i> sp. ND6	102 (cs)	P-7	-	Li et al. (2004)	NC_005244
pPGH1	Phenol and its methylated derivatives	<i>Pseudomonas putida</i> H	220	Na	Na	Herrmann et al. (1987)	AF052749-51, X80765, X91145, Y09450

(continued)

Table 1 (continued)

Plasmid	Substrate(s) degraded by the host strain	Host strain	Size (kb)	Inc group	Conjugation	References	GenBank accession numbers
pV1150	Phenol and its methylated derivatives	<i>Pseudomonas</i> sp. CF600	Na ^b	P-2	+	Shingler et al. (1989)	X52805, X60657, X60835, X68033, M33263, M60276
pEST1026	Phenol	<i>Pseudomonas putida</i> EST1020	109	Na	+	Kivisaar et al. (1990)	M57500
pBVIE04 (pTOM)	Phenol, toluene, <i>o</i> -cresol, <i>m</i> -cresol, benzene	<i>Burkholderia vietnamiensis</i> G4	107 (cs)	Na	+	Mahenthiralingam et al. (2005), O'Sullivan et al. (2007), Shields et al. (1995)	NC_009228
pRE1	Phthalate	<i>Arthrobacter keyseri</i> 12B	130	Na	Na	Eaton (2001)	AF331043
pBS1010	<i>p</i> -Toluenesulfonate	<i>Comamonas testosteroni</i> BS1310	130	Na	Na	Top et al. (2000)	–
pAL1 (linear)	Quinaldine	<i>Arthrobacter nitroguajacolicus</i> Rū61a	113 (cs)	Na	+	Parschat et al. (2007)	NC_009453
pKJ1	Toluene	<i>Pseudomonas</i> sp. TA8	225	Na	+	Yano and Nishi (1980)	–
pTK0	Toluene	<i>Pseudomonas putida</i> PPK1	150	Na	–	Keshavarz et al. (1985)	–
pDK1	Xylene, toluene	<i>Pseudomonas putida</i> HS1	180	Na	+	Kunz and Chapman (1981), Shaw and Williams (1988)	AF019635
pWW0	Xylene, toluene	<i>Pseudomonas putida</i> mt-2	117 (cs)	P-9	+	Greated et al. (2002)	NC_003350
pWW53	Xylene, toluene	<i>Pseudomonas putida</i> MT53	108 (cs)	P-7	–	Yano et al. (2007)	NC_008275

^acs, complete nucleotide sequence^bNa, data not available

plasmids. As IncP-1 plasmids are highly transferable and the most promiscuous plasmids characterized to date, they can act as efficient shuttle vectors for locally adapted genes, resulting in the rapid local adaptation of both phylogenetically related and distinct populations present in the same bacterial community in response to, for example, exposure to xenobiotics.

Of the 440 microorganisms in the UM-BBD database, one bacterial genus is predominant (131 representatives), namely *Pseudomonas*. The other most abundant genera in this database are *Rhodococcus* (22), *Sphingomonas* (20), *Arthrobacter* (17), and *Burkholderia* (14). **Table 1** contains 18 strains of *Pseudomonas* (11 of them are *P. putida*), 4 strains of *Rhodococcus*, 3 strains of *Burkholderia*, 2 strains of *Shingomonas/Novosphingobium* and *Arthrobacter*, implicating that catabolic (mega)plasmids play a crucial role in making the representatives of this genera such powerful degraders.

2 Aromatic Catabolic Traits Conferred by Megaplastids in Gram Negative Bacteria

The best studied and probably the most ubiquitous microbial strategy for aerobic degradation of aromatic compounds involves two critical steps: first, hydroxylation of adjacent carbon atoms of the aromatic ring and second, the ring cleavage of the resulting catecholic intermediates (Dagley 1986). In the case of phenol degradation, the aromatic ring is first monohydroxylated by phenol hydroxylase (PH, phenol 2-monooxygenase) at *ortho* position to the preexisting hydroxyl group (see **Fig. 1** for catabolic pathways). The next step is catalyzed by either catechol 1,2-dioxygenase (C12O, initiating the *ortho* pathway leading to formation of succinyl-CoA and acetyl-CoA) or catechol 2,3-dioxygenase (C23O, initiating the *meta* pathway leading to formation of pyruvate and acetaldehyde). Two different types of PHs have been identified: single-component (sPH) and multicomponent (mPH). In plasmid **pEST1026**, the *pheBA* operon codes for C12O and sPH, necessary for phenol degradation by its host strain (Kivisaar et al. 1990). In the case of **pPGH1** (Herrmann et al. 1995) and **pVI150** (Bartilson et al. 1990), which are found in *P. putida* host strains that also degrade methylated phenols, mPH and C23O are encoded by the *phl* and *dmp* operons, respectively.

Bacteria express a range of enzymes for the initial attack of different aromatic compounds – benzoate, benzene, toluene, xylenes, cresols, naphthalene, salicylate, mandelate, aniline, and others. These peripheral (upper) pathways bring about a convergence of many different aromatic compounds to a limited number of ring cleavage substrates such as catechol, protocatechuate, gentisate, hydroquinone, etc., which are further degraded by central (lower) pathways (Dagley 1986). In the natural hosts of the plasmids **pWW0**, **pWW53**, **pDK1**, and **pNL1**, degradation of toluene and xylenes is mediated by the *xyl* genes, the pathway proceeds via catechol or its methylated derivatives and C23O as the ring fission enzyme (Burlage et al. 1989, Romine et al. 1999, Shaw and Williams 1988, Tsuda and Genka 2001)

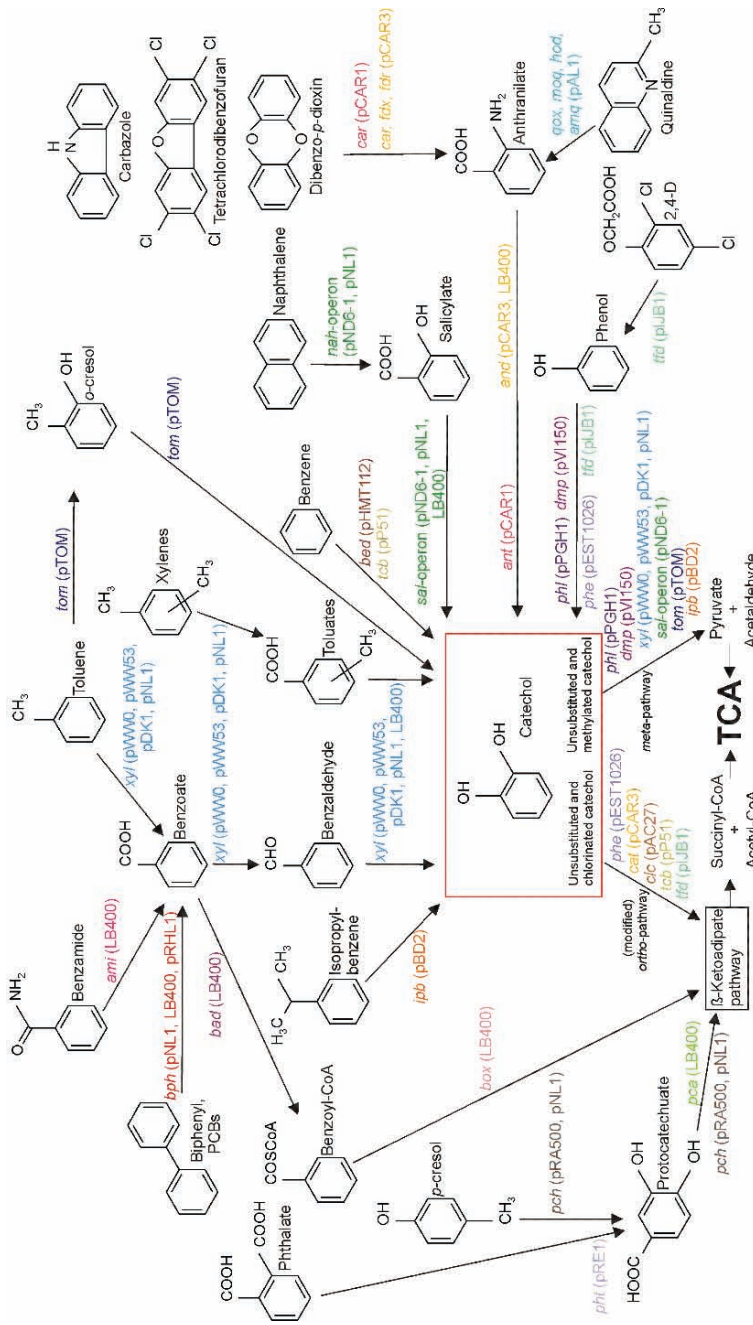


Fig. 1 Aromatic biodegradation pathways coded by catabolic genes and operons located on megaplasmids. Chlorine and methyl substitutions are not shown in the case of 2,4-dichlorophenol, di- and trichlorobenzenes, and catechol. TCA tricarboxylic acid cycle

(for detailed information about the most extensively studied catabolic megaplastids, pWW0 and its relatives, see the articles by Greated et al. 2002 and Ramos et al. 1997). In *P. putida* NCIB9869, *p*-cresol is degraded via the intermediate protocatechuate by the products of the *pch* genes coded on the plasmid **pRA500** (Hopper and Kemp 1980). In *Pseudomonas putida* ML2, benzene is degraded to catechol by the peripheral degradation pathway, coded by the *bedC1C2BA* and *bedD* genes located on the plasmid **pHMT112** (Fong et al. 2000). Catechol is further channeled through the *ortho* pathway to tricarboxylic acid cycle (TCA cycle) intermediates. In pHMT112, the *bed* upper pathway genes are encompassed on a class-I-type transposable element, Tn5542, and are located some distance away from the lower pathway genes. This suggests an evolutionary route by which the complete catabolic pathway arose from the accretion of the *ortho* pathway genes, followed by the acquisition of a set of genes specifying the upper pathway enzymes. The same mechanism, by which gene clusters can be mobilized as gene cassettes and joined with others to form novel catabolic pathways, has been suggested also in the case of plasmid pP51, described later.

Novosphingobium aromaticivorans (formerly *Sphingomonas aromaticivorans*) F199 has the ability to degrade an unusual variety of aromatic compounds, including toluene, all isomers of xylene, *p*-cresol, naphthalene, biphenyl, dibenzothiophene, fluorene, salicylate, and benzoate (Romine et al. 1999). Nearly half of one of its plasmids, **pNL1**, encodes pathways for the complete catabolism and transport of these compounds, bearing *bph*, *xyl*, *nah*, and *pch* genes among others. In the GenBank annotation of **pNL2**, the second plasmid of F199 (NC_009427), there are genes coding for catechol degradation *ortho*-pathway enzymes, as well. Thus, this plasmid probably also contributes to the degradation of aromatic compounds by its host strain.

Burkholderia is another effective genus with respect to biodegradation of aromatic compounds. *B. xenovorans* LB400 is one of the most important aerobic polychlorinated biphenyl (PCB) degraders yet discovered (Chain et al. 2006). PCBs are widespread, highly toxic, and persistent xenobiotic organic pollutants. LB400 oxidizes more than 20 PCB congeners, with up to six chlorine substitutions on the biphenyl rings, and has become a model organism to study PCB biodegradation. LB400 has one of the largest known, multireplicon bacterial genomes, consisting of two chromosomes (the "core" chromosome of 4.9 Mb and the small chromosome of 3.36 Mb) and one megaplastid (1.47 Mb). One characteristic of the genus *Burkholderia* is high genomic plasticity (due to the presence of multiple insertion sequences), diversity, and specialization. It was discovered that >20% of the LB400 sequence was recently acquired by HGT. According to in silico prediction, there are at least 11 central and 20 peripheral pathways coded in LB400, among the highest in any sequenced bacterial genome, indicating an unusually high metabolic versatility. More than 170 genes encoding these pathways are spread over the three replicons. Four peripheral pathways for the degradation of salicylate (*nahW*), biphenyl (*bph* operon), benzaldehyde (*xylC*), and benzamide (*amiE*) are encoded completely on the megaplastid. Two peripheral pathways for the degradation of phenylalkanoates (*phaC*) and phthalate (*bphB*) are partially encoded on the megaplastid.

The peripheral pathway for the degradation of anthranilate (*and* operon) has two paralogous copies – one on the small chromosome and one on the megaplasmid. Three central pathways for the degradation of protocatechuate (*pcaII*, paralogs also on both chromosomes), phenylacetyl-CoA (*paaC/paaH*), and homogentisate (2 copies of *hmgA*, *hmgB*, *iciR*, paralogs also on both chromosomes) are partially encoded on the megaplasmid (Chain et al. 2006). In addition, benzoate is degraded by two different pathways in the strain LB400 (Denef et al. 2006): a *ben-cat* pathway and two paralogous *box* pathways (on the large chromosome and on the megaplasmid). The peripheral *ben*-pathway is encoded on the small chromosome and is responsible for turning benzoate into catechol; the latter is degraded by the *cat* central *ortho* pathway, encoded on the large chromosome. An alternative route for the degradation of benzoate (as well as other aromatic compounds) has been discovered recently, which does not involve ring cleavage dioxygenases; instead, aromatic acids are processed as coenzyme A (CoA) thioesters (Gescher et al. 2005). The initial step is catalyzed by specific CoA ligases, in case of LB400 coded by the *badA* gene (two paralogs in the large chromosome and the megaplasmid). The aromatic ring of the CoA-activated aromatic acid (benzoate) is then reduced to a dihydroxylated intermediate by a dioxygenase (BoxAB). All intermediates in this pathway are CoA thioesters and the ring cleavage by BoxC does not require molecular oxygen. The pathway encoded by the *box* operon leads to formation of β -keto adipyl-CoA, a known intermediate of the classical β -keto adipate pathway. The enzymes of the two copies of *box* pathways share 70–90% amino acid sequence identity to each other. As discussed earlier, redundancy in metabolic pathways and genes seems to provide more efficient degradative capacities conferring a selective advantage (Laemmli et al. 2004, Ledger et al. 2006).

The rhizosphere bacterium *Burkholderia vietnamiensis* strain G4, a member of the ecologically versatile and biotechnologically important *Burkholderia cepacia* complex (Coenye and Vandamme 2003), can proficiently degrade many toxic aromatic hydrocarbons (Nelson et al. 1987). Its genome consists of three chromosomes (3.65, 2.41, and 1.241 Mb) and five megaplasms pBVIE01–05 (398, 266, 226, 107, and 88 kb). One of these plasmids, **pBVIE04** or **pTOM**, carries all the genes required to degrade phenol, toluene, *o*-cresol, *m*-cresol, and benzene (O’Sullivan et al. 2007, Shields et al. 1989, 1995), including toluene *ortho*-monooxygenase (TomA) and C23O (TomB).

2.1 Degradation of Chloroaromatics Mediated by Megaplasms

The aerobic degradation of chloroaromatic compounds appears to be governed by similar principles (Schlomann 1994) – chlorosubstituted catechols as central intermediates are converted to β -keto adipate in a series of reactions similar to the catechol *ortho*-cleavage pathway, designated as the modified *ortho* pathway. In *P. putida* AC867, 3-chlorobenzoate degradation leads to the formation of chlorocatechol, which is further degraded by chlorocatechol 1,2-dioxygenase, chloromuconate

cycloisomerase, and dienelactone hydrolase coded by the *clcABD* operon present in the plasmid **pAC27** (Chatterjee and Chakrabarty 1984). The host strain of the plasmid pP51, *Pseudomonas* sp. strain P51, is able to use di- and trichlorobenzene(s) as sole carbon and energy sources. The *tcb* genes, responsible for this phenotype, are located in two regions of **pP51**. The upper-pathway gene cluster *tcbAaA-bAcAdB*, converting chlorobenzoates to the respective chlorocatechol, is similar in size, organization, and functional characteristics to sets of genes for other bacterial multicomponent dioxygenases, such as the toluene degradation genes *todC1C2BA* on the *P. putida* F1 chromosome (Zylstra et al. 1988), the biphenyl degradation genes *bphA1A2A3A4B* on the *Burkholderia xenovorans* LB400 megaplastid (Mondello 1989), the xylene/toluene degradation genes *xylXYZ* on the plasmid pWW0 (Harayama et al. 1986), or the aforementioned *bed* catabolic genes of pHMT112. In the case of *tod* and *bph*, the upper and lower pathway genes are contiguous and form a complete operon, whereas in case of the *bed* and *tcb* genes, the upper pathway genes are located on a transposon, away from the lower pathway genes. The latter catabolic genes are a part of Tn5280 (van der Meer et al. 1991a, b). The pP51 gene cluster *tcbCDEF* codes for the modified *ortho*-pathway for chlorocatechol degradation.

The herbicide 2,4-dichlorophenoxyacetic acid (2,4-D) is degraded by several bacteria belonging to different phylogenetic groups. The best-studied 2,4-D degradation genes are *tfd*-like, and the most extensively studied *tfd* genes are located on the 88 kb IncP-1 plasmid pJP4 of *Ralstonia eutropha* JMP134 (Trefault et al. 2004). The peripheral pathway coded by the *tfdA* and *tfdB/tfdB_{II}* genes is responsible for converting 2,4-D and 3-chlorobenzoate to the respective chlorocatechol, the latter is further degraded by the modified *ortho*-pathway encoded by the isofunctional *tfdCDEF* and *tfdD_{II}C_{II}E_{II}F_{II}* operons (Figs. 2 and 3). Both gene clusters are required for efficient degradation of these compounds by the strain JMP134 (Plumeier et al. 2002). The expression of the catabolic *tfd* genes is regulated by the transcriptional activator TfdR (Leveau and van der Meer 1996). Two other sequenced 2,4-D plasmids are the 77 kb **pEST4011** of *Achromobacter xylosoxidans* EST4002 (Vedler et al. 2004) and the 99.5 kb **pIJB1** of *Burkholderia cepacia* 2a (Poh et al. 2002, Xia et al. 1998) (both belonging to IncP-1). The two last plasmids contain *tfd* genes homologous to the ones of pJP4, but organized differently. The sequenced region of the plasmid **pTV1** (AB028643) contains *tfd* genes almost identical to those of pEST4011 and pIJB1.

2.2 Degradation of Phenolic Compounds Mediated by the IncP-2 Plasmid pVII50 and the IncP-1 Plasmid pIJB1

IncP-2 plasmids CAM (Rheinwald et al. 1973) and OCT (Fennewald et al. 1978), controlling camphor and octane degradation by their host strains, were the first catabolic plasmids isolated, but due to their large size and difficulties of isolation,

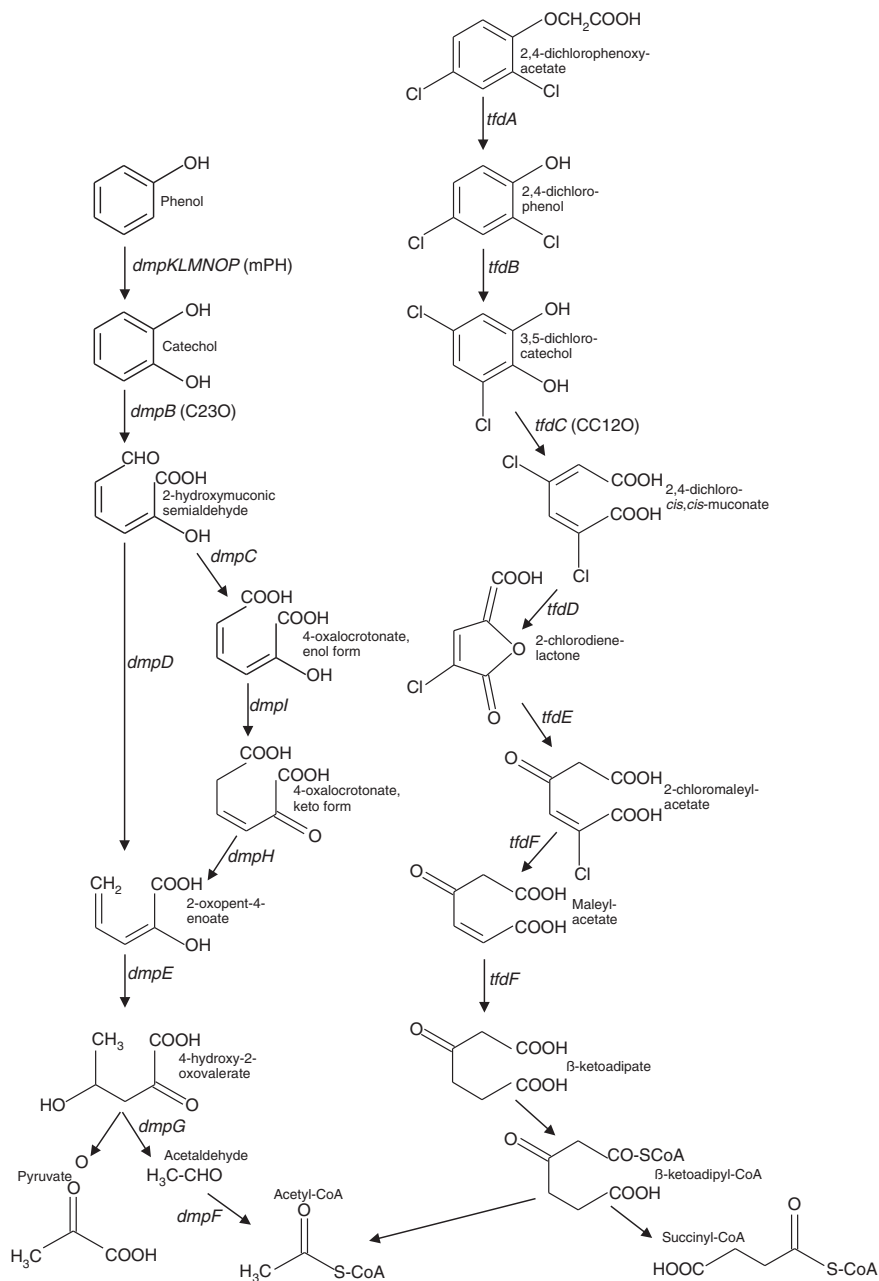


Fig. 2 On the left, the meta degradation pathway of phenol and its methylated derivatives determined by the *dmp* genes coded on the plasmid pV1150 of *Pseudomonas* sp. CF600. The names are given beside the compounds. *dmpKLMNOP* codes for a multicomponent phenol hydroxylase (mPH), *dmpB* codes for catechol 2,3-dioxygenase (C23O), *dmpC* codes for 2-hydroxymuconic semialdehyde dehydrogenase, *dmpI* codes for 4-oxalocrotonate isomerase, *dmpH* codes for 4-oxalocrotonate decarboxylase, *dmpD* codes for 2-hydroxymuconic semialdehyde hydrolase, *dmpE* codes for 2-oxopent-4-enoate hydratase, *dmpG* codes for 4-hydroxy-2-oxovalerate aldolase,

only their catabolic regions have been sequenced so far. The same is true for the IncP-2 megaplasmid **pVI150**, briefly mentioned earlier, bearing the *dmp* genes for phenol degradation. No information about the backbone of this plasmid is available (aside from the fact that the plasmid is conjugative), but there is much information regarding the catabolic gene organization and regulation as well as the chemistry and enzymology of the phenol degradation pathway conferred by pVI150 (Shingler et al. 1989, 1992, 1993; Shingler and Moore 1994, Shingler and Pavel 1995; Sze and Shingler 1999; Sze et al. 2001).

The host strain of pVI150, *Pseudomonas* sp. CF600, can efficiently grow with phenol, cresols, or 3,4-dimethylphenol as the sole carbon and energy source. The catabolic pathway for the degradation of these substrates into TCA cycle intermediates (Fig. 2) is entirely encoded on the contiguous 15 kb region of pVI150. The *dmp* catabolic genes are all encoded by the *dmpKLMNOPQBCDEFGHI* operon. Phenol is first monohydroxylated by mPH, encoded by the *dmpKLMNOP* genes. The aromatic ring of the resulting catechol is then cleaved by the *dmpB* gene product, C23O, thus initiating the *meta* pathway leading to the formation of pyruvate and acetyl-CoA. The expression of the *dmp* catabolic genes is tightly regulated by the product of the divergently transcribed *dmpR* regulatory gene (Shingler et al. 1993, Shingler and Moore 1994, Shingler and Pavel 1995). In the absence of the pathway substrates, only very low levels of expression of the *dmp* catabolic genes occur. The positive regulator DmpR belongs to the NtrC class of transcriptional activators, which act in conjunction with σ^{54} holoenzyme RNA polymerase. The DmpR regulatory protein is activated by aromatic effector molecules. Interestingly, the effector specificity range of DmpR is broad and includes, in addition to pathway substrates (phenol and its methylated derivatives) and intermediates (e.g., catechol and its methylated derivatives), compounds that are not catabolized by the strain CF600, for example, benzene, benzoate, benzaldehyde, etc. The aromatic catabolic pathways, like other catabolic processes, have to function efficiently within the general metabolic context of the host cell. The degradative pathways are not only subject to the specific regulatory mechanism responsive to substrate availability, but also integrated into the dominant global regulatory systems signaling the nutritional and energy status of the host cell (excellently reviewed in Shingler 2003). In the case of CF600, the physiological control, mediated by the alarmone (p)ppGpp, is superimposed on the specific regulatory system by aromatic compounds. Low levels of (p)ppGpp present when there is an abundant

←
Fig. 2 (continued) *dmpF* codes for acetaldehyde dehydrogenase. On the right, the modified *ortho* degradation pathway of 2,4-dichlorophenoxyacetate (2,4-D) determined by the *tfd* genes coded on the plasmid pIJB1 of *Burkholderia cepacia* 2a. The names are given beside the compounds. *tfdA* codes for 2,4-D/ α -ketoglutarate dioxygenase, *tfdB* codes for 2,4-dichlorophenol hydroxylase, *tfdC* codes for chlorocatechol 1,2-dioxygenase (CC12O), *tfdD* codes for chloromuconate cycloisomerase, *tfdE* codes for chlorodienelactone hydrolase, *tfdF* codes for (chloro)maleylacetate reductase. The last two enzymes of the β -keto adipate pathway, β -keto adipate CoA-transferase and acetyl-CoA C-acyltransferase, are chromosomally encoded

supply of nutrients cause silencing of the DmpR-mediated transcriptional response. Upon transition between exponential and stationary phase or during growth on poor carbon sources, the concentration of the alarmone increases in response to nutrient limitation (Sze and Shingler 1999). Furthermore, the optimal performance of DmpR requires another global regulator, integration host factor (IHF), which stimulates DmpR-activated transcriptional output (Sze et al. 2001). Thus, we have here a fascinating example of how global regulatory circuits also modulate the performance of plasmid-borne catabolic genes.

IncP-1 plasmids and their backbones are among the best-studied replicons. Among them there are many representatives of antibiotic resistance factors as well as catabolic plasmids. They are divided into four subgroups: α , β , γ , and δ (Adamczyk and Jagura-Burdzy 2003, Vedler et al. 2004), the amino acid identities between the backbone gene products of different subgroups range from about 38–87%. The complete nucleotide sequence has been determined in case of the 99,448 bp IncP-1 megaplasmiid **pIJB1**, briefly mentioned in the previous section. This plasmid contains the transposon Tn5530, bearing the *tfd* genes homologous to a set of genes on pJP4, but organized differently (Poh et al. 2002). **Figure 3** shows the comparison of the physical map of pIJB1 (according to the GenBank annotation DQ065837) with the other two sequenced 2,4-D plasmids pJP4 and pEST4011, the representatives of IncP1 β and δ subgroups, respectively. The catabolic transposon Tn5530 is also present on pEST4011, although in the opposite orientation with respect to the *tra* region and with the duplication of a 6 kb region. The latter took place in the laboratory during enrichment procedures performed in order to stabilize the plasmid (Vedler et al. 2004). The plasmid pIJB1 contains also, outside of the transposon Tn5530, a set of *bph* genes, not present either in pJP4 or pEST4011, and the *merPADE* operon, part of the mercury resistance operon *merTPADE* present on pJP4. The sequence and organization of the *bph* genes is most similar to the *bph* genes of Tn4371, a member of the family of genomic islands related to IncP and Ti plasmids (Toussaint et al. 2003). This transposon allows its host, *Ralstonia oxalatica* A5, to degrade biphenyl and 4-chlorobiphenyl.

It has been pointed out by many authors, and it is obvious also in this case, that genes and gene cassettes of different origin are recruited during assemblage of catabolic plasmids. The three 2,4-D plasmids discussed earlier seem to be formed independently. In the case of pIJB1 and pEST4011, the full set of 2,4-D degradation genes was inserted into the plasmid backbone as a part of the transposon Tn5530. As this transposon is in the opposite orientation in case of pIJB1 and pEST4011, it is the proof of independent origin of these two plasmids.

Surprisingly, the backbone of pIJB1 seems to be a fusion of IncP-1 β and δ backbones. As the result, pIJB1 contains two copies of the *trfA* gene coding for the replication initiation protein, and of the *trbABCDE* genes involved in mating pair formation during conjugation. An excellent overview of IncP-1 backbone functions and their regulation was given by Adamczyk and Jagura-Burdzy (2003).

3 Aromatic Catabolic Traits Conferred by Megaplastids in Gram Positive Bacteria

Actinomycetales are an order of gram positive bacteria that live in a broad range of environments, including soil and water. The representatives of two genera *Rhodococcus* and *Arthrobacter* are among the most frequently isolated aerobic bacterial genera. They are metabolically and ecologically diverse and have the ability to survive in environmentally harsh conditions for extended periods of time. They have been shown to degrade a wide range of organic compounds. Their assimilatory abilities have been attributed to a diversity of enzymatic activities on the one hand and to cell walls rich in mycolic acids on the other. The latter supposedly facilitate the uptake of hydrophobic compounds (Gurtler et al. 2004).

Rhodococcus sp. RHA1 is a potent PCB-degrading soil actinomycete that catabolizes a wide range of compounds and has one of the largest bacterial genomes consisting of a linear chromosome (7.804 Mb) and three linear plasmids – 1.123 Mb (**pRHL1**), 443 kb (**pRHL2**), and 332 kb (**pRHL3**) (McLeod et al. 2006). RHA1's 9,145 predicted protein-encoding genes are exceptionally rich in oxygenases (203), many of which occur in the numerous pathways predicted to degrade aromatic compounds. The catabolism of aromatic compounds in rhodococci is organized as in pseudomonads. RHA1 encodes at least 26 peripheral and 8 central aromatic pathways, and 11 peripheral pathways are encoded by plasmids (i.e., three times the density of the chromosome), suggesting that plasmids have a significant catabolic role. The plasmids also contain more insertion sequences (6), transposase genes (120), and pseudogenes (2) than the chromosome. Taken together, these results are consistent with the plasmids' role as reservoirs and workshops for the evolution of novel catabolic capabilities. In contrast to LB400, in which recent HGT appears to have played a crucial role in genome shaping, RHA1 appears to have primarily gained its large genome and diverse metabolic capacity through more ancient gene duplications and/or acquisitions.

Rhodococcus erythropolis strain BD2 harbors a linear 210-kb plasmid **pBD2** carrying the *ipb* genes for isopropylbenzene degradation (Stecker et al. 2003), the respective degradation pathway goes through 3-isopropylcatechol. The deduced proteins of this degradation pathway are 94–100% identical to RHA1's biphenyl degradation pathway enzymes encoded by **pRHL1** *bph* genes, thus indicating that the respective operons have been distributed among gram positive soil bacteria via linear-plasmid-mediated HGT.

Arthrobacter keyseri strain 12B harbors a 130 kb catabolic plasmid **pRE1** bearing the *pht* operon that encodes the conversion of phthalate to protocatechuate (Eaton 2001). Phthalate (benzene-1,2-dicarboxylate) is a central intermediate in the bacterial degradation of phthalate esters (used as plasticizers) as well as of certain fused-ring polycyclic aromatic hydrocarbons (found in fossil fuels). Gram negative bacteria transform phthalate through protocatechuate as well, but the enzymes of these pathways are not closely related although they catalyze similar reactions.

Arthrobacter nitroguajacolicus Rü61a is able to utilize quinaldine (2-methylquinoline) by degrading this compound first to anthranilate by the upper pathway

enzymes, encoded on the 113 kb catabolic megaplastid **pAL1** (Parschat et al. 2007) by the *qox*, *moq*, *hod*, and *amq* genes. It has been proposed that the lower pathway for the mineralization of anthranilate involves catechol formation and *ortho* ring cleavage; the respective genes are also encoded on pAL1.

References

- Adamczyk M, Jagura-Burdzy G (2003) Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol* 50:425–453
- Anson JG, Mackinnon G (1984) Novel *Pseudomonas* plasmid involved in aniline degradation. *Appl Environ Microbiol* 48:868–869
- Bartilson M, Nordlund I, Shingler V (1990) Location and organization of the dimethylphenol catabolic genes of *Pseudomonas* CF600. *Mol Gen Genet* 220:294–300
- Burlage RS, Hooper SW, Saylor GS (1989) The TOL (pWW0) catabolic plasmid. *Appl Environ Microbiol* 55:1323–1328
- Chain PS, Denev VJ, Konstantinidis KT et-al. (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc Natl Acad Sci USA* 103:15280–15287
- Chatterjee DK, Chakrabarty AM (1982) Genetic rearrangements in plasmids specifying total degradation of chlorinated benzoic acids. *Mol Gen Genet* 188:279–285
- Chatterjee DK, Chakrabarty AM (1984) Restriction mapping of a chlorobenzoate degradative plasmid and molecular cloning of the degradative genes. *Gene* 27:173–181
- Chatterjee DK, Kellogg ST, Hamada S, Chakrabarty AM (1981) Plasmid specifying total degradation of 3-chlorobenzoate by a modified *ortho* pathway. *J Bacteriol* 146:639–646
- Coenye T, Vandamme P (2003) Diversity and significance of *Burkholderia* species occupying diverse ecological niches. *Environ Microbiol* 5:719–729
- Dagley S (1986) Biochemistry of aromatic hydrocarbon degradation in *Pseudomonas*. In: Sokatch JR (ed) *The Bacteria*, vol X. Academic Press, London, pp 527–556
- Denev VJ, Klappenbach JA, Patrauchan MA et-al. (2006) Genetic and genomic insights into the role of benzoate-catabolic pathway redundancy in *Burkholderia xenovorans* LB400. *Appl Environ Microbiol* 72:585–595
- Dennis JJ (2005) The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol* 16:291–298
- Eaton RW (2001) Plasmid-encoded phthalate catabolic pathway in *Arthrobacter keyseri* 12B. *J Bacteriol* 183:3689–3703
- Eaton RW, Timmis KN (1986) Characterization of a plasmid-specified pathway for catabolism of isopropylbenzene in *Pseudomonas putida* RE204. *J Bacteriol* 168:123–131
- Fennelwald M, Prevatt W, Meyer R, Shapiro J (1978) Isolation of IncP-2 DNA from *Pseudomonas aeruginosa*. *Plasmid* 1:164–173
- Fong KP, Goh CB, Tan HM (2000) The genes for benzene catabolism in *Pseudomonas putida* ML2 are flanked by two copies of the insertion element IS1489, forming a class-I-type catabolic transposon, Tn5542. *Plasmid* 43:103–110
- Frantz B, Chakrabarty AM (1987) Organization and nucleotide sequence determination of a gene cluster involved in 3-chlorocatechol degradation. *Proc Natl Acad Sci USA* 84:4460–4464
- Gescher J, Eisenreich W, Worth J, Bacher A, Fuchs G (2005) Aerobic benzoyl-CoA catabolic pathway in *Azoarcus evansii*: studies on the non-oxygenolytic ring cleavage enzyme. *Mol Microbiol* 56:1586–1600
- Ghosal D, You IS (1988) Nucleotide homology and organization of chlorocatechol oxidation genes of plasmids pJP4 and pAC27. *Mol Gen Genet* 211:113–120
- Ghosal D, You IS, Chatterjee DK, Chakrabarty AM (1985) Genes specifying degradation of 3-chlorobenzoic acid in plasmids pAC27 and pJP4. *Proc Natl Acad Sci USA* 82:1638–1642

- Great A, Lambertsen L, Williams PA, Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* 4:856–871
- Gurtler V, Mayall BC, Seviour R (2004) Can whole genome analysis refine the taxonomy of the genus *Rhodococcus*? *FEMS Microbiol Rev* 28:377–403
- Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep* 2:376–381
- Harayama S, Reki M, Timmis KN (1986) Genetic analysis of a relaxed substrate specificity aromatic ring dioxygenase, toluate 1,2-dioxygenase, encoded by TOL plasmid pWW0 of *Pseudomonas putida*. *Mol Gen Genet* 202:226–234
- Harwood CS, Parales RE (1996) The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* 50:553–590
- Herrmann H, Janke D, Krejsa S, Kunze I (1987) Involvement of the plasmid pPGH1 in the phenol degradation of *Pseudomonas putida* strain H. *FEMS Microbiol Lett* 43:133–137
- Herrmann H, Müller C, Schmidt I, Mahnke J, Petruschka L, Hahnke K (1995) Localization and organization of phenol degradation genes of *Pseudomonas putida* strain H. *Mol Gen Genet* 247:240–246
- Hopper DJ, Kemp PD (1980) Regulation of enzymes of the 3,5-xyleneol-degradative pathway in *Pseudomonas putida*: evidence for a plasmid. *J Bacteriol* 142:21–26
- Jahnke M, Elbanna T, Klintworth R, Auling G (1990) Mineralization of orthonilic acid is a plasmid-associated trait in *Alcaligenes* sp. O-1. *J General Microbiol* 136:2241–2249
- Keshavarz T, Lilly MD, Clarke PH (1985) Stability of a catabolic plasmid in continuous culture. *J General Microbiol* 131:1193–1203
- Kivisaar M, Horak R, Kasak L, Heinaru A, Habicht J (1990) Selection of independent plasmids determining phenol degradation in *Pseudomonas putida* and the cloning and expression of genes encoding phenol monooxygenase and catechol 1,2-dioxygenase. *Plasmid* 24:25–36
- Kosono S, Maeda M, Fuji F, Arai H, Kudo T (1997) Three of the seven *bphC* genes of *Rhodococcus erythropolis* TA421, isolated from a termite ecosystem, are located on an indigenous plasmid associated with biphenyl degradation. *Appl Environ Microbiol* 63:3282–3285
- Kunz DA, Chapman PJ (1981) Isolation and characterization of spontaneously occurring TOL plasmid mutants of *Pseudomonas putida* HS1. *J Bacteriol* 146:952–964
- Laemmli C, Werlen C, van der Meer JR (2004) Mutation analysis of the different *tfd* genes for degradation of chloroaromatic compounds in *Ralstonia eutropha* JMP134. *Arch Microbiol* 181:112–121
- Ledger T, Pieper DH, Gonzalez B (2006) Chlorophenol hydroxylases encoded by plasmid pJP4 differentially contribute to chlorophenoxyacetic acid degradation. *Appl Environ Microbiol* 72:2783–2792
- Leveau JH, van der Meer JR (1996) The *tfdR* gene product can successfully take over the role of the insertion element-inactivated TfdT protein as a transcriptional activator of the *tfdCDEF* gene cluster, which encodes chlorocatechol degradation in *Ralstonia eutropha* JMP134 (pJP4). *J Bacteriol* 178:6824–6832
- Li W, Shi J, Wang X, Han Y, Tong W, Ma L, Liu B, Cai B (2004) Complete nucleotide sequence and organization of the naphthalene catabolic plasmid pND6-1 from *Pseudomonas* sp. strain ND6. *Gene* 336:231–240
- Lloyd-Jones G, de Jong G, Ogden RC, Duetz WA, Williams PA (1994) Recombination of the *bph* (biphenyl) catabolic genes from plasmid pWW100 and their deletion during growth on benzoate. *Appl Environ Microbiol* 60:691–696
- Maeda K, Nojiri H, Shintani M, Yoshida T, Habe H, Omori T (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. *J Mol Biol* 326:21–33
- Mahenthalingam E, Urban TA, Goldberg JB (2005) The multifarious, multireplicon *Burkholderia cepacia* complex. *Nat Rev Microbiol* 3:144–156

- McLeod MP, Warren RL, Hsiao WW et-al. (2006) The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci USA* 103:15582–15587
- Mondello FJ (1989) Cloning and expression in *Escherichia coli* of *Pseudomonas* strain LB400 genes encoding polychlorinated biphenyl degradation. *J Bacteriol* 171:1725–1732
- Nelson MJ, Montgomery SO, Mahaffey WR, Pritchard PH (1987) Biodegradation of trichloroethylene and involvement of an aromatic biodegradative pathway. *Appl Environ Microbiol* 53:949–954
- Nojiri H, Shintani M, Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl Microbiol Biotechnol* 64:154–174
- Nojiri N, Sota M, Shintani M (2008) Catabolic plasmids involved in the degradation of polycyclic aromatic hydrocarbons and heteroaromatic compounds. doi:10.1007/7171_2008_9
- O’Sullivan LA, Weightman AJ, Jones TH, Marchbank AM, Tiedje JM, Mahenthalingam E (2007) Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environ Microbiol* 9:1017–1034
- Osborn AM, Boltner D (2002) When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 48:202–212
- Parschat K, Overhage J, Strittmatter AW, Henne A, Gottschalk G, Fetzner S (2007) Complete nucleotide sequence of the 113-kilobase linear catabolic plasmid pAL1 of *Arthrobacter nitroguajacolicus* Ru61a and transcriptional analysis of genes involved in quinaldine degradation. *J Bacteriol* 189:3855–3867
- Plumeier I, Perez-Pantoja D, Heim S, Gonzalez B, Pieper DH (2002) Importance of different *tfd* genes for degradation of chloroaromatics by *Ralstonia eutropha* JMP134. *J Bacteriol* 184:4054–4064
- Poh RP, Smith AR, Bruce IJ (2002) Complete characterisation of Tn5530 from *Burkholderia cepacia* strain 2a (pIJB1) and studies of 2,4-dichlorophenoxyacetate uptake by the organism. *Plasmid* 48:1–12
- Ramos JL, Marques S, Timmis KN (1997) Transcriptional control of the *Pseudomonas* TOL plasmid catabolic operons is achieved through an interplay of host factors and plasmid-encoded regulators. *Annu Rev Microbiol* 51: 341–373
- Rheinwald JG, Chakrabarty AM, Gunsalus IC (1973) A transmissible plasmid controlling camphor oxidation in *Pseudomonas putida*. *Proc Natl Acad Sci USA* 70:885–889
- Romine MF, Stillwell LC, Wong KK et-al. (1999) Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *J Bacteriol* 181:1585–1602
- Schlomann M (1994) Evolution of chlorocatechol catabolic pathways. Conclusions to be drawn from comparisons of lactone hydrolases. *Biodegradation* 5:301–321
- Shaw LE, Williams PA (1988) Physical and functional mapping of two cointegrate plasmids derived from RP4 and TOL plasmid pDK1. *J Gen Microbiol* 134:2463–2474
- Shields MS, Montgomery SO, Chapman PJ, Cuskey SM, Pritchard PH (1989) Novel pathway of toluene catabolism in the trichloroethylene-degrading bacterium G4. *Appl Environ Microbiol* 55:1624–1629
- Shields MS, Reagin MJ, Gerger RR, Campbell R, Somerville C (1995) TOM, a new aromatic degradative plasmid from *Burkholderia (Pseudomonas) cepacia* G4. *Appl Environ Microbiol* 61:1352–1356
- Shingler V (2003) Integrated regulation in response to aromatic compounds: from signal sensing to attractive behaviour. *Environ Microbiol* 5:1226–1241
- Shingler V, Bartilson M, Moore T (1993) Cloning and nucleotide sequence of the gene encoding the positive regulator (DmpR) of the phenol catabolic pathway encoded by pVI150 and identification of DmpR as a member of the NtrC family of transcriptional activators. *J Bacteriol* 175:1596–1604
- Shingler V, Franklin FCH, Tsuda M, Holroyd D, Bagdasarian M (1989) Molecular analysis of a plasmid-encoded phenol hydroxylase from *Pseudomonas* CF600. *J General Microbiol* 135:1083–1092

- Shingler V, Moore T (1994) Sensing of aromatic compounds by the DmpR transcriptional activator of phenol-catabolizing *Pseudomonas* sp. strain CF600. *J Bacteriol* 176:1555–1560
- Shingler V, Pavel H (1995) Direct regulation of the ATPase activity of the transcriptional activator DmpR by aromatic compounds. *Mol Microbiol* 17:505–513
- Shingler V, Powlowski J, Marklund U (1992) Nucleotide sequence and functional analysis of the complete phenol/3,4-dimethylphenol catabolic pathway of *Pseudomonas* sp. strain CF600. *J Bacteriol* 174:711–724
- Shintani M, Urata M, Inoue K, Eto K, Omori T, Yamane H, Nojiri H (2007) The *Sphingomonas* plasmid pCAR3 is involved in complete mineralization of carbazole. *J Bacteriol* 189:2007–2020
- Stecker C, Johann A, Herzberg C, Averbhoff B, Gottschalk G (2003) Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J Bacteriol* 185:5269–5274
- Sze CC, Shingler V (1999) The alarmone (p)ppGpp mediates physiological-responsive control at the sigma 54-dependent Po promoter. *Mol Microbiol* 31:1217–1228
- Sze CC, Laurie AD, Shingler V (2001) In vivo and in vitro effects of integration host factor at the DmpR-regulated sigma (54)-dependent Po promoter. *J Bacteriol* 183:2842–2851
- Zylstra GJ, McCombie WR, Gibson DT, Finette BA (1988) Toluene degradation by *Pseudomonas putida* F1: genetic organization of the *tod* operon. *Appl Environ Microbiol* 54:1498–1503
- Tan HM, Mason JR (1990) Cloning and expression of the plasmid-encoded benzene dioxygenase genes from *Pseudomonas putida* ML2. *FEMS Microbiol Lett* 60:259–264
- Tan HM, Fong KP (1993) Molecular analysis of the plasmid-borne *bed* gene cluster from *Pseudomonas putida* ML2 and cloning of the *cis*-benzene dihydrodiol dehydrogenase gene. *Can J Microbiol* 39:357–362
- Thomas CM (2000) Paradigms of plasmid organization. *Mol Microbiol* 37:485–491
- Thomas CM, Smalla K (2000) Trawling the horizontal gene pool. *Microbiol Today* 27:24–27
- Top EM, Moenne-Loccoz Y, Pembroke T, Thomas CM (2000) Phenotypic traits conferred by plasmids. In: Thomas CM (ed) *The horizontal gene pool: bacterial plasmid and gene spread*. Harwood, New York, pp 249–285
- Toussaint A, Merlin C, Monchy S, Benotmane MA, Leplae R, Mergeay M, Springael D (2003) The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl Environ Microbiol* 69:4837–4845
- Trefault N, De la Iglesia R, Molina AM et al. (2004) Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. *Environ Microbiol* 6:655–668
- Tsuda M, Genka H (2001) Identification and characterization of Tn4656, a novel class II transposon carrying a set of toluene-degrading genes from TOL plasmid pWW53. *J Bacteriol* 183:6215–6224
- Vallaeyts T, Albino L, Soulas G, Wright AD, Weightmann AJ (1998) Isolation and characterization of a stable 2,4-dichlorophenoxyacetic acid degrading bacterium, *Variovorax paradoxus*, using chemostat culture. *Biotechnol Lett* 20:1073–1076
- van der Meer JR, Zehnder AJ, de Vos WM (1991a) Identification of a novel composite transposable element, Tn5280, carrying chlorobenzene dioxygenase genes of *Pseudomonas* sp. strain P51. *J Bacteriol* 173:7077–7083
- van der Meer JR, van Neerven AR, de Vries EJ, de Vos WM, Zehnder AJ (1991b) Cloning and characterization of plasmid-encoded genes for the degradation of 1,2-dichloro-, 1,4-dichloro-, and 1,2,4-trichlorobenzene of *Pseudomonas* sp. strain P51. *J Bacteriol* 173:6–15
- Vedler E, Vahter M, Heinaru A (2004) The completely sequenced plasmid pEST4011 contains a novel IncP1 backbone and a catabolic transposon harboring *tfd* genes for 2,4-dichlorophenoxyacetic acid degradation. *J Bacteriol* 186:7161–7174
- Williams PA, Sayers JR (1994) The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*. *Biodegradation* 5: 195–217

- Winstanley C, Taylor SC, Williams PA (1987) pWW174: a large plasmid from *Acinetobacter calcoaceticus* encoding benzene catabolism by the beta-ketoadipate pathway. *Mol Microbiol* 1:219–227
- Xia XS, Aathithan S, Oswiecimska K, Smith AR, Bruce IJ (1998) A novel plasmid pIJB1 possessing a putative 2,4-dichlorophenoxyacetate degradative transposon Tn5530 in *Burkholderia cepacia* strain 2a. *Plasmid* 39:154–159
- Yano H, Garruto CE, Sota M et-al. (2007) Complete sequence determination combined with analysis of transposition/site-specific recombination events to explain genetic organization of IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J Mol Biol* 369:11–26
- Yano K, Nishi T (1980) pKJ1, a naturally occurring conjugative plasmid coding for toluene degradation and resistance to streptomycin and sulfonamides. *J Bacteriol* 143:552–560

Catabolic Plasmids Involved in the Degradation of Polycyclic Aromatic Hydrocarbons and Heteroaromatic Compounds

Hideaki Nojiri(✉), Masahiro Sota, and Masaki Shintani

Contents

Introduction	56
1 PAH Degradation	57
1.1 Bacteria with PAH Degradation Ability	57
1.2 Naphthalene-degradative Plasmid NAH7	58
1.2.1 Overview of the Genetic Structure of NAH7	60
1.2.2 Backbone Functions	61
1.2.3 Genes for Naphthalene Degradation	62
1.2.4 Naphthalene Catabolic Transposon Tn4655	63
1.3 Other PAH-degradative Plasmids	65
2 Heteroaromatic Compounds	67
2.1 Isolation and Genetic Analyses of Carbazole-degrading Bacteria	67
2.2 Carbazole-degradative Plasmid pCAR1 in Pseudomonads	70
2.2.1 Overview of the Genetic Structure of pCAR1	70
2.2.2 Replication and Maintenance Functions	72
2.2.3 Transfer Function and Behavior in the Natural Environment	72
2.2.4 Crosstalk between the Host Chromosome and pCAR1	73
2.3 Carbazole-degradative Plasmid pCAR3 in Sphingomonads	77
2.3.1 Overview of the Genetic Structure of pCAR3	77
2.3.2 Carbazole Degradation System in pCAR3: A Sphingomonad-specific <i>car</i> Gene Cluster	79
2.3.3 Other Catabolic Genes on pCAR3	79
2.3.4 Basic Plasmid Core Function	80
3 Concluding Remarks	80
References	81

H. Nojiri and M. Shintani
Biotechnology Research Center, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113-8657, Japan
e-mail: anojiri@mail.ecc.u-tokyo.ac.jp

M. Sota
Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844-3051, USA

Abstract Catabolic plasmids containing xenobiotic-degradative genes play an important role in distributing catabolic functions in nature. In the case of polycyclic aromatic compounds (PAHs) and heteroaromatic compounds, many degrading bacteria have a large catabolic plasmid that contains the genes for all or most of the mineralization pathways of these compounds. Although the list of catabolic plasmids for PAHs and heteroaromatic compounds is increasing continuously, in this chapter, we focus on three extensively investigated catabolic plasmids: naphthalene catabolic plasmid NAH7 from *Pseudomonas putida* G7, carbazole catabolic plasmid pCAR1 from *Pseudomonas resinovorans* CA10, and carbazole catabolic plasmid pCAR3 from *Novosphingobium* sp. KA1. We summarize biochemical investigations to elucidate the xenobiotic degrading functions and basic plasmid functions (e.g., incompatibility, replication, transfer). In addition, complete nucleotide sequences of these three catabolic plasmids have been determined and have provided important and interesting clues to understanding plasmid functions. We discuss the genomic features of these plasmids in conjunction with the results of physiological investigations.

Introduction

Industrial activity over the past century has resulted in the release of quantities of xenobiotic chemicals into the ecosphere. Although some xenobiotics are highly recalcitrant, metabolically versatile microbial populations have adapted flexibly to a wide range of such recalcitrant xenobiotics, regardless of origin, resulting in the development of microbial degradation ability. Although adequate time is sometimes necessary to detect the degradation, almost all natural and synthetic aromatic compounds seem to be mineralized or partially converted. This unique diversity of bacterial functions depends on the diversity of bacterial genomes, which have been built by the recruitment of DNA from other organisms (or environments), followed by small local changes in the nucleotide sequences of the genome and intragenomic reshuffling of segments of genomic sequences.

In addition to a number of nongenetic factors, various specific gene(s) or gene product(s) are involved in the generation of genetic variation and in the modulation of the frequency of genetic variation. As the mediators of genetic exchange between bacterial strains, mobile genetic elements such as transferable plasmids, (conjugative) transposons, integrons, genomic islands, or phages, which are all able to move within and/or between genomes, are thought to play an important role. Although it is quite difficult to quantitatively estimate their respective involvement in the diversification of the bacterial genome, transferable plasmids are believed to be the most important because of their high frequency of horizontal transfer and their large size, which allow them to transfer many genes. In fact, a variety of xenobiotic-degradative genes have been found on plasmids (Top et al. 2000; Nojiri et al. 2004; Ogawa et al. 2004).

The first plasmids involved in the degradation of organic xenobiotics (termed *catabolic plasmids*) were described approximately 35 years ago; since then, the list has been

increasing continuously. Recent genomic approaches have enhanced the understanding of plasmid structure, evolution, and occurrence. Plasmids have various mobile genetic elements. Catabolic plasmids are commonly rather large (>50 kb), and genetic analyses have sometimes shown that such plasmids contain complete sets of catabolic genes for the conversion of xenobiotic aromatics to biotic compounds. Thus far, more than 100 catabolic plasmids have been identified from various bacteria, but only a restricted number of these have been investigated in detail. For example, some of the catabolic plasmids from *Pseudomonas* strains have been analyzed and characterized in detail and classified into four incompatibility (Inc) groups: IncP-1, IncP-2, IncP-7, and IncP-9. For several *Pseudomonas* catabolic plasmids, in addition to detailed analyses of degradation pathway-specific phenomena such as transcriptional regulation and catabolite repression, the determination of the complete nucleotide sequences has also been undertaken. These results have provided basic information on the plasmid function as a form of machinery for genetic exchange across taxa. There is not as much information on catabolic plasmids from sphingomonads and Gram-positive strains as there is on *Pseudomonas* catabolic plasmids.

In this chapter, we summarize the information on catabolic plasmids involved in the degradation of polycyclic aromatic hydrocarbon (PAH) and heterocyclic aromatic compounds, mainly carbazole.

1 PAH Degradation

1.1 *Bacteria with PAH Degradation Ability*

PAHs are now ubiquitous in nature because of human activities, especially in areas contaminated with petroleum, which is a complex mixture of hydrocarbons and other organic compounds. These compounds confer acute toxicity and are also mutagenic and/or carcinogenic to living creatures (Mastrangelo et al. 1996; Marston et al. 2001; Xue and Warshawsky 2005). However, a variety of microbial species can use PAHs as a sole source of carbon and energy (Kanaly and Harayama 2000; van Hamme et al. 2003). Several bacteria that can decompose PAHs such as naphthalene (two rings), phenanthrene (three), anthracene (three), pyrene (four), and chrysene (four) have been reported: Gram-negative bacteria such as *Burkholderia* (Laurie and Lloyd-Jones 1999), *Comamonas* (Goyal and Zylstra 1997), *Novosphingobium* (Romine et al. 1999), *Pseudomonas* (Stuart-Keil et al. 1998; Bosch et al. 1999; Park et al. 2003; Ma et al. 2006), *Ralstonia* (Zhou et al. 2002), *Sphingobium* (Kim and Zylstra 1999; Liu et al. 2004; Basta et al. 2005), and Gram-positive bacteria such as *Bacillus* (Kazunga and Aitken 2000), *Mycobacterium* (Kim et al. 2006), *Rhodococcus* (Uz et al. 2000; Dean-Ross et al. 2001; Kulakov et al. 2005; Kimura et al. 2006), and *Staphylococcus* (Mallick et al. 2007). Although no strains have been found that use high-molecular-weight PAHs (five or more rings) as a sole source of carbon and energy, some strains such as *Mycobacterium*

vanbaalenii PYR-1 and *Sphingomonas yanoikuyae* JAR02 can degrade the five-ring PAH benzo[*a*]pyrene in the presence of other energy sources (Moody et al. 2004; Rentz et al. 2008).

Aerobic degradation pathways of PAHs have been studied extensively. Usually, the initial step in PAH metabolism is to introduce oxygen atoms into an aromatic ring, leading to the production of dihydrodiol compounds. This reaction is catalyzed by a multicomponent enzyme, the so-called aromatic ring hydroxylating dioxygenase or Rieske non-heme iron oxygenase, which is composed of a reductase, ferredoxin, and iron–sulfur protein (Harayama et al. 1992). The dioxygenases are considered to be key enzymes for the degradation of PAHs, heterocyclic aromatic hydrocarbons, and their chlorinated derivatives. A variety of dioxygenases that have different substrate specificities have been identified and characterized. The ring of a PAH that the enzymes initially attack determines the resulting metabolic cascade of the PAH (Kanaly and Harayama 2000; Moody et al. 2004; Rentz et al. 2008). The intermediates thus formed by the initial dioxygenation are usually converted to the central intermediates of the aromatic compound degradation, e.g., catechols and protocatechuates, which can be further metabolized to merge into the TCA cycle. Although limited information is available to describe the metabolic cascades of PAHs with more than three rings (Kanaly and Harayama 2000; Moody et al. 2004; Rentz et al. 2008), detailed characterization for naphthalene, anthracene, and phenanthrene, as well as genetic and enzymatic features associated with the degradation, are available. Entire sets of genes for the degradation of these PAHs are often carried by large plasmids (Table 1), and similar genes and/or gene organization have been identified in PAH degradation gene clusters from different sources, indicating the wide dissemination of PAH degradation genes by these plasmids in bacterial populations.

1.2 Naphthalene-degradative Plasmid NAH7

Catabolic plasmids such as SAL (for salicylate), TOL (for toluene), and CAM (for camphor) were discovered in the 1970s. A self-transferable plasmid with the genes required for naphthalene degradation from *P. putida* G7, designated NAH7, was reported in 1973 by Dunn and Gunsalus (1973). NAH7 was assigned to the IncP-9 incompatibility group of plasmids from *Pseudomonas* strains on the basis of its incompatibility with another IncP-9 plasmid (White and Dunn 1978) and its close relationship with the TOL plasmid pWW0 from *P. putida* mt-2 (Benson and Shapiro 1978; Greated et al. 2002) for both the backbone and catabolic genes (Lehrbach et al. 1983; Harayama et al. 1987). Since its discovery, the biochemistry and genetics of the degradation pathway encoded on NAH7, as well as those on pWW0, have been thoroughly investigated by many researchers. The information thus obtained has provided the foundation for studies of PAH degradation. However, only recently have the complete nucleotide sequences of these plasmids been determined (Greated et al. 2002; Sota et al. 2006).

Table 1 PAH degradative plasmids

Plasmids	Degradable compounds	Size (kb)	Inc group	Catabolic genes	Transfer	Source	References
NAH7	Naphthalene, salicylate	82	P-9	<i>nah</i>	+	<i>Pseudomonas putida</i> G7	Sota et al. (2006)
NPL-1	Naphthalene, salicylate	100	P-9	<i>nah</i>	ND	<i>P. putida</i> BS202	Izmailkova et al. (2006)
P8C	Phenanthrene, naphthalene, salicylate	110	P-9	ND	ND	<i>P. putida</i> 8C	Izmailkova et al. (2006)
pCg1	Naphthalene, salicylate	88	P-9	<i>nah</i>	+	<i>P. putida</i> Cg1	Park et al. (2003)
pDTG1	Naphthalene, salicylate	83	P-9	<i>nah</i>	+	<i>P. putida</i> NCIB 9816-4	Dennis and Zylstra (2004)
pFKY1	Naphthalene, salicylate	200	P-9	<i>nah</i>	+	Oil-contaminated soil ^a	Ono et al. (2007)
pFKY4	Naphthalene, salicylate	80	P-9	<i>nah</i>	+	Oil-contaminated soil ^a	Ono et al. (2007)
pKA1	Naphthalene, salicylate	100	NA	ND	+	<i>P. fluorescens</i> 5R	Sanseverino et al. (1993)
pLP6a	Phenanthrene, anthracene, naphthalene, salicylate	63	NA	<i>nah</i>	ND	<i>P. fluorescens</i> LP6a	Foght and Westlake (1996)
pND6-1	Naphthalene, salicylate	102	P-7	<i>nah</i>	-	<i>Pseudomonas</i> sp. ND6	Li et al. (2004)
pNLI	Naphthalene, salicylate, toluene, xylene, biphenyl	184	NA	<i>nah</i> , <i>bph</i> , <i>xyl</i> , <i>pch</i>	+	<i>Novosphingobium aromaticivorans</i> F199	Romine et al. (1999)
pNUO1	Naphthalene	> 750	NA	ND	ND	<i>Rhodococcus opacus</i> M213	Uz et al. (2000)
pPHN	Phenanthrene, salicylate	112	NA	ND	ND	<i>Staphylococcus</i> sp. PN/Y	Mallick et al. (2007)
pWK301	Naphthalene, dibenzofuran, dibenzo- <i>p</i> -dioxin	1,100	NA	ND	ND	<i>R. opacus</i> SAO101	Kimura et al. (2006)
pZL	Phenanthrene, salicylate	60	NA	ND	ND	<i>Sphingomonas</i> sp. ZL5	Liu et al. (2004)
NA	Naphthalene, salicylate	> 250	NA	<i>Nag</i>	+	<i>Ralstonia</i> sp. U2	Zhou et al. (2002)
NA	Naphthalene, salicylate	75	NA	<i>Dox</i>	ND	<i>Pseudomonas</i> sp. C18	Denome et al. (1993)

ND Not determined, NA Not assigned

^aIsolated from unknown host using the exogenous isolation method

1.2.1 Overview of the Genetic Structure of NAH7

Plasmid NAH7 is 82,232 bp in size, with an average G + C content of 55.8% (Fig. 1). These values are lower than those for chromosomes of *Pseudomonas* strains such as *P. aeruginosa* PAO1 (67.1%) and *P. putida* KT2440 (62.3%), although IncP-9 plasmids have been identified only from *Pseudomonas* strains. NAH7 has a total of 84 open reading frames (ORFs), of which 59 and 25 encode proteins with known and unknown functions, respectively. Among the former 59 ORFs, 23 are involved in naphthalene degradation (*nah*) (see below), and other ORFs are predicted to function

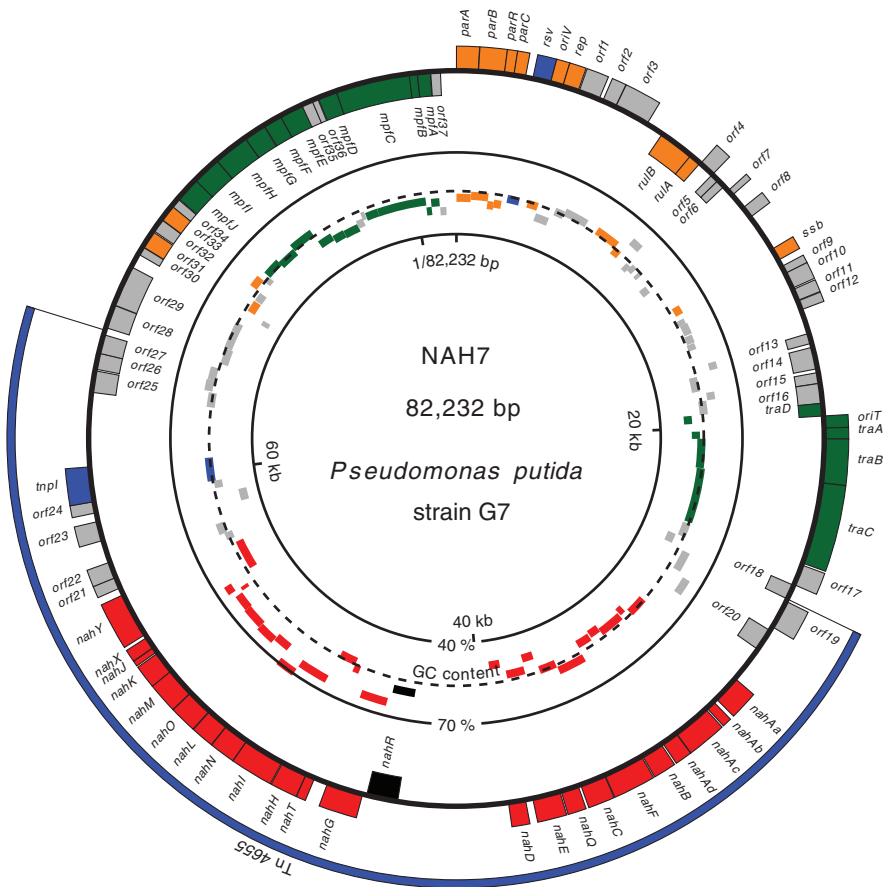


Fig. 1 The circular gene map of NAH7 (Sota et al. 2006). Genes or ORFs outside the circle are oriented clockwise; those inside are oriented counterclockwise. The (putative) functions of genes or ORFs are shown in color as follows: orange, maintenance or DNA processing; dark green, conjugative transfer; blue, transposition or integration; red, degradation; black, regulation; gray, unknown function (homologous to hypothetical protein). Bars inside the circular gene map indicate the G + C contents of ORFs shown in the same color. The broken circle indicates the average G + C content of the entire NAH7 (55.8%)

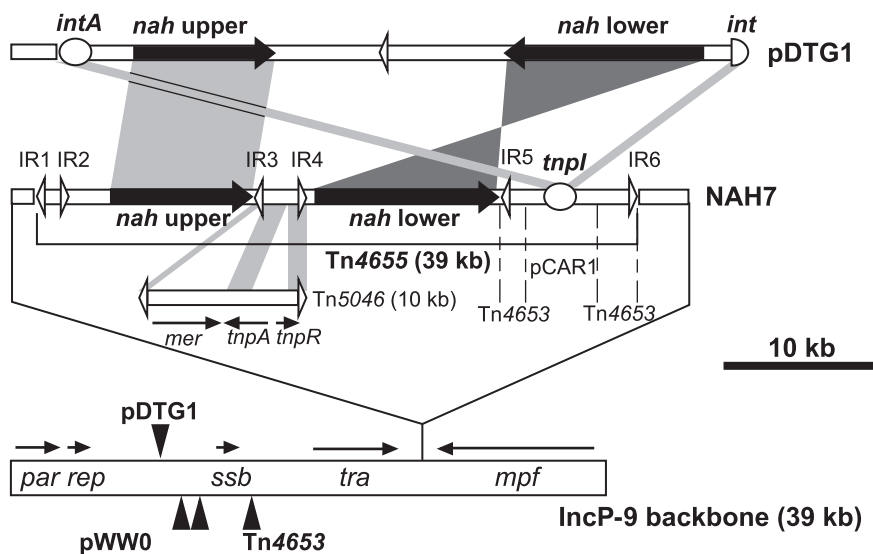


Fig. 2 Mosaic structure of NAH7 and comparison of insertions on NAH7 and pDTG1. The 39-kb IncP-9 backbone is shown at the bottom, and the locations of insertions on pDTG1 and pWW0 are indicated by *black arrowheads*. *Horizontal arrows* indicate the transcriptional direction of the indicated genes or gene clusters. *White arrowheads* indicate the IR and IR-like sequences of the class II transposons. Ovals and a half oval indicate the intact (*intA*, *tnpI*) and truncated (*int*) integrase genes, respectively. Homologous regions are indicated by light (same strand) and dark (different strand) shading. Regions similar to pCAR1 and Tn4653 are shown. The gene functions are as follows: *par*, plasmid partitioning; *rep*, plasmid replication; *ssb*, single-strand binding protein; *tra* and *mpf*; plasmid transfer; *tnpA* and *tnpR*, transposition and resolution; *mer*, resistance to mercury

in plasmid replication, maintenance, and transfer (e.g., *rep*, *par*, *tra*, *mpf*). Little is known about the functions of most NAH7 ORFs, except for *nah* and *rep*. A total of 36.4 kb of NAH7 are conserved in the IncP-9 plasmids pDTG1 (Dennis and Zylstra 2004) and pWW0 (Greated et al. 2002), and detailed comparisons of the three plasmids have allowed researchers to define the 39.0-kb sequence of NAH7 as the IncP-9 backbone (Fig. 2). The backbone of NAH7 has 94% and 70% identity with that of pDTG1 and pWW0, respectively, clearly indicating the closer phylogenetic relationship of NAH7 to pDTG1 than to pWW0. The comparison also indicated that IncP-9 plasmids have evolved by the independent acquisition of catabolic gene clusters at different locations in the backbone (Fig. 2) (Dennis 2005; Sota et al. 2006).

1.2.2 Backbone Functions

IncP-9 plasmids are especially common in *Pseudomonas* strains, but have a broad host range (Jacoby 1986). However, little is known of the replication machinery or of the transfer and maintenance of IncP-9 plasmids. It was approximately 30 years after the discovery of IncP-9 plasmids that a molecular analysis of the IncP-9

replication system was first reported by a British group. Greated et al. (2000) investigated the replication and stable inheritance of an IncP-9 mini-replicon, pMT2, a derivative of the resistance plasmid pM3. They identified the genes for replication (*rep*), mating pair formation (*mpf*), and stable inheritance (*par* and *mrs*) in pMT2. The expression of the *rep* gene was controlled by negative regulation through the Rep protein itself. The presence of *rep* and *oriV* (origin of vegetative replication) is sufficient for autonomous replication of the IncP-9 plasmid in *Pseudomonas*, but not in *Escherichia coli* (Sevastyanovich et al. 2005). An insertion between the *rep* and *par* genes strongly inhibits the replication of pMT2 in *E. coli*, suggesting the close interaction of these gene products. This was confirmed a few years later by the same group using experiments that demonstrated that ParB and its target, the *par* operon promoter, are required for replication in *E. coli*. It was proposed that ParB binding to the target promoter changes *rep* expression sufficient to activate *oriV*; this proposal is consistent with the fact that increased Rep production allows the replication of the *rep-oriV* minimal replicon (Krasowiak et al. 2006).

The functions of genes for the IncP-9 conjugation system were predicted from sequence similarities to those of well-studied groups of plasmids such as IncW and IncP-1. Interestingly, the genetic organization of the IncP-9 conjugation system is similar to that of the IncN and IncW plasmids pKM101 and R388, respectively (Fernandez-Lopez et al. 2006), suggesting that these conjugation systems might have evolved from the same origin. Four *tra* genes (*traA* to *traD*) and 12 *mpf* genes (*mpfR* and *mpfA* to *mpfJ*) have been identified in IncP-9 plasmids (Fig. 1). The IncP-9 transfer genes are transcribed as at least three units: *traABC*, *traD*, and *mpfR* to *mpfJ* (Fig. 1) (Lambertsen et al. 2004). Transcription of the former two units is repressed by TraA, and that of *mpf* genes is regulated by MpfR, which is not related to any known regulatory protein of plasmid transfer systems. Because all the *mpf* genes and other unknown ORFs are organized as an operon in all three IncP-9 plasmids that have been completely sequenced thus far (Greated et al. 2002; Dennis and Zylstra 2004; Dennis 2005; Sota et al. 2006), the products of these unknown ORFs may also be involved in DNA transfer functions. Detailed functions of the IncP-9 conjugative transfer systems remain to be investigated.

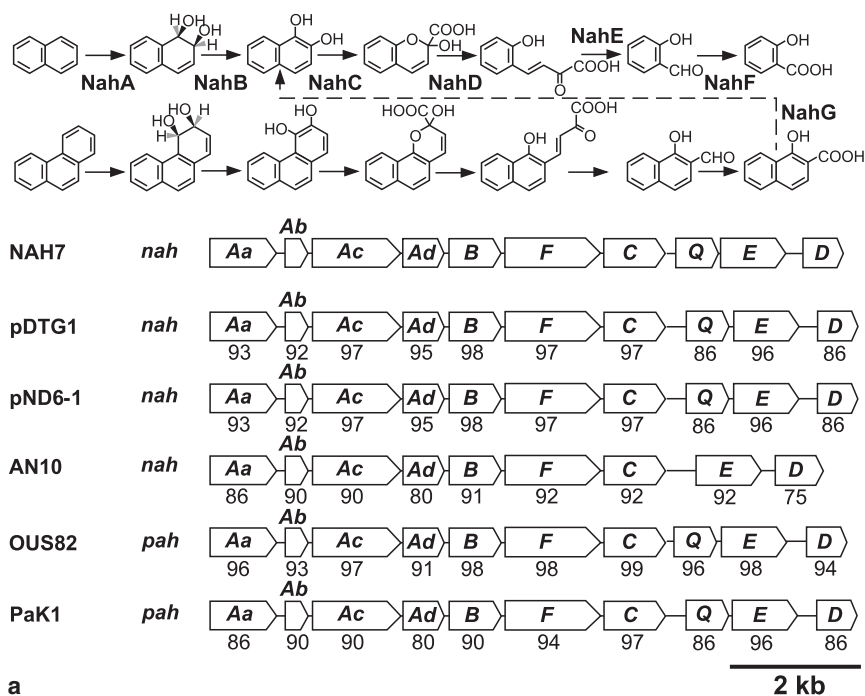
1.2.3 Genes for Naphthalene Degradation

The *nah* genes in NAH7 are organized in two major segments: the upper and lower pathways. The gene products that are encoded by the upper pathway operon (*nahAa* to *nahD*) convert naphthalene to salicylate: NahAa, naphthalene 1,2-dioxygenase reductase; NahAb, naphthalene 1,2-dioxygenase ferredoxin; NahAc, naphthalene dioxygenase large subunit; NahAd, naphthalene dioxygenase small subunit; NahB, naphthalene *cis*-dihydrodiol dehydrogenase; NahF, salicylaldehyde dehydrogenase; NahC, 1,2-dihydroxynaphthalene dioxygenase; NahE, hydroxybenzalpyruvate hydratase-aldolase; and NahD, 2-hydroxychromene-2-carboxylate dehydrogenase (Fig. 3a) (Yen and Serdar 1988). The enzymes that are encoded by the lower *nah* operon (*nahG* to *nahY*) convert salicylate to pyruvate and acetaldehyde via the catechol

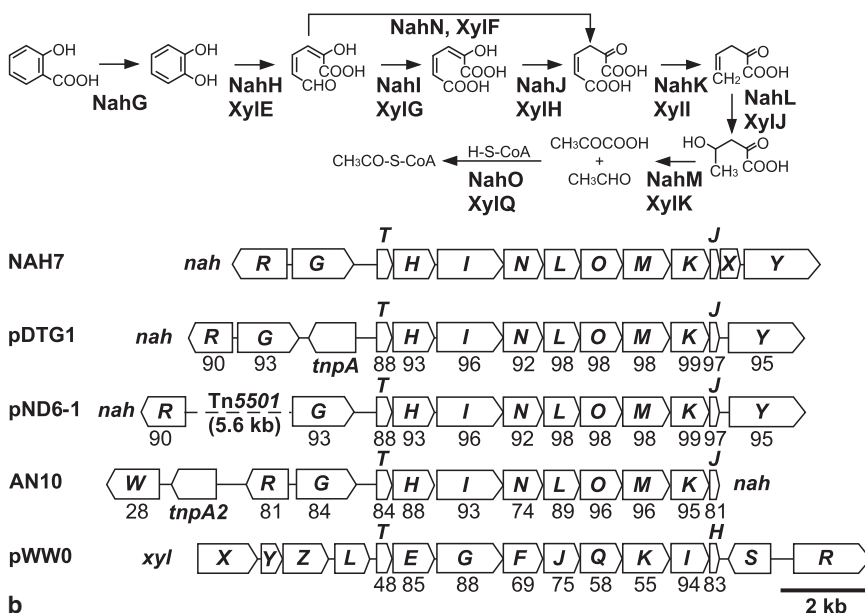
meta-cleavage pathway: NahG, salicylate hydroxylase; NahT, chloroplast-type ferredoxin; NahH, catechol 2,3-dioxygenase; NahI, hydroxymuconic semialdehyde dehydrogenase; NahN, hydroxymuconic semialdehyde hydrolase; NahL, 2-oxopent-4-enoate hydratase; NahO, acetaldehyde dehydrogenase; NahM, 2-oxo-4-hydroxypentanoate aldolase; NahK, 4-oxalocrotonate decarboxylase; and NahJ, oxalocrotonate tautomerase (Fig. 3b) (Yen and Serdar 1988). The *nahY* gene product that is encoded in the lower *nah* operon is involved in chemotaxis to naphthalene (Grimm and Harwood 1999), and the function of the *nahX* gene remains to be elucidated. The expression of the two operons is controlled by NahR, a LysR-type transcriptional regulator (Yen and Gunsalus 1985), which is much larger (374 residues) than other NahR proteins (300 residues). These extra residues in NahR of NAH7 might have been attached by a frameshift mutation near the 3' end of the *nahR* gene. The comparison of both operons on NAH7 with other plasmid- and chromosome-borne operons involved in the degradation of naphthalene and/or phenanthrene showed their high similarity in nucleotide sequences and gene organizations, indicating that these operons have the same origin (Fig. 3). This is consistent with the fact that the enzymes for naphthalene degradation can also usually metabolize phenanthrene (and occasionally anthracene) and *vice versa* (Fig. 3a) (Menn et al. 1993; Sanseverino et al. 1993; Goyal and Zylstra 1997). A noteworthy feature is the similarity of the lower *nah* operon of NAH7 to the *xylTEGFJQKIH* genes of the catechol *meta*-cleavage pathway used for toluene degradation (Fig. 3b). This strongly suggests that both naphthalene and toluene degradation pathways have evolved from a common ancestor: the naphthalene and toluene degradation cascades were established by the addition of the upper *nah* operon and *nahGR(X)Y* genes and the upper *xyl* operon and *xylRSXYZL* genes, respectively, to the common catechol *meta*-cleavage pathway distributed by horizontal gene transfer (Williams and Sayers 1994).

1.2.4 Naphthalene Catabolic Transposon Tn4655

Long before the NAH7 sequence became available, Tsuda and Iino (1990) demonstrated that all of the *nah* genes are carried on a large class II transposon (Fig. 1). The transposon, designated Tn4655, is bracketed by 38-bp inverted repeats (IRs) similar to those of the Tn1721-like transposons, but lacks its own transposition gene. The transposition of Tn4655 is restored when the transposase of a cognate (Tn1721-like) transposon is supplied *in trans*. However, Tn4655 has a site-specific resolution system to resolve a co-integrate of the donor and recipient molecules of the transposon, which is formed at the first step of the transposition. This resolution system is completely different from those of the Tn1721-like transposons. The sequence analysis of NAH7 and previous genetic examination of the site-specific resolution of Tn4655 revealed that this resolution system is mediated by an integrase, TnpI, and its recognition site, *attI*. Following detailed molecular analysis, the 119-bp *attI* sequence required for the integration/resolution reaction by TnpI was determined (Sota et al. 2006). Curiously, homologs of the *attI-orf24-tnpI* module of NAH7 have also been identified on the naphthalene-degradative plasmids pDTG1



a



b

Fig. 3 Genetic organization of PAH-degradative operons from *Pseudomonas* strains and catabolic pathways of naphthalene and phenanthrene mediated by NAH7-like gene products. *Nah*, *pah*, and *xyl* are genes for the degradation of naphthalene, phenanthrene, and toluene/xylene, respectively. The percent amino acid identity of the NAH7-specified gene products to the corresponding gene

(IncP-9) (Dennis and Zylstra 2004) and pND6-1 (IncP-7) (Li et al. 2004) and the carbazole-degradative plasmid pCAR1 (IncP-7, see below) (Maeda et al. 2003), but not on any molecules in the databases. This implies the specific function of the *attI-orf24-tnpI* module in incorporating catabolic genes into the IncP-7 and IncP-9 plasmids (see below).

Tn4655 has a total of six copies of IR-like sequences that show high similarity with the IRs of class II transposons (Fig. 2). Two of these (IR5 and IR6) are similar to those of Tn4653 on pWW0, and both ends of the intervening segment (9.6 kb) also show high similarity to the partial sequence of Tn4653 (> 90%) (Fig. 2). The remaining internal segment within the 9.6-kb region carries the *attI-orf24-tnpI* cluster and is > 85% identical to the corresponding sequence on pCAR1 (Fig. 2). The region bracketed by IR3 and IR4 resembles a mercury resistance class II transposon, Tn5046 (Mindlin et al. 2001). However, the *tnpA* and *tnpR* genes of the transposon are truncated, and no mercury resistance genes are present between IR3 and IR4. These observations suggest that Tn4655 was created after several rearrangements that were mediated by other transposons, including its close relative, Tn4653. Class II transposons generate a 5-bp duplication of their target sequences upon transposition (Sherratt 1989). Such a 5-bp duplication is found only just outside of IR1 and IR6, suggesting that the transposition of the entire Tn4655 transposon or its ancestor into an IncP-9 plasmid resulted in the generation of NAH7.

1.3 Other PAH-degradative Plasmids

Although many PAH-degradative plasmids have been reported (Table 1), only a few plasmids have been entirely sequenced. To date, complete sequences of four naphthalene-degradative plasmids, three from *Pseudomonas* (NAH7, pDTG1, and pND6-1) and one from *Novosphingobium* (pNL1), are available in the databases (Romine et al. 1999; Dennis and Zylstra 2004; Li et al. 2004; Sota et al. 2006). To our knowledge, no plasmid that encodes the degradation of PAHs that have four or more rings has been reported, although large plasmids have been identified in bacteria that degrade such PAHs (Moody et al. 2004; Basta et al. 2005). The sequences of two naphthalene-degrading plasmids, pDTG1 and pND6-1, were published in 2004. The respective plasmids belong to the IncP-9 and IncP-7 groups; therefore, their backbone sequences are entirely different. However, both plasmids have virtually



Fig. 3 (continued) products from the other operons is shown. **a** Upper pathways and operons. The Nah proteins can also catalyze phenanthrene degradation by the steps shown. **b** Lower (*meta*-cleavage) pathway and operons. The *tnpA* and *tnpA2* genes in pDTG1 and AN10, respectively, are ISPre1 derivatives. Tn5501 is a cryptic class II transposon identified in *P. putida* H. Accession numbers of the sequence data are as follows: NAH7, AB237655; pDTG1, AF491307; pND6-1, AY208917; *P. stutzeri* AN10, AF039533 (upper) and AF039534 (lower); *P. putida* OUS82, AB004059; *P. aeruginosa* PaK1, D84146; and pWW0, AJ344068. The genes in *P. stutzeri* AN10, *P. putida* OUS82, and *P. aeruginosa* PaK1 are carried on their chromosome

identical > 46-kb segments that contain all of the *nah* and pDTG1 have very similar sequences in both the backbone and the upper and lower *nah* operons, the *nah*-containing segments on these plasmids are considered to have different origins (Sota et al. 2006). This is consistent with the fact that the segments on NAH7 and pDTG1 are situated at sites located far apart in the IncP-9 backbone (Fig. 2) (Sota et al. 2006). In addition, there is no evidence to indicate that the *nah*-containing segments on pDTG1 and pND6-1 are carried on transposons or their remnants. These observations raised a question of how the large *nah*-carrying segments were incorporated into these plasmids and Tn4655. A possible explanation is that the TnpI integrase, which is responsible for the co-integrate resolution reaction of Tn4655, might have been associated with the integration of the *nah* genes into these plasmids for the following two reasons: (a) the TnpI-like integrases are found near catabolic genes in IncP-7 and IncP-9 plasmids, including the three naphthalene catabolic plasmids, and (b) the segment that contains all of the *nah* genes on pDTG1 is bracketed by two directly repeated copies of the *tnpI*-like gene (*intA* and *int*), although one (*int*) copy is truncated (Fig. 2). The latter observation is consistent with a site-specific integration event between two *att-tnpI*-containing fragments. Intriguingly, most of the plasmids that are presently known to metabolize PAHs and/or related aromatic compounds were identified in *Pseudomonas* strains, and most of them are classified as IncP-2, IncP-7, or IncP-9 plasmids (Top et al. 2000; Nojiri et al. 2004; Ogawa et al. 2004; Izmalkova et al. 2006; Ma et al. 2006), suggesting the important role of these groups of plasmids and strains in the degradation of these compounds.

The self-transmissible plasmid pNL1 from *Novosphingobium* (formerly *Sphingomonas*) *aromaticivorans* F199 is the first entirely sequenced catabolic plasmid from an α -proteobacterial strain (Romine et al. 1999). The plasmid carries the genes for the degradation of biphenyl, naphthalene, *m*-xylene, and *p*-cresol, which are distributed among 15 gene clusters. Three *nah* genes (*nahDEF*) have been identified in the plasmid, but are located at three separate loci and have quite low identity with the corresponding genes of *Pseudomonas* strains. This suggests that the *nah* genes in pNL1 derive from a different origin than the *Pseudomonas* genes. The complete sequence of another pNL1-like plasmid, pCAR3 from *Novosphingobium* sp. KA1, was published in 2007 (Shintani et al. 2007) (see below for details). Sphingomonad plasmids that contain genes for the degradation of various xenobiotics and polycyclic aromatic compounds have replication machinery similar to those of pNL1 (Basta et al. 2005), suggesting an important role of the pNL1-like plasmids in the dissemination of catabolic genes among the Sphingomonadaceae. This is supported by the fact that the naphthalene sulfonate catabolic genes on a 180-kb plasmid, pBN6 from *S. xenophaga* BN6, show the highest similarity to those on pNL1 or in *Sphingobium* sp. P2 and share the same organization in three different, noncontiguous regions as those on pNL1 or in *Sphingobium* sp. P2 (Pinyakong et al. 2003; Keck et al. 2006). Interestingly, in plasmids from α -Proteobacteria, including pNL1, pCAR3, and those that encode for symbiosis such as pSymA (Barnett et al. 2001), genes or gene clusters that work cooperatively are scattered. The reason for the scattered gene positioning is a mystery and of particular interest.

2 Heteroaromatic Compounds

The bacterial degradation of heteroaromatic compounds has been investigated in trials to develop biodegradation strategies for crude oil components and other toxic contaminants such as the herbicide atrazine. In particular, model compounds of crude oil components containing heteroatoms, e.g., dibenzothiophene, and dibenzofuran, have been used as substrates in screening for heteroaromatic degraders. Because dibenzo-*p*-dioxin and dibenzofuran are parental compounds of so-called dioxin, which is highly toxic and widely distributed, these compounds were also used to screen for dioxin-degrading strains. Carbazole has a planar structure similar to that of dioxin, and is degraded through a degradation pathway homologous to that of dibenzo-*p*-dioxin and dibenzofuran (Nojiri and Omori 2002). Therefore, carbazole was also used as a model substrate in the isolation of dioxin-degrading bacterial strains. Heteroaromatic compound-degradative plasmids have been found through such studies (Table 2). In this section, we focus on the two carbazole-degradative plasmids that are well investigated as catabolic plasmids and originate from *Pseudomonas* and *Novosphingobium* strains.

2.1 Isolation and Genetic Analyses of Carbazole-degrading Bacteria

Carbazole (dibenzopyrrole diphenylenimine, C₁₂H₉N) is used in the manufacture of many products, including dyes, reagents, explosives, insecticides, lubricants, and as a color inhibitor in detergents. However, carbazole is also an environmental pollutant, and there are environmental concerns regarding its release because it is both mutagenic and toxic (Arcos and Argus 1968). Even though carbazole itself is not highly toxic, it readily undergoes radical reactions to generate genotoxic hydroxynitrocarbazole. Among the carbazole degraders thus far isolated from soil, freshwater, and activated sludge samples, > 60% of the strains are classified as either *Pseudomonas* or *Sphingomonas* (Nojiri and Omori 2007).

Carbazole is degraded to anthranilic acid via an initial dioxygenation (specifically termed *angular dioxygenation*; Nojiri and Omori 2002, 2007), *meta*-cleavage, and hydrolysis (Fig. 4). Anthranilic acid is then converted to catechol. Catechol is metabolized through the β -ketoadipate pathway via *ortho*-cleavage in *Pseudomonas resinovorans* CA10; catechol metabolism via *meta*-cleavage has been also been reported (as reviewed in Nojiri and Omori 2007).

The *car* gene cluster that is involved in carbazole conversion to anthranilic acid was first isolated from *P. resinovorans* CA10, and then all structural genes were isolated and characterized (Fig. 5a) (Sato et al. 1997a, b; Nojiri et al. 2001, 2002b). Efforts to isolate *car* genes from various origins showed that most of the *car* genes could be divided into at least two types, which are mainly distributed in the genera *Pseudomonas* and *Sphingomonas* (Fig. 5; Inoue et al. 2004, 2005). Such genus-specific distribution of highly homologous *car* gene clusters suggests the presence of at least two mobile genetic elements involved in the horizontal transfer of *car*

Table 2 Catabolic plasmids for heteroaromatic compound

Plasmids	Degradable compounds	Source	Size (kb)	Inc group	Catabolic genes	Transfer	References
NIC	Nicotine, nicotine	<i>Pseudomonas convexa</i> Pc1	ND ^a	ND ^a	ND ^a	+	Thacker et al. (1978)
pADP-1	Atrazine	<i>Pseudomonas</i> sp. ADP	109	P-1β	<i>atzABCDEF</i>	+	de Souza et al. (1998), Martinez et al. (2001)
pCAR1	Carbazole	<i>P. resinovorans</i> CA10	199	P-7	<i>car, ant</i>	+	Nojiri et al. (2001), Maeda et al. (2003)
pCAR2 ^b	Carbazole	<i>P. putida</i> HS01	200	P-7	<i>car, ant</i>	+	Shintani et al. (2005a, b)
pCAR3	Carbazole	<i>Sphingomonas</i> sp. KA1	255	ND ^a	<i>car-I, car-II, fdr, fdx, fnu/dbf, and, lig</i>	-	Habe et al. (2002), Shintani et al. (2007)
pDBF1	Dibenzofuran, fluorene	<i>Terrabacter</i> sp. DBF63	160	ND ^a	<i>fnu/dbf, phi pca</i>	-	Nojiri et al. (2002a); Habe et al. (2003–2005)
pSWIT02	Dibenzo- <i>p</i> -dioxin, dibenzofuran	<i>Sphingomonas wittichii</i> RW1	220	ND ^a	<i>dsnA1A2</i>	ND ^a	Basta et al. (2004)
pYK3	Dibenzofuran	<i>Terrabacter</i> sp. YK3	ND ^a	ND ^a	<i>dfdA1A2A3A4</i>	ND ^a	Iida et al. (2002)
N/A ^c	Dibenzofuran	<i>Sphingomonas</i> sp. HH69	240	ND ^a	<i>dsnA1A2</i> homolog	+	Basta et al. (2004)

^aNot determined or not available^bThe differences between pCAR1 and pCAR2 are a single point mutation within *tmpA1* gene encoding transposase of IS*Pre1* and the presence of a fourth copy of IS*Pre1* at the adjacent position of IS*Pre4* (Takahashi et al. unpublished results), suggesting that pCAR1 and pCAR2 are virtually the same^cNot assigned

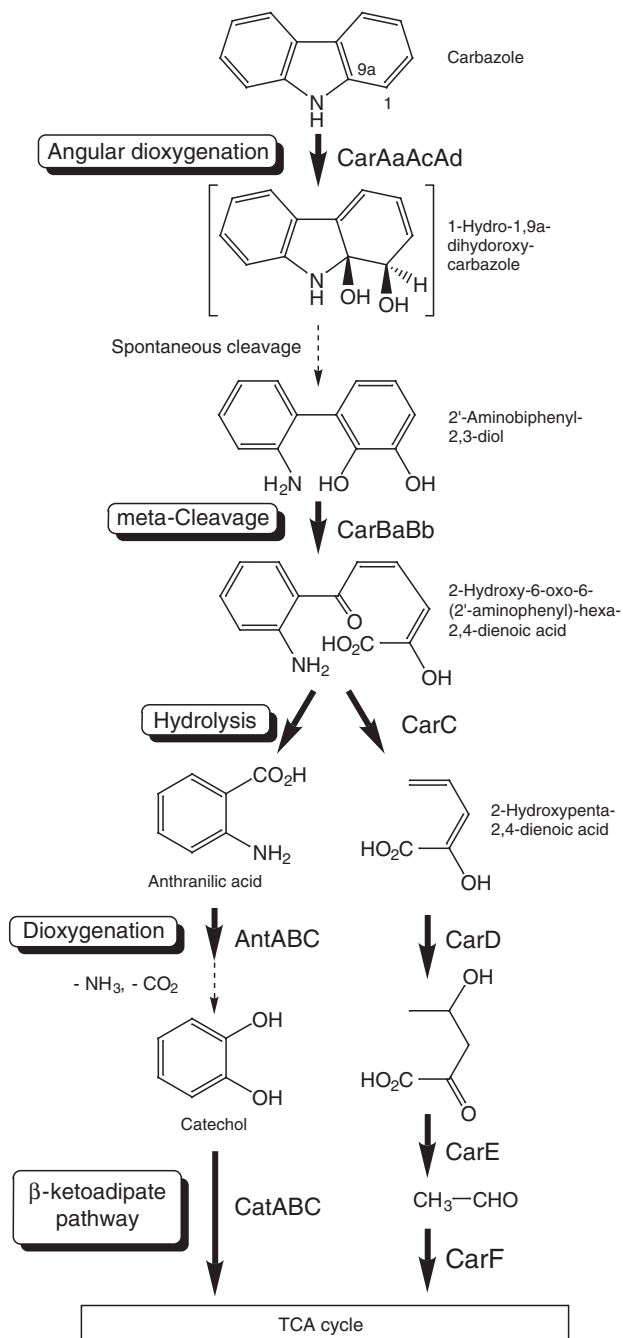


Fig. 4 Carbazole degradation pathway and degradative enzymes established in *Pseudomonas resinovorans* CA10. Compounds shown in brackets are unstable and have not been detected directly. *Solid and broken arrows* indicate enzymatic and spontaneous reactions, respectively

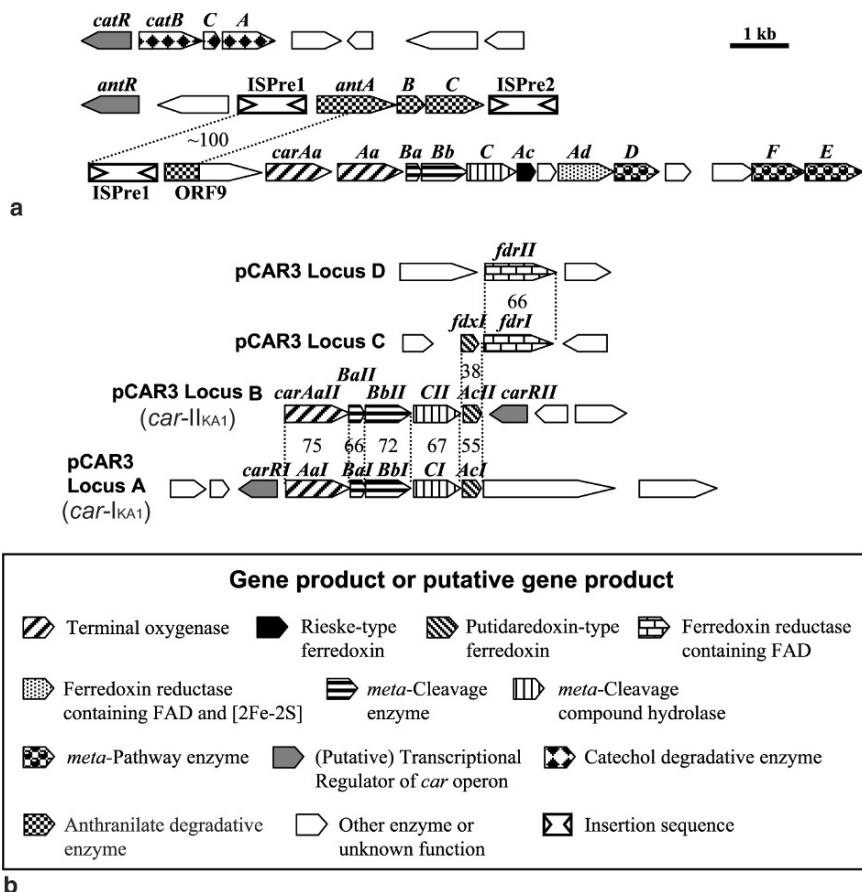


Fig. 5 Genetic structures of the gene clusters involved in carbazole degradation by (a) *Pseudomonas resinovorans* CA10 and (b) *Novosphingobium* sp. KA1. Amino acid sequence identities between the homologous proteins are also shown. The pentagons in the physical map indicate the size, location, and direction of transcription of the ORFs derived from the nucleotide sequence data

genes among *Pseudomonas* strains and *Sphingomonas* strains. In fact, carbazole-degradative pCAR1 and pCAR3 were found in *P. resinovorans* CA10 and *Novosphingobium* sp. KA1, respectively (Nojiri et al. 2001; Habe et al. 2002).

2.2 Carbazole-degradative Plasmid pCAR1 in *Pseudomonads*

2.2.1 Overview of the Genetic Structure of pCAR1

The entire nucleotide sequence of the circular plasmid pCAR1 was determined to elucidate the mechanism by which the *car* gene cluster may have been distributed in nature (Maeda et al. 2003). In the 199,035-bp sequence of pCAR1, a total of 190

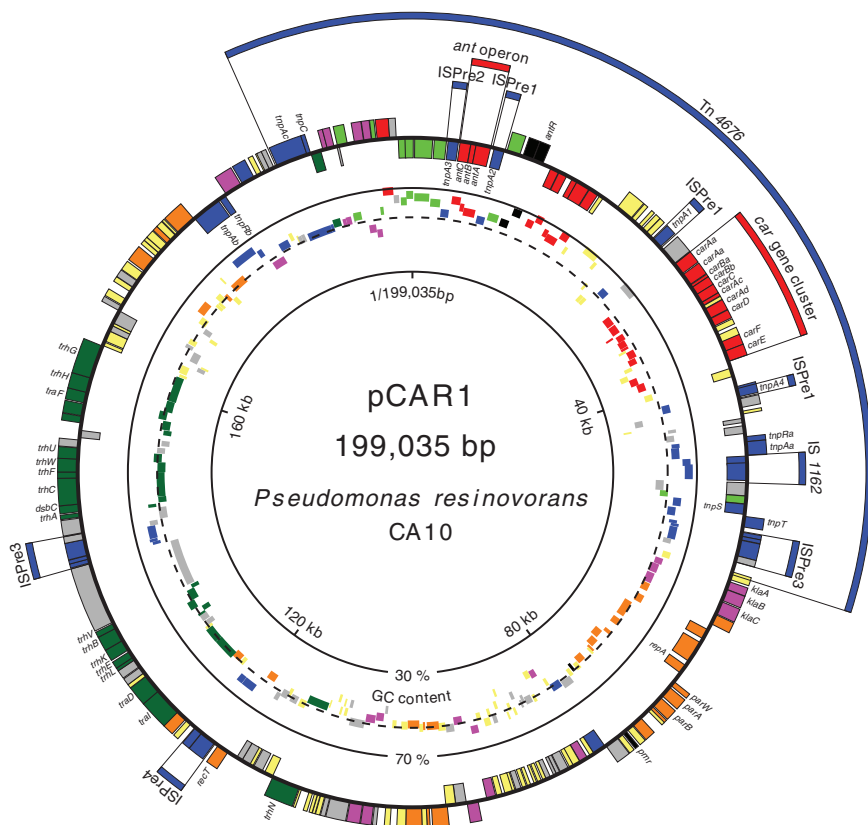


Fig. 6 The circular gene map of pCAR1 (Maeda et al. 2003). Genes or ORFs outside the circle are oriented clockwise; those inside are oriented counterclockwise. The (putative) functions of genes or ORFs are shown in color as follows: *orange*, maintenance or DNA processing; *dark green*, conjugative transfer; *blue*, transposition or integration; *red*, degradation; *light green*, transport; *black*, regulation; *magenta*, other known functions; *gray*, unknown function (homologous to hypothetical proteins); *yellow*, unknown function (no homology). *Bars* inside the circular gene map indicate the G + C contents of ORFs shown in same color. The broken circle indicates the average G + C content of the entire pCAR1 (56%)

ORFs were annotated. The gene organization and G + C contents of the respective ORFs are summarized in Fig. 6. The average G + C content of pCAR1 is 56%, whereas that of the genome of CA10 is 61.9%. Within pCAR1, there are several regions with notably different G + C contents. The average G + C content of the 28-kb region from ORF185 to ORF14 that contains the *ant* operon is higher (64%) than the average, whereas that of the *car* gene cluster and its flanking region (from ORF13 to ORF37) is approximately 51%.

The origin for vegetative replication, *oriV*, and Rep and Par proteins appear to be closely related to those of the naphthalene-degradative pND6-1 from *Pseudomonas* sp. ND6 (Li et al. 2004) and the toluene plasmid pWW53 from *Pseudomonas putida* MT53 (Yano et al. 2007). The potential tellurite resistance genes *klaABC* are also

related to those in the *Pseudomonas* plasmids. In contrast, the putative transfer genes of pCAR1 show low, but significant, homology (29–59% identity) with *trh* and *tra* genes that are involved in the conjugative transfer of plasmids or genomic islands from the Enterobacteriaceae.

In pCAR1, the *tnp* genes encode proteins that show > 70% overall lengthwise identity with transposition machinery proteins in the toluene/xylene-degrading transposon Tn4651 of the TOL plasmid pWW0 (Tsuda and Iino 1987). Both *car* and *ant* gene clusters are found within a 72.8-kb Tn4676 sequence that is defined by flanking *tnpAcC* and *tnpST* genes and bordered by a 46-bp inverted repeat (Fig. 6). Within Tn4676 and its flanking region, there are remnants of numerous mobile genetic elements such as the duplicated transposase genes that are highly homologous to *tnpR* of Tn4653 (Tsuda and Iino 1988) and the multiple candidates of IRS for Tn4676 and a Tn4653-like element. These findings suggest an evolutionary relationship between pCAR1 and pWW0.

2.2.2 Replication and Maintenance Functions

An approximately 2.7-kb DNA region containing *repA* and its flanking region (see Fig. 6) appears to be sufficient for the replication of pCAR1 in *P. putida* cells (Shintani et al. 2006). The RepA protein of pCAR1 shows high overall lengthwise identity with the corresponding proteins of pND6-1 (98%) and pWW53 (99%). In the *repA*-flanking region, a putative *oriV* could be proposed by analogy with *oriC* of *E. coli* (Schaper and Messer 1995). Deletion and complementation analyses have revealed that *oriV* of pCAR1 is located within a 345-bp DNA region, and that RepA could act *in trans* for the *oriV* region in *P. putida* cells (Shintani et al. 2006). The Inc group of pCAR1 could not be determined on the basis of nucleotide sequence alone because of the weak sequence similarities of *oriV* and RepA to the corresponding components of other plasmids. Nevertheless, incompatibility tests and Southern hybridization analyses with the *repA*-specific gene probe demonstrated that pCAR1 is an IncP-7 plasmid. pCAR1 is the first IncP-7 plasmid for which the entire nucleotide sequence was determined.

The *parWAB* genes are located upstream of the *repA* gene of pCAR1 and are oriented in the opposite direction (Fig. 6). RT-PCR analyses showed that *parWAB* is transcribed as a single transcriptional unit in CA10 cells. Deletion and gene disruption analyses revealed that *parWAB* is sufficient for the stable maintenance of pCAR1, although the role(s) of the ParW protein and/or *parW* gene region is unknown.

2.2.3 Transfer Function and Behavior in the Natural Environment

The conjugative transfer from *P. resinovorans* CA10 to *P. resinovorans* and *P. putida* strains was detected at frequencies of 3×10^{-1} and 3×10^{-3} per donor cell, respectively (Shintani et al. 2005a, 2006), suggesting that pCAR1 is a self-transmissible plasmid. Although the conjugal transfer of pCAR1 from CA10 to other *Pseudomonas* recipient

cells was not detected, mating analyses of pCAR2, which is an IncP-7 carbazole-degradative plasmid from *P. putida* HS01 that has a genetic structure highly homologous to that of pCAR1 (Table 2), showed that it could transfer from HS01 to *P. chlororaphis*, *P. fluorescens*, *P. putida*, *P. resinovorans*, and *P. stutzeri* (Shintani et al. 2005a). Recently, the nucleotide sequence of pCAR2 was determined. The nucleotide sequences of pCAR2 and pCAR1 reveal only negligible differences such as a single point mutation within the *tnpA1* transposase gene of ISPre1 and an additional copy of ISPre1 at the position adjacent to ISPre4 (Takahashi et al. unpublished results). Thus, pCAR1 and pCAR2 can be regarded as virtually the same plasmid. Hence, it can be concluded that pCAR1 has the capacity for conjugal transfer at least within the genus *Pseudomonas*. Very recently, Shintani et al. (2008) reported the conjugal transfer of pCAR1 from derivatives of *P. putida* KT2440 to *Stenotrophomonas*-like bacteria in a natural conjugation experiment in river water samples. This is the first report that the IncP-7 plasmid transfers into a non-*Pseudomonas* bacterium.

2.2.4 Crosstalk between the Host Chromosome and pCAR1

Does a chromosomal master switch control expression of the car and ant operons? The transcriptome of *P. putida* KT2440 harboring pCAR1 during growth with carbazole or succinate was analyzed (Miyakoshi et al. 2007). KT2440 is a plasmid-free derivative of *P. putida* mt-2 (original host of IncP-9 plasmid pWW0), and its complete genomic sequence is available (Nelson et al. 2002). Although KT2440 cannot use anthranilate as its sole source of carbon and energy (Jiménez et al. 2002), the conjugation of pCAR1 allows KT2440 to use either carbazole or anthranilate as its sole source of carbon and nitrogen. Microarray analysis showed that the levels of transcription of 38 plasmid genes and 139 chromosomal genes differed by more than twofold ($P < 0.05$). All structural genes constituting the pCAR1-borne *car* and *ant* operons and their regulatory gene *antR* (Urata et al. 2004; Miyakoshi et al. 2006) were upregulated during growth with carbazole. Chromosomal β -ketoacid pathway genes (*cat* and *pca* gene clusters) are induced, and, in conjunction with Car and Ant enzymes, the induced enzymes constitute a carbazole mineralization pathway to TCA cycle intermediates.

Upstream of the transcriptional start site of *antR*, at positions -24 and -12, there are conserved GG and GC doublets, which are characteristic binding motifs for the σ factor RpoN, suggesting that *antR* transcription is RpoN dependent. In fact, *rpoN* disruption followed by complementation analysis *in trans* confirmed the involvement of RpoN in *antR* induction in response to anthranilate (Miyakoshi et al. 2007). Generally, the transcription of RpoN-dependent promoters requires additional transcriptional activators (Buck et al. 2000). Therefore, the absence of an RpoN-dependent regulator gene in pCAR1 (Maeda et al. 2003) implies that *antR* transcription requires an unidentified regulator whose gene should be encoded on the host chromosome. This presumed regulatory system of the *car* and *ant* operons is an interesting parallel to the toluene-degradative *xyl meta* pathway operon regulatory

system consisting of the AraC/XylS-family regulator XylS and the RpoN-dependant NtrC-family regulator XylR (Ruíz et al. 2004). Toluene-degradative Xyl operons are located on conjugative plasmid pWW0, and XylS and XylR are encoded in the vicinity of the *xyl meta* pathway operon on pWW0. XylS corresponds to AntR in the carbazole degradation system of pCAR1, and an unidentified chromosomal regulator corresponds to XylR.

parI induction by the carriage of pCAR1. To analyze the change in the expression of the KT2440 chromosome in response to the carriage of pCAR1, the transcriptome of KT2440(pCAR1) was compared with that of the isogenic plasmid-free strain KT2440 growing under the same culture conditions (succinate minimal medium) (Miyakoshi et al. 2007). As a result, only 10 chromosomal genes in KT2440(pCAR1) were differentially expressed by at least twofold ($P < 0.05$; Table 3). Unexpectedly, a hypothetical gene, PP3700, showed an exceptional 41.8-times induction in the presence of pCAR1. The PP3700 product has a helix–turn–helix DNA-binding motif of the Xre family of transcriptional regulators in the N-terminal region. In addition, the PP3700 protein possesses the Walker-type ATPase motif, which is a characteristic of the ParA family of ATPases involved in the active partitioning of low-copy-number plasmids and the segregation of chromosomes upon cell division (Gerdes et al. 2000). PP3700 is designated *parI* (inducible *parA* homolog). Although the physiological role of ParI has not been clarified, ParI might function in pCAR1 maintenance in KT2440 cells.

pCAR1-encoded MvaT-like regulator Pmr. pCAR1 has an ORF (ORF70) that encodes an MvaT family transcriptional regulator and is designated *pmr* (plasmid-encoded MvaT-like regulator; Fig. 6). The MvaT family is a novel class of H-NS

Table 3 Differential expression of KT2440 chromosomal genes affected by pCAR1

ORF ^a	Gene name ^a	Change (<i>n</i> -fold) ^b	Description of product ^a
PP0457	<i>rplB</i>	2.1	Ribosomal protein L2
PP0789	<i>ampD</i>	2.5	<i>N</i> -acetyl-anhydromuramyl-L-alanine amidase AmpD
PP1149		5.4	Hypothetical protein
PP2161		3.0	Hypothetical protein
PP3700	<i>parI</i>	41.8	Hypothetical protein
PP4870		5.0	Azurin
PP1560		-2.0	Hypothetical protein
PP3921		-3.2	Hypothetical protein
PP3991		-5.0	Hypothetical protein
PP4117		-3.6	Hypothetical protein

^aFrom the annotated genome (Nelson et al. 2002) as indicated by The Institute for Genomic Research (<http://www.tigr.org>)

^bThe values indicate the mean levels of up-regulation (positive values) and down-regulation (negative values) of gene expression in KT2440(pCAR1) compared to that in KT2440. Student's *t*-test for each perfect match probe and each biological replicate, followed by the Bonferroni correction, was used to identify genes showing significantly different expression patterns ($P < 0.05$)

^cPP3700 was named as *parI* by Miyakoshi et al. (2007) (see text for the detail)

nucleotide-associated proteins that was recently found within *Pseudomonas* genomes (Tendeng et al. 2003). The KT2440 chromosome has five copies of *mvaT* homologs, and conjugal transfer of pCAR1 to KT2440 cells involves the introduction of a sixth copy of an *mvaT*-like gene. Conjugal transfer of pCAR1 may result in a change in the content and/or proportion of MvaT homologs in the KT2440(pCAR1) cell. Thus, it is possible that Pmr plays an important role as a global regulator in constructing a *de novo* transcriptional network between pCAR1 and its host chromosome.

To identify the regulon under the control of Pmr, the transcriptome of a *pmr* disruptant of *P. putida* KT2440 harboring pCAR1 [KT2440(pCAR1 Δ pmr)] was compared to that of the parental strain KT2440(pCAR1). Although no genes were upregulated by *pmr* disruption, 42 chromosomal genes and 17 pCAR1 genes were downregulated (Tables 4 and 5; Miyakoshi, Shintani, Terabayashi et al. unpublished results). Most noteworthy among the genes affected by *pmr* disruption are the *parAB*, which are involved in the active partitioning of pCAR1 (Shintani et al. 2006), and chromosomally encoded *parI* genes.

Table 4 Differential expression of *P. putida* KT2440 chromosomal genes affected by *pmr* disruption

ORF ^a	Gene name ^a	Change (n-fold) ^b	P-value ^b	Description of product ^a
PP0004	<i>gidA</i>	-2.1	5.0E-03	Glucose-inhibited division protein A
PP0008		-2.7	4.3E-03	Ribonuclease P protein component
PP0440	<i>tuf-1</i>	-2.0	1.2E-02	Translation elongation factor Tu
PP0441		-3.1	1.6E-02	Preprotein translocase, SecE subunit
PP0442		-3.1	8.2E-04	Transcription antitermination protein NusG
PP0455	<i>rplD</i>	-2.1	3.9E-02	Ribosomal protein L4
PP0461	<i>rplP</i>	-2.4	3.3E-03	Ribosomal protein L16
PP0841		-2.3	2.1E-03	rrf2 family protein
PP0844		-2.1	4.3E-03	Iron-binding protein IscA
PP0964		-2.1	9.1E-03	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
PP1213		-2.2	4.8E-02	Aspartyl-tRNA synthetase
PP1250		-2.6	1.4E-02	Group II intron-encoding maturase
PP1252		-3.4	5.5E-04	Group II intron-encoding maturase
PP1427	<i>algT</i>	-5.6	3.8E-06	RNA polymerase sigma factor AlgT
PP1428	<i>mucA</i>	-3.0	1.8E-03	σ factor AlgU negative regulatory protein MucA
PP1429	<i>algN</i>	-3.2	1.3E-03	Sigma factor AlgU regulatory protein AlgN
PP1607		-2.4	4.0E-02	Acetyl-CoA carboxylase, carboxyl transferase, alpha subunit
PP1623	<i>rpoS</i>	-2.5	4.8E-02	RNA polymerase sigma factor RpoS
PP1624		-3.6	2.0E-02	Group II intron-encoding maturase
PP1776		-2.7	3.3E-02	Mannose-6-phosphate isomerase/mannose-1-phosphate guanylyltransferase
PP1996		-2.6	6.3E-05	Acetyl-CoA carboxylase, carboxyl transferase, beta subunit
PP2452		-2.9	9.3E-03	Hypothetical protein

(continued)

Table 4 (continued)

ORF ^a	Gene name ^a	Change (n-fold) ^b	P-value ^b	Description of product ^a
PP2471		-2.4	1.0E-03	Integration host factor, alpha subunit
PP2628		-2.6	2.6E-03	ABC transporter, ATP-binding protein
PP2827		-2.2	3.6E-03	Alcohol dehydrogenase, zinc-containing
PP3100		-3.1	3.4E-02	Hypothetical protein
PP3455		-3.4	1.2E-03	Multidrug efflux RND membrane fusion protein
PP3601		-2.8	1.1E-02	D-Galactarate dehydratase, putative
PP3700		-40.5	2.4E-04	Hypothetical protein
PP3927		-2.5	1.7E-04	Hypothetical protein
PP4175		-2.1	3.0E-02	3-Oxoacyl-(acyl-carrier-protein) synthase I
PP4201		-2.4	1.8E-02	Electron transfer flavoprotein, alpha subunit
PP4378	<i>fliC</i>	-2.4	1.0E-03	Flagellin FlhC
PP4395	<i>flgM</i>	-2.0	2.9E-03	Negative regulator of flagellin synthesis FlgM
PP4470		-2.1	5.7E-03	Alginate biosynthesis transcriptional activator
PP4519		-3.0	4.7E-04	Agglutination protein
PP4965		-2.1	1.1E-02	Transketolase
PP5301		-2.4	2.9E-02	DNA-directed RNA polymerase, omega subunit
PP5302		-2.5	9.5E-03	Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase
PP5336		-2.4	2.4E-04	Phosphoribosylaminoimidazole carboxylase, catalytic subunit
PP5413	<i>atpD</i>	-2.1	3.6E-04	ATP synthase F1, beta subunit
PP5419		-2.5	1.7E-03	ATP synthase F0, A subunit

^aFrom the annotated genome (Nelson et al. 2002) as indicated by The Institute for Genomic Research (<http://www.tigr.org>)

^bThe values indicate the mean levels of down-regulation (negative values) of gene expression in KT2440(pCAR1 Δ pmr) compared to that in KT2440(pCAR1). Student's *t* test for each perfect match probe and each biological replicate, followed by the Bonferroni correction, was used to identify genes showing significantly different expression patterns ($P < 0.05$)

Signal transduction from pCAR1 carriage to parI induction. Whereas ParA and ParB proteins have an important role in the active partitioning of pCAR1 concomitant with chromosome segregation (Shintani et al. 2006), their homologous proteins regulate the expression of other genes as DNA-binding proteins (Gerdes et al. 2000; Funnell and Slavcev 2004). Because the expression of both *parAB* and *parI* were reduced in Pmr disruptant cells, it is possible that ParA and/or ParB proteins are involved in the induction of the *parI* promoter (*parIp*).

The transcriptional start point of the *parI* gene was determined by primer extension analysis, and reporter gene analyses revealed that the region up to at least -50 from the transcriptional start point is essential for the activation of *parIp* in response to the carriage of pCAR1 (Miyakoshi et al. 2007). In addition, the palindromic sequence located in the region from -50 to -38 plays an important role in the activation of *parIp*. Reporter gene analyses in several host cells expressing ParA, ParB, or ParI, showed that *parIp* is induced by the addition of ParA or ParI *in trans* in the KT2440 cell, but that the direct activator protein of *parIp* is the ParI protein itself (Miyakoshi et al. 2007). These results indicate the mechanism of *parI* induction by

Table 5 Differential expression of pCAR1 genes affected by *pmr* disruption

Gene ^a	Change (<i>n</i> -fold) ^b	<i>P</i> -value ^b	Description of product ^a
<i>mpA3</i>	-5.4	9.3E-13	Transposase
<i>mpA2</i>	-2.0	5.7E-05	Transposase
<i>mpA4</i>	-2.0	3.1E-07	Transposase
<i>parA</i>	-10.0	1.2E-15	Partitioning protein
<i>parB</i>	-8.5	1.8E-24	Partitioning protein
ORF67	-2.2	3.9E-10	Hypothetical protein
ORF70 (<i>pmr</i>)	-7.8	1.5E-14	Transcriptional regulator
ORF71	-2.5	4.1E-10	Hypothetical protein
ORF95a	-2.0	7.0E-06	DNA binding protein
ORF100	-2.4	6.3E-17	Hypothetical protein
ORF101	-2.5	1.3E-13	Cobalamine biosynthesis protein
ORF102	-2.1	3.0E-10	Cobalamine biosynthesis protein
ORF104	-2.3	1.1E-14	Hypothetical protein
ORF106	-2.6	7.2E-10	Hypothetical protein
ORF108	-2.7	4.2E-13	Hypothetical protein
ORF145	-3.6	2.1E-28	Putative DNA nickase
ORF146	-2.8	2.2E-09	Putative DNA primerase

^aFrom the annotated pCAR1 genome (Maeda et al. 2003)

^bThe values indicate the mean levels of down-regulation (negative values) of gene expression in KT2440(pCAR1Δ*pmr*) compared to that in KT2440(pCAR1). Students *t* test for each perfect match probe and each biological replicate, followed by the Bonferroni correction, was used to identify genes showing significantly different expression patterns ($P < 0.05$)

pCAR1 carriage (by the presence of *Pmr*): *Pmr* induces the expression of *ParA* and *ParB*; *ParA* induces the activation of *parIp* in an indirect manner; and *ParI* itself activates *parIp* to induce more *ParI* protein.

Pmr binding to chromosomally encoded MvaT homologs. MvaT-family transcriptional regulators constitute homomultimeric and heteromultimeric forms in vivo (Vallet-Gely et al. 2005). The KT2440 chromosome has five genes that encode MvaT homologs (PP0017, PP1366 [TurA], PP2947, PP3693, and PP3765 [TurB]). Pull-down assays indicate that *Pmr* can bind with not only its own protein but also three chromosomal MvaT homologs: PP1366 protein (TurA), PP2947 protein, and PP3756 protein (TurB) (Terabayashi et al. unpublished results). These results suggest the possibility that *Pmr* forms both homomultimeric and heteromultimeric forms in vivo.

2.3 Carbazole-degradative Plasmid pCAR3 in *Sphingomonads*

2.3.1 Overview of the Genetic Structure of pCAR3

The 254,797-bp nucleotide sequence of pCAR3 was determined by shotgun sequencing followed by manual gap closing (Shintani et al. 2007). pCAR3 contains 263 ORFs, including 26 ORFs reported previously (Inoue et al. 2004, 2005; Urata et al. 2006) (Fig. 7). The average G + C content of pCAR3 is 62.5%, which is similar to that of other sphingomonad genomes (61.6–67.8%; Balkwill et al. 1997). The average

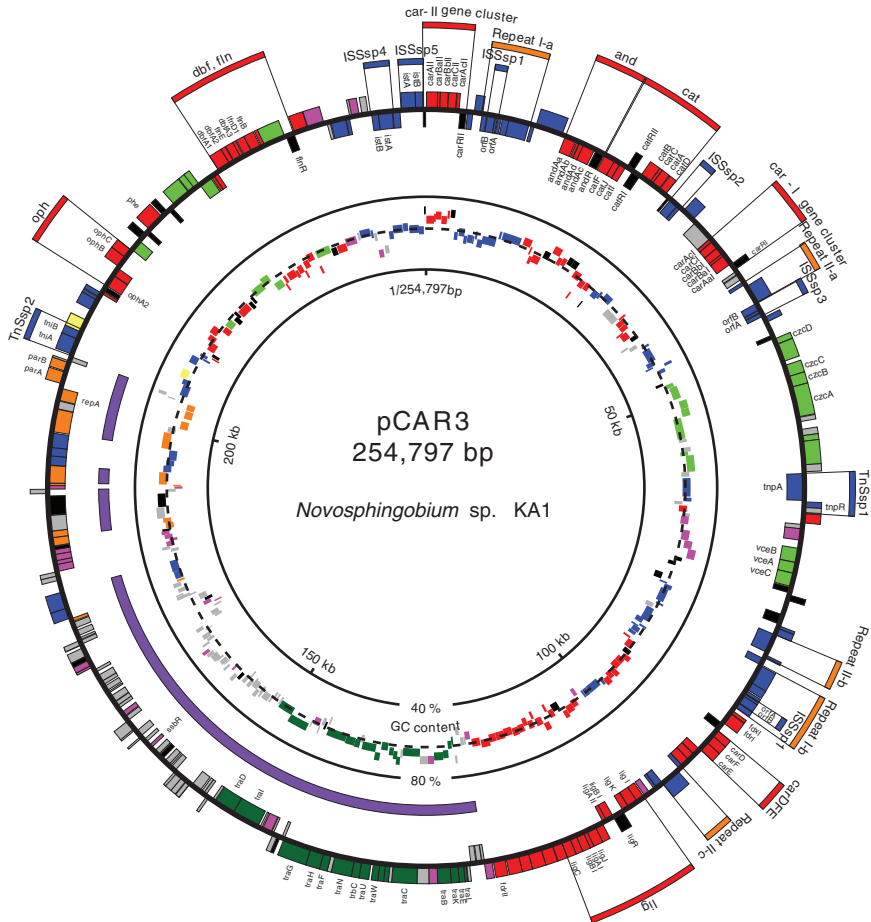


Fig. 7 The circular gene map of pCAR3 (Shintani et al. 2007). Genes or ORFs outside the circle are oriented clockwise; those inside are oriented counterclockwise. The (putative) functions of genes or ORFs are shown in color as follows: *orange*, maintenance or DNA processing; *dark green*, conjugative transfer; *blue*, transposition or integration; *red*, degradation; *light green*, transport; *black*, regulation; *magenta*, other known functions; *gray*, unknown function (homologous to hypothetical protein); *yellow*, unknown function (no homology). *Bars* inside the circular gene map indicate the G + C contents of ORFs shown in same color. The broken circle indicates the average G + C content of the entire pCAR3 (62.5%). *Bold purple lines* show the regions homologous to pNL1 (Romine et al. 1999)

G + C content of the 5-kb region containing the *car-II* operon (from ORF264 to ORF6) is high (70%), whereas that of the 5-kb region from ORF174 to ORF180 is low (53%). The sequences and organizations of ORF116 to ORF191 and ORF210 to ORF216 show homologies with those of pNL1 from *N. aromaticivorans* F199 (39–95% identity at the amino acid sequence level for each gene product; Fig. 7).

2.3.2 Carbazole Degradation System in pCAR3: A *Sphingomonas*-specific *car* Gene Cluster

Although the homology score is relatively low (<60% identity between counterparts at nucleotide sequence levels), *car* gene cluster homologs that show similarity in gene organization and phylogeny with the *Pseudomonas car* gene cluster have been found in the genus *Sphingomonas* and related genera (Habe et al. 2002; Kilbane et al. 2002; Inoue et al. 2004). *Sphingomonas car* gene clusters do not contain the NAD(P)H:ferredoxin oxidoreductase gene that is involved in the initial dioxygenation (Fig. 5b). Interestingly, although *Sphingomonas*-type CarAa (subunit of terminal oxygenase in the initial oxygenase system) shows significant homology with *Pseudomonas*-type CarAa (> 55% identity), ferredoxin encoded within *Sphingomonas car* gene cluster is unrelated to the *Pseudomonas*-type CarAc (Rieske ferredoxin), but shows similarity to the putidaredoxin-type ferredoxins (Fig. 5). *Sphingomonas*-type CarAa can receive electrons actually from putidaredoxin-type ferredoxin (termed also CarAc) and catalyze oxygenation for aromatics, suggesting that the ferredoxin selectivity differs between *Pseudomonas*-type CarAa and *Sphingomonas*-type CarAa (Inoue et al. 2004; Urata et al. 2006).

In addition to the *car* gene cluster first isolated from KA1 (redesignated *car*-I_{KA1}), an additional copy of the *car* gene cluster (*car*-II_{KA1} gene cluster) was found on plasmid pCAR3 (Fig. 5b; Urata et al. 2006; Shintani et al. 2007). In addition, NAD(P)H:ferredoxin oxidoreductase genes (*fdxI* and *fdxII*) and a third putidaredoxin-type ferredoxin gene were found on different loci of pCAR3 (Figs. 5b and 6b). RT-PCR analyses revealed that each *car* gene cluster constituted an operon, and their expression was induced in response to exposure to carbazole, although the *fdxI*-*fdxI* and *fdxII* genes were expressed constitutively (Urata et al. 2006). The genes involved in the mineralization of anthranilate (*and*) were identified on other loci of pCAR3 (Fig. 7). The products of the *and* gene cluster (*andAaAbAdAc-andR*) showed 39–68% identity with And proteins that encode anthranilate 1,2-dioxygenase of *Burkholderia cepacia* DBO1 (Chang et al. 2003). The well-established β -ketoadipate pathway genes, the *cat* gene clusters, were also found in pCAR3. The *meta*-cleavage pathway genes, *carDFE*, involved in 2-hydroxypenta-2,4-dienoate mineralization, were also identified (Fig. 7). These gene clusters were inducibly or constitutively expressed; thus, pCAR3 contains the complete set of genes responsible for carbazole mineralization. This contrasts with pCAR1, which does not carry the genes to metabolize catechol through the β -ketoadipate pathway.

2.3.3 Other Catabolic Genes on pCAR3

On the pCAR3 genome are gene clusters that show the highest homologies with the fluorene-degradative *fln/dbf* gene cluster of *Terrabacter* sp. DBF63 (Kasuga et al. 2001; Habe et al. 2004), the phthalate-degradative *oph* gene clusters of *B. cepacia* DBO1 (Chang and Zylstra 1998), and the protocatechuate-degradative gene clusters such as the *lig* gene clusters of *S. paucimobilis* SYK-6 (Masai et al. 1999) or *fld* genes of *Sphingomonas* sp. strain LB126 (Wattiau et al. 2001). Although RT-PCR

analyses showed that some of the *dbf/fln*, *oph*, and *lig* gene clusters are expressed constitutively, KA1 could not grow on fluorene or dibenzofuran (Shintani et al. 2007). The function of the possible gene products of *dbf/fln*, *oph*, and *lig* genes has not been analyzed, and it cannot be concluded whether these gene clusters are involved in any catabolic functions of the host cell.

2.3.4 Basic Plasmid Core Function

On the basis of sequence similarity, ORF212, ORF213, and ORF214 of pCAR3 are presumed to encode a replication initiator protein and partitioning proteins and were designated *repA*, *parA*, and *parB*, respectively (Shintani et al. 2007). The *oriV* on pCAR3 was predicted to be located between *repA* and *parA*. Detailed comparison and analysis of the *repA* and *parAB* genes and their products in pCAR3 and pNL1 revealed that the replication of these plasmids seems to be controlled by a single protein, RepAb (pNL1) or RepA (pCAR3), with the target iteron sequences (Shintani et al. 2007). This iteron-based replication is distinguishable from the replication of the *repABC*-type plasmids that are predominantly found in the α -proteobacterial group Rhizobiaceae. The RepA and RepB proteins of the latter type of replicon are homologs of the ParAB family of replicon-partitioning proteins (Bignell and Thomas 2001), and RepC is the replication initiation protein (Ramírez-Romero et al. 2000). The *repABC*-type replicons do not seem to possess direct repeats considered as iterons (MacLellan et al. 2005); therefore, pCAR3 and pNL1 can be assigned to a new Inc group that is phylogenetically distant from other groups of catabolic plasmids (Shintani et al. 2007).

The deduced amino acid sequences of 16 ORFs (ORF120 to ORF123, ORF126, ORF127 to ORF133, ORF135 to ORF137, and ORF 143 to ORF 144) show 60–89% homology to *tra* or *trw* genes that are encoded on pNL1 (Romine et al. 1999). Because pNL1 was reported to be a conjugative plasmid (Romine et al. 1999; Basta et al. 2004), the transferability of pCAR3 was tested using filter mating with a pCAR3-cured KA1-derivative strain as a recipient. However, no transconjugant has yet been obtained. Thus, pCAR3 may be defective in conjugative transfer (Shintani et al. 2007).

3 Concluding Remarks

Bacterial degradation of PAHs and heteroaromatics has been extensively studied for a long period; in the last decade, information on the degradation pathways, genes, and enzymes has increased rapidly. For PAH degradation, many recent studies have attempted to characterize the degradation system for PAHs with four or more rings. In fact, the metabolic pathways for pyrene, fluoranthene, benzo[*a*]pyrene, and others have been proposed recently (Kanaly and Harayama 2000). The isolation of new xenobiotic-degrading bacteria followed by genetic analyses will provide new examples of catabolic plasmids. In addition, recent genomewide research has provided information on the occurrence of plasmids with unknown functions. The current incompatibility grouping

is very useful in the classification of plasmids isolated from *Pseudomonas* and *Enterobacteriaceae*, but is sometimes not applicable for other plasmids. Therefore, a classification that includes all plasmids based on the plasmid basic functions (e.g., incompatibility, host range, transferability) and nucleotide sequences will be necessary to better understand plasmid diversity. Lack of the information on the plasmid basic function is remarkable for plasmids isolated from α -Proteobacteria. To the best of our knowledge, the incompatibility grouping has not been decided for this important class of plasmids, although many catabolic plasmids have recently been reported for α -Proteobacteria, including sphingomonads (Basta et al. 2004, 2005; Liu et al. 2004; Keck et al. 2006; Miyazaki et al. 2006). Together with nucleotide sequencing, basic investigations of the machinery for replication, maintenance, and transfer will be important to understand the (degradation) functions of catabolic plasmids from sphingomonads.

However, it should be noted that the activity of xenobiotic degradation conferred by catabolic plasmids might be dependent on the host cell. Although the catabolic gene(s) on plasmids is distributed among various bacterial strains via horizontal gene transfer, previous functional investigations of the xenobiotic-degrading systems conferred by catabolic plasmids have mostly been carried out using a single (in many cases, an original host) or a restricted number of host bacterial cell backgrounds in optimal laboratory conditions. A re-evaluation of the current information on the functions of particular catabolic systems (in bacterial populations or the natural environment) is needed from an ecological point of view. For this purpose, the accumulation of information on the genetic structures of various whole genomes and plasmids will greatly help to reveal the crosstalk between plasmid-borne catabolic systems and chromosome-borne housekeeping systems. Together with the 2,4-dichlorophenoxyacetic acid-degradative pJP4 (IncP-1; Trefault et al. 2004), toluene-degradative pWW0 (IncP-9; Greated et al. 2002), and NAH7 (IncP-9; Sota et al. 2006), pCAR1 has become one of the most extensively investigated IncP degradative plasmids. To elucidate the real performance of the plasmid, it will be important to clarify the alteration of plasmid functions under various host cell backgrounds and estimate plasmid dynamics in the natural environment such as in soil and water. For this purpose, because most of the host ranges of various plasmids have been estimated using filter-mating experiments, studies of the real host range of plasmids are also necessary to understand the ecological relationship between degradative plasmids and possible hosts in the natural environment.

Acknowledgments This work was partly supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN) in Japan and a Grant-in-Aid (Hazardous Chemicals) from the Ministry of Agriculture, Forestry and Fisheries of Japan (HC-07-2325-5).

References

- Arcos JC, Argus MF (1968) Molecular geometry and carcinogenic activity of aromatic compounds. New perspectives. *Adv Cancer Res* 11:305–471
- Balkwill DL, Drake GR, Reeves RH, Fredrickson JK, White DC, Ringelberg DB, Chandler DP, Romine MF, Kennedy DW, Spadoni CM (1997) Taxonomic study of aromatic-degrading

- bacteria from deep terrestrial-subsurface sediments and description of *Sphingomonas aromati-civorans* sp. nov., *Sphingomonas subterranea* sp. nov., and *Sphingomonas stygia* sp. nov. *Int J Syst Bacteriol* 47:191–201
- Barnett MJ, Fisher RF, Jones T, Komp C, Pia Abola A, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh KC, Davis RW, Federspiel NA, Long SR (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc Natl Acad Sci USA* 98:9883–9888
- Basta T, Keck A, Klein J, Stolz A (2004) Detection and characterization of conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas* strains. *J Bacteriol* 186:3862–3872
- Basta T, Buerger S, Stolz A (2005) Structural and replicative diversity of large plasmids from sphingomonads that degrade polycyclic aromatic compounds and xenobiotics. *Microbiology* 151:2025–2037
- Benson S, Shapiro J (1978) TOL is a broad-host-range plasmid. *J Bacteriol* 135:278–280
- Bignell C, Thomas CM (2001) The bacterial ParA-ParB partitioning proteins. *J Biotechnol* 91:1–34
- Bosch R, Garcia-Valdes E, Moore ER (1999) Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene* 236:149–157
- Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD (2000) The bacterial enhancer-dependent 54 (N) transcription factor. *J Bacteriol* 182:4129–4136
- Chang HK, Zylstra GJ (1998) Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1. *J Bacteriol* 180:6529–6537
- Chang HK, Mohseni P, Zylstra GJ (2003) Characterization and regulation of the genes for a novel anthranilate 1,2-dioxygenase from *Burkholderia cepacia* DBO1. *J Bacteriol* 185:5871–5881
- de Souza ML, Wackett LP, Sadowsky MJ (1998) The *atzABC* genes encoding atrazine catabolism are located on a self-transmissible plasmid in *Pseudomonas* sp. strain ADP. *Appl Environ Microbiol* 64:2323–2326
- Dean-Ross D, Moody JD, Freeman JP, Doerge DR, Cerniglia CE (2001) Metabolism of anthracene by a *Rhodococcus* species. *FEMS Microbiol Lett* 204:205–211
- Dennis JJ (2005) The evolution of IncP catabolic plasmids. *Curr Opin Biotechnol* 16:291–298
- Dennis JJ, Zylstra GJ (2004) Complete sequence and genetic organization of pDTG1, the 83 kilobase naphthalene degradation plasmid from *Pseudomonas putida* strain NCIB 9816–4. *J Mol Biol* 341:753–768
- Denome SA, Stanley DC, Olson ES, Young KD (1993) Metabolism of dibenzothiophene and naphthalene in *Pseudomonas* strains: complete DNA sequence of an upper naphthalene catabolic pathway. *J Bacteriol* 175:6890–6901
- Dunn NW, Gunsalus IC (1973) Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *J Bacteriol* 114:974–979
- Fernandez-Lopez R, Garcillan-Barcia MP, Revilla C, Lazaro M, Vielva L, de la Cruz F (2006) Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol Rev* 30:942–966
- Foght JM, Westlake DW (1996) Transposon and spontaneous deletion mutants of plasmid-borne genes encoding polycyclic aromatic hydrocarbon degradation by a strain of *Pseudomonas fluorescens*. *Biodegradation* 7:353–366
- Funnell BE, Slavcev RA (2004) Partition systems of bacterial plasmids. In: Funnell E, Philips GJ (eds) *Plasmid biology*. ASM, WA, pp 81–103
- Gerdes K, Møller-Jensen J, Jensen RB (2000) Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol* 37:455–466
- Goyal AK, Zylstra GJ (1997) Genetics of naphthalene and phenanthrene degradation by *Comamonas testosteroni*. *J Ind Microbiol Biotechnol* 19:401–407
- Greated A, Titok M, Krasowiak R, Fairclough RJ, Thomas CM (2000) The replication and stable-inheritance functions of IncP-9 plasmid pM3. *Microbiology* 146:2249–2258

- Greated A, Lambertsen L, Williams PA, Thomas CM (2002) Complete sequence of the IncP-9 TOL plasmid pWW0 from *Pseudomonas putida*. *Environ Microbiol* 4:856–871
- Grimm AC, Harwood CS (1999) NahY, a catabolic plasmid-encoded receptor required for chemotaxis of *Pseudomonas putida* to the aromatic hydrocarbon naphthalene. *J Bacteriol* 181:3310–3316
- Habe H, Ashikawa Y, Saiki Y, Yoshida T, Nojiri H, Omori T (2002) *Sphingomonas* sp. strain KA1, carrying a carbazole dioxygenase gene homologue, degrades chlorinated dibenzo-*p*-dioxins in soil. *FEMS Microbiol Lett* 211:43–49
- Habe H, Miyakoshi M, Chung JS, Kasuga K, Yoshida T, Nojiri H, Omori T (2003) Phthalate catabolic gene cluster is linked to the angular dioxygenase gene in *Terrabacter* sp. strain DBF63. *Appl Microbiol Biotechnol* 61:44–54
- Habe H, Chung JS, Kato H, Ayabe Y, Kasuga K, Yoshida T, Nojiri H, Yamane H, Omori T (2004) Characterization of the upper pathway genes for fluorene metabolism in *Terrabacter* sp. strain DBF63. *J Bacteriol* 186:5938–5944
- Habe H, Chung JS, Ishida A, Kasuga K, Ide K, Takemura T, Nojiri H, Yamane H, Omori T (2005) The fluorene catabolic linear plasmid in *Terrabacter* sp. strain DBF63 carries the β -ketoadipate pathway genes, *pcaRHGBDCFIJ*, also found in proteobacteria. *Microbiology* 151:3713–3722
- Harayama S, Reik M, Wasserfallen A, Bairoch A (1987) Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. *Mol Gen Genet* 210:241–247
- Harayama S, Kok M, Neidle EL (1992) Functional and evolutionary relationships among diverse oxygenases. *Annu Rev Microbiol* 46:565–601
- Iida T, Mukouzaka Y, Nakamura K, Kudo T (2002) Plasmid-borne genes code for an angular dioxygenase involved in dibenzofuran degradation by *Terrabacter* sp. strain YK3. *Appl Environ Microbiol* 68:3716–3723
- Inoue K, Widada J, Nakai S, Endoh T, Urata M, Ashikawa Y, Shintani M, Saiki Y, Yoshida T, Habe H, Omori T, Nojiri H (2004) Divergent structures of carbazole degradative *car* operons isolated from Gram-negative bacteria. *Biosci Biotechnol Biochem* 68:1467–1480
- Inoue K, Habe H, Yamane H, Omori T, Nojiri H (2005) Diversity of carbazole-degrading bacteria having the *car* gene cluster: isolation of a novel Gram-positive carbazole-degrading bacterium. *FEMS Microbiol Lett* 245:145–153
- Izmalkova TY, Mavrodi DV, Sokolov SL, Kosheleva IA, Smalla K, Thomas CM, Boronin AM (2006) Molecular classification of IncP-9 naphthalene degradation plasmids. *Plasmid* 56:1–10
- Jacoby GA (1986) Resistance plasmids of *Pseudomonas*. In: Sokatch JR (ed) *The bacteria: a treatise on structure and function*. Academic, NY, pp 265–293
- Jiménez JI, Miñambres B, García JL, Díaz E (2002) Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ Microbiol* 4:824–841
- Kanally RA, Harayama S (2000) Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J Bacteriol* 182:2059–2067
- Kasuga K, Habe H, Chung JS, Yoshida T, Nojiri H, Yamane H, Omori T (2001) Isolation and characterization of the genes encoding a novel oxygenase component of angular dioxygenase from the Gram-positive dibenzofuran-degrader *Terrabacter* sp. strain DBF63. *Biochem Biophys Res Commun* 283:195–204
- Kazunga C, Aitken MD (2000) Products from the incomplete metabolism of pyrene by polycyclic aromatic hydrocarbon-degrading bacteria. *Appl Environ Microbiol* 66:1917–1922
- Keck A, Conradt D, Mahler A, Stolz A, Mattes R, Klein J (2006) Identification and functional analysis of the genes for naphthalenesulfonate catabolism by *Sphingomonas xenophaga* BN6. *Microbiology* 152:1929–1940
- Kilbane JJ II, Daram A, Abbasian J, Kayser KJ (2002) Isolation and characterization of *Sphingomonas* sp. GTIN11 capable of carbazole metabolism in petroleum. *Biochem Biophys Res Commun* 297:242–248
- Kim E, Zylstra GJ (1999) Functional analysis of genes involved in biphenyl, naphthalene, phenanthrene, and *m*-xylene degradation by *Sphingomonas yanoikuyae* B1. *J Ind Microbiol Biotechnol* 23:294–302

- Kim SJ, Kweon O, Freeman JP, Jones RC, Adjei MD, Jhoo JW, Edmondson RD, Cerniglia CE (2006) Molecular cloning and expression of genes encoding a novel dioxygenase involved in low- and high-molecular-weight polycyclic aromatic hydrocarbon degradation in *Mycobacterium vanbaalenii* PYR-1. *Appl Environ Microbiol* 72:1045–1054
- Kimura N, Kitagawa W, Mori T, Nakashima N, Tamura T, Kamagata Y (2006) Genetic and biochemical characterization of the dioxygenase involved in lateral dioxygenation of dibenzofuran from *Rhodococcus opacus* strain SAO101. *Appl Microbiol Biotechnol* 73:474–484
- Krasowiak R, Sevastyanovich Y, Konieczny I, Bingle LE, Thomas CM (2006) IncP-9 replication initiator protein binds to multiple DNA sequences in oriV and recruits host DnaA protein. *Plasmid* 56:187–201
- Kulakov LA, Chen S, Allen CC, Larkin MJ (2005) Web-type evolution of *Rhodococcus* gene clusters associated with utilization of naphthalene. *Appl Environ Microbiol* 71:1754–1764
- Lambertsen LM, Molin S, Kroer N, Thomas CM (2004) Transcriptional regulation of pWWO transfer genes in *Pseudomonas putida* KT2440. *Plasmid* 52:169–181
- Laurie AD, Lloyd-Jones G (1999) The *phn* genes of *Burkholderia* sp. strain RP007 constitute a divergent gene cluster for polycyclic aromatic hydrocarbon catabolism. *J Bacteriol* 181:531–540
- Lehrbach PR, McGregor I, Ward JM, Broda P (1983) Molecular relationships between *Pseudomonas* Inc P-9 degradative plasmids TOL, NAH, and SAL. *Plasmid* 10:164–174
- Li W, Shi J, Wang X, Han Y, Tong W, Ma L, Liu B, Cai B (2004) Complete nucleotide sequence and organization of the naphthalene catabolic plasmid pND6-1 from *Pseudomonas* sp. strain ND6. *Gene* 336:231–240
- Liu Y, Zhang J, Zhang Z (2004) Isolation and characterization of polycyclic aromatic hydrocarbons-degrading *Sphingomonas* sp. strain ZL5. *Biodegradation* 15:205–212
- Ma Y, Wang L, Shao Z (2006) *Pseudomonas*, the dominant polycyclic aromatic hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in horizontal gene transfer. *Environ Microbiol* 8:455–465
- MacLellan SR, Smallbone LA, Sibley CD, Finan TM (2005) The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of α -proteobacterial plasmids. *Mol Microbiol* 55:611–623
- Maeda K, Nojiri H, Shintani M, Yoshida T, Habe H, Omori T (2003) Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicism and the presence of large catabolic transposon Tn4676. *J Mol Biol* 326:21–33
- Mallick S, Chatterjee S, Dutta TK (2007) A novel degradation pathway in the assimilation of phenanthrene by *Staphylococcus* sp. strain PN/Y via meta-cleavage of 2-hydroxy-1-naphthoic acid: formation of trans-2,3-dioxo-5-(2;-hydroxyphenyl)-pent-4-enoic acid. *Microbiology* 153:2104–2115
- Marston CP, Pereira C, Ferguson J, Fischer K, Hedstrom O, Dashwood WM, Baird WM (2001) Effect of a complex environmental mixture from coal tar containing polycyclic aromatic hydrocarbons (PAH) on the tumor initiation, PAH-DNA binding and metabolic activation of carcinogenic PAH in mouse epidermis. *Carcinogenesis* 22:1077–1086
- Martinez B, Tomkins J, Wackett LP, Wing R, Sadowsky MJ (2001) Complete nucleotide sequence and organization of the atrazine catabolic plasmid pADP-1 from *Pseudomonas* sp. strain ADP. *J Bacteriol* 183:5684–5697
- Masai E, Shinohara S, Hara H, Nishikawa S, Katayama Y, Fukuda M (1999) Genetic and biochemical characterization of a 2-pyrone-4,6-dicarboxylic acid hydrolase involved in the protocatechuate 4,5-cleavage pathway of *Sphingomonas paucimobilis* SYK-6. *J Bacteriol* 181:55–62
- Mastrangelo G, Fadda E, Marzia V (1996) Polycyclic aromatic hydrocarbons and cancer in man. *Environ Health Perspect* 104:1166–1170
- Menn FM, Applegate BM, Saylor GS (1993) NAH plasmid-mediated catabolism of anthracene and phenanthrene to naphthoic acids. *Appl Environ Microbiol* 59:1938–1942
- Mindlin S, Kholodii G, Gorlenko Z, Minakhina S, Minakhin L, Kalyaeva E, Kopteva A, Petrova M, Yurieva O, Nikiforov V (2001) Mercury resistance transposons of Gram-negative environmental bacteria and their classification. *Res Microbiol* 152:811–822

- Miyakoshi M, Urata M, Habe H, Omori T, Yamane H, Nojiri H (2006) Differentiation of carbazole catabolic operons by replacement of the regulated promoter via transposition of an insertion sequence. *J Biol Chem* 281:8450–8457
- Miyakoshi M, Shintani M, Terabayashi T, Kai S, Yamane H, Nojiri H (2007) Transcriptome analysis of *Pseudomonas putida* KT2440 harboring the completely sequenced IncP-7 plasmid pCAR1. *J Bacteriol* 189:6849–6860
- Miyazaki R, Sato Y, Ito M, Ohtsubo Y, Nagata Y, Tsuda M (2006) Complete nucleotide sequence of an exogenously isolated plasmid, pLB1, involved in γ -hexachlorocyclohexane degradation. *Appl Environ Microbiol* 72:6923–6933
- Moody JD, Freeman JP, Fu PP, Cerniglia CE (2004) Degradation of benzo[a]pyrene by *Mycobacterium vanbaalenii* PYR-1. *Appl Environ Microbiol* 70:340–345
- Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VAP, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Chris Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen J, Timmis KN, Dusterhöft A, Tümmler B, Fraser CM (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ Microbiol* 4:799–808
- Nojiri H, Omori T (2002) Molecular bases of aerobic bacterial degradation of dioxins: involvement of angular dioxygenation. *Biosci Biotechnol Biochem* 66:2001–2016
- Nojiri H, Omori T (2007) Carbazole metabolism by pseudomonads. In: Ramos JL, Filloux A (eds) *Pseudomonas*, vol 5. Springer, NY, pp 107–145
- Nojiri H, Sekiguchi H, Maeda K, Urata M, Nakai S, Yoshida T, Habe H, Omori T (2001) Genetic characterization and evolutionary implications of a *car* gene cluster in the carbazole degrader *Pseudomonas* sp. strain CA10. *J Bacteriol* 183:3663–3679
- Nojiri H, Kamakura M, Urata M, Tanaka T, Chung JS, Takemura T, Yoshida T, Habe H, Omori T (2002a) Dioxin catabolic genes are dispersed on the *Terrabacter* sp. DBF63 genome. *Biochem Biophys Res Commun* 296:233–240
- Nojiri H, Maeda K, Sekiguchi H, Urata M, Shintani M, Yoshida T, Habe H, Omori T (2002b) Organization and transcriptional characterization of catechol degradation genes involved in carbazole degradation by *Pseudomonas resinovorans* strain CA10. *Biosci Biotechnol Biochem* 66:897–901
- Nojiri H, Shintani M, Omori T (2004) Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl Microbiol Biotechnol* 64:154–174
- Ogawa N, Chakrabarty AM, Zaborina O (2004) Degradative plasmids. In: Phillips G, Funnel B (eds) *Plasmid biology*. ASM, WA, pp 341–376
- Ono A, Miyazaki R, Sota M, Ohtsubo Y, Nagata Y, Tsuda M (2007) Isolation and characterization of naphthalene-catabolic genes and plasmids from oil-contaminated soil by using two cultivation-independent approaches. *Appl Microbiol Biotechnol* 74:501–510
- Park W, Jeon CO, Hohnstock-Ashe AM, Winans SC, Zylstra GJ, Madsen EL (2003) Identification and characterization of the conjugal transfer region of the pCg1 plasmid from naphthalene-degrading *Pseudomonas putida* Cg1. *Appl Environ Microbiol* 69:3263–3271
- Pinyakong O, Habe H, Yoshida T, Nojiri H, Omori T (2003) Identification of three novel salicylate 1-hydroxylases involved in the phenanthrene degradation of *Sphingobium* sp. strain P2. *Biochem Biophys Res Commun* 301:350–357
- Ramírez-Romero MA, Soberón N, Pérez-Oseguera A, Téllez-sosa J, Cevallos MA (2000) Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. *J Bacteriol* 182:3117–3124
- Rentz JA, Alvarez PJ, Schnoor JL (2008) Benzo[a]pyrene degradation by *Sphingomonas yanoikuyae* JAR02. *Environ Pollut* 151:669–677
- Romine MF, Stillwell LC, Wong KK, Thurston SJ, Sisk EC, Sensen C, Gaasterland T, Fredrickson JK, Saffer JD (1999) Complete sequence of a 184-kilobase catabolic plasmid from *Sphingomonas aromaticivorans* F199. *J Bacteriol* 181:1585–1602

- Ruíz R, Aranda-Olmedo MI, Domínguez-Cuevas R, Ramos-González MI, Marqués S (2004) Transcriptional regulation of the toluene catabolic pathway. In: Ramos JL (ed) *Pseudomonas*, vol 2. Kluwer, NY, pp 509–537
- Sanseverino J, Applegate BM, King JM, Saylor GS (1993) Plasmid-mediated mineralization of naphthalene, phenanthrene, and anthracene. *Appl Environ Microbiol* 59:1931–1937
- Sato S, Nam JW, Kasuga K, Nojiri H, Yamane H, Omori T (1997a) Identification and characterization of genes encoding carbazole 1,9a-dioxygenase in *Pseudomonas* sp. strain CA10. *J Bacteriol* 179:4850–4858
- Sato S, Ouchiyama N, Kimura T, Nojiri H, Yamane H, Omori T (1997b) Cloning of genes involved in carbazole degradation of *Pseudomonas* sp. strain CA10: nucleotide sequences of genes and characterization of *meta*-cleavage enzymes and hydrolase. *J Bacteriol* 179:4841–4849
- Schaper S, Messer W (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. *J Biol Chem* 270:17622–17626
- Sevastyanovich YR, Titok MA, Krasowiak R, Bingle LE, Thomas CM (2005) Ability of IncP-9 plasmid pM3 to replicate in *Escherichia coli* is dependent on both *rep* and *par* functions. *Mol Microbiol* 57:819–833
- Sherratt D (1989) Tn3 and related transposable elements: site-specific recombination and transposition. In: Berg DE, Howe MM (eds) *Mobile DNA*. American Society for Microbiology, WA, pp 163–184
- Shintani M, Habe H, Tsuda M, Omori T, Yamane H, Nojiri H (2005a) Recipient range of IncP-7 conjugative plasmid pCAR2 from *Pseudomonas putida* HS01 is broader than from other *Pseudomonas* strains. *Biotechnol Lett* 27:1847–1853
- Shintani M, Yoshida T, Habe H, Omori T, Nojiri H (2005b) Large plasmid pCAR2 and class II transposon Tn4676 are functional mobile genetic elements to distribute the carbazole/dioxin-degradative *car* gene cluster in different bacteria. *Appl Microbiol Biotechnol* 67:370–382
- Shintani M, Yano H, Habe H, Omori T, Yamane H, Tsuda M, Nojiri H (2006) Characterization of the replication, maintenance, and transfer features of the IncP-7 plasmid pCAR1, which carries genes involved in carbazole and dioxin degradation. *Appl Environ Microbiol* 72:3206–3216
- Shintani M, Urata M, Inoue K, Eto K, Habe H, Omori T, Yamane H, Nojiri H (2007) The *Sphingomonas* plasmid pCAR3 is involved in complete mineralization of carbazole. *J Bacteriol* 189:2007–2020
- Shintani M, Fukushima N, Tezuka M, Yamane H, Nojiri H (2008) Conjugative transfer of IncP-7 carbazole degradative plasmid, pCAR1, in river water samples. *Biotechnol Lett* 30:117–122
- Sota M, Yano H, Ono A, Miyazaki R, Ishii H, Genka H, Top EM, Tsuda M (2006) Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and its transposon Tn4655 suggests catabolic gene spread by a tyrosine recombinase. *J Bacteriol* 188:4057–4067
- Stuart-Keil KG, Hohnstock AM, Drees KP, Herrick JB, Madsen EL (1998) Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar-contaminated site are homologous to pDTG1 from *Pseudomonas putida* NCIB 9816–4. *Appl Environ Microbiol* 64:3633–3640
- Tendeng C, Soutourina OA, Danchin A, Bertin PN (2003) MvaT proteins in *Pseudomonas* spp.: a novel class of H-NS-like proteins. *Microbiology* 149:3047–3050
- Thacker R, Rørvig O, Kahlon P, Gunsalus IC (1978) NIC, a conjugative nicotine-nicotinate degradative plasmid in *Pseudomonas convexa*. *J Bacteriol* 135:289–290
- Top EM, Moenne-Loccoz Y, Pembroke T, Thomas CM (2000) Phenotypic traits conferred by plasmids. In: Thomas CM (ed) *The horizontal gene pool*. Harwood, Amsterdam, The Netherlands, pp 249–286
- Trefault N, de la Iglesia R, Molina AM, Manzano M, Ledger T, Pérez-Pantoja D, Sánchez MA, Stuardo M, González B (2004) Genetic organization of the catabolic plasmid pJP4 from *Ralstonia eutropha* JMP134 (pJP4) reveals mechanisms of adaptation to chloroaromatic pollutants and evolution of specialized chloroaromatic degradation pathways. *Environ Microbiol* 6:655–668
- Tsuda M, Iino T (1987) Genetic analysis of a transposon carrying toluene-degrading genes on a TOL plasmid pWW0. *Mol Gen Genet* 210:270–276

- Tsuda M, Iino T (1988) Identification and characterization of Tn4653, a transposon covering the toluene transposon Tn4651 on TOL plasmid pWW0. *Mol Gen Genet* 213:72–77
- Tsuda M, Iino T (1990) Naphthalene degrading genes on plasmid NAH7 are on a defective transposon. *Mol Gen Genet* 223:33–39
- Urata M, Miyakoshi M, Kai S, Maeda K, Habe H, Omori T, Yamane H, Nojiri H (2004) Transcriptional regulation of the *ant* operon, encoding two-component anthranilate 1,2-dioxygenase, on the carbazole-degradative plasmid pCAR1 of *Pseudomonas resinovorans* strain CA10. *J Bacteriol* 186:6815–6823
- Urata M, Uchimura H, Noguchi H, Sakaguchi T, Takemura T, Eto K, Habe H, Omori T, Yamane H, Nojiri H (2006) Plasmid pCAR3 contains multiple gene sets involved in the conversion of carbazole to anthranilate. *Appl Environ Microbiol* 72:3198–3205
- Uz I, Duan YP, Ogram A (2000) Characterization of the naphthalene-degrading bacterium, *Rhodococcus opacus* M213. *FEMS Microbiol Lett* 185:231–238
- Vallet-Gely I, Donovan KE, Fang R, Joung JK, Dove SL (2005) Repression of phase-variable cup gene expression by H-NS-like proteins in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 102:11082–11087
- van Hamme JD, Singh A, Ward OP (2003) Recent advances in petroleum microbiology. *Microbiol Mol Biol Rev* 67:503–549
- Wattiau P, Bastiaens L, van Herwijnen R, Daal L, Parsons JR, Renard ME, Springael D, Cornelis GR (2001) Fluorene degradation by *Sphingomonas* sp. LB126 proceeds through protocatechuic acid: a genetic analysis. *Res Microbiol* 152:861–872
- White GP, Dunn NW (1978) Compatibility and sex specific phage plating characteristics of the TOL and NAH catabolic plasmids. *Genet Res* 32:207–213
- Williams PA, Sayers JR (1994) The evolution of pathways for aromatic hydrocarbon oxidation in *Pseudomonas*. *Biodegradation* 5:195–217.
- Xue W, Warshawsky D (2005) Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol Appl Pharmacol* 206:73–93
- Yano H, Garruto CE, Sota M, Ohtsubo Y, Nagata Y, Zylstra GJ, Williams PA, Tsuda M (2007) Complete sequence determination combined with analysis of transposition/site-specific recombination events to explain genetic organization of IncP-7 TOL plasmid pWW53 and related mobile genetic elements. *J Mol Biol* 369:11–26
- Yen KM, Gunsalus IC (1985) Regulation of naphthalene catabolic genes of plasmid NAH7. *J Bacteriol* 162:1008–1013
- Yen KM, Serdar CM (1988) Genetics of naphthalene catabolism in pseudomonads. *Crit Rev Microbiol* 15:247–268
- Zhou NY, Al-Dulayymi J, Baird MS, Williams PA (2002) Salicylate 5-hydroxylase from *Ralstonia* sp. strain U2: a monooxygenase with close relationships to and shared electron transport proteins with naphthalene dioxygenase. *J Bacteriol* 184:1547–1555

Sinorhizobium meliloti Megaplasms and Symbiosis in *S. meliloti*

Frédérique Barloy-Hubler(✉) and Mohamed Jebbar

Contents

1	<i>Sinorhizobium meliloti</i> Genome Architecture	92
2	Megaplasmid Genomic Features.....	93
2.1	pSymA and pSymB Replication and Mobilization.....	93
2.2	pSymB, a Second Chromosome?.....	95
2.3	pSymB and Cell Division	95
2.4	pSymB and Adaptation to the Environment	96
2.5	pSymA, a Symbiotic Accessory Megaplasmid?.....	97
2.6	<i>S. meliloti</i> Genome Dynamism	98
3	Plasmid Functions.....	100
3.1	Transport and Catabolism of Sugar and Others Components.....	100
3.2	Exopolysaccharides.....	103
3.3	Nodulation and the Plant–Microbe Relationship.....	104
3.4	Nitrogen Fixation.....	107
4	Conclusions on the Significance of <i>S. meliloti</i> Megaplasms	109
	References.....	110

Abstract *Sinorhizobium meliloti* is a Gram-negative soil bacterium forming a symbiotic nitrogen-fixing relationship with legumes such as *Medicago sativa*. All strains analyzed so far contain three replicons: one chromosome and two inherently stable megaplasms, the maintenance of which is not only due to their contribution to the cell viability but also to enhancement of the competitiveness of the cells in their natural environment. pSymB contains both plasmid and chromosomal features and is designed as a second chromosome, whereas pSymA is considered as a symbiotic accessory megaplasmid, as it can be cured without affecting *S. meliloti* viability. *S. meliloti* genome architecture was shown to be highly dynamic, as the three replicons

F. Barloy-Hubler

CNRS UMR 6026. Interactions Cellulaires et Moléculaires, Groupe B@sic, Université de Rennes 1, IFR140 GFAS, Campus de Beaulieu, Av. du Général Leclerc, 35042, Rennes, France
e-mail: fhubler@univ-rennes1.fr

M. Jebbar

CNRS UMR 6197. Laboratoire de Microbiologie des Environnements Extrêmes, Université de Bretagne Occidentale (Brest), Institut Universitaire Européen de la Mer (IUEM), 29280 Plouzané, France

continuously cointegrate and excise. Many of the genes identified on pSymA and pSymB in *S. meliloti* are involved in the formation and functioning of nitrogen-fixing root nodules. Genes located on pSymA are necessary for nodulation and nitrogen fixation, while those located on pSymB are involved in exopolysaccharide synthesis and uptake of various nutrients. Functional analyses of the genome have contributed to our understanding of the influence of the megaplasmids on *S. meliloti*'s metabolic and symbiotic abilities as well as of its successful occupation of natural niches, soil survival, plant colonization, and atmospheric dinitrogen fixation.

1 *Sinorhizobium meliloti* Genome Architecture

The existence of large plasmids in bacteria that form associations with plants was described over 20 years ago (Rosenberg et al. 1982). *Sinorhizobium meliloti* (formerly *Rhizobium meliloti*) is a Gram-negative soil bacterium forming a symbiotic nitrogen-fixing relationship with legumes such as alfalfa (*Medicago sativa*). All strains of *S. meliloti* analysed so far contain three replicons: one chromosome and two megaplasmids (Fig. 1). The smaller megaplasmid is called pSymA, pNod-Nif, or pRmeSU47a and the larger one pSymB, pExo, or pRmeSU47b (Sobral et al. 1991; Van Sluys et al. 2002). However, the sizes of these replicons differ among strains: pSymA of strain 1021 is 1.35 Mb large vs. 1.63 Mb in strain ATCC 9930 and pSymB of strain 1021 is 1.68 Mb large compared with 1.82 Mb for strain ATCC 9930 (Galibert et al. 2001; Guo et al. 2005). The two megaplasmids both play roles in symbiosis: pSymA carries symbiotic genes essential for nitrogen fixation and root nodulation (Banfalvi et al. 1981; Rosenberg et al. 1981) and pSymB genes are important for interactions with host legumes (Finan et al. 1986). The complete genome of *S. meliloti* has been sequenced and annotated in 2001 (Barnett et al. 2001a; Capela et al. 2001; Finan et al. 2001; Galibert et al. 2001). Differences in the GC contents of the replicons imply that their evolutionary histories are different; the G + C content of pSymA (60.4%) is lower than the chromosome (62.7%) and pSymB (62.4%). Because of this difference, it has been suggested that pSymB was probably acquired before pSymA (Galibert et al. 2001; Downie and Young 2001). In addition to these three replicons, many indigenous *S. meliloti* strains possess one or more smaller accessory plasmids. One of these accessory plasmids, pSmeSM11a (144,170 bp, 160 genes) was recently sequenced (Stiens et al. 2006). One-third of the whole pSmeSM11a contains a 42,367-bp long region very similar to pSymA loci (*smA1076 to smA1169*). However, synteny is not completely preserved since a pSymA region (10 kb from *smA1092 to smA1115*) is missing on pSmeSM11a. It was speculated that in *S. meliloti* strain SM11, a part of pSymA was transferred to pSmeSM11a. Another accessory plasmid of strain SM11, pSmeSM11b, contains a 5,188 bp segment, which contains 84% identity to a pSymB region (*smb20458 to smb20463*) involved in sugar metabolism and/or in polysaccharide catabolism. This locus is flanked by complete and partial transposable elements (Stiens et al. 2007).

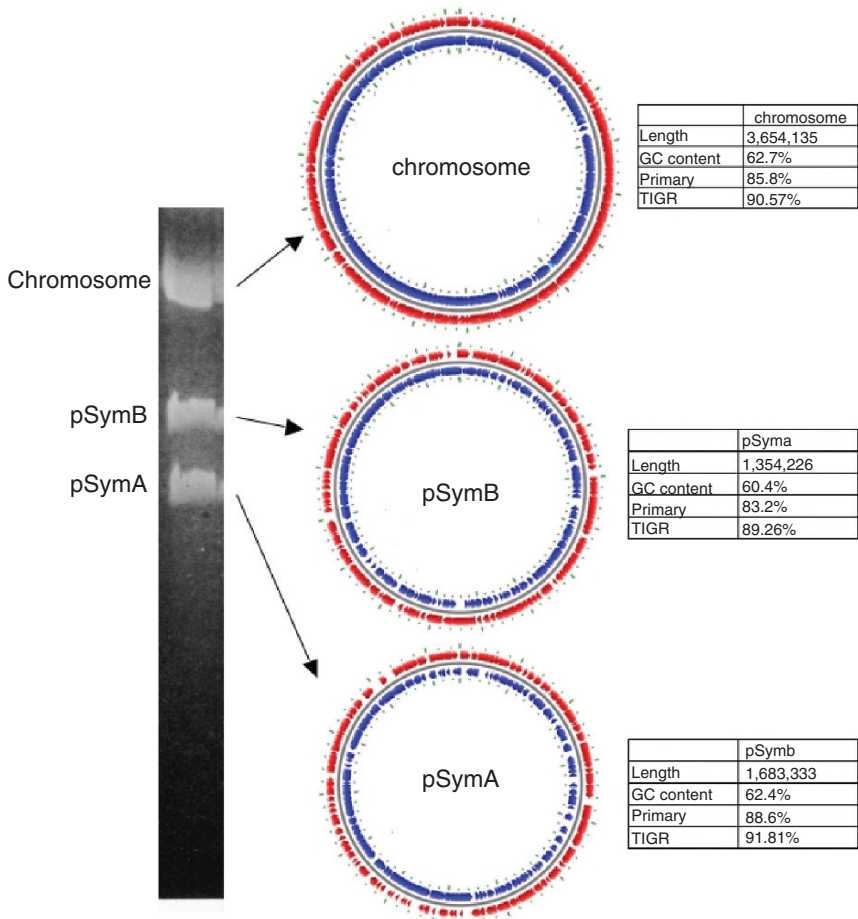


Fig. 1 The three replicons of *S. meliloti* strain 1021: PFGE-electrophoretic separation, ORF maps, and basic genomic information is shown

2 Megaplasmid Genomic Features

2.1 *pSymA* and *pSymB* Replication and Mobilization

The origins of replication of both megaplasmids have been genetically characterized (Chain et al. 2000; MacLellan et al. 2005). Both contain *repABC* genes that control replication and correct segregation during cell division. The *repABC* genes are encoded in a direct series raising the possibility that the locus is encoded as a single transcriptional unit (Ramirez-Romero et al. 2000). The *repA* and *repB* genes encode homologues of the ParAB family of replication partitioning proteins (Bignell and Thomas 2001), and *repC* is believed to encode the replication initiation

(replicase) protein, because its expression is necessary and sufficient for autonomous replication of *repABC* basic replicons (Bartosik et al. 1998; Ramirez-Romero et al. 2000; MacLellan et al. 2005).

The two replicons encode strong incompatibility (*inc*) determinants: *inc α 2* and *inc γ* on pSymA and *inc α 1* on pSymB (MacLellan et al. 2005; MacLellan et al. 2006). The *inc γ* centromere-like site is a palindrome that interacts with *repA2B2* genes to segregationally stabilize pSymA in a population of dividing cells. The *inc α 1* and *inc α 2* incompatibility determinants mediate incompatibility in a replicon-specific manner, through a *cis*-acting 56-nt untranslated-RNA named *incA* (for intergenic nucleotide sequence that influences plasmid compatibility gene A) (Izquierdo et al. 2005). These RNA molecules possibly act as a negative posttranscriptional regulator of *repC* (key replication initiator).

pSymB has another autonomously replicating sequence (ARS) of 800 bp, which was mapped within a region nonessential for pSymB replication (Margolin and Long 1993). This sequence required *trans*-acting factors from another unidentified portion of pSymB to function. Since *S. meliloti* derivatives that lack the ARS region retain pSymB, it was suggested that pSymB contains multiple replication origins. The ARS and *repABC* regions shared no homologous sequence motifs, and they are clearly distinct regions (MacLellan et al. 2005). It would be interesting to test whether the *repABC* genes are sufficient to initiate replication at the ARS site.

Recent studies using Cy3-labeled FISH localization showed that in *S. meliloti*, the replication origins of the chromosome and two megaplasmids are located at the cell poles (Kahng and Shapiro, 2003). The *repABC* replicator regions for the pSymA and pSymB megaplasmids were also scanned for DnaA-binding sites (Sibley et al. 2006), since the specific interaction between the chromosomal origin of replication (*oriC*) and DnaA ATPase is required for bidirectional replication (Fuller et al. 1981; Fuller et al. 1984). DnaA has also been described as being a transcriptional factor able to repress or activate expression from several other genes (Messer et Weigel, 1997). Five and eleven putative DnaA-binding sites were identified in the pSymA and pSymB *repABC* regions, respectively. Thus, it was proposed that the chromosomal DnaA protein may be involved in regulating megaplasmidic *repABC* gene expression (Sibley et al. 2006). If experimentally confirmed, this DnaA-dependence may explain the coordinated replication of the three *S. meliloti* replicons.

In several *Rhizobium* species, megaplasmids have been shown to be self-transmissible at variable frequencies under laboratory conditions, but only in few cases has transfer between native bacteria in soil microcosms been demonstrated (Kinkle and Schmidt 1991). A conjugatively transmissible or mobilizable plasmid must contain origins of conjugative transfer (*oriT*, *mob*). DNA hybridization experiments showed that pSymA and pSymB contain at least, four and one *mob* DNA regions, respectively (Herrera-Cervera et al. 1998). Multiple *oriTs* in megaplasmids allow efficient DNA mobilization to rhizobia as well as to phylogenetically distant gram-negative bacteria (Herrera-Cervera et al. 1998), and other studies investigating self-transfer of *S. meliloti* megaplasmids showed also that transfer of entire megaplasmids occurs at low frequency in nodules, where conditions are comparable to optimal laboratory conditions (Pretorius-Guth et al. 1990). According to these results, it is

most probable that plasmid transfer does occur within the nodules of coinfecting seedlings. This implies that genetically modified strains released in the field could potentially transfer their plasmids to other soil strains. Hence, the GM strains should be altered to prevent uncontrolled plasmid transfer.

2.2 *pSymB, a Second Chromosome?*

The pSymB replicon comprises 25% of the total DNA and coding capacity of the whole genome (Wong et al. 2002) with 1,683,333 bp and an overall GC % of 62.4. Ninety percent of this sequence is predicted to be protein-coding, constituting 1,570 ORFs with a mean length of 959 bp. The distribution of ORFs on the forward and reverse strands is symmetrical, with 52% on the forward strand and 48% on the reverse strand (Finan et al. 2001). This replicon contains both plasmid (*oriV*, *repABC* genes, absence of *rrn* genes) and chromosomal features (the sole tRNA^{arg} (CCG) of the genome, *minCDE* and *ftsK* genes). pSymB resembles the *S. meliloti* chromosome signature in terms of nucleotide composition (GC contents and dinucleotide relative abundances). Most of pSymB genes are nonessential to cell viability as a large portion of this replicon can be deleted without loss of viability, the smallest derivative constructed being 450 kb in size. However, strains cured of pSymB cannot be produced and a ca. 200 kb region was recalcitrant to deletion (Charles and Finan 1991)

Results suggested that genetic exchange and recombination have played a significant role in natural populations of *S. meliloti*. Using a whole-replicon nearest-neighbour analysis, Wong and Golding (2003) have shown that strain 1021 pSymB has a complex evolutionary history, with sequence matches to diverse groups of organisms. They estimate that ca. 13% of pSymB genes have been involved in horizontal gene transfer and that 87% are the result of duplication or have orthologs in Rhizobiaceae. Hence, it was proposed that this replicon may have a plasmidic origin and may have acquired chromosome-like features after long-term residence and exposure to the *S. meliloti* transcription/translation machinery. Another scenario is that a chromosomal DNA segment was converted into a megaplasmid or ancillary chromosome, which then lost some genes and gained others. However, distinct features of pSymB (gene specialization, low abundance of IS elements, and high proportion of orphan genes argue against a chromosomal origin for pSymB (Finan et al. 2001). Unlike the chromosome, pSymB does not have a region syntenic with *A. tumefaciens* chromosomes, and its origin and evolutionary history is still unclear. Its designation as a *S. meliloti* second chromosome was proposed and is still a matter of debate.

2.3 *pSymB and Cell Division*

In *S. meliloti*, the *minCDE* genes are located in pSymB. Orthologous *minCDE* loci are carried by the *Agrobacterium tumefaciens* linear replicon and the *Brucella melitensis* chromosome II, pointing to the possibility that these genes are essential

for the segregation of nonchromosomal replicons in alpha-proteobacteria (Young et al. 2006). In *S. meliloti*, the *minCDE* genes are transcribed as an operon. The 174-nt 5'-UTR region upstream of *minC* contains a presumed RpoD as well as two potential CtrA-binding sites (Barnett et al. 2001a). These genes are expressed in free-living cells where they play a role in cell septa localization/formation. Expression of the *minCDE* genes is also detected during infection and symbiotic processes (Cheng et al. 2007). However, Δ *minCDE* mutants revealed the existence of additional systems that control cell septation in *S. meliloti*, since the *min* genes are not essential for cell viability or division in this organism. The fact that the *min* genes are dispensable in *S. meliloti* is consistent with their megaplasmid location. Moreover, the *minCDE* genes are located on plasmid p42e of *Rhizobium etli* (Gonzalez et al. 2006) and on plasmid pRL11 of *Rhizobium leguminosarum* (Young et al. 2006). Furthermore, even if the *minCDE* genes are not absolutely required for the bacteroid differentiation process, they are expressed in nodules, demonstrating the existence of redundancy in the septa formation pathway(s) in *S. meliloti*.

pSymB also carries a gene called *ftsK2* (homolog of SpoIIIIE) located close to the *repA3B3* replication locus. FtsK is involved in septa formation and recombinational resolution of dimeric chromosomes. However, another *ftsK* gene (*ftsK1*) was found on the chromosome and this copy is sufficient for cell division, as *ftsK2* deletion derivatives showed no obvious growth defects (Finan et al. 2001).

2.4 *pSymB* and Adaptation to the Environment

Genes encoding transport systems constitute the largest (12%) class of genes in the *S. meliloti* genome with 387 predicted transport systems encompassing 849 genes (Fig. 2). An important finding to emerge from the analysis of the pSymB nucleotide sequence is the large number of solute transport systems as well as genes involved in the biosynthesis and export of surface polysaccharides (essential for nodule invasion and adaptation to rhizosphere environments) and genes involved in nitrate/nitrite reduction. Transport systems represent 18.9% of the pSymB orfeome against 10.4% for the chromosome. In addition, pSymB contains genes important for catabolic activities (Finan et al. 2001). Thus, it appears that this replicon allows the bacteria to take up and presumably oxidize many different compounds from the environment, playing an important role in the survival of the bacteria under diverse nutritional living conditions. Although it is noticeable that pSymB has few genes, if any, directly involved in nodulation and nitrogen fixation, it apparently plays a role in the adaptation to the endosymbiotic lifestyle.

The majority of *pSymB* genes appears to have a constitutive expression, comparable to chromosomal genes expression (Barnett et al. 2004). However, *pSymB* genes are up-regulated during osmotic stress, and a 200-kb region of pSymB required for adaptation to saline shock was identified (Dominguez-Ferreras et al. 2006). Among the salt stress-induced *pSymB* genes are genes involved in trehalose transport and utilization: the trehalose 6-phosphate synthase gene (*otsA*), an operon (*smb20095 to smb20100*) containing a putative trehalose synthase gene (*smb20099*),

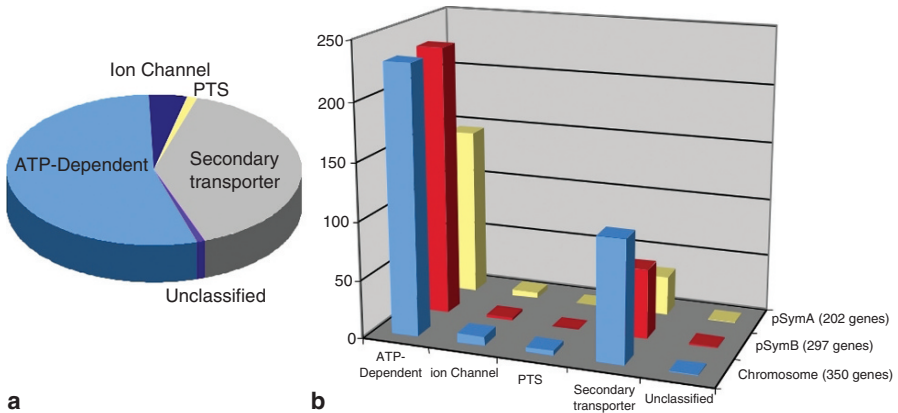


Fig. 2 Distribution of transport systems in *S. meliloti*. **(a)** Breakdown of *S. meliloti* transporters by category. **(b)** Distribution of the transport systems by category on the three *S. meliloti* replicons. Data are from <http://www.membranetransport.org/transporter2.php?oOID=smel1>

and the *thuA*, *thuE*, and *thuG* genes. If for a given gene both plasmidic and chromosomal alleles are present, the pSymB allele is induced under high-salt conditions whereas the chromosomal allele is not (for instance, the pSymB *glgA*, *glgB2*, and *glgX2* genes involved in glycogen metabolism were expressed but the chromosomal alleles *glgA1*, *glgB1*, and *glgX1* were not). Indeed, 41% of salt up-regulated genes are carried on pSymB (vs. 22% on pSymA and 37% on the chromosome), suggesting that this plasmid is essential for *S. meliloti* adaptation to hyperosmotic stress. It appears that acquisition of pSymB significantly extended the metabolic capabilities of *S. meliloti* by allowing this bacterium to metabolize a large variety of small compounds encountered in the soil or in the plant rhizosphere. An increased capacity in synthesizing polysaccharides may also have significantly improved the colonization potential of these microbes.

2.5 *pSymA*, a Symbiotic Accessory Megaplasmid?

pSymA is 1,354,226 nt long and 1,293 putative genes were identified, yielding a coding capacity of 83.6%. pSymA proteins linked to nitrogen metabolism are organized in a 53-kb segment encoding a complete pathway for denitrification (a NapAB-type periplasmic dissimilatory nitrate reductase, a nitrite reductase, and a nitric oxide reductase encoded by the *nor* operon). Other pSymA proteins important in denitrification are: NnrU, AzuI, HemN (involved in heme maturation) and two regulatory proteins (NnrR and NnrS). Previously identified genes encoding nitrous oxide reductase (*nos*) are also located on this replicon together with putative genes for nitrate transport (*nrtAB*) (Barnett et al. 2001b). pSymA encodes two tRNAs: one specifies methionine redundant with respect to a chromosomal allele. The other one with a UCA anticodon specifies selenocysteine and is unique in the genome.

The gene *selC*, encoding the seryl tRNA, is adjacent to genes encoding Sella (selenocysteine synthase), SelD (required to modify seryl tRNA to selenocysteine tRNA), and SelB (selenocysteine-specific elongation factor). A transposon separates *selA* and *selB* from *selC* and *selD*. Housekeeping functions are also encoded on pSymA but they are redundant (GroES, GroEL, UvrD2, DnaE3, RpoE6, DNA ligase, and DNA-damage inducible protein). A number of pSymA ORFs may be involved in stress responses (three cold shock proteins, one heat shock protein, a hydroperoxidase, and haloperoxidases). Many genes found on pSymA are similar to interesting genes of other bacteria but have no obvious symbiotic function. pSymA is clearly specialized in nodulation and nitrogen fixation. The many pSymA genes involved in nitrogen metabolism point to a significant role in providing versatility for dealing with nitrogen in many oxidation states and chemical combinations. Physiologically, low oxygen conditions characterize the nodule environment and may also be encountered by rhizobia in soil. Thus it is interesting that symbiotic genes are linked to other genes likely to be useful under low oxygen conditions.

pSymA genes involved in nitrogen fixation are clustered within a 275-kb region that includes *nod* genes required for synthesis of Nod factor as well as the *nol* and *noe* genes, which are encoded in six operons (Schlaman, 1998). Previously discovered *nif* and *fix* genes for symbiotic nitrogen fixation also lie within the 275-kb region (Kaminski et al. 1998). Most of the sequences located on pSymA are transcribed only at the bacteroid stage (Barnett et al. 2004). It has been proposed that genes expressed in bacteroids are not robustly influenced by translational selection (Patriarca et al. 2002). Indeed, results show that genes located on pSymA are less adapted to optimal codon usage (Karlin et al. 2003; Peixoto et al. 2003) and differ more with the respect to synonymous substitutions, despite the fact that some pSymA-encoded proteins are present at the same levels as some chromosomally-encoded proteins in bacteroids (Natera et al. 2000).

pSymA can be cured without affecting growth on either rich or minimal succinate media (Oresnik et al. 2000), although the cured strain is unable to use certain carbon sources such as gluconate and will not grow with ammonium as the sole nitrogen source. These observations suggest that pSymA plays a role in carbon and ammonium assimilation. Comparative proteome analysis of wild-type strain 2011 and a pSymA-cured derivative show 60 differentially (up- and down-) expressed proteins and suggest that pSymA has a role in the regulation of the expression of genes from the other replicons (Chen et al. 2000). Finally, recent studies showed that a pSymA-cured derivative of *S. meliloti* is unable to cause rice growth inhibition as the wild type does. Plant inhibitory genes are probably associated with pSymA, as no phenotype was observed in a pSymB-cured mutant (Perrine et al. 2005).

2.6 *S. meliloti* Genome Dynamism

S. meliloti genome architecture was shown to be highly dynamic as the three replicons continuously cointegrate and excise (Guo et al. 2003). Comparison of the genome sequences of different strains revealed considerable genomic plasticity,

although overall genomic organization is similar within a given species (Boucher et al. 2001). Although pSymA could be cured without affecting viability (at least under laboratory conditions), studies indicated that *in vitro* constructed *S. meliloti* containing only one or two replicons grew more slowly in minimal medium than the wild-type strain, arguing that the three-replicon architecture provides a selective advantage (Guo et al. 2003).

The genome of bacteria belonging to the plant-interacting *Rhizobiaceae* family is characterized by a large number of mobile genetic elements (MGEs), especially many insertion sequence (IS) elements, transposases, and transposition-related genes (Freiberg et al. 1997). These transposable elements are considered to be responsible for the major part of the genetic variability found in these bacteria (Laberge et al. 1995). Up to now, 21 types of ISs have been identified on the *S. meliloti* genome: 4 are chromosome-specific, 4 are pSymA-specific and 1 is pSymB-specific (Wheatcroft and Laberge 1991; Mahillon and Chandler 1998; Zekri et al. 1998a; Zekri and Toro 1998b; Selbitschka et al. 1999). Some of them have been identified inside or flanking the symbiotic genes (Ruvkun et al. 1982; Dusha et al. 1987). The occurrence of multiple ISs and MGEs has a strong impact on the structure and stability of the *S. meliloti* genome. Accumulation of IS elements have been reported for pSymA. This replicon carries most of the variable genes and the highest percentage of mobile genetic elements of all three replicons (3.6% for pSymA, 0.9% for pSymB, and 2.6% for the chromosome). As transposable elements tend to disrupt essential cellular genes, their accumulation in pSymA is consistent with the fact that this replicon is not essential for cell survival. Additionally *S. meliloti* strain 1021 contains a group II intron (RmInt1) (Toro et al. 2003) only found on megaplasmids. Two copies are located on pSymB and one on pSymA. RmInt1 mobility is based on retrohoming via an RNA intermediate (Martinez-Abarca et al. 2004). RmInt1 is self-splicing and upon acquisition by conjugation spreads in the *S. meliloti* genome, primarily by retrohoming to a target sequence in ISRm2011–2 (Nisa-Martinez et al. 2006). IS Rm2011–2 seems to be conserved throughout *S. meliloti* genome evolution, since almost 100% of the isolates tested carry this IS in the three replicons. Ten percent of the *S. meliloti* isolates tested may have lost RmInt1 early in their evolution, and it is possible that this intron was acquired and lost repeatedly during the evolution of *S. meliloti*. RmInt1 is also present in other *Sinorhizobium* and *Rhizobium* species, acquired by vertical inheritance from a common ancestor and by independent horizontal transfer events. RmInt1 is mobile in related taxa of bacteria that interact with plants and tends to evolve toward an inactive form by fragmentation, with loss of the 3'-terminus including the intron-encoded protein (Fernandez-Lopez 2005).

Genes responsible for nodulation and nitrogen fixation (*sym* loci) are thought to have been acquired through a horizontal transfer event. It has been suggested that the differences between pSymA and chromosomal genes (GC content and codon usage) are due to a "recent" horizontal transfer event (Galibert et al. 2001). The ability of the *sym* loci to transfer across divergent chromosomal lineages has been shown in the laboratory (Herrera-Cervera et al. 1998), in alfalfa nodules (Pretorius-Güth et al. 1990) as well as in agricultural populations of a relative of *S. meliloti*, *Mesorhizobium loti* (Sullivan et al. 1995). Transfer of plasmids containing the *sym* loci restores the nodulating ability of a *S. meliloti nod-nif* deletion-mutant (Wong et al. 1983). Additionally,

transfer of a 290-kb pSym derivative containing nitrogenase and *nod* genes into the pathogen *A. tumefaciens* results in specific root deformation in *Medicago sativa* and *Melilotus alba* host plants susceptible to nodulation by *S. meliloti*, suggesting that these pSym genes are important for host-specific nodulation (Truchet et al. 1984).

Loss or gain of megaplasmids (or portions of megaplasmids) among strains in natural populations may occur frequently. Recent screening for DNA sequences absent in the *S. meliloti* 1021 laboratory strain but present in natural isolates identified multiple gains and losses of DNA segments on the two megaplasmids and, to a lesser extent, on the chromosome as well (Guo et al. 2005). These gains and losses might have contributed to the phylogenetic incongruences of these three replicons. Horizontal gene transfer (HGT), found in many natural bacterial populations, including nitrogen-fixing bacteria (Silva et al. 2005; Vinuesa et al. 2005) has also been postulated. However, recent studies support the lack of lateral transfer among major genomic subdivisions (Wernegreen et al. 1997) but frequent horizontal transfer events among congenics (Wernegreen and Riley 1999). Although there is no direct evidence for natural genetic exchange, pSymA contains a cluster of genes encoding a type IV pilus (Galibert et al. 2001), a surface structure involved in adhesion to host cells, infection by bacteriophages and in conjugative DNA transfer among strains (Ashelford et al. 2003).

Analysis of 49 strains shows that unique sequences are more often found in megaplasmids than in chromosomes (43 unique sequences for pSymA, 46 for pSymB, and 34 for the chromosome) (Sun et al. 2006). The proportion of orphan genes was significantly higher on the megaplasmids than on the chromosome, with 11.5% on pSymA and 12.3% on pSymB. Comparative analysis of 49 natural *S. meliloti* strains showed that divergence among chromosomal genes is lower than among megaplasmid ones (Sun et al. 2006). There are two hypotheses concerning this phenomenon: (1) genes on megaplasmids accumulate more mutations due to less functional constraints than genes on the chromosome. These observations are consistent with the results of laboratory studies showing that extensive regions of megaplasmids can be deleted with no consequences for fitness (Charles and Finan 1991; Oresnik et al. 2000). (2) The second hypothesis is that genes on pSymA and pSymB are highly niche-specific and that their divergence is correlated with host specialization and life-style variability. The most divergent gene of *S. meliloti* is *smbExoF3*, located on pSymB and encoding a putative outer-membrane protein. This might have major functional implications in host recognition and niche specialization.

3 Plasmid Functions

3.1 Transport and Catabolism of Sugar and Others Components

The two megaplasmids pSymA and pSymB of *S. meliloti* carry many genes that are involved in uptake and assimilation of various organic and inorganic nutrients such as dicarboxylate acids, sugars, aromatic compounds, osmoprotectants (betaines,

ectoines), iron, and inorganic phosphate. These compounds are not often obviously associated with the host–bacteria relationship; they have been shown to play a trophic role, being used as nutrient sources by members of the family *Rhizobiaceae* (Boivin et al. 1991). pSymB is known to encode thiamine biosynthesis genes (Finan et al. 1986) and to contain genes that are involved in the synthesis of an exopolysaccharide required for effective nodulation (Leigh et al. 1985; Finan et al. 1986). This plasmid also carries three *dct* genes, *dctA*, and *dctBD*, encoding the C₄-dicarboxylate transport (Dct) system (Ronson et al. 1984). The *dctA*-encoded product consists of a dicarboxylate carrier protein (Yarosh et al. 1989) required both for symbiotic nitrogen fixation and for the transport of succinate, fumarate, malate, and aspartate (Watson et al. 1988). The *dctB* and *dctD* genes encode proteins sharing homology with other proteins comprising two-component kinase regulatory systems (Watson 1990). The expression of *dctA*, which is transcribed in the opposite direction to *dctBD*, is regulated both by DctB and DctD proteins and by tricarboxylic acid cycle intermediates succinate, fumarate, and malate (Watson et al. 1988; Watson 1990). In many rhizobia, mutations in *dctBD* lead to ineffective nodules but nodules induced by *S. meliloti* strains with mutations in the *dctBD* genes can fix nitrogen at nearly wild-type rate (Engelke et al. 1989; Yarosh et al. 1989; Watson 1990; Yurgel and Kahn 2004). Thus, an alternative symbiotic regulatory mechanism operates in *S. meliloti*, which is able to activate the *dctA* gene in nodules and is still not understood.

The rhizosphere is a thin zone that surrounds the plants roots, where plant secretions enrich this environment with organic compounds like sugars. Arabinose, a pentose sugar, is a component of plant cell walls that is found in the rhizosphere. This carbon source supports the growth of many rhizobia (Stowers 1985), and in *S. meliloti*, a gene cluster (*araABCDEF*) on pSymB necessary for arabinose transport and catabolism, was genetically characterized (Poysti et al. 2007). *araABC* encode components of an ABC-type transporter, while *araD*, *araE*, and *araF* encode an hydratase, an aldehyde dehydrogenase, and a dehydratase, respectively (Poysti et al. 2007). Also, α -galactoside uptake was investigated by screening for genes under control of the regulatory proteins NodD3, SyrM, and SyrA that activate genes required to form nitrogen fixing nodules (Mulligan and Long 1989). Using this approach, a gene encoding a periplasmic-binding protein (*agpA*) was characterized (Mulligan and Long 1989). AgpA is required for uptake and utilization of α -galactosides such as melibiose and raffinose. The *agpA* gene is upregulated in the presence of α -galactosides and downregulated by the *syrA* gene product and also by glucose and succinate (Gage and Long 1998). Substrates of α -galactoside transport system, of which AgpA is a component, may have a role in growth and survival of bacterial cells in the rhizosphere as the seeds of many legume are especially rich in α -galactosides (Obendorf 1997). *S. meliloti* might use these sugars to support growth, if they were to be released by germinating seeds or the roots of young plants.

Many other functions are encoded on pSymB, such as gene products that are required for growth on the short-chain fatty acids acetoacetate and hydroxybutyrate, which are intermediates in the polyhydroxybutyrate (PHB) degradation pathway (*bhbA-D*) (Charles et al. 1997). It seems that PHB, a rhizobial carbon storage compound, is used as a carbon and energy source to support bacterial cell

proliferation within the infection thread, where nutrients from the plant may be limited (Charles et al. 1997).

Aromatic compounds constitute another class of carbon source of which degradation pathways are widespread among soil bacteria, including *S. meliloti* (Harwood and Parales 1996). These compounds accumulate primarily as the result of the degradation of different plant-derived aromatic polymers, including lignin. Many aromatic substrates such as lignin-derived monomers may be converted to protocatechuate or catechol, which then are degraded via the β -keto adipate pathway to tricarboxylic acid intermediates (Harwood and Parales 1996). The β -keto adipate pathway was characterized in *S. meliloti*, where it is organized in two transcriptional units named *pcaDCHGB* and *pcaIJF* that mapped on the pSymB replicon (MacLean et al. 2006). The *pcaDCHGB* operon encodes enzymes that transform protocatechuate to β -keto adipate, while the *pcaIJF* operon encodes another set of enzymes involved in the transformation of β -keto adipate into tricarboxylic acid intermediates. In *S. meliloti*, a gene encoding a LysR-type transcriptional regulator PcaQ adjacent to and transcribed in opposite direction to the *pcaDCHGB* operon regulates the expression of this operon from a promoter mapped upstream of *pcaD* (MacLean et al. 2006). The expression of the *pcaIJF* operon required an IclR-type regulator encoded by *pcaR* gene and β -keto adipate is considered as the *in vivo* metabolite inducer responsible for *pcaIJF* expression in *S. meliloti* (MacLean et al. 2006).

Secondary metabolites like betaines (*N*-methylammoniums) and ectoines (Tetrahydropyrimidines) produced by plants and bacteria are used by *S. meliloti* both as powerful osmoprotectants and as a carbon, nitrogen, and energy source (Goldmann et al. 1991; Talibart et al. 1997; Jebbar et al. 2005). Genes essential for betaine catabolism are located on pSymA in the symbiotic region (Goldmann et al. 1991), while a gene cluster encompassing nine genes (*ehuABCD-eutABCDE*) located on pSymB is involved in ectoine uptake and catabolism (Jebbar et al. 2005). The *ehuABCD* genes encode components of a specific ectoine-hydroxyectoine ABC-type transporter, and the *eutABCDE* operon encodes enzymes allowing ectoine degradation in *S. meliloti*.

S. meliloti megaplasmids also contain loci involved in the assimilation of inorganic nutrients such as phosphate and iron. A phosphate ABC-type transport system, which is encoded by the *phoCDET* genes carried by pSymB, was characterized in *S. meliloti*; this system is presumably used by bacteria in the soil environment to scavenge inorganic phosphate available at very low concentrations (Bardin et al. 1996). Rhizobactin 1021, a siderophore that chelates ferric iron, is produced under iron stress by *S. meliloti* (Lynch et al. 2001). The genetic organization of the cluster, encompassing eight genes that function in the regulation, biosynthesis, and transport of rhizobactin 1021, was identified on pSymA in *S. meliloti* (Lynch et al. 2001). Six of the genes (*rhbABCDEF*), which constitute an operon, function in the biosynthesis of the siderophore and are repressed under iron-replete conditions. One gene of the cluster named *rhtA* encodes the outer membrane receptor protein for the siderophore, and the transcription of both the *rhbABCDEF* operon and the *rhtA* gene is upregulated by the *rhrA* gene product, which shares homologies with the AraC-type regulator family (Lynch et al. 2001). During symbiosis, the genes *rhbABCDEF* and *rhtA* are

inactive (Lynch et al. 2001). Rhizobactin 1021 is most likely an element that helps free-living *S. meliloti* cells to compete for iron in iron-depleted soils.

3.2 Exopolysaccharides

Rhizobial megaplastids have been characterized extensively with regard to symbiosis. Among the known genes required for nitrogen fixation (Finan et al. 1986), the genes responsible for synthesis of different types of exopolysaccharides (EPS) play an important role. Most of these genes are located on pSymB. Exopolysaccharide-deficient mutants of *S. meliloti* fail to invade the host legume and to fix nitrogen (Finan et al. 1985; Leigh et al. 1985). The best studied rhizobial EPS is an acidic exopolysaccharide termed succinoglycan (EPS I) produced by *S. meliloti* (Reinhold et al. 1994). EPS I is a high-molecular-weight polymer composed of repeating octasaccharide subunits, whereby each subunit is composed of one galactose and seven glucose residues, joined by β -1,3, β -1,4 and β -1,6 glycosidic linkages, and carries succinyl, acetyl, and pyruvyl modifications (Reinhold et al. 1994). *S. meliloti* has a cryptic capacity to produce a second exopolysaccharide (ESP II) named galactoglucan (Her et al. 1990) composed of repeating disaccharide units. Each unit contains an acetylated glucose and a pyruvylated galactose linked by α -1,3 and β -1,3 glycosidic bonds (Her et al. 1990). Galactoglucan is able to substitute for succinoglycan in symbiosis (Glazebrook and Walker 1989; Zhan et al. 1989).

Genes directing the biosynthesis of succinoglycan form a 25-kb cluster (Glucksmann et al. 1993b) on pSymB, composed of nineteen genes (*exoP*, *exoN*, *exoO*, *exoM*, *exoA*, *exoL*, *exoK*, *exoH*, *exoI*, *exoT*, *exoW*, *exoV*, *exoU*, *exoX*, *exoY*, *exoF*, *exoQ*, *exoZ* and *exoB*). *exoC* is the only gene involved in succinoglycan biosynthesis that is located on the *S. meliloti* chromosome (Finan et al. 1986). *exoB* and *exoC* mutations affect not only succinoglycan production but also affect the synthesis of other polymers (galactoglucan, lipopolysaccharide, and β -glucans), and have been shown to be defective in sugar nucleotide biosynthesis (Leigh and Lee 1988; Glazebrook and Walker 1989). Mutations in *exoA*, *exoF*, *exoL*, *exoM*, *exoU*, *exoW*, *exoP*, *exoQ*, *exoT*, and *exoY* completely abolish succinoglycan (Becker et al. 1993a; Becker et al. 1993b; Glucksmann et al. 1993b; Reuber and Walker 1993b).

exoB codes for a UDP-glucose-4-epimerase that converts UDP-glucose to UDP-galactose (Canter Cremers et al. 1990; Buendia et al. 1991), and *exoC* encodes a phosphoglucomutase that transforms glucose-6-phosphate into glucose-1-phosphate (Uttaro et al. 1990). An enzyme encoded by the *exoN* gene was found to be highly similar to a UDP-glucose pyrophosphorylase, and mutants in *exoN* produce less succinoglycan (Becker et al. 1993a; Glucksmann et al. 1993b). Synthesis of succinoglycan subunits is initiated by the products of the *exoY* and *exoF* genes that are needed for the addition of galactose to the membrane-bound isoprenoid lipid carrier (Reuber and Walker 1993b). The glucosyltransferases encoded by *exoA*, *exoL*, *exoM*, *exoO*, *exoU*, and *exoW* genes catalyse the subsequent addition of six glucose residues (Becker et al. 1993b; Glucksmann et al. 1993a). The ExoA protein

appears to be anchored in the membrane at its carboxyl terminus and this may facilitate a probable interaction with ExoY and ExoF that seem to be membrane bound (Long et al. 1988; Reed et al. 1991; Muller et al. 1993). Sequences analysis predicted that the ExoL, ExoO, and ExoU glucosyltransferases are localised in the cytoplasm, whereas the ExoM and ExoW proteins are anchored in the membrane (Becker et al. 1993a; Glucksmann et al. 1993b). Both *exoH* and *exoZ* products, which are required for the succinyl and acetyl modifications, respectively, are trans-membrane proteins (Leigh et al. 1987; Buendia et al. 1991; Becker et al. 1993a; Glucksmann et al. 1993b; Reuber and Walker 1993a). *exoV* mutants failed to produce EPS I and induced pseudonodules on alfalfa (Muller et al. 1993; Keller et al. 1995). The latter mutants accumulate subunits lacking pyruvyl modification. Thus, the modification is necessary for the polymerisation and secretion of succinoglycan (Becker et al. 1993b; Glucksmann et al. 1993b; Reuber and Walker 1993b). The products of the *exoP*, *exoQ*, and *exoT* genes appear to be needed for polymerisation and secretion of succinoglycan (Glucksmann et al. 1993b).

The second exopolysaccharide (Galactoglucan) produced by *S. meliloti* is directed by a 23-kb *exp* gene cluster carried on pSymB (Becker et al. 1997). The *exp* cluster is separated by about 200 kb from the *exo* cluster (Charles and Finan 1991). The production of this exopolysaccharide was observed only at low phosphate concentrations (Zhan et al. 1991) or in the presence of a mutation in either *expR* (Glazebrook and Walker 1989) or *mucR* (Zhan et al. 1989; Keller et al. 1995). ExpA7, ExpA8, ExpA9, and ExpA10 proteins are involved in the conversion of glucose-1-phosphate and dTTP to dTDP-rhamnose (Jiang et al. 1991; Koplín et al. 1993). Six other genes in the cluster encode β -glucosyltransferases (ExpA2 and ExpE2) and galactosyltransferases (ExpA3, ExpC, ExpE4, and ExpE7) which are involved in polymerisation of sugars (Becker et al. 1997). The *expE1* gene codes for a secreted protein that binds calcium ions. This protein displays homology to the NodO protein from *Rhizobium leguminosarum* bv. *Viciae*, which has pore-forming activity in membranes (Sutton et al. 1994). ExpD1 (ABC transporter protein of a proteinase export complex) and ExpD2 (membrane fusion protein of a proteinase export complex) represent components of a complex that might be involved in the secretion of ExpE1 (Becker et al. 1997). *expD1* and *expD2* mutants lack the ability to synthesise and secrete galactoglucan (Becker et al. 1997; Moreira et al. 2000). ExpG, which belongs to the MarR family of transcriptional regulators (Dehoux and Cossart 1995), may function as a transcriptional activator of *exp* gene expression (Becker et al. 1997).

3.3 *Nodulation and the Plant–Microbe Relationship*

The establishment of the symbiotic association between *S. meliloti* and its host legume alfalfa is mediated by the chemical signals exchanged between the symbiotic partners (Schultze et al. 1992). The plant excretes flavonoid molecules that activate *nod* genes located for the majority on the pSymA replicon. Most of the known *nod* gene products catalyse synthesis of an oligosaccharide signal (Nod Factor) that is structurally modified in a host-specific manner

The soil microorganism *S. meliloti* induces the formation of nodules on the roots of leguminous hosts, in which it fixes atmospheric nitrogen. The genes controlling functions in symbiotic association like nodulation (*nod* genes) and nitrogen fixation (*fix* and *nif* genes) are located on pSymA in *S. meliloti* (Banfalvi et al. 1981; Rosenberg et al. 1981; Meade et al. 1982; Ruvkun et al. 1982). This megaplasmid, which is not self-transmissible at detectable frequency, could be mobilized into *A. tumefaciens* by cloning a *mob* region from plasmid RP4 into pSymA; the transconjugants were able to form nodule-like structures on *Medicago sativa* roots (Wong et al. 1983). A mutant carrying a Tn5 insertion in the *nif* region (Ruvkun and Ausubel 1981), but not altered in its symbiosis properties, was used to clone a 290-kb pSymA fragment carrying *nif* and *nod* genes into an RP4-primed derived vector (Ruvkun and Ausubel 1981; Meade et al. 1982; Truchet et al. 1984). The resulting plasmid was introduced by conjugation into *A. tumefaciens* and the transconjugants were able to induce root deformations specifically on *S. meliloti* hosts legumes, which suggests that host-specificity genes were also carried by pSymA (Truchet et al. 1984). An approximately 90-kb region of the pSymA replicon contains the different structural and regulatory genes that constitute the *nod* regulon (Fig. 3 A and B). The *nod* regulon encompasses several operons that encode the enzymes that synthesize the Nod factors (Lipo-chitoooligosaccharides). Several *nif* and *fix* genes are located in the same region, interspersed with the elements that constitute the *nod* regulon. The correspondingly named *nif* genes include those encoding subunits of the nitrogenase complex (Ruvkun et al. 1982; Corbin et al. 1983), linked within 30 kb to nodulation (*nod*) genes (Meade et al. 1982; Schmidt et al. 1984).

A DNA fragment containing a cluster of four *nod* genes was cloned in pRK290 plasmid (Jacobs et al. 1985; Mulligan and Long 1985) and has been characterized by transposon mutagenesis and DNA sequencing analysis. The *nodD1* gene reads in opposite direction from an operon of at least three genes, *nodABC* (Egelhoff et al. 1985). Two additional *nodD* sequences (*nodD2* and *nodD3*) were localized in the vicinity of the other *nod* genes (Gottfert et al. 1986). The *nod/nif/fix* region of *S. meliloti* contains also 11 other *nod* genes, designated as *E*, *F*, *G*, *H*, *I*, *J*, *L*, *M*, *N*, *P* and *Q* (Debelle and Sharma 1986; Baev et al. 1991; Baev et al. 1992; Schwedock and Long 1992). The *nod* genes are organized in six transcriptional units coding for nodulation functions (*nodABCIJ*, *nodFEG*, *nodPQ*, *nodM-nolFG-nodN*, *nodL* and *nodH*) (Rostas et al. 1986; Baev et al. 1991; Baev et al. 1992; Schwedock and Long 1992). *nod* genes fall into two general categories: the common *nod* genes, such as *nodDABC*, and the host specific *nod* genes, such as *nodPQ*, *nodFEG*, and *nodH* (Rostas et al. 1986). The common *nod* genes are required for the induction of host cortical cell division and root hair curling by the bacteria during the initial stages of the infection process (Jacobs et al. 1985; Debelle et al. 1986). The expression of *nodABC* is activated by host plant inducers (flavonoids and betaines) and requires the products of *nodD1* and *nodD2* (Mulligan and Long 1985; Gottfert et al. 1986; Peters et al. 1986; Hartwig et al. 1990). The *nodD3* gene product is an inducer-independent activator of *nod* genes that requires another closely linked gene, *syrM* (Swanson et al. 1993). *SyrM* like the NodD proteins is a member of the LysR family of transcriptional activators (Henikoff et al. 1988). The *nodD3-syrM* genes constitute a self-amplifying positive regulatory circuit in free-living cells but

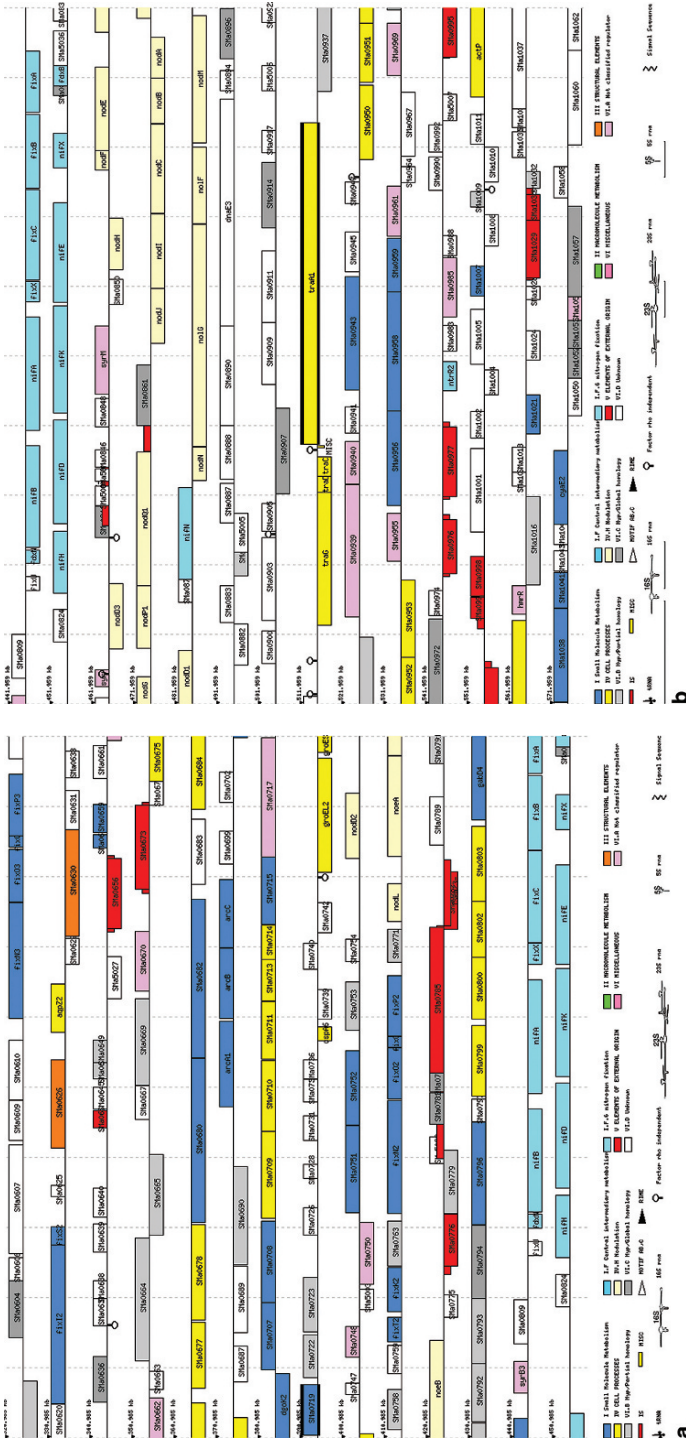


Fig. 3 (a and b): *nodIn* region on the symbiotic plasmid pSymAMap was drawn at <http://bioinfo.genopole-toulouse.prd.fr/annotation/fANT/bacterial/rhime/> see this page for color-coding scheme

not in planta, since *syrM* is expressed in the nitrogen fixation zone of mature nodules while *nodD3* is not (Sharma and Signer 1990; Swanson et al. 1993). NodD transcriptional activator proteins bind to a specific sequence in *nod* gene promoters called the *nod* box (Rostas et al. 1986; Fisher et al. 1988; Fisher and Long 1989), and the ability of NodD1, NodD3, and SyrM to bind *nod* promoters necessitates the activity of the molecular chaperone *GroEL* (Ogawa and Long 1995).

GroEL is thus probably implicated in the folding or assembly of active Nod proteins. A *nod* gene cluster (*nodM*, *nolFG*, and *nodN*) in *S. meliloti* is required for optimal nodulation of alfalfa (Baev et al. 1991). The *nodM* gene coding for glucosamine synthase is required for efficient Nod signal production. *nodL* was shown to play a role in the O-acetylation, which is required for increasing the efficiency of the Nod factor in evoking cortical cell division (Baev et al. 1992). The *nodM* and *nodN* genes have common functions, and NodM is a glucosamine synthase with the biochemical role of providing sufficient amounts of the sugar moiety for the synthesis of the glucosamine oligosaccharide Nod factors (Baev et al. 1991; Baev et al. 1992). The structural nodulation genes *nodABC* participate in the synthesis of the basic structure of the Nod-factor signals. NodC catalyses the polymerisation of UDP-glucosamine (Geremia et al. 1994), NodB is a chitooligosaccharide deacetylase (John et al. 1993), and NodA is an *N*-acyltransferase (Rohrig et al. 1994), which adds the acyl chain to the oligosaccharide core. The species-specific NodF and NodE proteins are homologous to acyl carrier proteins and β -ketoacyl synthases, which are involved in fatty acid synthesis (Shearman et al. 1986; Bibb et al. 1989). NodE and NodF are required for the synthesis of the C16 polyunsaturated chains that represent the acyl moiety of the Nod factor (Demont et al. 1993).

The sulfurylation of the Nod-factor precursor is catalysed by NodH enzyme, which utilizes 3'-phosphadenosine-5'-phosphosulphate (PAPS) as a sulfate donor; NodP and NodQ are, respectively, an ATP sulfurylase and an adenosine-5'-phosphosulphate (APS) kinase, enzymes that catalyze the transformation of ATP and SO_4^{2-} into PAPS (Schwedock and Long 1992). *S. meliloti* possesses two copies each of the *nodP* and *nodQ* genes, one is located on pSymA and the other on pSymB, and both copies have to be inactivated to impair nodule formation on alfalfa (Schwedock and Long 1992).

The molecular signal synthesised by *S. meliloti*, is a β -1,4 linked tri-, tetra- or pentamer of *N*-acetyl-glucosamine that is mono-*N*-acylated on the terminal non-reducing subunit and 6-*O*-sulfated on the reducing sugar (Lerouge et al. 1990; Schultze et al. 1992).

3.4 Nitrogen Fixation

The interaction between *S. meliloti* and its host legumes results in the formation of nodules on plant roots in which bacteria differentiate into bacteroids that fix atmospheric nitrogen, providing ammonia to the host plant. A number of genes that are essential for this process have been identified and mapped on the pSymA replicon (Fig. 3a, b).

In *S. meliloti*, expression of nitrogen-fixation genes is under the control of a complex regulatory circuit, which responds to the low oxygen concentration prevailing inside nodules (David et al. 1988). Oxygen limitation inside host nodules is required to synthesize and maintain the nitrogenase enzyme activity. *nifH*, *nifD*, and *nifK* code for subunits of the nitrogenase complex (Ruvkun et al. 1982; Corbin et al. 1983), *nifA* encodes a transcriptional activator (Szeto et al. 1984; Weber et al., 1985), and *nifB* encodes a protein involved in the biosynthesis of the MoFe cofactor (Buikema et al. 1987). Expression of genes required for nitrogen fixation are activated by two regulators, NifA and FixK, respectively, which are under the control of a two-component regulatory system, FixL/FixJ (Fig. 4) (Szeto et al. 1984; Aguilar et al. 1985; Buikema et al. 1987; Cosseau et al. 2002; Bobik et al. 2006). FixL, a sensor histidine kinase, autophosphorylates from ATP under microoxic conditions and transfers its phosphate to the FixJ transcriptional regulator protein (Gilles-Gonzalez et al. 1991; Reytrat et al. 1993). Phosphorylated FixJ activates transcription of the *nifA* and *fixK* genes. NifA activates 14 genes involved in nitrogen fixation (*fdxN*, *fdxB*, *fixABCX*, *nifB*, *fixU*, *nifHDKEX*, and *nifN*), one gene

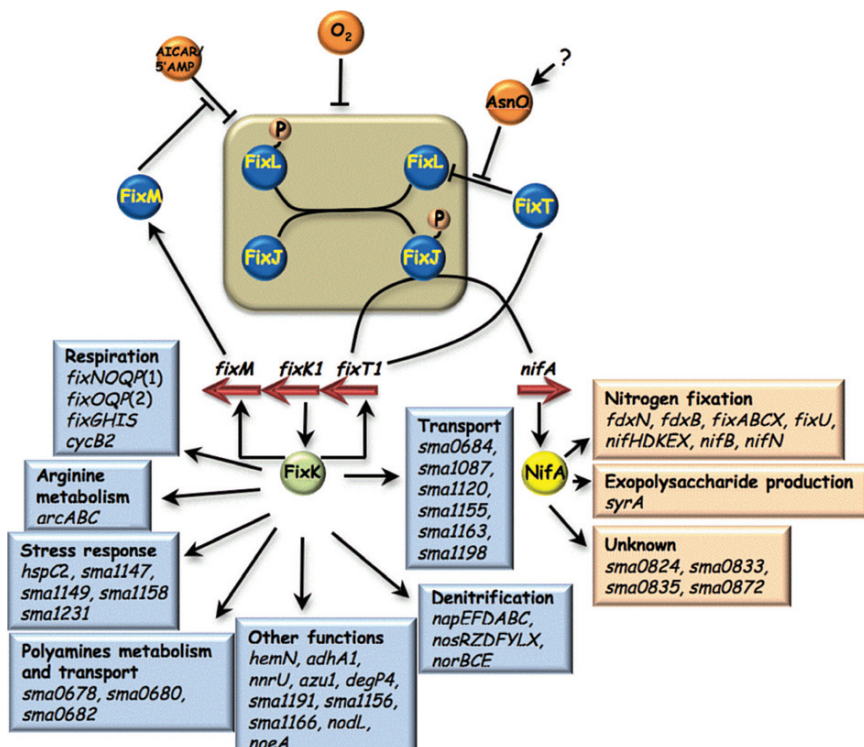


Fig. 4 Model for the FixLJ regulatory circuit in *S. meliloti*. Lines ending in bars represent repression. Lines ending in arrowheads indicate activation. The dark orange symbols are molecules. The dark blue, yellow, and green circles represent proteins. Adapted from Bobik et al. 2006 and Cosseau et al. 2002

involved in EPS production (*syrA*) and 4 genes with unknown function (Fig. 4) (Bobik et al. 2006). FixK, a member of the Crp/Fnr family, activates expression of 37 genes of unknown function and 60 others with attributed function (Bobik et al. 2006). Among the latter are the genes *fixOQP2*, *fixNOQP1*, *fixGHIS*, and *cycB2*, which are involved in the synthesis of a respiratory oxidase complex (Batut et al. 1989; Bobik et al. 2006). FixK also activates expression of regulatory genes such as *fixT1*, *fixT2*, *fixK*-like, and *fixM*. FixT inhibits expression of FixLJ-dependent genes by inhibiting FixL autophosphorylation (Foussard et al. 1997; Garnerone et al. 1999). The inhibiting effect of FixT was suppressed in an *asnO* mutant of *S. meliloti*. *asnO* encodes a protein homologous to glutamine-dependent asparagine synthetase (Berges et al. 2001). The role of *asnO* might be to couple expression of nitrogen fixation genes to the nitrogen needs of the cells (Berges et al. 2001). FixM, a flavoprotein oxidoreductase, modulates the negative effect of 5-amino-imidazole-4-carboxamide riboside (AICAR), a purine-related metabolite, and 5'AMP on the expression of respiratory and nitrogen fixation genes in *S. meliloti* (Fig. 4) (Cosseau et al. 2002). FixK controls the expression of many other genes like those involved in arginine metabolism (*arcABC*), and in denitrification (*napEFDABC*, *nosRZDFYLX*, *norBCE*) (Bobik et al. 2006).

S. meliloti is not only able to fix dinitrogen (N_2) in symbiosis with host legumes and to convert it into ammonia, but is also capable of dissimilatory reduction of nitrate or nitrite as the terminal electron acceptor coupled to proton translocation with the formation of dinitrogen. *S. meliloti*, like most denitrifiers, possesses the four reductases necessary to carry out the complete denitrification pathway. The genes (*napEFDABC*, *nosRZDFYLX*, *norECBQD*, *nirKV*, *nnrR*, *azuI*) involved in denitrification are located on the *nod/nif/fix* megaplastid, showing that pathways of nitrogen metabolism are genetically linked (Fig. 3a, b). The *nosRZDFYLX* gene cluster is required for N_2O reduction activity (Chan et al. 1997), but Tn5 insertions in the *nos* region did not affect the symbiotic N_2 fixation activity (Holloway et al. 1996). *S. meliloti* also carries on its pSymA replicon (Fig. 3a and b), the gene clusters *nap* (*napEFDABC*), *nor* (*norECBQD*), and *nir* (*nirKV*, *nnrR*, *azuI*) (Galibert et al. 2001; Chan and McCormick 2004), which are probably involved in nitrate (NO_3^-), nitrite (NO_2^-), and nitrous oxide (NO) reduction, respectively (Chan and McCormick 2004). Most of these genes are under the positive control of FixJL via FixK under microoxic conditions (Bobik et al. 2006). It might be that denitrification contributes to symbiosis only in nitrate-containing or oxygen-depleted soils but further experiments are needed to determine whether there is a close connection between N_2 production by denitrification and symbiotic N_2 fixation.

4 Conclusions on the Significance of *S. meliloti* Megaplastids

Many of the genes identified on pSymA and pSymB in *S. meliloti* are involved in the formation and functioning of nitrogen-fixing root nodules. Genes located on pSymA are necessary for nodulation and nitrogen fixation (Banfalvi et al. 1981;

Rosenberg et al. 1981), while those located on pSymB are involved in exopolysaccharide synthesis, symbiotic nitrogen fixation, and uptake and assimilation of various nutrients (Finan et al. 1986; Finan 1988; Watson et al. 1988; Glazebrook and Walker 1989; Watson 1990; Charles and Finan 1991; Finan et al. 2001). Functional analyses of the *S. meliloti* genome lead to a better understanding of the influence of the genes of both *S. meliloti* megaplasmids on the organism's metabolic and symbiotic abilities as well its successful occupation of natural niches, soil survival, plant colonization, and atmospheric dinitrogen fixation. pSymA could be cured and thus it is considered to be a plasmid and not chromosome while pSymB may be a true chromosome, since at least 200 kb from different regions of this replicon are necessary for cell viability. These two replicons are inherently stable and their maintenance within the cells may be due not only to their contribution to the cell viability but also to enhancement of the competitiveness of the cells in their natural environment.

References

- Aguilar OM, Kapp D, Pühler A (1985) Characterization of a *Rhizobium meliloti* fixation gene (*fixF*) located near the common nodulation region. *J Bacteriol* 164:245–254
- Ashelford KE, Day MJ, Fry JC (2003) Elevated abundance of bacteriophage infecting bacteria in soil. *Appl Environ Microbiol* 69:285–289.
- Baev N, Endre G, Petrovics G, Banfalvi Z, Kondorosi A (1991) Six nodulation genes of nod box locus 4 in *Rhizobium meliloti* are involved in nodulation signal production: *nodM* codes for d-glucosamine synthetase. *Mol Gen Genet* 228:113–124
- Baev N, Schultze M, Barlier I, Ha DC, Virelizier H, Kondorosi E, Kondorosi A. (1992) *Rhizobium nodM* and *nodN* genes are common *nod* genes: *nodM* encodes functions for efficiency of *nod* signal production and bacteroid maturation. *J Bacteriol* 174:7555–7565
- Banfalvi Z, Sakanyan V, Kocz C, Kiss A, Dusha I, Kondorosi A (1981) Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *R. meliloti*. *Mol Gen Genet* 184:318–325
- Bardin S, Dan S, Osteras M, Finan TM (1996) A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. *J Bacteriol* 178:4540–4547
- Barnett MJ, Hung DY, Reisenauer A, Shapiro L, Long SR (2001a) A homolog of the CtrA cell cycle regulator is present and essential in *Sinorhizobium meliloti*. *J Bacteriol* 183:3204–3210
- Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, Gurjal M, Hong A, Huizar L, Hyman RW, Kahn D, Kahn ML, Kalman S, Keating DH, Palm C, Peck MC, Surzycki R, Wells DH, Yeh KC, Davis RW, Federspiel NA, Long SR (2001b) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc Natl Acad Sci USA* 98:9883–9888
- Barnett MJ, Toman CJ, Fisher RF, Long SR (2004) A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proc Natl Acad Sci USA* 101:16636–16641
- Bartosik D, Baj J, Wlodarczyk M (1998) Molecular and functional analysis of pTAV320, a *repABC*-type replicon of the *Paracoccus versutus* composite plasmid pTAV1. *Microbiology* 144:3149–3157
- Batut J, Daveran-Mingot ML, David M, Jacobs J, Garnerone AM, Kahn D (1989) *fixK*, a gene homologous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. *EMBO J* 8:1279–1286

- Becker A, Kleickmann A, Keller M, Arnold W, Pühler A (1993a) Identification and analysis of the *Rhizobium meliloti* *exoAMONP* genes involved in exopolysaccharide biosynthesis and mapping of promoters located on the *exoHKLAMONP* fragment. *Mol Gen Genet* 241:367–379
- Becker A, Kleickmann A, Kuster H, Keller M, Arnold W, Pühler A (1993b) Analysis of the *Rhizobium meliloti* genes *exoU*, *exoV*, *exoW*, *exoT*, and *exoI* involved in exopolysaccharide biosynthesis and nodule invasion: *exoU* and *exoW* probably encode glucosyltransferases. *Mol Plant Microbe Interact* 6:735–744
- Becker A, Rüberg S, Küster H, Roxlau AA, Keller M, Ivashina T, Cheng HP, Walker GC, Pühler A. (1997) The 32-kilobase *exp* gene cluster of *Rhizobium meliloti* directing the biosynthesis of galactoglucan: genetic organization and properties of the encoded gene products. *J Bacteriol* 179:1375–1384
- Berges H, Checroun C, Guiral S, Garnerone AM, Boistard P, Batut J (2001) A glutamine-amidotransferase-like protein modulates FixT anti-kinase activity in *Sinorhizobium meliloti*. *BMC Microbiol* 1:6
- Bibb MJ, Biro S, Motamedi H, Collins JF, Hutchinson CR (1989) Analysis of the nucleotide sequence of the *Streptomyces glaucescens* *tml* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J* 8:2727–2736
- Bignell C, Thomas CM (2001) The bacterial ParA-ParB partitioning proteins. *J Biotechnol* 91:1–34
- Bobik C, Meilhoc E, Batut J (2006) FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J Bacteriol* 188:4890–4902
- Boivin C, Barran LR, Malpica CA, Rosenberg C (1991) Genetic analysis of a region of the *Rhizobium meliloti* pSym plasmid specifying catabolism of trigonelline, a secondary metabolite present in legumes. *J Bacteriol* 173:2809–2817
- Boucher C, Genin S, Arlat M (2001) Current concepts on the pathogenicity of phytopathogenic bacteria. *C R Acad Sci III* 324:915–922.
- Buendia AM, Enenkel B, Koplin R, Niehaus K, Arnold W, Pühler A (1991) The *Rhizobium meliloti* *exoZl* *exoB* fragment of megaplastid 2: ExoB functions as a UDP-glucose 4-epimerase and ExoZ shows homology to NodX of *Rhizobium leguminosarum* biovar viciae strain TOM. *Mol Microbiol* 5:1519–1530
- Buikema WJ, Klingensmith JA, Gibbons SL, Ausubel FM (1987) Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae* *nifB* genes. *J Bacteriol* 169:1120–1126
- Canter Cremers HC, Batley M, Redmond JW, Eydemis L, Breedveld MW, Zevehuizen LP, Pees E, Wijffelman CA, Lugtenberg BJ. (1990) *Rhizobium leguminosarum* *exoB* mutants are deficient in the synthesis of UDP-glucose 4 ϵ -epimerase. *J Biol Chem* 265:21122–21127
- Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, Boistard P, Becker A, Boutry M, Cadieu E, Dreano S, Gloux S, Godrie T, Goffeau A, Kahn D, Kiss E, Lelaure V, Masuy D, Pohl T, Portetelle D, Pühler A, Purnelle B, Ramsperger U, Renard C, Thebault P, Vandenbol M, Weidner S, Galibert F (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc Natl Acad Sci USA* 98:9877–9882
- Chain PS, Hernandez-Lucas I, Golding B, Finan TM (2000) *oriT*-directed cloning of defined large regions from bacterial genomes: identification of the *Sinorhizobium meliloti* pExo megaplastid replicator region. *J Bacteriol* 182:5486–5494
- Chan YK, McCormick WA, Watson RJ (1997) A new *nos* gene downstream from *nosDFY* is essential for dissimilatory reduction of nitrous oxide by *Rhizobium* (*Sinorhizobium*) *meliloti*. *Microbiology* 143:2817–2824
- Chan YK, McCormick WA (2004) Experimental evidence for plasmid-borne *nor-nir* genes in *Sinorhizobium meliloti* JJ1c10. *Can J Microbiol* 50:657–667
- Charles TC, Finan TM (1991) Analysis of a 1600-kilobase *Rhizobium meliloti* megaplastid using defined deletions generated *in vivo*. *Genetics* 127:5–20.
- Charles TC, Cai GQ, Aneja P (1997) Megaplastid and chromosomal loci for the PHB degradation pathway in *Rhizobium* (*Sinorhizobium*) *meliloti*. *Genetics* 146:1211–1220

- Chen H, Higgins J, Oresnik IJ, Hynes MF, Natera S, Djordjevic MA, Weinman JJ, Rolfe BG (2000) Proteome analysis demonstrates complex replicon and luteolin interactions in pSymA-derived derivatives of *Sinorhizobium meliloti* strain 2011. *Electrophoresis* 21:3833–3842.
- Cheng J, Sibley CD, Zaheer R, Finan TM (2007) A *Sinorhizobium meliloti* *minE* mutant has an altered morphology and exhibits defects in legume symbiosis. *Microbiology* 153:375–387
- Corbin D, Barran L, Ditta G (1983) Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. *Proc Natl Acad Sci USA* 80:3005–3009
- Cosseau C, Garnerone AM, Batut J (2002) The FixM flavoprotein modulates inhibition by AICAR or 5 ϵ AMP of respiratory and nitrogen fixation gene expression in *Sinorhizobium meliloti*. *Mol Plant Microbe Interact* 15:598–607
- David M, Daveran ML, Batut J, Dedieu A, Domergue O, Ghai J, Hertig C, Boistard P, Kahn D. (1988) Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. *Cell* 54:671–683
- Debelle F, Rosenberg C, Vasse J, Mailliet F, Martinez E, Dénarié J, Truchet G. (1986) Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti*. *J Bacteriol* 168:1075–1086
- Debelle F, Sharma SB (1986) Nucleotide sequence of *Rhizobium meliloti* RCR2011 genes involved in host specificity of nodulation. *Nucleic Acids Res* 14:7453–7472
- Dehoux P, Cossart P (1995) Homologies between salmolyisin and some bacterial regulatory proteins. *Mol Microbiol* 15:591
- Demont N, Debelle F, Aurelle H, Dénarié J, Prome JC (1993) Role of the *Rhizobium meliloti* *nodF* and *nodE* genes in the biosynthesis of lipo-oligosaccharidic nodulation factors. *J Biol Chem* 268:20134–20142
- Domínguez-Ferreras A, Pérez-Arnedo R, Becker A, Olivares J, Soto MJ, Sanjuán J (2006) Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *J Bacteriol* 188:7617–7625
- Downie JA, Young JP (2001) Genome sequencing. *The ABC of symbiosis*. *Nature* 412:597–598
- Dusha I, Kovalenko S, Banfalvi Z, Kondorosi A. (1987) *Rhizobium meliloti* insertion element *ISRm2* and its use for identification of the *fixX* gene. *J Bacteriol* 169:1403–1409.
- Egelhoff TT, Fisher RF, Jacobs TW, Mulligan JT, Long SR (1985) Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. *DNA* 4:241–248
- Engelke T, Jording D, Kapp D, Pühler A (1989) Identification and sequence analysis of the *Rhizobium meliloti* *dctA* gene encoding the C4-dicarboxylate carrier. *J Bacteriol* 171:5551–5560
- Fernández-López M, Muñoz-Adelantado E, Gillis M, Willems A, Toro N (2005) Dispersal and Evolution of the *Sinorhizobium meliloti* Group II RmInt1 Intron in Bacteria that Interact with Plants. *Mol Biol Evol* 22:1518–1528.
- Finan TM, Hirsch AM, Leigh JA, Johansen E, Kuldau GA, Deegan S, Walker GC, Signer ER. (1985) Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. *Cell* 40:869–877
- Finan TM, Kunkel B, De Vos GF, Signer ER (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. *J Bacteriol* 167:66–72
- Finan TM (1988) Genetic and physical analyses of group E *exo-* mutants of *Rhizobium meliloti*. *J Bacteriol* 170:474–477
- Finan TM, Weidner S, Wong K, Buhmester J, Chain P, Vorholter FJ, Hernandez-Lucas I, Becker A, Cowie A, Gouzy J, Golding B, Pühler A (2001) The complete sequence of the 1,683-kb pSymB megaplasmid from the N₂-fixing endosymbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci USA* 98:9889–9894
- Fisher RF, Egelhoff TT, Mulligan JT, Long SR (1988) Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes Dev* 2:282–293
- Fisher RF, Long SR (1989) DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible *nod* gene promoters. *J Bacteriol* 171:5492–5502

- Foussard M, Garnerone AM, Ni F, Soupene E, Boistard P, Batut J (1997) Negative autoregulation of the *Rhizobium meliloti* *fixK* gene is indirect and requires a newly identified regulator, FixT. *Mol Microbiol* 25:27–37
- Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* 387:394–401
- Gage DJ, Long SR (1998) alpha-Galactoside uptake in *Rhizobium meliloti*: isolation and characterization of *aggA*, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. *J Bacteriol* 180:5739–5748
- Galibert F, Finan TM, Long SR, Pühler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothé G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668–672
- Garnerone AM, Cabanes D, Foussard M, Boistard P, Batut J (1999) Inhibition of the FixL sensor kinase by the FixT protein in *Sinorhizobium meliloti*. *J Biol Chem* 274:32500–32506
- Geremia RA, Mergaert P, Geelen D, Van Montagu M, Holsters M (1994) The NodC protein of *Azorhizobium caulinodans* is an N-acetylglucosaminyltransferase. *Proc Natl Acad Sci USA* 91:2669–2673
- Gilles-Gonzalez MA, Ditta GS, Helinski DR (1991) A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*. *Nature* 350:170–172
- Glazebrook J, Walker GC (1989) A novel exopolysaccharide can function in place of the calcofluor-binding exopolysaccharide in nodulation of alfalfa by *Rhizobium meliloti*. *Cell* 56:661–672
- Glucksmann MA, Reuber TL, Walker GC (1993a) Family of glycosyl transferases needed for the synthesis of succinoglycan by *Rhizobium meliloti*. *J Bacteriol* 175:7033–7044
- Glucksmann MA, Reuber TL, Walker GC (1993b) Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. *J Bacteriol* 175:7045–7055
- Goldmann A, Boivin C, Fleury V, Message B, Lecoeur L, Maille M, Tepfer D. (1991) Betaine use by rhizosphere bacteria: genes essential for trigonelline, stachydrine, and carnitine catabolism in *Rhizobium meliloti* are located on pSym in the symbiotic region. *Mol Plant Microbe Interact* 4:571–578
- Gonzalez V, Santamaria RI, Bustos P, Hernandez-Gonzalez I, Medrano-Soto A, Moreno-Hagelsieb G, Janga SC, Ramirez MA, Jimenez-Jacinto V, Collado-Vides J, Dávila G (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons *Proc Natl Acad Sci USA* 103:3834–3839
- Gottfert M, Horvath B, Kondorosi E, Putnoky P, Rodriguez-Quinones F, Kondorosi A (1986) At least two *nodD* genes are necessary for efficient nodulation of alfalfa by *Rhizobium meliloti*. *J Mol Biol* 191:411–420
- Guo X, Flores M, Mavingui P, Fuentes SI, Hernández G, Dávila G, Palacios R (2003) Natural Genomic Design in *Sinorhizobium meliloti*: Novel Genomic Architectures. *Genome Res* 13:1810–1817
- Guo H, Sun S, Finan TM, Xu J (2005) Novel DNA sequences from natural strains of the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. *Appl Environ Microbiol* 71:7130–7138
- Hartwig UA, Maxwell CA, Joseph CM, Phillips DA (1990) Effects of alfalfa nod gene-inducing flavonoids on *nodABC* transcription in *Rhizobium meliloti* strains containing different *nodD* genes. *J Bacteriol* 172:2769–2773
- Harwood CS, Parales RE (1996) The b-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol* 50:553–590
- Henikoff S, Haughn GW, Calvo JM, Wallace JC (1988) A large family of bacterial activator proteins. *Proc Natl Acad Sci USA* 85:6602–6606

- Her GR, Glazebrook J, Walker GC, Reinhold VN (1990) Structural studies of a novel exopolysaccharide produced by a mutant of *Rhizobium meliloti* strain Rm1021. *Carbohydr Res* 198:305–312
- Herrera-Cervera JA, Sanjuán-Pinilla JM, Olivares J, Sanjuán J (1998) Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. *J Bacteriol* 180:4583–4590
- Holloway P, McCormick W, Watson RJ, Chan YK (1996) Identification and analysis of the dissimilatory nitrous oxide reduction genes, *nosRZDFY*, of *Rhizobium meliloti*. *J Bacteriol* 178:1505–1514
- Izquierdo J, Venkova-Canova T, Ramírez-Romero MA, Téllez-Sosa J, Hernández-Lucas I, Sanjuan J, Cevallos MA (2005). An antisense RNA plays a central role in the replication control of a *repC* plasmid. *Plasmid* 54:259–277
- Jacobs TW, Egelhoff TT, Long SR (1985) Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. *J Bacteriol* 162:469–476
- Jebbar M, Sohn-Bosser L, Bremer E, Bernard T, Blanco C (2005) Ectoine-induced proteins in *Sinorhizobium meliloti* include an Ectoine ABC-type transporter involved in osmoprotection and ectoine catabolism. *J Bacteriol* 187:1293–1304
- Jiang XM, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR (1991) Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar Typhimurium (strain LT2). *Mol Microbiol* 5:695–713
- John M, Rohrig H, Schmidt J, Wieneke U, Schell J (1993) *Rhizobium* NodB protein involved in nodulation signal synthesis is a chitooligosaccharide deacetylase. *Proc Natl Acad Sci USA* 90:625–629
- Kahng LS, Shapiro L (2003) Polar localization of replicon origins in the multipartite genomes of *Agrobacterium tumefaciens* and *Sinorhizobium meliloti*. *J Bacteriol* 185:3384–3391
- Kaminski PA, Batut J, Boistard P (1998). A survey of symbiotic nitrogen fixation by rhizobia. In: The Rhizobiaceae. Spaink HP, Kondoros A, Hooykaas PJJ (eds) Dordrecht, The Netherlands: Kluwer, pp 431–460.
- Karlin S, Barnett MJ, Campbell AM, Fisher RF, Mrazek J (2003) Predicting gene expression levels from codon biases in a-proteobacterial genomes *Proc Natl Acad Sci USA* 100:7313–7318
- Keller M, Roxlau A, Weng WM, Schmidt M, Quandt J, Niehaus K, Jording D, Arnold W, Pühler A. (1995) Molecular analysis of the *Rhizobium meliloti mucR* gene regulating the biosynthesis of the exopolysaccharides succinoglycan and galactoglucan. *Mol Plant Microbe Interact* 8:267–277
- Kinkle BK, Schmidt EL (1991) Transfer of the Pea Symbiotic Plasmid pJB5JI in Nonsterile Soil. *Appl Environ Microbiol* 57:3264–3269
- Koplin R, Wang G, Hotte B, Priefer UB, Pühler A (1993) A 3.9-kb DNA region of *Xanthomonas campestris* pv. *campestris* that is necessary for lipopolysaccharide production encodes a set of enzymes involved in the synthesis of dTDP-rhamnose. *J Bacteriol* 175:7786–7792
- Laberge S, Middleton AT, Wheatcroft R. (1995) Characterization, nucleotide sequence, and conserved genomic location of insertion sequence ISRm5 in *Rhizobium meliloti*. *J. Bacteriol* 177:3133–3142
- Leigh JA, Lee CC (1988) Characterization of polysaccharides of *Rhizobium meliloti* exo mutants that form ineffective nodules. *J Bacteriol* 170:3327–3332
- Leigh JA, Reed JW, Hanks JF, Hirsch AM, Walker GC (1987) *Rhizobium meliloti* mutants that fail to succinylate their calcofluor-binding exopolysaccharide are defective in nodule invasion. *Cell* 51:579–587
- Leigh JA, Signer ER, Walker GC (1985) Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc Natl Acad Sci USA* 82:6231–6235
- Lerouge P, Roche P, Faucher C, Maillat F, Truchet G, Promé JC, Dénarié J (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* 344:781–784

- Long S, Reed JW, Himawan J, Walker GC (1988) Genetic analysis of a cluster of genes required for synthesis of the calcofluor-binding exopolysaccharide of *Rhizobium meliloti*. *J Bacteriol* 170:4239–4248
- Lynch D, O'Brien J, Welch T, Clarke P, Cuív PO, Crosa JH, O'Connell M (2001) Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *J Bacteriol* 183:2576–2585
- MacLean AM, MacPherson G, Aneja P, Finan TM (2006) Characterization of the b-ketoadipate pathway in *Sinorhizobium meliloti*. *Appl Environ Microbiol* 72:5403–5413
- MacLellan SR, Smallbone LA, Sibley CD, Finan TM (2005) The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of α -proteobacterial plasmids. *Mol Microbiol* 55:611–623
- MacLellan SR, Zaheer R, Sartor AL, MacLean AM, Finan TM (2006) Identification of a megaplasmid centromere reveals genetic structural diversity within the *repABC* family of basic replicons. *Mol Microbiol* 59:1559–1575.
- Mahillon M, Chandler M (1998) Insertion Sequence. *Microbiol Mol Biol Rev* 62:725–774
- Margolin W, Long SR (1993) Isolation and characterization of a DNA replication origin from the 1,700-kilobase-pair symbiotic megaplasmid pSym-b of *Rhizobium meliloti*. *J Bacteriol* 175:6553–6561
- Martínez-Abarca F, Barrientos-Durán A, Fernández-López M, Toro N (2004) The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res* 20:2880–2888
- Meade HM, Long SR, Ruvkun GB, Brown SE, Ausubel FM (1982) Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J Bacteriol* 149:114–122
- Moreira LM, Becker JD, Pühler A, Becker A (2000) The *Sinorhizobium meliloti* ExpE1 protein secreted by a type I secretion system involving ExpD1 and ExpD2 is required for biosynthesis or secretion of the exopolysaccharide galactoglucan. *Microbiology* 146:2237–2248
- Muller P, Keller M, Weng WM, Quandt J, Arnold W, Pühler A (1993) Genetic analysis of the *Rhizobium meliloti* *exoYFQ* operon: ExoY is homologous to sugar transferases and ExoQ represents a transmembrane protein. *Mol Plant Microbe Interact* 6:55–65
- Mulligan JT, Long SR (1985) Induction of *Rhizobium meliloti* *nodC* expression by plant exudate requires *nodD*. *Proc Natl Acad Sci USA* 82:6609–6613
- Mulligan JT, Long SR (1989) A family of activator genes regulates expression of *Rhizobium meliloti* nodulation genes. *Genetics* 122:7–18
- Natera SH, Guerreiro N, Djordjevic MA (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. *Mol Plant Microbe Interact* 13:995–1009
- Nisa-Martínez R, Jiménez-Zurdo JI, Martínez-Abarca F, Muñoz-Adelantado E, Toro N (2007) Dispersion of the RmInt1 group II intron in the *Sinorhizobium meliloti* genome upon acquisition by conjugative transfer. *Nucleic Acids Res* 35:214–222
- Obendorf RL (1997). Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Sci Res* 7:63–74
- Ogawa J, Long SR (1995) The *Rhizobium meliloti* *groELc* locus is required for regulation of early *nod* genes by the transcription activator NodD. *Genes Dev* 9:714–729
- Oresnik IJ, Liu LL, Yost CK, Hynes MF (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability *J Bacteriol* 182:3582–3586
- Patriarca EJ, Tate R, Iaccarino M (2002) Key role of bacterial NH₄⁺ metabolism in Rhizobium-plant symbiosis. *Microbiol Mol Biol Rev* 66:203–222
- Peixoto L, Zavala A, Romero H, Musto H (2003) The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium meliloti*. *Gene* 320:109–116
- Perrine FM, Hocart CH, Hynes MF, Rolfe BG (2005) Plasmid-associated genes in the model micro-symbiont *Sinorhizobium meliloti* 1021 affect the growth and development of young rice seedlings. *Environ Microbiol* 7:1826–1838

- Peters NK, Frost JW, Long SR (1986) A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* 233:977–980
- Poysti NJ, Loewen ED, Wang Z, Oresnik IJ (2007) *Sinorhizobium meliloti* pSymB carries genes necessary for arabinose transport and catabolism. *Microbiology* 153:727–736
- Pretorius-Güth IM, Pühler A, Simon R. (1990) Conjugal Transfer of Megaplasmid 2 between *Rhizobium meliloti* Strains in Alfalfa Nodules. *Proc Natl Acad Sci USA* 94:5483–5488
- Ramírez-Romero MA, Soberón N, Pérez-Oseguera A, Téllez-Sosa J, Cevallos MA (2000) Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. *J Bacteriol* 182:3117–3124
- Reed JW, Capage M, Walker GC (1991) *Rhizobium meliloti* *exoG* and *exoJ* mutations affect the *exoX-exoY* system for modulation of exopolysaccharide production. *J Bacteriol* 173:3776–3788
- Reinhold BB, Chan SY, Reuber TL, Marra A, Walker GC, Reinhold VN (1994) Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. *J Bacteriol* 176:1997–2002
- Reuber TL, Walker GC (1993a) The acetyl substituent of succinoglycan is not necessary for alfalfa nodule invasion by *Rhizobium meliloti* Rm1021. *J Bacteriol* 175:3653–3655
- Reuber TL, Walker GC (1993b) Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. *Cell* 74:269–280
- Reyrat JM, David M, Blonski C, Boistard P, Batut J (1993) Oxygen-regulated *in vitro* transcription of *Rhizobium meliloti* *nifA* and *fixK* genes. *J Bacteriol* 175:6867–6872
- Rohrig H et-al. (1994) Biosynthesis of lipooligosaccharide nodulation factors: *Rhizobium* NodA protein is involved in *N*-acylation of the chitooligosaccharide backbone. *Proc Natl Acad Sci USA* 91:3122–3126
- Ronson CW, Astwood PM, Downie JA (1984) Molecular cloning and genetic organization of C4-dicarboxylate transport genes from *Rhizobium leguminosarum*. *J Bacteriol* 160:903–909
- Rosenberg C, Boistard P, Dénarié J, Casse-Delbart F (1981) Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol Gen Genet* 184:326–333
- Rosenberg C, Casse-Delbart F, Dusha I, David M, Boucher C (1982) Megaplasmids in the plant-associated bacteria *Rhizobium meliloti* and *Pseudomonas solanacearum*. *J Bacteriol* 150:402–406
- Rostas K, Kondorosi E, Horvath B, Simoncsits A, Kondorosi A (1986) Conservation of extended promoter regions of nodulation genes in *Rhizobium*. *Proc Natl Acad Sci USA* 83:1757–1761
- Ruvkun GB, Ausubel FM (1981) A general method for site-directed mutagenesis in prokaryotes. *Nature* 289:85–88
- Ruvkun GB, Long SR, Meade HM, van der Bos RC, Ausubel FM (1982a) *ISRm1*: a *Rhizobium meliloti* insertion sequence that transposes preferentially into nitrogen fixation genes. *J Mol Appl Genet* 1:405–418
- Ruvkun GB, Sundaresan V, Ausubel FM (1982b) Directed transposon Tn5 mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen fixation genes. *Cell* 29:551–559
- Schlaman HL, Phillips DA, Kondorosi E (1998). In *The Rhizobiaceae*. Spaink HP, Kondorosi A, Hooykaas PJJ (eds). Dordrecht, The Netherlands: Kluwer, pp 361–386
- Schmidt J, John M, Kondorosi E, Kondorosi A, Wieneke U, Schröder G, Schröder J, Schell J. (1984) Mapping of the protein-coding regions of *Rhizobium meliloti* common nodulation genes. *EMBO J* 3:1705–1711
- Schultze M, Quiclet-Sire B, Kondorosi E, Virelizer H, Glushka JN, Endre G, Géro SD, Kondorosi A. (1992) *Rhizobium meliloti* produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity. *Proc Natl Acad Sci USA* 89:192–196
- Schwedock JS, Long SR (1992) *Rhizobium meliloti* genes involved in sulfate activation: the two copies of *nodPQ* and a new locus, *saa*. *Genetics* 132:899–909
- Selbitschka W, Zekri S, Schröder G, Pühler A, Toro N (1999) The *Sinorhizobium meliloti* insertion sequence (IS) elements *ISRm102F34-ISRm7* and *ISRm220.13.5* belong to a new family of insertion sequence elements. *FEMS Microbiol Lett* 172:1–7

- Sharma SB, Signer ER (1990) Temporal and spatial regulation of the symbiotic genes of *Rhizobium meliloti* in planta revealed by transposon Tn5-*gusA*. *Genes Dev* 4:344–356
- Shearman CA, Rossen L, Johnston AW, Downie JA (1986) The *Rhizobium leguminosarum* nodulation gene *nodF* encodes a polypeptide similar to acyl-carrier protein and is regulated by *nodD* plus a factor in pea root exudate. *EMBO J* 5:647–652
- Sibley CD, MacLellan SR, Finan T (2006) The *Sinorhizobium meliloti* chromosomal origin of replication. *Microbiology* 152:443–455
- Silva C, Vinuesa P, Eguiarte LE, Souza V, Martínez-Romero E (2005) Evolutionary genetics and biogeographic structure of *Rhizobium gallicum* sensu lato, a widely distributed bacterial symbiont of diverse legumes. *Mol Ecol* 14:4033–4050
- Sobral BW, Honeycutt RJ, Atherly AG, McClelland M (1991) Electrophoretic separation of the three *Rhizobium meliloti* replicons. *J Bacteriol* 173:5173–5180
- Stiens M, Schneiker S, Keller M, Kuhn S, Pühler A, Schlüter A (2006) Sequence analysis of the 144-kilobase accessory plasmid pSmeSM11a, isolated from a dominant *Sinorhizobium meliloti* strain identified during a long-term field release experiment. *Appl Environ Microbiol* 72:3662–3672
- Stiens M, Schneiker S, Pühler A, Schlüter A (2007) Sequence analysis of the 181-kb accessory plasmid pSmeSM11b, isolated from a dominant *Sinorhizobium meliloti* strain identified during a long-term field release experiment. *FEMS Microbiol Lett* 271:297–309
- Stowers MD (1985) Carbon metabolism in *Rhizobium* species. *Annu Rev Microbiol* 39:89–108
- Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci USA* 12:8985–8989
- Sun S, Guo H, Xu J (2006) Multiple gene genealogical analyses reveal both common and distinct population genetic patterns among replicons in the nitrogen-fixing bacterium *Sinorhizobium meliloti*. *Microbiology* 152:3245–3259
- Sutton JM, Lea EJ, Downie JA (1994) The nodulation-signaling protein NodO from *Rhizobium leguminosarum* biovar viciae forms ion channels in membranes. *Proc Natl Acad Sci USA* 91:9990–9994
- Swanson JA, Mulligan JT, Long SR (1993) Regulation of *syrM* and *nodD3* in *Rhizobium meliloti*. *Genetics* 134:435–444
- Szeto WW, Zimmerman JL, Sundaresan V, Ausubel FM (1984) A *Rhizobium meliloti* symbiotic regulatory gene. *Cell* 36:1035–1043
- Talibart R, Jebbar M, Gouffi K, Pichereau V, Gouesbet G, Blanco C, Bernard T, Pocard J (1997) Transient accumulation of glycine betaine and dynamics of endogenous osmolytes in salt-stressed cultures of *Sinorhizobium meliloti*. *Appl Environ Microbiol* 63:4657–4663
- Toro N, Martínez-Abarca F, Fernández-Lopez M, Muñoz-Adelantado E (2003) Diversity of group II introns in the genome of *Sinorhizobium meliloti*. *Mol Genet Genomics* 268:628–636
- Truchet G, Rosenberg C, Vasse J, Julliot JS, Camut S, Dénarié J (1984) Transfer of *Rhizobium meliloti* pSym genes into *Agrobacterium tumefaciens*: host-specific nodulation by atypical infection. *J Bacteriol* 157:134–142
- Uttaro AD, Cangelosi GA, Geremia RA, Nester EW, Ugalde RA (1990) Biochemical characterization of avirulent *exoC* mutants of *Agrobacterium tumefaciens*. *J Bacteriol* 172:1640–1646
- Van Sluys MA, Monteiro-Vitorello CB, Camargo LE, Menck CF, Da Silva AC, Ferro JA, Oliveira MC, Setubal JC, Kitajima JP, Simpson AJ (2002) Comparative genomic analysis of plant-associated bacteria. *Annu Rev Phytopathol* 40:169–189
- Vinuesa P, Silva C, Lorite MJ, Izaguirre-Mayoral ML, Bedmar EJ, Martínez-Romero E (2005) Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from *rrs*, *atpD*, *recA*, and *nifH* sequences, and their use in the classification of Sesbania microsymbionts from Venezuelan wetlands. *Syst Appl Microbiol* 28:702–716
- Watson RJ (1990) Analysis of the C4-dicarboxylate transport genes of *Rhizobium meliloti*: nucleotide sequence and deduced products of *dctA*, *dctB*, and *dctD*. *Mol Plant Microbe Interact* 3:174–181

- Watson RJ, Chan YK, Wheatcroft R, Yang AF, Han SH (1988) *Rhizobium meliloti* genes required for C4-dicarboxylate transport and symbiotic nitrogen fixation are located on a megaplasmid. *J Bacteriol* 170:927–934
- Wernegreen JJ and Riley MA (1999) Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages *J Bacteriol* 150:402–406
- Wernegreen JJ, Harding EE, Riley MA (1997) *Rhizobium* gone native: unexpected plasmid stability of indigenous *Rhizobium leguminosarum*. *Proc Natl Acad Sci USA* 94:5483–5488
- Wheatcroft R, Laberge S (1991) Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence ISRm3: similarity between the putative transposase encoded by ISRm3 and those encoded by *Staphylococcus aureus* IS256 and *Thiobacillus ferrooxidans* IST2. *J Bacteriol* 173:2530–2538
- Wong CH, Pankhurst CE, Kondorosi A, Broughton WJ (1983) Morphology of root nodules and nodule-like structures formed by *Rhizobium* and *Agrobacterium* strains containing a *Rhizobium meliloti* megaplasmid. *J Cell Biol* 97:787–794
- Wong K, Golding GB (2003) A phylogenetic analysis of the pSymB replicon from the *Sinorhizobium meliloti* genome reveals a complex evolutionary history. *Can J Microbiol* (4):269–280
- Yarosh OK, Charles TC, Finan TM (1989) Analysis of C4-dicarboxylate transport genes in *Rhizobium meliloti*. *Mol Microbiol* 3:813–823
- Young JP, Crossman LC, Johnston AW, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson AR, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 7:R34
- Yurgel SN, Kahn ML (2004) Dicarboxylate transport by rhizobia. *FEMS Microbiol Rev* 28:489–501
- Zekri S, Soto M, Toro N (1998a) ISRm4–I and ISRm9, two novel insertion sequences from *Sinorhizobium meliloti*. *Gene* 207:93–96
- Zekri S, Toro N (1998b) A new insertion sequence from *Sinorhizobium meliloti* with homology to IS1357 from *Methylobacterium sp.* and IS1452 from *Acetobacter pasteurianus*. *FEMS Microbiol Lett* 158:83–87
- Zhan HJ, Levery SB, Lee CC, Leigh JA (1989) A second exopolysaccharide of *Rhizobium meliloti* strain SU47 that can function in root nodule invasion. *Proc Natl Acad Sci USA* 86:3055–3059
- Zhan HJ, Lee CC, Leigh JA (1991) Induction of the second exopolysaccharide (EPSb) in *Rhizobium meliloti* SU47 by low phosphate concentrations. *J Bacteriol* 173:7391–7394

The Megaplasmid pNGR234a of *Rhizobium* sp. Strain NGR234

Patrick Mavingui

Contents

1	Introduction.....	119
2	Biology of pNGR234a.....	120
3	pNGR234a Genomic Architecture	121
3.1	General Features	121
3.2	Repeated DNA Sequences and Genomic Rearrangement	123
4	Nodulation and Nitrogen Fixation Loci.....	124
4.1	Nodulation Genes.....	124
4.2	Nitrogen Fixation Gene Cluster.....	125
5	T3SS in Symbiotic Process.....	126
6	Global Transcriptional Analysis	127
7	Conclusion	127
	References.....	128

Abstract The symbiotic megaplasmid or pNGR234a of *Rhizobium* sp. NGR234 was the first rhizobial replicon to be sequenced. The circular replicon is 536,165 base pairs long and contains most loci involved in plant nodulation (*nod*) and nitrogen fixation (*nif*, *fix*). pNGR234a carries also genes encoding a type III secretion system and many repeated DNA sequences that participate in genomic rearrangement. This chapter focuses on the biology and genetics of pNGR234a, highlighting its contribution to the broad-host-range phenotype of *Rhizobium* sp. NGR234 strain.

1 Introduction

Rhizobium species NGR234 (also called *Sinorhizobium fredii* NGR234) is a gram-negative alpha-proteobacterium that was isolated in 1965 from nodules of the tropical legume *Lablab purpureus* in Papua New Guinea (Trinick 1980). It is able to nodulate

P. Mavingui

Université de Lyon, Université Lyon 1, CNRS, UMR 5557, Ecologie Microbienne,
and IFR 41, Villeurbanne, F-69622, France
e-mail: patrick.mavingui@univ-lyon1.fr

more than 112 genera in all three subfamilies of *Leguminosae* as well as the nonlegume *Parasponia andersonii* (Pueppke and Broughton 1999). This extensively broad host range of *Rhizobium* sp. NGR234 contrasts with narrow-host-range behaviour exhibited by the majority of rhizobia. For instance, *Sinorhizobium meliloti* nodulates legume plants belonging to only three genera, *Medicago*, *Melilotus* and *Trigonella* (Dénarié et al. 1992), while *Azorhizobium caulinodans* nodulation capacity is restricted to *Sesbania* species (Lewin et al. 1987). What gives *Rhizobium* sp. NGR234 this extremely broad host range is not fully understood and is the subject of ongoing intensive research.

Since the isolation of *R. sp.* NGR234 strain, several studies have investigated the composition of its genome with contrasting results, mainly due to the methodology used, notably Eckhardt gelsor pulse-field gel electrophoresis (PFGE). Overall, it is now well established that the NGR234 genome is partitioned in three replicons: the chromosome (>3,700 kb), the megaplasmid pNGR234b (>2,000 kb), and the symbiotic plasmid pNGR234a (536 kb) (Pankhurst et al. 1983; Morrison et al. 1983, 1984; Freiberg et al. 1997; Flores et al. 1998; Mavingui et al. 2002). Like many rhizobia, the plasmid pool of NGR234 represents a large part of the genome (40%). The pNGR234b that is not reviewed in this chapter carries some genes involved in the symbiotic process. These include *exo* and *exs* genes that are involved in the synthesis of exopolysaccharides (Staehelin et al. 2006) as well as *nodPQ* (Schwedock and Long 1990) genes that encode enzymes involved in the synthesis of PAPS (3'-phosphoadenosine 5'-phosphosulfate), a precursor of Nod-factor sulphation; these components are essential for nodule invasion (Battisti et al. 1992). Data on the gene content and the expression pattern of both pNGR234b and the chromosome of *R. sp.* NGR234 can be found in published studies (Viprey et al. 2000; Streit et al. 2004; Mavingui et al. 2005). Here we summarise general information on the symbiotic plasmid pNGR234a, especially focusing on genomic aspects that could help to understand the promiscuous behaviour of *Rhizobium* sp. NGR234 strain.

2 Biology of pNGR234a

Generally, most symbiotic genes of *Rhizobium* are carried by large plasmids called "symbiotic plasmids" (Banfalvi et al. 1981; Rosenberg et al. 1981), but *B. japonicum* and *M. Loti* harbour symbiotic *loci* in particular regions designated "symbiotic islands" located on their unique chromosomal replicon (Kündig et al. 1993; Sullivan and Ronson 1998). The symbiotic plasmid status was assigned to pNGR234a as its curing results in the inability of the resulting pSym-less strain to induce nodulation, whereas transfer of pNGR234a to compatible strains confers on bacterial transconjugants the ability to nodulate non-host plants or extend host range (Morrison et al. 1983, 1984; Broughton et al. 1984, 1986). Later on, sequencing of the 536,165 bp of pNGR234a was achieved (Freiberg et al. 1997) and its analysis confirmed the presence of most loci involved in plant nodulation (*nod*) and nitrogen fixation (*nif*, *fix*).

pNGR234a contains *repABC* genes that encode replicator proteins that are highly similar (up to 60%) to those found in plasmids of *Rhizobiaceae*, including *S. meliloti* pSymA and pSymB (Galibert et al. 2001; Barloy-Hubler and Jebbar 2008), *Rhizobium etli* pSym42d (Ramirez-Romero et al. 1997), *Rhizobium leguminosarum* bv. *leguminosarum* cryptic plasmid pRL8JI (Turner and Young 1995), *Agrobacterium rhizogenes* pRiA4b (Nishiguchi et al. 1987) and *Agrobacterium tumefaciens* plasmid pTiB6S (Tabata et al. 1989). RepA and RepB proteins are involved in plasmid segregation and copy number control, whereas RepC initiates the replication (Ramírez-Romero et al. 2000, 2001 Soberón et al. 2004; Chai and Winans 2005a). Freiberg et al. (1997) identified a locus within the intergenic region between the *repB* and *repC* genes as a probable *oriV* of pNGR234a. However, in *repABC* plasmids, recent data suggest that the *repB-repC* intergenic region contains a gene coding a 55-59 nucleotides small non translated RNA that constitutes a *trans* incompatibility factor (Venkova-Canova et al., 2004; MacLellan et al. 2005; Chai and Winans 2005b) while the *oriV* is placed in an A+T-rich region within the *repC* gene (for review see Cevallos et al. 2008). These studies identified only one replicator in the majority of rhizobial replicons, whilst more than one *oriV* were found in pSymB of *S. meliloti* (Margolin and Long 1993). Experiments are required to decipher the functionality of the pNGR234a replicator. In addition to replication loci, the pNGR234a sequence also contains conjugal transfer (*tra*) loci similar to those of *A. tumefaciens* Ti-plasmids (Freiberg et al. 1997). These *tra* loci include the 12-bp origin of transfer (*oriT*) and the *traI*, *traM* and *traR* genes that are quorum-sensing regulators of the acylated homoserine lactones (acyl-HCLs); a group of molecules used by a number of bacteria to modulate population density (Whitehead et al. 2001; Fuqua and Greenberg 2002). In *A. tumefaciens*, TraI is the HCL synthase, TraR is the regulatory activator and TraM is the repressor (Hwang et al. 1994, 1999). Similarly, pNGR234a TraI synthesizes an acyl-HSL (likely 3-oxo-C8-HSL), TraR activates expression of pNGR234a *tra* operons in response to 3-oxo-C8-HSL and TraM inhibits TraR function (He et al. 2003). These authors demonstrated that transcription of these *tra* functions affects NGR234 growth and transfer of pNGR234a to plasmid-less *A. tumefaciens* recipients.

3 pNGR234a Genomic Architecture

Sequence annotation of pNGR234a revealed a striking presence of highly repeated DNA sequences, suggesting that genomic rearrangements may occur in this replicon.

3.1 General Features

The symbiotic plasmid pNGR234a is a circular replicon of 536,165 base pairs (Fig. 1; Freiberg et al. 1997). It encodes for 416 open reading frames (ORFs) constituting 418 kb or 78% of the whole replicon. Most loci involved in plant

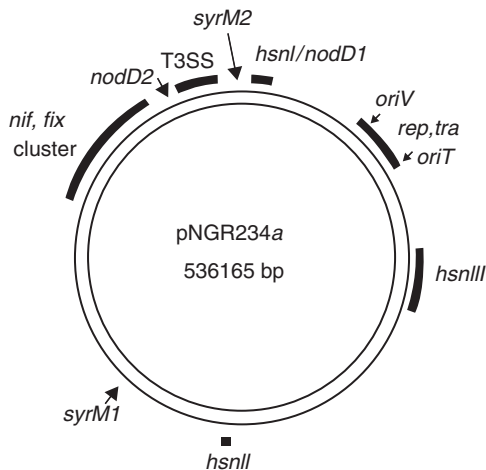


Fig. 1 Physical and genetic map of pNGR234a. Key genes and gene clusters (see text) are indicated by *arrows* or *black arcs* (Adapted from Freiberg et al. 1997)

nodulation and nitrogen fixation are located on this replicon, which is thus required for effective symbiosis. The G + C content of pNGR234a is 58.5%, which is significantly lower than the 62.2% of the entire genome (Broughton et al. 1972), suggesting that all or part of pNGR234a was probably acquired by lateral transfer from other microbes. Several genes with lower G + C (45–55%) are represented by loci encoding for nodulation factors and polysaccharide components. Sequence annotation did not reveal the presence of key genes required for bacterial growth, supporting the data of free-living survival of the pNGR234a-less derivative strain ANU265 (Morrison et al. 1984), which is in contrast to pNGR234b and pSymB of *S. meliloti* that harbour key determinants of cellular processes (Streit et al. 2004, Galibert et al. 2001).

Strikingly, a putative integrase/recombinase gene (*y4qk*) and many repeated DNA sequences are present on pNGR234a including complete operons, specific genes, regulatory elements, insertion (IS) and mosaic (MS) sequences. All together, repeated sequences accounted for a large portion (18%) of the whole replicon, which is more than that found in the *R. etli* pSym42d (10%) or the *S. meliloti* pSymA (<3%). Generally, IS/MS elements separate clusters of functionally related genes such as *nod-nif-fix* genes that have distinct G + C values, indicating horizontal gene transfer and movement of IS/MS elements have shaped the mosaic structure of rhizobial replicons. Accordingly, transposase expression and transposition activity were shown in some rhizobia (Ruvkun et al. 1982; Gay et al. 1985; Simon et al. 1991). Transcriptional analysis and IS entrapment experiments demonstrated that ISs located on pNGR234a were functional in both free-living conditions and in symbiosis (Perret et al. 1999; Hernandez-Lucas et al. 2006).

3.2 Repeated DNA Sequences and Genomic Rearrangement

Reiterated DNA sequences present in a replicon are potential sites for homologous recombination, leading to genomic rearrangement. Depending on the size, location and orientation of the repeated DNA sequences, different types of rearrangement events may be generated. Recombination between a pair of repeated sequences in direct orientation can lead to deletion or duplication of the region between them, whereas recombination involving inverted elements may produce an inversion. Rearrangements generated by homologous recombination between repeated sequences were found in many *Rhizobia* species by using particular experimental strategies or genetic elements (Flores et al. 1988; Brom et al. 1991; Romero et al. 1991; Flores et al. 1993, Romero and Palacios 1997). In *R. etli*, amplification and deletion of the 120-kb segment bracketed by two *nifHDK* operons occur in about 1 in 10^{-3} to 10^{-5} cells (Romero et al., 1991; Flores et al. 1993). Mavingui et al. (1998) have demonstrated the occurrence of amplification and deletion of a symbiotic region of 60 kb flanked by two *ISRtr1* copies in *Rhizobium tropici*.

The availability of the complete pNGR234a sequence and the presence of a huge number of repeated DNA elements on this replicon have prompted studies on the rearrangement events on this replicon. To investigate intra-genomic rearrangement on pNGR234a, Flores et al. (2000) analyzed first the whole sequence and identified reiterated sequences. Location and orientation of reiterated DNA sequences allowed them to predict the potential rearrangements that might be generated by homologous recombination. By employing an artificial selection coupled with PCR strategy, all the major 14 rearrangements (7 amplifications and 7 deletions) predicted were evidenced in a culture of the wild-type NGR234 strain (Fig. 2; Flores et al. 2000).

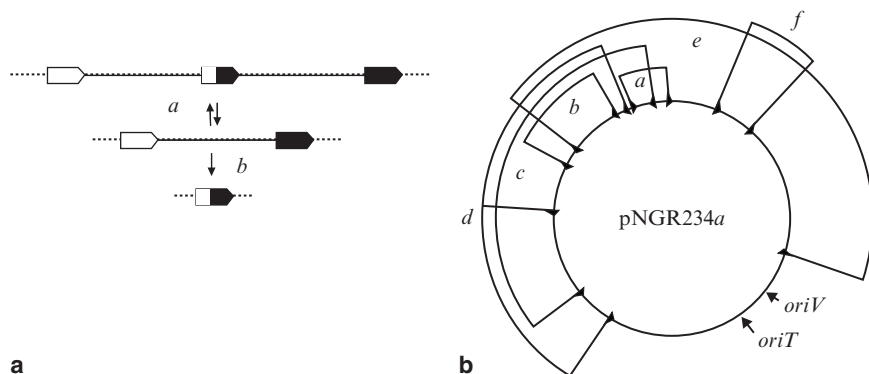


Fig. 2 Dynamics of intragenomic rearrangements predicted in pNGR234a. (a) Schematic representation of duplication (a) that is a reversible event (double arrows) and deletion (b), a non reversible event (one arrow), resulted from the recombination between repeats (white and black arrow heads). (b) The internal circle represents the pNGR234a replicon with direct repeats symbolized by arrowheads. Pairs of identical repeated sequences are connected by thick arcs that denote regions (a to f) prone to rearrangement. Position of origins of replication (*oriV*) and transfer (*oriT*) are indicated. (Adapted from Flores et al. 2000)

As mentioned earlier, strain NGR234 harbours three replicons: a chromosome, a megaplasmid pNGR234*b* and the pSym pNGR234*a*. It was shown that some repeated DNA sequences such as ISs of pNGR234*a* are shared by the other two replicons (Perret et al., 1997, 2000b). We have investigated if recombination may also occur between these repeats. Recombination between reiterated DNA sequences located on different replicons may lead to co-integration (Campbell 1969). The analysis of replicon profiles of individual colonies derived from the culture of NGR234 strain revealed alternative genomic architectures on Eckhardt gels and pulse field gel electrophoresis. Recombination mediated by ISs has led to co-integration between two replicons (pNGR234*a*-pNGR234*b*, pNGR234*a*-chromosome, pNGR234*b*-chromosome) as well as between the three replicons generating one large replicon (Mavingui et al. 2002). Similar co-integration events were also demonstrated between the replicons of *S. meliloti* strain 1021 (Guo et al. 2003).

In some cases, genomic rearrangement affects bacterial responses to different environments. For instance, gene amplification of *nod* and *nif* genes are associated with enhanced symbiotic properties of amplified derivatives of *R. tropici* CFN299 (Mavingui et al., 1997, 1998) and *S. meliloti* strain 1021 (Castillo et al. 1999). Conversely, deletion of the same genes abolishes nodulation and nitrogen fixation functions. However, the different co-integrated genome architectures adopted by both NGR234 and *S. meliloti* 1021 strains did not alter significantly their symbiotic capacities (Mavingui et al. 2002, Guo et al. 2003).

4 Nodulation and Nitrogen Fixation Loci

The pioneer investigations that detected *nod*, *nif* and *fix* genes on pNGR234*a* were performed through Southern hybridisations (Pankhurst et al. 1983; Broughton et al. 1984). Later, sequencing of the complete 536,165 bp of pNGR234*a* confirmed the presence of most *nod*, *nif* and *fix* genes on this replicon (Freiberg et al. 1997).

4.1 Nodulation Genes

In rhizobia-plant interactions, nodule formation is the result of a cellular differentiation process mediated by molecular signal exchange, which involves the expression of specific genes in both partners. In response to plant flavonoid compounds, the bacteria produce a family of lipo-chito-oligosaccharide molecules, called nodulation (Nod) factors, which in turn elicit nodule development on roots or stems of the legumes (Dénarié and Cullimore 1993; Schultze et al. 1994). In *R. sp.* NGR234 strain, most nodulation (*nod*, *nol*, and *noe*) genes involved in the production of Nod factors are located on pNGR234*a*, where they appeared dispersed throughout the replicon. These include the transcriptional regulators of the LysR family NodD1, NodD2, SyrM1 and SyrM2, all of which bind to conserved

NodD-dependant promoters designated *nod* boxes. A total of 19 *nod* boxes are linked to the promoter regions of *nod* genes on pNGR234a, 11 of which were shown to be symbiotically active (Freiberg et al. 1997; Perret et al. 1999). Genetically, structural host-specificity nodulation loci comprise *hsnI* (*nodZ*), *hsnII* (*nodSU*) and *hsnIII* (*nodABCIIJ-nolO-noeI* and *noeE*), which are involved in the synthesis of the core Nod factor and its decoration by the addition of carbomoyl, methyl, sulphate and acetyl groups (Fellay et al. 1995, Freiberg et al. 1997). In transcriptional analysis, Perret et al. (1999) showed that many of these *nod* genes are expressed both in determinate and indeterminate nodules. As mentioned earlier, pNGR234a does not encode *nodPQ* genes; two copies each of these genes are located on pNGR234b and the chromosome that contains also *nodE* and *nodG* (Perret et al. 1991).

As mentioned earlier, host-specific regulation of nodulation loci in rhizobia is mediated by a plant signal interacting with the NodD proteins that bind to *nod* boxes and activate in turn other *nod* genes. In NGR234 strain, *nodDI* is essential for plant nodulation; its inactivation is sufficient to generate a Nod⁻ mutant (Relic' et al. 1993). On the contrary, transfer of NGR234 *nodDI* to restricted-host-range rhizobia extends the nodulation capacity of the recipients (Bender et al. 1988; Horvath et al. 1987). Thus, it is believed that a large part of NGR234 broad host spectrum is linked to the ability of the *nodDI* product to interact with a huge number of flavonoids and related compounds (Perret et al. 2000a; Le Strange et al. 1990). Nevertheless, other *nod* regulators as well as *nod* genes involved in the modification Nod factor backbone mediated the production of a variety of compounds that contribute to the full, wild-type, host-range behaviour of NGR234 strain.

4.2 Nitrogen Fixation Gene Cluster

Once inside the nodule, expression of bacterial *nif* and *fix* genes are required for effective symbiosis. pNGR234a contains 43 *nif* and *fix* loci clustered in a single 55-kb region (Freiberg et al. 1997). These include the regulator gene *nifA*, two identical nitrogenase structural *nifHDK* operons involved in the synthesis of Mo-Fe nitrogenase and genes for electron transport and ferredoxin synthesis (*fixAB-CXfixBN*). A total of 16 NifA- σ 54 promoters were also present, and disruption of the *rpoN* gene that encodes σ 54 generates NGR234 derivatives unable to fix nitrogen (van Sooten et al. 1990). Several other expressed genes were demonstrated to be part of *nif* and *fix* cluster, although without assigned functions (Freiberg et al. 1997; Perret et al. 1999).

It should be mentioned that some *fix* genes are also present on the NGR234 chromosome, including homologues of *fixLJ* (David et al. 1988), *fixK* (Batut et al. 1989) and *fixNOPQ* (Preisig et al. 1993) whose products participate in regulation of nitrogen fixation, as well as *fixGHIS* (Kahn et al. 1989) that encode a membrane-bound complex with a cation pump. A copy of the sigma factor gene *rpoN* is also present.

5 T3SS in Symbiotic Process

Type III secretion systems (T3SSs) are used by many pathogenic and symbiotic gram-negative bacteria to export effector molecules into the milieu or directly into the cytoplasm of interacting eukaryotic cells (He 1998; Hueck 1998; Dale et al. 2002). Recently, a T3SS has been discovered in several *Rhizobium* species (Marie et al. 2001). Indeed, T3SSs were identified in *R. sp.* NGR234 (Freiberg et al. 1997), *M. loti* (Kaneko et al. 2000), *B. japonicum* (Göttfert et al. 2001) and *S. fredii* (de Lyra et al. 2000).

In *R. sp.* NGR234, T3SS determinants are located on the pSym pNGR234a (Fig. 3) in a region containing 29 predicted ORFs delimited by *ttsI*, a transcriptional activator, and *y4yS* (Viprey et al. 1998; Marie et al. 2004). Similar genetic organization was found in the T3SSs of the closest relative *S. fredii* (Jiang and Krishnan 2000; Marie et al. 2004; de Lyra et al. 2006). Viprey et al. (1998) reported an elegant study that evidenced polar mutations in two T3SS genes, *rhcN* and *y4xI*, which alter considerably the capacity of NGR234 to nodulate several compatible legumes. Sequence analysis suggests that RhcN is an ATPase that may provide energy for T3SS functioning, whereas *y4xI* shares features with transcription regulators. These authors further showed that disruption of *rhcN* and *y4xI* affects a number of downstream genes that encode components of the T3SS. Moreover, the above polar mutation also abolishes the secretion of two bacterial proteins (Y4xL and NolX) usually secreted upon induction by plant flavonoids (Viprey et al. 1998).

In the light of the aforementioned study, it was found that NGR234 secretes several other Nops through its T3SS. These include NopA and NopX thought to be pilus and translocon components, respectively (Marie et al. 2003), as well as NopB, which is associated with pilus-like appendages on the bacterial surface (Saad et al. 2005); these three proteins can interact in vitro (Saad et al. 2008). NopL is a putative effector protein that may interfere with the activation of plant defence (Bartsev et al. 2004), whereas NopT was shown to exhibit autoproteolytic activity in *Escherichia coli* or the human HEK 293T cell line (Dai et al. 2008). It is important to point out that several *tts* boxes were found in the promoter region of the predicted ORFs of the T3SS (Marie et al. 2004). The latest studies have demonstrated that under induction by plant flavonoids, the TtsI protein binds to *tts* boxes and controls downstream genes responsible for T3SS function as well as the synthesis of rhamnose-rich-lipo-polysaccharide, which is also important for fine-tuning of symbiosis (Wassem et al. 2008). Finally, it was established that abolition of Nop

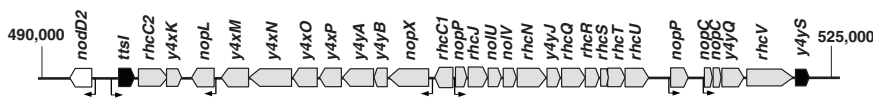


Fig. 3 Genetic organization of T3SS cluster on pNGR23a replicon. *Black boxes* delineate *bona fide* T3SS genes. *Bent arrows* labelled *nbx* and *tbx* indicate *nod* boxes and *tts* boxes, respectively (Adapted from Viprey et al. 1998 and Marie et al. 2004)

proteins affects, positively or negatively, the symbiotic interactions depending on the *nop* gene product and the legume species.

6 Global Transcriptional Analysis

The availability of complete DNA sequences of entire genomes or particular replicons has opened new avenues for large-scale gene expression analysis by the development of arrays of either DNA fragments or oligonucleotides (Schena et al. 1995; DeRisi et al. 1996; Lockhart et al. 1996).

Based on the complete DNA sequence of pNGR234a, its general coding capacity has been investigated in both free-living conditions and symbiosis. The first study was performed by Freiberg et al. (1997), in which PCR products of 113 predicted ORFs were blotted onto filter membranes and hybridized against RNAs extracted from NGR234 bacteroids isolated from nodules or cells cultivated in synthetic medium with or without flavonoids (daidzein, apigenin). Later on, Perret et al. (1999) extended the study to all of the 416 predicted ORFs and their intergenic regions. Overall, a total of 250 genes (60%) coded by pNGR234a were specifically expressed under symbiotic conditions, among which are genes regulated through active *nod* boxes or NifA-s54 promoters. After the addition of flavonoids to the bacterial culture medium, a rapid and strong induction of several genes responsible for Nod factorsynthesis occurs, reaching a maximal expression within an hour. This is in contrast to many other genes such as those located in the T3SS cluster, which respond 24 h or more later following flavonoid induction. Transcripts recovered from bacteroids indicate that *nif* and *fix* loci, including nitrogenase structural genes, are actively expressed in nodules, whereas flavonoid-inducible genes are repressed (Perret et al. 1999). Interestingly, some ORFs such as the *repC* gene involved in pNGR234a replication, the *nopL* gene (formerly *y4xL*), the product of which is secreted through T3SS, and many IS elements were transcribed in all conditions; notably in nodules as well as in both induced and noninduced free-living states (Freiberg et al. 1997; Viprey et al. 1998, Perret et al. 1999).

7 Conclusion

Rhizobium strain NGR234 has the broadest host range for legume nodulation known to date. It harbours three replicons, a chromosome and two megaplasmids, one of which is the symbiotic plasmid pNGR234a. As reviewed in this chapter, the presence of diverse symbiotic genes with concerted expression profil supports the idea that pNGR234a is essential for the promiscuous behaviour of NGR234 strain. However, additional genes that are important for efficient symbiosis are not pNGR234a-borne. The ongoing sequencing project of the other two replicons (chromosome and pNGR234b) at the Göttingen Genomics Laboratory

(<http://www.g2l.bio.uni-goettingen.de/project>) will allow deciphering the whole potential of the NGR234 genome for establishing symbiosis with as many as a hundred legume species.

References

- Banfalvi Z, Sakanyan V, Koncz G, Kiss A, Dusha I, Kondorosi A (1981) Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *Rhizobium meliloti*. *Mol Gen Genet* 184:318–325
- Barloy-Hubler F, Jebbar M (2008) *Sinorhizobium meliloti* megaplasmids and symbiosis in *S. meliloti*. doi:10.1007/7171_2008_16
- Bartsev AV, Deakin WJ, Boukli NM, McAlvin CB, Stacey G, Malnoe P, Broughton WJ, Staehelin C (2004) NopL, an effector protein of *Rhizobium* sp. NGR234, thwarts activation of plant defense reactions. *Plant Physiol* 134:871–879
- Battisti L, Lara JC, Leigh JA (1992) Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. *Proc Natl Acad Sci USA* 89:5625–5629
- Batut J, Daveran ML, David M, Jacobs J, Garnerone AM, Kahn D (1989). *fixK*, a gene homologous with *fur* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. *EMBO J* 8:1279–1286
- Bender GL, Nayudu M, Strange KL, Rolfe BG (1988) The *nodD1* gene from *Rhizobium* strain NGR234 is a key determinant in the extension of host range to nonlegume *Parasponia*. *Mol Plant-Microbe Interact* 1:259–266
- Brom S, Garcia de los Santos A, de Lourdes Girard M, Davila G, Palacios R, Romero D (1991) High-frequency rearrangements in *Rhizobium leguminosarum* bv. *phaseoli* plasmid. *J Bacteriol* 173:2411–2419
- Broughton WJ, Dilworth MJ, Passmore IK (1972) Base ratio determination using unpurified DNA. *Anal Biochem* 46:164–172
- Broughton WJ, Heycke N, Meyer ZA, Pankhurst CE (1984) Plasmid link *nif* and *nod* genes in fast-growing rhizobia that nodulate *Glycine max*, *Psophocarpus tetragonolobus*, and *Vigna unguiculata*. *Proc Natl Acad Sci USA* 82:3093–3097
- Broughton WJ, Wong CH, Lewin A, Samrey U, Myint H, Meyer H et-al. (1986) Identification of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus*, *Vigna*, and other legumes. *J Cell Biol* 102:1173–1182
- Campbell AM (1969) Episomes. Harper & Row, New York, NY
- Castillo M, Flores M, Mavingui P, Martinez-Romero E, Palacios R, Hernandez G (1999) Increase in alfalfa nodulation, nitrogen fixation, and plant growth by specific DNA amplification in *Sinorhizobium meliloti*. *Appl Environ Microbiol* 65:2716–2722
- Dai WJ, Zeng Y, Xie XP, Staehelin C (2008) Symbiosis-promoting and deleterious effects of NopT, a novel type 3 effector of *Rhizobium* sp. NGR234. *J Bacteriol* 190:5101–5110
- Dale C, Plague GR, Wang B, Ochman H, Moran NA (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. *Proc Natl Acad Sci USA* 99:12397–12402
- David M, Daveran ML, Batut J, Dedieu A, Domergue O, Ghai J et-al. (1988) Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. *Cell* 26:671–683
- de Lyra MCPP, Lopez-Baena FJ, Madinabeitia N, Vinardell JM, Espuny MR, Cubo MT, Bellogin RA, Ruiz-Sainz JE, Ollero FJ (2006) Inactivation of the *Sinorhizobium fredii* HH103 *rhcJ* gene abolishes nodulation outer proteins (Nops) secretion and decreases the symbiotic capacity with soybean. *Int Microbiol* 9:125–133
- de Lyra MCPP, Ollero FJ, Madinabeitia N, Espuny MR, Bellogin RA, Cubo MT, Ruiz-Sainz JE (2000) Characterization of a *nolT* mutant of *Sinorhizobium fredii* HH103: its role in type III

- secretion protein. In: Fourth European nitrogen fixation conference abstract book. Sevilla, Viceconsejería. 192
- Dénarié J, Cullimore J (1993) Lipo-oligosaccharide nodulation factors: a minireview new class of signaling molecules mediating recognition and morphogenesis. *Cell* 74:951–954
- Dénarié J, Debelle F, Rosenberg G (1992) Signaling and host range variation in nodulation. *Ann Rev Microbiol* 46:497–531
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14:457–460
- Fellay R, Perret X, Viprey V, Broughton WJ, Brenner SQ (1995) Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol Microbiol* 16:657–667
- Flores M, González V, Pardo MA, Laija A, Martínez E, Romero D, Piñero D, Dávila G, Palacios R (1988) Genomic instability in *Rhizobium phaseoli*. *J Bacteriol* 170:1191–1196
- Flores M, Brom S, Stepkowski T, Girard ML, Dávila G, Romero D, Palacios R (1993) Gene amplification in *Rhizobium*: identification and *in vivo* cloning of discrete amplifiable DNA regions (amplicons) from *Rhizobium leguminosarum* biovar *phaseoli*. *Proc Natl Acad Sci USA* 90:4932–4936
- Flores M, Mavingui P, Girard L, Peret X, Broughton WJ, Martínez-Romero E et-al. (1998) Three replicons of *Rhizobium* sp. strain NGR234 harbor symbiotic gene sequences. *J Bacteriol* 180:6052–6053
- Flores M, Mavingui P, Perret X, Broughton WJ, Romero D, Hernández G et-al. (2000) Prediction, identification, and artificial selection of DNA rearrangements in *Rhizobium*: Toward a natural genomic design. *Proc Natl Acad Sci USA* 97:9138–9143
- Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* 387:394–401
- Fuqua C, Greenberg EP (2002) Listening in on bacteria: acylhomoserine lactone signalling. *Nat Rev Mol Cell Biol* 3:685–695
- Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F et-al. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293:668–672
- Gay P, Le coq D, Steinmetz M, Berkelman T, Kado CI (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* 164:918–921
- Göttfert M, Rothlisberger S, Kundig C, Beck C, Marty R, Hennecke H (2001) Potential symbiosis-specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *J Bacteriol* 183:1405–1412
- Guo X, Flores M, Mavingui P, Fuentes SI, Hernández G, Dávila G, Palacios R (2003) Natural genomic design in *Sinorhizobium meliloti*: Novel genomic architectures. *Genome Res* 13:1810–1817
- He SY (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Ann Rev Phytopathol* 36:363–392
- He X, Chang W, Pierce DL, Ort Seib L, Wagner J, Fuqua C 2003. Quorum sensing in *Rhizobium* sp. strain NGR234 regulates conjugal transfer (*tra*) gene expression and influences growth rate. *J Bacteriol* 185:809–822
- Hernández-Lucas I, Ramirez-Trujillo JA, Gaitan MG, Guo X, Flores M, Martinez-Romero E, Perez-Rueda E, Mavingui P (2006) Isolation and characterization of functional insertion sequences of rhizobia. *FEMS Microbiol Lett* 261:25–31
- Horvath B, Bachem CW, Schell J, Kondorosi A (1987) Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant signal interacting with the *nodD* gene. *EMBO J* 6:841–848
- Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 62:379–433
- Hwang I, Li P-L, Zhang L, Piper KR, Cook DM, Tate ME, Farrand SK (1994) TraI, a LuxI homologue, is responsible for production of conjugation factor, the Ti plasmid *N*-acylhomoserine lactone autoinducer. *Proc Natl Acad Sci USA* 91:4639–4643

- Hwang I, Smyth AJ, Luo Z-Q, Farrand SK (1999) Modulating quorum sensing by antiactivation: TraM interacts with TraR to inhibit activation of Ti plasmid conjugal transfer genes. *Mol Microbiol* 34:282–294
- Jiang G, Krishnan HB (2000) *Sinorhizobium fredii* USDA257, a cultivar-specific soybean symbiont, carries two copies of *y4yA* and *y4yB*, two open reading frames that are located in a region that encodes the type III protein secretion system. *Mol Plant-Microbe Interact* 13:1010–1014
- Kahn D, David M, Domergue O, Daveran M, Ghai J, Hirsch PR, Batut J (1989) *Rhizobium meliloti* *fxGHI* sequence predicts involvement of specific cation pump in symbiotic nitrogen fixation. *J Bacteriol* 171:929–939
- Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto SA et-al. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* 7:331–338
- Kündig C, Hennecke H, Göttfert M (1993) Correlated physical and genetic map of *Bradyrhizobium japonicum* 110 genome. *J Bacteriol* 175:613–622
- Le Strange KK, Bender GL, Djordjevic MA, Rolfe BG, Redmond JW (1990) The *Rhizobium* strain NGR234 *nodD1* gene product responds to activation by simple phenolic compounds vanillin and isovanillin present in wheat seedling extracts. *Mol Plant-Microbe Interact* 3:214–220
- Lewin A, Rosenberg C, Stanley J, Downling DN, Manen J-F, Debellé F, Broughton WJ (1987) Multiple host-specificity loci in the broad host-range *Rhizobium* strain NGR234. In: Verma DPS, Brisson N (ed) Molecular genetics of plant-microbe interactions. Martinus Nijhoff, Dordrecht, The Netherlands, p 232–237
- Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680
- Margolin W, Long SR (1993) Isolation and characterization of a DNA replication origin from the 1,700-kilobase-pair symbiotic megaplasmid pSym-b of *Rhizobium meliloti*. *J Bacteriol* 175:6553–6561
- Marie C, Broughton WJ, Deakin WJ (2001) *Rhizobium* type III secretion systems: legume charmers or alarmers? *Curr Opin Plant Biol* 4:336–342
- Marie C, Deakin WJ, Viprey V, Kopcińska J, Golinowski W, Krishnan HB, Perret X, Broughton WJ (2003) Characterization of Nops, nodulation outer proteins, secreted via the type III secretion system of NGR234. *Mol Plant-Microbe Interact* 16:743–751
- Marie C, Deakin WJ, Reuhs TO, Diallo E, Reuhs B, Broughton W, Perret X (2004) TtsI, a key regulator of *Rhizobium* species NGR234, is required for type III-dependent protein secretion and synthesis of rhamnoseric polysaccharides. *Mol Plant-Microbe Interact* 17:958–966
- Mavingui P, Flores M, Romero D, Martinez-Romero E, Palacios R (1997) Generation of *Rhizobium* strains with improved symbiotic properties by random DNA amplification (RDA). *Nat Biotechnol* 15:564–569
- Mavingui P, Laeremans T, Flores M, Romero D, Martinez-Romero E, Palacios R (1998) Genes essential for nod factor production and nodulation are located on a symbiotic amplicon (AMPRtrCFN299pc60) in *Rhizobium tropici*. *J Bacteriol* 180:2866–2874
- Mavingui P, Flores M, Guo X, Dávila G, Perret X, Broughton WJ et-al. (2002) Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. *J Bacteriol* 184:171–176
- Mavingui P, Perret X, Broughton WJ (2005) Genomic architecture of the multiple replicons of the promiscuous *Rhizobium* species NGR234. In: Palacios R, Newton WE (eds) Genomes and genomics of nitrogen-fixing organisms. Springer, Dordrecht, The Netherlands, pp 83–98
- Morrison NA, Hau CY, Trinick MJ, Shine J, Rolfe BG (1983) Heat curing of a Sym plasmid in a fast-growing *Rhizobium* sp. that is able to nodulate legumes and the nonlegume *Parasponia* sp. *J Bacteriol* 153:527–531
- Morrison NA, Cen YH, Chen HC, Plazinski J, Rdige R, Rolfe BG (1984) Mobilization of a sym plasmid from a fast-growing cowpea *Rhizobium* strain. *J Bacteriol* 160:483–487
- Nishiguchi R, Takanami M, Oka A (1987) Characterization and sequence determination of the hairy root inducing plasmid pRiA4b. *Mol Gen Genet* 206:1–8

- Pankhurst CR, Broughton WJ, Bachem C, Kondorosi E, Kondorosi A (1983) Identification of nitrogen fixation and nodulation genes on a large plasmid from a broad host range *Rhizobium* sp. In: Puhler A (ed) Molecular genetics of the bacteria-plant microbe interaction. Springer-Verlag, Berlin Heidelberg, pp 169–176
- Perret X, Broughton WJ, Brenner S (1991) Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. *Proc Natl Acad Sci USA* 88:1923–1927
- Perret X, Viprey V, Freiberg C, Broughton WJ (1997) Structure and evolution of NGRRS-1, a complex, repeated element in the genome of *Rhizobium* sp. NGR234. *J Bacteriol* 179:7488–7496
- Perret X, Freiberg C, Rosenthal A, Broughton WJ, Fellay R (1999) High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234. *Mol Microbiol* 32:415–425
- Perret X, Staehelin C, Broughton WJ. (2000a) Molecular basis of symbiotic promiscuity. *Microbiol Mol Biol Rev* 64:180–201
- Perret X, Viprey V, Broughton WJ (2000b) Physical and genetic analysis of the broad-host range *Rhizobium* sp. NGR234. In: Triplett EW (ed) Prokaryotic nitrogen fixation. HorizonScientific Press, Wymondham, Norfolk, UK, pp 679–692
- Preisig O, Anthamatten D, Hennecke H (1993) Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. *Proc Natl Acad Sci USA* 90:3309–3313
- Pueppke SG, Broughton WJ (1999) *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host-ranges. *Mol Plant-Microbe Interact* 12:293–318
- Ramirez-Romero MA, Bustos P, Girard L, Rodriguez O, Cevallos MA, Davila G (1997) Sequence, localization and characteristics of the replicator region of the symbiotic plasmid of *Rhizobium etli*. *Microbiology* 143:2825–2831
- Relic' B, Talmont F, Kopcinska J, Golinowsky W, Promé JC, Broughton WJ (1993) Biological activity of *Rhizobium* sp. NGR234 *Nod*-factors on *Macropitilium atropurpureum*. *Mol Plant-Microbe Interact* 6:764–774
- Romero D, Palacios R (1997) Gene amplification and genomic plasticity in prokaryotes. *Ann Rev Genet* 31:91–111
- Romero D, Brom S, Martinez-Salazar J, Girard ML, Palacios R, Davila G (1991) Amplification and deletion of a *nod-nif* region in the symbiotic plasmid of *Rhizobium phaseoli*. *J Bacteriol* 173:2435–2441
- Rosenberg C, Boistard P, Dénarié J, Casse-Delbart F (1981) Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol Gen Genet* 184:326–333
- Ruvkun GB, Long SR, Meade HM, van den Bos RC, Ausubel FM (1982) IsRm1: a *Rhizobiummeliloti* insertion sequence that transposes preferentially into nitrogen fixation genes. *J Mol Appl Genet* 1:405–418
- Saad MM, Kobayashi H, Marie C, Brown IR, Mansfield JW, Broughton WJ, Deakin WJ (2005) NopB, a type III secreted protein of *Rhizobium* sp. strain NGR23, is associated with pilus-like surface appendages. *J Bacteriol* 187:1173–1181
- Saad MM, Staehelin C, Broughton WJ, Deakin WJ (2008) Protein–protein interactions within type III secretion system-dependent pili of *Rhizobium* sp. strain NGR23. *J Bacteriol* 190:750–7545
- Schena M, Shanon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Schultze M, Kondorosi E, Dénarié J, Buiré M, Kondorosi A (1994) Cell and molecular biology of *Rhizobium*–plant interactions. *Int Rev Cytol* 156:1–74
- Schwedock J, Long S (1990) ATP sulphurylase activity of the *nodP* and *nodQ* gene products of *Rhizobium meliloti*. *Nature* 348:644–647
- Simon R, Hotte B, Klauke B, Kosier B (1991) Isolation and characterization of insertion sequence elements from gram-negative bacteria by using new broad-host-range, positive selection vectors. *J Bacteriol* 173:1502–1508

- Staehelin C, Forsberg LS, D'Haese W, Gao M-Y, Carlson RW, Xie Z-P, Pellock BJ, Jones KM, Walker GC (2006) Exo-oligosaccharides of *Rhizobium* sp. strain NGR234 are required for symbiosis with various legumes. *J Bacteriol* 188:6168–6178
- Streit WR, Schmitz RA, Perret X, Staehelin C, Deakin WJ, Raasch C, Liesegang H, Broughton WJ (2004) An evolutionary hot spot: the pNGR234b replicon of *Rhizobium* sp. strain NGR234. *J Bacteriol* 186:535–542
- Sullivan JT, Ronson CW (1998) Evolution of rhizobia by acquisition of a 500 kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci USA* 95:5145–5149
- Tabata S, Hooykaas PJJ, Oka A (1989) Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3. *J Bacteriol* 171:1665–1672
- Trinick MJ (1980) Relationships among the fast-growing rhizobia of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* sp., *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with other rhizobia groups. *J Appl Bacteriol* 49:39–53
- Turner SL, Young PW (1995) The replicator region of the *Rhizobium leguminosarum* cryptic plasmid pRL8JI. *FEMS Microbiol Lett* 133:53–58
- Van Slooten JC, Cervantes E, Broughton WJ, Wong CH, Stanley J (1990) Sequence and analysis of the *rpoN* sigma factor gene of *Rhizobium* sp. strain NGR234, a primary coregulator of symbiosis. *J Bacteriol* 172:5563–5574
- Viprey V, Del Geco A, Golinowski W, Broughton WJ, Perret X (1998) Symbiotic implications of type III protein secretion machinery in *Rhizobium*. *Mol Microbiol* 28:1381–1389
- Viprey V, Rosenthal A, Broughton WJ, Perret X (2000) Genetic snapshots of the *Rhizobium* species NGR234 genome. *Genome Biol* 1:14.1–14.7
- Wassem R, Kobayashi H, Kambara K, Le Quéré A, Walker GC, Broughton WJ, Deakin WJ (2008) TtsI regulates symbiotic genes in *Rhizobium* species NGR234 by binding to tts boxes. *Mol Microbiol* 68:736–748
- Whitehead NA, Barnard AM, Slater H, Simpson NJ, Salmond GP (2001) Quorum-sensing in gram-negative bacteria. *FEMS Microbiol Rev* 25:365–404

Ti and Ri Plasmids

Katsunori Suzuki(✉), Katsuyuki Tanaka, Shinji Yamamoto,
Kazuya Kiyokawa, Kazuki Moriguchi, and Kazuo Yoshida

Contents

1	Introduction	134
2	Chimeric Structure	136
3	T-DNA	137
4	Virulence Genes	139
4.1	Regulation by VirA and VirG	139
4.2	<i>vir</i> Operon Genes for Processing and Transfer of T-DNA	139
4.3	Auxiliary <i>vir</i> Genes	141
5	Opine Utilization Genes	142
6	Replication and Stability Genes	142
6.1	<i>repABC</i> Genes	142
6.2	Incompatibility and Stability Enhancing Genes	142
7	Conjugation Genes	143
8	Perspectives	143
	References	144

Abstract *Agrobacterium* species harboring tumor-inducing (Ti) or hairy root-inducing (Ri) plasmids cause crown gall or hairy root diseases, respectively. These natural plasmids provide the basis for vectors to construct transgenic plants. The plasmids are approximately 200 kbp in size. Complete sequence analysis indicates that the pathogenic plasmids contain gene clusters for DNA replication, virulence, T-DNA, opine utilization, and conjugation. T-DNA genes have lower G + C content, which is presumably suitable for expression in host plant cells. Besides these genes, each plasmid contains a large number of unique genes. Even plasmids of the same opine type differ considerably in gene content and have highly chimeric structures. The plasmids seem to interact with each other and with plasmids of other members of the *Rhizobiaceae* and are likely to shuffle genes for infection between

K. Suzuki, K. Tanaka, S. Yamamoto, K. Kiyokawa, and K. Moriguchi
Graduate School of Science, Hiroshima University, Higashi-Hiroshima, 739-8526, Japan
e-mail: ksuzuki@hiroshima-u.ac.jp

K. Yoshida
Department of Nutrition Science, Yasuda Women's University, Hiroshima, Japan

Ti and Ri plasmids. Plasmid stability genes are discussed, which are important for plasmid evolution and construction of useful strains.

1 Introduction

A group of *Agrobacterium* species are the causative agent of crown gall disease and hairy root disease on dicot plants (Fig. 1). On the positive side from the point of view of biotechnology and molecular biology, the species with the infection system are also the most reliable tool for introducing chimeric DNA into plants. *A. tumefaciens*, *A. rhizogenes*, and *A. vitis*, also termed *Agrobacterium* biovar 1, biovar 2, and biovar 3, respectively, are the three major pathogenic *Agrobacterium* species. Recent bacterial classification schemes use the genus *Rhizobium* in place of the genus *Agrobacterium* (Young et al. 2003), since the strains are indistinguishable from *Rhizobium* strains except for their pathogenic and the nitrogen-fixing characteristics. In fact, the pathogenic strain ATCC11325, which is the type strain of *A. rhizogenes*, contains a symbiotic plasmid (Velázquez et al. 2005). In this paper, however, we use the genus *Agrobacterium* to highlight the pathogenic species. Most genes essential for pathogenicity are found on the large plasmids, approximately 200 kbp, called tumor-inducing (Ti) or root-inducing (Ri) plasmids. Introduction of the plasmids

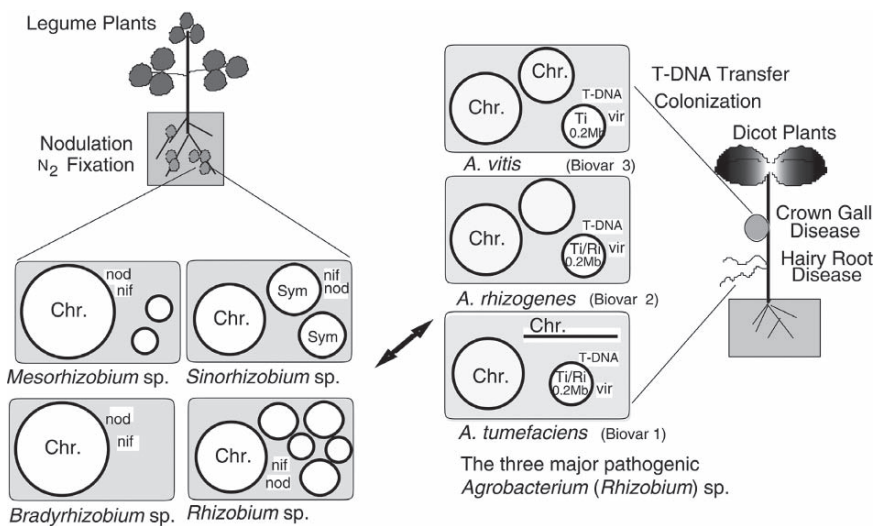


Fig. 1 Association of symbiotic and pathogenic strains of the order *Rhizobiales* with plants. The double arrow indicates that gene transfer between the pathogenic and nitrogen fixing bacteria takes place. The evolutionary relationships between the symbiotic plasmids are observable in the plasmid sequences. See text for details. Genomic structures of the species are based on Kaneko et al. (2000, 2002), González et al. (2006), Suzuki et al. (2001), Urbanczyk et al. (2003), and Tanaka et al. (2006)

into nonpathogenic *Agrobacterium* strains and *Rhizobium* strains makes them pathogenic (Klein and Klein 1953, Hooykaas et al. 1977). Better understanding of the plasmids should give insights useful for controlling the spread of the diseases and also improve plant and fungal transformation technology (Lacroix et al. 2006).

Ti and Ri plasmids are classified by organic compounds termed opines, which are produced by infected plant cells and delivered to the pathogen as nutrients. Opine synthesis is directed by the genes on the T-DNA portion of the plasmids in plant cells. The complete sequences of two nopaline-type plasmids, pTi-SAKURA and pTiC58; a mikimopine-type plasmid pRi1724; a cucumopine-type plasmid pRi2659; an agropine/mannopine-type Ti plasmid, pTiBo542; and a composite assembly of five octopine-type plasmids, pTiA6NC, pTi15955, pTiAch5, pTiR10, and pTiB6S3, have been reported (Suzuki et al. 2000, Goodner et al. 2001, Wood et al. 2001, Moriguchi et al. 2001, Mankin et al. 2007, Zhu et al. 2000). **Table 1** lists the general information for the six plasmids. In this paper, we describe the structural and functional properties of Ti and Ri plasmids based on these plasmid sequences as well as several partially sequenced plasmids. For a general review on Ti plasmids, refer to the article by Christie (2004a). For a general review with a biotechnological viewpoint, refer to the article by Gelvin (2003). More focused reviews are cited below in the sections dealing with specific subjects.

Table 1 List of Ti and Ri plasmids whose complete nucleotide sequences are available in the public DNA databases

Plasmid	<i>inc</i> group	Disease	Opine production ^a	Host range	Length (bp)	Accession number	Key reference
pTiC58	Rh-1	Crown gall	<u>Nopaline</u> Agrocinopine	Broad	214,234	AE007871	Goodner et al. (2001) Wood et al. (2001)
pTi-SAKURA	Rh-1	Crown gall	<u>Nopaline</u> Agrocinopine	Broad	206,479	AB016260	Suzuki et al. (2000)
Octopine type pTi ^b	Rh-1	Crown gall	<u>Octopine</u> agropine man- nopine	Broad	194,140	AF242881	Zhu et al. (2000)
pTiBo542	Rh-2	Crown gall	<u>Succinamopine</u> <u>Agropine</u> <u>Mannopine</u>	Broad	244,978	DQ058764	Oger et al.'s data in the DNA database
pRi1724	(Rh-3)	Hairy root	<u>Mikimopine</u>	Broad	217,594	AP002086	Moriguchi et al. (2001)
pRi2659	(Rh-3)	Hairy root	<u>Cucumopine</u>	Broad	202,297 ^c	EU186381 EF433766	Mankin et al. (2007) Collier et al.'s data in the DNA database

^a The opine type follows the opine(s) marked with underline

^b Composite sequence from five well-conserved plasmids

^c Calculated for this paper by assembling two nucleotide sequences from the two groups

2 Chimeric Structure

The overall G + C content of these plasmids is about 56%, which is slightly lower than that of *A. tumefaciens* total genomic DNA (Goodner et al. 2001, Wood et al. 2001). Ti and Ri plasmids consist of many segments with different G + C content (see Fig. 2 for example). A few segments are rich in A and T. T-DNA is especially abundant in A and T. Generally, functionally related genes are clustered in a segment with an even G + C content (Fig. 2). The data suggest a chimeric structure comprising segments from different sources. As illustrated in Fig. 3, there are five clusters common among Ti and Ri plasmids: (1) T-DNA, which is transferred to the host plants; (2) the virulence gene (*vir*) region, which directs the recognition of plant phenolic compounds, processing, and transfer of T-DNA; (3) the replication gene (*rep*) region, which is required for the plasmid replication; (4) *tra* and (5) *trb* regions, which direct conjugal transfer of the plasmid. Two or more regions that direct uptake and catabolism of a respective opine are present in each plasmid. An exception to the functional clustering is the *tra* and *trb* regions in Ti plasmids. These two regions are separated from each other by more than 60 kb in nopaline- and octopine- type plasmids (Suzuki et al. 2000; Zhu et al. 2000). We termed the

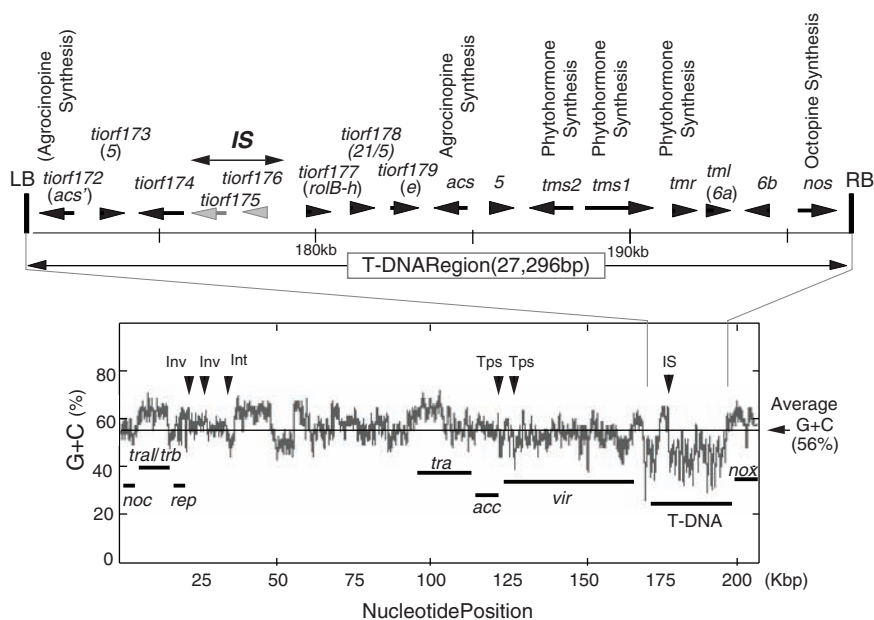


Fig. 2 Genetic organization and G + C distribution of the nopaline type plasmid pTi-SAKURA. *Lower panel*: G + C content (%) determined for a sliding window size of 400 bp. Functionally defined gene clusters are indicated by *thick horizontal bars*. *Arrowheads* indicate sequences for invertase (Inv), integrase (Int), transposase (Tps), and insertion sequence (IS). *Upper panel*: An exploded view of the T-DNA. Arrows give the size and direction of transcription of genes. See text for details

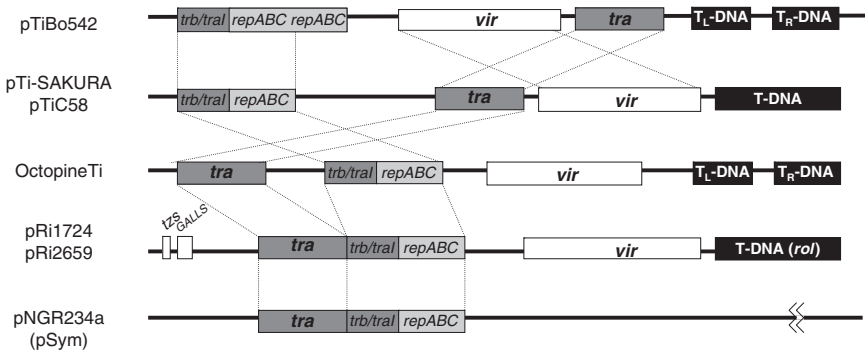


Fig. 3 Gene clusters commonly observable among Ti and Ri plasmids. Each box represents a gene cluster or a set of clusters. Functionally related segments are indicated by a box with the same shading level. The diagram is based on published sequence data and annotations (see [Table 1](#) for references and the database accession numbers). A symbiotic plasmid pNGR234a (536,165 bp; accession No. NC000914) was included as a reference replicon

large region the large variable region (VAR), since it is highly variable even between the two nopaline-type Ti plasmids (Suzuki et al. 2000; Goodner et al. 2001). The VAR region contains many genes with unknown function. Phylogenetic trees based on a *trb* gene and the *oriV* sequence indicate a closer relationship of Ri plasmids to the symbiotic plasmid pNGR234a. However, a tree based on a *vir* gene indicates that pRiA4b is closer to the nopaline type plasmids than to pRi1724 (see reviews by Yoshida et al. 2003, 2004). These data together with the G + C values indicate that the plasmids have exchanged and shuffled genes, including those for virulence.

As illustrated in [Fig. 1](#), strains in the order *Rhizobiales* have various types of genome constitutions. The nitrogen-fixing symbiosis system genes are observable in many genera either on plasmids or on chromosomes. Contrarily, the T-DNA/*vir* pathogenesis system has been found only on plasmids so far. Neither fragments of *vir* genes nor T-DNA genes were found on chromosomes. It would be interesting to survey chromosomal sequences for the presence of the pathogenesis gene system.

3 T-DNA

T-DNA is a DNA segment transferable to host cells and integrated into the host plant genomic DNA. Nopaline-type Ti plasmids and pRi1724 contain a single T-DNA region, whereas the agropine-type plasmid, pRiA4b, and octopine-type Ti plasmids contain two T-DNA regions (Suzuki et al. 2000, Goodner et al. 2001, Wood et al. 2001, Moriguchi et al. 2001, Zhu et al. 2000). The size of each T-DNA region(s) in a plasmid is around 20 kbp: 15,098 bp in pRi1724 and 27,296 bp in pTi-SAKURA. The two extreme ends of a T-DNA region are defined by 25-bp direct repeats, called right

border (RB) and left border (LB) sequences (Yadav et al. 1982). RB is essential for T-DNA transfer and is the initiation site for DNA processing. LB is dispensable for T-DNA transfer but is the termination site for single-strand DNA formation, which is initiated at RB. As shown in Fig. 2, the T-DNA coding regions have a G + C content of about 50% (Suzuki et al. 2000, Hattori et al. 2000, Zhu et al. 2000, Moriguchi et al. 2001), and the intergenic portions have less. The relatively low G + C content is likely suitable for efficient expression of T-DNA genes in plant cells. A portion in nopaline-type T-DNA exhibits high G + C content and contains an insertion sequence (Hattori et al. 2000, Suzuki et al. 2000).

Ti plasmids contain three T-DNA genes that direct synthesis of the plant hormones, cytokinin and auxin. Production of the two phytohormones in plants directed by the T-DNA genes primarily causes the tumor formation (for a review see Zambryski et al. 1989). Enzymes coded by *tmsA* and *tmsB* (also designated *iaaM* and *iaaH*, respectively) direct the conversion of tryptophan to indoleacetic acid. The *tmr* (*ipt*) gene directs production of cytokinin and *tml* (also designated gene *6b*) enhances cell division (Hooykaas et al. 1988). Gene *6b* from strain AKE10 can induce plant cell division even in media without phytohormones (Wabiko and Minemura 1996). The 6b protein localizes to the plant nucleus by interacting with NtSIP1 (Kitakura et al. 2002). The *rol* genes, *rolA*, *rolB*, and *rolC*, and *orf13* are conserved among Ri plasmid T-DNA regions. The *rolA* and *rolB* genes play primary roles in adventitious root induction (Zambryski et al. 1989) by a mechanism other than the production of phytohormones. Similar to the 6b protein of Ti T-DNA, RolB protein of pRi1724 is localized in the nucleus and interacts with 14-3-3-like proteins (Nt14-3-3) (Moriuchi et al. 2004). The 6b and Rol proteins are supposed to affect expression of plant genes that regulate plant cell division and morphology. Terakura et al. (2007) indicated that 6b protein binds specifically to histone H3 and has a histone-chaperon like activity, suggesting a relationship between alterations in nucleosome structure and the expression of growth-regulating genes.

Several genes that direct the synthesis of a different opine are observable in T-DNA. Two genes *nos* and *acs* in nopaline-type Ti plasmids encode nopaline synthase and agrocinopine synthase, respectively (Hattori et al. 2000). Two genes, *ocs* and *ags* in octopine-type Ti plasmids encode octopine synthase and agropine synthase, respectively. In addition, two genes, *mas1* and *mas2* in octopine-type plasmids direct synthesis of mannopines (Zhu et al. 2000). The T-DNA of pRi1724 harbors a mikimopine synthase gene (*mis*) and an unknown opine synthesis gene (Moriguchi et al. 2001). The opines are produced by expression of the genes in the infected plants. The product opines support bacterial growth as nutrients. The phenomenon is called “genetic colonization.” An agrocinopine-like compound is presumed to be an ancestral opine in the evolution of the T-DNA, because a homolog or fragment of the *acs* gene is present close to the right of LB of the T-DNA in most Ti and Ri plasmids (Paulus and Otten 1993; see also references for the complete nucleotide sequences given in Table 1) as shown, for example, in the ORF map in Fig. 2.

Expression of the T-DNA genes is repressed in *Agrobacterium* cells. However, the *ipt* gene is derepressed in a *ros* mutant *A. tumefaciens* cells. The *ros* gene codes

for a zinc finger motif-containing protein that binds to the 40-bp *ros*-box sequence in the operators of *ipt*, *virC/D* and the succinoglycan synthesis operon (for more details, see a review by Kado (2002)). It remains to be elucidated whether there are additional genes that repress T-DNA genes in bacterial cells.

4 Virulence Genes

In addition to T-DNA, virulence (*vir*) gene regions are essential for pathogenicity. The essential genes are coded in the core *vir* region consisting of four operons, *virB*, *virC*, *virD*, and *virE*, and two regulatory genes, *virA* and *virG*. The genes form a large cluster about 30 kbp in size. Both the nucleotide sequence of the genes in the core *vir* gene region and the order of the genes are well conserved among Ti and Ri plasmids. For more details on host recognition and DNA transfer, refer to the article by McCullen and Binns (2006).

4.1 Regulation by *VirA* and *VirG*

The two-component-system proteins, *VirA* and *VirG*, regulate virulence gene expression (Stachel and Zambryski 1986). The *vir* regulon is induced by low molecular weight phenolic compounds, such as acetosyringone, which are released from the host plant. The induction requires weakly acidic pH (pH 5.0–5.5) and moderate temperature (below 28°C) optimally in the presence of monosaccharides. The phenolics are detected by the transmembrane sensor kinase *VirA*, which phosphorylates the response regulator *VirG* (Fig. 4). Phosphorylated *VirG* binds to sequences called *vir* boxes upstream of *vir* regulon promoters and positively regulates *vir* gene expression (Roitsch et al. 1990). A periplasmic sugar binding protein, *ChvE* (Cangelosi et al. 1990, Shimoda et al. 1993), which is encoded by a chromosomal gene, associates with *VirA* for induction of *vir* genes.

4.2 *vir* Operon Genes for Processing and Transfer of T-DNA

T-DNA is processed by the action of the *vir* gene products (Fig. 4). The RB sequence is nicked by *VirD2* in the presence of *VirD1*, and simultaneously *VirD2* is linked covalently to the 5' end of the T-DNA strand (Scheffele et al. 1995). The *VirD2*-bound strand is then cleaved at the left border end and released from the plasmid. *VirC1* binds to a 24-bp sequence called overdrive (Peralta et al. 1986), which lies adjacent to RB and is required for efficient transfer (van Haaren et al. 1987, Toro et al. 1989). The single-stranded T-DNA covalently linked with *VirD2* is called T-complex. The opposite strand remains in the plasmid, and DNA polymerase fills in

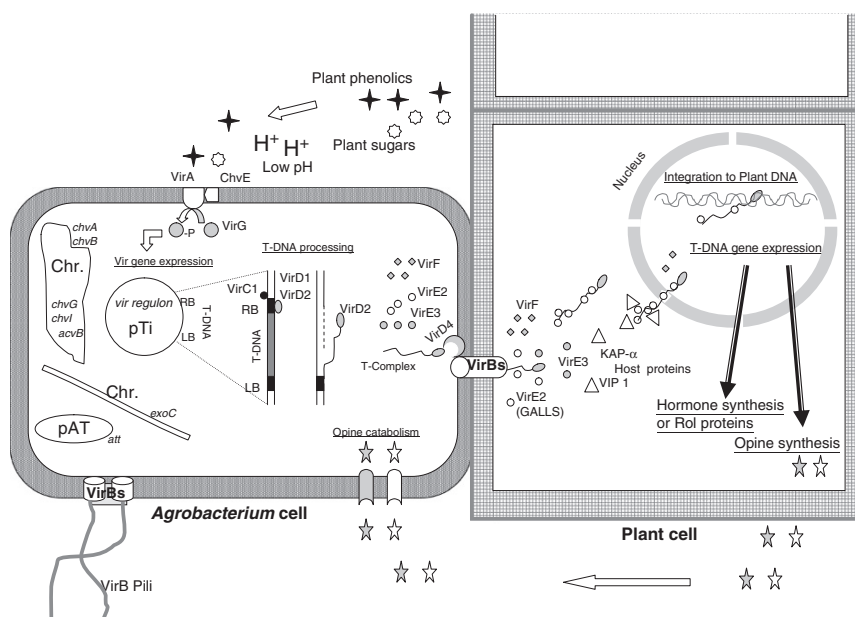


Fig. 4 Molecular mechanism of the interaction between pathogenic *Agrobacterium* cells and plant cells. See text for details

the single stranded portion on the plasmid. Ri plasmids contain 8-bp repeats, named TSS (T-DNA transfer stimulator sequence) (Hansen et al. 1992, Moriguchi et al. 2001), similar to overdrive. However, the function of TSS remains to be elucidated.

Proteins for the T-DNA transfer channel apparatus (see reviews by Christie 2004b, Schröder and Lanka 2005) are encoded by the *virB* operon, which contains 11 genes. Pili containing VirB2 as the major subunit form on the bacterial cell surface under conditions suitable for *vir* gene induction (Fullner et al. 1996). A coupling protein, VirD4, mediates the association of the T-complex with the VirB apparatus. The T-complex and several proteins including a single-stranded DNA binding protein, VirE2, are transferred to the cytoplasm of host plant cells via the VirB apparatus. VirE1 is a chaperone for VirE2 (Sundberg et al. 1996, Deng et al. 1999). In the plant cell cytoplasm, VirE2 covers the T-complex DNA and protects the single-stranded T-DNA. Import of the VirE2/T-complex into the plant nucleus through nuclear pores is mediated by nuclear localization signals (NLSs) in VirE2 and VirD2 (Citovsky et al. 1992, Howard et al. 1992). VirE2-interacting plant proteins were surveyed by yeast two hybrid screen, and termed as VirE2-interacting proteins (VIPs). A plant protein, VIP1, associates with VirE2 in the complex and helps its entry into the plant nucleus (Tzfira et al. 2001). Another plant protein, VIP2, is required for T-DNA integration in plants (Anand et al. 2007).

The two genes *virE1* and *virE2* are absent in pRiA4, pRi1724, and pRi2659 (Aoyama et al. 1989, Moriguchi et al. 2001, Mankin et al. 2007), although they are essential for Ti plasmid-directed tumorigenesis. The GALLS gene is present in the Ri plasmids and substitutes for *virE2* (Hodges et al. 2004, 2006). The GALLS protein does not resemble VirE2 and the gene is not located near the core *vir* region, but it has a NLS and is transferred from the bacteria to plant cells in the same manner as VirE2.

4.3 Auxiliary *vir* Genes

Several *vir* genes such as *virF*, *virJ*, and *virH* are not always essential for tumorigenesis. Presence and location of the auxiliary *vir* genes are variable depending on the plasmid. They are usually located near or within the core *vir* region and inducible by phenolics (Hattori et al. 2001). The host plant determinant gene *virF* is contained in octopine-type plasmids and many but not all nopaline-type plasmids (Schrammeijer et al. 1998). pTiC58, pTi-SAKURA, pTiBo542, and pRi1724 harbor a *virF* or *virF*-related gene. VirF is exported to host plant cells via the VirB apparatus, and then destabilizes the complex between ViE2 and VIP1 in the plant nucleus (Tzfira et al. 2004). Eventually, this VirF function enhances T-DNA integration in some plant species. Octopine-type plasmids and pTiBo542 have *virJ*, whereas nopaline-type plasmids and the two Ri plasmids, pRi1724 and pRi2659, do not. A chromosomal gene, *acvB* (Wirawan et al. 1993), is functionally equivalent to *virJ* (Pan et al. 1995). Thus, *virJ* is not necessary for tumorigenesis in the wild-type chromosomal background. Two genes *virH1* and *virH2* (also named *pinF1* and *pinF2*, respectively) encode cytochrome P450 family proteins, which have been supposed to detoxify plant substances. Actually, VirH2 converts the toxic inducer phenolic substance, ferrulic acid, to a less toxic noninducer, caffeic acid (Kalogeraki et al. 1999). VirH2 presumably supports infection and subsequent colonization in plants rich in ferrulic acid. Nopaline-type plasmids and the octopine-type plasmid contain both *virH* genes.

Nopaline-type plasmids contain *tzs* (trans-zeatin synthesis) gene (Suzuki et al. 2000, Goodner et al. 2001, Wood et al. 2001), which encodes a cytokinin biosynthetic prenyl transferase. The phytohormone is produced and released by the bacterium before and during infection. In plant transformation, pretreatment with auxin and cytokinin increases the transformation frequency, and auxin treatment suppresses silencing of gene expression (Dunoyer et al. 2006). Hormone production by enzymes encoded by *tzs* and hormone genes of T-DNA ensure infection and efficient expression of opine synthesis genes. The *tzs* gene is not present in octopine-type Ti plasmids (Zhu et al. 2000), whereas pRi1724 and pRi2659 contain *tzs* in the VAR region (Moriguchi et al. 2000, Mankin et al. 2007) as illustrated in Fig. 3. GALLS and *tzs* in the two Ri plasmids are located close to each other, but are far away (60 and 34 kbp in the two plasmids, respectively) from the core *vir* region. pTiBo542 harbors *virH1*, *virH2*, and *virJ*, but not *tzs*.

5 Opine Utilization Genes

In addition to the genes for opine synthesis, Ti and Ri plasmids encode genes for utilization of opines. Three operons, *noc*, *nox*, and *acc*, are responsible for uptake and catabolism of nopaline and agropinopine in nopaline-type plasmids (Hattori et al. 2000). More than 40 genes in octopine-type Ti plasmids are devoted to growth on opines, including octopine, agropine, and mannopine (Zhu et al. 2000). The opine utilization genes differ from plasmid to plasmid.

6 Replication and Stability Genes

6.1 *repABC* Genes

Ti plasmids are stably maintained at a low copy number comparable to chromosomal DNA copy number (Suzuki et al. 2001). Three genes, *repA*, *repB*, and *repC*, are sufficient for the replication and copy number control (Tabata et al. 1989). The *repC* gene is essential for replication, while *repA* and *repB* are required for stable plasmid inheritance. The copy number of pTiC58 is increased by binding of a conjugational regulatory protein TraR to a region called *tra*-box upstream from *repA* (Li and Farrand 2000). The *repABC* type replicators are commonly contained in large plasmids in members of the family *Rhizobiaceae*. In *A. tumefaciens*, strain C58, three replicons, pTiC58, a cryptic plasmid pAtC58 (543 kbp), and a linear chromosome (2,075 kbp) harbor *repABC* loci of their own (Goodner et al. 2001, Wood et al. 2001). As shown in Table 1, octopine-type and nopaline-type Ti plasmids belong to the same incompatibility group, IncRh-1. pTiBo542 and a vitopine-type plasmid pTiS4 belong to different groups, IncRh-2 and IncRh-4, respectively (Szegedi et al. 1996). Ti plasmids are compatible with Ri plasmids, which belong to group IncRh-3.

6.2 *Incompatibility and Stability Enhancing Genes*

It is known that many nopaline-type Ti plasmids are resistant to the exclusion pressure by incompatible plasmids. They fuse with incompatible plasmids upon encountering them in a cell, indicating that the nopaline-type plasmids are highly stable (Hooykaas et al. 1980). Most Ti plasmids are hard to cure. These characteristics cause serious difficulty in genetic manipulation of strains. However, there has been no report about stability-enhancing genes in Ti and Ri plasmids. In pTi-SAKURA, a locus containing two genes, *tiorf24* and *tiorf25*, enhances incompatibility of plasmids and increases stability of unstable plasmids (Yamamoto et al. 2007). Our additional data suggest that the plasmid stabilization mechanism is the toxin-antitoxin (TA)

addiction system, in which Tiorf24 and Tiorf25 are antitoxin and toxin, respectively. Tiorf25 protein contains a PIN domain, which is also found in the VapC protein encoded by the *vapBC* operon. Another locus pTi-SAKURA also helps to stabilize the plasmid (Yamamoto, personal communication). When Ti plasmids enter a cell that already contains an incompatible Ti plasmid, the resident Ti plasmid is usually expelled. However, a plasmid with increased stability is not always expelled but amalgamates with the incoming plasmid to form a larger plasmid. The resultant large co-integrate plasmids are likely to reduce their size subsequently to around 200 kbp, because all pathogenic strains so far examined possess a Ti and/or a Ri plasmid of this size. During this process, both regions that provide a selection advantage and also maintain stability determine which portions of the parental plasmids remain in the new plasmid.

Uraji et al. (2002) proposed a simple method to prepare Ti plasmid-less cells using a small *repABC* plasmid. The small plasmid remains in the resulting Ti plasmid-less cells and is easily removable when equipped with a counter-selectable marker gene, such as *sacB*. The Ti plasmid-less cells can then accept Ti plasmids with high efficiency and without formation of fusion plasmids. This method is applicable to many pathogenic strains, and thereby can help construction of new and useful strains for biotechnology (Tanaka et al. unpublished data).

7 Conjugation Genes

Ti plasmids harbor a *tra* gene region and a *tral/trb* operon. The *tra* gene region contains an origin of transfer (*oriT*) and DNA processing genes, while the *tral/trb* operon codes for components for formation of conjugal pili and mating pairs. TraI is an enzyme to synthesize *N*-acyl-homoserine lactone, which is the quorum-sensing signaling molecule for cell–cell communication (for a review see White and Winans 2007). Special opines induce the expression of the *tra* and *tral/trb* genes: octopine for octopine-type plasmids, agrocine A for nopaline type, and agrocine C and D for agropine-type plasmids (Dessaux et al. 1992). The *tral/trb* operon is located far away from the *tra* region in Ti plasmids, whereas the two regions are neighboring in Ri plasmids (Fig. 3), similar to those in several Sym plasmids such as pNGR234a.

8 Perspectives

In addition to the genes on Ti and/or Ri plasmids, a number of chromosomal genes are necessary for the pathogenicity of *Agrobacterium* species. Those chromosomal genes, such as the above mentioned *acvB*, *chvE*, and *ros* genes, are unevenly located on chromosomal replicon(s) (Suzuki et al. 2001, 2004). There has been some controversy about the involvement of an auxiliary plasmid pAtC58 in pathogenicity.

Nair et al. (2003) reexamined this and found that the plasmid is not necessary for but has a positive effect on *vir* gene induction. Evolutionary relationships between the chromosomal gene, auxiliary plasmid(s), and pathogenesis and between pathogenesis and symbiosis remain to be elucidated. In nitrogen-fixing strains of the order *Rhizobiales*, the nitrogen-fixing genes (*nif*) and nodulation genes (*nod*) are located on plasmids in some species and on a chromosomal replicon in others (see Fig. 1). In contrast to the *nif* and *nod* genes, the *vir* genes and T-DNA have been found exclusively on plasmids. This raises interesting questions regarding the physiological significance of the location of genes within a complex genome.

From a biotechnological viewpoint, *Agrobacterium* pathogenic strains and plasmids are important resources for developing new and useful strains. Sequencing more Ti and Ri plasmids would advance the engineering strategy aimed at removing parts of the T-DNA portion (armless), which is applicable at least to a group of closely related plasmids. Studies on their variability in general in addition to the molecular genetic analysis might help to solve problems of transformation-recalcitrant plants and to widen the range of hosts. Recently, we determined rDNA partial sequences and analyzed their variability among strains in the three pathogenic species (Bautista-Zapanta et al. 2007). It is hoped that these data will be useful both for future bioengineering efforts as well as for the tracking and prevention of *Agrobacterium*-associated plant diseases.

References

- Anand A, Krichevsky A, Schornack S, Lahaye T, Tzfira T et-al. (2007) *Arabidopsis* VIRE2 INTERACTING PROTEIN2 is required for *Agrobacterium* T-DNA integration in plants. *Plant Cell* 19:1695–1708
- Aoyama T, Takanami M, Oka A (1989) Signal structure for transcriptional activation in the upstream regions of virulence genes on the hairy-root-inducing plasmid A4. *Nucleic Acids Res* 17:8711–8725
- Bautista-Zapanta J, Arafat HH, Tanaka K, Samala H, Suzuki K (2007) Variation of 16S-23S internally transcribed spacer sequence and intervening sequence in rDNA among the three major *Agrobacterium* species. *Microbiol Res* (doi:10.1016)
- Cangelosi GA, Ankenbauer RG, Nester EW (1990) Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. *Proc Natl Acad Sci USA* 87:6708–6712
- Christie PJ (2004a) The *Agrobacterium* Ti plasmids. In: Funnell BE, Phillips GJ (eds) *Plasmid biology*, ASM Press, Washington, DC, pp 455–472
- Christie PJ (2004b) Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. *Biochim Biophys Acta* 1694:219–234
- Citovsky V, Zupan J, Warnick D, Zambryski P (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. *Science* 256:1802–1805
- Deng W, Chen L, Peng WT, Liang X, Sekiguchi S et-al. (1999) VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium*. *Mol Microbiol* 31:1795–1807
- Dessaux Y, Petit A, Teme J (1992) Opines in *Agrobacterium* biology. In: Verma DPS (ed) *Molecular signals in plant-microbe communications*, CRC Press, Boca Raton, FL, pp 109–136

- Dunoyer P, Himber C, Voinnet O (2006) Induction, suppression and requirement of RNA silencing pathways in virulent *Agrobacterium tumefaciens* infections. *Nat Genet* 38:258–263
- Fullner KJ, Lara JC, Nester EW (1996) Pilus assembly by *Agrobacterium* T-DNA transfer genes. *Science* 273:1107–1109
- Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the “gene-jockeying” tool. *Microbiol Mol Biol Rev* 67:16–37
- González V, Santamaría RI, Bustos P, Hernández-González I, Medrano-Soto A et-al. (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci USA* 103:3834–3839
- Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M et-al. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* 294:2323–2328
- Hansen G, Tempé J, Brevet J (1992) A T-DNA transfer stimulator sequence in the vicinity of the right border of pRi8196. *Plant Mol Biol* 20:113–122
- Hattori Y, Iwata K, Suzuki K, Uraji M, Ohta N et-al. (2001) Sequence characterization of the *vir* region of a nopaline type Ti plasmid, pTi-SAKURA. *Genes Genet Syst* 76:121–130
- Hattori Y, Uraji M, Suzuki K, Ohta N, Iwata K et-al. (2000) Gene list on a plant tumor-inducing plasmid, pTi-SAKURA in *Agrobacterium tumefaciens* MAFF301001. *DNA Res* 7:371–380
- Hodges LD, Cuperus J, Ream W (2004) *Agrobacterium rhizogenes* GALLS protein substitutes for *Agrobacterium tumefaciens* single-stranded DNA-binding protein VirE2. *J Bacteriol* 186:3065–3077
- Hodges LD, Vergunst AC, Neal-McKinney J, den Dulk-Ras A, Moyer DM et-al. (2006) *Agrobacterium rhizogenes* GALLS protein contains domains for ATP binding, nuclear localization, and type IV secretion. *J Bacteriol* 188:8222–8230
- Hooykaas PJJ, den Dulk-Ras H, Ooms G, Schilperoort RA (1980) Interactions between octopine and nopaline plasmids in *Agrobacterium tumefaciens*. *J Bacteriol* 143:1295–1306
- Hooykaas, PJJ, den Dulk-Ras H, Schilperoort RA (1988) The *Agrobacterium tumefaciens* T-DNA gene *6b* is an *onc* gene. *Plant Mol Biol* 11:791–794
- Hooykaas PJJ, Klapwijk PM, Nuti MP, Schilperoort RA, Rorsch A (1977) Transfer of the *Agrobacterium tumefaciens* Ti plasmid to avirulent *Agrobacteria* and to *Rhizobium ex planta*. *J Gen Microbiol* 98:477–484
- Howard EA, Zupan JR, Citovsky V, Zambryski PC (1992) The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell* 68:109–118
- Kado CI (2002) Negative transcriptional regulation of virulence and oncogenes of the Ti plasmid by bearing a conserved C2H2-zinc finger motif. *Plasmid* 48:179–185
- Kalogeraki VS, Zhu J, Eberhard A, Madsen EL, Winans SC (1999) The phenolic *vir* gene inducer ferulic acid is O-demethylated by the VirH2 protein of an *Agrobacterium tumefaciens* Ti plasmid. *Mol Microbiol* 34:512–522
- Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T et-al. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* 7:331–338
- Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T et-al. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 9:189–197
- Kitakura S, Fujita T, Ueno Y, Terakura S, Wabiko H, Machida Y (2002) The protein encoded by oncogene *6b* from *Agrobacterium tumefaciens* interacts with a nuclear protein of tobacco. *Plant Cell* 14:451–463
- Klein DT, Klein RM (1953) Transmissibility of tumor-inducing ability to avirulent crown-gall and related bacteria. *J Bacteriol* 66:220–228
- Lacroix B, Tzfira T, Vainstein A, Citovsky V (2006) A case of promiscuity: *Agrobacterium*'s endless hunt for new partners. *Trends Genet* 22:29–37
- Li PL, Farrand SK (2000) The replicator of the nopaline-type Ti plasmid pTiC58 is a member of the *repABC* family and is influenced by the TraR-dependent quorum-sensing regulatory system. *J Bacteriol* 182:179–188

- Mankin SL, Hill DS, Olhoft PM, Toren E, Wenck AR et-al. (2007) Disarming and sequencing of *Agrobacterium rhizogenes* strain K599 (NCPBP2659) plasmid pRi2659. *In Vitro Cell Dev Biol Plant* 43:521–535
- McCullen CA, Binns AN (2006) *Agrobacterium tumefaciens* and plant cell interactions and activities required for interkingdom macromolecular transfer. *Annu Rev Cell Dev Biol* 22:101–127
- Moriguchi K, Maeda Y, Satou M, Handayani NSN, Kataoka M et-al. (2001) The complete nucleotide sequence of a plant root-inducing (Ri) plasmid indicates its chimeric structure and evolutionary relationship between tumor-inducing (Ti) and symbiotic (Sym) plasmids in *Rhizobiaceae*. *J Mol Biol* 307:771–784
- Moriguchi K, Maeda Y, Satou M, Kataoka M, Tanaka N, Yoshida K (2000) Analysis of unique variable region of a plant root inducing plasmid, pRi1724, by the construction of its physical map and library. *DNA Res* 7:157–163
- Moriuchi H, Okamoto C, Nishihama R, Yamashita I, Machida Y, Tanaka N (2004) Nuclear localization and interaction of RolB with plant 14–3–3 proteins correlates with induction of adventitious roots by the oncogene *rolB*. *Plant J* 38:260–275
- Nair GR, Liu Z, Binns AN (2003) Reexamining the role of the accessory plasmid pAtC58 in the virulence of *Agrobacterium tumefaciens* strain C58. *Plant Physiol* 133:989–999
- Pan SQ, Jin S, Boulton MI, Hawes M, Gordon MP, Nester EW (1995) An *Agrobacterium* virulence factor encoded by a Ti plasmid gene or a chromosomal gene is required for T-DNA transfer into plants. *Mol Microbiol* 17:259–269
- Paulus F, Otten L (1993) Functional and mutated agrocinopine synthase genes on octopine T-DNAs. *Mol Plant Microbe Interact* 6:393–402
- Peralta EG, Hellmiss R, Ream W (1986) Overdrive, a T-DNA transmission enhancer on the *A. tumefaciens* tumour-inducing plasmid. *EMBO J* 5:1137–1142
- Roitsch T, Wang H, Jin SG, Nester EW (1990) Mutational analysis of the VirG protein, a transcriptional activator of *Agrobacterium tumefaciens* virulence genes. *J Bacteriol* 172:6054–6060
- Scheiffelle P, Pansegrau W, Lanka E (1995) Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of superhelical T-border DNA *in vitro*. *J Biol Chem* 270:1269–1276
- Schrammeijer B, Hemelaar J, Hooykaas PJJ (1998) The presence and characterization of a *virF* gene on *Agrobacterium vitis* Ti plasmids. *Mol Plant Microb Interact* 11:429–433
- Schröder G, Lanka E (2005) The mating pair formation system of conjugative plasmids – A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* 54:1–25
- Shimoda N, Toyoda-Yamamoto A, Aoki S, Machida Y (1993) Genetic evidence for an interaction between the VirA sensor protein and the ChvE sugar-binding protein of *Agrobacterium*. *J Biol Chem* 268:26552–26558
- Stachel SE, Zambryski PC (1986) *virA* and *virG* control the plant-induced activation of the T-DNA transfer process of *A. tumefaciens*. *Cell* 46:325–333
- Sundberg C, Meek L, Carroll K, Das A, Ream W (1996) VirE1 protein mediates export of the single-stranded DNA-binding protein VirE2 from *Agrobacterium tumefaciens* into plant cells. *J Bacteriol* 178:1207–1212
- Suzuki K, Hattori Y, Uraji M, Ohta N, Iwata K et-al. (2000) Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. *Gene* 242:331–336
- Suzuki K, Iwata K, Yoshida K (2001) Genome analysis of *Agrobacterium tumefaciens*: construction of physical maps for linear and circular chromosomal DNAs, determination of copy number ratio and mapping of chromosomal virulence genes. *DNA Res* 8:141–152
- Suzuki K, Uraji M, De Costa D, Hattori Y, Ohta N et-al. (2004) An overview of the agrobacterial genome. *Endocytobiosis Cell Res* 15:143–150
- Segedi E, Czako M, Otten L (1996) Further evidence that vitopine-type pTis of *Agrobacterium vitis* represent a novel group of Ti plasmids. *Mol Plant-Microbe Interact* 9:139–143
- Tabata S, Hooykaas PJ, Oka A (1989) Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3. *J Bacteriol* 171:1665–1672

- Tanaka K, Urbanczyk H, Matsui H, Sawada H, Suzuki K (2006) Construction of physical map and mapping of chromosomal virulence genes of the biovar 3 *Agrobacterium* (*Rhizobium vitis*) strain K-Ag-1. *Genes Genet Syst* 81:373–380
- Terakura S, Ueno Y, Tagami H, Kitakura S, Machida C et-al. (2007) An oncoprotein from the plant pathogen *Agrobacterium* has histone chaperone-like activity. *Plant Cell* 19:2855–2865
- Toro N, Datta A, Carmi OA, Young C, Prusti RK, Nester EW (1989) The *Agrobacterium tumefaciens* *virC1* gene product binds to overdrive, a T-DNA transfer enhancer. *J Bacteriol* 171:6845–6849
- Tzfira T, Vaidya M, Citovsky V (2001) VIP1, an *Arabidopsis* protein that interacts with *Agrobacterium* VirE2, is involved in VirE2 nuclear import and *Agrobacterium* infectivity. *EMBO J* 20:3596–3607
- Tzfira T, Vaidya M, Citovsky V (2004) Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. *Nature* 431:87–92
- Uraji M, Suzuki K, Yoshida K (2002) A novel plasmid curing method using incompatibility of plant pathogenic Ti plasmids in *Agrobacterium tumefaciens*. *Genes Genet Syst* 77:1–9
- Urbanczyk H, Suzuki K, Yoshida K, Kondo K (2003) Physical and gene maps of *Agrobacterium* biovar 2 strains and their relationship to biovar 1 chromosomes. *Microbiology* 149:3035–3042
- Velázquez E, Peix A, Zurdo-Piñeiro JL, Palomo JL, Mateos PF et-al. (2005) The coexistence of symbiosis and pathogenicity-determining genes in *Rhizobium rhizogenes* strains enables them to induce nodules and tumors or hairy roots in plants. *Mol Plant Microbe Interact* 18:1325–1332
- van Haaren MJ, Sedee NJ, Schilperoort RA, Hooykaas PJ (1987) Overdrive is a T-region transfer enhancer which stimulates T-strand production in *Agrobacterium tumefaciens*. *Nucleic Acids Res* 15:8983–8997
- Wabiko H, Minemura M (1996) Exogenous phytohormone-independent growth and regeneration of tobacco plants transgenic for the *6b* gene of *Agrobacterium tumefaciens* AKE10. *Plant Physiol* 112:939–951
- White CE, Winans SC (2007) Cell-cell communication in the plant pathogen *Agrobacterium tumefaciens*. *Philos Trans R Soc Lond B Biol Sci* 362:1135–1148
- Wirawan IG, Kang HW, Kojima M (1993) Isolation and characterization of a new chromosomal virulence gene of *Agrobacterium tumefaciens*. *J Bacteriol* 175:3208–3212
- Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP et-al. (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* 294:2317–2323
- Yadav NS, Vanderlayden J, Bennett DR, Barnes WM, Chilton MD (1982) Short direct repeats flank the T-DNA on a nopaline Ti plasmid. *Proc Natl Acad Sci USA* 79:6322–6326
- Yamamoto S, Uraji M, Tanaka K, Moriguchi K, Suzuki K (2007) Identification of pTi-SAKURA DNA region conferring enhancement of plasmid incompatibility and stability. *Genes Genet Syst* 82:197–206
- Yoshida K, Uraji M, Hattori Y, Moriguchi K, Suzuki K et-al. (2003) The first complete sequencing analyses of plant tumor-inducing plasmid Ti and root-inducing plasmid Ri indicate their chimeric structures and unique evolutionary relationships. *Recent Res Devel Plant Cell Physiol* 1:53–61
- Yoshida K, Uraji M, Hattori Y, Moriguchi K, Suzuki K et-al. (2004) Genome structure and evolution of giant plant pathogenic plasmids in *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes*. *Endocytobiosis Cell Res* 15:371–378
- Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H (2003) Classification and nomenclature of *Agrobacterium* and *Rhizobium*. *Int J Syst Evol Microbiol* 53:1689–1695
- Zambryski P, Tempe J, Schell J (1989) Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. *Cell* 56:193–201
- Zhu JP, Oger PM, Schrammeijer B, Hooykaas PJJ, Farrand SK, Winans SC (2000) The basis of crown gall tumorigenesis. *J Bacteriol* 182:3885–3895

The Virulence Plasmids of *Shigella flexneri*

Charles J. Dorman (✉)

Contents

1	<i>Shigella</i> and Bacillary Dysentery.....	152
2	The <i>Shigella</i> Virulence Plasmid.....	153
3	The Entry Region and <i>Shigella</i> Virulence.....	155
4	The Virulence Gene Regulatory Cascade.....	156
5	Virulence Plasmid Structural Instability.....	161
6	Plasmid Replication.....	161
7	Transfer Genes.....	162
8	Postsegregational Killing System.....	162
9	Plasmid Partitioning Systems.....	163
	References.....	164

Abstract *Shigella* species and enteroinvasive *Escherichia coli* possess a large virulence plasmid that carries the genes necessary for invasion and colonization of the epithelial cell layer of the human gut resulting in dysentery. The plasmid is a hybrid molecule with a complex evolutionary history. The genes required for expression of the type III secretion system and many of the effector proteins involved in the early stage of invasion are grouped within a segment of A + T-rich DNA known as the entry region. Other A + T-rich genes coding for additional effectors and regulatory proteins are scattered around the plasmid. There is a central role for the H-NS nucleoid-associated protein in repressing these genes and their expression is activated via a complicated regulatory cascade involving at least three plasmid-located genes (*virF*, *virB* and *mxiE*) and a multitude of chromosomally located regulatory loci. One of the regulatory proteins, VirB, is closely related to the family of ParB-like plasmid partitioning factors. The virulence plasmid possesses two complete plasmid partitioning systems (ParA–ParB–*parS* and StbA–StbB) and two toxin-antidote postsegregational killing systems (CcdA–CcdB and MvpT–MvpA). The modern virulence plasmid is incapable of self-transmission via

C.J. Dorman

Department of Microbiology, School of Genetics and Microbiology, Moyne Institute of Preventive Medicine, University of Dublin, Trinity College, Dublin 2, Ireland
e-mail: cjorman@tcd.ie

conjugation but it retains vestiges of a now-defunct plasmid transfer system that is related to that of the F plasmid.

1 *Shigella* and Bacillary Dysentery

Bacillary dysentery, or shigellosis, is a disease of humans in which the colonic epithelium is invaded by bacteria and subjected to inflammatory destruction (Kotloff et al. 1999; Parsot 2005; Phalipon and Sansonetti 2007; Wei et al. 2003). The bacteria responsible are Gram-negative facultative intracellular pathogens of the genus *Shigella*. Similar symptoms can be caused by enteroinvasive *Escherichia coli* (EIEC). There are four sub-groups of *Shigella*. These are *S. boydii*, *S. dysenteriae*, *S. flexneri*, and *S. sonnei*. *S. dysenteriae* type 1 expresses a potent cytotoxin and causes epidemics, whereas *S. flexneri* and *S. sonnei* are responsible for endemic shigellosis (Niyogi 2005; Parsot 2005; Phalipon and Sansonetti 2007; Sansonetti 2006). EIEC and the *Shigella* species are regarded as a single pathovar of *E. coli* (Lan et al. 2004).

Bacillary dysentery is a highly infectious disease that can be spread directly via the faecal–oral route by infected individuals or indirectly through contaminated food or water (DuPont et al. 1989). In severe cases, dysentery involves fever accompanied by abdominal cramps and bloody diarrhea. These symptoms arise as a result of the destruction of the colonic and rectal mucosa following invasion of the cells of the epithelium by the bacteria and the activation of an inflammatory immune response in the host (LaBrec et al. 1964; Philpott et al. 2000). Bacillary dysentery is predominantly a disease of the poor and is associated with poor sanitation and hygiene. Approximately 160 million cases are reported worldwide annually, with children under the age of 5 being particularly at risk; there are up to one million deaths from the disease each year (Kotloff et al. 1999). In developing countries the chief causative agents are *S. dysenteriae* type 1 and *S. flexneri*. However, it can also occur in developed countries. Here, outbreaks are frequently due to infection by *S. sonnei*, and, once again, young children are particularly at risk (Niyogi 2005).

Shigella cannot invade the cells of the colonic epithelium through their apical surfaces, so it uses the antigen-sampling M cells as a means of gaining access to the basolateral surfaces of its target cells (Sansonetti and Phalipon 1999; Wassef et al. 1989) (Fig. 1). Macrophage stationed below the epithelial cell layer engulf the bacteria, but *Shigella* escapes from the phagocytic vacuole and triggers apoptosis, killing the macrophage (Fernandez-Prada et al. 2000; Chen et al. 1996; Zychlinsky et al. 1994). The dying macrophage release cytokines that recruit polymorphonuclear cells to sites of infection (Navarre and Zychlinsky 2000; Perdomo et al. 1994a, b; Zychlinsky and Sansonetti 1997). This has the effect of destabilizing the epithelial layer and granting access to the basolateral layer to even more bacteria.

Shigella uses a type III secretion system (TTSS) to gain entry to the non-phagocytic epithelial cells (Allaoui et al. 1993b; Andrews and Maurelli 1992; Blocker et al. 1999, 2001; Parsot et al. 1995; Sasakawa et al. 1993; Venkatesan

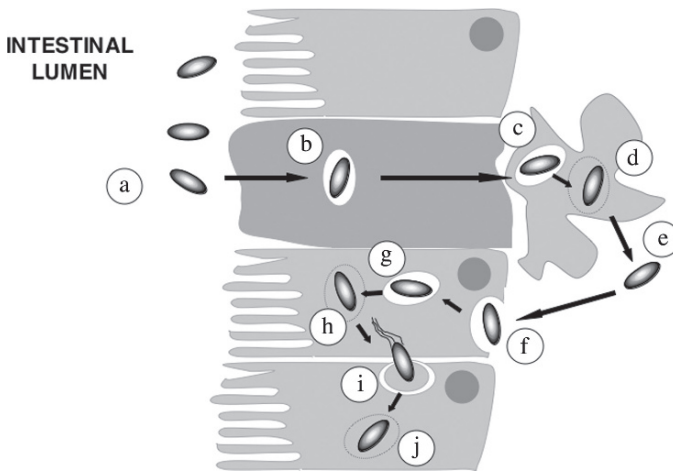


Fig. 1 Summary of the invasive process. Bacteria move from the intestinal lumen (a) into M cells (b), allowing them to traverse the epithelial layer where they encounter macrophage (c). The *Shigellae* induce apoptosis in the macrophage (d) and gain access (e) to the basolateral pole of intestinal epithelial cells. Invasion (f) is facilitated by the type III secretion system (TTSS), and bacteria appear in vacuoles in the cytoplasm (g). The vacuoles are lysed (h), and the liberated bacteria recruit host actin which is polymerized to provide a mechanism of propulsion that sends the microbes through the cell membrane into host adjacent cells (i). The bacteria then escape the double membrane of the vacuole (j) leaving them free to repeat the process, with concomitant damage to the cells of the epithelial layer. The release of cytokines by damaged macrophage and epithelial cells induces and inflammatory response that further weakens the integrity of the epithelial barrier, facilitating further invasion by bacteria

et al. 1992). Effector proteins injected via the TTSS subvert signaling pathways in the host cells in ways that lead to the engulfment of the bacterium (Allaoui et al. 1993a; Baudry et al. 1988; Ménard et al. 1994a; Niebuhr et al. 2000; Page et al. 1999; 2001; Sasakawa et al. 1989; Venkatesan and Buysse 1990). Further manipulation of the host by bacterially encoded factors permits cell-to-cell spreading within the colonic epithelium (Bernardini et al. 1989; Goldberg 2001; Prévost et al. 1992; Sansonetti et al. 1994; Vasselon et al. 1992). The infected epithelial cells release more cytokines that attract more host defence cells, exacerbating the inflammation at the point of infection (Jung et al. 1995; Sansonetti et al. 1999). *Shigella* even has the ability to control the level of the host immune response to the infection in ways that benefit the pathogen and damage the host (Phalipon and Sansonetti 2007).

2 The *Shigella* Virulence Plasmid

It was discovered several decades ago that many of the essential genes required for *Shigella* virulence are located on a large plasmid (Hale et al. 1983; Sansonetti et al. 1982, 1983; Sasakawa et al. 1986). Since those original observations were published, the virulence genes, their products, and the regulatory mechanisms that

control their expression have all been the subject of intensive research. There have also been investigations of the plasmid replication systems and its mechanisms for controlling daughter plasmid segregation at cell division.

Virulence plasmids have been sequenced from a number of *Shigella* strain and data from the sequencing projects have contributed useful information to studies of *Shigella* evolution (Buchrieser et al. 2000; Jiang et al. 2005; Jin et al. 2002; Venkatesan et al. 2001; Wei et al. 2003) (Table 1). Analysis of the DNA sequences of the virulence plasmids pINV-2457T (serotype 2a strain 2457T), pWR100 and pWR501 (serotype 5 strain M90T), and pCP301 (serotype 2a strain 301) shows that these four are largely identical, apart from some minor differences in insertion sequence (IS) elements (which make up approximately 25% of the plasmids) (Wei et al. 2003). Plasmid pWR501 is slightly larger than pWR100 (Table 1) because of the presence of a Tn501 insertion (Venkatesan et al. 2001). The virulence plasmid of *S. sonnei* is similar in size and structure (Jiang et al. 2005). The number and variety of insertion sequences seen in *Shigella* virulence plasmids are among the highest that have been described for any bacterial plasmids (Buchrieser et al. 2000; Jiang et al. 2005; Venkatesan et al. 2001). The IS elements are believed to have made important contributions to the evolution of the modern virulence plasmid with its mosaic structure. They are also likely to be a cause of structural instability within the plasmid, either due to transposition of the elements or their participation in homologous recombination (Yang et al. 2007).

The evolution of the virulence plasmid and the relationship between *Shigella* and *E. coli* have been studied in considerable detail. Phylogenetic studies using DNA sequence analysis have shown that EIEC and *Shigella* are a single pathovar of *E. coli* (Lan et al. 2004). Although some studies have suggested that the virulence plasmid entered a single *E. coli* ancestor once to generate the modern *Shigella*/EIEC lineage (Escobar-Paramo et al. 2003), most evidence indicates that *Shigella*/EIEC has arisen due to the transfer of diverse ancestral virulence plasmids to *E. coli* on several occasions over a long period of time and that this has produced the modern *Shigella* species with their observed genome diversity combined with similar pathogenic traits (Yang et al. 2007). Phylogenetic analyses of the virulence plasmid have shown that it occurs in at least two forms designated pINV A and pINV B (Lan et al. 2001, 2003).

Table 1 *Shigella* virulence plasmids discussed in this chapter

Plasmid	Source serotype and strain	Size (bp)	Accession number	Reference
pINV-2457T	Serotype 2a strain 2457T	~218,000	Not available ^a	Wei et al. (2003)
pWR100	Serotype 5a strain M90T	213,494	AL391753	Buchrieser et al. (2000)
pWR501	Serotype 5a strain M90T	221,851	AF348706	Venkatesan et al. (2001)
pCP301	Serotype 2a strain 301	221,618	AF386526	Jin et al. (2002)
pSS	<i>S. sonnei</i> ^b	214,396	CP000038	Jiang et al. (2005)

^aRefer to URL (<http://www.genome.wisc.edu/sequencing/sflex.htm>) for updates on the progress with the sequencing and annotation of this plasmid

^bAll *S. sonnei* form a single serotype; strain unknown

A third and distinct pINV form is found in *S. dysenteriae* type 1 (Lan et al. 2003). The pINV A and pINV B forms are incompatible, an observation that is consistent with an early report that the *Shigella* virulence plasmids belong to two incompatibility groups (Makino et al. 1988). This incompatibility is not a function of differences in origins of replication; both plasmid forms belong to the IncFII incompatibility group (see Sect. 6) and have identical *copA* sequences, the determinant of IncFII replication-based incompatibility. Instead, the basis of pINV A/B incompatibility appears to lie in the MvpT/MvpA toxin/antidote postsegregational killing mechanism (see Sect. 8). Small differences between the systems may result in an antidote-toxin mismatch, which causes incompatibility (Lan et al. 2003).

3 The Entry Region and *Shigella* Virulence

The entry region contains the genes that encode the TTSS and many of the major effector proteins required by *Shigella* to cause invasive disease (Buchrieser et al. 2000; Dorman 2004a). The region is 31 kb pairs in length and contains 37 open reading frames (Fig. 2). Most of the genes are organized within operons and this facilitates coregulation at the level of transcription; the locations of the main promoters have been determined. The DNA of the entry region has an unusually high A + T content, suggesting an origin outside the *Enterobacteriaceae* (Buchrieser et al. 2000; Jiang et al. 2005; Venkatesan et al. 2001). All of the genes in the segment have an A + T content in the range 65–70%, which is in stark contrast to the average for the *Shigella* chromosome of 49.1% (Wei et al. 2003). There are vestigial transposable elements at the ends of the entry region (Buchrieser et al. 2000), and their presence is consistent with the view that the sequence has been acquired by a horizontal transfer mechanism. The structural components of the TTSS are encoded by the *mxi* and *spa* genes. An exception is the *mxiE* gene, which codes for a regulatory protein (Kane et al. 2002; Mavris et al. 2002a, b). The *ipa* and *ipg* genes code for important effector proteins that are injected into the host cell by the TTSS (Baharani et al. 1997; Ménard et al. 1994b; Watarai et al. 1995). Expression of the virulence genes is controlled by environmental signals. Optimal conditions for their transcription in bacteria growing in the laboratory include a temperature of 37°C (i.e. human body temperature), neutral pH (possibly a signal that the bacteria have successfully traversed the acid environment of the stomach), and osmotic pressure equivalent to that found in the human gut (Maurelli et al. 1984; Porter and Dorman 1994; Nakayama and Watanabe 1995, 1998). Although the proteins to be secreted are expressed in bacteria growing in broth, they are stored and not released in large amounts until there is physical contact with the host cells. This physical contact enhances the activity of the TTSS, which feeds back positively onto the expression of the plasmid-located virulence genes (Parsot 2005). Once the bacteria have entered the host cell, the main virulence genes on the plasmid become repressed (Lucchini et al. 2005). The intracellular signals and the molecular mechanisms responsible for this repression have not yet been characterized.

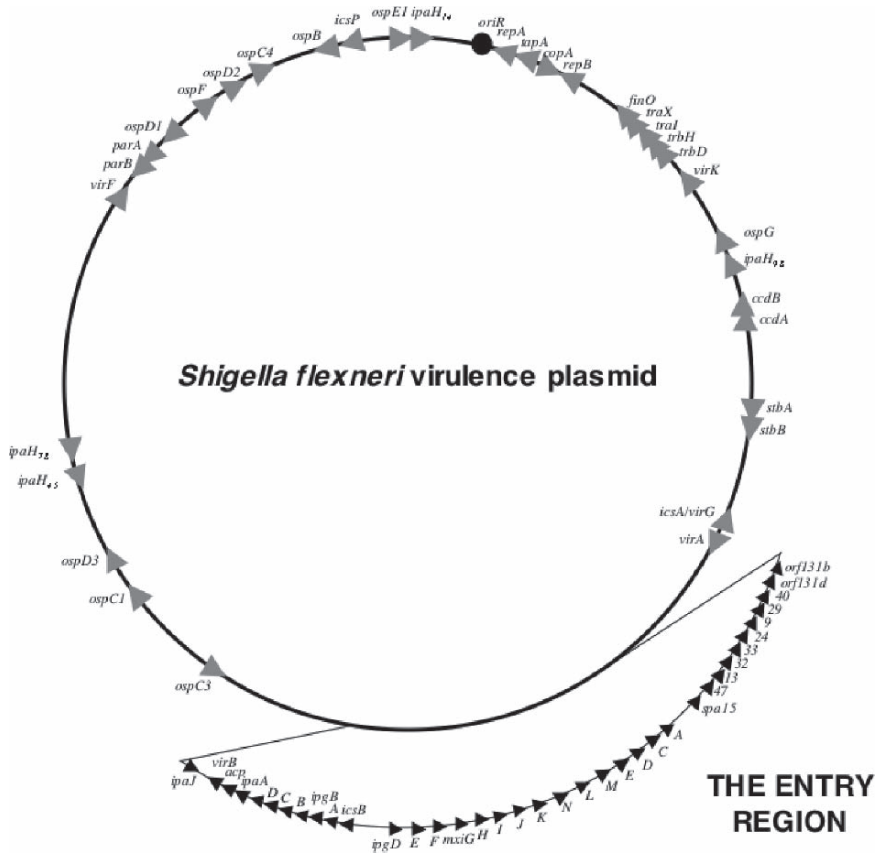


Fig. 2 Genetic map of the large virulence plasmid. Individual genes are represented by arrowheads and the origin of plasmid replication is represented by a filled circle at *oriR*. The entry region is shown in an expanded format and its operon structure should be considered in association with the more detailed diagram in Fig. 3. Virulence genes within the entry region are represented by black arrowheads while genes located elsewhere on the plasmid are shown in grey. The diagram is not to scale

4 The Virulence Gene Regulatory Cascade

The chromosomally encoded H-NS nucleoid-associated protein has been identified as a repressor of every promoter in the *S. flexneri* virulence gene regulatory cascade where its involvement has been assessed (Beloin and Dorman 2003). This DNA binding protein is a global regulator of transcription in Gram-negative bacteria, usually acting as a repressor (Atlung and Ingmer 1997; Dorman 2004, 2007a; Dorman et al. 1999; Tendeng and Bertin 2003). H-NS binds to A + T-rich DNA and it has been described as having a preference for DNA sequences with intrinsic curvature (Rimsky et al. 2001). Recent research has identified an A + T-rich consensus

sequence, 5'-TCGATATATT-3', for H-NS binding (Bouffartigues et al. 2007). It is possible that H-NS binds to this site, or to closely related sequences, and then polymerizes along the DNA to render it transcriptionally silent. The H-NS protein consists of an amino-terminal oligomerization domain and a carboxy-terminal nucleic acid binding domain, with a flexible linker connecting the two (Bertin et al. 1999; Dorman 2004b, 2007a; Dorman et al. 1999; Esposito et al. 2002; Rimsky 2004; Stella et al. 2005). The protein is active as a dimer and data from single-molecule and bulk experiments indicate that it is capable of cross-linking different parts of the same DNA molecule or segments of separate DNA molecules through the formation of DNA–H-NS–DNA bridges (Dame 2005; Dame et al. 2006; Dorman 2007b) (Fig. 3). Mutants deficient in the H-NS protein display upregulated expression of the virulence gene regulon in bacteria grown under nonpermissive conditions (Dorman et al. 1990; Hromockyj et al. 1992; Maurelli and Sansonetti 1988).

The entry region, with the high A + T content in its DNA and its flanking sequences derived from IS elements, resembles many pathogenicity islands (Dobrindt et al. 2004; Hacker et al. 2004). It differs from these in not being associated with a tRNA gene and in being located on a plasmid rather than the chromosome. The involvement of the H-NS protein in repressing transcription of the virulence genes in the entry region is another feature that is shared with other blocks of A + T-rich genes that have been acquired horizontally in *Salmonella enterica*, including the genes of its major pathogenicity islands (Dorman 2007a; Lucchini et al. 2006; Navarre et al. 2006). It has been argued that the ability of H-NS to silence imported genes in this way may win time for the bacterium to evolve mechanisms for safely expressing them (Dorman 2007b; Navarre et al. 2006). If this is so, the regulatory methods that have emerged in *S. flexneri* to oppose H-NS are indeed elaborate.

Genetic analysis has identified a regulatory cascade in which an AraC-like transcription factor called VirF activates transcription of a second regulatory gene (*virB*) whose product in turn upregulates the promoters of the key structural virulence genes and operons on the plasmid (Adler et al. 1989; Dorman and Porter 1998; Dorman et al. 2001). Transcription of the virulence genes occurs at 37°C but is repressed at lower temperatures (Maurelli et al. 1984) and thermoregulation is exerted at both the *virF* and the *virB* regulatory genes (Dorman 2004a; Dorman and Porter 1998; Dorman et al. 2001) (Fig. 3). Osmolarity also plays a role, with osmotic pressure equivalent to that found in the lower human gut being required for optimal transcription (Porter and Dorman 1994). The osmosensing two-component regulatory system EnvZ/OmpR has been implicated in the regulation of the virulence genes and expression of the virulent phenotype but details of the molecular mechanism are lacking (Bernardini et al. 1990). The system is also sensitive to pH: a pH value of 7.4 is required for full upregulation of virulence gene transcription (Nakayama and Watanabe 1995, 1998). The pH signal is sensed by the CpxA/CpxR two-component regulatory system and transmitted by the CpxR DNA binding protein to the *virF* regulatory gene (Nakayama and Watanabe 1998).

The *virF* gene is located outside the entry region and its promoter is repressed by the H-NS nucleoid-associated protein through a mechanism that involves the

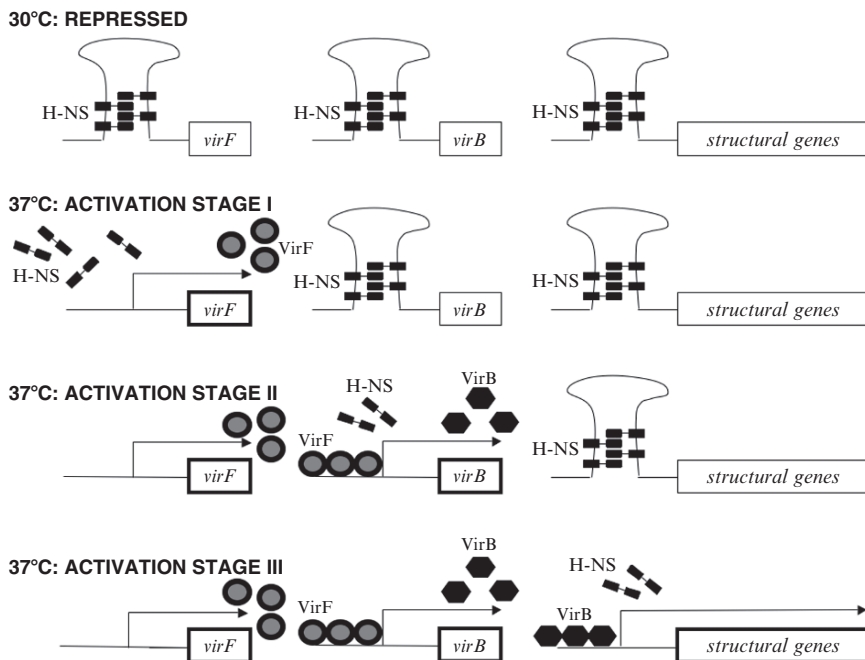


Fig. 3 Activation of the virulence genes. At 30°C, the regulatory and structural virulence genes are repressed by the H-NS protein. This dimeric protein forms repression complexes by cross-linking DNA at target promoters. Following a shift to the permissive temperature (37°C), the repression complex at the promoter of the *virF* regulatory gene is disrupted by temperature-induced changes in the conformation of the DNA (Stage I). The *virF* regulatory protein then activates the promoter of the *virB* gene, partly by displacing H-NS (Stage II). The *virB* protein then acts to derepress the promoters of the major structural gene operons by dislodging H-NS (Stage III), leading to expression of the genes coding for the type III secretion system and its associated effector proteins

formation of a DNA–H-NS–DNA bridged structure (Falconi et al. 1998; Prosseda et al. 2004) (Fig. 3). Repression is relieved when the temperature increases to 37°C. The rising temperature alters DNA curvature within the A + T-rich H-NS-bound DNA sequence and this is thought to dislodge the repressor. The *virF* promoter is also regulated by Fis, another nucleoid-associated protein. In this case, the regulatory input is positive (Falconi et al. 2001). Evidence from negative dominance genetic experiments indicates that the VirF protein forms dimers, a property it shares with many AraC-like proteins (Porter and Dorman 2002). Translation of *virF* mRNA is modulated by the tRNA modification genes *miaA* and *tgt* genes and this may link virulence gene expression to iron metabolism in the bacterium (Durand et al. 2000, 2003). Further evidence of a link to iron metabolism comes from the observation that the iron-responsive small RNA, RyhB, controls the expression of the VirB regulatory protein, and hence gene expression in the virulence regulon (Murphy and Payne 2007). Expression of RyhB is itself controlled by the iron master regulator protein, Fur (Oglesby et al. 2005).

The VirF protein is required to activate transcription of *virB*, another gene that is repressed by H-NS (Tobe et al. 1993) (Figs. 3 and 4). The activation mechanism includes a requirement for DNA negative supercoiling at the *virB* promoter and the manipulation of DNA supercoiling is known to alter expression of the VirB regulon (Dorman et al. 1990; McNairn et al. 1995; Ni Bhriain and Dorman 1993). The manipulation of promoter DNA supercoiling by artificial means can upregulate *virB* transcription even in the absence of a thermal signal, but the gene retains its dependency on the VirF protein (Tobe et al. 1995). Removing H-NS from the cell does not abrogate the dependency of *virB* on VirF for transcription; this shows that VirF plays an active role in the upregulation mechanism and does not act simply to remove the repressor (Tobe et al. 1995).

The VirB protein positively regulates the structural genes in the virulence cascade (Figs. 3 and 4). These genes are repressed by the H-NS protein, and VirB probably upregulates them by displacement of the repressor. At the time of writing, the mechanism has only been studied in the case of the *icsB* promoter of *S. flexneri* 2a strain 2457T. Here, VirB appears to form a nucleoprotein complex upstream of the promoter through a mechanism that displaces H-NS (Turner and Dorman 2007). The DNA sequence that is bound by VirB at *icsB* resembles the *parS* elements to which plasmid partition proteins bind. In addition, VirB shows amino acid sequence homology to ParB-like plasmid-partitioning proteins (Adler et al. 1989; Beloin et al. 2002). This is also true of the InvE protein in *S. sonnei*, a homolog of VirB, which also binds to a DNA element with homology to *parS*-like sequences (Taniya et al. 2003).

The *mxiE* gene is located within one of the operons responsible for expression of the Mxi-Spa TTSS (Fig. 4). It encodes an AraC-like transcription factor. Mutants harboring knockout mutations in *mxiE* are avirulent (Kane et al. 2002). MxiE controls the expression of at least 11 genes on the virulence plasmid that lie outside the entry region. The protein also regulates at least one gene on the *S. flexneri* chromosome; this locus is related to the *ipaH* loci that are found at various positions on the virulence plasmid (Kane et al. 2002; Mavris et al. 2002b). IpaH proteins contain leucine-rich repeats that are also features of Toll-like receptors in mammalian cells and NB/LRR proteins in plants, suggesting that IpaH proteins allow the bacterium to modulate the immune response of the host during the infection (Aderem and Ulevitch 2000; Kjemtrup et al. 2000). Consistent with this is the observation that the IpaH_{7,8} protein encoded by the *S. flexneri* virulence plasmid is required for escape by the bacterium from macrophage vacuoles (Fernandez-Prada et al. 2000).

MxiE is active only when it forms a complex with the IpgC cofactor protein. However, this cofactor may be unavailable owing to the formation of alternative independent complexes with the IpaB and IpaC effector proteins while these are still in the bacterial cell (Ménard et al. 1994b). The free form of IpgC becomes available when IpaB and IpaC secretion begins and this is a signal for transcription activation by MxiE of its regulon of secreted protein genes (Mavris et al. 2002a). MxiE-dependent promoters possess a so-called MxiE box with the consensus sequence 5'-GTATCGTTTTTTTAnAG-3' located between positions -33 and -49

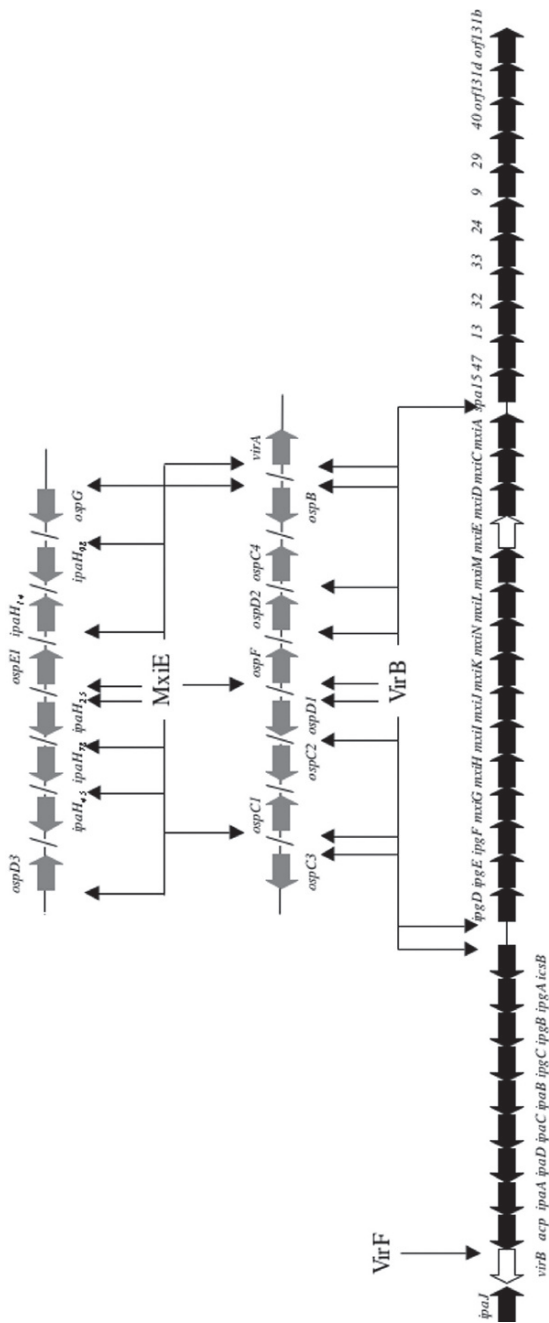


Fig. 4 Regulation of the virulence genes on the large plasmid. The virF arac-like transcription factor activates the promoter of the *virB* gene within the entry region. This regulatory gene is represented by a white horizontal arrow. The structural virulence genes of the entry region are represented by horizontal black arrows, and those virulence genes that are located elsewhere on the plasmid are shown in grey. The entry region contains three major operators, each of which is activated by the virB regulatory protein. Positive regulatory inputs are indicated by vertical arrows that point to the target promoters. VirB upregulates at least nine genes outside the entry region. The slash marks between the grey genes indicate that these are not necessarily adjacent to one another on the plasmid (see Fig. 2 for a complete illustration of the relative locations of the plasmid genes). The *mxiE* regulator is encoded by a gene within the entry region (shown in white) and it positively controls the transcription of at least 12 genes, 4 of which are also under the control of *virB*. The diagram is not to scale

with respect to the known or presumed transcription start sites of eight promoters (Mavris et al. 2002b). It regulates some genes that are also activated by VirB (Le Gall et al. 2005; Parsot 2005) (Fig. 2); the significance of this dual regulation is not clear. The *mxiE* gene also plays a regulatory role in *cis* because it is the site of a transcriptional slippage event that modulates transcription and translation of its downstream neighbor (*mxiD*) in the virulence operon (Penno and Parsot 2006) (Fig. 4). Interestingly, the *mxiE* gene itself relies for expression on transcriptional slippage between two overlapping reading frames, MxiEa and MxiEb (Penno et al. 2005). Transcriptional slippage is also a feature that modulates the expression of several of the structural virulence genes on the plasmid (Penno et al. 2006).

5 Virulence Plasmid Structural Instability

It has been shown that the virulence plasmids of EIEC and *S. flexneri* can integrate into the bacterial chromosome at a specific site within the *metB* gene. Integration leads to methionine auxotrophy, presumably due to inactivation of the *metB* gene, and to general silencing of plasmid-associated virulence gene expression with associated loss of a virulent phenotype (Colonna et al. 1995; Zagaglia et al. 1991). Normal expression of the virulence genes, and restoration of methionine prototrophy, follow RecA-dependent excision of the plasmid (Zagaglia et al. 1991). The gene silencing effect seems to be focused on the *virB* regulator. Its expression is silenced, whereas that of the upstream *virF* gene is normal. Silencing can be overcome by ectopic expression of *virB* in trans or by inactivation of the *hns* gene (Colonna et al. 1995). It is tempting to speculate that the *virB* locus within the integrated plasmid has been placed in an H-NS-DNA nucleoprotein complex analogous to those associated with horizontally acquired virulence genes within the pathogenicity islands of *Salmonella enterica* serovar Typhimurium (Navarre et al. 2006; Lucchini et al. 2006).

Inactivation of the *virB* and *virF* genes has been reported to stabilize the structure of the virulence plasmid, as though expression *per se* is a destabilizing influence (Schuch and Maurelli 1997). These data come from experiments in which the bacteria were grown under laboratory conditions that upregulate the virulence genes. The plasmid was either lost completely or physically rearranged through either RecA-dependent recombination or insertion sequence transposition (Mills et al. 1992; Schuch and Maurelli 1997). Perhaps, silencing of the virulence genes following integration into the chromosome might provide a mechanism for the preservation of the virulence plasmid while the bacterium is growing in isolation from the host.

6 Plasmid Replication

The large virulence plasmid has a replication system that is closely related to that of plasmid R100, which is a member of incompatibility group IncFII; there is also similarity between some virulence plasmid replication functions and those of plasmid

R1, another member of IncFII (Jiang et al. 2005; Makino et al. 1988; Nordstrom 2006; Silva et al. 1988; Venkatesan et al. 2001). Plasmid R100 was isolated originally from a strain of *S. flexneri* 2a (Buchrieser et al. 2000; Nakaya et al. 1960; Venkatesan et al. 2001), whereas plasmid R1 was isolated originally from *Salmonella* (Datta and Kontomichalou 1965). The replication functions of the virulence plasmid match those of R100 (or R1) in all important respects and include an origin, *oriR*, that is bound by the RepA replication initiation protein (Fig. 2). Plasmid copy number is governed (a) by CopA antisense RNA regulation acting to inhibit translation of the TapA leader peptide, thus preventing translation-coupling to, and hence translation of, the RepA protein and (b) through transcriptional repression of the *repA* gene by the RepB repressor protein that is encoded by a gene located downstream and on the opposite DNA strand from *copA* (Buchrieser et al. 2000; Nordstrom 2006; Venkatesan et al. 2001) (Fig. 2). A DnaA box is located downstream of the *repA* gene in the space before *oriR*, in common with other IncFII systems (Buchrieser et al. 2000) (Fig. 2). As might be expected, given the similarity of the replication systems, the large virulence plasmid is incompatible with plasmids from incompatibility group IncFII (Silva et al. 1988).

7 Transfer Genes

Early work indicated that the virulence plasmid is incapable of self-transmission by conjugation (Makino et al. 1988; Sansonetti et al. 1982). DNA sequence analysis has revealed an incomplete set of full-length and truncated transfer genes, in keeping with the earlier functional studies. These genes show a high level of DNA sequence homology with some of the *tra* genes of the F plasmid (Buchrieser et al. 2000) and are located in the sector of the plasmid between the *virK* virulence gene and the origin of replication, *oriR* (Fig. 2). The same situation has been reported for each of the *S. flexneri* and *S. sonnei* virulence plasmids that have been sequenced; interestingly, *S. sonnei* strain Ss046 also has the F *tra* sequences on its chromosome (Yang et al. 2007).

8 Postsegregational Killing System

The *S. flexneri* virulence plasmid has at least two toxin–antidote systems that are related to well-characterized examples of postsegregational killing systems. The first of these shows strong DNA sequence similarity to the *ccdA* and *ccdB* system from the F plasmid (Afif et al. 2001; Jaffe et al. 1985) and is located upstream of the *ipaH*_{9,8} virulence gene (Fig. 2). The well-characterized CcdAB system from *E. coli* is known to target and inhibit DNA gyrase, thus blocking DNA replication in the plasmid-free segregant (Van Melderen 2002).

The second toxin–antidote system consists of the *mvpT* and *mvpA* genes that encode a toxin and antidote, respectively (maintenance of virulence plasmid, Toxin and Antidote). The *mvpA* and *mvpT* genes were originally known as STBORF1 and

STBORF2 and are located within and on the complementary DNA strand to the *trbH* gene in the defunct plasmid transfer region (Fig. 2) (Radnedge et al. 1997; Sayeed et al. 2000).

The virulence plasmid is more likely to be physically lost by the bacterium during growth at 37°C, the permissive temperature for expression of the invasive phenotype. It seems that at this temperature, the normal plasmid maintenance systems are overridden, leading to accelerated rates of plasmid loss. This is not seen at lower temperatures such as 30°C (Sayeed et al. 2005). Plasmid loss rates of 2–3% per generation have been reported in cultures growing at 37°C; the *mvp* toxin–antidote system then kills the plasmid-cured cells. It has been pointed out that such heavy reliance on a toxin–antidote system to control plasmid loss represents an unusual strategy among bacterial plasmids due to the heavy metabolic burden involved (Sayeed et al. 2005).

9 Plasmid Partitioning Systems

The *S. flexneri* plasmid has two functional active partitioning systems (something that is in keeping with its mosaic structure) located on opposite sides of the plasmid (Fig. 2). One of these, *parAB* is related to the partition system of phage/plasmid P1 and consists of the genes *parA* and *parB* and the *cis*-acting site *parS* (Sergueev et al. 2005). The relationship is sufficiently close for the P1 prophage and the pWR100 plasmid to be incompatible (Sergueev et al. 2005). The *cis*-acting *parS* motif is bound by the ParB protein and by the integration host factor, IHF, a nucleoid-associated protein that binds and bends DNA through angles of up to 180° (Rice et al. 1996). Interestingly, IHF is also involved in modulating the expression of the virulence genes within the entry region (Porter and Dorman 1997).

The second plasmid partitioning system is encoded by the *stbA* and *stbB* genes whose products have significant amino-acid sequence identity to the StbA and StbB proteins of plasmid R100; DNA sequence analysis of the pWR100 virulence plasmid also identified a possible *cis*-acting site associated with the StbA/StbB system (Buchrieser et al. 2000).

It should be recalled that the virulence plasmid expresses a protein, VirB (InvE in *S. sonnei*), that is similar to the ParB proteins encoded by the P1/P7 prophage, and that the DNA sequence to which VirB/InvE binds is similar to the *parS* sequences in the prophage partitioning systems (Adler et al. 1989; Taniya et al. 2003; Turner and Dorman 2007). There is no evidence for a ParA-like partner for VirB, nor is there evidence that VirB plays any direct role in plasmid partitioning in the modern virulence plasmid. Instead, VirB seems to have been co-opted to play a role as a positive regulator of virulence structural gene promoters through the displacement of the H-NS repressor. This points to the evolvability of gene regulatory circuits and the possibility that a DNA-binding protein can be diverted from one function (e.g. plasmid partitioning) to another (e.g. virulence gene regulation). Such possibilities illustrate the complexities of biological systems and their abilities

to exploit their essentially modular nature to develop effective solutions to the problems that arise as a result of continuous genome shuffling.

Acknowledgements I thank Niamh Ní Bhriain for a critical reading of the manuscript. Research in my laboratory is supported by grants from Science Foundation Ireland, the Wellcome Trust, and the Health Research Board.

References

- Aderem A, Ulevitch RJ (2000) Toll-like receptors in the induction of the innate immune response. *Nature* 406:782–787
- Adler B, Sasakawa C, Tobe T, Makino S, Komatsu K, Yoshikawa M (1989) A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. *Mol Microbiol* 3:627–635
- Afif H, Allali N, Couturier M, Van Melderen L (2001) The ratio between CcdA and CcdB modulates the transcriptional repression of the *ccd* poison-antidote system. *Mol Microbiol* 41:73–82
- Allaoui A, Ménard R, Sansonetti PJ, Parsot C (1993a) Characterization of the *Shigella flexneri* *ipgD* and *ipgF* genes, which are located in the proximal part of the *mxi* locus. *Infect Immun* 61:1707–1714
- Allaoui A, Sansonetti PJ, Parsot C (1993b) MxiD, an outer membrane protein necessary for the secretion of the *Shigella flexneri* Ipa invasins. *Mol Microbiol* 7:59–68
- Andrews GP, Maurelli AT (1992) *mxiA* of *Shigella flexneri* 2a, which facilitates export of invasion plasmid antigens, encodes a homolog of the low-calcium-response protein, LcrD, of *Yersinia pestis*. *Infect Immun* 60:3287–3295
- Atlung T, Ingmer H (1997) H-NS: a modulator of environmentally regulated gene expression. *Mol Microbiol* 24:7–17
- Bahrani FK, Sansonetti PJ, Parsot C (1997) Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect Immun* 65:4005–4010
- Baudry B, Kaczorek M, Sansonetti PJ (1988) Nucleotide sequence of the invasion plasmid antigen B and C genes (*ipaB* and *ipaC*) of *Shigella flexneri*. *Microb Pathog* 4:345–357
- Beloin C, Dorman CJ (2003) An extended role for the nucleoid structuring protein H-NS in the virulence gene regulatory cascade of *Shigella flexneri*. *Mol Microbiol* 47:825–838
- Beloin C, McKenna S, Dorman CJ (2002) Molecular dissection of VirB, a key regulator of the virulence cascade of *Shigella flexneri*. *J Biol Chem* 277:15333–15344
- Bernardini ML, Mounier J, d’Hauteville H, Coquis-Rondon M, Sansonetti PJ (1989) Identification of *icsA*, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc Natl Acad Sci USA* 86:3867–3871
- Bernardini ML, Fontaine A, Sansonetti PJ (1990) The two-component regulatory system *ompR-envZ* controls the virulence of *Shigella flexneri*. *J Bacteriol* 172:6274–6281
- Bertin P, Benhables N, Krin E, Laurent-Winter C, Tendeng C, Turlin E, Thomas A, Danchin A, vasseur R (1999) The structural and functional organization of H-NS-like proteins is evolutionarily conserved in Gram-negative bacteria. *Mol Microbiol* 31: 319–329
- Blocker A, Gounon P, Larquet E, Niebuhr K, Cabiliaux V, Parsot C, Sansonetti P (1999) The tripartite type III secretion of *Shigella flexneri* inserts IpaB and IpaC into host membranes. *J Cell Biol* 147:683–693
- Blocker A, Jouihri N, Larquet E, Gounon P, Ebel F, Parsot C, Sansonetti PJ, Allaoui A (2001) Structure and composition of the *Shigella flexneri* “needle complex”, a part of its type III secretion. *Mol Microbiol* 39:652–663
- Bouffartiges E, Buckle M, Badaut C, Travers A, Rimsky S (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nat Struct Mol Biol* 14:441–448

- Buchrieser C, Glaser P, Rusniok C, Nedjari H, D'Hauteville H, Kunst F, Sansonetti P, Parsot C (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. *Mol Microbiol* 38:760–771
- Chen Y, Smith MR, Thirumalai K, Zychlinsky A (1996) A bacterial invasin induces macrophage apoptosis by directly binding ICE. *EMBO J* 15:3853–3860
- Colonna B, Casalino M, Fradiani PA, Zagaglia C, Naitza S, Leoni L, Prosseda G, Coppo A, Ghelardini P, Nicoletti M (1995) H-NS regulation of virulence gene expression in enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome. *J Bacteriol* 177:4703–4712
- Dame RT (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol Microbiol* 56:858–870
- Dame RT, Noom MC, Wuite GJ (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* 444:387–390
- Datta N, Kontomichalou P (1965) Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. *Nature* 208:239–241
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2:414–424
- Dorman CJ (2004a) Virulence gene regulation in *Shigella*. In: Curtiss III R, Ingraham JL, Kaper JB, Maloy S, Neidhardt FC, Riley MM, Squires CJ, Wanner BL, Bock A (eds) *Escherichia coli* and *Salmonella*: cellular and molecular biology, 3rd edn., *EcoSal* online. Posted 29th December, 2004 at <http://www.ecosal.org>. American Society for Microbiology, Washington DC
- Dorman CJ (2004b) H-NS: a universal regulator for a dynamic genome. *Nature Rev Microbiol* 2:391–400
- Dorman CJ (2007a) H-NS, the genome sentinel. *Nat Rev Microbiol* 5:157–161
- Dorman CJ (2007b) Probing bacterial nucleoid structure with optical tweezers. *BioEssays* 29:212–216
- Dorman CJ, Porter ME (1998) The *Shigella* virulence gene regulatory cascade, a paradigm of bacterial gene control mechanisms. *Mol Microbiol* 29:677–684
- Dorman CJ, Ní Bhriain N, Higgins CF (1990) DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*. *Nature* 344:789–792
- Dorman CJ, Hinton JCD, Free A (1999) Homo- and hetero-oligomerization among H-NS-like nucleoid-associated proteins in bacteria. *Trends Microbiol* 7:124–128
- Dorman CJ, McKenna S, Beloin C (2001) Regulation of virulence gene expression in *Shigella flexneri*, a facultative intracellular pathogen. *Int J Med Microbiol* 291:89–96
- DuPont HL, Levine MM, Hornick RB, Formal SB (1989) Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis* 159:1126–1128
- Durand JMB, Björk GR (2003) Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of *Shigella flexneri*: a possible role in adaptation of virulence. *Mol Microbiol* 47:519–527
- Durand JMB, Dagberg B, Uhlin BE, Björk GR (2000) Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of *Shigella flexneri*: the expression of the *virF* gene. *Mol Microbiol* 35:924–935
- Escobar-Paramo P, Giudicelli C, Parsot C, Denamur E (2003) The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J Mol Evol* 57:140–148
- Esposito D, Petrovic A, Harris R, Ono S, Eccleston JF, Mbabaali A, Haq I, Higgins CF, Hinton JC, Driscoll PC, Ladbury JE (2002) H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein. *J Mol Biol* 324:841–850
- Falconi M, Colonna B, Prosseda G, Micheli G, Gualerzi CO (1998) Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. *EMBO J* 17:7033–7043
- Falconi M, Prosseda G, Giangrossi M, Beghetto E, Colonna B (2001) Involvement of FIS in the H-NS-mediated regulation of *virF* gene of *Shigella* and enteroinvasive *Escherichia coli*. *Mol Microbiol* 42:439–452

- Fernandez-Prada CM, Hoover DL, Tall BD, Hartman AB, Kopelowitz J, Venkatesan MM (2000) *Shigella flexneri* IpaH 7.8 facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. *Infect Immun* 68:3608–3619
- Goldberg MB (2001) Actin-based motility of intracellular bacterial pathogens. *Microbiol Mol Biol Rev* 65:595–626
- Hacker J, Hochhut B, Middendorf B, Schneider G, Buchrieser C, Gottschalk G, Dobrindt U (2004) Pathogenomics of mobile genetic elements of toxigenic bacteria. *Int J Med Microbiol* 293:453–461
- Hale TL, Sansonetti PJ, Schad PA, Austin S, Formal SB (1983) Characterization of virulence plasmids and plasmid-associated outer membrane proteins in *Shigella flexneri*, *Shigella sonnei*, and *Escherichia coli*. *Infect Immun* 40:340–350
- Hromockyj AE, Tucker SC, Maurelli AT (1992) Temperature regulation of *Shigella* virulence: identification of the repressor gene *virR*, an analogue of *hns*, and partial complementation by tyrosyl transfer RNA (tRNA^{Tyr}). *Mol Microbiol* 6:2113–2124
- Jaffe A, Ogura T, Hiraga S (1985) Effects of the *ccd* function of the F plasmid on growth. *J Bacteriol* 163:841–849
- Jiang Y, Yang F, Zhang X, Yang J, Chen L, Yan Y, Nie H, Xiong Z, Wang J, Dong J, Xue Y, Xu X, Zhu Y, Chen S, Jin Q (2005) The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*. *Plasmid* 54:149–159
- Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J (2002) Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res* 30:4432–4441
- Jung HC, Eckmann L, Yang SK, Panja A, Fierer J, Morzycka-Wroblewska E, Kagnoff MF (1995) A distinct array of proinflammatory cytokines is expressed in human epithelial cells in response to bacterial invasion. *J Clin Invest* 95:55–62
- Kane CD, Schuch R, Day WA Jr, Maurelli AT (2002) MxiE regulates intracellular expression of factors secreted by the *Shigella flexneri* 2a type III secretion system. *J Bacteriol* 184:4409–4419
- Kjemtrup S, Nimchuk Z, Dangel JL (2000) Effector proteins of phytopathogenic bacteria: bifunctional signals in virulence and host recognition. *Curr Opin Microbiol* 3:73–78
- Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Sverdlow DL, Sansonetti PJ, Adak GK, Levine MM (1999) Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ* 77:651–666
- LaBrec EH, Schneider H, Mangani TJ, Formal SB (1964) Epithelial cell penetration as an essential step in the pathogenesis of bacillary dysentery. *J Bacteriol* 88:1503–1518
- Lan R, Lumb B, Ryan D, Reeves PR (2001) Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect. Immun.* 69:6303–6309
- Lan R, Stevenson G, Reeves PR (2003) Comparison of two major forms of the *Shigella* virulence plasmid pINV: positive selection is a major force driving the divergence. *Infect Immun* 71:6298–6306
- Lan R, Alles MC, Donohue K, Martinez MB, Reeves PR (2004) Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect Immun* 72:5080–5088
- Le Gall T, Mavris M, Martino MC, Bernardini ML, Denamur E, Parsot C (2005) Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of *Shigella flexneri*. *Microbiol* 151:951–962
- Lucchini S, Liu H, Jin Q, Hinton JC, Yu J (2005) Transcriptional adaptation of *Shigella flexneri* during infection of macrophages and epithelial cells: insights into the strategies of a cytosolic bacterial pathogen. *Infect Immun* 73:88–102
- Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JCD (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog* 2:748–752
- Makino S, Sasakawa C, Yoshikawa M (1988) Genetic relatedness of the basic replicon of the virulence plasmid in shigellae and enteroinvasive *Escherichia coli*. *Microb Pathog* 5:267–274

- Maurelli AT, Sansonetti PJ (1988) Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. Proc Natl Acad Sci USA 85:2820–2824
- Maurelli AT, Blackmon B, Curtiss RIII (1984) Temperature-dependent expression of virulence genes in *Shigella* species. Infect Immun 43:195–201
- Mavris M, Page AL, Tournebize R, Demers B, Sansonetti PJ, Parsot C (2002a) Regulation of transcription by the activity of the *Shigella flexneri* type III secretion apparatus. Mol Microbiol 43:1543–1553
- Mavris M, Sansonetti PJ, Parsot C (2002b) Identification of the cis-acting site involved in activation of promoters regulated by activity of the type III secretion apparatus in *Shigella flexneri*. J Bacteriol 184:6751–6759
- McNairn E, Ní Bhriain N, Dorman CJ (1995) Over-expression of the *Shigella flexneri* genes coding for DNA topoisomerase IV compensates for loss of DNA topoisomerase I: Effect on virulence gene expression. Mol Microbiol 15:507–517
- Ménard R, Sansonetti P, Parsot C (1994a) The secretion of the *Shigella flexneri* Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. EMBO J 13:5293–5302
- Ménard R, Sansonetti PJ, Parsot C, Vasselon T (1994b) Extracellular association and cytoplasmic partitioning of the IpaB and IpaC invasins of *S. flexneri*. Cell 79:515–525
- Mills JA, Venkatesan MM, Baron LS, Buysse JM (1992) Spontaneous insertion of an *IS1*-like element into the *virF* gene is responsible for avirulence in opaque colonial variants of *Shigella flexneri* 2a. Infect Immun 60:175–182
- Murphy ER, Payne SM (2007) RyhB, an iron-responsive small RNA molecule, regulates *Shigella dysenteriae* virulence. Infect Immun 75:3470–3477
- Nakaya R, Nakamura A, Murata Y (1960) Resistance transfer agents in *Shigella*. Biochem Biophys Res Commun 3:654–659
- Nakayama S, Watanabe H (1995) Involvement of *cpxA*, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of *Shigella sonnei virF* gene. J Bacteriol 177:5062–5069
- Nakayama S, Watanabe H (1998) Identification of *cpxR* as a positive regulator essential for expression of the *Shigella flexneri virF* gene. J Bacteriol 180:3522–3528
- Navarre WW, Zychlinsky A (2000) Pathogen-induced apoptosis of macrophages: a common end for different pathogenic strategies. Cell Microbiol 2:265–273
- Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella enterica* Sv. Typhimurium. Science 313:236–238
- Ni Bhriain N, Dorman CJ (1993) Identification and characterization of a *topA* mutation in *Shigella flexneri*. Mol Microbiol 7:351–358
- Niebuhr K, Jouihri N, Allaoui A, Gounon P, Sansonetti PJ, Parsot C (2000) IpgD, a protein secreted by the type III secretion machinery of *Shigella flexneri*, is chaperoned by IpgE and implicated in entry focus formation. Mol Microbiol 38:8–19
- Niyogi SK (2005) Shigellosis. J Microbiol 43:133–143
- Nordstrom K (2006) Plasmid R1 – replication and its control. Plasmid 55:1–26
- Oglesby AG, Murphy ER, Iyer VR, Payne SM (2005) Fur regulates acid resistance in *Shigella flexneri* via RyhB and *ydeP*. Mol Microbiol 58:1354–1367
- Page AL, Ohayon H, Sansonetti PJ, Parsot C (1999) The secreted IpaB and IpaC invasins and their cytoplasmic chaperone IpgC are required for intercellular dissemination of *Shigella flexneri*. Cell Microbiol 1:183–193
- Page AL, Fromont-Racine M, Sansonetti PJ, Legrain P, Parsot C (2001) Characterization of the interaction partners of secreted proteins and chaperones of *Shigella flexneri*. Mol Microbiol 42:1133–1145
- Parsot C (2005) *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. FEMS Microbiol Lett 252:11–18
- Parsot C, Ménard R, Gounon P, Sansonetti PJ (1995) Enhanced secretion through the *Shigella flexneri* Mxi-Spa translocon leads to assembly of extracellular proteins into macromolecular structures. Mol Microbiol 16:291–300

- Penno C, Parsot C (2006) Transcriptional slippage in *mxiE* controls transcription and translation of the downstream *mxiD* gene, which encodes a component of the *Shigella flexneri* type III secretion apparatus. *J Bacteriol* 188:1196–1198
- Penno C, Sansonetti P, Parsot C (2005) Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in *Shigella flexneri*. *Mol Microbiol* 56:204–214
- Penno C, Hachani A, Biskri L, Sansonetti P, Allaoui A, Parsot C (2006) Transcriptional slippage controls production of type III secretion apparatus components in *Shigella flexneri*. *Mol Microbiol* 62:1460–1468
- Perdomo JJ, Gounon P, Sansonetti PJ (1994a) Polymorphonuclear leukocyte transmigration promotes invasion of colonic epithelial monolayer by *Shigella flexneri*. *J Clin Invest* 93:633–643
- Perdomo OJ, Cavaillon JM, Huerre M, Ohayon H, Gounon P, Sansonetti PJ (1994b) Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. *J Exp Med* 180:1307–1319
- Phalipon A, Sansonetti PJ (2007) *Shigella's* ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? *Immunol Cell Biol* 85:119–129
- Philpott DJ, Edgeworth JD, Sansonetti PJ (2000) The pathogenesis of *Shigella flexneri* infection: lessons from in vitro and in vivo studies. *Philos Trans R Soc Lond B Biol Sci* 355:575–586
- Porter ME, Dorman CJ (1994) A role for H-NS in the thermo-osmotic regulation of virulence gene expression in *Shigella flexneri*. *J Bacteriol* 176:4187–4191
- Porter ME, Dorman CJ (1997) Positive regulation of *Shigella flexneri* virulence genes by integration host factor. *J Bacteriol* 179:6537–6550
- Porter ME, Dorman CJ (2002) In vivo DNA-binding and oligomerization properties of the *Shigella flexneri* AraC-like transcriptional regulator VirF as identified by random and site-specific mutagenesis. *J Bacteriol* 184:531–539
- Prévost MC, Lesourd M, Arpin M, Vernel F, HELLIO R, Sansonetti PJ (1992) Unipolar reorganization of F-actin layer at bacterial division and bundling of actin filaments by plastin correlate with movement of *Shigella flexneri* within HeLa cells. *Infect Immun* 60:4088–4099
- Prosseda G, Falconi M, Giangrossi M, Gualerzi CO, Micheli G, Colonna B (2004) The *virF* promoter in *Shigella*: more than just a curved DNA stretch. *Mol Microbiol* 51:523–537
- Radnedge L, David MA, Youngren B, Austin SJ (1997) Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*. *J Bacteriol* 179:3670–3675
- Rice PA, Yang S, Mizuuchi K, Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87:1295–1306
- Rimsky S (2004) Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Curr Opin Microbiol* 7:109–114
- Rimsky S, Zuber F, Buckle M, Buc H (2001) A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol* 42:1311–1323
- Sansonetti PJ (2006) Shigellosis: an old disease in new clothes? *PLoS Med* 3:e354
- Sansonetti PJ, Phalipon A (1999) M cells as ports of entry for enteroinvasive pathogens: mechanisms of interaction, consequences for the disease process. *Semin Immunol* 11:193–203
- Sansonetti PJ, Arondel J, Huerre M, Harada A, Matsushima K (1999) Interleukin-8 controls bacterial transepithelial translocation at the cost of epithelial destruction in experimental shigellosis. *Infect Immun* 67:1471–1480
- Sansonetti PJ, Kopecko DJ, Formal SB (1982) Involvement of a plasmid in the invasive ability of *Shigella flexneri*. *Infect Immun* 35:852–860
- Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Collins HH Jr, Formal SB (1983) Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. *Infect Immun* 39:1392–1402
- Sansonetti PJ, Mounier J, Prévost MC, Mege RM (1994) Cadherin expression is required for the spread of *Shigella flexneri* between epithelial cells. *Cell* 5:829–839
- Sasakawa C, Kamata K, Sakai T, Murayama SY, Makino S, Yoshikawa M (1986) Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*. *Infect Immun* 51:470–475

- Sasakawa C, Adler B, Tobe T, Okada N, Nagai S, Komatsu K, Yoshikawa M (1989) Functional organization and nucleotide sequence of virulence Region-2 on the large virulence plasmid in *Shigella flexneri* 2a. *Mol Microbiol* 3:1191–1201
- Sasakawa C, Komatsu K, Tobe T, Suzuki T, Yoshikawa M (1993) Eight genes in region 5 that form an operon are essential for invasion of epithelial cells by *Shigella flexneri* 2a. *J Bacteriol* 175:2334–2346
- Sayed S, Reaves L, Radnedge L, Austin SJ (2000) The stability region of the large virulence plasmid of *Shigella flexneri* encodes an efficient postsegregational killing system. *J Bacteriol* 182:2416–2421
- Sayed S, Brendler T, Davis M, Reaves L, Austin S (2005) Surprising dependence on postsegregational killing of host cells for maintenance of the large virulence plasmid in *Shigella flexneri*. *J Bacteriol* 187:2768–2773
- Schuch R, Maurelli AT (1997) Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression. *Infect Immun* 65:3686–3692
- Sergueev K, Dabrazhynetskaya A, Austin S (2005) Plasmid partition system of the P1_{par} family from the pWR1000 virulence plasmid of *Shigella flexneri*. *J Bacteriol* 187:3369–3373
- Silva RM, Saadi S, Maas WK (1988) A basic replicon of virulence-associated plasmids of *Shigella* spp. and enteroinvasive *Escherichia coli* is homologous with a basic replicon in plasmids of IncF groups. *Infect Immun* 56:836–842
- Stella S, Spurio R, Falconi M, Pon CL, Gualerzi CO (2005) Nature and mechanism of the in vivo oligomerization of nucleoid protein H-NS. *EMBO J* 24:2896–2905
- Taniya T, Mirobe J, Nakayama S-I, Mingshan Q, Okuda K, Watanabe H (2003) Determination of the InvE binding site required for expression of IpaB of the *Shigella sonnei* virulence plasmid: involvement of a ParB BoxA-like sequence. *J Bacteriol* 185:5158–5165
- Tendeng C, Bertin PN (2003) H-NS in Gram-negative bacteria: a family of multifaceted proteins. *Trends Microbiol* 11:511–518
- Tobe T, Yoshikawa M, Mizuno T, Sasakawa C (1993) Transcriptional control of the invasion regulatory gene *virB* of *Shigella flexneri*: activation by VirF and repression by H-NS. *J Bacteriol* 175:6142–6149
- Tobe T, Yoshikawa M, Sasakawa C (1995) Thermoregulation of *virB* transcription in *Shigella flexneri* by sensing of changes in local DNA superhelicity. *J Bacteriol* 177:1094–1097
- Turner EC, Dorman CJ (2007) H-NS antagonism in *Shigella flexneri* by VirB, a virulence gene transcription regulator that is closely related to plasmid partition factors. *J Bacteriol* 189:3403–3413
- Van Melderen L (2002) Molecular interactions of the CcdB poison with its bacterial target, the DNA gyrase. *Int J Med Microbiol* 291:537–544
- Vasselon T, Mounier J, Hellio R, Sansonetti PJ (1992) Movement along actin filaments of the perijunctional area and de novo polymerization of cellular actin are required for *Shigella flexneri* colonization of epithelial Caco-2 cell monolayers. *Infect Immun* 60:1031–1040
- Venkatesan MM, Buysse JM (1990) Nucleotide sequence of invasion plasmid antigen gene *ipaA* from *Shigella flexneri* 5. *Nucleic Acids Res* 18:1648
- Venkatesan MM, Buysse JM, Oaks EV (1992) Surface presentation of *Shigella flexneri* invasion plasmid antigens requires the products of the *spa* locus. *J Bacteriol* 174:1990–2001
- Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR (2001) Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infect Immun* 69:3271–3285
- Wassef JS, Keren DF, Mailloux JL (1989) Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of shigellosis. *Infect Immun* 57:858–863
- Watarai M, Tobe T, Yoshikawa M, Sasakawa C (1995) Contact of *Shigella* with host cells triggers release of Ipa invasins and is an essential function of invasiveness. *EMBO J* 14:2461–2470
- Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, Plunkett G III, Rose DJ, Darling A, Mau B, Perna NT, Payne SM, Runyen-Janecky LJ, Zhou S, Schwartz DC, Blattner FR (2003) Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71:2775–2786

- Yang J, Huan N, Chen L, Zhang X, Yang F, Xu X, Zhu Y, Yu J, Jin Q (2007) Revisiting the molecular evolutionary history of *Shigella* spp. *J Mol Evol* 64:71–79
- Zagaglia C, Casalino M, Colonna B, Conti C, Calconi A, Nicoletti M (1991) Virulence plasmids of enteroinvasive *Escherichia coli* and *Shigella flexneri* integrate into a specific site on the host chromosome: integration greatly reduces expression of plasmid-carried virulence genes. *Infect Immun* 59:792–799
- Zychlinsky A, Sansonetti PJ (1997) Apoptosis as a proinflammatory event: what can we learn from bacteria-induced cell death? *Trends Microbiol.* 5:201–204
- Zychlinsky A, Kenny B, Ménard R, Prévost MC, Holland IB, Sansonetti PJ (1994) IpaB mediates macrophage apoptosis induced by *Shigella flexneri*. *Mol Microbiol* 11:619–627

The Three Bears and Virulence-Associated Plasmids in the Genus *Yersinia*

Sarah L. Howard, Philippa C.R. Strong, and Brendan W. Wren(✉)

Contents

1	The Pathogenic <i>Yersiniae</i>	172
1.1	The Plague Bacillus <i>Yersinia pestis</i>	173
1.2	The Enteropathogenic <i>Yersiniae</i> , <i>Y. pseudotuberculosis</i> and <i>Y. enterocolitica</i>	173
2	Mother Bear (pYV).....	174
2.1	Yops and the Secretion System	175
2.2	<i>Yersinia</i> Adhesin A	176
3	Daddy Bear (pMT1).....	177
3.1	Ymt the Putative Murine Toxin.....	177
3.2	F1 Antigen	178
4	Baby Bear (pPCP1).....	179
4.1	Pla (Plasminogen Activator)	179
4.2	Pesticin and Immunity	180
4.3	Origin and Mobility of the Baby and Daddy Plasmid	180
5	Other Plasmids Found in Pathogenic <i>Yersinia</i>	181
6	Conclusion	182
	References.....	183

Abstract The *Yersiniae* are Gram-negative rods belonging to the family *Enterobacteriaceae* and represent a useful genus to study the evolution of virulence embodied in their plasmid DNA content (Wren 2003). *Yersinia* species range from the benign to the highly virulent, exemplified by the plague bacillus *Yersinia pestis* whose effects during the Black Death wiped out a third of the European population and helped to shape Western civilization. Acquisition of plasmids is central to the development of virulence in the *Yersiniae*. All three human pathogenic species including *Y. pestis* and the enteropathogenic *Yersiniae*, *Y. pseudotuberculosis*, and *Y. enterocolitica*, harbor a 70-kb virulence plasmid (pYV), which is essential for infection in lymph tissues as well as to overcome host defense mechanisms. Additionally, *Y. pestis* has two further plasmids, a large (100-kb pMT1) and a small plasmid (9.6-kb pPla), which are essential for transmission via the flea vector (summarized in [Table 1](#));

S.L. Howard, P.C.R. Strong, and B.W. Wren
Department of Infectious and Tropical Diseases, London School of Hygiene
and Tropical Medicine, London, UK
e-mail: brendan.wren@lshtm.ac.uk

Table 1 Plasmids important in the virulence of pathogenic *Yersinia*e

Plasmid	Other designations	Size (kb)
<i>Yersinia</i> virulence plasmid, pYV	Mother plasmid pCD, pCD1, pIB1, pLCR, pYVe8081, pYVa127/90, pYVe227	68–72
Plasmid encoding murine toxin, pMT1	Daddy plasmid pMT, pFra	80–110
Plasminogen activating plasmid, pPCP1	Baby plasmid pPCP, pPst, pPla	9.6

Table 2 Sequenced plasmids from the pathogenic *Yersinia*e

<i>Yersinia</i> strain sequenced	Plasmid	Size (bp)	Accession number
<i>Y. enterocolitica</i> 8081	pYVe8081	67,721	NC_008791
<i>Y. enterocolitica</i> e227	pYVe227	69,673	NC_002120
<i>Y. enterocolitica</i> a127/90	pYVa127/90	66,591	NC_004564
<i>Y. pestis</i> Antiqua	pCD	70,299	NC_008122
<i>Y. pestis</i> Antiqua	pMT	96,471	NC_008120
<i>Y. pestis</i> Antiqua	pPCP	10,777	NC_008121
<i>Y. pestis</i> CO92	pCD	70,305	NC_003131
<i>Y. pestis</i> CO92	pMT1	96,210	NC_003134
<i>Y. pestis</i> CO92	pPCP1	9,612	NC_003132
<i>Y. pestis</i> KIM	pMT1	100,990	NC_004838
<i>Y. pestis</i> Nepal516	pMT1	100,918	NC_008118
<i>Y. pestis</i> Nepal516	pPCP	10,778	NC_008119
<i>Y. pestis</i> Pestoides F	pCD	71,507	NC_009377
<i>Y. pestis</i> Pestoides F	pMT1	137,010	NC_009378
<i>Y. pestis</i> 91001	pCD1	70,159	NC_005813
<i>Y. pestis</i> 91001	pMT1	106,642	NC_005815
<i>Y. pestis</i> 91001	pPCP1	9,609	NC_005816
<i>Y. pseudotuberculosis</i> IP 31758	pYpsIP31758.1	153,140	NC_009705
<i>Y. pseudotuberculosis</i> IP 31758	pYpsIP31758.2	58,679	NC_009704
<i>Y. pseudotuberculosis</i> IP32953	pYptb32953	27,702	NC_006154
<i>Y. pseudotuberculosis</i> IP32953	pYV	68,526	NC_006153

these two plasmids are absent from *Y. pseudotuberculosis* and *Y. enterocolitica*. To date, sequencing has been performed on plasmids isolated from several strains of *Y. pestis* and *Y. pseudotuberculosis* and one strain of *Y. enterocolitica* (Table 2). This chapter will describe the salient features of these plasmids named after the three bears, daddy plasmid, baby plasmid, and the mother of all plasmids, pYV.

1 The Pathogenic *Yersinia*e

The genus *Yersinia* belongs to the family *Enterobacteriaceae* and comprises of 11 species. Of these *Y. pestis*, *Y. pseudotuberculosis*, and *Y. enterocolitica* are the only species pathogenic to humans.

1.1 The Plague *Bacillus Yersinia pestis*

Y. pestis has a complex life cycle involving a mammalian reservoir (primarily rodents) and a flea vector (Perry and Fetherston 1997). The bacterium ensures transmission by forming a cohesive aggregate that blocks the foregut of infected fleas (Bacot and Martin 1914). This results in futile attempts to feed on a new host where a blocked flea regurgitates infected blood into the bite site, effectively injecting the bacteria under the animal's skin (Bacot and Martin 1914). Upon infection, some of the bacilli are engulfed by macrophages and carried to the regional lymph nodes draining the infection site. The bulk of the bacteria penetrate tissue directly protected by plasminogen activator, eventually accessing lymphatic vessels draining regional lymph nodes (Lahteenmaki et al. 1998), where they multiply, giving rise to classical symptoms of infection, followed by bacteremia (Perry and Fetherston 1997). This is known as bubonic plague. If the infection then progresses to the lungs, pneumonic plague develops, which is highly infectious and fatal.

1.2 The Enteropathogenic *Yersiniae*, *Y. pseudotuberculosis* and *Y. enterocolitica*

The enteropathogenic *Yersiniae* are widely found in the environment, for example in soil, and are a common cause of animal infections, affecting a variety of mammalian and avian species (Naktin and Beavis 1999). In humans, the infection causes gastroenteritis, following the consumption of contaminated food or water. After ingestion, the bacteria pass into the small intestine, where they translocate across the intestinal epithelium via Peyer's patches. They then migrate to the mesenteric lymph nodes and are subsequently found in the liver and spleen where they replicate externally to the host cells (Lian et al. 1987). After multiplication, rapid inflammation ensues, giving rise to symptoms associated with gastroenteritis, such as mesenteric lymphadenitis and terminal ileitis.

Y. enterocolitica comprises a biochemically and genetically heterogeneous collection of organisms that has been divided into six biogroups that are differentiated by biochemical tests (1A, 1B, 2, 3, 4, and 5). The six biotypes can be separated into three further groups according to their lethality after intravenous challenge of mice: a nonpathogenic group (biogroup 1A); a mildly pathogenic group that is unable to kill mice (biogroups 2–5), and a highly pathogenic, mouse-lethal group (biogroup 1B). Biogroup 1A lacks the mother plasmid pYV and appears to be more distantly related to the other biogroups, whereas biogroup 1B forms a geographically distinct group of strains frequently isolated in North America (so-called "New World" strains) and biogroups 2–5 are predominantly isolated in Europe and Japan ("Old World" strains). Most commonly, biogroups 1B and 2–5 have been associated with disease in humans and biogroup 1A was thought to be avirulent owing to lack of classical virulence determinants. However, there is growing epidemiological evidence that 1A strains can cause disease in humans. *Y. pseudotuberculosis* is subgrouped into 21

different serological groups on the basis of variation in the O-antigen of its lipopolysaccharide (LPS). All *Y. pestis* strains fail to express an O-antigen.

2 Mother Bear (pYV)

The mother plasmid pYV, also known as pCD1, pIB1, and pLCR, is a ~70-kb plasmid encoding a type III secretion system (T3SS) and proteins known as the *Yersinia* outer membrane proteins (Yops) (Cornelis et al. 1998; Huang et al. 2006) (summarized in Table 3). It is found in all three human pathogenic species of *Yersinia* and plays a role in avoidance of the immune system and in toxicity (Wren 2003). Only the putatively avirulent 1A biotype of *Y. enterocolitica* lacks the mother plasmid. Yops are important virulence determinants in the pathogenesis of the *Yersinae*. The currently known Yops include YopB, YopD, YopO, YopH, YopM, YopT, YopJ, YopE, YopK, YopN, and LcrV (Viboud and Bliska 2005). The *Y. enterocolitica* Yops, YopO, YopP, and YopQ are termed YpkA, YopJ, and YopK, respectively, in *Y. pseudotuberculosis* (Cornelis et al. 1998).

Table 3 Known virulence determinants carried by plasmids pYV, pMT1, and pPCP1

Plasmid	Virulence determinants
pYV	<i>yadA</i> , adhesin <i>yopB</i> , translocator <i>yopD</i> , translocator <i>yopE</i> , effector <i>yopM</i> , effector <i>yopT</i> , effector <i>yopJ</i> , effector <i>ypkA</i> , effector <i>yopH</i> , effector <i>yopK</i> , translocator <i>yopN</i> , secretion regulator <i>lcrV</i> , translocator/effector <i>lcrG</i> , secretion regulator <i>yscF</i> , injectisome needle <i>yscA-E</i> , secretion/structural <i>yscG-Y</i> , secretion/structural <i>syncE,H,T,N,D</i> , chaperones <i>tyeA</i> , secretion regulator <i>virF</i> , transcription regulator
pMT1	<i>ymt</i> , murine toxin <i>cafI</i> , F1 antigen <i>cafIM</i> , usher/chaperone <i>cafIA</i> , anchor/structural <i>cafIR</i> , regulator
pPCP1	<i>pla</i> , plasminogen activator and coagulase <i>pst</i> , pesticin <i>pim</i> , pesticin immunity

2.1 *Yops and the Secretion System*

When incubated at 37°C, the temperature of mammalian host eukaryotic cells, *Yersinia* assemble several Yop secretion (Ysc) injectisomes (syringe-like organelles) at their surface and synthesise a stock of intracellular Yops (Cornelis 2002a). However, the secretion channels of the injectisomes remain closed and there is feedback inhibition to prevent the deleterious accumulation of Yops until close contact with a target host cell. The initial interaction between the host cell and the bacterium is mediated by bacterial outer membrane adhesins and integrins on the surface of the host. This contact enables the injectisome channels to become unblocked and secretion into the host cytosol to commence (Cornelis 2002a). On the mother plasmid is a T3SS, which is encoded by the *ysc* genes and forms a Yop secretion apparatus or Ysc injectisome. The *ysc* genes are located on four loci that have been termed *virA*, *virB*, *virC*, and *virD* (Cornelis et al. 1998). The Ysc injectisome comprises a protein pump basal body structure that spans the bacterial envelope and a needlelike appendage that extends out from the bacterium (Cornelis 2002a). The needle is 600–800 Å long and 60–70 Å wide and is formed by the polymerization of YscF, a 6-kDa monomer, which is secreted by the Ysc apparatus (Cornelis 2002a; Cornelis 2002b). The whole injectisome structure comprises 27 different proteins and is responsible for the translocation of the Yop effectors into the host cell cytosol. However, the proteins must be at least partially unfolded, as the diameter of the needle is too narrow for fully folded globular proteins to enter (Cornelis 2002a).

It is believed that YopB and YopD form a pore within the eukaryotic cell membrane which allows the Yop effectors to pass across the plasma membrane and into the host cytosol (Broms et al. 2007). LcrV, which is at the tip of the injectisome needle, is required for the correct assembly of the translocation pore and forms a tripartite complex with YopB and YopD (Mueller et al. 2005; Aili et al. 2007). Yops B and D have hydrophobic domains which enable them to act as transmembrane proteins (Cornelis 2002a). LcrV (also known as the V antigen) is encoded on the same operon as YopB and YopD but has no particular hydrophobicity (Cornelis 2002b). To date, it is not known how the tip of the injectisome maintains contact with the plasma membrane; one suggestion is that the needle actually passes through the membrane (Viboud and Bliska 2005). In addition to their role in the formation of the pore, LcrV and YopD have regulatory effects on other components of the system (Cornelis 2002b). LcrV is a low-calcium-response *Yersinia* virulence protein associated with the suppression of tumor necrosis factor- α (TNF- α), interferon- γ , and the induction of immunosuppressive IL-10 (Badie et al. 2004; Viboud and Bliska 2005; Heesemann et al. 2006). Secretion of Yops and LcrV appears to be regulated in response to environmental Ca²⁺ by YopN, TyeA (translocation of Yops into eukaryotic cells A), and LcrG (Nilles et al. 1997; Iriarte et al. 1998). Maximal production of Yops is under conditions of low calcium and high temperature (37°C), and millimolar concentrations of Ca²⁺ cause expression to be downregulated and secretion blocked (Nilles et al. 1997). There is, however, a paradox with the calcium response, as the concentration of Ca²⁺ in the extracellular medium is such that Yop

expression should not be induced. Only intracellularly is Ca^{2+} at a low enough concentration, and hence *in vivo* it may be that physical cell contact is the signal that induces Yop secretion rather than Ca^{2+} concentration (Cornelis et al. 1998).

In the host, cytosol effector Yops collectively act to inhibit several innate immune defense mechanisms, such as phagocytosis, production of proinflammatory signaling molecules, and activation of the adaptive immune system (Aili et al. 2007). Yops also play a role in inducing apoptosis (Lemaitre et al. 2006). It is because of the activity of these proteins that *Yersinia* can replicate in the extracellular medium (Aili et al. 2007). Owing to the targeting of the innate immune system, the host has to develop a strong adaptive immune response to clear *Yersinia* infection, with CD4 and CD8 T cells playing an essential role (Heesemann et al. 2006). The secretion of some Yops requires the presence of small cytosolic chaperones called the Syc (specific Yop chaperone) family of proteins. These are small acidic proteins that only bind to specific partner Yops. Yop secretion is severely reduced in their absence (Cornelis 2002a). The precise function of these chaperones has yet to be established.

Six effector Yops have been identified so far; these are YopH, YopE, YopT, YpkA, YopJ, and YopM. YopE, YopH, YopT and YpkA target the regulation of the actin cytoskeleton and in doing so help *Yersiniae* evade phagocytosis by macrophages and polymorphonuclear leukocytes. YopH is a phosphotyrosine phosphatase and its action against macrophages is rapid; within 2 min, YopH-mediated dephosphorylation can be observed. It targets an array of signaling pathways and has a role in inhibiting phagocytosis as well as counteracting the oxidative burst of macrophages and neutrophils (Cornelis 2002a). YopE is a GTPase-activating protein that is involved in antiphagocytotic activity by targeting the regulation of the host actin cytoskeleton (Huang et al. 2006; Aili et al. 2007). It has been suggested that YopE can also form a Yop–lipid complex, which upon host cell contact is released and directly targets the plasma membrane (Isaksson et al. 2006). YopJ (YopP in *Y. enterocolitica*) is a deubiquinating cysteine protease and has been shown to have a role in macrophage apoptosis, as well as disrupting host cell signaling pathways involved in proinflammatory cytokine production (Lemaitre et al. 2006). Studies have demonstrated that YopJ has a dominant role in suppressing T-cell priming in the mouse model (Heesemann et al. 2006). YopM is a strongly acidic protein that migrates to the nucleus of the host cell via a vesicle-associated pathway and is the only Yop effector not to have an enzymatic activity. Although the precise function of YopM is yet to be established, it is thought to be an important *Yersinia* virulence factor, possibly with a role in resistance to the innate immune response (Viboud and Bliska 2005).

2.2 *Yersinia Adhesin A*

YadA (*Yersinia* adhesin A) is an outer membrane adhesin encoded on pYV, which is important for attachment of the bacterium to host cells. *YadA* is not under the control of low calcium response (*lcr*), so *YadA* is maximally expressed at 37°C regardless of Ca^{2+} concentration. It is a 45-kDa protein that mediates intimate bacterial

attachment to host cells independent of other virulence determinants such as the Yops (Bliska et al. 1993; Cornelis et al. 1998). It has been shown that YadA can bind to a variety of surface molecules including collagens, fibronectin, laminins, and mucus (Marra and Isberg 1997; Cornelis et al. 1998). YadA can also mediate internalization of *Yersinia* possibly via interaction with the β -integrins of the host cell. There is a difference between *Y. enterocolitica* and *Y. pseudotuberculosis* with regard to reliance on YadA. Mutation of *yadA* in *Y. enterocolitica* results in decreased virulence, whereas a mutation in *Y. pseudotuberculosis yadA* does not affect virulence. This could be due to additional adhesion factors in *Y. pseudotuberculosis* that render YadA redundant (Cornelis et al. 1998).

Although the mother plasmid is highly conserved between the three pathogenic *Yersinia* species, they are not the same. *YadA* is a pseudogene in *Y. pestis* due to a frameshift, but is functional in the other two species. Similarly, owing to the insertion of IS elements, *yopM* is also nonfunctional in *Y. pestis* (Cornelis et al. 1998; Chain et al. 2004). The mother plasmid has an important role in the virulence of all three human pathogenic species of *Yersinia*, either by the adhesion of the bacteria to the surface of the host cell or by effect on the innate and adaptive immune response.

3 Daddy Bear (pMT1)

pMT1, the daddy plasmid, is the largest and the least studied of the three plasmids found in *Y. pestis* ranging in size from 60 to 100 kb. During the late 1990s, two groups independently determined the complete DNA sequence of pMT1 isolated from *Y. pestis* KIM5 (Hu et al. 1998; Lindler et al. 1998); these groups revealed that pMT1 was approximately 100,990 bp in length and had an overall GC content of 52%. The daddy plasmid carries five important genes that have been associated with virulence located within a region spanning 16 kb of the plasmid: F1 capsular antigen (Caf1), F1 capsule anchoring protein (Caf1A), F1 capsule chaperone protein (Caf1M), F1 capsule regulatory protein (Caf1R), and plague *Yersinia* murine toxin (Ymt) (Table 3). Depending upon the criteria used to identify predicted coding sequences (CDSs), a further 70–115 CDSs in addition to the five genes mentioned above may be encoded on pMT1 (Hu et al. 1998; Lindler et al. 1998). To date, none of the additional CDSs have been characterized. However, loci adjacent to *caf1* and *ymt* may be involved in virulence (Welkos et al. 2004).

3.1 *Ymt* the Putative Murine Toxin

Ymt, first characterized in 1955 (Ajl et al. 1955), is a 586–amino acid protein that encodes phospholipase D (PLD) (Rudolph et al. 1999). Phospholipase D is one of a group of enzymes that have assorted functions and collectively known as the PLD superfamily. PLD enzymes can be found in a wide range of organisms including

humans. Ymt has two conserved motifs of HXKX₄DX₆GG/S that are characteristic of the PLD superfamily. *In vivo*, Ymt is located in the cytoplasm of the bacterium and is thought to be released during cell death when the bacteria lyse during the end stages of septicemic plague resulting in circulatory collapse (Rudolph et al. 1999; Hinnebusch et al. 2000).

Originally, Ymt was named murine toxin owing to its high toxicity to mice and rats, and for this reason it was thought to play a role in the lethality of plague-infected mice. Yet, when Ymt was tested in other animals such as dogs, monkeys, and guinea pigs, it was not toxic (Perry and Fetherston 1997), and when administered intravenously to mice, *Y. pestis ymt*⁻ did not significantly reduce virulence. It is now believed to be nonessential for virulence in mice (Hinnebusch et al. 2000). Hinnebusch et al. noted that Ymt expression was greatest at 26°C rather than 37°C and this led them to investigate whether Ymt may be necessary for the flea stage of the transmission cycle (Bacot and Martin 1914). Hinnebusch et al. also demonstrated that Ymt is necessary for *Y. pestis* to initially colonize the digestive tract of fleas. Studies using *Y. pestis ymt*H188N (a strain in which a single amino acid in one of the conserved catalytic domains had been substituted resulting in > 99% reduction of phospholipase activity) prevented blockage in all 319 fleas examined, and *Y. pestis ymt*H188N was eliminated from up to 95% of fleas within 24 h. In comparison, the wild-type strain caused blockage rates of between 24 and 38% in fleas and infection rates of 50–80% after 24 h. Furthermore, it was demonstrated that Ymt, which is located intracellularly, may protect *Y. pestis* from a toxin or neutralizing agent present in the flea midgut (Hinnebusch et al. 2002). The precise mechanism of protection has yet to be determined.

3.2 F1 Antigen

F1 antigen is a 15.5-kDa protein encoded by *caf1*, which forms a large gel-like capsule or envelope around *Y. pestis* (Baker et al. 1952) which is only expressed at 37°C and comprising linear fibers of Caf1 subunits (Zavialov et al. 2003). Cloning and sequencing of the F1 antigen and region on pMT1 revealed an operon (F1 operon) consisting of four genes, *caf1*, *caf1M*, *caf1A*, and *caf1R* (Galyov et al. 1990). Caf1A and Caf1M share many similarities to proteins required for pili biogenesis. Caf1M shares significant amino acid similarity to the chaperone protein PapD of uropathogenic *E. coli* and was shown to be required for the synthesis of the capsular antigen by acting as a chaperone during folding and secretion of F1 antigen (Galyov et al. 1991; Chapman et al. 1999; MacIntyre et al. 2001). In addition, the outer membrane protein Caf1A is also required for assembly and anchoring of F1 (Karlyshev et al. 1992a), and Caf1R is involved in the regulation of capsule gene expression (Karlyshev et al. 1992b).

F1 antigen is an important protective antigen (Oyston et al. 1995; Andrews et al. 1996); however, isogenic mutants of *caf1* do not increase the LD₅₀ significantly in models such as mice and guinea pigs, but did result in prolonged survival of the infected animal (Drozdov et al. 1995). F1 antigen is thought to be involved in the

antiphagocytic activity of *Y. pestis* (Cavanaugh and Randall 1959). Work by Du et al. using J774 macrophage cells revealed that 95% of *Y. pestis* EV76 were able to resist phagocytosis. However, in a strain that was unable to express the F1 antigen, nearly 30% was phagocytosed (Du et al. 2002). The method by which F1 antigen blocks phagocytic uptake is clearly different to that of Yop effector proteins (described earlier in this chapter), exported by the T3SS present on pYV. F1 antigen is expressed on the surface of the bacterium and it is unlikely that it enters host cells. Furthermore, although F1 antigen aids in the resistance of phagocytosis, it does not seem to have a major effect on macrophage as they can still ingest other prey such as yeast (Du et al. 2002). Thus the precise role that F1 antigen plays in resistance to phagocytosis has yet to be determined. More recently, the F1 antigen has been shown to inhibit bacterial uptake by three different types of human respiratory tract epithelial cell lines by preventing both adhesion and internalization (Liu et al. 2006).

4 Baby Bear (pPCP1)

The baby plasmid pPCP1 has been associated with the following four biochemical activities: a bacteriocin called pesticin, a protein that provides immunity to pesticin, a plasminogen activator, and a coagulase (Table 3). However, Sodeinde et al. mapped the loci that encode these proteins and discovered that both coagulase and plasminogen-activator proteins were encoded by the same gene, *pla* (the precursor pre-Pla is cleaved to produce two proteins a-Pla and b-Pla) (Sodeinde and Goguen 1988). The absence of pPCP1 from strains of *Y. pestis* was shown to increase the lethal dose required in guinea pigs by 50% when it was injected subcutaneously or intraperitoneally, yet had no effect on virulence when such strains were injected via an intravenous route (Brubaker et al. 1965).

4.1 *Pla* (Plasminogen Activator)

Pla is an outer membrane protein. Amino acid deduction from the nucleotide sequence of *pla* showed homology (59%) to *E. coli* OmpT but did not show significant homology to the sequences of other known bacterial or eukaryote plasminogen activators (Sodeinde and Goguen 1989). Both coagulases and plasminogen activators are thought to be involved in the contribution to virulence in pathogenic bacteria. Plasminogen activators are thought to aid in dissemination of invading bacteria through the activation of plasminogen and, in turn, plasmin, which acts to degrade fibrin resulting in the break up of clots surrounding the foci of an infection. Coagulase activity has been shown to be low in *Y. pestis* (Sodeinde and Goguen 1989); thus it was proposed that plasminogen activator could be the primary activity of Pla. Mutation of *pla* in *Y. pestis* was shown to reduce virulence significantly and disease was restricted to the site of injection with no spread to the spleen or liver in mice when administered subcutaneously (Sodeinde et al. 1992). This suggests

that Pla aids bacterial spread of *Y. pestis*, causing systemic infection. However, this is not the sole function of Pla. It is a multifunctional protein and is thought to enhance attachment to mammalian extracellular matrices and promote invasion of human epithelial cells (Lahteenmaki et al. 1998; Cowan et al. 2000; Lahteenmaki et al. 2001). The primary route of infection for *Y. pestis* is subcutaneously via a flea bite; however, tiny aerosol droplets, which if inhaled can lead to pneumonic plague, can also transmit *Y. pestis* from person to person. Pneumonic plague is often fatal if not treated rapidly. Recent work by Lathem et al. has demonstrated that when a *Y. pestis pla* mutant was used to infect mice intranasally, only 50% of the mice developed terminal plague compared to all mice infected with the wild-type strain after one week (Lathem et al. 2007). Deletion of *pla* resulted in reduced growth of *Y. pestis* in the lungs but did not affect dissemination from the lungs to the liver and spleen, indicating that Pla controls the proliferation of *Y. pestis* within the lungs but has no effect on systemic spread. The precise targets of Pla within lung tissue is not currently known; however, this protease may act on fibrin clots that form during infection of the lungs, breaking them up to expedite evasion (Lathem et al. 2007).

4.2 *Pesticin and Immunity*

Baby plasmid also encodes a bacteriocin (*pst*) called pesticin. Bacteriocins are bacterial products that kill bacteria from different closely related species. Pesticin has been shown to be active against serotype I strains of *Y. pseudotuberculosis* and some strains of *Y. enterocolitica* as well as isolates of *Y. pestis* which are pesticin negative (Brubaker and Surgalla 1961; Hu et al. 1972). Pesticin functions by interfering with cell-wall biosynthesis and acts as a lysozyme that hydrolyses a bond between *N*-acetylglucosamine and *N*-acetylmuramic acid within the glycan backbone of murein, thereby degrading the cell wall resulting in instability and cell lysis (Ferber and Brubaker 1979). Pesticin enters sensitive cells via the FyuA receptor which is usually used to translocate yersiniabactin, a ferric siderophore (Rakin et al. 1994). The immunity protein (*pim*), which is located in the periplasm (Pils et al. 1996), is thought to inactivate imported pesticin before it can hydrolyze murein, thereby protecting the bacterium from lysis.

4.3 *Origin and Mobility of the Baby and Daddy Plasmid*

A conjugation apparatus appears to be absent in *Yersinia* plasmids, but horizontal transfer has occurred, as is apparent by the acquisition of several plasmids by *Y. pestis*. Since the acquisition of baby and daddy plasmids are likely to have played an important role in the evolution of *Y. pestis* from *Y. pseudotuberculosis* and its emergence from an enteric pathogen to a highly virulent bacterium disseminated primarily via the subcutaneous route, it is important to try to discover the origin and capacity for

mobility of these two plasmids. Prentice et al. have shown that the daddy plasmid sequence from strains of *Y. pestis* CO92 and KIM5 share at least 52% similarity to plasmid pHCM2, isolated from the exclusively human pathogen *Salmonella enterica* serovar Typhi (which was originally isolated from a typhoid patient in Ho Chi Minh city) (Prentice et al. 2001). The daddy plasmid appears to be a “chimera” of several insertion elements, bacteriophage, and plasmids. The region of pMT1 forming a mosaic of DNA identities and insertion sequences is also the region that encodes the known *Y. pestis* virulence genes F1 and murine toxin. Markedly, there is almost no similarity to pHCM2 in this region. *Ymt* and the *caf* operon are absent in pHCM2, and sequencing of pMT1 revealed that the GC ratio of this 16-kb region was lower (45.8%) than that of the rest of the plasmid, which is indicative of insertion of mobile genetic elements such as pathogenicity islands. Thus it is likely that the acquisition of both *Ymt* and the *caf* operon may have resulted from several homologous recombination events involving many mobile genetic elements (Lindler et al. 1998; Prentice et al. 2001). The origin and evolution of pMT1 and pHCM2 are unknown; it is possible that the plasmids merged from a patient coinfecting with typhoid and plague, or there may be an unidentified common ancestor.

Studies have shown that pMT1 can readily integrate into the bacterial chromosome of *Y. pestis* (Protsenko et al. 1991). Integration of this plasmid is not specific to one region of the bacterial chromosome and does not result in any loss of functions encoded by the daddy plasmid. The presence of *IS100* elements at approximately 10 sites throughout the *Y. pestis* chromosome and on pMT1 may be a possible explanation for the integration of the daddy plasmid onto the chromosome via homologous recombination (Protsenko et al. 1991).

Very little is known about the origin of the baby plasmid. The origin of replication and a region upstream of the bacteriocins of pPCP1 from *Y. pestis* and pColE1 plasmid from *E. coli* have a high level of sequence similarity (Pils1 et al. 1996; Hu et al. 1998). In addition, the regions flanking pesticin are very similar to other bacteriocins such as colicins 5, K, and E1. It is possible that exchange of bacteriocins has occurred through homologous recombination between these regions. However, the baby plasmid possesses plasmidogen activator gene *pla*, which is absent in other bacteriocin-carrying plasmids.

5 Other Plasmids Found in Pathogenic *Yersinia*

In addition to the baby, mother, and daddy plasmids, a number of other plasmids are occasionally found in the pathogenic *Yersinia* (Table 4). Often these plasmids, which are found in strains isolated from patients or the environment, are cryptic. Typically, infection with *Y. pseudotuberculosis* presents with abdominal pain, fever, and diarrhea, although the most common presentation is pseudoappendicitis (Zhou et al. 2006). However, during the 1950s an epidemic of *Y. pseudotuberculosis* spread through areas of Russia causing scarlet fever-type symptoms including skin rashes and toxic shock syndrome. Strains causing these atypical clinical presentations

Table 4 Examples of cryptic plasmids found in the pathogenic *Yersinia*

Plasmid	Strain	Size (bp) if known	References
pCRY	<i>Y. pestis</i> Microtus 91001	21,742	Song et al. (2004)
p29930	<i>Y. enterocolitica</i> 29930 1A	Unknown	Strauch et al. (2003)
pYC	<i>Y. pestis</i>	5,919	Dong et al. (2000)
19 kb plasmid	<i>Y. pestis</i>	19,000	Chu et al. (1998)
pYpsIP31758.1	<i>Y. pseudotuberculosis</i>	153,140	Eppinger et al. (2007)
pYpsIP31758.2	<i>Y. pseudotuberculosis</i>	58,679	Eppinger et al. (2007)

were referred to as Far East scarlet-like fever (FESLF) *Y. pseudotuberculosis*. In order to ascertain genetic differences that may be attributable to these unusual symptoms, a strain of *Y. pseudotuberculosis* (IP31758) isolated from a stool sample of a patient from the former Soviet Union suffering from FESLF has recently been sequenced (Eppinger et al. 2007). Analysis revealed that the mother plasmid was absent from this strain (this may not be unusual as pYV can frequently be lost during *in vitro* cultivation) and possessed two novel plasmids, pYpsIP31758.1 (~153 kb) and pYpsIP31758.2 (~58 kb). Sequencing of both plasmids identified a number of genes that could contribute to the unusual FESLF pathogenicity of IP31758. pYpsIP31758.1 carries a type IVB *icm/dot* secretion system that has not been discovered previously in *Yersinia*, an operon encoding toxoflavin, which has been associated with fitness and pathogenicity in *Burkholderia glumae* (Kim et al. 2004), and several further genes associated with environmental fitness and regulation (Eppinger et al. 2007). The smaller plasmid encodes a plasmid transfer system that may act for both plasmids (which is absent in pYpsIP31758.1). In addition, pilin-encoding gene clusters that are different from those carried on the chromosome are present on both plasmids (Eppinger et al. 2007). Although a number of potential virulence determinants carried on these two novel plasmids may play a role in the FESLF pathogenicity infection, both plasmids are not always found in *Y. pseudotuberculosis* strains causing this unusual disease syndrome (Eppinger et al. 2007).

6 Conclusion

Plasmids have played a central role in the evolution of the *Yersinia* from benign soil-dwelling bacteria to enteropathogens and the plague bacillus (see Fig. 1). There are few such clear-cut examples in nature where plasmids have had such a dramatic effect on the virulence of a bacterium exemplified by the mother plasmid that enabled the discovery of the original and prototype example of a T3SS with its multiple effector molecules. But what separates the mild enteropathogenic *Y. enterocolitica* and *Y. pseudotuberculosis* from the devastating plague bacillus? The baby and daddy plasmids are unlikely to contribute to the hypervirulent state of *Y. pestis* and are more likely to be important in the transmission of the pathogen in its newly acquired flea-borne vector. Differences among the three pathogenic *Yersinia* cannot be explained exclusively through their plasmid content, and other genetic forces on the chromosome including gene gain, gene rearrangements, and gene loss also

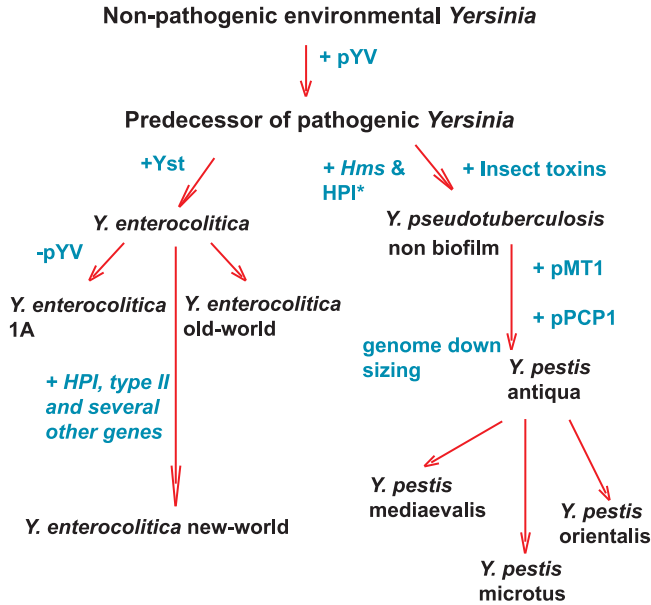


Fig. 1 The evolution of pathogenic *Yersiniae* (adapted from Wren Nature reviews on Microbiology). Acquisition (and loss) of the three plasmids pYV, pPCP1, and pMT1. Yst refers to heat-stable toxin, type II refers to type II secretion system, Hms refers to Haemin storage locus, and HPI* refers to the high pathogenicity island (different islands in *Y. enterocolitica* and *Y. pseudotuberculosis*)

account for virulence and their specific life styles (Wren 2003). Overall, the *Yersinia* represents an ideal genus to study the origin and role of plasmids as well as the evolution of virulence (Wren 2003). It remains to be determined whether we have exhausted information from these three plasmids and whether further virulence/transmissibility/fitness factors remain to be discovered.

In truth, the three bear plasmids are pretty nasty characters with their respective virulence payloads. Daddy bear with the biggest plasmid has toxins, capsules, and yet-to-be-discovered determinants; the delinquent baby bear has the plasminogen activator; and pYV, the mother of all plasmids, carries multiple effector molecules – perhaps Goldilocks should avoid the bears’ house all together?

References

Aili M, Isaksson EL, Carlsson SE, Wolf-Watz H, Rosqvist R, Francis MS (2007) Regulation of *Yersinia* Yop-effector delivery by translocated YopE. *Int J Med Microbiol* 298(3–4):183–192

Ajl SJ, Reedal JS, Durrum EL, Warren J (1955) Studies on plague. I. Purification and properties of the toxin of *Pasteurella pestis*. *J Bacteriol* 70(2):158–169

Andrews GP, Heath DG, Anderson GW Jr, Welkos SL, Friedlander AM (1996) Fraction 1 capsular antigen (F1) purification from *Yersinia pestis* CO92 and from an *Escherichia coli* recombinant strain and efficacy against lethal plague challenge. *Infect Immun* 64(6):2180–2187

Bacot AW, Martin CJ (1914) Observations of the mechanisms of transmission of plague by fleas. *J Hyg* 13:423–439

- Badie G, Heithoff DM, Mahan MJ (2004) LcrV synthesis is altered by DNA adenine methylase overproduction in *Yersinia pseudotuberculosis* and is required to confer immunity in vaccinated hosts. *Infect Immun* 72(11):6707–6710
- Baker EE, Sommer H, Foster LE, Meyer E, Meyer KF (1952) Studies on immunization against plague. I. The isolation and characterization of the soluble antigen of *Pasteurella pestis*. *J Immunol* 68(2):131–145
- Bliska JB, Copass MC, Falkow S (1993) The *Yersinia pseudotuberculosis* adhesin YadA mediates intimate bacterial attachment to and entry into HEP-2 cells. *Infect Immun* 61(9):3914–3921
- Broms JE, Francis MS, Forsberg A (2007) Diminished LcrV secretion attenuates *Yersinia pseudotuberculosis* virulence. *J Bacteriol* 189(23):8417–8429
- Brubaker RR, Surgalla MJ (1961) Pesticins. I. Pesticinbacterium interrelationships, and environmental factors influencing activity. *J Bacteriol* 82:940–949
- Brubaker RR, Beesley ED, Surgalla MJ (1965) *Pasteurella pestis*: role of pesticin I and iron in experimental plague. *Science* 149(3682):422–424
- Cavanaugh DC, Randall R (1959) The role of multiplication of *Pasteurella pestis* in mononuclear phagocytes in the pathogenesis of flea-borne plague. *J Immunol* 83:348–363
- Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M, Medigue C, Simonet M, Chenal-Francois V, Souza B, Dacheux D, Elliott JM, Derbise A, Hauser LJ, Garcia E (2004) Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 101(38):13826–13831
- Chapman DA, Zavialov AV, Chernovskaya TV, Karlyshev AV, Zav'yalova GA, Vasiliev AM, Dudich IV, Abramov VM, Zav'yalov VP, MacIntyre S (1999) Structural and functional significance of the FGL sequence of the periplasmic chaperone Caf1M of *Yersinia pestis*. *J Bacteriol* 181(8):2422–2429
- Chu MC, Dong XQ, Zhou X and Garon CF (1998). A cryptic 19-kilobase plasmid associated with U.S. isolates of *Yersinia pestis*: a dimer of the 9.5-kilobase plasmid. *Am J Trop Med Hyg* 59(5):679–86
- Cornelis GR (2002a) *Yersinia* type III secretion: send in the effectors. *J Cell Biol* 158(3):401–408
- Cornelis GR (2002b) The *Yersinia* Ysc-Yop 'type III' weaponry. *Nat Rev Mol Cell Biol* 3(10):742–752
- Cornelis GR, Boland A, Boyd AP, Geuijen C, Iriarte M, Neyt C, Sory MP, Stainier I (1998) The virulence plasmid of *Yersinia*, an antihost genome. *Microbiol Mol Biol Rev* 62(4):1315–1352
- Cowan C, Jones HA, Kaya YH, Perry RD, Straley SC (2000) Invasion of epithelial cells by *Yersinia pestis*: evidence for a *Y. pestis*-specific invasin. *Infect Immun* 68(8):4523–4530
- Dong XQ, Lindler LE and Chu MC (2000). Complete DNA sequence and analysis of an emerging cryptic plasmid isolated from *Yersinia pestis*. *Plasmid* 43(2):144–8
- Drozdzov IG, Anisimov AP, SamoiloVA SV, Yezhov IN, Yeremin SA, Karlyshev AV, Krasilnikova VM, Kravchenko VI (1995) Virulent non-capsulate *Yersinia pestis* variants constructed by insertion mutagenesis. *J Med Microbiol* 42(4):264–268
- Du Y, Rosqvist R, Forsberg A (2002) Role of fraction 1 antigen of *Yersinia pestis* in inhibition of phagocytosis. *Infect Immun* 70(3):1453–1460
- Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, Fayolle C, Lindler LE, Carniel E, Ravel J (2007) The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. *PLoS Genet* 3(8):e142
- Ferber DM, Brubaker RR (1979) Mode of action of pesticin: N-acetylglucosaminidase activity. *J Bacteriol* 139(2):495–501
- Galyov EE, Smirnov O, Karlyshev AV, Volkovoy KI, Denesyuk AI, Nazimov IV, Rubtsov KS, Abramov VM, Dalvadyanz SM, Zav'yalov VP (1990) Nucleotide sequence of the *Yersinia pestis* gene encoding F1 antigen and the primary structure of the protein. Putative T and B cell epitopes. *FEBS Lett* 277(1–2):230–232
- Galyov EE, Karlyshev AV, Chernovskaya TV, Dolgikh DA, Smirnov O, Volkovoy KI, Abramov VM, Zav'yalov VP (1991) Expression of the envelope antigen F1 of *Yersinia pestis* is mediated

- by the product of *cafIM* gene having homology with the chaperone protein PapD of *Escherichia coli*. FEBS Lett 286(1–2):79–82
- Heesemann J, Sing A, Trulzsch K (2006) *Yersinia*'s stratagem: targeting innate and adaptive immune defense. Curr Opin Microbiol 9(1):55–61
- Hinnebusch BJ, Cherepanov P, Du Y, Rudolph A, Dixon JD, Schwan T, Forsberg A (2000) Murine toxin of *Yersinia pestis* shows phospholipase D activity but is not required for virulence in mice. Int J Med Microbiol 290(4–5):483–487
- Hinnebusch BJ, Rudolph AE, Cherepanov P, Dixon JE, Schwan TG, Forsberg A (2002) Role of *Yersinia* murine toxin in survival of *Yersinia pestis* in the midgut of the flea vector. Science 296(5568):733–735
- Hu PC, Yang GC, Brubaker RR (1972) Specificity, induction, and absorption of pesticin. J Bacteriol 112(1):212–219
- Hu PC, Elliott J, McCreedy P, Skowronski E, Garnes J, Kobayashi A, Brubaker RR, Garcia E (1998) Structural organization of virulence-associated plasmids of *Yersinia pestis*. J Bacteriol 180(19):5192–5202
- Huang XZ, Nikolich MP, Lindler LE (2006) Current trends in plague research: from genomics to virulence. Clin Med Res 4(3):189–199
- Iriarte M, Sory MP, Boland A, Boyd AP, Mills SD, Lambermont I, Cornelis GR (1998) TyeA, a protein involved in control of Yop release and in translocation of *Yersinia* Yop effectors. Embo J 17(7):1907–1918
- Isaksson EL, Olsson J, Bjornfot A, Rosqvist R, Aili M, Carlsson KE, Francis MS, Hamberg M, Klumperman J, Wolf-Watz H (2006) Effector translocation is mediated by a surface located protein complex of *Y. pseudotuberculosis*. In: Ninth international symposium on *Yersinia*, Lexington, Kentucky
- Karlyshev AV, Galyov EE, Smirnov O, Guzayev AP, Abramov VM, Zav'yalov VP (1992a) A new gene of the *fI* operon of *Y. pestis* involved in the capsule biogenesis. FEBS Lett 297(1–2):77–80
- Karlyshev AV, Galyov EE, Abramov VM, Zav'yalov VP (1992b) CafIR gene and its role in the regulation of capsule formation of *Y. pestis*. FEBS Lett 305(1):37–40
- Kim J, Kim JG, Kang Y, Jang JY, Jog GJ, Lim JY, Kim S, Suga H, Nagamatsu T, Hwang I (2004) Quorum sensing and the LysR-type transcriptional activator ToxR regulate toxoflavin biosynthesis and transport in *Burkholderia glumae*. Mol Microbiol 54(4):921–934
- Lahteenmaki K, Virkola R, Saren A, Emody L, Korhonen TK (1998) Expression of plasminogen activator *pla* of *Yersinia pestis* enhances bacterial attachment to the mammalian extracellular matrix. Infect Immun 66(12):5755–5762
- Lahteenmaki K, Kukkonen M, Korhonen TK (2001) The *Pla* surface protease/adhesin of *Yersinia pestis* mediates bacterial invasion into human endothelial cells. FEBS Lett 504(1–2):69–72
- Latham WW, Price PA, Miller VL, Goldman WE (2007) A plasminogen-activating protease specifically controls the development of primary pneumonic plague. Science 315(5811):509–513
- Lemaitre N, Sebbane F, Long D, Hinnebusch BJ (2006) *Yersinia pestis* YopJ suppresses tumor necrosis factor alpha induction and contributes to apoptosis of immune cells in the lymph node but is not required for virulence in a rat model of bubonic plague. Infect Immun 74(9):5126–5131
- Lian CJ, Hwang WS, Kelly JK, Pai CH (1987) Invasiveness of *Yersinia enterocolitica* lacking the virulence plasmid: an in-vivo study. J Med Microbiol 24(3):219–226
- Lindler LE, Plano GV, Burland V, Mayhew GF, Blattner FR (1998) Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. Infect Immun 66(12):5731–5742
- Liu F, Chen H, Galvan EM, Lasaro MA, Schifferli DM (2006) Effects of Psa and F1 on the adhesive and invasive interactions of *Yersinia pestis* with human respiratory tract epithelial cells. Infect Immun 74(10):5636–5644
- MacIntyre S, Zyrianova IM, Chernovskaya TV, Leonard M, Rudenko EG, Zav'yalov VP, Chapman DA (2001) An extended hydrophobic interactive surface of *Yersinia pestis* CafIM chaperone is essential for subunit binding and F1 capsule assembly. Mol Microbiol 39(1):12–25

- Marra A, Isberg RR (1997) Invasin-dependent and invasin-independent pathways for translocation of *Yersinia pseudotuberculosis* across the Peyer's patch intestinal epithelium. *Infect Immun* 65(8):3412–3421
- Mueller CA, Broz P, Muller SA, Ringler P, Erne-Brand F, Sorg I, Kuhn M, Engel A, Cornelis GR (2005) The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* 310(5748):674–676
- Naktin J, Beavis KG (1999) *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *Clin Lab Med* 19(3):523–536, vi
- Nilles ML, Williams AW, Skrzypek E, Straley SC (1997) *Yersinia pestis* LcrV forms a stable complex with LcrG and may have a secretion-related regulatory role in the low-Ca²⁺ response. *J Bacteriol* 179(4):1307–1316
- Oyston PC, Williamson ED, Leary SE, Eley SM, Griffin KF, Titball RW (1995) Immunization with live recombinant *Salmonella typhimurium* aroA producing F1 antigen protects against plague. *Infect Immun* 63(2):563–568
- Perry RD, Fetherston JD (1997) *Yersinia pestis*—etiologic agent of plague. *Clin Microbiol Rev* 10(1):35–66
- Pilsil H, Killmann H, Hantke K, Braun V (1996) Periplasmic location of the pesticin immunity protein suggests inactivation of pesticin in the periplasm. *J Bacteriol* 178(8):2431–2435
- Prentice MB, James KD, Parkhill J, Baker SG, Stevens K, Simmonds MN, Mungall KL, Churcher C, Oyston PC, Titball RW, Wren BW, Wain J, Pickard D, Hien TT, Farrar JJ, Dougan G (2001) *Yersinia pestis* pFra shows biovar-specific differences and recent common ancestry with a *Salmonella enterica* serovar Typhi plasmid. *J Bacteriol* 183(8):2586–2594
- Protsenko OA, Filippov AA, Kutyrev VV (1991) Integration of the plasmid encoding the synthesis of capsular antigen and murine toxin into *Yersinia pestis* chromosome. *Microb Pathog* 11(2):123–128
- Rakin A, Saken E, Harmsen D, Heesemann J (1994) The pesticin receptor of *Yersinia enterocolitica*: a novel virulence factor with dual function. *Mol Microbiol* 13(2):253–263
- Rudolph AE, Stuckey JA, Zhao Y, Matthews HR, Patton WA, Moss J, Dixon JE (1999) Expression, characterization, and mutagenesis of the *Yersinia pestis* murine toxin, a phospholipase D superfamily member. *J Biol Chem* 274(17):11824–11831
- Sodeinde OA, Goguen JD (1988) Genetic analysis of the 9.5-kilobase virulence plasmid of *Yersinia pestis*. *Infect Immun* 56(10):2743–2748
- Sodeinde OA, Goguen JD (1989) Nucleotide sequence of the plasminogen activator gene of *Yersinia pestis*: relationship to ompT of *Escherichia coli* and gene E of *Salmonella typhimurium*. *Infect Immun* 57(5):1517–1523
- Sodeinde OA, Subrahmanyam YV, Stark K, Quan T, Bao Y, Goguen JD (1992) A surface protease and the invasive character of plague. *Science* 258(5084):1004–1007
- Song Y, Tong Z, Wang J, Wang L, Guo Z, Han Y, Zhang J, Pei D, Zhou D, Qin H, Pang X, Han Y, Zhai J, Li M, Cui B, Qi Z, Jin L, Dai R, Chen F, Li S, Ye C, Du Z, Lin W, Wang J, Yu J, Yang H, Wang J, Huang P and Yang R (2004). Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res* 11(3): 179–97
- Strauch E, Goelz G, Knabner D, Konietzny A, Lanka E and Appel B (2003). A cryptic plasmid of *Yersinia enterocolitica* encodes a conjugative transfer system related to the regions of CloDF13 Mob and IncX Pil. *Microbiology* 149(Pt 10):2829–45
- Viboud GI, Bliska JB (2005) *Yersinia* outer proteins: role in modulation of host cell signaling responses and pathogenesis. *Annu Rev Microbiol* 59:69–89
- Welkos SL, Andrews GP, Lindler LE, Snellings NJ, Strachan SD (2004) Mu dII(Ap lac) mutagenesis of *Yersinia pestis* plasmid pFra and identification of temperature-regulated loci associated with virulence. *Plasmid* 51(1):1–11
- Wren BW (2003) The yersinia—A model genus to study the rapid evolution of bacterial pathogens. *Nat Rev Microbiol* 1(1):55–64
- Zavialov AV, Berglund J, Pudney AF, Fooks LJ, Ibrahim TM, MacIntyre S, Knight SD (2003) Structure and biogenesis of the capsular F1 antigen from *Yersinia pestis*: preserved folding energy drives fiber formation. *Cell* 113(5):587–596
- Zhou D, Han Y, Yang R (2006) Molecular and physiological insights into plague transmission, virulence and etiology. *Microbes Infect* 8(1):273–284

Virulence Megaplasmid in *Bacillus anthracis* and Their Relatives in the *Bacillus cereus* Group

Agnès Fouet (✉) and Marie Moya

Contents

1	Introduction.....	188
2	pXO1 and pXO1-like Plasmids	188
2.1	pXO1.....	189
2.2	pXO1-like Plasmids.....	193
3	pXO2 and pXO2-like Plasmids	197
3.1	pXO2.....	198
3.2	pXO2-like Plasmids.....	199
4	Concluding Remarks.....	202
	References.....	203

Abstract *Bacillus anthracis* is the etiological agent of anthrax. In the 1980s, a link was discovered between plasmids and major virulence factors of *B. anthracis*. Indeed, the three toxin components are encoded by pXO1, a 181-kb plasmid, and the poly-glutamate capsule biosynthetic operon is carried by pXO2, a 97-kb plasmid. The functions encoded by a few other genes were described after screening or selecting for specific phenotypes, such as regulation of virulence factor synthesis. Despite a renewal of interest in *B. anthracis* at the end of the twentieth century, which prompted further research and led to complete sequencing of both plasmids in 1999, only few genes have been fully characterized. These include genes involved in replication, sporulation, and germination. Yet, 40% of the open reading frames (ORFs) are of unknown function, and for most of the others a function has only been predicted *in silico*. *B. anthracis* is, on a genetic basis, a *Bacillus cereus*. *B. cereus* strains also harbor megaplasmid, some sharing core sequences with pXO1 and pXO2. These plasmids also encode virulence factors or specific environmental adaptive pathways.

A. Fouet and M. Moya
Institut Pasteur, Toxines et Pathogénie Bactérienne, CNRS, URA 2172, France
e-mail: afouet@pasteur.fr

1 Introduction

Bacillus anthracis, a Gram-positive, non-motile, aerobic, facultatively anaerobic, spore-forming, rod-shaped bacterium, is the etiological agent of anthrax (Mock and Fouet 2001). Anthrax is primarily a disease of herbivores, but all mammals including humans are susceptible. Both the resistant form found in the soil and the infective form are spores. The disease is initiated by the entry of spores, by various routes, into the host body. After entering the host, the spores germinate, and the bacilli multiply and synthesize the virulence factors including the major ones: a tripartite toxin and a polyglutamate capsule. This results in toxemia and septicemia, leading to the death of the animal. The bacilli are then shed into the environment where, in the presence of oxygen, they can sporulate, and the cycle can reinitiate.

Fully virulent strains of *B. anthracis* harbor two large plasmids, pXO1 (~181 kbp) and pXO2 (~97 kbp), considered to be major virulence determinants, since strains lacking either one of these plasmids are attenuated in most animal models. This is the consequence of pXO1 harboring the toxin genes and pXO2 the capsule biosynthetic operon (Mikesell et al. 1983; Vodkin and Leppla 1983; Green et al. 1985; Uchida et al. 1985). The synthesis of these virulence factors is induced or enhanced in the presence of CO₂ and at 37°C, conditions mimicking the mammal environment (Bartkus and Leppla 1989; Sirard et al. 1994; Fouet and Mock 1996).

B. anthracis belongs to the *B. cereus* group, which, among other species, includes *Bacillus cereus sensu stricto* and *Bacillus thuringiensis*. *B. cereus* is ubiquitous in the environment, and some strains are opportunistic pathogens, causing gastrointestinal and non-gastrointestinal infections (Kotiranta et al. 2000). *B. thuringiensis* is an entomopathogen that can occasionally be an opportunistic human pathogen (Hernandez et al. 1998). On a genetic basis, i.e., as shown by multilocus enzyme electrophoresis and by sequence analysis of nine chromosomal genes, *B. anthracis* is a *B. cereus* (Helgason et al. 2000). Plasmids are found in most *B. cereus* strains and they harbor genes encoding many specific virulence factors and some important phenotypic traits. Some of these plasmids share pXO1 or pXO2 backbones (Table 1).

In this chapter, we will describe our current knowledge on pXO1 and pXO2 and the genes that have been characterized, which, despite the availability of their sequences since 1999, are but a few (Okinaka et al. 1999a, b). We will also discuss the *B. cereus* plasmids that display similarity to the *B. anthracis* plasmids. However, other megaplasmids, which may be very important, such as those harboring the *cry* genes encoding the entomopathogen crystal proteins (González et al. 1981), will not be dealt with in this chapter (for a review see Schnepf et al. 1998).

2 pXO1 and pXO1-like Plasmids

The fact that *B. anthracis* cells cured of a large plasmid produced no detectable lethal toxin or edema activity and very little protective antigen (PA, a toxin component, see Sect. 2.1) serological activity suggested the involvement of a plasmid, initially termed

Table 1 Plasmids discussed in this article

Plasmid name	Size (bp)	Host organism	pXO - like	GenBank Accession no.	Reference
pXO1	181,677	<i>B. anthracis</i>		AE017336	Okinaka et al. (1999a)
pBc10987	208,369	<i>B. cereus</i> ATCC 10987	pXO1	AE017195	Rasko et al. (2004)
pBCXO1	191,110	<i>B. cereus</i> G9241	pXO1	DQ889680	Hoffmaster et al. (2004)
pPER272	272,145	<i>B. cereus</i> AH817	pXO1		Rasko et al. (2007)
pPER818	272,145	<i>B. cereus</i> AH818	pXO1	DQ889678	Rasko et al. (2007)
pPER820	272,145	<i>B. cereus</i> AH820	pXO1	DQ889677	Rasko et al. (2007)
pCER270	204,529	<i>B. cereus</i> AH817	pXO1	DQ889676	Rasko et al. (2007)
pBtoxis	127,923	<i>B. thuringiensis israelensis</i>	pXO1	AL731825 (EMBL)	Berry et al. (2002)
pXO2	94,829	<i>B. anthracis</i>		AE017335	Okinaka et al. (1999b)
pAW63	71,777	<i>B. thuringiensis kurstaki</i>	pXO2	DQ025752	Van der Auwera et al. (2005)
pBT97-27	77,112	<i>B. thuringiensis</i> 9727	pXO2	CP000047	Rasko et al. (2005)

pBA1, in the toxin component synthesis (Mikesell et al. 1983). Furthermore, transformation of cured strains by pBA1 restored these activities. The first toxin gene to be cloned, from a plasmid-derived library, was that encoding the protective antigen. This confirmed that at least one toxin component gene was harbored by a plasmid in *B. anthracis* (Vodkin and Leppla 1983). This plasmid, since called pXO1, has been studied further as well as *B. cereus* megaplasms that share a highly conserved core region with it.

2.1 pXO1

The 181-kb pXO1 plasmid was sequenced and 143 open reading frames (ORFs) were identified (Okinaka et al. 1999a). For about 70% of the predicted ORF products (108 of 143 ORFs), no putative functions could be assigned. Later *in silico* analysis resulted in a significant reduction in the percentage of totally unknown ORFs, from ~70 to ~40% (Ariel et al. 2002). Importantly, within the isolates of *B. anthracis* for which the genome sequence is available, the pXO1 sequences are, except for one inversion (see below), highly similar. Apart from a small number of Single Nucleotide Polymorphisms (SNPs), Single Nucleotide Repeats (SNRs), and Variable Nucleotide Tandem Repeats (VNTRs), no large insertions or deletions have been observed (Read et al. 2002). A striking feature is the existence of an inversion of an approximately 40-kb DNA segment (Thorne 1993). This was deduced from the difference observed in restriction maps made for pXO1s originating from different Sterne strains. Sterne strains, originally described as noncapsulated strains, are pXO1⁺ pXO2⁻ strains used as veterinary vaccines (Sterne 1937). Finding this region in both orientations indicates that it is transpositionally active.

Sequence analysis showed that this region, 44.8 kb in size, is bordered by inverted copies of IS1627. It resides between ORF96 and ORF127, both of which encode hypothetical transposases (Okinaka et al. 1999a). The IS-flanked segment contains, among others, the three toxin genes (*pagA*, *cya* and *lef*; ORFs 110, 122 and 107, respectively), encoding the protective antigen (PA), the edema factor (EF), and the lethal factor (LF), respectively. This region thus displays many features that define pathogenicity islands (PAIs) and is now referred to as such.

The three toxin components, PA, LF, and EF, are secreted proteins and their secretion requires the extracellular chaperone PrsA, (Williams et al. 2003). There are three chromosomally encoded putative PrsA proteins (PrsAA, PrsAB, and PrsAC) in *B. anthracis*. Each of them is able to complement the activity of *Bacillus subtilis* PrsA with respect to cell viability and secretion of recombinant PA. PA is known to elicit a protective immune response against anthrax (Gladstone 1946). PA is expressed as a pre-protein and binds to ubiquitous cell receptors (for a review see Collier and Young 2003). After binding, PA is proteolytically processed into its active form and oligomerizes. Up to three molecules of EF and/or LF can bind to the heptamer and be endocytosed. LF affects host cells by proteolytic cleavage of mitogen-activated protein kinases (MAPKKs) and EF by conversion of intracellular ATP into cAMP, resulting in edema, necrosis, and hypoxia (for a review see Moayeri and Leppla 2004).

The three structural genes, *cya* (EF), *lef* (LF), and *pagA* (PA) are coordinately regulated by CO₂ and temperature (Sirard et al. 1994). Furthermore, the transcription reaches its maximum upon entry of the cells into stationary phase. Different regulators have been shown to influence toxin gene transcription (for a review see Fouet and Mock 2006). By introducing a transcriptional fusion containing a *pagA* promoter in a plasmid-less strain of *B. anthracis*, it was shown that a toxin gene regulator is harbored by pXO1 (Bartkus and Leppla 1989). AtxA, encoded by a gene in the PAI on pXO1 was initially described as a toxin gene trans-activator (Uchida et al. 1993a; Koehler et al. 1994). It was shown to also activate the capsule operon transcription (Guignot et al. 1997; Uchida et al. 1997) to activate a chromosomally encoded S-layer gene and repress a second chromosomal S-layer gene (Mignot et al. 2003) and, finally, to influence positively or negatively the transcription of many genes carried by all genetic elements (Bourgogne et al. 2003). The mechanism of action of AtxA is still not understood. So far, the binding of AtxA to promoter regions of genes it controls has not been shown. *pagA* is the first gene of a bicistronic operon. The second gene, *pagR*, encodes a weak autorepressor (Hoffmaster and Koehler 1999). Interestingly, PagR is an intermediate between AtxA and the S-layer genes it controls; furthermore PagR binds to the S-layer gene promoters as well as to that of *pagA* (Mignot et al. 2003).

Other interesting ORFs were found on the PAI. Pallen identified the ORF98 product as a member of the novel ESAT-6/WX100 superfamily (based on similar motifs found in *Mycobacterium tuberculosis* proteins), members of which may constitute a new Gram-positive secretion system (in spite of the absence of a signal or anchoring sequence) (Pallen 2002). As the *Mycobacterium* orthologs have been reported to be relevant to the virulence of pathogenic mycobacteria, the same may be true for the *B. anthracis* ortholog, particularly in view of the location of ORF98 within the PAI.

The ORF105 product (67 amino acids) is a truncated version of AbrB (a central pleiotropic transcriptional regulator). The full-sized paralog encoded in the *B. anthracis* chromosome is involved in the temporal regulation of toxin gene expression (Saile and Koehler 2002; Strauch et al. 2005).

A pXO1-cured strain differs from the parental strain in colony morphology, grows more slowly on minimal medium than its pXO1 counterpart, and sporulates earlier and at a higher frequency than the parental strain (Thorne 1993). Interestingly, the PAI harbors a tricistronic *ger* operon, consisting of the genes *gerXB*, *gerXA*, and *gerXC*, (ORF 114, 113 and 112 respectively), located between the *pagA* and *atxA* genes (Guidi-Rontani et al. 1999). A *ger* null mutant strain did not germinate efficiently within phagocytic cells or in vivo and was less virulent than the parental strain. However, all known in vitro germination pathways are utilized by *B. anthracis* endospores, which lack the pXO1 virulence plasmid and thus the *gerX* germinant receptor locus; its cognate ligand remains unknown (Fisher and Hanna 2005).

The ORF118 product (150 amino acids), previously defined as an *atxA* gene product, has a shorter (136-residue) paralog that is encoded on pXO2 (pXO2-61). These proteins are highly homologous to the signal sensor domain of a chromosomally encoded major sporulation sensor histidine kinase (BA2291) (White et al. 2006). The initiation of sporulation is controlled by the phosphorelay signal transduction system. The signals that initiate the phosphorelay reactions are recognized and interpreted by several sporulation sensor histidine kinases, the nature of the signals remaining a mystery (Brunsing et al. 2005). The sporulation sensor kinase BA2291 is converted from an activator to an inhibitor of sporulation in *B. anthracis* by these plasmid-encoded signal sensor domains (White et al. 2006). The fact that each virulence plasmid-encoded sensor domain is located within a pathogenicity island suggests that the regulation of the function of BA2291 by these domains may be one of the links coordinating pathogenesis and the inhibition of sporulation (White et al. 2006).

BXA0205 encodes a Rap-Phr system; members of the Rap family of phosphatases are involved in the regulatory cascade leading to sporulation (Aronson et al. 2005; Bongiorno et al. 2006). The BXA0205Rap protein, initially described as a regulator of extracellular proteases in *B. anthracis*, was shown to dephosphorylate the Spo0F response regulator intermediate of the phosphorelay signal transduction system. This system regulates the initiation of the developmental pathway in response to environmental, metabolic, and cell cycle signals. The C-terminal pentapeptide generated through an export-import processing pathway from the associated BXA0205Phr protein inhibits the activity of the Rap protein. Overexpression or lack of the Phr pentapeptide results in severe inhibition of sporulation.

No origin of replication could be defined on pXO1 by sequence homology to known replication initiator proteins. An analysis based on Guanosine Cytidine (GC) skew was used and a 5-kb fragment capable of auto-replication was thus subcloned (Tinsley and Khan 2006). One particular coding sequence, renamed *repX* (also known as pXO1-45), is required for the initiation of replication (Tinsley and Khan 2006). It has similarities to the GTPases of the FtsZ family. The functional replication origin

appears to be a 158-bp fragment, located approximately 2 kb from *repX*, indicating that the corresponding protein acts in *trans*. The origin contains an inverted repeat sequence of 24-bp with one extra C in one arm of *B. anthracis* Sterne pXO1; in the other *B. anthracis* pXO1 plasmid sequences, as well as the *B. cereus* pXO1-like plasmids, the palindrome has another 2-bp mismatch (Rasko et al. 2007). The effect, if any, of these differences on plasmid replication is unclear.

pXO1 (ORF142, *topX*) (Fouet et al. 1994). The protein sequence suggests a hybrid structure, with the *N*-terminal portion resembling more a topoisomerase III and the *C*-terminal a topoisomerase I. This protein could thus either be involved in plasmid topology as a topoisomerase I (and therefore potentially in expression of pXO1-harbored genes) or nonexclusively in the replication and segregation of pXO1 as a topoisomerase III.

Of the 143 ORF products, about 50% contain either a secretion signal sequence (cleavable or uncleavable) and/or a transmembrane segment(s) (Ariel et al. 2002). The ORF45, ORF130, and ORF141 products contain lipoprotein signal sequences. The secreted or anchored proteins include the tripartite toxin (surface-layer homology) harboring proteins, hydrolytic enzymes, additional putative virulence factors, and several unknown proteins. SLH proteins are noncovalently anchored to pyruvylated peptidoglycan through the SLH domain, which is composed of three repeats of about 50 amino acids, termed SLH motifs (Mesnage et al. 2000). Only a single pXO1-derived protein, the ORF79 product, seems to have a Gram-positive LPXTG anchoring signal. These proteins are anchored covalently to the peptidoglycan and this reaction is catalyzed by enzymes named sortases (for a review see Dramsi et al. 2005).

The ORF54 product is a 404-amino-acid SLH protein. The ORF65 product is a 202-amino-acid secreted protein. The secondary structure of this protein resembles (with > 90% certainty) that of the *Bordetella pertussis* virulence factor cell adhesin p.69/pentactin (Everest et al. 1996). The ORF67 product is a 562-amino-acid membrane-anchored (via its *N*-terminal helices) protein that is probably involved in protein-protein interactions. Part of the sequence resembles a cell adhesion domain from *Clostridium acetobutylicum*, which harbors tetratricopeptide repeats and is involved in protein-protein interactions. The ORF90 product is a 652-amino-acid membrane-anchored protein that carries its three SLH motifs in the *N*-terminus. The SLH domain is followed by five coiled-coil segments, a lysine-rich region, and a leucine zipper motif. The *C*-terminal part is rich in myosin-like repeats.

Genes related to T4SSs (type IV secretion systems) are present on both plasmids as well as on the chromosome of *B. anthracis*. Indeed, in addition to the eight previously known genes, 14 novel T4SS-related genes (plus two chaperone-like genes) have been detected, (Grynberg et al. 2007). Phylogenetically, the T4SS genes are distant cousins of two branches of T4SSs: the *Actinobacillus actinomycetemcomitans* widespread colonization island-like systems (eight and three genes on pXO1 and pXO2, respectively) and the canonical T4SSs (three, four, and one genes on pXO1, pXO2, and the chromosome, respectively). In recent years, evidence has been gathered that demonstrates the presence of T4SSs not only in Gram-negative species, but also in Gram-positive bacteria, although only in a few cases has it been shown experimentally that the genes are really involved in secretion or/and in conjugation (Grohmann et al. 2003). Even though the mechanisms of action, especially in Gram-positive species and also in *B. anthracis*,

remain unclear, the discovery of T4SS-like genes on *B. anthracis* plasmids opens a new direction in studies on bacterial secretion or/and conjugation.

More ORFs with putative functions of biological interest are also present on plasmid pXO1 and, for example, the ORF53 product is a 131-amino-acid protein with an amino acid sequence unique to *B. anthracis*. It may function as a putative metal-dependent hydrolase.

2.2 *pXO1-like Plasmids*

Numerous *B. cereus* genome and plasmid sequencing projects have significantly increased our understanding of the biological role of plasmids in *B. cereus* and raised questions regarding the classifications within the *B. cereus* group (Rasko et al. 2007). The plasmids of the *B. cereus* complex can be separated into three groups on the basis of their nucleotide sequence similarity to the *B. anthracis* plasmids: pXO1-like, pXO2-like, and others. The first group includes plasmids that are similar to pXO1 and includes pBc10987 (*B. cereus* ATCC 10987), pBCXO1 (*B. cereus* G9241), pPER818 (*B. cereus* AH818), pPER820 (*B. cereus* AH820), collectively referred to as pPER272 and pCER270 (*B. cereus* AH187) (Fig. 1) (Rasko et al. 2007). Plasmids from this pXO1-like group range in size from ~181 to 272 kb and, interestingly, *B. anthracis* pXO1 is the smallest of the group.

Historically, conservation of pXO1 sequences in other *Bacillus* including multiple *B. cereus* isolates led to the conclusion that some strains from the *B. cereus* group harbor pXO1-like plasmids (Pannucci et al. 2002). For instance, three isolates contain sequences that are similar to more than one-half of the pXO1 ORF sequences examined. Groups of ORFs (2–6) are in a conserved order compared to that on pXO1, and DNA sequences of amplified fragments display between 80 and 98% similarity to those of pXO1. These sequences are located on megaplastids of approximately 300 kb (Pannucci et al. 2002).

Comparison of the *B. cereus* pXO1-like plasmids to *B. anthracis* plasmid pXO1 revealed a highly conserved core region. This region of approximately 50 kb contains genes that have been shown to be involved in plasmid replication and maintenance. Indeed, alignment of the RepX homolog from each of the pXO1-like plasmids demonstrates that this protein is highly conserved (> 98% identity among the examined isolates) and is encoded in that region (Tinsley and Khan 2006).

Four pXO1-like plasmids are described in more detail. pBtoxis is not one of them because it shares little homology with pXO1. It is included in Fig. 1 because, when sequenced, it was compared to pXO1 (Berry et al. 2002) and later it was used for the sake of comparison when pBc10987 was sequenced and shown to be a pXO1-like plasmid (Rasko et al. 2004).

2.2.1 pBc10987

B. cereus ATCC 10987, originally isolated from a dairy source, harbors a 208-kb pXO1-like plasmid, pBc10987 (Rasko et al. 2004). pBc10987 shows surprising

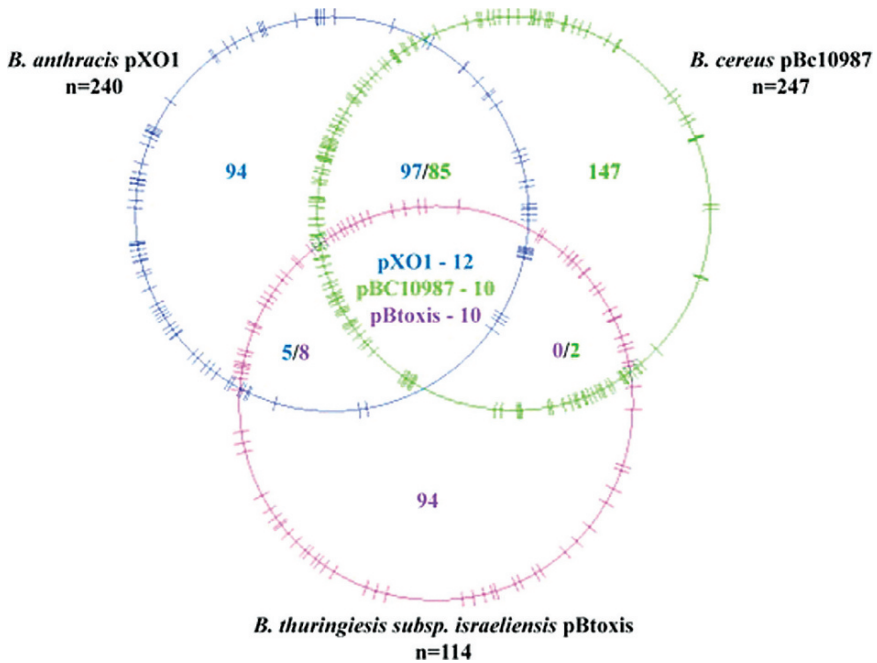


Fig. 1 Venn diagram illustrating the number of putative proteins associated with each plasmid and the number shared with the intersecting plasmid. Tick marks that are on each circle represent the location of the unique proteins. Blue, *B. anthracis* Ames pXO1; green, *B. cereus* ATCC10987 pBc10987; magenta, *B. thuringiensis* subsp. *israeliensis* pBtoxis (from Rasko et al. 2004)

similarity to the plasmid pXO1 of *B. anthracis* (40% nucleotide identity, 65% proteins are homologous and 50% are in a syntenic location) (Fig. 1). pBc10987 lacks the pathogenicity islands (PAIs), but its neighborhood is conserved. pBc10987 contains genes potentially involved in adaptation to either an environmental or a pathogenic lifestyle, different from that of *B. anthracis*. Environmental adaptations of pBc10987 include a copper-requiring tyrosinase, arsenite resistance and its associated regulators, as well as an amino acid transport system, proteins promoting life in the soil or the insect. The pathogenic adaptations comprise two potential novel toxins, BCEA0165, a major intrinsic protein (MIP) family channel protein, and BCEA0203, a possible metalloprotease. Unlike the PAI in pXO1, this region in pBc10987 is not flanked by mobile genetic elements.

An additional copy of *abrB*, a pleiotropic transition state regulator, is present both on pXO1 and on pBc10987, *B. anthracis* and *B. cereus* ATCC 10987 carrying also chromosomal copies of *abrB*. In *B. anthracis*, the chromosomally encoded AbrB has been shown to negatively regulate toxin production (Saile and Koehler 2002 Strauch et al. 2005). While the pBc10987 and pXO1 AbrB proteins are similar (80% amino acid identity), they differ significantly at their *N*-terminus, in that the pXO1 protein is 27 amino acids shorter, rendering it potentially inactive

(Saile and Koehler 2002). pBc10987 *abrB* may be involved in control of plasmid gene expression.

Besides *AbrB*, there are two more examples of possible genetic exchange between the chromosome and the plasmid of the *B. cereus* ATCC 10987. Identical copies of a transposable element are present on both the chromosome and the plasmid. There are four genes associated with the Tn554-like element, one of which is *bclA*. In *B. anthracis*, multiple functions have been attributed to *BclA*. It is the major spore surface antigenic protein (Sylvestre et al. 2002) and has been shown to play a role in the length determination of the exosporium associated filaments (Sylvestre et al. 2003). Divergent proteins of the same family as *BclA* are encoded on the chromosomes of *B. cereus*. Interestingly, in *B. anthracis* a single copy of *bclA* is located on the chromosome, whereas in *B. cereus* ATCC 10987 one copy of a *bclA*-like gene is present on the chromosome and one on the plasmid.

2.2.2 pBCXO1

B. cereus G9241 was isolated from a patient with an anthrax-like disease and possesses unique biochemical and phenotypic traits (Hoffmaster et al. 2004). This isolate harbors pBCXO1, a plasmid that is 99.6% identical to the *B. anthracis* plasmid pXO1. This similarity extends to the pXO1 virulence genes, protective antigen (99.7% amino acid identity), lethal factor (99% amino acid identity), and edema factor (96% amino acid identity), as well as the known virulence regulatory proteins *AtxA* (100% amino acid identity) and *PagR* (98.6% amino acid identity) (Hoffmaster et al. 2004). This high level of identity between pXO1 and pBCXO1 suggests that both isolates acquired their plasmid from a common ancestor or that transfer occurred into *B. cereus* G9241 from *B. anthracis*. Two sets of data argue in favor of an acquisition, after strain separation, of pXO1 or at least of the PAI. The first one is experimental: although it required the aid of a mobilizing plasmid, the successful transfer of pXO1 from *B. anthracis* to a close relative has been observed (Battisti et al. 1985). The second one is observational: *PlcR*, a transcriptional regulator, most often an activator that is found in all *B. cereus* strains, controls the synthesis of many *B. cereus* virulence factors such as hemolysins and phospholipases, and many other secreted enzymes (Slamti et al. 2004). However, in *B. anthracis*, *plcR* is nonfunctional, with the same mutation in all strains tested. The simultaneous presence of active *PlcR* and *AtxA* regulators has been shown to inhibit sporulation in *B. anthracis*. The introduction of pXO1-encoded *AtxA* is thought to have led to the selection for the *B. anthracis* *plcR* nonsense mutation (Mignot et al. 2001). This is supported by the fact that some *B. cereus* strains have *plcR* mutations, but they are all different from that observed in *B. anthracis* (Slamti et al. 2004). Strikingly, in *B. cereus* G9241 *PlcR* is functional, suggesting a post strain-separation acquisition of the *atxA*-harboring pBCXO1. It has to be tested whether in *B. cereus* G9241 both *PlcR* and *AtxA* are active and, furthermore, whether their regulons are identical. Indeed, the incompatibility between the presence of both regulators and sporulation in *B. anthracis*

is not defined as yet and could depend on the expression of only a few genes. In light of the discovery of this important organism, functional studies of *B. cereus* G9241 regulatory pathways must be performed to develop a better understanding of this evolutionary event.

B. cereus G9241 also contains two additional plasmids: a 52-kb linear element with similarity to phages of the *B. cereus* group, including that of *B. cereus* ATCC 14579, and pBC218, a large circular plasmid, which contains a putative polysaccharide biosynthetic gene cluster thought to be responsible for the production of the observed encapsulation of *B. cereus* G9241 (Hoffmaster et al. 2004). Intriguingly, pBC218 carries a second copy of *atxA*; however, the gene product is only 78% identical to the *B. anthracis* homolog. Interestingly, a homolog of pXO1-118 is found close to this gene (see Sect. 2.1). Genes coding for a homolog of the protective antigen peptide and the lethal factor are also found on pBC218; however, the plasmid does not contain a homolog of the edema factor. The lethal factor is significantly truncated compared to that of *B. anthracis* pXO1 and is probably not functional. In contrast, the pBC218-encoded protective antigen peptide is conserved with respect to its of *B. anthracis* counterpart (55% identities, 71% similarities). The presence of these additional toxin subunits raises some interesting questions about their putative function in the pathogenesis of this organism.

2.2.3 pPER272, pCER270

The plasmids from *B. cereus* isolates producing emetic toxin or linked to periodontal disease were sequenced and analyzed (Rasko et al. 2007). Two periodontal isolates examined contained almost identical ~272 kb plasmids, named pPER272. The emetic toxin-producing isolate contained one ~270 kb plasmid, named pCER270, encoding the cereulide biosynthesis gene cluster. Sequence analyses of these *B. cereus* plasmids revealed a high degree of sequence similarity to the *B. anthracis* pXO1 plasmid, especially in a putative replication region. These plasmids form a newly defined group of pXO1-like plasmids. Sequence analysis resulted in the identification of unique virulence-associated regions in pCER270 which contain, in addition to the cereulide synthetase gene cluster previously described (Ehling-Schulz et al. 2006), a region shared with pPER272 and the chromosomes of the *B. cereus* group strains. The genes of this region include sporulation and germination genes, as well as a formaldehyde detoxification locus. The cereulide synthetase gene cluster is harbored by a 24 kb DNA segment that comprises seven ORFs. These encode a typical nonribosomal peptide synthetase, two enzyme modules for activation and incorporation of monomers in the growing chain, a putative hydrolase and an ABC transporter.

Examination of pPER272 did not result in the conclusive identification of classical virulence factors. pPER272 contains a region previously described in pBc10987, which replaces pXO1 PAI (Rasko et al. 2004). This region encodes an MIP channel homolog and other membrane proteins that may allow interaction with epithelial cells in the oral cavity. Adjacent to this shared region, pPER272 contains a unique

additional 90 kb of sequence not found in pXO1. It would be difficult to determine whether pBc10987 has lost the pPER272 unique region over time in the environment or whether pPER272 acquired the additional region to adapt to other ecological niches. Alternatively, the shared pPER272/pBc10987 region may encode factors associated with periodontal virulence, and the categorization of pBc10987 as an environmental plasmid is misleading. Further molecular analysis is required to determine whether any virulence determinants are encoded on pPER272 and/or pBc10987.

The highly conserved core region of the pXO1-like plasmids, a backbone, can be used to analyze the evolutionary link between the plasmids. It is clear that plasmids pXO1 and pBCXO1 are closely related, and the other plasmids are significantly divergent. pBc10987 is more closely related to the pXO1/pBCXO1 cluster than the pathogenic pXO1-like plasmids, pPER272 and pCER270, suggesting that it has diverged from the pXO1/pBCXO1 group more recently than the pPER272 and pCER270 plasmids. pBc10987, pPER272, and pCER270 share an extended core region that is not shared with pXO1, suggesting that pXO1 may have either evolved further or may represent a more ancestral form of the plasmid. In either case, pXO1 is more distantly related than the other pXO1-like plasmids to the other members of that plasmid group.

3 pXO2 and pXO2-like Plasmids

The second group of plasmids does not show any similarity to pXO1 and includes pBC218 (*B. cereus* G9241), pXO2 (*B. anthracis*), and the pXO2-like plasmids such as pAW63 (*B. thuringiensis* AW63) and pBT9727 (*B. thuringiensis* serovar Konkukian strain 97-27). These non-pXO1-like plasmids range in size from ~54 to 466 kb. pXO1-like plasmids have a copy number between 1 and 3, whereas plasmid pXO2 has a copy number ranging from two to five copies per chromosome (estimated from seven *B. anthracis* genome projects). The difference in copy number may be attributed to differences in plasmid size (181 kb vs. 95 kb for *B. anthracis* pXO1 and pXO2, respectively) or replication and maintenance mechanisms as described by Tinsley and Khan (2004, 2006).

It was observed that all *B. anthracis* capsulated strains harbored a 60-Mda plasmid, in contrast to those that were noncapsulated (Uchida et al. 1985). Furthermore, two classes of rough, noncapsulated strains were characterized, those that reverted and those that did not (Green et al. 1985). The latter class had lost the megaplasmsid pXO2. pXO2 was introduced by transduction or mating experiments into *B. cereus* strains. After the acquisition of pXO2, these strains produced a capsule, indicating that pXO2 carries genes required for capsule synthesis (Green et al. 1985; Uchida et al. 1985). In contrast to pXO1, no pXO2 plasmid has been found in non-*B. anthracis* strains. However, plasmids with a pXO2 backbone have been described in *B. thuringiensis* strains. The common backbone has been thoroughly analyzed for the pXO2-like plasmid pAW63 and will be described below.

3.1 *pXO2*

The first complete sequence of *pXO2* (96 kb) was obtained from a *B. anthracis* Pasteur strain (*pXO1*⁻, *pXO2*⁺) and encodes 104 genes of which 78 do not have functional assignments (75%) (Okinaka et al. 1999b). Co-analysis of *pXO2* and plasmids possessing the same backbone led to the description of a PAI on *pXO2* (Van der Auwera et al. 2005). It is a 37-kb segment that harbors the capsule biosynthetic operon and the genes encoding its regulators (see below). Furthermore, it has a lower G + C content than the rest of *pXO2* and is rich in mobile genetic elements, including cryptic or defective ones.

pXO2 harbors a five-gene operon, *capBCADE*, ORFs 58–54, that is responsible for capsule biosynthesis (Candela et al. 2005). The capsule is composed of poly- γ -D-glutamate. It protects the vegetative bacterial cells from phagocytosis, thereby evading the host immune response and increasing systemic sepsis (Makino et al. 1989). *Cap* mutants are, like *pXO2*⁻ strains, of attenuated virulence. Four genes, *capB*, *capC*, *capA*, and *capE*, of the five-gene operon are necessary and sufficient for poly- γ -D-glutamate synthesis in physiological conditions (Candela et al. 2005). The corresponding proteins are membrane-associated proteins, most probably organized as a membranous complex, and the function of each has been tentatively assigned by genetic experiments, sequence analysis, and biochemical studies (for a review see Candela and Fouet 2006). PgsB, the equivalent of *CapB* in *B. subtilis*, may be the polymerase; it belongs to the poly(γ -glutamate) ligase family and can by itself catalyse poly-glutamate synthesis in vitro, (Eveland et al. 1997; Urushibata et al. 2002). *B. subtilis* *CapB*-like and *CapC*-like proteins form a tight complex and may have polymerase activity in situ (Ashiuchi et al. 1999; Urushibata et al. 2002). A mixture of *CapB*-like, *CapC*-like, and *CapA*-like has the highest ATPase activity, suggesting that these proteins are organized in a complex (Ashiuchi et al. 1999). *CapA* could be the polyglutamate transporter (Ashiuchi et al. 2001). Finally, *CapE* is necessary, when proteins are produced in physiological concentrations, for poly-glutamate synthesis and must interact with *CapA* to exert its activity (Candela et al. 2005). The product of the fifth gene, *capD*, is a γ -glutamyl-transpeptidase (Candela and Fouet 2005). *CapD* catalyzes the covalent anchoring of polyglutamate to the peptidoglycan, most probably to its peptidic moiety via Dap (DL-mesodiaminopimelic acid). Interestingly, a *capD* mutant is of attenuated virulence, probably because of the absence of covalent anchoring of the capsular material to the surface of the bacteria. *CapD* was initially described as a polyglutamate depolymerase (Uchida et al. 1993b). In fact, as a glutamyl-transpeptidase, it transfers glutamate onto an acceptor, which can be either an amino acid or water. In the latter case, the reaction leads to depolymerization. Degradation of the capsular material by *CapD* results in high- and low-molecular-weight forms of the capsule, both of which may be essential for infection (Makino et al. 2002).

The *capBCADE* operon is under the control of a number of transcriptional activators. Two of them, *AcpA* and *AcpB* (ORFs 64 and 53 respectively), are encoded by *pXO2* and their genes are in fact located close to the *cap* operon, *acpB*, belonging to the *cap* operon (Vietri et al. 1995; Drysdale et al. 2004; Drysdale et al. 2005a). The toxin gene

regulator AtxA, encoded by pXO1, is implicated in *cap* operon transcription activation (Guignot et al. 1997; Uchida et al. 1997). *acpA* and *acpB* have been shown to be themselves under the control of AtxA (Bourgogne et al. 2003). Although their binding to the *cap* operon promoter has not been reported, our hypothesis is that, like PagR for the S-layer genes, AcpA and AcpB are intermediates between AtxA and the *cap* operon (Fouet and Mock 2006). AcpA and AcpB are not functional equivalents and AcpB may have a more general role in virulence (Drysdale et al. 2005b).

An autolysin, AmiA (ORF42) (Mesnage and Fouet 2002). This enzyme, a N-acetylmuramoyl-L-alanine amidase, is a bi-modular protein, composed of an SLH domain and an activity domain. Interestingly, peptidoglycan hydrolysis by AmiA does not require cell-wall binding. Four amino-acid residues have been shown to be essential for catalysis. *amiA* transcription is, like the *cap* operon, activated by AtxA and by AcpA (Bourgogne et al. 2003). AtxA transcriptional activation of *amiA* may, again, be due to AtxA activating *acpA* transcription.

The pXO2 replication region was initially described after analyzing a hot spot of Tn917 insertions in pXO2 (Hoover 1998). This region contains two ORFs whose products are similar to plasmid replication proteins, e.g., RepR of *E. faecalis* pAMβ1 and *Streptococcus pyogenes* pIP501 plasmids and RepB of *E. faecalis* pAD1 plasmid. An 8-kb DNA fragment containing these ORFs is capable of replication in *B. anthracis* (Hoover 1998). This fragment also harbors sequences with features of plasmid origins, such as an AT-rich region with 8-bp repeats, 14 copies in one orientation and 4 in the other. More recently, a 2.5-kb fragment was cloned that contains the same ORFs and a 60-nucleotide region essential for the initiation of replication. Interestingly, the product of one of the ORFs, RepS (previously termed RepR), specifically binds to the 60-bp fragment (Tinsley et al. 2004).

pXO2 is believed to be incapable of autonomous transfer although it has been shown to be mobilizable by the conjugative plasmid pXO14 from *B. thuringiensis toumanoffi* (Reddy et al. 1987), suggesting that pXO2 carries a set of conjugative genes that is incomplete or disrupted in one or several key components (see below).

ORF BXB0104 encodes an interrupted topoisomerase. It is shorter than TopX from pXO1 with which it shares sequence similarity (673 vs. 870 amino acid residues; 20.2% identity and 33.9% similarity). Because the 267th codon is a stop codon, there is probably no functional protein synthesized. The topoisomerase is therefore not required for pXO2 stability, nor for expression of the other genes it harbors. It could be an evolutionary ghost.

Recently, seven genes potentially related to T4SSs (type IV secretion systems) were found on pXO2 plasmid (see [Sect 2.1](#))

3.2 *pXO2-like Plasmids*

Characterization of many *B. cereus* plasmids has led to the description of two pXO2-like plasmids ([Fig. 2](#)).

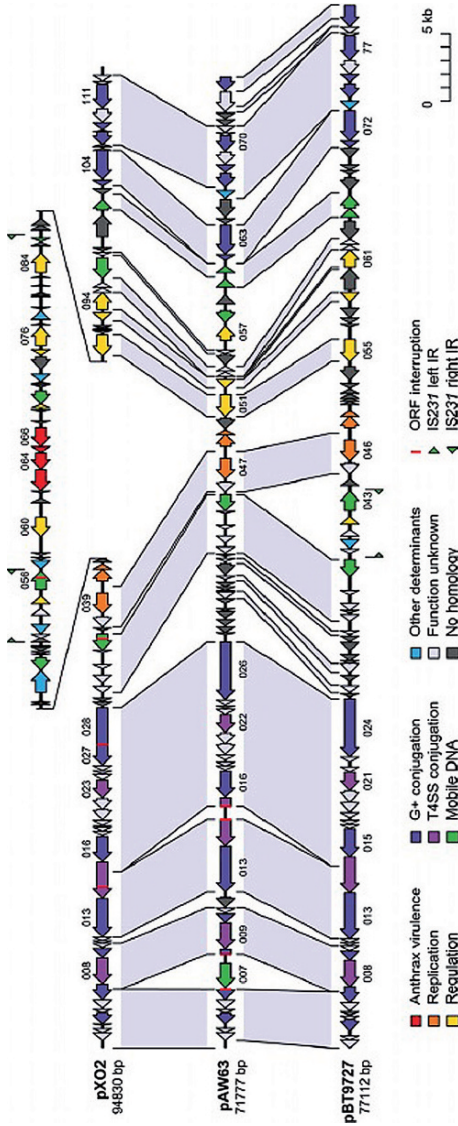


Fig. 2 Linear alignment of pXO2, pAW63, and pBT9727. Cds are represented by block arrows. Several CDS numbers (see Table 1 in Van der Auwera et al. 2005) are indicated for reference on each plasmid, just above or below their representation. Predicted functions/homologies are indicated by the color key featured below. Well conserved segments of the plasmids are paired by shaded regions (> 40% amino acid identity; percentages for specific CDS pairs can be found in Tables 1 and 2 in Van der Auwera et al. 2005). The proposed PAI of pXO2 is raised above the rest of the sequence for clarity. Scale is indicated by the bar in the lower right-hand corner (from Van der Auwera et al. 2005)

pAW63, pXO2 and pBT9727 are highly similar (Fig. 2). Out of the 76 ORFs predicted for pAW63, 50 ORFs showed strong similarity (with an average of 81.1% amino acid identity) to ORFs found on pXO2 from *B. anthracis* and 49 ORFs showed strong similarity (with an average of 70.5% amino acid identity) to pBT9727 from *B. thuringiensis* subsp. *konkukian* (serotype H34) with 42/76 (55.3%) of these ORFs being shared by all three plasmids (Van der Auwera et al. 2005). This is consistent with previous observations of sequence similarities between these plasmids (Pannucci et al. 2002).

3.2.1 pAW63

The broad-host-range conjugative plasmid pAW63 was identified in *B. thuringiensis* serovar *kurstaki*. A 4.1-kb fragment contains all the information for autonomous replication (Wilcks et al. 1999). It harbors four ORFs, displaying strong similarity. The largest, Rep63A, is similar to replication proteins of pAM β 1 and pIP501. The second largest is similar to proteins involved in control of plasmid copy number in plasmids pAD1 and pCF10. The two other genes display no obvious similarities, but share location, orientation, and, with respect to their products, size and hydrophilicity, with putative genes encoding stability functions on plasmids pAD1 and pCF10. An 85-bp region downstream of Rep63A also has strong similarity to the origins of replication of pAM β 1. pAW63 belongs to the pAM β 1 family of theta-replicating plasmids. In fact, the similarity of the pXO2 replicon to the pAW63 replicon led to the conclusion that pXO2 belongs to the same family (Tinsley et al. 2004).

Apart from the replicative system described above, most of the ORFs are involved in conjugation including T4SS proteins and regulators (Van der Auwera et al. 2005).

The comparison of the *tra* regions of pAW63, pBT9727 and pXO2 has provided several valuable clues on the relative importance of specific genes in the conjugative process (Van der Auwera et al. 2005). While pAW63 is known to be fully functional and quite efficient as a conjugative plasmid, pXO2 is transferable but not self-transmissible (Reddy et al. 1987). Both pXO2 and pBT9727 were shown to possess a set of potential conjugation genes closely related to those of pAW63. The unique interruptions found in the pXO2 homolog of the VirD4 element and in that of the putative cell surface protein encoded by ORF26 on pAW63 may prove to be particularly significant, especially if pBT9727 is shown to be fully conjugative (Van der Auwera et al. 2005).

3.2.2 pBT9727

B. thuringiensis 97-27 has been shown to produce crystal protein in sporulated culture by direct microscopic examination (Hernandez et al. 1998). Detailed sequence analysis of pBT9727, the sole plasmid found in *B. thuringiensis* 97-27,

did not reveal any Cry toxin genes with similarity to already known *B. thuringiensis* toxin genes. Interestingly, pBT9727 shows similarity to *B. anthracis* pXO2 (GeneBank CP_000047; Van der Auwera et al. 2005). Comparison of the predicted coding regions of the two plasmids revealed that pBT9727 shares 89% (82/92) of its putative coding sequences with pXO2. Their replication proteins are almost identical and the predicted origin of replication is well conserved (Tinsley et al. 2004). As inferred in the previous paragraph, the level of protein similarity, combined with the conservation of gene order, suggests that these plasmids might have diverged recently. The pXO2 PAI is replaced on pBT9727 with genes encoding hypothetical proteins and putative mobile elements.

The phylogenetic analysis of the three predicted proteomes and the study of the structural features involved in the replication control center have made it possible to roughly reconstruct the molecular evolution of the plasmid trio. pBT9727 was the first to diverge from the common stem, leaving pXO2 and pAW63 to branch off from an intermediate form later. The existence of the PAI was inferred from these analyses (Van der Auwera et al. 2005). The recognition of pXO2 as a PAI-equipped virulence plasmid would also be consistent with the observation that most of the major known pathogens possess such PAIs, many of these carried by virulence plasmids.

4 Concluding Remarks

Analysis of the chromosomes of the *B. cereus* group members shows that they are very similar, both in sequence and synteny (Rasko et al. 2005). The plasmids play a significant role in virulence. This is a consequence of the megaplasmids harboring the major virulence factor genes. This has been recently illustrated by analysis of the in vivo progression of strains lacking one or the other major virulence factor in *B. anthracis* (Glomski et al. 2007a; Glomski et al., 2007b). Major differences can be observed between Tox⁺ Cap⁻ and Tox⁻ Cap⁺ strains. For one thing, the pattern of invasion of the various organs is different. For example, the most striking difference was observed in the spleen; Tox⁺,Cap⁻ and Tox⁻,Cap⁺ bacteria may be differently controlled in the spleen through phagocytosis-dependent mechanisms.

The plasmids could also play a significant adaptive role in the pathogen host range, in the ecology of the isolate, and in the observed clonality (Han et al. 2006; Ehling-Schulz et al. 2005). A *B. cereus* isolate having acquired a pXO1-like plasmid may then contain a unique combination of plasmid and chromosomal factors, allowing successful exploitation of an environmental niche. Examples of highly successful *B. cereus*-group clones are *B. anthracis*, which infects mammals (Mock and Fouet 2001), and *B. thuringiensis*, which specifically infects lepidopteran worms (Schnepf et al. 1998).

Acknowledgments M Moya was funded by DGA no. 04 34 025 and CRSSA (Ministère de la Défense/DGA) no. 04 07 005/00

References

- Ariel N, Zvi A, Grosfeld H, Gat O, Inbar Y, Velan B, Cohen S, Shafferman A (2002) Search for potential vaccine candidate open reading frames in the *Bacillus anthracis* virulence plasmid pXO1: in silico and in vitro screening. *Infect Immun* 70:6817–6827
- Aronson AI, Bell C, Fulroth B (2005) Plasmid-encoded regulator of extracellular proteases in *Bacillus anthracis*. *J Bacteriol* 187:3133–3138
- Ashiuchi M, Soda K, Misono H (1999) A poly-gamma-glutamate synthetic system of *Bacillus subtilis* IFO 3336: gene cloning and biochemical analysis of poly-gamma-glutamate produced by *Escherichia coli* clone cells. *Biochem Biophys Res Commun* 263:6–12
- Ashiuchi M, Nawa C, Kamei T, Song JJ, Hong SP, Sung MH, Soda K, Yagi T, Misono H (2001) Physiological and biochemical characteristics of poly gamma-glutamate synthetase complex of *Bacillus subtilis*. *Eur J Biochem* 268:5321–5328 Corrigendum: *Eur J Biochem* 268:6003
- Bartkus JM, Leppla SH (1989) Transcriptional regulation of the protective antigen gene of *Bacillus anthracis*. *Infect Immun* 57:2295–2300
- Battisti L, Green BD, Thorne CB (1985) Mating system for plasmid transfer of plasmids among *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*. *J Bacteriol* 162:543–550
- Berry C, O'Neil S, Ben-Dov E, Jones AF, Murphy L, Quail MA, Holden MT, Harris D, Zaritsky A, Parkhill J (2002) Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl Environ Microbiol*. 68:5082–5095
- Bongiorni C, Stoessel R, Shoemaker D, Perego M (2006) Rap phosphatase of virulence plasmid pXO1 inhibits *Bacillus anthracis* sporulation. *J Bacteriol* 188:487–498
- Bourgogne A, Drysdale M, Hilsenbeck SG, Peterson SN, Koehler TM (2003) Global effects of virulence gene regulators in a *Bacillus anthracis* strain with both virulence plasmids. *Infect Immun* 71:2736–2743
- Brunsing RL, La Clair C, Tang S, Chiang C, Hancock LE, Perego M, Hoch JA (2005) Characterization of sporulation histidine kinases of *Bacillus anthracis*. *J Bacteriol* 187:6972–6981
- Candela T, Fouet A (2005) *Bacillus anthracis* CapD, belonging to the gamma-glutamyltranspeptidase family, is required for the covalent anchoring of capsule to peptidoglycan. *Mol Microbiol* 57:717–726
- Candela T, Fouet A (2006) Poly-gamma-glutamate in bacteria. *Mol Microbiol* 60:1091–1098
- Candela T, Mock M, Fouet A (2005) CapE, a 47-amino-acid peptide, is necessary for *Bacillus anthracis* polyglutamate capsule synthesis. *J Bacteriol* 187:7765–7772
- Collier RJ, Young JA (2003) Anthrax toxin. *Annu Rev Cell Dev Biol* 19:45–70
- Dramsi S, Trieu-Cuot P, Bierre H (2005) Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. *Res Microbiol* 156:289–297
- Drysdale M, Bourgogne A, Hilsenbeck SG, Koehler TM (2004) *atxA* controls *Bacillus anthracis* capsule synthesis via *acpA* and a newly discovered regulator, *acpB*. *J Bacteriol* 186:307–315
- Drysdale M, Bourgogne A, Koehler TM (2005a) Transcriptional analysis of the *Bacillus anthracis* capsule regulators. *J Bacteriol* 187:5108–5114
- Drysdale M, Heninger S, Hutt J, Chen Y, Lyons CR, Koehler TM (2005b) Capsule synthesis by *Bacillus anthracis* is required for dissemination in murine inhalation anthrax. *EMBO J* 24:221–227
- Ehling-Schulz M, Svensson B, Guinebretiere MH, Lindback T, Andersson M, Schulz A, Fricker M, Christiansson A, Granum PE, Martlbauer E, Nguyen-The C, Salkinoja-Salonen M, Scherer S (2005) Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. *Microbiology* 151:183–197
- Ehling-Schulz M, Fricker M, Grallert H, Riek P, Wagner M, Scherer S (2006) Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1. *BMC Microbiol* 6:20
- Eveland SS, Pompliano DL, Anderson MS (1997) Conditionally lethal *Escherichia coli* murein mutants contain point defects that map to regions conserved among murein and folyl poly- γ -glutamate ligases: identification of a ligase superfamily. *Biochemistry* 36:6223–6229

- Everest P, Li J, Douce G, Charles I, De Azavedo J, Chatfield S, Dougan G, Roberts M (1996) Role of the *Bordetella pertussis* P.69/pertactin protein and the P.69/pertactin RGD motif in the adherence to and invasion of mammalian cells. *Microbiology* 142:3261–3268
- Fisher N, Hanna P (2005) Characterization of *Bacillus anthracis* germinant receptors in vitro. *J Bacteriol* 187:8055–8062
- Fouet A, Mock M (1996) Differential influence of the two *Bacillus anthracis* plasmids on regulation of virulence gene expression. *Infect Immun* 64:4928–4932
- Fouet A, Mock M (2006) Regulatory networks for virulence and persistence of *Bacillus anthracis*. *Curr Opin Microbiol* 9:160–166
- Fouet A, Sirard JC, Mock M (1994) *Bacillus anthracis* pXO1 virulence plasmid encodes a type I DNA topoisomerase. *Mol Microbiol* 11:471–479
- Gladstone GP (1946) Immunity to anthrax – protective antigen present in cell-free culture filtrates. *Brit J Exp Pathol* 27:394–418
- Glomski IJ, Corre JP, Mock M, Goossens PL (2007a) Noncapsulated toxinogenic *Bacillus anthracis* presents a specific growth and dissemination pattern in naive and protective antigen-immune mice. *Infect Immun*. 75:4754–4761
- Glomski IJ, Piris-Gimenez A, Huerre M, Mock M, Goossens PL (2007b) Primary involvement of pharynx and peyer's patch in inhalational and intestinal anthrax. *PLoS Pathog*. 3(6):e76
- González JM Jr, Dulmage HT, Carlton BC (1981) Correlation between specific plasmids and delta-endotoxin production in *Bacillus thuringiensis*. *Plasmid* 5:351–365
- Green BD, Battisti L, Koehler TM, Thorne CB, Ivins BE (1985) Demonstration of a capsule plasmid in *Bacillus anthracis*. *Infect Immun* 49:291–297
- Grohmann E, Muth G, Espinosa M (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol* 67:277–301
- Grynberg M, Li Z, Szczurek E, Godzik A (2007) Putative type IV secretion genes in *Bacillus anthracis*. *Trends Microbiol* 15:191–195
- Guidi-Rontani C, Pereira Y, Ruffie S, Sirard JC, Weber-Levy M, Mock M (1999) Identification and characterization of a germination operon on the virulence plasmid pXO1 of *Bacillus anthracis*. *Mol Microbiol* 33:407–414
- Guignot J, Mock M, Fouet A (1997) AtxA activates the transcription of genes harbored by both *Bacillus anthracis* virulence plasmids. *FEMS Microbiol Lett* 147:203–207
- Han CS, Xie G, Challacombe JF, Altherr MR, Bhotika SS, Bruce D, Campbell CS, Campbell ML, Chen J, Chertkov O, Cleland C, Dimitrijevic M, Doggett NA, Fawcett JJ, Glavina T, Goodwin LA, Hill KK, Hitchcock P, Jackson PJ, Keim P, Kewalramani AR, Longmire J, Lucas S, Malfatti S, McMurry K, Meincke LJ, Misra M, Moseman BL, Mundt M, Munk AC, Okinaka RT, Parson-Quintana B, Reilly LP, Richardson P, Robinson DL, Rubin E, Saunders E, Tapia R, Tesmer JG, Thayer N, Thompson LS, Tice H, Ticknor LO, Wills PL, Brettin TS, Gilna P (2006) Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. *J Bacteriol* 188:3382–3390
- Helgason E, Økstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolstø AB (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—One species on the basis of genetic evidence. *Appl Environ Microbiol* 66:2627–2630
- Hernandez E, Ramisse F, Ducoureau JP, Cruel T, Cavallo JD (1998) *Bacillus thuringiensis* subsp. konkukian (serotype H34) superinfection: case report and experimental evidence of pathogenicity in immunosuppressed mice. *J Clin Microbiol* 36:2138–2139
- Hoffmaster AR, Koehler TM (1999) Autogenous regulation of the *Bacillus anthracis* pag operon. *J Bacteriol* 181:4485–4492
- Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW, Maiden MCJ, Priest FG, Barker M, Jiang LX, Cer RZ, Rilstone J, Peterson SN, Weyant RS, Galloway DR, Read TD, Popovic T, Fraser CM (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc Natl Acad Sci USA* 101:8449–8454
- Hoover TA (1998) Characterization of a region of *Bacillus anthracis* capsule-encoding plasmid pXO2 capable of autonomous replication, In: Abstracts of the 3rd International Conference on Anthrax, University of Plymouth, Plymouth, UK. DERA Chemical and

- Biological Defence Sector, Porton Down, UK, and the Society for Applied Microbiology, Plymouth, UK, p 52
- Koehler TM, Dai Z, Kaufman-Yarbray M (1994) Regulation of the *Bacillus anthracis* protective antigen gene:CO2 and a *trans*-acting element activate transcription from one of two promoters. *J Bacteriol* 176:586–595
- Kotiranta A, Lounatmaa K, Haapasalo M (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. *Microbes Infect* 2:189–198
- Makino SI, Uchida I, Terakado N, Sasakawa C, Yoshikawa M (1989) Molecular characterization and protein analysis of the *cap* region, which is essential for encapsulation in *Bacillus anthracis*. *J Bacteriol* 171:722–730
- Makino S, Watarai M, Cheun HI, Shirahata T, Uchida I (2002) Effect of the lower molecular capsule released from the cell surface of *Bacillus anthracis* on the pathogenesis of anthrax. *J Infect Dis* 186:227–233
- Mesnager S, Fouet A (2002) Plasmid encoded autolysin in *Bacillus anthracis*: modular structure and catalytic properties. *J Bacteriol* 184:331–334
- Mesnager S, Fontaine T, Mignot T, Delepierre M, Mock M, Fouet A (2000) Bacterial SLH-domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall-polysaccharide pyruvylation. *EMBO J* 19:4473–4484
- Mignot T, Mock M, Robichon D, Landier A, Lereclus D, Fouet A (2001) The incompatibility between the PlcR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis*. *Mol Microbiol* 42:1189–1198
- Mignot T, Mock M, Fouet A (2003) A plasmid-encoded regulator couples the synthesis of toxins and surface structures in *Bacillus anthracis*. *Mol Microbiol* 47:917–927
- Mikesell P, Ivins BE, Ristroph JD, Dreier TM (1983) Evidence for plasmid-mediated toxin production in *Bacillus anthracis*. *Infect Immun* 39:371–376
- Moayeri M, Leppla SH (2004) The roles of anthrax toxin in pathogenesis. *Curr Opin Microbiol* 7:19–24
- Mock M, Fouet A (2001) Anthrax. *Annu Rev Microbiol* 55:647–671
- Okinaka RT, Cloud K, Hampton O, Hoffmaster AR, Hill KK, Keim P, Koehler TM, Lamke G, Kumano S, Mahillon J, Manter D, Martinez Y, Ricke D, Svensson R, Jackson PJ (1999a) Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J Bacteriol* 181:6509–6515
- Okinaka R, Cloud K, Hampton O, Hoffmaster A, Hill K, Keim P, Koehler T, Lamke G, Kumano S, Manter D, Martinez Y, Ricke D, Svensson R, Jackson P (1999b) Sequence, assembly and analysis of pXO1 and pXO2. *J Appl Microbiol* 87:261–262
- Pallen MJ (2002) The ESAT-6/WXG100 superfamily – and a new Gram-positive secretion system? *Trends Microbiol* 10:209–212
- Pannucci J, Okinaka RT, Sabin R, Kuske CR (2002) *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. *J Bacteriol* 184:134–141
- Pannucci J, Okinaka RT, Williams E, Sabin R, Ticknor LO, Kuske CR (2002) DNA sequence conservation between the *Bacillus anthracis* pXO2 plasmid and genomic sequence from closely related bacteria. *BMC Genomics* 3:34
- Rasko DA, Ravel J, Økstad OA, Helgason E, Cer RZ, Jiang L, Shores KA, Fouts DE, Tourasse NJ, Angiuoli SV, Kolonay J, Nelson WC, Kolstø AB, Fraser CM, Read TD (2004) The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucleic Acids Res* 32:977–988
- Rasko DA, Altherr MR, Han CS, Ravel J (2005) Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol Rev* 29:303–329
- Rasko DA, Rosovitz MJ, Økstad OA, Fouts DE, Jiang L, Cer RZ, Kolstø AB, Gill SR, Ravel J (2007) Complete sequence analysis of novel plasmids from emetic and periodontal *Bacillus cereus* isolates reveals a common evolutionary history among the *B-cereus*-group plasmids, including *Bacillus anthracis* pXO1. *J Bacteriol* 189:52–64
- Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang LX, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296:2028–2033

- Reddy A, Battisti L, Thorne CB (1987) Identification of self-transmissible plasmids in four *Bacillus thuringiensis* subspecies. *J Bacteriol* 169:5263–5270
- Saile E, Koehler TM (2002) Control of anthrax toxin gene expression by the transition state regulator *abrB*. *J Bacteriol* 184:370–380
- Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol Mol Biol Rev* 62:775–790
- Sirard JC, Mock M, Fouet A (1994) The three *Bacillus anthracis* toxin genes are coordinately regulated by bicarbonate and temperature. *J Bacteriol* 176:5188–5192
- Slamti L, Perchat S, Gominet M, Vilas-Boas G, Fouet A, Mock M, Sanchis V, Chaufaux J, Gohar M, and Lereclus D (2004) Distinct mutations in PlcR explain why some strains of the *Bacillus cereus* group are nonhemolytic. *J Bacteriol*. 186:3531–3538
- Sterne M (1937) Avirulent anthrax vaccine. *Onderstepoort J Vet Sci Animal Ind* 21:41–43
- Strauch MA, Ballar P, Rowshan AJ, Zoller KL (2005) The DNA-binding specificity of the *Bacillus anthracis* AbrB protein. *Microbiology* 151:1751–1759
- Sylvestre P, Couture-Tosi E, Mock M (2002) A collagen-like surface glycoprotein is a structural component of the *Bacillus anthracis* exosporium. *Mol Microbiol* 45:169–178
- Sylvestre P, Couture-Tosi E, Mock M (2003) Polymorphism in the collagen-like region of the *Bacillus anthracis* BclA protein leads to variation in exosporium filament length. *J Bacteriol* 185:1555–1563
- Thorne CB (1993) *Bacillus anthracis*. In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington DC, pp 113–124
- Tinsley E, Khan SA (2006) A novel FtsZ-like protein is involved in replication of the anthrax toxin-encoding pXO1 plasmid in *Bacillus anthracis*. *J Bacteriol* 188:2829–2835
- Tinsley E, Naqvi A, Bourgogne A, Koehler TM, Khan SA (2004) Isolation of a minireplicon of the virulence plasmid pXO2 of *Bacillus anthracis* and characterization of the plasmid-encoded RepS replication protein. *J Bacteriol* 186:2717–2723
- Uchida I, Sekizaki T, Hashimoto K, Terakado N (1985) Association of the encapsulation of *Bacillus anthracis* with a 60 megadalton plasmid. *J Gen Microbiol* 131:363–367
- Uchida I, Hornung JM, Thorne CB, Klimpel KR, Leppla SH (1993a) Cloning and characterization of a gene whose product is a *trans*-activator of anthrax toxin synthesis. *J Bacteriol* 175:5329–5338
- Uchida I, Makino S, Sasakawa C, Yoshikawa M, Sugimoto C, Terakado N (1993b) Identification of a novel gene, *dep*, associated with depolymerization of the capsular polymer in *Bacillus anthracis*. *Mol Microbiol* 9:487–496
- Uchida I, Makino S, Sekizaki T, Terakado N (1997) Cross-talk to the genes for *Bacillus anthracis* capsule synthesis by *atxA*, the gene encoding the *trans*-activator of anthrax toxin synthesis. *Mol Microbiol* 23:1229–1240
- Urushibata Y, Tokuyama S, Tahara Y (2002) Characterization of the *Bacillus subtilis* *ywsC* gene, involved in gamma-polyglutamic acid production. *J Bacteriol* 184:337–343
- Van der Auwera GA, Andrup L, Mahillon J (2005) Conjugative plasmid pAW63 brings new insights into the genesis of the *Bacillus anthracis* virulence plasmid pXO2 and of the *Bacillus thuringiensis* plasmid pBT9727. *BMC Genomics* 26:103
- Vietri NJ, Marrero R, Hoover TA, Welkos SL (1995) Identification and characterization of a *trans*-activator involved in the regulation of encapsulation by *Bacillus anthracis*. *Gene* 152:1–9
- Vodkin MH, Leppla SH (1983) Cloning of the protective antigen gene of *Bacillus anthracis*. *Cell* 34:693–697
- White AK, Hoch JA, Grynberg M, Godzik A, Perego M (2006) Sensor domains encoded in *Bacillus anthracis* virulence plasmids prevent sporulation by hijacking a sporulation sensor histidine kinase. *J Bacteriol* 188:6354–6360
- Wilcks A, Smidt L, Økstad OA, Kolstø AB, Mahillon J, Andrup L (1999) Replication mechanism and sequence analysis of the replicon of pAW63, a conjugative plasmid from *Bacillus thuringiensis*. *J Bacteriol* 181:3193–3200
- Williams RC, Rees ML, Jacobs MF, Pragai Z, Thwaite JE, Baillie LW, Emmerson PT, Harwood CR (2003) Production of *Bacillus anthracis* protective antigen is dependent on the extracellular chaperone, PrsA. *J Biol Chem* 278:18056–18062

Megaplasms in *Cupriavidus* Genus and Metal Resistance

Max Mergeay(✉), Sébastien Monchy, Paul Janssen, Rob Van Houdt, and Natalie Leys

Contents

1	Introduction: Plasmids with Heavy Metal Resistance	210
2	Megaplasms for Resistance to Heavy Metals	212
2.1	Other Biotopes Where Metal-Resistant <i>C. metallidurans</i> were Found	218
3	Plasmid pMOL28 Encoding for Nickel, Cobalt, Chromate, and Mercury Resistance is Related to pHG1 of the Facultative Chemolithotroph <i>C. eutrophus</i> H16 and to pRALTA of the Betarhizobium <i>C. taiwanensis</i>	218
3.1	Backbone Common to Various Megaplasms	219
3.2	The Enigmatic RHS Island	220
4	Plasmid pMOL30 Shares Many Orthologs with pBVIE01: A Plasmid of the Betarhizobium and Opportunistic Pathogen <i>Burkholderia vietnamiensis</i> G4.....	221
5	Most of the Metal Resistance and Response Genes in Both Plasmids are Gathered in Genomic Islands	222
6	Peculiar Mobile Genetic Elements in pMOL28 and pMOL30	224
6.1	Bona Fide IS Elements in pMOL28 and pMOL30	224
6.2	Mercury Transposons: Tn4378 and Tn4380	225
6.3	TnCme2	225
6.4	A Binome also Found in <i>B. vietnamiensis</i> G4	227
7	Metal Resistance Genes in pMOL28 and pMOL30	227
7.1	pMOL28	227
7.2	pMOL30	228
7.3	Organization of the Metal Resistance Genes in pMOL28 and pMOL30: Special Features	231
8	Transcriptomics	232
9	Plasmids and the Evolution of Metal Resistance Genes	233
	References	234

M. Mergeay, S. Monchy, P. Janssen, R. Van Houdt, and N. Leys
Molecular & Cellular Biology, Belgian Center for Nuclear Energy, SCK•CEN,
2400 MOL, Belgium
e-mail: mmergeay@sckcen.be

Abstract Megaplasms carrying genetic determinants for the response and the resistance to heavy metals are regularly found in bacteria belonging to the genus *Cupriavidus/Ralstonia* and isolated from industrial biotopes rich in heavy metals. The *C. metallidurans* strain CH34 and its representative plasmids pMOL28 (171 kb), which carries the resistance to mercury, chromate, nickel and cobalt, and pMOL30 (234 kb), which carries the resistance to cadmium, zinc, cobalt, lead, mercury and copper, were fully annotated. The plasmid pMOL28 contains a backbone that is quite similar to the backbone of plasmid pHG1 from *C. eutrophus* H16 and of pRALTA from *C. taiwanensis*. Three putative genomic islands were discovered on pMOL28. One of them, CMGI-28a, carries heavy metal resistance genes (*mer*, *cnr*, and *chr*) and is flanked by *IS1071* elements that have undergone further rearrangements stabilizing the island and its metal resistance determinants. The backbone of pMOL30 is related to the large plasmid pBVIE01 of *Burkholderia vietnamiensis* G4 with an especially high identity between the *parAB* genes. Plasmid pMOL30 contains two large putative genomic islands comprising most of the genes involved in the response or resistance to heavy metals: CMGI-30a with the *czc*, *pbr*, and *mer* genes and CMGI-30b with the *sil* and *cop* genes.

1 Introduction: Plasmids with Heavy Metal Resistance

Plasmid-borne resistance to heavy metals has been described or suggested since the late sixties. The first determinants for metal resistance were described in small or medium size plasmids (transposons) of *Staphylococcus aureus* and *Escherichia coli* (Novick and Roth 1968) containing resistance to mercury, cadmium, lead, bismuth, or arsenic (Hedges and Baumberg 1973; Mobley et al. 1983; Nakahara et al. 1977; Ni'Bhriain et al. 1983; Novick and Roth 1968; Peyru et al. 1969; Silver 1996; Silver and Misra 1984; Smith and Novick 1972; Summers and Silver 1972; Summers and Silver 1978). A copper-resistant plasmid was also found in an *E. coli* strain isolated from copper-fed pig manure (Tetaz and Luke 1983).

Strains with multiple resistance to heavy metals were isolated from industrial biotopes (sediments in a non-ferro metallurgical plant) with high content in zinc, lead, and cadmium (Mergeay et al. 1978; Mergeay et al. 1985) for a review see (Mergeay 2000) and (Mergeay et al. 2003). These strains carried megaplasms that contained genetic determinants for heavy metal resistance. In the representative strain CH34, which carries the megaplasms pMOL28 and pMOL30 (Fig. 1a), genes for metal resistance have been localized by plasmid curing, genetic transfer, molecular cloning and analysis of plasmid deletions and insertions. The *czcCBA* genes (resistance to Cd(II), Zn(II) and Co(II)) of pMOL30 (Nies et al. 1987; Nies 1995; Nies et al. 1989b; van der Lelie et al. 1997) and later the *cnrCBA* genes of pMOL28 (Grass et al. 2005; Grass et al. 2000; Liesegang et al. 1993; Nies et al. 1990a; Nies et al. 1989a; Siddiqui et al. 1989; Siddiqui et al. 1988; Tibazarwa et al. 2000) were the first metal-resistance genes that were identified as belonging to the RND (for Resistance Nodulation Division) family of tri-component efflux systems

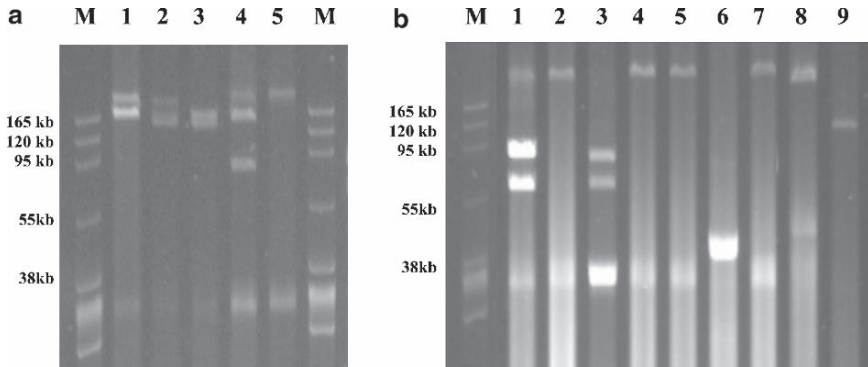


Fig. 1 Megaplasms present in *C. metallidurans* (a) and *R. pickettii* (b) strains isolated from spacecraft related biotopes. Panel A: *C. metallidurans* type strain CH34 (1), strains isolated from drinking water systems: 0103308-2 (2), 0200393-2 (3), 0502478-1 (4), NE12 (5); Panel B: *R. pickettii* strains isolated from surfaces of space robots 46V1 (1), 48V1 (2), 49V2 (3), and cooling water systems R21D (4), R21F (5), R21H (6), R42JB (7), 4RB8 (8), 6BR8 (9). M is BAC-Tracker Supercoiled DNA ladder (Epicentre)

implying an inner membrane pump, a membrane fusion protein, and an outer membrane protein (Dong and Mergeay 1994; Liesegang et al. 1993; Nies 1995; Nies et al. 1989b; Saier et al. 1994). Plasmid-borne determinants for resistance to chromate were thereafter soon identified in plasmid pMOL28 (Juhnke et al. 2002; Nies et al. 1990a; Nies et al. 1989a; Nies et al. 1998; Peitzsch et al. 1998). Later pMOL30 determinants for resistance to lead (Borremans et al. 2001) and copper (Monchy et al. 2006b) were also sequenced. The *ncc* operon (resistance to nickel, cobalt, and cadmium) on the large plasmid pTOM9 harbored by *Cupriavidus metallidurans* strain 31A (formerly known as *Alcaligenes xylosoxydans*) was the third metal resistance operon that was identified as belonging to the RND family and is closely related to *cnr*, although the *ncc*-mediated resistance to nickel is by far more efficient than the *cnr* one (Schmidt and Schlegel 1994; Schmidt et al. 1991). Another nickel-resistance determinant, *nre*, is closely linked to *ncc* on pTOM9 (Grass et al. 2001; Schmidt and Schlegel 1994; Schmidt et al. 1991). Both *ncc* and *nre* genes were found to be present in pMOL30 but were likely not functional. Today, pMOL28 (171 kb) and pMOL30 (234 kb) from strain CH34 have been fully sequenced and annotated and were also studied at the transcriptomic level after various challenges by heavy metals (Monchy et al. 2007).

This review will focus on the knowledge acquired on megaplasms pMOL28 and pMOL30 especially in regards to the general structure of the megaplasms, their possible affiliation to families, the role of mobile genetic elements (MGEs) in the mosaic architecture of the plasmids, and the specific genes involved in the responses to high concentrations of heavy metals.

A valuable approach was the comparison of the synteny between these plasmids and their genes with others plasmids and replicons of sequenced betaproteobacterial strains belonging to the genera *Cupriavidus*, *Ralstonia*, and *Burkholderia*. Synteny

studies, which were made mainly thanks to the MAGE Genoscope annotation platform (Vallenet et al. 2006), focuses on the conserved arrangements of genes and give also some clues about the origin of plasmids and even their evolution.

Before describing the plasmids pMOL28 and pMOL30, some comments on the taxonomy and the ecology of strain CH34 and closely related strains found in different geographical locations are in order as they are important for understanding the specialization of these plasmids.

2 Megaplasmids for Resistance to Heavy Metals

Strain CH34 was first assigned to the genus *Alcaligenes* (Mergeay et al. 1985), then to the genus *Ralstonia* and, finally, as type strain to the genus *Cupriavidus* (Brim et al. 1999; Goris et al. 2001; Mergeay et al. 2003; Vandamme and Coenye 2004; Vanechoutte et al. 2004; Yabuuchi et al. 1995). Various other betaproteobacteria were isolated from metal-rich biotopes. Most of them displayed a high level of resistance to heavy metals, their DNA hybridized with probes derived from strain CH34 for *czc*, *cnr*, *ncc*, or *mer* genes (Table 1) and carried megaplasmids (some of them being proficient for conjugation) containing these genes. But up to now, only the plasmids of *C. metallidurans* CH34 (pMOL28 and pMOL30) have been thoroughly studied, although they are poorly self-transmissible unlike other plasmids of the genus *Cupriavidus* (Diels et al. 1989, Diels et al. 1995b).

These strains were selected by their ability to grow in the presence of high concentrations of multiple heavy metals and to hybridize with *czc* or *ncc* probes. Bacteria were mainly assigned to the genus *C. metallidurans*, *C. basilensis*, and *C. campinensis* (Goris et al. 2001). Most of the biotopes of these strains contained high levels of heavy metals and often at concentrations higher than the minimum inhibitory concentration (MIC) observed on plates.

The size of the *Czc*⁺, *Cnr*⁺, or *Ncc*⁺ plasmids found in these strains ranged from 180 to 450 kb. Some of these plasmids are conjugative and display transfer frequencies much higher than those found in strain CH34 (Diels et al. 1995b).

Conjugative plasmids carrying metal resistance determinants may also easily be isolated from industrial biotopes by using “exogenous plasmid isolation” (Bale et al. 1988; Diels et al. 1993) where an appropriate plasmid-free recipient such as the plasmid-free derivative AE0104 of *C. metallidurans* CH34 is directly mixed with soil extracts. Metal-resistant transconjugants then obtained on selective plates (*C. Lodewyckx*, *Y. Wang*, and *M. Mergeay*, unpublished data) and further analyzed for plasmid presence, size, further transfer capabilities and hybridization with specific probes (*czc*, *cnr*, *mer*, *chr*, *cop*, *pbr*, or others) (*N. Leys*, *A. Provoost*, *A. Bossus*, and *M. Mergeay*, unpublished data). Strain AE0104 is an appropriate recipient due to apparent lack of efficient DNA restriction mechanisms. This method, although efficient and easy to use, is limited to the detection of plasmids compatible with a *Cupriavidus/Ralstonia*/betaproteobacteria background.

Table 1 List of plasmid genes involved in metal response or resistance

Plasmid	Gene Cluster	Resistance	Gene name	Size (a.a.)	(Putative) function	Island/Transposon	(Predicted) cellular localization		
pMOL28	<i>mer</i>	Hg(II)	<i>merR</i>	144	Regulatory protein	CMGI-28a/Tn4378	Cytoplasmic		
			<i>merT</i>	116	Mercury transport protein	CMGI-28a/Tn4378	Inner membrane		
			<i>merP</i>	87	Mercury captation protein	CMGI-28a/Tn4378	Periplasmic		
			<i>merA</i>	561	Mercuric reductase	CMGI-28a/Tn4378	Cytoplasmic		
			<i>merD</i>	121	Putative regulatory protein	CMGI-28a/Tn4378	Cytoplasmic		
			<i>merE</i>	78	?	CMGI-28a/Tn4378	?		
			<i>wrf-2</i>	329	Putatively involved in the transposition	CMGI-28a/Tn4378	Cytoplasmic		
			<i>tnpR</i>	186	Resolvase from Tn3 transposon family	CMGI-28a/Tn4378	Cytoplasmic		
			<i>tnpA</i>	989	Transposase from Tn3 transposon family	CMGI-28a/Tn4378	Cytoplasmic		
			<i>chr</i>	CrO ₄ ²⁻	<i>chrF1</i>	140	Regulatory protein	CMGI-28a	Cytoplasmic
					<i>chrE</i>	113	Processing of chromium-glutathione-complexes	CMGI-28a	Cytoplasmic
					<i>chrC</i>	197	Iron superoxide dismutase (Fe-SOD)	CMGI-28a	Cytoplasmic
					<i>chrA1</i>	374	Efflux pump	CMGI-28a	Inner membrane
			<i>cnr</i>	Co(II), Ni(II)	<i>chrB1</i>	297	Regulatory protein	CMGI-28a	Cytoplasmic
<i>chrI</i>	164	Regulatory protein (repressible by various metals)			CMGI-28a	Cytoplasmic			
<i>cnrY</i>	95	Regulatory protein (antisigma factor)			CMGI-28a	Inner membrane			
<i>cnrX</i>	148	Regulatory protein (antisigma)			CMGI-28a	Inner membrane			
<i>cnrH</i>	191	Sigma factor family ECF			CMGI-28a	Cytoplasmic			
<i>cnrC</i>	418	Three components proton antiporter efflux system			CMGI-28a	Outer membrane			
<i>cnrB</i>	395	Three components proton antiporter efflux system			CMGI-28a	Membrane fusion protein			
<i>cnrA</i>	1076	Cation efflux pump: three components proton antiporter efflux system			CMGI-28a	Inner membrane			
<i>cnrT</i>	351	Cation Diffusion Facilitator			CMGI-28a	Outer membrane			

(continued)

Table 1 (continued)

Plasmid	Gene Cluster	Resistance	Gene name	Size	(Putative) function	Island/Transposon	(Predicted) cellular localization
pMOL30	<i>mer</i>	Hg(II)	<i>merP</i>	88	MerP, periplasmic protein (functional?)	CMGI-30a	Periplasm
			<i>merT</i>	115	MerT, inner membrane protein (functional?)	CMGI-30a	Inner membrane
			<i>merR</i>	132	MerR, regulatory protein (functional)	CMGI-30a	Cytoplasm
	<i>czc</i>	Co(II), Zn(II), Cd(II)	<i>czcM</i>	152	Mg(II) transport ATPase	CMGI-30a	Inner membrane
			<i>czcN</i>	216	regulation	CMGI-30a	Inner membrane
			<i>czcI</i>	115	regulation	CMGI-30a	Periplasm
			<i>czcC</i>	377	Three components proton antiporter efflux system	CMGI-30a	Outer membrane
			<i>czcB</i>	520	Three components proton antiporter efflux system	CMGI-30a	Membrane fusion protein
			<i>czcA</i>	1063	Cation efflux pump: three components proton antiporter efflux system	CMGI-30a	Inner membrane
			<i>czcD</i>	316	Cation Diffusion Facilitator (CDF)	CMGI-30a	Inner membrane
			<i>czcR</i>	225	Regulator from the two components sensor/regulator family	CMGI-30a	Cytoplasm
			<i>czcS</i>	476	Sensor from the two components sensor/regulator family	CMGI-30a	Inner membrane
	<i>pbr</i>			<i>czcE</i>	131	Copper-binding protein	CMGI-30a
			<i>czcJ</i>	92	Regulator of the Cop Q/Cyco/MerQ family	CMGI-30a	Periplasm
			<i>ompP</i>	355	Porin	CMGI-30a	Outer membrane
			<i>czcP</i>	722	P-Type ATPase involved in Cd(II), Zn(II) efflux	CMGI-30a	Inner membrane
		Pb(II)	<i>pbrD</i>	241	Pb(II) binding protein	CMGI-30a	Cytoplasm
			<i>pbrB/C</i>	358	PbrB lipoprotein phosphatase/PbrC prelipoprotein signal peptidase	CMGI-30a	PbrB membrane associated periplasmic, PbrC inner membrane

Table 1 (continued)

Plasmid	Gene Cluster	Resistance	Gene name	Size (a.a.)	(Putative) function	Island/Transposon	(Predicted) cellular localization
			<i>pbrA</i>	799	Cation efflux P-type ATPase	CMGI-30a	Inner membrane
			<i>pbrR</i>	145	Regulatory protein	CMGI-30a	Cytoplasm
			<i>pbrT</i>	642	Pb(II)-uptake transport protein	CMGI-30a	Inner membrane
			<i>pbrU</i>	312 + 125	Permease of the major facilitator superfamily	CMGI-30a	Inner membrane
	<i>mer</i>	Hg(II)	<i>tnpA</i>	988	Transposase from Tn3 transposon family	Tn4380-CMGI-30a	Cytoplasm
			<i>tnpR</i>	182	putative resolvase from Tn3 transposon family	Tn4380-CMGI-30a	Cytoplasm
			<i>urf-2</i>	333	Putatively involved in the transposition	Tn4380-CMGI-30a	Cytoplasm
			<i>merE</i>	78	?	Tn4380-CMGI-30a	Cytoplasm?
			<i>merD</i>	121	Putative regulatory protein	Tn4380-CMGI-30a	Cytoplasm
			<i>merA</i>	561	mercuric reductase	Tn4380-CMGI-30a	Cytoplasm
			<i>merP</i>	91	Mercury captation protein	Tn4380-CMGI-30a	Periplasm
			<i>merT</i>	116	Mercury transport protein	Tn4380-CMGI-30a	Inner membrane
			<i>merR</i>	144	regulatory protein	Tn4380-CMGI-30a	Cytoplasm
	<i>ncc</i>	No phenotype	<i>nccC</i>	418	Three components proton antiporter efflux system	CMGI-30b	Outer membrane
			<i>nccB1</i>	93	Three components proton antiporter efflux system(fragment)	CMGI-30b	Membrane fusion protein
			<i>nccB2</i>	294	Three components proton antiporter efflux system (fragment)	CMGI-30b	Membrane fusion protein
			<i>nccA</i>	1076	Three components proton antiporter efflux system	CMGI-30b	Inner membrane efflux pump
			<i>nreB</i>	408	Transport (major facilitator superfamily)	CMGI-30b	Membrane
			<i>mmrQ</i>	81	Regulator of the cop Q/czc/mmrQ family	CMGI-30b	Periplasm
	<i>sil</i>	Ag(I), Cu(II)	<i>silA</i>	1056	Three components proton antiporter efflux system	CMGI-30b	Inner membrane

Table 1 (continued)

Plasmid	Gene Cluster	Resistance	Gene name	Size (a.a.)	(Putative) function	Island/Transposon	(Predicted) cellular localization
			<i>silB</i>	521	Three components proton antiporter efflux system	CMGI-30b	Membrane fusion protein
			<i>silC</i>	435	Three components proton antiporter efflux system	CMGI-30b	Outer membrane
			<i>silD</i>	131	conserved hypothetical protein	CMGI-30b	Periplasm
			<i>orf178</i>	131	conserved hypothetical protein (strongly induced by Cd, Cu, Ni & Zn)	CMGI-30b	Periplasm
<i>cop</i>		Cu(I), Cu(II)	<i>copW</i>		Unknown	CMGI-30b	Periplasm
			<i>copE</i>	211	Unknown	CMGI-30b	Cytoplasm
			<i>copH</i>	153	Cu(II) binding as a Cu(II) sensor, homolog to czcE	CMGI-30b	Periplasm
			<i>copQ</i>	82	regulator of the <i>cop Q/czcJ/mmrQ</i> family	CMGI-30b	Periplasm
			<i>copL</i>	421	Unknown	CMGI-30b	Cytoplasm?
			<i>copO</i>	63	Putative conserved chaperone	CMGI-30b	Inner membrane
			<i>copF</i>	805	P-Type ATPase involved in Cu(I) efflux	CMGI-30b	Inner membrane
			<i>copG</i>	137	Involved in survival at the MIC?	CMGI-30b	Periplasm
			<i>copJ</i>	174	Putative cytochrome c protein	CMGI-30b	Periplasm
			<i>copI</i>	158	Putative oxido-reductase protein	CMGI-30b	Periplasm
	<i>copDI</i>	305	Canal involved in Cu(II)/Cu(I) uptake	CMGI-30b	Inner membrane		
	<i>copCI</i>	128	Cu(II) binding protein	CMGI-30b	Periplasm		
	<i>copBI</i>	495	Cu(I)-binding protein	CMGI-30b	Outer membrane		
	<i>copAI</i>	614	Multi-Cu(II) oxidase		CMGI-30b	protein Cytoplasm/ periplasm	
	<i>copRI</i>	228	Regulator from the two components sensor/regulator family		CMGI-30b	Cytoplasm	
	<i>copSI</i>	463	Sensor from the two components sensor/regulator family		CMGI-30b	Inner membrane	

Table 1 (continued)

Plasmid	Gene Cluster	Resistance	Gene name	Size (a.a.)	(Putative) function	Island/Transposon	(Predicted) cellular localization
			<i>copN</i>	164	Unknown (contains a SAM radical)	CMGI-30b	Cytoplasm
			<i>copK</i>	94	Cu(I) binding protein	CMGI-30b	periplasm
			<i>copM</i>	136	Unknown	CMGI-30b	Periplasm or inner membrane associated
			<i>copT</i>	261	Putative cytochrome, similar to the cytochrome domain of PbrT	CMGI-30b	Periplasm
			<i>copV</i>	117	Unknown	CMGI-30b	Cytoplasm?
<i>gtr</i>	No phenotype		<i>gtrBI</i>	373	Glycosyltransferase involved in cell wall biogenesis	CMGI-30a	Inner membrane
			<i>gtrAJ</i>	127	Glycosyltransferase involved in cell wall biogenesis	CMGI-30a	Inner membrane
			<i>gtrMI</i>	541	Glycosyltransferase, family 39	CMGI-30a	Inner membrane
			<i>gtrM3</i>	561	Glycosyltransferase, family 39	CMGI-30a	Inner membrane
			<i>gtrM2</i>	541	Glycosyltransferase, family 39	CMGI-30b	Inner membrane
			<i>gtrA2</i>	127	Glycosyltransferase involved in cell wall biogenesis	CMGI-30b	Inner membrane
			<i>gtrB2</i>	367	Glycosyltransferase involved in cell wall biogenesis	CMGI-30b	Inner membrane
			<i>ompP2</i>	355	Porin; inducible by Cu(II)	CMGI-30b	Outer membrane
			<i>orf-231</i>	528	Hypothetical protein metal-inducible	CMGI-30b	?

(1) number of amino-acids

2.1 Other Biotopes Where Metal-Resistant *C. metallidurans* were Found

Cupriavidus and *Ralstonia* appear to be (key) players in the natural microbial contamination of ultra-clean spacecraft assembly rooms and in spacecraft atmosphere and water reservoirs. Several reports have shown that the air of such spacecraft assembly rooms often contain *Cupriavidus* bacteria (La Duc et al. 2003). Furthermore, *R. pickettii* strains have also been found prior-to-flight on surfaces of space robots (La Duc et al. 2003) and even in an in-flight space vehicle. *R. pickettii* and *C. metallidurans* strains have been found in the cooling (Bernardini et al. 2005, Pierson et al. unpublished, Pyle et al. unpublished) and drinking water systems of the space station (Bruce et al. unpublished). The study of the survival and proliferation strategies of these *Cupriavidus* and *Ralstonia* bacteria in these biotopes is of importance to improve contamination prevention, monitoring, and disinfection tools for manned space capsules in the future.

Our analysis showed that all the *Cupriavidus* and *Ralstonia* clean-room isolates are able to use a wide variety of substrates as carbon sources, including organic cleaning solvents such as ethanol and acetone (N. Leys et al. unpublished). Some of the test strains were able to form biofilms on plastic and metal materials used as spacecraft structural materials, which could provide a nutritional and resistance advantage. In addition, all isolates were showed to have accumulated moderate resistances to an extraordinary collection of physical and chemical antimicrobial agents (N. Leys et al. unpublished). All strains were found to be resistant to a whole range of heavy metals, and *czc* determinants were identified (by hybridization, phenotypic analysis, or gene transfer) (N. Leys et al. unpublished). Moreover, the resistance properties are probably encoded in “mobile” DNA fragments, as several megaplas-mids were detected in these strains (Fig. 1). This phenotype is very similar to that of *C. metallidurans* CH34. The accumulation of a wide variety of moderate resistances has given these bacteria the advantage to survive in the poor and repeatedly changing acute and chronically “life-threatening” stresses more easily and to adapt to new harsh man-made environments such as ultra-clean spacecraft biotopes.

3 Plasmid pMOL28 Encoding for Nickel, Cobalt, Chromate, and Mercury Resistance is Related to pHG1 of the Facultative Chemolithotroph *C. eutrophus* H16 and to pRALTA of the Betarhizobium *C. taiwanensis*

Plasmid pMOL28 (Monchy et al. 2007) is easily distinguished by its capacity to confer resistance to nickel (Liesegang et al. 1993) and to chromate (Nies et al. 1990a 1993). pMOL28 contains a quite complete set of conjugative genes (similar to that of pHG1 from *C. eutrophus* H16, which is self-transferring at high frequency), although it is self-transferred at a low frequency. A derivative of pMOL28, pMOL50 that

is self-transferred at high frequency was described and even used for the chromosomal mapping in *C. metallidurans* CH34 (Sadouk and Mergeay 1993). However, the molecular nature of the event that gave rise to pMOL50 is still elusive.

3.1 Backbone Common to Various Megaplasmid

The recent annotation (Monchy et al. 2007) of pMOL28 shows an extensive synteny between pMOL28 and two other (much larger) plasmids or megaplasmid: pHG1 from the facultative chemolithoautotrophic bacterium *C. eutrophus* H16 (Schwartz et al. 2003) and pRALTA (root symbiosis and nitrogen fixation) from the betarhizobium *C. taiwanensis* (Amadou et al. 2008; Chen et al. 2001; Chen et al. 2003). This illustrates how the plasmids contribute to defining life styles of very similar bacteria colonizing very different niches. The synteny starts from the *parAB repA* region (replication functions) and extends counterclockwise for 80 kb, (almost half of pMOL28), to within a few kilobases of the metal island (Fig. 2). The last gene of the synteny is *sbcC* involved in DNA repair or maintenance of the neighboring genes. In fact, a fourth megaplasmid shares this synteny as well: pBVIE02 (266 kbp; GC content: 61%; accession number NC009227) of *Burkholderia vietnamiensis* G4.

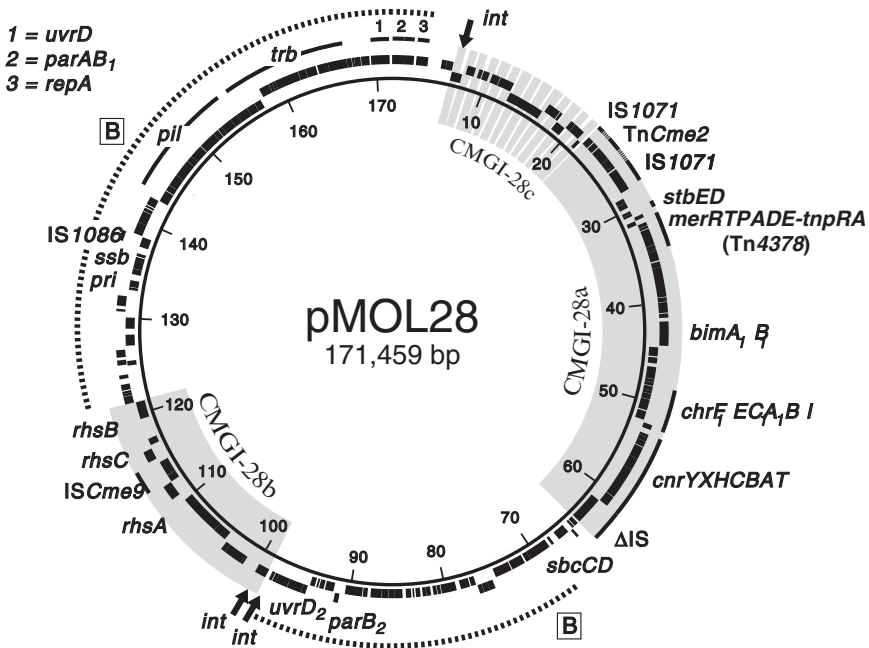


Fig. 2 Map of the large plasmid pMOL28 present in *C. metallidurans* CH34. Genomic islands are highlighted in grey; plasmid backbone (B) in synteny with pHG1 is marked by dotted line

There are 55 genes of pBVIE02 that have orthologs in pHG1 and pMOL28 (and likely with pRALTA), although the percentage of identity is much lower than between pMOL28 and pHG1 and some gene blocks are rearranged. Plasmid pBVIE02 also contains RND genes with some similarity to *cnr* and *czc*.

Thus there is clearly a family of megaplasms in related betaproteobacteria that share a substantial number of common genes. These genes constitute a backbone including basic functions for replication, maintenance, and transfer. Yet the synteny between pMOL28 and the three other megaplasms is interrupted by a very peculiar 23 kb region (CMGI-28b) that will be discussed below and for which no equivalent has been observed in the other plasmids or in pMOL30. Besides the above mentioned *parAB repA* genes located at one extremity and the *sbcC* gene located at the other extremity, the pMOL28 backbone contains also *uvrD*, *trb* (transfer), *pil* (piliation) genes, a second region involved in replication as well as some genes of unknown function (Fig. 2). The percentage of identity between the backbone genes of pMOL28 and pHG1 or pRALTA is quite variable: up to 80%, but very low for *parAB repA*. This would strongly suggest that pMOL28 is fully compatible (i.e. may stably coexist) with pHG1 and pRALTA. In this context, it is interesting that pHG1 could be transferred to and stably maintained in derivatives of *C. metallidurans* carrying pMOL28 (Gerstenberg et al. 1982).

3.2 The Enigmatic RHS Island

Figure 2 shows that a 23 kb (13 genes) fragment (called CMGI-28b) is inserted in the PMOL28 backbone near the *uor D2* gene. This insertion contains two genes encoding for a site-specific recombinase at one extremity and three *rhs*-like genes, which are rich in YD motifs but whose function is unknown, at the other extremity. RHS elements are proteins having nonessential functions believed to play an important role in the cell's natural ecology. The protein sequence derived from the first described *rhs* gene includes a highly conserved 141-kDa domain containing multiple, 22-residues, tandem repeats, followed by divergent C-terminal domains (Feulner et al. 1990; Hill et al. 1994). The most strongly conserved motif of the 22-residues repeat is a tyrosine-aspartate (YD) dipeptide.

The *rhs*-rich regions like those found in *C. metallidurans* CH34, *R. solanacearum* (megaplasmsid) or *Shigella sonnei* and various enterics share the following characteristics: the presence of three to five *rhs* or *rhs*-like genes with YD motifs (tyrosyl-aspartyl residues), the presence of insertion sequence elements (IS) or recombinase genes, the presence of genes for *rhs* accompanying proteins and the presence of a few genes of unknown function. Rhs proteins are clearly associated with the outer membrane and some *rhs* elements encode for carbohydrate receptors. Would the association between MGEs and recombinase genes with outer membrane determinants suggest that environmental signals turn on either genomic rearrangements or even the acquisition of MGEs carrying genes for adaptation to changing environments? Would the RHS (for Recombination Hot Spot) region not be a clue to explain the changes

that modified the poorly self-transferable pMOL28 into the highly conjugative pMOL50? (Mergeay 2000; Sadouk and Mergeay 1993; Taghavi et al. 1997).

4 Plasmid pMOL30 Shares Many Orthologs with pBVIE01: A Plasmid of the Betarhizobium and Opportunistic Pathogen *Burkholderia vietnamiensis* G4

In contrast to pMOL28, the affiliation of pMOL30 to a family of plasmids sharing the same backbone was not so evident. Yet the syntenic approach to the annotation of pMOL30 revealed many orthologs with plasmid pBVIE01, the largest plasmid of *B. vietnamiensis* G4 (O’Sullivan et al. 2007) (398 kb; 58% GC; accession number NC009230), for which most of the genes are hypothetical and which has no clear phenotype. Around 45 orthologs with pBVIE01 were found in the two pMOL30 regions situated between the genomic islands of pMOL30 (Fig. 3). High

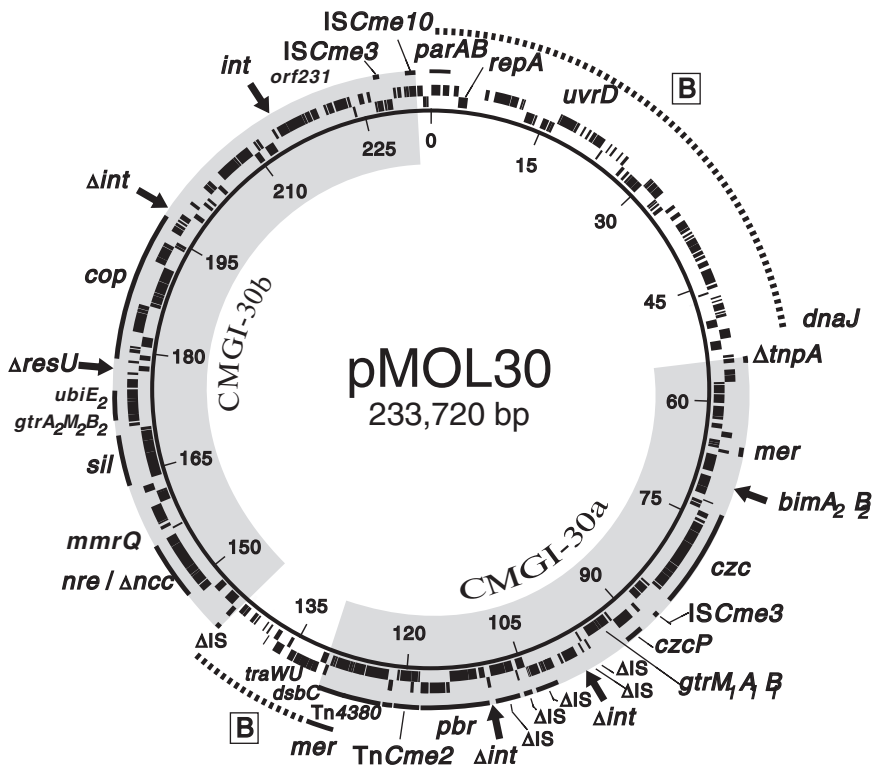


Fig. 3 Map of the large plasmid pMOL30 (233720 bp) present in *C. metallidurans* CH34. Genomic islands are highlighted in grey, and plasmid backbone (B) in synteny with pBVIE01 is marked by dotted line

identities (above 80%) were found for the replication region (*parAB repA uvrD* genes) and the *traW traU trbC dsbC* genes. This similarity for the replication genes is striking, suggesting that the plasmids pMOL30 and pBVIE01 may well be incompatible, although they are quite different for most of their genetic content. The main synteny region between pBVIE01 and pMOL30 spans 65 kb and contains 28 putative orthologs somewhat interspersed with other genes mostly of unknown function. The order of these genes is conserved in pBVIE01 (Fig. 3). This would suggest the existence of a backbone principally based on the backbone of an ancestor of pBVIE01 where pBVIE01-like genes would have been progressively removed and replaced by the genes that we see now on pMOL30.

The backbone of pMOL30 is roughly defined by the regions that are somewhat syntenic with pBVIE01 and contains the genes for partition and replication, as mentioned earlier. These genes are very similar to their pBVIE01 orthologs (Fig. 3). In contrast, there are very few genes or ORFs involved in plasmid transfer. Only three bona fide genes that are located between the two islands are present, which are also closely related to their pBVIE01 counterparts. It is quite likely that pMOL30 is not self-transmissible and that the low-frequency transfer of pMOL30 into other hosts observed is mediated via conjugation genes located on other replicons. In this respect, it should be mentioned that the chromosome of *C. metallidurans* CH34 contains two sets of *traF traR traG trb* conjugative genes nested in the genomic islands CMGI-2 and CMGI-3 (CMGI for *Cupriavidus metallidurans* Genomic Island), which belong to the Tn4371 family. These genes are highly identical to the Ti/IncP1 plasmid conjugative genes (Toussaint et al. 2003). The IncP1 plasmids are broad host range plasmids that can efficiently mobilize non-conjugative plasmids and even chromosomal genes

The regions of the pMOL30 backbone that are not syntenic with pBVIE01 contain many small ORFs encoding for “unique” proteins (hypothetical proteins with no counterpart in the databases). Most of these ORFs seem to be partial or correspond to deleted genes or gene remnants.

5 Most of the Metal Resistance and Response Genes in Both Plasmids are Gathered in Genomic Islands

A glance at the maps of pMOL28 and pMOL30 (Figs. 2 and 3) shows that the metal resistance genes – and also more generally all the genes overexpressed in the presence of heavy metals – are clustered in regions that are clearly separated from the backbone of the plasmids. It is tempting to check whether these regions could be a kind of genomic islands that jumped onto the plasmid backbone.

The term genomic island (GI) was first coined to designate genomic regions similar to the pathogenicity islands (PAI) but without recognizable pathogenicity determinants. When islands containing aromatic degradation or plant symbiosis genes were identified, they were called catabolic and symbiotic islands, respectively (reviewed in (Hacker et al. 2004)). The chromosome of *C. metallidurans*

contains also a 105 kb genomic island that is almost 100% identical with the PAGI2-C island found in *P. aeruginosa*, isolated from cystic fibrosis patients (Larbig et al. 2002).

GIs are also commonly defined as clusters of genes in prokaryotic genomes with one or a set of following properties:

- Presence of a site-specific recombinase gene and an adjacent tRNA gene at one extremity
- Presence of flanking insertion sequence elements
- A base composition and/or phylogeny which differ from the bulk of the genome, indicating their foreign origin and, hence, their acquisition through horizontal transfer
- A higher content in hypothetical genes than the neighboring regions
- Clustering of genes characteristic of MGEs: recombinase genes, IS elements
- The conservation of GI between different unrelated hosts
- A high concentration of genes involved in resistance, catabolism, or other specialized metabolic pathways

Most often, however, as for most PAIs, intercellular mobility has not been demonstrated experimentally. Under the above broad definition, the term GI covers a spectrum of integrated MGEs that are not well-defined, which may be defective or have lost their intercellular mobility determinants and are inferred to be or to have been mobile (or to be mobilized in trans by a proficient MGE).

According to this broad definition, three islands may be defined in pMOL28 and two in pMOL30, which are highlighted in grey in [Figs. 2 and 3](#). These five GIs are clearly distinguished from the plasmid backbones.

The genomic island CMGI-28a ([Fig. 2; Table 1](#)) of pMOL28 contains the genetic determinants for resistance to mercury (Tn4378), chromate (*chrFECABI*), and cobalt/nickel (*cnrYXHCBAT*). This island is delimited by two derivatives of the Tn3 family (to which *IS1071* belongs) derivatives. At the *cnr* extremity, the derivative is partially deleted while at the other extremity *IS1071* is inactivated by Tn*Cme2*, which will be described below. CMGI-28b (the *rhs* island) was described above.

The putative CMGI-28c spans 15 kb (17 ORFs) starting at an *int* gene and finishing at the contact with CMGI-28a. Most genes of CMGI-28c are hypothetical and do not have any counterpart in a variety of plasmid backbones. It may be possible that CMGI-28c and 28a would constitute one island.

The island CMGI-30a of pMOL30 ([Table 1; Fig. 3](#)) appears to be delimited by a truncated *tnpA* of Tn4380 at one side and by a full-length Tn4380 at the other side: CMGI30a shelters some *mer* genes, all the *czc* genes, the *pbr* genes, and the *mer* genes from Tn4380, respectively. Nearby the *czc* genes, there are also some membrane *gtr* genes that are likewise overexpressed in the presence of heavy metals (Monchy et al. 2007). This association between RND resistance determinants and membrane repair genes in the same cluster is also found in CMGI-30b.

The island CMGI-30b is separated from CMGI-30a ([Fig. 3](#)) by a short stretch of the pMOL30 backbone (with putative orthologs also found in pBVIE01) and is flanked by IS elements or derivatives like the other plasmid “metallic” islands. At

one side, a deleted IS₃-like element is recognizable, at the other side lie *ISRme3* and *ISRme10*.

CMGI-30b contains a block with *sil* and *ncc-nre* genes (Schmidt and Schlegel 1994) as well as *gtr* genes involved in membrane repair or maintenance just as in the *czc* region. Both blocks also contain *mmrQ* and *czcJ*, similar ORFs that are strongly induced by heavy metals. However, the *ncc/nre* genes seem to be partially deleted or damaged since no nickel resistance phenotype is associated with these genes.

The *cop* genes (at least 21 ORFs) (Table 1) also belong to CMGI-30b. They make up a compact block flanked by genes encoding for deleted tyrosine recombinases (integrases) (*resU* and *int*).

The rest of the putative island CMGI-30b is enigmatic: it extends from the position 195 kb to 230 kb and is divided into two parts joined by an *int* gene (encoding for a tyrosine recombinase). The first part contains some (deleted) catabolic genes but most of the genes in both parts encode for hypothetical proteins (some conserved). There is apparently no syntenic evidence that these genes are involved in the response to heavy metals. Yet *orf231* (Table 1) is overexpressed in presence of metals (Cu(II) 5x, Ni(II) 6x, and Zn (II) 7x induction). The corresponding protein (Orf231) is so far unique to pMOL30 (Monchy et al. 2007).

CMGI-30a and CMGI-30b islands contain various genes encoding for tyrosine recombinases (*integrases*) but most of them appear to be partial. In fact, these genes seem to delimitate subregions or putative “nested islands” (A. Toussaint, pers. comm.) like the *pbr* region and the *czc gtr* “island” in CMGI-30a and the *cop* cluster in CMGI-30b. A second “nested island” in CMGI-30b seems to shelter RND genes and other genes that have a counterpart in the *czc* “nested island” of CMGI-30a.

This would suggest a kind of recruitment of heavy metal response or resistance genes from different hosts into CMGI-30a and CMGI-30b via small GIs or structures reminiscent of integrons. If this hypothesis is correct, the deletions in some *int* genes helped to stabilize the recruited genes after their acquisition.

6 Peculiar Mobile Genetic Elements in pMOL28 and pMOL30

6.1 Bona Fide IS Elements in pMOL28 and pMOL30

Plasmid pMOL28 contains, besides *IS1071* (inactivated by *TnCme2*) and other IS elements located in the *rhs* region, another IS: *IS1086* (IS₃₀ family) (Dong et al. 1992) (Fig. 2) that is lacking in pMOL50, the highly conjugative derivative of pMOL28 (Taghavi et al. 1997)

Plasmid pMOL30 carries two copies of *ISRme3* (one located between *czcP* and the main *czc* cluster, and the other at the end of the putative island CMGI-30b) and one copy of *ISRme10*. It also contains a substantial number of deleted IS ORFs related to the IS₃ family that lie in CMGI-30a. Some of them lie in a small cluster

flanked by deleted “*int*” (tyrosine recombinase) genes. A few of these ORFs are also highly inducible by heavy metals (see below).

6.2 Mercury Transposons: Tn4378 and Tn4380

Mercury transposons Tn4378 and Tn4380 (Diels et al. 1985) are very similar, and it is not possible to discriminate their resistance genes (quasi 100% of identity). Their *tnpA tnpR* genes are also nearly identical. They are bounded by 38-bp inverted repeats (IR) and clearly belong to the Tn501 family.

Tn4380 seems to have played a role in the construction of the CMGI-30a island and so helped to gather heavy metal resistance determinants in this island.

The mercury resistance genes *merRTPADEorf2* have a narrow spectrum of detoxification of mercury compounds (Liebert et al. 1999): mercury chloride, merbromin, and thiomersal. Genes *merP*, *merD*, and *merA* of Tn4378 were further studied at the structural level (Champier et al. 2004; Rossy et al. 2004; Serre et al. 2004)

The transposon core *tnpA tnpR* of Tn4380 is surprisingly also observed at two copies (100% identity) in the chromosome 2 but without the *mer* genes. These two transposons, which were designated TnCme3, are located in opposite orientations in the megaplasmid and separated by 143 kb. Both are associated with four genes encoding for proteins that are not classically associated with transposons.

The conjugative transfer of plasmid pMOL28 of *C. metallidurans* CH34 can be enhanced to a frequency of circa 1 transfer per 5×10^5 acceptor cells through cotransfer with plasmid RP4 and the formation of a cointegrate via the transposition of Tn4378 (Diels et al. 1985) (Fig. 4), while apparently the transfer of pMOL30 can be enhanced to a frequency of circa 1 transfer per 10^3 acceptor cells with RP4 and the formation of a cointegrate via the transposition of Tn4380. The corresponding frequencies of transposition of Tn4378 and Tn4380 on plasmid RP4 and 3.1 E-4 and 2.3 E-2 respectively.

6.3 TnCme2

The TnCme2 structure is found in both plasmids and in both cases it knocks out a gene, *IS1071* in pMOL28 (Fig. 2) and *pbrU* in pMOL30 (Fig. 3). TnCme2 is a real transposon, as it contains inverted repeats and it seems to generate a direct repeat of 5–6 bp of the target sequence (Fig. 5).

TnCme2 contains four genes: a transposase (*tnmA*), two other genes (*tnmB* and *tnmC*) encoding hypothetical proteins (but some hints suggest that the corresponding proteins act on DNA) and a gene (*gspA*) encoding for a major protein of type II secretion system, GspA (also called LstA). At the moment, there is no clue why such a gene is associated with a transposon and if it confers a special phenotype.

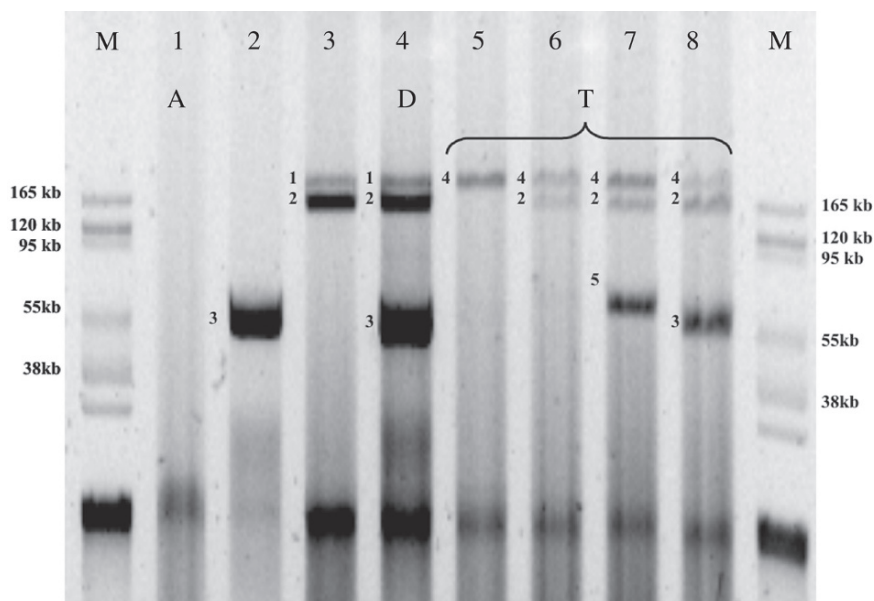


Fig. 4 Enhanced conjugative transfer of plasmid pMOL28 of *C. metallidurans* CH34^T through the formation of a cointegrate with plasmid RP4. A donor strain AE136 (D, lane 4) was constructed by supplementing a methionine, leucine, and tryptophan auxotroph derivative of *C. metallidurans* CH34^T strain containing plasmids pMOL30 (¹) and pMOL28 (coding for *cnr* Ni²⁺ resistance) (²) (lane 3), with the plasmid RP4 (³) from *E. coli* strain CM104 (lane 2). This donor strain was conjugated with a plasmid free acceptor strain AE104 (A, lane 1), also derived from of *C. metallidurans* CH34^T. By selection on minimal agar medium containing 1mM of free Ni²⁺ ions, different transconjugants (T, lanes 5–8) were obtained, where the original pMOL28 (²) and RP4 (³) plasmid can be detected as well as the cointegrate of the two plasmids pMOL28-RP4 (⁴) and the enlarged RP4 plasmid resolved from the cointegrate containing a copy of the pMOL28 mercury transposon Tn4378 (⁵). These Nickel-resistant transconjugants (lanes 5–8) were checked for the absence of pMOL30 plasmid and resistance markers (Zn(II), Cd(II) and Cu(II)). M is BAC-Tracker Supercoiled DNA ladder (Epicentre)

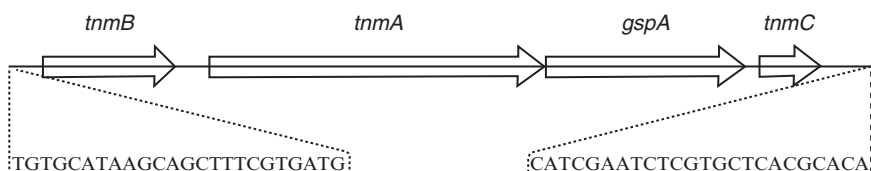


Fig. 5 Structure of Tn*Cme2*, which contains a transposase (*tnmA*), two other genes (*tnmB* and *tnmC*) encoding hypothetical proteins and a gene (*gspA*) encoding for a major type II secretion system component GspA. The sequence of the inverted repeats is shown

As it is the case for Tn*Cme3* (see above in the section about mercury transposons), it seems that transposons are associated with genes that do not necessarily provide a straightforward selective advantage (e.g., antibiotic and metal resistance, degradation of recalcitrant carbon sources).

Furthermore, there are ten other copies of Tn*Cme2* in the genome of *C. metallidurans* CH34, 7 in the main chromosome and 3 in the chromosome 2. In at least three cases, Tn*Cme2* has knocked out key genes of integrated islands in these replicons, just as in the case of the insertion of Tn*Cme2* in pMOL28 at one extremity (IS1071) of the CMGI-28 island. Therefore, Tn*Cme2* seems to be specialized in stabilizing the genomic islands acquired or captured by *C. metallidurans* CH34.

6.4 A Binome also Found in *B. vietnamiensis* G4

The MAGE-mediated syntenic annotation of pMOL30 with the genome of *B. vietnamiensis* G4 revealed a binome of two genes: one tyrosine recombinase gene and one gene of unknown function that are highly conserved in pMOL28 and in pMOL30. These genes, for the time being designated as *bimA1B1* and *bimA2B2* (*bim* for binome involved in mobility), respectively, for pMOL28 and pMOL30, are also transcriptionally coupled and located in a genomic island near a Tn4380 transposase (total or partial). The interesting thing is that the *B. vietnamiensis* G4 equivalent is present in four copies (in four different replicons) perfectly conserved and associated with three other genes to form a specific MGE in *B. vietnamiensis* G4, called Tn*Bvie1*. The other genes for Tn*Bvie1* are a gene for nickel uptake, a transposase and a gene of unknown function. The role of this MGE and of the *bim* genes is unknown.

7 Metal Resistance Genes in pMOL28 and pMOL30

7.1 pMOL28

The putative genomic island CMGI-28a (Fig. 2; Table 1) contains the *mer*, *cnr* (resistance to cobalt and nickel) (Liesegang et al. 1993), and *chr* operons. The *cnrYXHCBAT* genes encode two resistance mechanisms: a RND system of cation/proton antiporter efflux mediated by CnrC, CnrB, and CnrA with the support of CnrT, a cation diffusion facilitator (CDF) (Nies 2003). This association RND/CDF was also observed in the *czc* region. As shown in Fig. 6, the regulation is mainly governed by *cnrH* encoding a sigma70_{ECF} protein (Grass et al. 2005; Grosse et al. 2007; Lonetto et al. 1994) and the antisigma factors CnrY and CnrX (Collard et al. 1993; Grass et al. 2000; Kim et al. 2008; Tibazarwa et al. 2000).

Resistance to chromate is closely linked to the *cnr* genes and includes six genes *chrFECBAI* (Juhnke et al. 2002; Nies et al. 1989a; Nies et al. 1990b; Peitzsch et al. 1998). Three genes (*chrFBA*) have chromosomal counterparts. *chrC* encodes a superoxide dismutase (SOD). In fact, this six-gene organization, which includes the SOD, has counterparts in *B. xenovorans* LB400 and *B. vietnamiensis* G4. A special feature of the *chr* cluster is that *chrI* is downregulated by Cu(II), Cd(II), Co(II), Ni(II), Pb(II), and Zn (II). This gene is one of the rare genes that was found to be

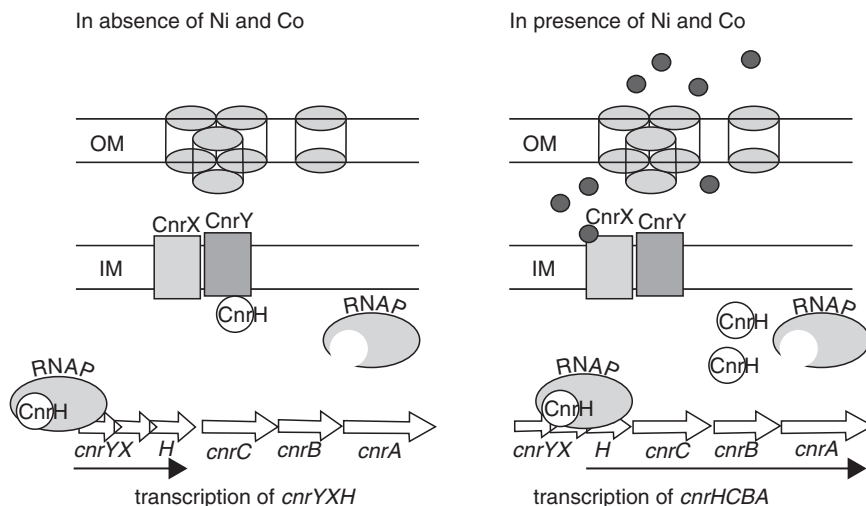


Fig. 6 Model for the regulation of the pMOL28 *cnr* operon. In the absence of inducer (Ni(II) or Co(II)), a low transcription of *cnrYXH* proceeds from *PcnrY* located upstream *cnrY*, with most of produced CnrH being trapped by the antisigma factor CnrY. In presence of Ni(II) or Co(II), the pressure on CnrH is released and full transcription of *cnrH* and the resistance genes *cnrCBA* may proceed from *PcnrH* (from Tibazarwa et al, 2000)

repressed in transcriptomic studies (Monchy et al. 2007) and could have a pleiotropic function that would extend to other operons or clusters than the *chr* cluster.

7.2 *pMOL30*

Resistance to Cd(II), Zn(II), and Co(II) mediated by the *czcCBA* genes is the most striking characteristic of pMOL30 at least from a phenotypic point of view. The *czcCBA* genes also became the reference for RND-mediated efflux of cations (Dong and Mergeay 1994; Nies et al. 1987; Nies 1995; Nies 2003; Nies et al. 1989b; Nies et al. 2006; Nies and Silver 1989). The basic operon was soon extended to other *czc* genes (*czcDRSE* (Mergeay et al. 2003; Nies 1992; Zoropogui et al. 2008) and *czcI* (Diels et al. 1995a)). Neighboring genes *czcN*, *czcP*, *czcM*, *czcJ* were also found to be involved in the response to Cd(II) and Zn(II) ions mostly via analysis of transcriptomic or microarray data (Grosse et al. 2004; Grosse et al. 1999; Monchy et al. 2007; Monchy et al. 2006a; Nies et al. 2006) (D. Nies, pers. comm.). However, the real role of most of these genes is still mysterious, especially their putative synergy or coordination with the better known *czcCBA* and *czcD* genes.

The *czc* genes are contained in the CMGI-30a putative genomic island (Fig. 3; Table 1), which also contains the *pbr* genes (Borremans et al. 2001). This *pbr* operon is the only operon encoding resistance to lead that was described at the

molecular level up to now, although some other plasmid-borne resistances have been demonstrated or suggested (Trajanovska et al. 1997). The pMOL30 mediated detoxification of lead mainly relies on the P1-ATPase mediated efflux (Borremans et al. 2001; Monchy et al. 2006a; Rensing et al. 1998). The *pbr* locus of plasmid pMOL30 consists of six genes encoding the Pb(II) uptake protein PbrT, the regulator PbrR (Hobman 2007a; Permina et al. 2006), the P-type Pb(II) efflux ATPase PbrA, the lipoprotein PbrB, the signal peptidase PbrC, and the Pb(II)-binding protein PbrD. The operon is a divergon with *pbrR*, *pbrT*, and *pbrU* in one orientation and the *pbrAB/CD* genes in the opposite. Both *pbrT* and *pbrU* are induced by heavy metals (Monchy et al. 2007) and *pbrU* was up to now an unnoticed gene, partially because it was knocked out by Tn*Cme2*. Yet *pbrU* would encode for a transport protein of the Major Facilitation Superfamily (MFS) that is very well conserved through a broad variety of bacteria, has its own promoter recognized by PbrR (Fig. 6) and is overexpressed by Cd(II) and Pb(II) like the other *pbr* genes (Monchy et al. 2007); however, it is not known if *pbrU* would encode a functional product, if the transposon were to be precisely excised, and which substrate it would process. Figure 7 displays a tentative mechanism for the resistance to Pb(II) (Borremans et al. 2001; D. van der Lelie, pers. comm.).

The metal resistance genes of CMGI-30b (Fig. 3) are located in two distinct regions:

- 1) The RND gene determinants *silCBAD* and *nre/ncc* are located in the region between nt 145 and 175 delimited by one CMGI-30b extremity and the deleted *resU* gene (tyrosine recombinase) (Fig. 3). They are associated with some *gtr* genes (membrane maintenance or repair) and *mmrQ*, a *czcJ*-like gene, just as in the *czc* region. This suggests a kind of functional association between RND-mediated detoxification systems, membrane repair, a porin OmpP, and a small protein that is strongly induced in the presence of heavy metals (Monchy et al. 2007). Genes involved in the biosynthesis of flagella are also found within these clusters. As already mentioned above, the *ncclnre* genes (Grass et al. 2001; Schmidt and Schlegel 1994) do not appear to be functional, since no nickel resistance phenotype is linked to pMOL30 and no induced expression has been shown by proteomics or transcriptomics (Monchy et al. 2007).

Sil proteins are expressed in presence of copper and of silver (Mergeay et al. 2003).

- 2) The *cop* region of CMGI-30b is flanked by two deleted *int* genes and contains 22 or 23 ORFs for which 21 seem to play a role to the response to copper or are inducible by this or other metals (Mergeay et al. 2003; Monchy et al. 2007; Monchy et al. 2006b; Monchy et al. 2006a). The *cop* genes encode for at least two resistance mechanisms, one mediated by the CopF efflux P1-ATPase and the other, a periplasmic detoxification system, mediated by the proteins CopA,B,C,D and tuned by the regulators CopR and CopS.

The whole *cop* region was cloned in a cosmid that, when transferred in the plasmid-free derivative of *C. metallidurans* CH34 restored only partially the copper resistance phenotype, although the MIC seemed to be the same than in the wild-type,

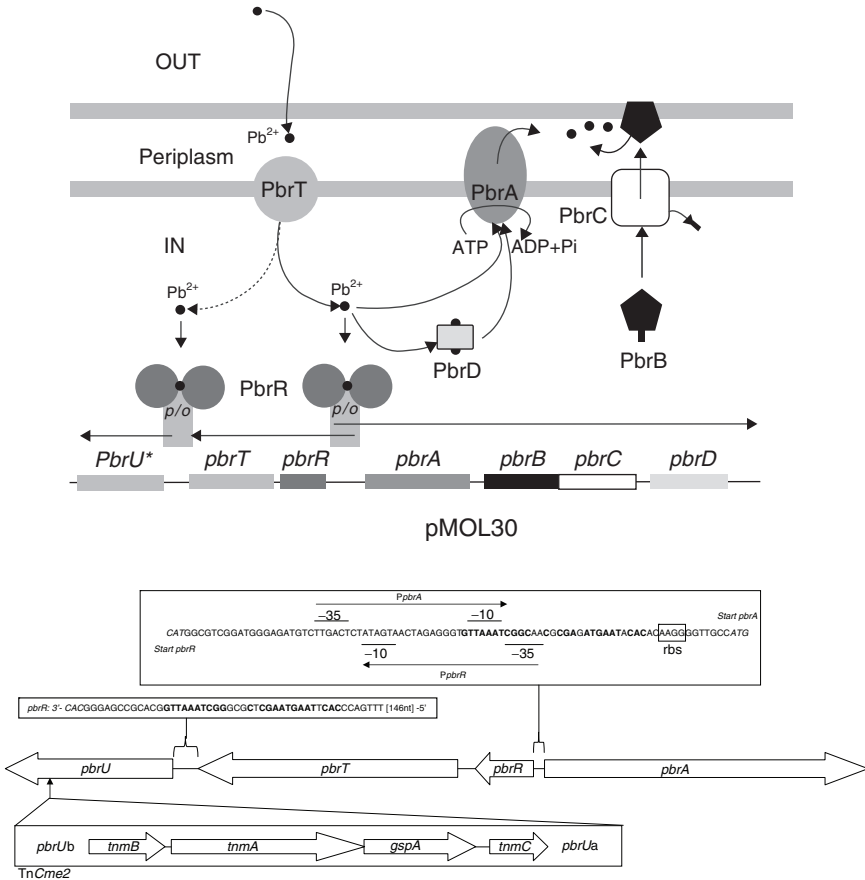


Fig. 7 Model for the mechanism of resistance to lead. The lead resistance operon includes the genes *pbrU*, *pbrT*, *pbrR*, *pbrA*, *pbrB/C* and *pbrD*. *pbrU* encodes a transporter of the major facilitator superfamily, the substrate of which is unknown. *pbrU* has been inactivated by a Tn*Cme2* insertion (see also Fig. 5). *pbrT* is involved in the transport of Pb(II) into the cytoplasm. In the presence of Pb(II), *pbrR* acts as an activator on two promoter sites (shown in the lower part of the figure) controlling the transcription of *pbrU* as well as of the divergon *pbrRT* and *pbrABCD*. The *pbrD* gene product seems to sequester Pb(II) in the cytoplasm and probably delivers it to the efflux ATPase PbrA. PbrB is most likely an outer membrane protein that is involved in the further fate of extruded Pb(II) via interaction with phosphate ions. PbrB is transported to the outer membrane by a mechanism involving the PbrC peptidase activity. It is likely that the complex *pbr* resistance mechanism provides a means of minimizing the costly reentry of the toxic ion. The upper part of the figure is published with permission of the authors

at least for cultures not preinduced by Cu(II) (Monchy et al. 2006b). It seems that other pMOL30 genes are required for full resistance to copper ions. In this respect, the neighboring *sil* genes may play a role by offering a third mechanism of detoxification, based on RND mediated efflux. Furthermore, almost all of the genes located between *sil* and *cop* are overexpressed in the presence of Cu(II). Thus, it is plausible that a series of almost 35 genes are required for maximal expression of

copper resistance. Besides *copRSABCD* and *copF*, most of the other *cop* genes appear to be unique to the *Cupriavidus* genus or related betaproteobacteria or even to the strain CH34, and their role is still the subject of ongoing investigations, especially at the level of the protein structure as in the case of CopH (Sendra et al. 2006), and CopK (Bersch et al, 2008; Tricot et al. 2005).

7.3 Organization of the Metal Resistance Genes in pMOL28 and pMOL30: Special Features

The information we have about the clusters involved in heavy metal resistance or response that are located on the plasmids pMOL28 and pMOL30 suggests different types of organization.

7.3.1 Recruiting Existing Determinants

First, the plasmids contain resistance genes that have been observed in a large variety of replicons: examples are the genes encoding P1-ATPases (*copF*, *pbrA*, *czcP*), the *merRTPADEorf2* genes, and the cluster *czcCBA* (RND) adjacent to two-component regulatory genes *czcRS* or to the *copABCD* genes. These rather ubiquitous clusters constitute the first conspicuous part of the metal response tool box and could have been recruited by horizontal gene transfer. However, some genes of these operons may differ from their current (mostly chromosomal) counterparts. A striking case is the *copB* gene of pMOL30. Its gene product contains an amplified methionine-rich motif (Mergeay et al. 2003; Monchy et al. 2006b), which may constitute a more specific adaptation to very high concentrations of heavy metals that are typical of the biotopes where *C. metallidurans* strains were found.

7.3.2 Clustering

Another feature is the clustering of different resistance mechanisms as for example RND, CDF, and ATPase genes in the *czc* cluster or the periplasmic *copABCD* genes, an ATPase, and even RND (*sil*) in the *cop* region sensu lato (CMGI-30b). Synchronizing the detoxification of all cellular compartments would clearly be an appropriate response to heavy toxic pressures. The *cop* “nested island” seems also to be such a case of clustering.

7.3.3 Synergies

Synergies are also obvious between plasmid-borne genes and genes located on chromosome 2 with the main example being the synergy between the ZntA ATPase (chromosome 2) and CzcCBA (pMOL30) (Legatzki et al. 2003; Monchy et al. 2006a).

7.3.4 Unique Genes

Besides these expected combinations and synergies, there are also many genes that are quite rare (and observed mostly in related betaproteobacteria) or even unique to these plasmids and for which the functions are still poorly known: *czcJ*, *mmrQ*, and *copQ* are quite similar and encode for very small proteins that are highly induced in the presence of heavy metals (Monchy et al. 2007). *czcE* (Zoropogui et al. 2008) and *copH* (Sendra et al. 2006) encode similar proteins (CzcE, CopH) that fix copper ions. The function of unique genes of the *cop* region (e.g., *copK*, *V*, *T*, and *M*) is still unknown or not clear (Mergeay et al. 2003).

8 Transcriptomics

For the assignment of functions or phenotypes of plasmid ORFs, many of which are hypothetical or even unique to *C. metallidurans*, important tools are provided by the high-throughput approaches of transcriptomics and proteomics (Mergeay et al. 2003; Noel-Georis et al. 2004). RT-PCR based studies on copper-challenged cultures showed that most of the *cop* ORFs, even the unique and very small ones, were upregulated in the presence of Cu(II) (Hobman et al. 2007b; Monchy et al. 2006b), providing a first clue on the function of apparently unique *cop* genes. The induction of the pMOL30-borne P1-ATPases *pbrA*, *czcP*, and *copF* was also measured via RT-PCR (Monchy et al. 2006a). Whole-genome oligonucleotide microarray studies of *C. metallidurans* CH34 cultures, which were challenged by different heavy metals during a 30 min shock, have provided interesting data in this respect. Table 2 shows that, depending on the metal tested, between 40 and 200 genes display at least a twofold induction. An important fraction of these genes is provided by the plasmids pMOL28 and pMOL30 that contribute for the majority of the 20–30 most induced genes. Clearly, transcriptomic data illustrate that pMOL28 and especially pMOL30 are specialized in the response to heavy metals, even if the four replicons contribute to the diversity of overexpressed genes.

Table 2 Number of differentially upregulated (more than twofold) genes in CH34 by heavy metal shock as indicated by whole-genome oligonucleotide microarray

	Cd(II)	Zn(II)	Ni(II)	Cu(II)	Pb(II)	Co(II)	Se(IV)
CH34 genes	190	88	82	104	183	37	119
pMOL28 genes	27	8	9	14	18	6	1
pMOL28 genes within first 20 ^a	3	1	3	6	4	6	0
pMOL30 genes	57	57	28	35	30	9	2
pMOL30 genes within first 20 ^a	15	15	13	10	3	6	1

^aFirst 20 are the most induced genes

Most of the metal-induced plasmid genes reside, as expected, in the genomic islands, with a few exceptions of genes involved in the putative transfer or mobility of pMOL28.

A very striking observation is the cross-regulation of most of the operons. Structural *mer* genes are overexpressed in presence of Cd(II). Nickel is a major inducer of the pMOL30 *cop* genes and Cu(II) also induces the *cnr* genes of pMOL28. (This phenomenon is less pronounced in the case of the *czc* genes that seem to respond specifically to Cd(II) and Zn(II)). This effect deserves more attention and is also observed on metal resistance operons that are located on the chromosomes. Different regulatory networks are perhaps intertwined (Hobman et al. 2007b; Kershaw et al. 2005). Although bona fide IS elements do not react to various metals, some small ORFs corresponding to partial IS ORFs (catalytic or DNA binding orfs) or recombinase genes concentrated in a 12 kb region of pMOL30 are overexpressed in the presence of various metals. One of them corresponds to the C-terminal part of the catalytic subunit of a IS3 family transposase and is induced 14-fold by Cd(II), 3.7-fold by Ni(II), 6.4 by Zn(II), 2.4 by Pb, and 2.1 by Cu(II). The adjacent ORF encoding for a partial DNA-binding subunit of the same kind of IS is also induced by Cd(II), Ni(II), and Zn(II) (two to threefold). The concentration of these deleted genes and the induction of some of them by heavy metals are intriguing: could it be a clue to the evolution towards regulatory functions?

9 Plasmids and the Evolution of Metal Resistance Genes

Plasmids pMOL28 and pMOL30 are especially rich in heavy metal resistance genes and seem to allow their host to occupy specific niches characterized by high levels of bioavailable toxic metals.

The great diversity of genes involved in the response to heavy metals that are gathered in plasmid pMOL28 and mainly in pMOL30 raises the question of their origin: some resistance genes seem to be unique to *C. metallidurans*. Of course, uniqueness is always tentative pending the addition of new isolates to the database. In this respect, gene rearrangements, combinations, and reshuffling supported by conjugative mobility would explain most of the acquisitions of specialized genes.

Nevertheless, the fundamental question remains and we may still ask if it is applicable to *C. metallidurans* or a related strain: in extreme chemical conditions such as those in the biotopes (volcanic or industrial) where *C. metallidurans* and related strains were found, can new genes arise in these or closely related bacteria as an effect of selection pressure? Could a plasmid be a cradle or a proof bank of new ORFs that could improve adaptation to heavy metals?

Are some resistance genes involved in the survival of the bacteria in the presence of mixtures of heavy metals at high concentrations as is the case in metallurgical plants and especially at concentrations higher than the MIC?

From the genetic point of view, plasmids pMOL28 and pMOL30 have backbones that recall some other plasmids especially in the genera *Cupriavidus* and *Burkholderia*.

Islands rich in heavy metal resistance genes are grafted on these backbones. IS transposases and tyrosine recombinases seem to play distinct roles in the construction of these islands as may be deduced from their location: transposase genes (often inactivated) at the borders of the putative islands and tyrosine recombinase genes (also often deleted) being inside and delimitating blocks of genes. The fact that these genes are inactivated or deleted indicates that selection pressures acted on two levels: first in the acquisition step via gene mobility and second in stabilizing the acquired modules via the deactivation of the mobility agents.

Plasmids pMOL28 and pMOL30 still maintain most of their secrets but the first rounds of annotation supported by transcriptomic data show clearly a structure and an organization at different levels as well as a specialization in the response to heavy metals.

Acknowledgments This work was partially supported by a grant of ESA (European Space Agency) and Belgian Scientific Policy via the MISSEX project. The genome of *C. metallidurans* was sequenced by the Joint Genome Institute. Thanks are due to Claudine Medigue and the team of the MAGE Genoscope annotation platform for help and precious advice. We are also grateful to Daniel van der Lelie for comments and suggestions and to Pieter Monsieurs for statistic reassessment of transcriptomic data. The technical assistance of Ann Provoost and Albert Bossus is gratefully acknowledged.

The first author is deeply grateful to Ed Schwartz for his comprehension and patience in difficult circumstances.

This work is dedicated to the memory of Larissa Hendrickx (†3/3/2008) and of Jozef Gerits († 24/3/2008).

References

- Amadou C, Pascal G, Mangenot S, Glew M, Bontemps C, Capela D, Carrère S, Cruveiller S, Dossat C, Lajus A, Marchetti M, Poinot V, Rouy Z, Servin B, Saad M, Schenowitz C, Barbe V, Batut J, Médigue C, Masson-Boivin C (2008) Genome sequence of the β -rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res* doi:10.1101/gr.076448.108
- Bale MJ, Fry JC, Day MJ (1988) Transfer and occurrence of large mercury resistance plasmids in river epilithon. *Appl Environ Microbiol* 54:972–978.
- Bersch B, Favier A, Schanda P, van Aelst S, Vallaëys T, Covès J, Mergeay M, Wattiez R (2008) Molecular structure and metal-binding properties of the periplasmic CopK protein expressed in *Cupriavidus metallidurans* CH34 during copper challenge. *J Mol Biol* 380:386–403.
- Borremans B, Hobman JL, Provoost A, Brown NL, van Der Lelie D (2001) Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. *J Bacteriol* 183:5651–5658.
- Brim H, Heyndrickx M, de Vos P, Wilmotte A, Springael D, Schlegel H, Mergeay M (1999) Amplified rDNA restriction analysis and further genotypic characterisation of metal-resistant soil bacteria and related facultative hydrogenotrophs. *Syst Appl Microbiol* 22:258–268.
- Champier L, Duarte V, Michaud-Soret I, Covès J (2004) Characterization of the MerD protein from *Ralstonia metallidurans* CH34: a possible role in bacterial mercury resistance by switching off the induction of the *mer* operon. *Mol Microbiol* 52:1475–1485.
- Chen WM, Moulin L, Bontemps C, Vandamme P, Bena G, Boivin-Masson C (2003) Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. *J Bacteriol* 185:7266–7272.
- Chen WM, Laevens S, Lee TM, Coenye T, De Vos P, Mergeay M, Vandamme P (2001) *Ralstonia taiwanensis* sp. nov., isolated from root nodules of Mimosa species and sputum of a cystic fibrosis patient. *Int J Syst Evol Microbiol* 51:1729–1735.

- Collard JM, Provoost A, Taghavi S, Mergeay M (1993) A new type of *Alcaligenes eutrophus* CH34 zinc resistance generated by mutations affecting regulation of the *cnr* cobalt-nickel resistance system. *J Bacteriol* 175:779–784.
- Diels L, Springael D, van der Lelie N, Top E, Mergeay M (1993) Use of DNA probes and plasmid capture in a search for new interesting environmental genes. *Sci Total Environ* 139–140:471–478.
- Diels L, Dong Q, van der Lelie D, Baeyens W, Mergeay M (1995a) The *czc* operon of *Alcaligenes eutrophus* CH34: from resistance mechanism to the removal of heavy metals. *J Ind Microbiol* 14:142–153.
- Diels L, Van Roy S, Somers K, Willems I, Doyen W, Mergeay M, Springael D, Leysen R (1995b) The use of bacteria immobilized in tubular membrane reactors for heavy metal recovery and degradation of chlorinated aromatics. *J Memb Sci* 100:249–258.
- Diels L, Faelen M, Mergeay M, and Nies D (1985) Mercury transposons from plasmids governing multiple resistance to heavy metals in *Alcaligenes eutrophus* CH34. *Arch Intern Physiol Bioch* 93:27–28.
- Diels L, Sadouk A and Mergeay M (1989) Large plasmids governing multiple resistance to heavy metals: a genetic approach. *Toxicol Environ Chem* 23:79–89.
- Dong Q, Mergeay M (1994) *czc/cnr* efflux: a three-component chemiosmotic antiport pathway with a 12-transmembrane-helix protein. *Mol Microbiol* 14:185–187.
- Dong Q, Sadouk A, van der Lelie D, Taghavi S, Ferhat A, Nuyten JM, Borremans B, Mergeay M, Toussaint A (1992) Cloning and sequencing of IS1086, an *Alcaligenes eutrophus* insertion element related to IS30 and IS4351. *J Bacteriol* 174:8133–8138.
- Feulner G, Gray JA, Kirschman JA, Lehner AF, Sadosky AB, Vlazny DA, Zhang J, Zhao S, Hill CW (1990) Structure of the *rhsA* locus from *Escherichia coli* K-12 and comparison of *rhsA* with other members of the *rhs* multigene family. *J Bacteriol* 172:446–456.
- Gerstenberg C, Friedrich B, Schlegel HG (1982) Physical evidence for plasmids in autotrophic, especially hydrogen-oxidizing bacteria. *Arch Microbiol* 133:90–96.
- Goris J, De Vos P, Coenye T, Hoste B, Janssens D, Brim H, Diels L, Mergeay M, Kersters K, Vandamme P (2001) Classification of metal-resistant bacteria from industrial biotopes as *Ralstonia campinensis* sp. nov., *Ralstonia metallidurans* sp. nov. and *Ralstonia basilensis* Steinle et al. 1998 emend. *Int J Syst Evol Microbiol* 51:1773–1782.
- Grass G, Grosse C, Nies DH (2000) Regulation of the *cnr* cobalt and nickel resistance determinant from *Ralstonia* sp. strain CH34. *J Bacteriol* 182:1390–1398.
- Grass G, Fricke B, Nies DH (2005) Control of expression of a periplasmic nickel efflux pump by periplasmic nickel concentrations. *Biomaterials* 18:437–448.
- Grass G, Fan B, Rosen BP, Lemke K, Schlegel HG, Rensing C (2001) NreB from *Achromobacter xylosoxidans* 31A is a nickel-induced transporter conferring nickel resistance. *J Bacteriol* 183:2803–2807.
- Grosse C, Friedrich S, Nies DH (2007) Contribution of Extracytoplasmic Function Sigma Factors to Transition Metal Homeostasis in *Cupriavidus metallidurans* Strain CH34. *J Mol Microbiol Biotechnol* 12:227–240.
- Grosse C, Anton A, Hoffmann T, Franke S, Schleuder G, Nies DH (2004) Identification of a regulatory pathway that controls the heavy-metal resistance system *Czc* via promoter *czcNp* in *Ralstonia metallidurans*. *Arch Microbiol* 182:109–118.
- Grosse C, Grass G, Anton A, Franke S, Santos AN, Lawley B, Brown NL, Nies DH (1999) Transcriptional organization of the *czc* heavy-metal homeostasis determinant from *Alcaligenes eutrophus*. *J Bacteriol* 181:2385–2393.
- Hacker J, Hochhut B, Middendorf B, Schneider G, Buchrieser C, Gottschalk G, Dobrindt U (2004) Pathogenomics of mobile genetic elements of toxigenic bacteria. *Int J Med Microbiol* 293:453–461.
- Hedges RW, Baumberg S (1973) Resistance to arsenic compounds conferred by a plasmid transmissible between strains of *Escherichia coli*. *J Bacteriol* 115:459–460.
- Hill CW, Sandt CH, Vlazny DA (1994) *Rhs* elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol Microbiol* 12:865–871.
- Hobman JL (2007a) MerR family transcription activators: similar designs, different specificities. *Mol Microbiol* 63:1275–1278.

- Hobman JL, Yamamoto K, Oshima T (2007b) Transcriptomic Responses of Bacterial Cells to Sublethal Metal Ion Stress. In: Nies DH, Silver S (eds) *Molecular microbiology of heavy metals*. Springer Verlag, Berlin Heidelberg, pp 73–115.
- Juhnke S, Peitzsch N, Hubener N, Grosse C, Nies DH (2002) New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34. *Arch Microbiol* 179:15–25.
- Kershaw CJ, Brown NL, Constantinidou C, Patel MD, Hobman JL (2005) The expression profile of *Escherichia coli* K-12 in response to minimal, optimal and excess copper concentrations. *Microbiology* 151:1187–1198.
- Kim KH, Jung EJ, Im H, Lelie DV, Kim EE (2008) Expression, purification, and crystallization and preliminary X-ray crystallographic analysis of CnrX from *Cupriavidus metallidurans* CH34. *J Microbiol Biotechnol* 18:43–47.
- La Duc MT, Nicholson W, Kern R, Venkateswaran K (2003) Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility. *Environ Microbiol* 5:977–985.
- Larbig KD, Christmann A, Johann A, Klockgether J, Hartsch T, Merkl R, Wiehlmann L, Fritz HJ, Tummler B (2002) Gene islands integrated into tRNA(Gly) genes confer genome diversity on a *Pseudomonas aeruginosa* clone. *J Bacteriol* 184:6665–6680.
- Legatzki A, Grass G, Anton A, Rensing C, Nies DH (2003) Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. *J Bacteriol* 185:4354–4361.
- Liebert CA, Hall RM, Summers AO (1999) Transposon Tn21, flagship of the floating genome. *Microbiol Mol Biol Rev* 63:507–522.
- Liesegang H, Lemke K, Siddiqui RA, Schlegel HG (1993) Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. *J Bacteriol* 175:767–778.
- Lonetto MA, Brown KL, Rudd KE, Buttner MJ (1994) Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proc Natl Acad Sci USA* 91:7573–7577.
- Mergeay M (2000) Bacteria adapted to industrial biotopes: the metal resistant *Ralstonia*. In: Hengge-Aronis G, Storz and R. Hengge-Aronis (eds) “Bacterial Stress Responses”. ASM Press, Washington D.C. USA, Chap.26 pp., 403–414.
- Mergeay M, Houba C, Gerits J (1978) Extrachromosomal inheritance controlling resistance to cadmium, cobalt, copper and zinc ions: evidence from curing in a *Pseudomonas*. *Arch Int Physiol Biochim* 86:440–442.
- Mergeay M, Nies D, Schlegel HG, Gerits J, Charles P, Van Gijsegem F (1985) *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J Bacteriol* 162:328–334.
- Mergeay M, Monchy S, Vallaëys T, Auquier V, Benotmane A, Bertin P, Taghavi S, Dunn J, van der Lelie D, Wattiez R (2003) *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol Rev* 27:385–410.
- Mobley HL, Chen CM, Silver S, Rosen BP (1983) Cloning and expression of R-factor mediated arsenate resistance in *Escherichia coli*. *Mol Gen Genet* 191:421–426.
- Monchy S, Vallaëys T, Bossus A, Mergeay M (2006a) Metal efflux P1-ATPase genes of *Cupriavidus metallidurans* CH34: a transcriptomic approach. *Intern J Environ Anal Chem* 86:677–692.
- Monchy S, Benotmane MA, Janssen P, Vallaëys T, Taghavi S, van der Lelie D, Mergeay M (2007) Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals. *J Bacteriol* 189:7417–7425.
- Monchy S, Benotmane MA, Wattiez R, van Aelst S, Auquier V, Borremans B, Mergeay M, Taghavi S, van der Lelie D, Vallaëys T (2006b) Transcriptomic and proteomic analyses of the pMOL30-encoded copper resistance in *Cupriavidus metallidurans* strain CH34. *Microbiology* 152:1765–1776.
- Nakahara H, Ishikawa T, Sarai Y, Kondo I, Kozukue H, Silver S (1977) Linkage of mercury, cadmium, and arsenate and drug resistance in clinical isolates of *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 33:975–976.
- Ni’Bhriain NN, Silver S, Foster TJ (1983) Tn5 insertion mutations in the mercuric ion resistance genes derived from plasmid R100. *J Bacteriol* 155:690–703.

- Nies A, Nies DH, Silver S (1989a) Cloning and expression of plasmid genes encoding resistances to chromate and cobalt in *Alcaligenes eutrophus*. *J Bacteriol* 171:5065–5070.
- Nies A, Nies DH, Silver S (1990b) Nucleotide sequence and expression of a plasmid-encoded chromate resistance determinant from *Alcaligenes eutrophus*. *J Biol Chem* 265:5648–5653.
- Nies D, Mergeay M, Friedrich B, Schlegel HG (1987) Cloning of plasmid genes encoding resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus* CH34. *J Bacteriol* 169:4865–4868.
- Nies DH (1992) CzcR and CzcD, gene products affecting regulation of resistance to cobalt, zinc, and cadmium (*czc* system) in *Alcaligenes eutrophus*. *J Bacteriol* 174:8102–8110.
- Nies DH (1995) The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *Escherichia coli*. *J Bacteriol* 177:2707–2712.
- Nies DH (2003) Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol Rev* 27:313–339.
- Nies DH, Silver S (1989) Plasmid-determined inducible efflux is responsible for resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus*. *J Bacteriol* 171:896–900.
- Nies DH, Nies A, Chu L, Silver S (1989b) Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*. *Proc Natl Acad Sci USA* 86:7351–7355.
- Nies DH, Koch S, Wachi S, Peitzsch N, Saier MH, Jr. (1998) CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate antiporters. *J Bacteriol* 180:5799–5802.
- Nies DH, Rehbein G, Hoffmann T, Baumann C, Grosse C (2006) Paralogs of genes encoding metal resistance proteins in *Cupriavidus metallidurans* strain CH34. *J Mol Microbiol Biotechnol* 11:82–93.
- Noel-Georis I, Vallaëys T, Chauvaux R, Monchy S, Falmagne P, Mergeay M, Wattiez R (2004) Global analysis of the *Ralstonia metallidurans* proteome: prelude for the large-scale study of heavy metal response. *Proteomics* 4:151–179.
- Novick RP, Roth C (1968) Plasmid-linked resistance to inorganic salts in *Staphylococcus aureus*. *J Bacteriol* 95:1335–1342.
- O'Sullivan LA, Weightman AJ, Jones TH, Marchbank AM, Tiedje JM, Mahenthalingam E (2007) Identifying the genetic basis of ecologically and biotechnologically useful functions of the bacterium *Burkholderia vietnamiensis*. *Environ Microbiol* 9:1017–1034.
- Peitzsch N, Eberz G, Nies DH (1998) *Alcaligenes eutrophus* as a bacterial chromate sensor. *Appl Environ Microbiol* 64:453–458.
- Permina EA, Kazakov AE, Kalinina OV, Gelfand MS (2006) Comparative genomics of regulation of heavy metal resistance in Eubacteria. *BMC Microbiol* 6:49.
- Peyru G, Wexler LF, Novick RP (1969) Naturally occurring penicillinase plasmids in *Staphylococcus aureus*. *J Bacteriol* 98:215–221.
- Rensing C, Sun Y, Mitra B, Rosen BP (1998) Pb(II)-translocating P-type ATPases. *J Biol Chem* 273:32614–32617.
- Rossey E, Champier L, Bersch B, Brutscher B, Blackledge M, Covès J (2004) Biophysical characterization of the MerP-like amino-terminal extension of the mercuric reductase from *Ralstonia metallidurans* CH34. *J Biol Inorg Chem* 9:49–58.
- Sadouk A, Mergeay M (1993) Chromosome mapping in *Alcaligenes eutrophus* CH34. *Mol Gen Genet* 240:181–187.
- Saier MH, Jr., Tam R, Reizer A, Reizer J (1994) Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Mol Microbiol* 11:841–847.
- Schmidt T, Schlegel HG (1994) Combined nickel-cobalt-cadmium resistance encoded by the *ncc* locus of *Alcaligenes xylosoxidans* 31A. *J Bacteriol* 176:7045–7054.
- Schmidt T, Stoppel RD, Schlegel HG (1991) High-Level Nickel Resistance in *Alcaligenes xylosoxidans* 31A and *Alcaligenes eutrophus* KTO2. *Appl Environ Microbiol* 57:3301–3309.
- Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B, Gottschalk G (2003) Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplastid encoding key enzymes of H₂-based lithoautotrophy and anaerobiosis. *J Mol Biol* 332:369–383.

- Sendra V, Cannella D, Bersch B, Fieschi F, Menage S, Lascoux D, Covès J (2006) CopH from *Cupriavidus metallidurans* CH34. A novel periplasmic copper-binding protein. *Biochemistry* 45:5557–5566.
- Serre L, Rossey E, Pebay-Peyroula E, Cohen-Addad C, Covès J (2004) Crystal structure of the oxidized form of the periplasmic mercury-binding protein MerP from *Ralstonia metallidurans* CH34. *J Mol Biol* 339:161–171.
- Siddiqui RA, Schlegel HG, Meyer M (1988) Inducible and constitutive expression of pMOL28-encoded nickel resistance in *Alcaligenes eutrophus* N9A. *J Bacteriol* 170:4188–4193.
- Siddiqui RA, Benthin K, Schlegel HG (1989) Cloning of pMOL28-encoded nickel resistance genes and expression of the genes in *Alcaligenes eutrophus* and *Pseudomonas* spp. *J Bacteriol* 171:5071–5078.
- Silver S (1996) Bacterial resistances to toxic metal ions--a review. *Gene* 179:9–19.
- Silver S, Misra TK (1984) Bacterial transformations of and resistances to heavy metals. *Basic Life Sci* 28:23–46.
- Smith K, Novick RP (1972) Genetic studies on plasmid-linked cadmium resistance in *Staphylococcus aureus*. *J Bacteriol* 112:761–772.
- Summers AO, Silver S (1972) Mercury resistance in a plasmid-bearing strain of *Escherichia coli*. *J Bacteriol* 112:1228–1236.
- Summers AO, Silver S (1978) Microbial transformations of metals. *Annu Rev Microbiol* 32:637–672.
- Taghavi S, Mergeay M, van der Lelie D (1997) Genetic and physical maps of the *Alcaligenes eutrophus* CH34 megaplasmid pMOL28 and its derivative pMOL50 obtained after temperature-induced mutagenesis and mortality. *Plasmid* 37:22–34.
- Tetaz TJ, Luke RK (1983) Plasmid-controlled resistance to copper in *Escherichia coli*. *J Bacteriol* 154:1263–1268.
- Tibazarwa C, Wuertz S, Mergeay M, Wyns L, van Der Lelie D (2000) Regulation of the *cnr* cobalt and nickel resistance determinant of *Ralstonia eutropha* (*Alcaligenes eutrophus*) CH34. *J Bacteriol* 182:1399–1409.
- Toussaint A, Merlin C, Monchy S, Benotmane MA, Leplae R, Mergeay M, Springael D (2003) The biphenyl- and 4-chlorobiphenyl-catabolic transposon Tn4371, a member of a new family of genomic islands related to IncP and Ti plasmids. *Appl Environ Microbiol* 69:4837–4845.
- Trajanovska S, Britz ML, Bhave M (1997) Detection of heavy metal ion resistance genes in gram-positive and gram-negative bacteria isolated from a lead-contaminated site. *Biodegradation* 8:113–124.
- Tricot C, van Aelst S, Wattiez R, Mergeay M, Stalon V, Wouters J (2005) Overexpression, purification, crystallization and crystallographic analysis of CopK of *Cupriavidus metallidurans*. *Acta Crystallographica Section F* 61:825–827.
- Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Médigue C (2006) MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res* 34:53–65.
- van der Lelie D, Schwuchow T, Schwidetzky U, Wuertz S, Baeyens W, Mergeay M, Nies DH (1997) Two-component regulatory system involved in transcriptional control of heavy-metal homeostasis in *Alcaligenes eutrophus*. *Mol Microbiol* 23:493–503.
- Vandamme P, Coenye T (2004) Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int J Syst Evol Microbiol* 54:2285–2289.
- Vanechoutte M, Kampf P, De Baere T, Falsen E, Verschraegen G (2004) *Wautersia* gen. nov., a novel genus accommodating the phylogenetic lineage including *Ralstonia eutropha* and related species, and proposal of *Ralstonia* [*Pseudomonas*] *syzygii* (Roberts et al. 1990) comb. nov. *Int J Syst Evol Microbiol* 54:317–327.
- Yabuuchi E, Kosako Y, Yano I, Hotta H, Nishiuchi Y (1995) Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. Nov.: Proposal of *Ralstonia pickettii* (Ralston, Palleroni and Doudoroff 1973) comb. Nov., *Ralstonia solanacearum* (Smith 1896) comb. Nov. and *Ralstonia eutropha* (Davis 1969) comb. Nov. *Microbiol Immunol* 39:897–904.
- Zoropogui A, Gambarelli S, Covès J (2008) CzcE from *Cupriavidus metallidurans* CH34 is a copper-binding protein. *Biochem Biophys Res Commun* 365:735–739.

Megaplasמידs of Aerobic Hydrogenotrophic and Carboxidotrophic Bacteria

Edward Schwartz

Contents

1	Introduction.....	240
1.1	Hydrogenotrophy: Knallgas Without the Bang.....	240
1.2	Carboxidotrophy: Thriving on Poison.....	242
2	Megaplasמיד pHG1 of <i>Ralstonia eutropha</i> H16.....	243
2.1	Plasmid Functions and General Organization of pHG1.....	243
2.2	Hydrogenase Genes.....	243
2.3	The <i>cbb</i> _p Operon.....	251
2.4	Denitrification Genes.....	253
2.5	Catabolism of Aromatic Substances.....	255
2.6	The pHG1 Junkyard Region.....	255
2.7	Megaplasמיד–Chromosome Synergy.....	257
3	Megaplasמיד pHCG3 of <i>Oligotropha carboxidovorans</i> OM5.....	258
3.1	Plasmid Functions and General Organization of pHCG3.....	259
3.2	The <i>cox</i> Genes.....	259
3.3	The <i>cbb</i> Genes.....	263
3.4	The <i>hox</i> Genes.....	264
3.5	Megaplasמיד–Chromosome Synergy.....	264
	References.....	265

Abstract The hydrogen-oxidizing bacterium *Ralstonia eutropha* H16 and the carbon monoxide-oxidizing bacterium *Oligotropha carboxidovorans* OM5 carry key genetic determinants for their respective forms of lithoautotrophic metabolism on megaplasמידs. In *R. eutropha* H16 genetic information for the H₂-oxidizing system and for CO₂ fixation via the Calvin–Benson–Bassham cycle is located on the 452-kb megaplasמיד pHG1. In addition, pHG1 harbors clusters of genes for denitrification and for degradation of aromatic compounds. The 133-kb megaplasמיד pHCG3 is the genetic basis for CO oxidation in *O. carboxidovorans* OM5. Aside

E. Schwartz

Institut für Biologie, Mikrobiologie, Humboldt-Universität zu Berlin, Chauseestr. 117,
10115 Berlin, Germany
e-mail: edward.schwartz@rz.hu-berlin.de

from the aerobic CO dehydrogenase, pHCG3 also encodes Calvin cycle enzymes and a dimeric hydrogenase. In both cases there is an interdigitation of megaplasmid-based and chromosomally encoded functions, indicating that these megaplasmids are, although not strictly essential for viability, an integral part of the genome.

1 Introduction

Lithotrophs are prokaryotic organisms with the capacity to use inorganic matter as a source of reducing power to fuel their electron transport chains. This group is phylogenetically heterogeneous, including both bacteria and archaea, and the various representatives exploit a wide range of substances, including H₂, S, H₂S, NH₃, Fe²⁺, and CO (Kelly and Wood 2006; Schlegel and Bowien 1989). Correspondingly diverse are the habitats of lithotrophic microbes. While some highly specialized representatives are restricted to niches with extreme conditions, such as submarine hydrothermal vents, numerous other lithotrophs are ubiquitous in soils, lakes, and seawater. Many lithotrophs are also autotrophs, that is, they are capable of fixing CO₂. The latter organisms are entirely independent of organic substrates.

Early studies revealed that in two groups of lithoautotrophs the key genetic determinants for lithoautotrophic metabolism are carried on giant plasmids: the H₂-based or hydrogenotrophic bacteria and the CO-oxidizing bacteria, also known as carboxidotrophs. Despite the fact that other groups of lithotrophs such as the sulfur-oxidizing bacteria and the nitrifiers have been extensively studied – and in recent years have also been the subject of several genome projects – no cases of megaplasmid-linked lithotrophy genes have been reported for these groups to date. Therefore, this chapter confines itself to the H₂- and CO-based lithotrophs. In the following, a brief introduction to the physiology and ecology of these organisms is given.

1.1 Hydrogenotrophy: Knallgas Without the Bang

Aerobic H₂-oxidizing microbes are ubiquitous in terrestrial and aquatic environments. These organisms obtain energy and reducing power by coupling H₂ oxidation with O₂ reduction in the equivalent of a “Knallgas reaction”:



This strongly exergonic reaction ($\Delta G^{0'} = -237.2 \text{ kJ mol}^{-1}$) is an attractive energy source for biological systems. Among the better characterized aerobic H₂-oxidizing bacteria are *Alcaligenes hydrogenophilus* M50, *Azohydromonas lata*, *Acidovorax facilis*, *Variovorax paradoxus* H-4, *Aquaspirillum autotrophicum* SA32, *Paracoccus pantotrophus*, *Rhodococcus opacus* MR11, *Cupriavidus metallidurans* CH34, and *Ralstonia eutropha* H16 (Friedrich and Friedrich 1990; Schwartz and Friedrich

2006). All of the aforementioned strains are facultative lithotrophs capable of growth either on organic compounds or, in the absence of such substrates, on H_2 as the sole energy source. Some of the organisms, such as *Ralstonia eutropha* H16, can even grow mixotrophically, utilizing H_2 and organic compounds simultaneously. H_2 -based lithotrophy as a facultative growth mode is found in bacteria not usually thought of as lithotrophs. Some phototrophic bacteria, such as *Rhodobacter capsulatus*, grow lithotrophically in the dark on H_2 (Madigan and Gest 1979; Siefert and Pfennig 1979). *R. capsulatus* is also a diazotroph, and facultative growth on H_2 as sole energy source is not uncommon in diazotrophs. The reason for this is simple: H_2 is a byproduct of N_2 fixation. Many N_2 -fixers are equipped with H_2 -oxidizing enzymes, with which they can recover at least a part of the energy that would otherwise escape in the form of H_2 . It is then not surprising that some of these organisms have adapted their H_2 -oxidizing system to utilize external H_2 on a facultative basis and can thereby take advantage of an additional substrate for survival in the soil. *Bradyrhizobium japonicum* USDA 110, an endosymbiotic N_2 -fixing bacterium, can grow in the soil (outside of the root nodule) on H_2 as sole energy source (Hanus et al. 1979). A systematic survey of N_2 -fixers capable of lithotrophic growth on H_2 is lacking.

The examples noted above make it clear that hydrogenotrophy is not correlated with taxonomic clades. Even at the species level there is no such correlation. For example, some *Ralstonia eutropha* strains grow on H_2 and others (e.g., *Ralstonia eutropha* JMP134) do not.

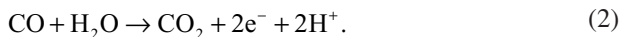
All the "classical" hydrogenotrophs mentioned earlier were discovered using enrichment culture techniques. Such experiments tell us little about microbial communities "in the wild." Data from systematic ecological investigations on the distribution of H_2 oxidizers in the natural microbial populations of defined habitats are scant. One such study found that up to 30% of the culturable isolates from temperate-zone meadow and forest soils were capable of H_2 -based lithoautotrophy (Klüber et al. 1995). Another survey employed both culture-based and culture-independent techniques to estimate the abundance of H_2 -oxidizing lithoautotrophic bacteria in the rhizosphere of rice plants (Lechner and Conrad 1997). These experiments showed that a gram of fresh rice roots contained about 7.4×10^6 bacteria that could grow on a mixture of H_2 and CO_2 compared to an estimate of 2.6×10^7 derived from quantitative PCR using primers specific for a hydrogenase gene. In general, the available data suggest that the distribution of H_2 -oxidizers in the soil is heterogeneous, with the majority of the population living in more or less close association with N_2 -fixing organisms, that is, in the immediate vicinity of root nodules (La Favre and Focht 1983). H_2 is an obligate byproduct of biological N_2 -fixation (see above). Consequently, N_2 -fixing microbes are a major biogenic source of H_2 . While some diazotrophs are capable of oxidizing the internally produced H_2 , many cannot. Studies indicate that over 50% of symbiotic N_2 -fixing strains lack H_2 -oxidizing enzymes (Arp 1992). The latter so-called Hup^- strains are responsible for the substantial amounts of H_2 released by root nodules. It has been estimated that legumes nodulated with Hup^- diazotrophs release 5,000 L of H_2 per day and hectare during the peak growth period (Dong et al. 2003). Little or none of this H_2 escapes into the atmosphere

(Conrad and Seiler 1979). Most of the H_2 is oxidized within a radius of a few centimeters of the nodule (Cunningham et al. 1986; La Favre and Focht 1983). A recent study surveyed soil samples for the presence of H_2 -oxidizing bacteria (Maimaiti et al. 2007). In soil samples taken from locations with no Hup^- -nodulated legumes in the vicinity, no H_2 -oxidizing bacteria were isolated. In contrast, 3 of 146 isolates from a sample taken from a field of legumes nodulated with a Hup^- diazotroph oxidized H_2 (Maimaiti et al. 2007)

R. eutropha H16 is probably the best studied H_2 -oxidizing bacterium. The organism conserves the energy of the Knallgas reaction with the help of H_2 -oxidizing enzymes called hydrogenases. *R. eutropha* H16 produces two distinct energy-conserving hydrogenases: a membrane-bound hydrogenase (MBH) and a soluble or cytoplasmic enzyme (SH).

1.2 Carboxidotrophy: Thriving on Poison

CO is a highly toxic compound for organisms with respiratory metabolism. Paradoxically, a group of lithoautotrophs can thrive on CO as a sole source of reducing power. Some of them also use this gas as a carbon source. At levels of about 0.05–0.35 ppm (Crutzen and Gidel 1983), CO is a minor component in the present-day atmosphere. This seemingly insignificant fraction represents the atmospheric equilibrium resulting from a global flux of gigantic proportions. Abiogenic, biogenic, and anthropogenic processes contribute to the global balance of CO. A total of 2,500–2,600 Tg per year CO are released into the atmosphere (Khalil et al. 1999). Aerobic CO-oxidizing microorganisms in the soil are a major sink, consuming approx. 300 Tg per year from the atmospheric CO pool (King 1999). In addition, these microbes consume a large fraction of the 200 Tg per year released into the soil by plant roots. Marine CO oxidizers take up about 40–50 Tg per year of the CO produced in the oceans. Despite renewed efforts in recent years, only a minority of the organisms participating in CO oxidation have been identified (King and Weber 2007). Among the better known aerobic CO oxidizers are *Oligotropha carboxidovorans* OM5, *Pseudomonas carboxydohydrogena*, *Carbophilus carboxydus*, *Zavarzinia compransoris*, *Hydrogenophaga pseudoflava*, *Pseudomonas thermocarboxidovorans*, *Streptomyces thermoautotrophicus*, and *Bacillus schlegelii* (Meyer 1989). These organisms obtain energy from the following reaction:



The oxidation of CO is catalyzed by the enzyme CO dehydrogenase (CODH). In *Oligotropha carboxidovorans* OM5 – and presumably in several other carboxidotrophs – the electrons generated in the reaction are transferred to a specialized electron transport chain, thereby conserving energy.

This chapter focuses on two organisms as representatives of the H_2 - and CO-based lifestyles: the H_2 -oxidizing bacterium *Ralstonia eutropha* H16 and the CO-oxidizing

bacterium *Oligotropha carboxidovorans* OM5. In both these organisms the key genetic determinants for lithotrophic metabolism are carried on giant plasmids.

2 Megaplasmiid pHG1 of *Ralstonia eutropha* H16

It was mainly the pioneering work of Friedrich and coworkers that led to the discovery that *R. eutropha* H16 and related strains contain megaplasmiids and that in some cases these megaplasmiids carry the genetic determinants for the H₂-oxidizing system (Andersen et al. 1981; Friedrich et al. 1981; Gerstenberg et al. 1982) (Table 1). In the decades following this discovery, numerous studies on the 452,156-bp megaplasmiid pHG1 have provided a wealth of information on H₂-based lithoautotrophy.

2.1 Plasmid Functions and General Organization of pHG1

The main replication protein of pHG1 appears to be encoded by the *repA* gene (Fig. 1). The *R. eutropha* H16 RepA is similar to the orthologs on the *Cupriavidus metallidurans* CH34 megaplasmiid pMOL28 and on the *Cupriavidus taiwanensis* megaplasmiid pSym (Mergeay et al. 2008; Monchy et al. 2007). pHG1 is a conjugative plasmid and a large cluster of genes for mating pair formation and conjugative DNA processing occupy a 27-kb-large region (Schneider et al. 1988; Schwartz et al. 2003). This cluster shows extensive synteny to sets of genes on the above-mentioned plasmids pMOL28 and pSym (Monchy et al. 2007; Mergeay et al. 2008). Other maintenance systems on pHG1 include the partitioning functions *parA* and *parB*, a *xerC/xerD*-type gene for a multimer resolvase, and a *pemI/pemK*-type addiction system (Schwartz et al. 2003).

Groups of functionally related genes on pHG1 form loose clusters bracketed by mobile elements. The main clusters are the *hox* regulon encoding the hydrogenases, the *cbp_p* genes encoding the Calvin-cycle enzymes, the denitrification genes, and a set of determinants for the degradation of aromatic compounds. The mobile elements flanking the gene clusters could be the result of (1) transposition events leading to insertion of an entire gene cluster into the megaplasmiid, (2) transposition of mobile elements into nonessential sequences (e.g., mobile elements) between the clusters, where such insertions are most likely to be tolerated, or (3) both mechanisms.

2.2 Hydrogenase Genes

The genetic determinants for the H₂-oxidizing enzymes of *R. eutropha* H16 occupy a 90-kb region of pHG1 (Fig. 1). This region contains three clusters of hydrogenase genes separated by blocks of unrelated sequence (Kortlüke et al. 1992; Schwartz

Table 1 Megaplasmsids of H₂- and CO-oxidizing lithotrophic bacteria^a

Plasmid	Host strain	Size (kb)	Phenotype	Self-transmissible	Accession No.	Reference
Hydrogenotrophs						
pHG1	<i>Ralstonia eutropha</i> H16	452,156	Hox, Cfx, Dnt	+	AY305378	Friedrich et al. (1981); Andersen et al. (1981); Schwartz et al. (2003)
pHG2	<i>Ralstonia eutropha</i> TF93	450	Hox	+		Friedrich et al. (1981); Hogrefe and Friedrich (1984)
pHG3	<i>Ralstonia eutropha</i> N9A	450	Hox	+		Friedrich et al. (1981); Hogrefe and Friedrich (1984)
pHG4	<i>Ralstonia eutropha</i> G27	450	Hox	+		Friedrich et al. (1981)
pHG7	<i>Ralstonia eutropha</i> H20	430	Hox	+		Hogrefe et al. (1984)
pHG10	<i>Ralstonia eutropha</i> ATCC17704	500	Hox	+		Hogrefe and Friedrich (1984)
pRMB1	<i>Ralstonia eutropha</i> ATCC17706	ND	Hox	-		Behki et al. (1983)
pHG12-b	<i>Ralstonia eutropha</i> ATCC17707	500	Hox	+		Cangelosi and Wheelis (1984)
pHG12-a	<i>Ralstonia eutropha</i> ATCC17707	550	ND	ND		Hogrefe and Friedrich, unpublished
pHG21-a	<i>Ralstonia hydrogenophila</i> M50	420	Hox	+		Friedrich et al. (1984); Ohi et al. (1979)
pHG21-b	<i>Ralstonia hydrogenophila</i> M50	360	ND	ND		Friedrich et al. (1984)
ND	<i>Ralstonia hydrogenophila</i> 4a-2	238	Hox	-		Timotius and Schlegel (1987)
pMOL28	<i>Cupriavidus metallidurans</i> CH34	171,459	Cnr, Chr, Mer	+	X90708	Monchy et al. (2007)
pMOL30	<i>Cupriavidus metallidurans</i> CH34	233,720	Czc, Pbr, Mer, Sil, Cop	-	X71400	Monchy et al. (2007)
pHG22-a	<i>Acidovorax facilis</i> DSM 620	440	Hox, Cfx	-		Warrelmann and Friedrich (1989)
pHG22-b	<i>Acidovorax facilis</i> DSM 620	200	ND	ND		Warrelmann and Friedrich (1989)
pHG20	<i>Acidovorax facilis</i> ATCC17695	270	ND	ND		Warrelmann and Friedrich (1989)
pHG23	<i>Acidovorax facilis</i> ATCC11228	215	ND	ND		Warrelmann and Friedrich (1989)
pHG24	<i>Acidovorax facilis</i> ATCC15376	325	ND	ND		Warrelmann and Friedrich (1989)
pHG31-a	<i>Rhodococcus opacus</i> MR11	140	Tl	ND		Sensfuss et al. (1986)
pHG201	<i>Rhodococcus opacus</i> MR11	270	Hox	+		Kalkus et al. (1990)
pHG202	<i>Rhodococcus opacus</i> MR11	400	ND	ND		Kalkus et al. (1990)

pHG203	<i>Rhodococcus opacus</i> MR11	420	ND	ND	ND	Kalkus et al. (1990)
pACRY01	<i>Acidiphilium cryptum</i> JF-5	203,589	ND	ND	CP000689	
pACRY02	<i>Acidiphilium cryptum</i> JF-5	187,422	C	ND	CP000690	
plasmid 1	<i>Paracoccus denitrificans</i> PD1222	653,815	ND	ND	CP000491	
pHG16-b	<i>Paracoccus denitrificans</i> ATCC 17741	450	ND	ND		Gerstenberg et al. (1982)
pHG42	<i>Paracoccus pantotrophus</i> GB17	110	ND	ND		Chandra and Friedrich (1986)
pHG41	<i>Paracoccus pantotrophus</i> GB17	450	ND	ND		Chandra and Friedrich (1986)
pBTA11	<i>Bradyrhizobium japonicum</i> BTA11	228,826	ND	ND	CP000495	Giraud et al. (2007)
Carboxidotrophs						
pHG3	<i>Oligotropha carboxidovorans</i> OM5	133,058	Cox, Hox, Cfx	ND	X82447	Gerstenberg et al. (1982); Kraut and Meyer (1988); Fuhrmann et al. (2003)
pHCG5-a	<i>Oligotropha carboxidovorans</i> OM4	158	ND	ND		Kraut and Meyer (1988)
pHCG5-b	<i>Oligotropha carboxidovorans</i> OM4	128	ND	ND		Kraut and Meyer (1988)
pHCG2-a	<i>Hydrogenophaga pseudoflava</i> DSM11084	110	ND	ND		Gerstenberg et al. (1982)
pHCG1-a	<i>Carbophilus carboxidus</i>	558	ND	ND		Gerstenberg et al. (1982)
pHCG1-b	<i>Carbophilus carboxidus</i>	428	ND	ND		Gerstenberg et al. (1982)
pHCG1-c	<i>Carbophilus carboxidus</i>	129	ND	ND		Gerstenberg et al. (1982)
pHCG6-a	<i>Azomonas</i> B1	254	ND	ND		Kraut and Meyer (1988)
pHCG6-b	<i>Azomonas</i> B1	200	ND	ND		Kraut and Meyer (1988)

^aThis table lists replicons between 100 and 1,000 kb in size irrespective of the carriage of lithotrophy-related genetic information. *Abbreviations:* *Hox* Growth on H₂; *Cfx* CO₂ fixation; *Dnt* Denitrification; *Cox* Growth on CO; *Cnr* Cobalt and nickel resistance; *Czc* Cobalt, zinc, and cadmium resistance; *Pbr* Lead resistance; *Mer* Mercury resistance; *Sil* Silver resistance; *Tl* Thallium resistance; *ND* no data

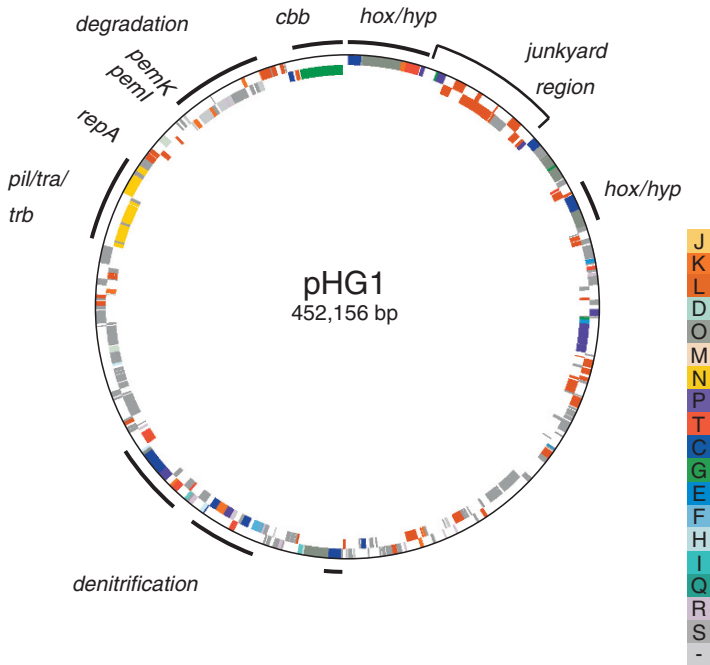


Fig. 1 Circle diagram of *R. eutropha* H16 megaplasmid pHG1. The outer and inner concentric rings represent coding regions (CDSs) transcribed in the clockwise and counterclockwise directions, respectively. CDSs are color-coded on the basis of the COG functional categories. The horizontal bar gives the color-coding used for the COG functional categories. The letter codes for these categories are as follows: J, translation and ribosomal structure; K, transcription; L, replication, recombination, repair; D, cell division, partitioning; O, posttranslational modification; M, cell envelope, outer membrane biogenesis; N, cell motility and secretion; P, inorganic ion transport and metabolism; T, signal transduction; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme metabolism; I, Lipid metabolism; Q, sec. metabolite biosynthesis; R, general function; S, function unknown; –, not in COGs. Arcs on the perimeter of the circle mark gene clusters. See text for details

et al. 2003; Tran-Betcke et al. 1990): One cluster encodes the MBH and another, the SH. A third cluster contains genes which could encode a second heterodimeric hydrogenase. However, since MBH⁻ SH⁻ double mutants do not grow on H₂, the physiological significance of the third cluster is unclear.

2.2.1 The Structure of the MBH and Its Complex Maturation Pathway

The MBH cluster consists of a series of 23 genes and encodes (1) the catalytic subunits of the enzyme, (2) accessory proteins for its maturation, (3) a set of general hydrogenase maturation proteins, (4) hydrogenase regulatory proteins, and (5) a high-affinity

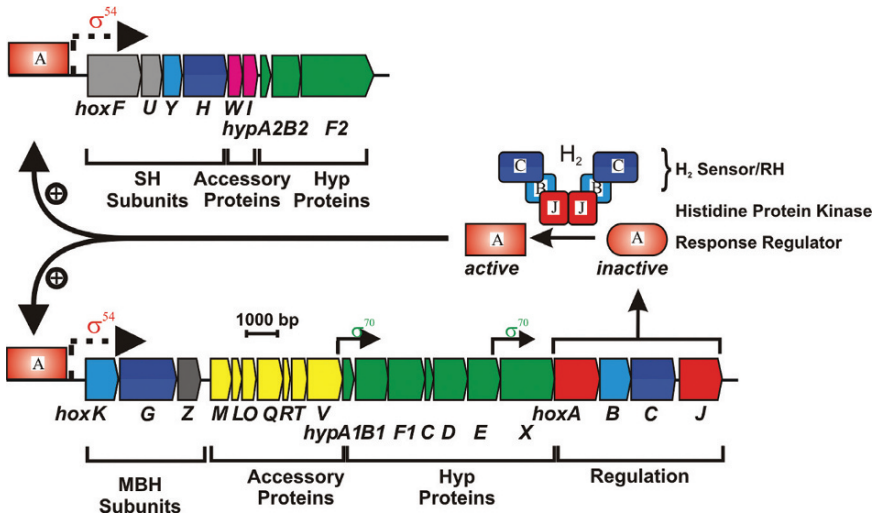


Fig. 2 Organization and regulation of the *R. eutropha* H16 *hox* regulon. Colored arrows represent the hydrogenase-related genes. Position and orientation of promoters are indicated by bent arrows. σ^{54} - and σ^{70} -dependent promoters are labeled accordingly. Symbols for the proteins HoxA, HoxB, HoxC, and HoxJ are labeled “A”, “B”, “C,” and “J”, respectively. Arrows marked with a “+” indicate a positive effect on transcription. See text for details

Ni permease (Fig. 2). The *R. eutropha* H16 MBH enzyme is a representative of the large family of heterodimeric [NiFe] hydrogenases (group 1 in the system of Vignais and coworkers (Vignais et al. 2001)). The enzyme consists of a small (34.6 kDa) subunit, the product of the gene *hoxK*. HoxK is an iron–sulfur protein, which coordinates two [4Fe4S] centers and a [3Fe4S] cluster (Ludwig et al. submitted). The large subunit (67.1 kDa) contains the active site [NiFe] center (Kortlüke et al. 1992; Schneider et al. 1983). In the native enzyme, Fe forms a complex with Ni and three diatomic ligands: two cyanide ions and one carbon monoxide molecule. The active site is buried deep inside the protein, suggesting that there are hydrophobic channels to facilitate access by H_2 molecules. Such gas channels have been experimentally demonstrated for the [NiFe] hydrogenase of *Desulfovibrio fructosovorans* (Montet et al. 1997). The presence of multiple redox centers in the enzyme suggests that an intramolecular electric circuit conducts electrons away from the active site to the primary electron acceptor (Fig. 3). The latter is a di-heme, *b*-type cytochrome encoded by the third gene of the MBH cluster, *hoxZ* (Bernhard et al. 1996; Kortlüke et al. 1992). The function of HoxZ is twofold: (1) It couples the MBH to an electron transport chain, thus mediating the conservation of energy arising from H_2 oxidation. (2) It mediates attachment of the MBH to the periplasmic aspect of the cytoplasmic membrane (Bernhard et al. 1997).

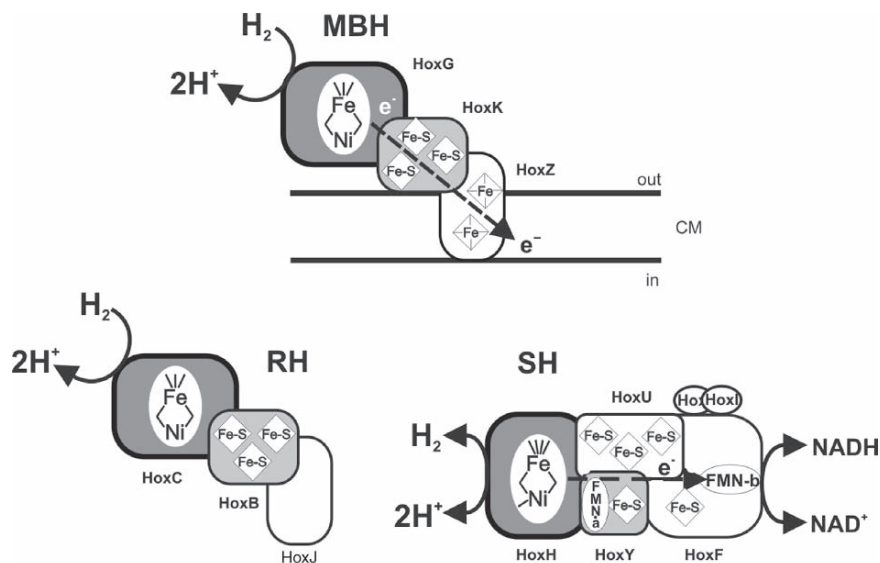


Fig. 3 The three hydrogenases of *R. eutropha* H16. Schematic representation of the subunit structure and the cofactor composition of MBH (membrane-bound hydrogenase), SH (Soluble hydrogenase) and RH (regulatory hydrogenase). The small and large subunits of the hydrogenase modules are shown in *light gray* and *dark gray*, respectively. A *dashed arrow* indicates electron transport. *CM* cytoplasmic membrane, *FMN* flavin mononucleotide. See text for details

Expression studies showed that deletions removing the region downstream of the MBH structural genes resulted in the formation of catalytically inactive hydrogenase polypeptides, indicating that posttranscriptional maturation processes – mediated by the products of genes in the MBH cluster – are required for hydrogenase activity (Kortlüke et al. 1992). The downstream accessory genes encode six proteins: HoxL (11.9 kDa), HoxO (17.7 kDa), HoxQ (30.9 kDa), HoxR (8.6 kDa), HoxT (19.7 kDa), and HoxV (41.5 kDa). Knock-out mutants defective for the corresponding genes belong to two groups: (1) Lesions in *hoxO* and *hoxQ* are totally devoid of MBH activity. (2) Deletions in *hoxL*, *hoxR*, *hoxT*, and *hoxV* contain reduced levels of MBH activity (Bernhard et al. 1996; Schubert et al. 2007). The molecular roles of the gene products are still unclear, but a picture is beginning to emerge. A detailed study on HoxO and HoxQ points to a role in chaperoning the small subunit until it undergoes assembly with the large subunit (Schubert et al. 2007). Ternery complexes of the small subunit precursor, HoxO and HoxQ, were identified in the cytoplasm of both wild-type cells and mutants devoid of the large subunit (Schubert et al. 2007). HoxV forms a complex with HoxL, and HoxL, in turn, forms a complex with immature HoxG (preHoxG) *in vivo* (Ludwig et al. submitted). According to the present model, HoxV and HoxL play a specific role in MBH maturation, acting as intermediate carriers of the $[\text{Fe}(\text{CN})_2\text{CO}]$ complex, which is destined for the active site in HoxG.

The formation of the $[\text{Fe}(\text{CN}^-)_2\text{CO}]$ complex is mediated by the products of the *hyp* genes (Fig. 2). Related proteins have been identified in many H_2 -oxidizing bacteria that contain $[\text{NiFe}]$ hydrogenases. In the MBH cluster of *R. eutropha* H16, the genes *hypA1*, *hypB1*, *hypC1*, *hypD1*, *hypE1*, *hypF1*, and *hypX* encode a complete set of Hyp proteins (Dernedde et al. 1996). Knock-out mutants defective in the *hyp* genes can be sorted into two phenotypic classes: (1) Lesions in *hypC1*, *hypD1*, and *hypE1* lack MBH and SH activity altogether. (2) Deletions in *hypA1*, *hypB1*, and *hypF1* are phenotypically unaltered (Dernedde et al. 1996). Paralogs of the latter three genes, designated *hypA2*, *hypB2*, and *hypF2*, are located downstream of the SH structural genes (Wolf et al. 1998) (Fig. 2). Deletion mutants defective for any one of these alleles are phenotypically wild-type. However, double knock-out mutants lacking both copies of a given protein are totally devoid of MBH activity. In the single-site mutants defective for *hypB1*, *hypD1*, and *hypE1* as well as in the double knock-out mutants, nickel-free precursor forms of the hydrogenase large subunit accumulate in the cytoplasm (Dernedde et al. 1996; Wolf et al. 1998). These findings are compatible with a model for Hyp-mediated $[\text{NiFe}]$ cofactor assembly based on extensive work on hydrogenase maturation in *E. coli* (reviewed by Böck et al. 2006; Forzi and Sawers 2007). According to this model, a HypC/HypD complex serves as a scaffold for processing Fe. Active site Fe is complexed with three diatomic ligands: two CN^- s and a CO. In an ATP-dependent reaction, HypE and HypF catalyze the transfer of two cyanides from carbamoyl phosphate to Fe waiting on the HypC/HypD complex. After addition of the third diatomic ligand, CO, the Fe complex is transferred to the hydrogenase large subunit. A recent study on the *R. eutropha* MBH revealed that an additional protein is involved in Fe processing: the $[\text{Fe}(\text{CN}^-)_2\text{CO}]$ complex is transferred from HypC/HypD to HoxV and then to preHoxG. Finally, HypA and HypB mediate incorporation of a Ni atom, completing the bimetallic center. The *E. coli* model does not address the question of the origin of the carbonyl ligand. Recent experiments on the hydrogenase 2 of *E. coli* and on the RH of *R. eutropha* H16 rule out the possibility that the carbonyl group originates in carbamoyl phosphate and indicate that free CO may be the source (Forzi and Sawers 2007; Lenz et al. 2007).

The final step in the maturation of MBH is the proteolytic processing of a C-terminal peptide of the large subunit. The $[\text{NiFe}]$ metallocenter-containing protein is the substrate for this reaction, which is catalyzed by a specific protease encoded by the gene *hoxM*. The mature MBH dimer undergoes translocation through the cytoplasmic membrane. As is the case for other cofactor-containing proteins, translocation is mediated by the Tat system (Bernhard et al. 2000; Schubert et al. 2007).

2.2.2 The SH: A Modular Enzyme that Couples NAD^+ Reduction to H_2 Oxidation

The second energy-transducing enzyme encoded on pHG1 is a cytoplasmic hydrogenase (Fig. 3). The SH is a multisubunit molecular machine, which couples H_2 oxidation to the reduction of NAD^+ . This capability gives *R. eutropha* H16 an enormous

metabolic advantage over H_2 -oxidizing lithotrophs that lack such an enzyme and, hence, must generate pyridine nucleotides via reverse electron flow through their respiratory chain. The SH is composed of an H_2 -oxidizing module consisting of the subunits HoxY (22.9 kDa) and HoxH (54.7 kDa), and an NADH dehydrogenase module consisting of the subunits HoxU (26 kDa) and HoxF (66.8 kDa) (Schneider and Schlegel 1976; Tran-Betcke et al. 1990). Biochemical determination of iron and sulfur content and EPR measurements, on the one hand, and the interpretation of conserved motifs in the deduced amino acid sequences, on the other, suggest the following structural model for the SH (Bleijlevens et al. 2004; Burgdorf et al. 2005a; Massanz and Friedrich 1999; Schneider et al. 1979; Tran-Betcke et al. 1990) (Fig. 3): HoxY is an iron–sulfur protein that contains a single Fe–S center and one flavin mononucleotide (FMN) group. HoxH harbors the active-site [NiFe] bimetallic center. HoxU contains up to three Fe–S centers. HoxF contains an FMN group and a Fe–S cluster in addition to the NAD(H) binding site. As in the case of MBH, the redox cofactors of the SH constitute an intramolecular circuit, conducting electrons from the active site of the hydrogenase module to NAD^+ . Depending on the purification conditions, either a tetrameric or a hexameric form of the SH is isolated. The hexameric form contains, in addition to HoxF, HoxU, HoxY, and HoxH, two identical HoxI subunits. HoxI provides a binding site for NADPH and could mediate the reductive activation of the enzyme by NADPH (Burgdorf et al. 2005b).

Like the MBH, the SH undergoes a posttranslational modification process, which converts inactive precursors to catalytically active holoenzyme. The fact that the full set of Hyp proteins is required for wild-type levels of SH activity indicates that maturation of the two enzymes proceeds in part via a common pathway. Likewise, the Ni–Fe-containing subunits of both enzymes undergo C-terminal proteolytic processing as the final maturation step. However, each enzyme has its own specific endopeptidase. For the SH this enzyme is encoded by the gene *hoxW* (Thiemermann et al. 1996). It is, nevertheless, certain that the maturation pathways of the two enzymes differ, since the MBH requires some special accessory proteins, for example, HoxV and HoxL, in addition to the Hyp proteins. Much remains to be done to clarify the specific steps in the maturation of the two enzymes.

2.2.3 Regulation of the H_2 -Oxidizing System: A Specialized Hydrogenase Controls Hydrogenase Gene Expression

The MBH and SH are organized in two large operons (Fig. 2). The MBH operon contains the structural and accessory genes for the MBH (*hoxKGZMLOQRTV*), *hyp* genes (*hypA1B1F1C1D1E1X*), the regulatory gene (*hoxA*), and possibly also the genes *hoxB*, *hoxC*, and *hoxJ*. Genetic data indicate that the corresponding transcript encompasses at least 17,000 nt (Schwartz et al. 1999). The SH operon consists of the structural and accessory genes for the SH (*hoxFUYHWI*) and the *hyp* genes (*hypA2B2F2*). The two operons are coordinately expressed. Regulation of expression is at the level of transcription and is governed by the NtrC-type response regulator HoxA in conjunction with its cognate histidine kinase HoxJ (Eberz and Friedrich

1991; Lenz and Friedrich 1998; Schwartz et al. 1998). These two proteins form the basis of an H_2 -dependent signal transduction chain. The MBH and SH operons are transcribed from σ^{54} promoters that are strictly dependent on HoxA. Since the two operons are separated by approx. 57 kb, they constitute a regulon.

Perhaps the most fascinating aspect of the H_2 -oxidizing system of *R. eutropha* H16 is the H_2 -sensing protein that mediates the initial response to the presence of H_2 (Fig. 3). The sensory protein is in essence a specialized [NiFe] hydrogenase and, hence, was designated regulatory hydrogenase (RH) (Kleihues et al. 2000). The subunits of the RH, HoxB and HoxC, are isologous to the small and large subunits, respectively, of the dimeric Ni–Fe hydrogenases (Kleihues et al. 2000) and have a similar subunit composition: HoxC contains a [NiFe] center and HoxB is an iron–sulfur protein (Pierik et al. 1998b). The exact composition of the iron–sulfur clusters is not known, but spectroscopic evidence indicates that they are different from their counterparts in standard hydrogenases (Buhrke et al. 2005). In accord with the cytoplasmic localization of the RH, HoxB lacks an N-terminal signal peptide. The C-terminal end of the protein carries a 55 amino acid extension that is absent in the small subunits of energy-transducing hydrogenases. The latter extension is essential for oligomerization. Functional RH is a dimer of heterodimers: (HoxBC)₂. This module in turn forms a complex with four molecules of HoxJ (Bernhard et al. 2001; Buhrke et al. 2004). The experimentally determined size of the (HoxBC)₂(HoxJ)₄ complex (382 kDa) is in good agreement with the predicted value (380 kDa) (Buhrke et al. unpublished results). The N-terminal part of HoxJ contains a CAP domain and a PAS domain (Buhrke et al. 2004). The CAP domain determines binding to the RH module (Lenz and Friedrich, unpublished results). PAS domains have been found in sensor proteins responsive to light, oxygen, and changes in redox state (Taylor and Zhulin 1999).

In the presence of H_2 , redox-related changes have been observed in the UV–visible spectra of the iron–sulfur clusters of HoxB, indicating that oxidation of H_2 actually takes place, with a concomitant transfer of electrons (Buhrke et al. 2005; Friedrich et al. 2005). At present it is not known how the primary signal is transmitted to HoxJ, the histidine kinase. In any case, the PAS domain is essential for signal transmission (Buhrke et al. 2004). Unlike canonical two-component systems, the H_2 -triggered alteration in HoxJ suppresses the net phosphorylating activity of the enzyme. This leads to a net decrease in the pool of phospho-HoxA. Unphosphorylated HoxA induces transcription from the *hox* promoters. In addition to control by the H_2 -dependent signaling pathway, transcription of the hydrogenase genes is susceptible to catabolite repression, which varies in degree depending on the quality of the organic growth substrates (Lenz and Friedrich 1998).

2.3 The *cbb*_p Operon

The genes for the fixation of CO_2 via the reductive pentose phosphate pathway (Calvin–Benson–Bassham cycle) are located in a 12.8-kb region adjacent to the MBH gene cluster (Fig. 1). The 12 *cbb* genes, which form an operon transcribed

from a single promoter, allow *R. eutropha* H16 to grow lithoautotrophically on H₂ and CO₂ and organoautotrophically on formate and CO₂ (Bowien and Kusian 2002) (Fig. 4). The first two genes, *cbbL_p* and *cbbS_p*, encode the large and small subunits of a red-type form I (L₈S₈) ribulose-1,5-bisphosphate carboxylase/oxygenase. *cbbP_p* codes for phosphoribulokinase, the second key enzyme of the cycle. *cbbE_p*, *cbbF_p*, *cbbT_p*, *cbbG_p*, *cbbK_p*, and *cbbA_p* code for ribulose-phosphate 3-epimerase, fructose-1,6-bisphosphatase, transketolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, and fructose-1,6-bisphosphate aldolase. The *R. eutropha* H16 fructose-1,6-bisphosphatase is bifunctional: it has sedoheptulose-1,7-bisphosphatase activity. The fructose-1,6-bisphosphate aldolase has sedoheptulose-1,7-bisphosphate aldolase activity. The *cbb_p* operon also contains a gene (*cbbZ_p*) encoding a 2-phosphoglycolate phosphatase. The functions of the *cbbX_p* and *cbbY_p* gene products are not known. A second, complete copy of the operon, designated *cbb_c*, is located on chromosome 2 (Kusian et al. 1995). Adjacent to the *cbb_c* operon is a gene encoding a LysR-type regulator, which controls the expression

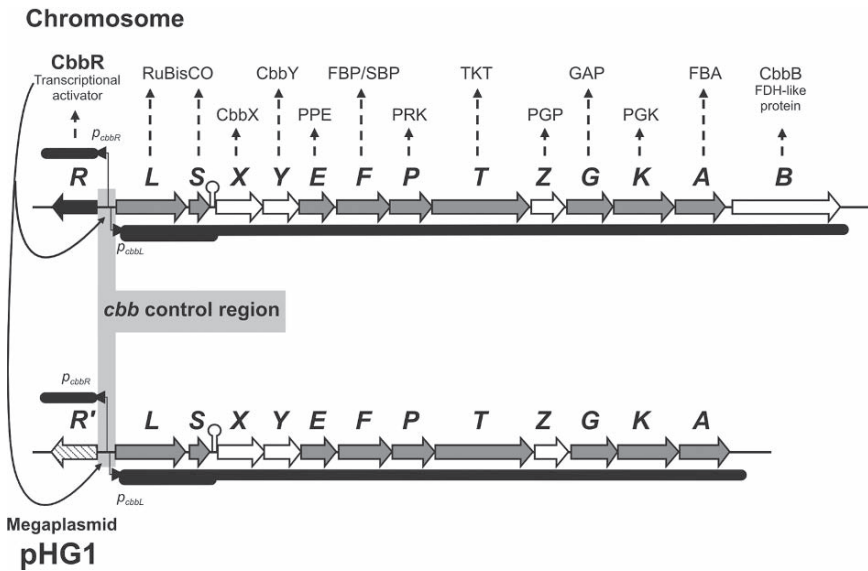


Fig. 4 Organization and regulation of the *cbb* operons of *R. eutropha* H16. The gray arrows indicate genes for enzymes of the Calvin–Benson–Bassham cycle. The solid black arrow and the hatched arrow represent the intact and defective copies of the regulatory gene *cbbR*. The divergent promoters P_{cbbR} and P_{cbbL} are marked by arrowheads. Thick bars indicate transcripts. Putative hairpin structures between *cbbS* and *cbbX* are represented by loops. Abbreviations: *RuBisCO* ribulose-1,5-bisphosphate carboxylase/oxygenase; *CbbX*, *CbbY*, products of unknown function; *PPE* pentose-5-phosphate 3-epimerase; *FBP* fructose-1,6-/sedoheptulose-1,7-bisphosphatase; *PRK* phosphoribulokinase; *TKT* transketolase; *PGP* 2-phosphoglycolate phosphatase; *GAP* glyceraldehyde-3-phosphate dehydrogenase; *PGK* phosphoglycerate kinase; *FBA* fructose-1,6-/sedoheptulose-1,7-bisphosphate aldolase; *CbbB* formate dehydrogenase (FDH)-like protein. (With kind permission from Springer Science+Business Media: Archives of Microbiology, Genetics and control of CO₂ assimilation in the chemoautotroph *Ralstonia eutropha*. 178 (2002) p. 86, Bowien, B. and Kusian, B., Fig. 1)

of both operons. At least one effector molecule, phoshoenol pyruvate (PEP), interacts with CbbR to signal the carbon status of the cell (Grzeszik et al. 2000).

2.4 Denitrification Genes

R. eutropha H16 can grow under anoxic conditions utilizing NO_3^- as an alternative electron acceptor for its respiratory chain – an important metabolic option for any organism that has to cope with fluctuating O_2 levels in its environment. *R. eutropha* H16 performs the complete denitrification sequence:



The genetic determinants for the various components of the denitrification apparatus are encoded on megaplasmid pHG1 as well as on the two chromosomes. Thirty denitrification-related genes are clustered in a 76-kb-large region of pHG1 (Cramm et al. 1997; Schwartz et al. 2003; Siddiqui et al. 1993) (Fig. 1). The first step in the denitrification sequence, the conversion of NO_3^- to NO_2^- is catalyzed by two membrane-bound, dissimilatory nitrate reductase isoenzymes (Fig. 5). The megaplasmid

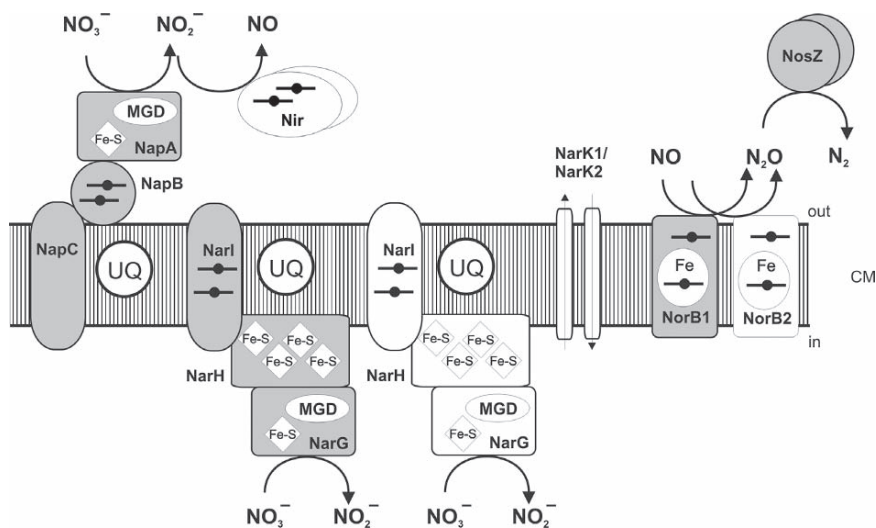


Fig. 5 Overview of the denitrification apparatus of *R. eutropha* H16. The hatched area represents the cytoplasmic membrane (CM). Gray symbols represent megaplasmid-encoded components. Unfilled symbols are the products of chromosomal genes. Iron-sulfur centers are indicated by diamonds (Fe-S). Unshaded circles symbolize ubiquinone (UQ). Heme groups are indicated by a dot on a bar. Molybdopterin guanine dinucleotide cofactors (MGD) are labeled. See text for details

genes *narG1*, *narH1*, and *narI1* encode the three subunits of one of the isoenzymes. The membrane-bound nitrate reductase is a molybdoenzyme and the products of the two downstream genes, *nifM* and *mobB*, may participate in biosynthesis of the molybdenum cofactor. The duplicate genes *narK1* and *narK2* encode representatives of the major facilitator superfamily, which may be involved in nitrate/nitrite transport. The products of genes *narX* and *narL* suggest a typical two-component regulatory system that could direct expression of the *nar* genes.

pHG1 encodes a third nitrate reductase that resides in the periplasm (Siddiqui et al. 1993). It has been postulated that this enzyme is important during the transition from aerobic to anaerobic conditions. Genes *napA* and *napB* encode the two core subunits of this enzyme. The deduced amino acid sequences suggest that NapA contains a molybdopterin guanine dinucleotide cofactor (MGD) and a [4Fe-4S] center and that NapB is a *c*-type cytochrome. The product of a third gene, *napC*, may serve as membrane anchor and/or electron donor.

Thus, three enzymes, two of which are encoded on pHG1, engage, either simultaneously or under different growth conditions, in the dissimilatory reduction of nitrate to nitrite. The next step in the denitrification sequence, the conversion of nitrite to nitric oxide, is catalyzed by nitrite reductase. Only a fragmented gene for this enzyme is present on pHG1. A unique copy of a gene encoding a functional *cd₁*-type nitrite reductase (*nirS*) is located on chromosome 2 (Pohlmann et al. 2006; Rees et al. 1997).

The product of the reduction of nitrite, the toxic gas NO, is in turn reduced by two nitric oxide reductase isoenzymes (Cramm et al. 1997). One of them is encoded by the megaplasmid gene *norB1*, the other by a chromosomal allele *norB2*. Characterization of the purified isoenzyme derived from the megaplasmid allele revealed that it belongs to the qNOR subgroup and that it contains two *b*-type heme groups and one non-heme Fe (Cramm et al. 1999). Based on sequence similarity, the chromosomally encoded isoenzyme is likely to have similar properties.

Another gene of the pHG1 denitrification cluster, *flhp*, encodes a flavohemoprotein (Fhp) (Cramm et al. 1994). While the precise role of the *R. eutropha* Fhp is unknown, it appears to be involved in nitric oxide metabolism.

The final step in denitrification, the reduction of N₂O to N₂, is catalyzed by a nitrous oxide reductase encoded by the *nosZ* gene of pHG1 (Zumft et al. 1992). N₂O reductases are interesting on account of their unusual copper centers (Zumft 2005). The neighboring genes *nosD*, *nosF*, *nosY*, *nosL*, and *nosX* may be involved in the maturation of NosZ.

The regulation of the *R. eutropha* H16 denitrification genes is complex. Monitoring of the activity of denitrification gene promoters showed that the *nirS* and *nosZ* promoters are coordinately regulated in response to O₂ (Pohlmann et al. 2000). The megaplasmid and chromosomal *nor* operons are transcribed from σ^{54} -dependent promoters controlled by the cognate NO-sensing regulators NorR1 and NorR2. These functionally interchangeable regulators respond to the presence of NO, activating the *nor* promoters (Pohlmann et al. 2000). Induction of the *nor* genes was delayed with respect to the *nir* and *nos* promoters, which is congruent with the finding that *nor* expression is dependent on NO.

Interestingly, pHG1 also encodes a general metabolic function that is not related to denitrification but is required for anaerobic growth. The genes *nrdD* and *nrdG*

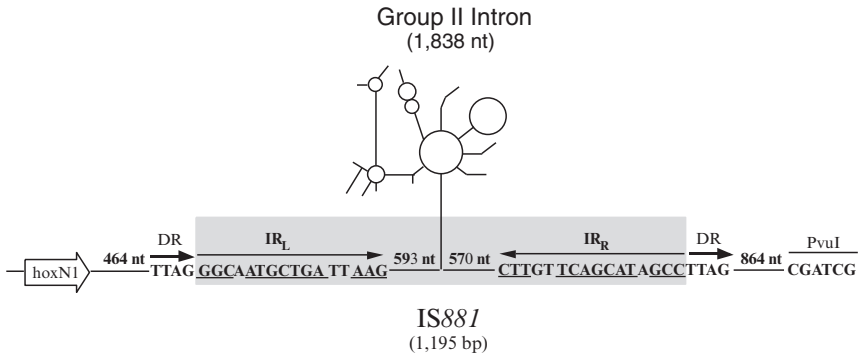
code for an anaerobic class III ribonucleotide reductase and its activase (Siedow et al. 1999). This enzyme system synthesizes 2'-deoxyribonucleotides under anaerobic conditions via a radical mechanism.

2.5 Catabolism of Aromatic Substances

It has been known for some time that *R. eutropha* H16 can grow on benzoate and other aromatic compounds. It was, therefore, not surprising that the megaplasmid sequencing project uncovered a 25-kb-large segment of pHG1 encoding enzymes involved in the degradation of aromatics (Schwartz et al. 2003) (Fig. 1). This region presents a somewhat confusing picture. On the one hand, the deduced amino acid sequences of the gene products add up to incomplete pathways in which key enzymes are obviously lacking. On the other hand, there is redundancy, with two or more genes encoding the same function. For instance, the products of ORFs PHG394 and PHG405 predict proteins with similarity to muconate cycloisomerases and the deduced product of ORF PHG404 is isologous to muconolactone isomerases. Both enzymes catalyze steps in standard *ortho* degradation pathways. However, genes for other key enzymes such as catechol-1,2-dioxygenase and 3-oxoadipate enol-lactone hydrolase are missing. A complete *ortho* pathway for the degradation of benzoate is encoded on chromosome 1 (Pohlmann et al. 2006). Nevertheless, three of the degradation genes on pHG1 could represent a biologically relevant branch of a chromosomally encoded pathway. The deduced products of genes *mmlH*, *mmlI*, and *mmlJ* are obviously similar to three enzymes of the closely related strain *R. eutropha* JMP134: a 4-methylmuconolactone transporter, a methylmuconolactone methyl isomerase, and a methylmuconolactone isomerase (Erb et al. 1998). 4-Methylmuconolactone, a “dead-end” product of the degradation of methylated catechols via standard *ortho* pathways, is secreted by aromatic-degrading bacteria. Thus, the extension of the *ortho* pathway by the above mentioned gene products would allow *R. eutropha* H16 to exploit an additional growth substrate (Schwartz et al. 2003).

2.6 The pHG1 Junkyard Region

A 72-bp segment of pHG1 (nt 22,390–94,364) contains a multitude of full-length and partial transposase genes (17 ORFs) and phage-type integrases/recombinases (22 ORFs) (Schwartz et al. 2003). This region, termed “junkyard region,” is, due to its repetitive structure, probably inherently instable. The most remarkable feature of the junkyard region is a 3,033-bp IS element belonging to the IS5 family (Fig. 6). This mobile element, designated IS881, is bracketed by imperfect inverted repeats and flanked by 4-bp direct repeats. The latter suggest that the present flanking sequences correspond to the target site prior to insertion of IS881. Inspection of the IS881 transposase gene revealed that its open reading frame is interrupted by a 1,838-bp sequence inserted



a

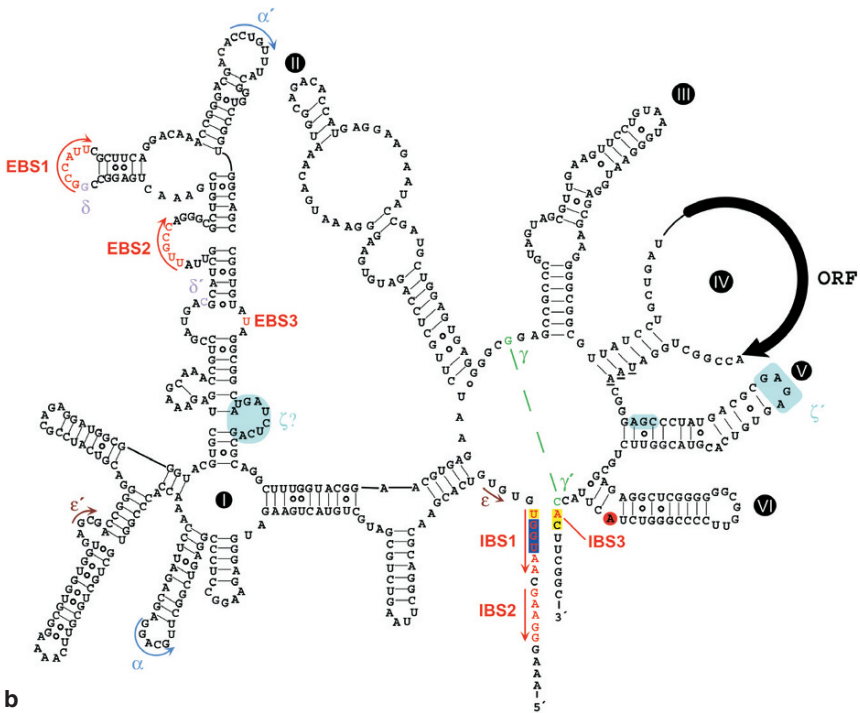


Fig. 6 (a) Diagrammatic representation of IS881 and the flanking sequences on pHG1 (nt 21,335–26,767). The IS element is highlighted in *gray*. The terminal inverted repeats (IRs) are overlined with arrows and underlined to indicate mismatches. The direct repeats (DRs) of TTAG created by the insertion are overlined with thick arrows. A restriction site for PvuI is overlined. The location of gene *hoxN1* is given for orientation and the position of the group II intron is indicated schematically. (b) Secondary structure model of the group II intron *R.e.II*. Exon-binding sites (EBS) and intron-binding sites (IBS) are highlighted in *red*. Other potential 3D interactions are indicated by color-coded pairs of Greek letters (α - α' , γ - γ' , δ - δ' , ϵ - ϵ' , ζ - ζ'). Ribozyme domains are labeled with Roman numerals in *black circles*. The thick *black arrow* in domain *dIV* labeled “ORF” is the coding sequence for an intron-encoded protein (IEP). The boxed nucleotides indicate Trp (*blue*) and Tyr (*yellow*) codons of the IS881 transposase gene. Non-Watson–Crick basepairs are indicated by circles (Panel b reprinted from Schwartz et al. 2003, with permission from Elsevier)

within the codon for Tyr. The conceptual transcript of the inserted DNA bears the hallmarks of group II introns, such as are found in organelles of fungi, plants, and protists, as well as in bacteria and some methanogenic archaea. Typical features of group II introns, which are also present in the sequence in IS88I, include a six-domain architecture and the 5' splice site N*GUGYG (Fig. 6) (Bonen and Vogel 2001; Martínez-Abarca et al. 2000). The group II intron within IS88I was designated *R.e.II*. A prerequisite for the mobility of group II introns is the so-called intron-encoded protein (IEP), a multifunctional enzyme with reverse transcriptase and RNA maturase activities. RT-PCR-based screening failed to detect splicing of *R.e.II* in vivo (Eitinger, personal communication). A frameshift mutation within the IEP ORF predicts a truncated gene product, suggesting that *R.e.II* is crippled with respect to retrohoming and retrotransposition. Since, on the one hand, *R.e.II* disrupts the transposase ORF of IS88I and, on the other hand, related IS elements with intact transposase genes are not present in the *R. eutropha* H16 genome, it is likely that IS88I is also a cripple.

2.7 Megaplasmid–Chromosome Synergy

The megaplasmid pHG1 carries the genetic information for biochemical pathways that significantly expand the metabolic versatility of the host organism. Interestingly, for some key processes, the ensemble of genes on the megaplasmid is not complete and “self-contained.” Rather, chromosomal information is required for the corresponding metabolic system to operate. A remarkable example is the growth of *R. eutropha* H16 on nitrate under anaerobic conditions. A set of denitrification enzymes for the sequence of reactions that convert NO_3^- to N_2O is encoded on chromosome 2. These enzymes are sufficient for the growth-supporting utilization of NO_3^- as an alternative electron acceptor for respiration. However, a basal requirement for anaerobic metabolism in general is the capacity to synthesize deoxyribonucleotides in the absence of O_2 . In *R. eutropha* H16, the only enzyme system that can catalyze this reaction under anaerobic conditions in the absence of vitamin B12 is the anaerobic ribonucleotide reductase/activase encoded by the genes *nrdD* and *nrdG* on pHG1. Thus, the chromosomal denitrification genes are biologically irrelevant without the megaplasmid pHG1. On the contrary, pHG1 encodes a nitrate reductase, an NO reductase, and an N_2O reductase, but not a nitrite reductase, the enzyme responsible for converting nitrite to NO. This means that without the chromosomally encoded nitrite reductase, there would be no intrinsic substrate for the megaplasmid-encoded NO reductase and N_2O reductase.

Another example of mutual interdependence of megaplasmid and chromosome is the utilization of methylmuconolactone. Three genes on the megaplasmid code for uptake and degradation of this substrate to methylenol-lactone. The catabolism of the latter intermediate is dependent on chromosomally encoded enzymes.

Finally, pHG1 carries a set of genes for RuBisCO and the other enzymes of the Calvin–Benson–Bassham cycle. A duplicate set of genes is located on chromosome 2. The transcription of both the megaplasmid and the chromosomal genes is dependent

on the positive regulator CbbR encoded by the sole intact *cbbR* gene on chromosome 2. A megaplasmid copy of *cbbR* is recognizable, but it is no longer functional due to multiple mutations.

The examples mentioned above suggest that megaplasmid pHG1 is not merely a transient visitor in its host strain *R. eutropha* H16, but rather an integral part of the genome. The functions encoded on megaplasmid and chromosomes mesh with and complement each other. In the case of the *cbbR* genes, we have a snapshot of the ongoing coevolution that is fine-tuning the three replicons of *R. eutropha* H16.

3 Megaplasmid pHCG3 of *Oligotropha carboxidovorans* OM5

In *Oligotropha carboxidovorans* OM5 the genes for the carbon monoxide dehydrogenase (CODH) and for other key enzymes of lithoautotrophic metabolism are located on the 133,058-bp megaplasmid pHCG3 (Fig. 7).

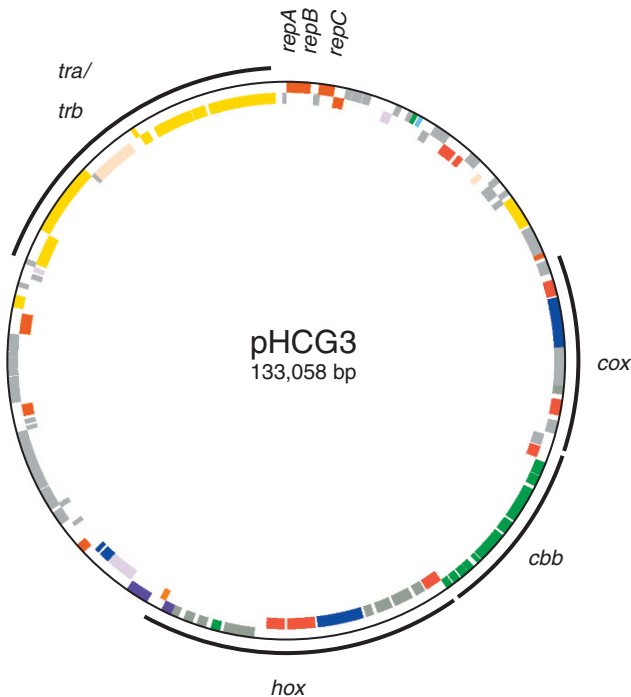


Fig. 7 Circle diagram of *O. Carboxidovorans* OM5 megaplasmid pHCG3. The outer and inner concentric rings represent coding regions (CDSs) transcribed in the clockwise and counterclockwise directions, respectively. CDSs are color-coded on the basis of the COG functional categories. See Fig. 1 for a key to the color-coding used for the COG functional categories. Arcs on the perimeter of the circle mark gene clusters. See text for details

3.1 Plasmid Functions and General Organization of pHCG3

pHCG3 is a *repABC*-type replicon (Fuhrmann et al. 2003). The three-gene set (*repA*, *repB*, and *repC*) is characteristic of a plasmid backbone that is widespread in alpha-proteobacterial plasmids. The first two genes, *repA* and *repB*, encode partitioning proteins, which ensure the equal distribution of the plasmid copies at cell division (Bignell and Thomas 2001). The product of gene *repC* is a DNA-binding protein that is believed to mediate initiation of replication (Romero and Brom 2004). RepC sequences have been used as a basis for analyzing the phylogenetic relationships of megaplasms. A comparison of the pHCG3 RepC with its orthologs revealed the highest similarities to RepCs of rhizobial megaplasms (Fig. 8). Likewise, the nucleotide sequences of the *repB*–*repC* intergenic regions of pHCG3 and rhizobial megaplasms are clearly conserved. The *repB*–*repC* intergenic region of *repABC*-type replicators contains signal sequences for plasmid partitioning functions (not *oriV* sites as is often erroneously claimed in sequence annotations) (Romero and Brom 2004) and may also encode a small RNA involved in the control of plasmid copy number (MacLellan et al. 2005; MacLellan et al. 2006). In this context it is interesting that strains of *Bradyrhizobium japonicum* and other *Rhizobiaceae* are capable of oxidizing CO. This is not surprising considering that root nodules are a major biogenic source of CO. Root nodules contain leghemoglobin and CO is a product of the degradation of heme originating from leghemoglobin catabolism (King and Weber 2007). Furthermore, two of the most closely related RepC proteins are derived from the plasmids pTB1 of *Roseobacter denitrificans* OCh 114 and pSD25 of *Ruegeria* sp. PR1b. Both strains are marine bacteria that inhabit the upper layers of the euphotic zone, where photochemical decomposition of dissolved organic matter leads to the production of significant amounts of CO. The recently published genome sequence of *R. denitrificans* reveals that this organism carries genes for an aerobic CODH (Swingley et al. 2007).

pHCG3 carries a set of *tra/trb* genes for conjugative transfer in a 25 kb-large region. The deduced products of these genes show the highest similarity to orthologs of the *Agrobacterium tumefaciens* Ti plasmid (Fuhrmann et al. 2003). No experimental data are available on conjugative transfer of pHCG3.

Aside from the *tra/trb* cluster, pHCG3 carries three groups of functionally related genes. These genes are arranged in three compact, contiguous clusters. The *cox* genes encode the CODH, the *cbb* genes encode Calvin-cycle enzymes, and the *hox* genes code for a membrane-bound NiFe hydrogenase. Outside of these four clusters are numerous ORFs, the functions of which are largely unknown (Fuhrmann et al. 2003).

3.2 The *cox* Genes

The genetic determinants for the CODH are located on a 14.5-kb segment of pHCG3 (Fig. 7). The cluster consists of 12 genes. The three closely spaced or overlapping genes *coxM*, *coxS*, and *coxL* encode the three subunits of the CODH. *coxM*

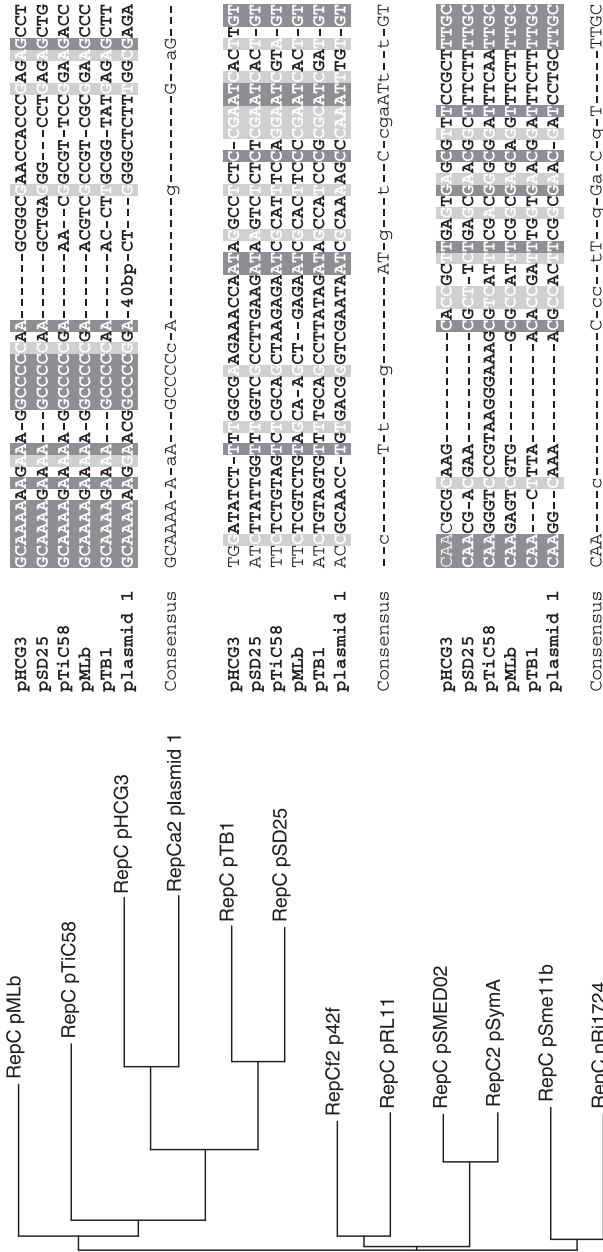


Fig. 8 (a) Dendrogram based on an alignment of deduced RepC sequences. Alignments were done using the EBI Clustal W resource (1.83) (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). RepC amino acid sequences were deduced from *repC* genes of pHCG3 (NC_005873), pSD25 of *Ruegeria* sp. PR1b (AF416331), pTiC58 of *Agrobacterium tumefaciens* C58 (AF126445), pMLb of *Mesorhizobium loti* MAFF303099 (AP003017), pTB1 of *Roseobacter denitrificans* Och 114 (CP000464) and plasmid 1 of *Mesorhizobium* sp. BNC1 (CP000389), p42f of *Rhizobium etli* CFN 42 (CP000138), pRL11 of *Rhizobium leguminosarum* bv. *viciae* 3841 (AM236085), pSMED02 of *Sinorhizobium medicae* WSM419 (CP000740), pSymA of *Sinorhizobium meliloti* 1021 (AE006469), pSme11b (pSmeSM11b) of *Sinorhizobium meliloti* SM11 (EF066650), and pRi1724 of *Agrobacterium rhizogenes* (AP002086) (GenBank accession nos. of the plasmids are given in parentheses). (b) Alignment of nucleotide sequences of the *repB-repC* intergenic region. The five closest relatives of pHCG3 (based on the RepC alignments) were chosen for this analysis. Matching positions in all six sequences are highlighted in dark gray. Positions with 5 out of 6 matches are highlighted in light gray. The consensus sequence is given below the alignment (lower case letters indicate positions with a mismatch)

encodes the medium-sized (30.2 kDa) subunit, which is an FAD-containing flavo-protein. *coxS* and *coxL* code for the small (17.8 kDa) and large (88.7 kDa) subunits, respectively. The former is an iron–sulfur protein and the latter is a molybdoprotein harboring the active site. High resolution crystal structures together with the results of EXAFS and EPR spectroscopy have revealed a fascinating picture of the intricate structure of the enzyme (Dobbek et al. 1999; Meyer et al. 2000; Dobbek et al. 2001). The CODH holoenzyme is a dimer of identical heterotrimers (Fig. 9). Each heterotrimer consists of one molecule each of the three subunits, with the iron–sulfur protein wedged in between the molybdoprotein and the flavoprotein. The heterotrimers make contact with each other via the molybdoproteins at dimerization interfaces consisting of several α -helices and β -sheets.

The active site of the enzyme contains the dinuclear heterometal cofactor [CuSMoO₂] complexed with an organic cofactor known as molybdopterin–cytosin dinucleotide (MCD) buried deep in the central part of the large subunit. To reach the active site, molecules of substrate must pass through a channel 17 Å long and 6 Å in diameter. The Mo ion (Mo(VI) in the oxidized form and Mo(IV) in the reduced form of the enzyme) is coordinated by five ligands with distorted square pyramidal geometry. Two of the ligands are enedithiolate sulfur atoms of the MCD, two others have been modeled as

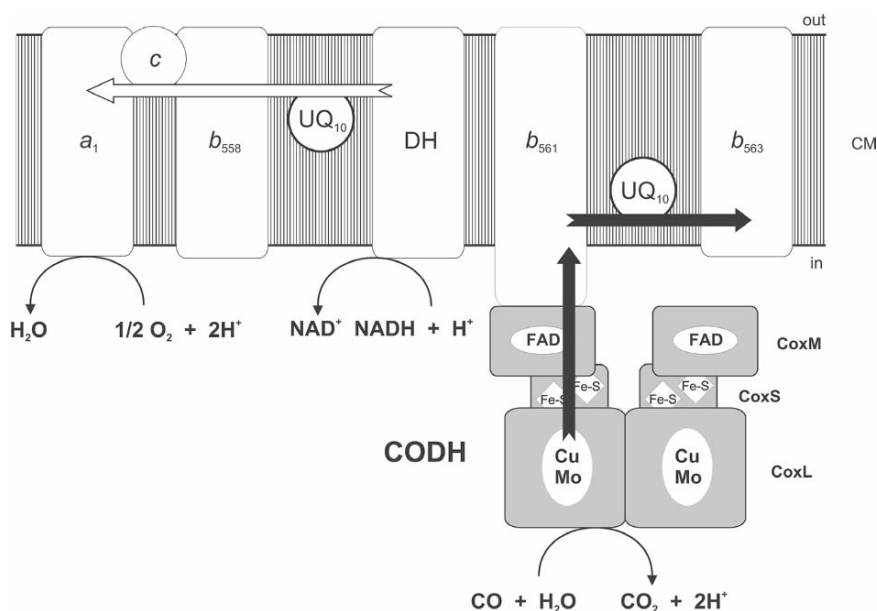


Fig. 9 CO-insensitive and CO-sensitive electron transport chains of *O. carboxidovorans* OM5. Schematic representation showing the megaplasmid- (gray) and chromosomally encoded (unshaded) components of electron transport. Iron–sulfur clusters of the CO dehydrogenase (CODH) are symbolized by diamonds, other cofactors by ellipses. *Solid* and *open* arrows indicate CO-insensitive and CO-sensitive electron transport, respectively. *CM* cytoplasmic membrane, *FAD* flavin adenine dinucleotide, *UQ₁₀* ubiquinone-10. See text for details. Based on Meyer et al. (1993)

oxo groups. The fifth ligand is a sulfur atom. This sulfur bridges the Mo and a copper atom. A region of electron density near the Mo atom was originally interpreted as Se and later recognized as Cu (Dobbek et al. 1999; Dobbek et al. 2001; Dobbek et al. 2002). This assignment was subsequently confirmed by EXAFS studies (Dobbek et al. 2002; Gnida et al. 2003). The Cu atom is covalently bound to cysteine residue 388 of the large subunit. Thus, Mo and Cu constitute the active site bimetallic center with S as bridging ligand. The organic component of the active site, MCD, consists of a pterin moiety covalently bound to a cytosine dinucleotide and is noncovalently bound to the large subunit.

CoxS, the small subunit, contains two [2Fe-2S] centers. The [2Fe-2S] cluster in the C-terminal domain of the protein is proximal to the [CuSmO₂] cofactor of the molybdoprotein (Dobbek et al. 1999). The second cluster is located in the N-terminal domain of CoxS, distal to the active site. This cluster is on the surface of the subunit in the interface that makes contact with the flavoprotein. The flavoprotein, CoxM, contains a single molecule of noncovalently bound FAD. Two conserved glycine pairs are thought to mediate binding of the cofactor.

The overall structure of the CODH suggests an intramolecular electronic circuit, in which electrons liberated at the active site flow via the [2Fe-2S] clusters to FAD, before being transferred to the membrane-bound cytochrome *b*₅₆₁. The distances between the redox components of the CODH as derived from the crystal structure are compatible with this idea: Mo–Fe_{proximal}: 14.6 Å; Fe_{proximal}–Fe_{distal}: 12.4 Å; Fe_{distal}–FAD_{CH₃7a}: 8.7 Å (Meyer et al. 2000). The electron transport pathways in the two heterotrimers operate independently of each other.

Heterologous expression studies, in which the CODH structural genes *coxMSL* were introduced into different strains of *E. coli*, yielded only inactive enzyme, indicating that posttranslational maturation mediated by specialized accessory proteins is a prerequisite for catalytically active CODH. Further support for this interpretation came from experiments with a derivative of *O. carboxidovorans* OM5 cured of pHCG3 and carrying *coxMSL* on an expression vector. This system produced CODH containing Fe and FAD but lacking intact MCD cofactor (Meyer et al. 2000). Downstream of *coxMSL* are a set of four closely spaced or overlapping genes designated *coxD*, *coxE*, *coxF*, and *coxG* (Santiago et al. 1999). The predicted products of these genes (33.3, 44.2, 29.3, and 21.5 kDa) are similar to CODH-related ORFs of other organisms (Santiago et al. 1999). Interposon mutants defective for *coxD*, *coxE*, and *coxF* produced CODH species containing both iron–sulfur clusters, the MCD cofactor and FAD, but lacking catalytic activity (Meyer et al. 2000). Characterization of CODH isolated from the *coxD* and *coxE* mutants revealed that Cu was absent from the active site, pointing to a role for the corresponding gene products in adding this metal to the bimetallic center (Fuhrmann et al. 2003). The *coxG* mutant contains wild-type CODH, but the enzyme accumulates in the cytoplasm, indicating that *coxG* is involved in mediating attachment of CODH to the membrane. The phenotype of the *coxF* mutant remains to be analyzed.

The genes *coxB* and *coxC* are located upstream of *coxM* and could code for proteins of 35.4 and 42.8 kDa, respectively (Santiago et al. 1999). The predicted product of *coxC* carries a signal peptide at its N-terminus followed by a so-called MHYT domain (Galperin et al. 2001; Santiago et al. 1999). This domain consists

of six transmembrane helices. The helix-forming stretches are arranged in pairs, with each pair bracketing an arginine-rich stretch of amino acids. The MHYT domain is characteristic of bacterial signaling proteins. A second feature that points to a regulatory role for CoxC is the C-terminal LytTR DNA-binding domain (Nikolskaya and Galperin 2002). This signature is found in various bacterial response regulators. It has been postulated that CoxC acts as a CO sensor (Fuhrmann et al. 2003; Galperin et al. 2001). The deduced sequence of CoxB predicts an N-terminal signal peptide. Aside from this there are no clues as to its function.

The genes *coxH*, *coxI*, and *coxK* lie downstream of *coxG* (Santiago et al. 1999). CoxH is a paralog of CoxC (27% identity) and has a similar domain architecture. This suggests that CoxH, like CoxC, could be involved in CO sensing. However, a mutant defective for *coxH* is not impaired in CO-dependent induction of CODH (Santiago et al. 1999). CoxI (36.6 kDa) is related to the *Rhodobacter capsulatus* XdhC protein, which is required for insertion of the molybdopterin cofactor into the molybdoenzyme xanthine dehydrogenase (Leimkühler and Klipp 1999). CoxK (30.9 kDa) is a membrane protein of unknown function.

Transposon mutagenesis has revealed three other megaplasmid genes involved in the expression of active CODH (Santiago et al. 1999). One of these genes is isologous to the *E. coli lon* gene, which codes for the Lon protease. Another mutant carried a transposon insertion in a *cycH* ortholog. *cycH* genes have been found in various bacteria and they are involved in cytochrome *c* biogenesis. A third insertion was mapped to an open reading frame designated *orfX*. The function of the corresponding gene product is unknown.

Analysis of transcripts from the *cox* gene cluster showed that *coxB*, *coxC*, *coxL*, *coxD*, *coxG*, *coxI*, and *coxK* are coexpressed in a strictly CO-dependent fashion (Santiago et al. 1999). However, transcript levels during growth on organic substrates are significantly lower, suggesting a catabolite repression mechanism. Additional studies are needed to clarify this and other open questions regarding regulation.

3.3 The *cbb* Genes

A cluster of 13 genes encoding the enzymes of the Calvin–Benson–Bassham cycle for CO₂ fixation is located in a 13.3-kb region adjacent to the *cox* cluster (Fuhrmann et al. 2003) (Fig. 7). The first gene of the group, *cbbR*, encodes a LysR-type regulator that probably controls the expression of the other genes of the cluster. *cbbL* and *cbbS* encode the large and small subunits of the ribulose-1,5-bisphosphate carboxylase/oxygenase. *cbbP* codes for phosphoribulokinase, the second key enzyme of the cycle. *cbbF*, OC145, OC146, *cbbK*, *cbbA*, OC152, and *cbbZ* code for fructose-1,6-bisphosphatase, transketolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, fructose-1,6-bisphosphate aldolase, ribulose-phosphate 3-epimerase, and phosphoglycolate phosphatase. The functions of the *cbbX* and *cbbY* gene products are not known.

3.4 The *hox* Genes

A 23-kb region of pHCG3 contains genes for a dimeric NiFe hydrogenase (Fuhrmann et al. 2003), enabling the organism to utilize H₂ as an alternative energy source for lithotrophic growth (Fig. 7). The small (35.4 kDa) and large (67.4 kDa) subunits of the enzyme are encoded by the genes *hoxS* and *hoxL*, respectively. *hoxZ* codes for a *b*-type cytochrome, the characteristic primary electron acceptor for membrane-bound hydrogenases of aerobic bacteria. A set of *hyp* genes predicts the general pathway for [NiFe] center assembly, suggesting that maturation of the *O. carboxidovorans* hydrogenase proceeds via mechanisms similar to those operating in *E. coli* and *R. eutropha* H16 (see above). Membrane-bound hydrogenase accessory genes resembling those in *R. eutropha* H16 are also present. The regulatory components encoded in the cluster are reminiscent of *R. eutropha* H16 regulators, pointing to an analogous mechanism of transcriptional regulation. The genes *hoxA* and *hupT* encode an NtrC-like response regulator and a histidine protein kinase, respectively. The products deduced from ORFs OC166 and OC165 have the typical features of the small and large subunits, respectively, of regulatory hydrogenases. The *O. carboxidovorans* proteins are similar to their counterparts in *R. eutropha* H16. The product of ORF OC174 is a high-affinity nickel transporter belonging to the UreH family (Eitinger et al. 2005). The genes immediately downstream of OC174 encode a NikR-type nickel-responsive transcriptional regulator and an outer-membrane nickel transporter (Rodionov et al. 2006).

3.5 Megaplasmid–Chromosome Synergy

In the case of the megaplasmid pHG1 of *R. eutropha* H16, evidence was discussed suggesting that the megaplasmid is an integral part of the genome (see Sect. 2.7). The CODH-dependent electron transport chain of *O. carboxidovorans* OM5 points to a similar status for pHCG3. *O. carboxidovorans* OM5 has a branched electron transport chain (Cypionka and Meyer 1983a, b). One branch contains the cytochromes *b*₅₅₈, *c*, and *a*₁. This part of the electron transport chain is susceptible to inhibition by CO. The second branch contains the cytochromes *b*₅₆₁ and *b*₅₆₃. Cytochrome *b*₅₆₁ is the primary electron acceptor of the CODH and is also essential for anchoring the CODH to the inner aspect of the cytoplasmic membrane (Rohde et al. 1984, 1985) (Fig. 9). Cytochrome *b*₅₆₃ is a CO-insensitive terminal oxidase and is, therefore, an essential prerequisite for utilization of CO by the organism. The latter two cytochromes are chromosomally encoded (Meyer et al. 1993). Without these cytochromes, the megaplasmid borne CODH genes are biologically irrelevant. On the contrary, in the absence of CODH, a biological role of the specialized cytochromes is questionable. Thus, megaplasmid and chromosomally encoded components the CO oxidation system interdigitate.

Acknowledgments The author is grateful to R. Cramm, H. Dobbek, T. Eitinger, B. Friedrich and O. Lenz for critical reading of the manuscript and to O. Lenz and B. Kusian for providing illustrations.

References

- Andersen K, Tait RC, King WR (1981) Plasmids required for utilization of molecular hydrogen by *Alcaligenes eutrophus*. *Arch Microbiol* 129:384–390
- Arp DJ (1992) Hydrogen recycling in symbiotic bacteria. In: Stacey GS, Burris RH, Evans HJ (eds) Biological nitrogen fixation. Chapman & Hall, New York, pp 432–460
- Behki RM, Selvaraj G, Iyer VN (1983) Hydrogenase and ribulose-1, 5-bisphosphate carboxylase activities of *Alcaligenes eutrophus* ATCC 17706 associated with an indigenous plasmid. *Can J Microbiol* 29:767–774
- Bernhard M, Schwartz E, Rietdorf J, Friedrich B (1996) The *Alcaligenes eutrophus* membrane-bound hydrogenase gene locus encodes functions involved in maturation and electron transport coupling. *J Bacteriol* 178:4522–4529
- Bernhard M, Benelli B, Hochkoeppler A, Zannoni D, Friedrich B (1997) Functional and structural role of the cytochrome *b* subunit of the membrane-bound hydrogenase complex of *Alcaligenes eutrophus* H16. *Eur J Biochem* 248:179–186
- Bernhard M, Friedrich B, Siddiqui RA (2000) *Ralstonia eutropha* TF93 is blocked in Tat-mediated protein export. *J Bacteriol* 182:581–588
- Bernhard M, Buhrke T, Bleijlevens B, De Lacey AL, Fernandez VM, Albracht SP, Friedrich B (2001) The H₂ sensor of *Ralstonia eutropha*. Biochemical characteristics, spectroscopic properties, and its interaction with a histidine protein kinase. *J Biol Chem* 276:15592–15597
- Bignell C, Thomas CM (2001) The bacterial ParA–ParB partitioning proteins. *J. Biotechnol* 91:1–34
- Bleijlevens B, Buhrke T, van der Linden E, Friedrich B, Albracht SP (2004) The auxiliary protein HypX provides oxygen tolerance to the soluble [NiFe]-hydrogenase of *ralstonia eutropha* H16 by way of a cyanide ligand to nickel. *J Biol Chem* 279:46686–46691
- Böck A, King PW, Blokesch M, Posewitz MC (2006) Maturation of hydrogenases. *Adv Microb Physiol* 51:1–71
- Bonen L, Vogel J (2001) The ins and outs of group II introns. *Trends Genet* 17:322–331
- Bowien B, Kusian B (2002) Genetics and control of CO₂ assimilation in the chemoautotroph *Ralstonia eutropha*. *Arch Microbiol* 178:85–93
- Buhrke T, Lenz O, Porthun A, Friedrich B (2004) The H₂-sensing complex of *Ralstonia eutropha*: interaction between a regulatory [NiFe] hydrogenase and a histidine protein kinase. *Mol Microbiol* 51:1677–1689
- Buhrke T, Löscher S, Lenz O, Schlotter E, Zebger I, Andersen LK, Hildebrandt P, Meyer-Klaucke W, Dau H, Friedrich B, Haumann M (2005) Reduction of unusual iron-sulfur clusters in the H₂-sensing regulatory Ni-Fe hydrogenase from *Ralstonia eutropha* H16. *J Biol Chem* 280:19488–19495
- Burgdorf T, Lenz O, Buhrke T, van der Linden E, Jones AK, Albracht SP, Friedrich B (2005a) [NiFe]-hydrogenases of *Ralstonia eutropha* H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation. *J Mol Microbiol Biotechnol* 10:181–96
- Burgdorf T, van der Linden E, Bernhard M, Yin QY, Back JW, Hartog AF, Muijsers AO, de Koster CG, Albracht SP, Friedrich B (2005b) The soluble NAD⁺-Reducing [NiFe]-hydrogenase from *Ralstonia eutropha* H16 consists of six subunits and can be specifically activated by NADPH. *J Bacteriol* 187:3122–3132
- Cangelosi GA, Wheelis ML (1984) Regulation by molecular oxygen and organic substrates of hydrogenase synthesis in *Alcaligenes eutrophus*. *J Bacteriol* 159:138–144
- Chandra TS, Friedrich CG (1986) Tn5-induced mutations affecting sulfur-oxidizing ability (Sox) of *Thiosphaera pantotropha*. *J Bacteriol* 166:446–452

- Conrad R, Seiler W (1979) Role of hydrogen bacteria during the decomposition of hydrogen by soil. *FEMS Microbiol Lett* 6:143–145
- Cramm R, Siddiqui RA, Friedrich B (1994) Primary sequence and evidence for a physiological function of the flavohemoprotein of *Alcaligenes eutrophus*. *J Biol Chem* 269:7349–7354
- Cramm R, Siddiqui RA, Friedrich B (1997) Two isofunctional nitric oxide reductases in *Alcaligenes eutrophus* H16. *J Bacteriol* 179:6769–6777
- Cramm R, Pohlmann A, Friedrich B (1999) Purification and characterization of the single-component nitric oxide reductase from *Ralstonia eutropha* H16. *FEBS Lett* 460:6–10
- Crutzen PJ, Gidel LT (1983) A two-dimensional photochemical model of the atmosphere. 2. the tropospheric budgets of the anthropogenic chlorocarbons CO, CH₄, CH₃Cl and the effect of various NO_x sources on tropospheric ozone. *J Geophys Res* 88:6641–6661
- Cunningham SD, Kapulnik Y, Phillips DA (1986) Distribution of hydrogen-metabolizing bacteria in alfalfa field soil. *Appl Environ Microbiol* 52:1091–1095
- Cypionka H, Meyer O (1983a) Carbon monoxide-insensitive respiratory chain of *Pseudomonas carboxydovorans*. *J Bacteriol* 156:1178–1187
- Cypionka H, Meyer O (1983b) The cytochrome composition of carboxydotrophic bacteria. *Arch Microbiol* 135:293–298
- Dernedde J, Eitinger T, Patenge N, Friedrich B (1996) *hyp* gene products in *Alcaligenes eutrophus* are part of a hydrogenase-maturation system. *Eur J Biochem* 235:351–358
- Dobbek H, Gremer L, Meyer O, Huber R (1999) Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selenylcysteine. *Proc Natl Acad Sci USA* 96:8884–8889
- Dobbek H, Gremer L, Meyer O, Huber R (2001) CO dehydrogenase. In: Messerschmidt A, Huber R, Poulos T, Wieghardt, K (eds) *Handbook of Metalloproteins*. Wiley, Chichester, pp 1136–1147
- Dobbek H, Gremer L, Kiefersauer R, Huber R, Meyer O (2002) Catalysis at a dinuclear [CuSMo(=O)OH] cluster in a CO dehydrogenase resolved at 1.1-Å resolution. *Proc Natl Acad Sci USA* 99:15971–15976
- Dong Z, Wu L, Kettlewell B, Caldwell CD, Layzell DB (2003) Hydrogen fertilization of soils – is this a benefit of legumes in rotation? *Plant Cell Environ* 26:1875–1879
- Eberz G, Friedrich B (1991) Three *trans*-acting regulatory functions control hydrogenase synthesis in *Alcaligenes eutrophus*. *J Bacteriol* 173:1845–1854
- Eitinger T, Suhr J, Moore L, Smith JA (2005) Secondary transporters for nickel and cobalt ions: theme and variations. *Biometals* 18:399–405
- Erb RW, Timmis KN, Pieper DH (1998) Characterization of a gene cluster from *Ralstonia eutropha* JMP134 encoding metabolism of 4-methylmuconolactone. *Gene* 206:53–62
- Forzi L, Sawers RG (2007) Maturation of [NiFe]-hydrogenases in *Escherichia coli*. *Biometals* 20:565–578
- Friedrich B, Buhrke T, Burgdorf T, Lenz O (2005) A hydrogen-sensing multiprotein complex controls aerobic hydrogen metabolism in *Ralstonia eutropha*. *Biochem Soc Trans* 33:97–101
- Friedrich B, Friedrich CG (1990) Hydrogenases in lithoautotrophic bacteria. In: Codd GA, Dijkhuizen L, Tabita FR (eds) *Autotrophic microbiology and one-carbon metabolism*. Kluwer, Dordrecht, pp 55–92
- Friedrich B, Hogrefe C, Schlegel HG (1981) Naturally occurring genetic transfer of hydrogen-oxidizing ability between strains of *Alcaligenes eutrophus*. *J Bacteriol* 147:198–205
- Friedrich B, Friedrich CG, Meyer M, Schlegel HG (1984) Expression of hydrogenase in *Alcaligenes* spp. is altered by interspecific plasmid exchange. *J Bacteriol* 158:331–333
- Friedrich B, Kortlüke C, Hogrefe C, Eberz G, Silber B, Warrelmann J (1986) Genetics of hydrogenase from aerobic lithoautotrophic bacteria. *Biochimie* 68:133–145
- Fuhrmann S, Ferner M, Jeffke T, Henne A, Gottschalk G, Meyer O (2003) Complete nucleotide sequence of the circular megaplasmid pHCG3 of *Oligotropha carboxidovorans*: function in the chemolithoautotrophic utilization of CO, H₂ and CO₂. *Gene* 322:67–75
- Galperin MY, Nikolskaya AN, Koonin EV (2001) Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett* 203:11–21

- Gerstenberg C, Friedrich B, Schlegel HG (1982) Physical evidence for plasmids in autotrophic, especially hydrogen-oxidizing bacteria. *Arch Microbiol* 133:90–96
- Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang WS, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Verméglio A, Médigue C, Sadowsky M (2007) Legume symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* 316:1307–1312
- Gnida M, Ferner R, Gremer L, Meyer O, Meyer-Klaucke W (2003) A novel binuclear [CuSMo] cluster at the active site of carbon monoxide dehydrogenase: characterization by X-ray absorption spectroscopy. *Biochemistry* 42:222–230
- Grzeszik C, Jeffke T, Schäferjohann J, Kusian B, Bowien B (2000) Phosphoenolpyruvate is a signal metabolite in transcriptional control of the *cbb* CO₂ fixation operons in *Ralstonia eutropha*. *J Mol Microbiol Biotechnol* 2:311–320
- Hanus FJ, Maier RJ, Evans HJ (1979) Autotrophic growth of H₂-uptake-positive strains of *Rhizobium japonicum* in an atmosphere supplied with hydrogen gas. *Proc Natl Acad Sci USA* 76:1788–1792
- Hogrefe C, Friedrich B (1984) Isolation and characterization of megaplastid DNA from lithoautotrophic bacteria. *Plasmid* 12:161–169
- Hogrefe C, Römermann D, Friedrich B (1984) *Alcaligenes eutrophus* hydrogenase genes (Hox). *J Bacteriol* 158:43–48
- Kalkus J, Reh M, Schlegel HG (1990) Hydrogen autotrophy of *Nocardia opaca* strains is encoded by linear megaplastids. *J Gen Microbiol* 136:1145–1151
- Kelly DP, Wood AP (2006) The chemolithotrophic bacteria. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds) *The prokaryotes*, 3rd edn. Springer, New York, pp 441–456
- Khalil MAK, Pinto JP, Shearer MJ (1999) Atmospheric carbon monoxide *Chemosphere* 1:xi–xiii
- King GM (1999) Characteristics and significance of atmospheric carbon monoxide consumption by soils. *Chemosphere* 1:53–63
- King GM, Weber CF (2007) Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat Rev Microbiol* 5:107–118
- Kleihues L, Lenz O, Bernhard M, Buhrke T, Friedrich B (2000) The H₂ sensor of *Ralstonia eutropha* is a member of the subclass of regulatory [NiFe] hydrogenases. *J Bacteriol* 182:2716–2724
- Klüber HD, Lechner S, Conrad R (1995) Characterization of populations of aerobic hydrogen-oxidizing soil bacteria. *FEMS Microbiol Ecol* 16:167–175
- Kortlüke C, Horstmann K, Schwartz E, Rohde M, Binsack R, Friedrich B (1992) A gene complex coding for the membrane-bound hydrogenase of *Alcaligenes eutrophus* H16. *J Bacteriol* 174:6277–6289
- Kraft I, Bock E (1984) Plasmids in *Nitrobacter*. *Arch Microbiol* 140:79–82
- Kraut M, Meyer O (1988) Plasmids in carboxidotrophic bacteria: physical and restriction analysis. *Arch Microbiol* 149:540–546
- Kusian B, Bednarski R, Husemann M, Bowien B (1995) Characterization of the duplicate ribulose-1,5-bisphosphate carboxylase genes and *cbb* promoters of *Alcaligenes eutrophus*. *J Bacteriol* 177:4442–4450
- La Favre JS, Focht DD (1983) Conservation in soil of H₂ liberated from N₂ fixation by Hup nodules. *Appl Environ Microbiol* 46:304–311
- Lechner S, Conrad R (1997) Detection in soil of aerobic hydrogen-oxidizing bacteria related to *Alcaligenes eutrophus* by PCR and hybridization assays targeting the gene of the membrane-bound (NiFe) hydrogenase. *FEMS Microbiol Ecol* 22:193–206
- Leimkühler S, Klipp W (1999) Role of XDHC in Molybdenum cofactor insertion into xanthine dehydrogenase of *Rhodobacter capsulatus*. *J Bacteriol* 181:2745–2751

- Lenz O, Friedrich B (1998) A novel multicomponent regulatory system mediates H₂ sensing in *Alcaligenes eutrophus*. *Proc Natl Acad Sci USA* 95:12474–12479
- Lenz O, Zebger I, Hamann J, Hildebrandt P, Friedrich B (2007) Carbamoylphosphate serves as the source of CN₂, but not of the intrinsic CO in the active site of the regulatory [NiFe]-hydrogenase from *Ralstonia eutropha*. *FEBS Lett* 581:3322–3326
- MacLellan SR, Smallbone LA, Sibley CD, Finan TM (2005) The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of α -proteobacterial plasmids. *Mol Microbiol* 55:611–23
- MacLellan SR, Zaheer R, Sartor AL, MacLean AM, Finan TM (2006) Identification of a megaplasmid centromere reveals genetic structural diversity within the *repABC* family of basic replicons. *Mol Microbiol* 59:1559–1575
- Madigan MT, Gest H (1979) Growth of the photosynthetic bacterium *Rhodospseudomonas capsulata* chemoautotrophically in darkness with H₂ as the energy source. *J Bacteriol* 137:524–530
- Maimaiti J, Zhang Y, Yang J, Cen YP, Layzell DB, Peoples M, Dong Z (2007) Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environ Microbiol* 9:435–444
- Martínez-Abarca F, García-Rodríguez FM, Toro N (2000) Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol Microbiol* 35:1405–1412
- Massanz C, Friedrich B (1999) Amino acid replacements at the H₂-activating site of the NAD-reducing hydrogenase from *Alcaligenes eutrophus*. *Biochemistry* 38:14330–14337
- Mergeay M, Monchy S, Janssen P, Van Houdt R, Leys N (2008) Megaplasmids in *Cupriavidus* genus and metal resistance. 10.1007/7171_2008_19
- Meyer O (1989) Aerobic, carbon monoxide-oxidizing bacteria. In: Schlegel HG, Bowien B (eds) *Autotrophic bacteria*. Science Tech Publishers, Madison, pp 331–350
- Meyer O, Frunzke K, Mörsdorf G (1993) Biochemistry of aerobic utilization of carbon monoxide. In: Murrel JC, Kelley DP (eds) *Microbial growth on C₁ compounds*. Intercept, Andover, pp 433–459
- Meyer O, Gremer L, Ferner R, Ferner M, Dobbek H, Gnida M, Meyer-Klaucke W, Huber R (2000) The role of Se, Mo and Fe in the structure and function of carbon monoxide dehydrogenase. *Biol Chem* 381:865–876
- Monchy S, Benotmane MA, Janssen P, Vallaeyts T, Taghavi S, van der Lelie D, Mergeay M (2007) Plasmids pMOL28 and pMOL30 of *Cupriavidus metallidurans* are specialized in the maximal viable response to heavy metals. *J Bacteriol* 189:7417–7425
- Montet Y, Amara P, Volbeda A, Vernede X, Hatchikian EC, Field MJ, Frey M, Fontecilla-Camps JC (1997) Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics. *Nat Struct Biol* 4:523–526
- Nikolskaya AN, Galperin MY (2002) A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. *Nucleic Acids Res* 30:2453–2459
- Ohi K, Takada N, Komemushi S, Okazaki M, Miura Y (1979) A new species of hydrogen-utilizing bacterium. *J Gen Appl Microbiol* 25:53–58
- Pierik AJ, Schmelz M, Lenz O, Friedrich B, Albracht SP (1998) Characterization of the active site of a hydrogen sensor from *Alcaligenes eutrophus*. *FEBS Lett* 438:231–235
- Pohlmann A, Cramm R, Schmelz K, Friedrich B (2000) A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol Microbiol* 38:626–638
- Pohlmann A, Fricke WF, Reinecke F, Kusian B, Liesegang H, Cramm R, Eitinger T, Ewering C, Pötter M, Schwartz E, Strittmatter A, Voss I, Gottschalk G, Steinbüchel A, Friedrich B, Bowien B (2006) Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. *Nat Biotechnol* 24:1257–1262
- Rees E, Siddiqui RA, Köster F, Schneider B, Friedrich B (1997) Structural gene (*nirS*) for the cytochrome *cdI* nitrite reductase of *Alcaligenes eutrophus* H16. *Appl Environ Microbiol* 63:800–802

- Rodionov DA, Hebbeln P, Gelfand MS, Eitinger T (2006) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J Bacteriol* 188:317–327
- Rohde M, Mayer F, Meyer O (1984) Immunocytochemical localization of carbon monoxide oxidase in *Pseudomonas carboxydovorans*. The enzyme is attached to the inner aspect of the cytoplasmic membrane. *J Biol Chem* 259:14788–14792
- Rohde M, Mayer F, Jacobitz S, Meyer O (1985) Attachment of CO dehydrogenase to the cytoplasmic membrane is limiting the respiratory rate of *Pseudomonas carboxydovorans*. *FEMS Microbiol Lett* 28:141–144
- Romero D, Brom S (2004) The Symbiotic Plasmids of the Rhizobiaceae. In: Funnel BE, Phillips GJ (eds) Plasmid Biology. ASM Press, Washington DC, pp 271–290
- Santiago B, Schübel U, Egelseer C, Meyer O (1999) Sequence analysis, characterization and CO-specific transcription of the *cox* gene cluster on the megaplastid pHCG3 of *Oligotropha carboxydovorans*. *Gene* 236:115–124
- Schlegel HG, Bowien B (eds) (1989) Autotrophic bacteria. Science Tech Publishers, Madison
- Schneider B, Nies A, Friedrich B (1988) Transfer and expression of lithoautotrophy and denitrification in a host lacking these metabolic activities. *Appl Environ Microbiol* 54:3173–3176
- Schneider K, Schlegel HG (1976) Purification and properties of soluble hydrogenase from *Alcaligenes eutrophus* H16. *Biochim Biophys Acta* 452:66–80
- Schneider K, Cammack R, Schlegel HG, Hall DO (1979) The iron-sulphur centres of soluble hydrogenase from *Alcaligenes eutrophus*. *Biochim Biophys Acta* 578:445–461
- Schneider K, Pinkwart M, Jochim K (1983) Purification of hydrogenases by affinity chromatography on Procion Red-agarose. *Biochem J* 213:391–398
- Schubert T, Lenz O, Krause E, Volkmer R, Friedrich B (2007) Chaperones specific for the membrane-bound [NiFe]-hydrogenase interact with the Tat signal peptide of the small subunit precursor in *Ralstonia eutropha* H16. *Mol Microbiol* 66:453–467
- Schwartz E, Friedrich B (2006) The H₂-metabolizing prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds) The prokaryotes, 3rd edn. Springer, New York, pp 496–563
- Schwartz E, Gerischer U, Friedrich B (1998) Transcriptional regulation of *Alcaligenes eutrophus* hydrogenase genes. *J Bacteriol* 180:3197–3204
- Schwartz E, Buhrke T, Gerischer U, Friedrich B (1999) Positive transcriptional feedback controls hydrogenase expression in *Alcaligenes eutrophus* H16. *J Bacteriol* 181:5684–5692
- Schwartz E, Henne A, Cramm R, Eitinger T, Friedrich B, Gottschalk G (2003) Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplastid encoding key enzymes of H₂-based lithoautotrophy and anaerobiosis. *J Mol Biol* 332:369–383
- Sensfuss C, Reh M, Schlegel HG (1986) No correlation exists between the conjugative transfer of the autotrophic character and that of plasmids in *Nocardia opaca* strains. *J Gen Microbiol* 132:997–1007
- Siddiqui RA, Warnecke-Eberz U, Hengsberger A, Schneider B, Kostka S, Friedrich B (1993) Structure and function of a periplasmic nitrate reductase in *Alcaligenes eutrophus* H16. *J Bacteriol* 175:5867–5876
- Siedow A, Cramm R, Siddiqui RA, Friedrich B (1999) A megaplastid-borne anaerobic ribonucleotide reductase in *Alcaligenes eutrophus* H16. *J Bacteriol* 181:4919–4928
- Siefert E, Pfennig N (1979) Chemoautotrophic growth of rhodospseudomonas species with hydrogen and chemotrophic utilization of methanol and formate. *Arch Microbiol* 122:177–182
- Starkenburger SR, Larimer FW, Stein LY, Klotz MG, Chain PS, Sayavedra-Soto LA, Poret-Peterson AT, Gentry ME, Arp DJ, Ward B, Bottomley PJ (2008) The complete genome sequence of *Nitrobacter hamburgensis* X14 and a comparative genomic analysis of species within the Genus *Nitrobacter*. *Appl Environ Microbiol* 74:2852–2863
- Swingley WD, Sadekar S, Mastrian SD, Matthies HJ, Hao J, Ramos H, Acharya CR, Conrad AL, Taylor HL, Dejesa LC, Shah MK, O'hallachain ME, Lince MT, Blankenship RE, Beatty JT, Touchman JW (2007) The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. *J Bacteriol* 189:683–690

- Taylor BL, Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* 63:479–506
- Thiemermann S, Dervede J, Bernhard M, Schroeder W, Massanz C, Friedrich B (1996) Carboxyl-terminal processing of the cytoplasmic NAD-reducing hydrogenase of *Alcaligenes eutrophus* requires the *hoxW* gene product. *J Bacteriol* 178:2368–2374
- Timotius K, Schlegel HG (1987) [Nickel-resistant bacteria isolated from sewage]. *Nachrichten der Akademie der Wissenschaften in Göttingen. II Mathematisch-Physikalische Klasse*. 3:1–9
- Tran-Betcke A, Warnecke U, Böcker C, Zaborosch C, Friedrich B (1990) Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16. *J Bacteriol* 172:2920–2929
- Vignais PM, Billoud B, Meyer J (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* 25:455–501
- Warrelmann J, Friedrich B (1989) Large plasmids in wild type strains of *Pseudomonas facilis*. *Endocytobiosis Cell Res* 6:213–217
- Wolf I, Buhrke T, Dervede J, Pohlmann A, Friedrich B (1998) Duplication of *hyp* genes involved in maturation of [NiFe] hydrogenases in *Alcaligenes eutrophus* H16. *Arch Microbiol* 170:451–459
- Zumft WG (2005) Biogenesis of the bacterial respiratory Cu_A , Cu-S enzyme nitrous oxide reductase. *J Mol Microbiol Biotechnol* 10:154–166
- Zumft WG, Dreusch A, Löchelt S, Cuypers H, Friedrich B, Schneider B (1992) Derived amino acid sequences of the *nosZ* gene (respiratory N_2O reductase) from *Alcaligenes eutrophus*, *Pseudomonas aeruginosa* and *Pseudomonas stutzeri* reveal potential copper-binding residues. Implications for the Cu_A site of N_2O reductase and cytochrome-*c* oxidase. *Eur J Biochem* 208:31–40

The Megaplasmid pAO1 of *Arthrobacter Nicotinovorans* and Nicotine Catabolism

Petra Ganas, Gabor L. Igloi, and Roderich Brandsch(✉)

Contents

1	Introduction.....	271
2	General Features of the Megaplasmid pAO1.....	272
3	Predicted Metabolic Capabilities Other than Nicotine Degradation.....	274
4	Nicotine Catabolism.....	275
4.1	From Nicotine to Nicotine Blue.....	275
4.2	A Gene Module for γ -N-Methylaminobutyrate Catabolism.....	277
4.3	A Gene Cluster for Molybdenum Cofactor Biosynthesis and Molybdenum Holoenzyme Assembly.....	278
4.4	Nicotine Uptake and Extrusion of Nicotine Metabolites.....	279
4.5	Pathway Regulators.....	280
	References.....	281

Abstract The 165-kb megaplasmid pAO1 of *Arthrobacter nicotinovorans* enables its host to grow on nicotine as carbon, nitrogen and energy source. The plasmid has a modular structure with DNA segments flanked by mobile genetic elements. The presence of gene modules specific for L-nicotine, D-nicotine and γ -N-methylaminobutyrate catabolism, for the biosyntheses of the molybdenum cofactor required by the molybdenum enzymes of the pathway, for the two subunits of a small multi-drug resistance pump and for gene products for the protection from oxidative stress generated by oxidation of nicotine blue, underscores the notion that this plasmid evolved specifically to fulfil the task of nicotine degradation.

1 Introduction

L-Nicotine, the main alkaloid synthesized by the tobacco plant, and the small amounts of D-nicotine and D-nornicotine, can constitute 3% of the dry weight of leaves (Armstrong et al. 1999). It is produced by the plant in defence against predators,

P. Ganas, G.L. Igloi, and R. Brandsch
Institut für Biochemie und Molekularbiologie Albert-Ludwigs University,
Hermann-Herder-Str. 7, 79104, Freiburg, Germany
e-mail: roderich.brandsch@biochemie.uni-freiburg.de

is highly toxic for mammals and has been used as an insecticide. The availability of nicotine in natural habitats spurred the evolution of pathways for its degradation, providing for the recycling of carbon and nitrogen bound in the molecule. This is a typical task of the microbial soil community with its extraordinary adaptive ability for the degradation of natural, and recently, man made organic compounds. Catabolic pathways of organic compounds are often encoded by plasmids (Igloi and Brandsch 2003). In *Arthrobacter nicotinovorans* (formerly known as *A. oxydans* and reclassified as *A. nicotinovorans*, Kodama et al. 1992) the ability to degrade nicotine is linked to the presence of the megaplasmid pAO1.

2 General Features of the Megaplasmid pAO1

The 165,139-nucleotide pAO1 genome (Igloi and Brandsch 2003) has an overall G + C content of 61%, which is slightly less than that of the genome of *A. nicotinovorans* (62%, Kodama et al. 1992) but within the range of the genus *Arthrobacter* (Jones and Keddie 1992). The plasmid shows a modular structure, assembled from DNA segments flanked by transposons similar to Tn554 and Tn552 of *Staphylococcus aureus* and by IS1473 and IS1473-related elements (Igloi and Brandsch 2003) (Fig. 1). They have counterparts on the recently released genome sequences of *Arthrobacter aurescens* TC1 (Mongodin et al. 2006, Gen Bank Accession Number CP000509) and *Arthrobacter* sp strain FB24 (http://genome.jgi-psf.org/draft_microbs/art_f/art_f.download ftp.html), and illustrate the dynamic nature of plasmid genomes.

The DNA module flanked by Tn554 and IS1473 (Menéndez et al. 1997) contains a large number of genes involved in nicotine catabolism (*nic*-genes) (Igloi and

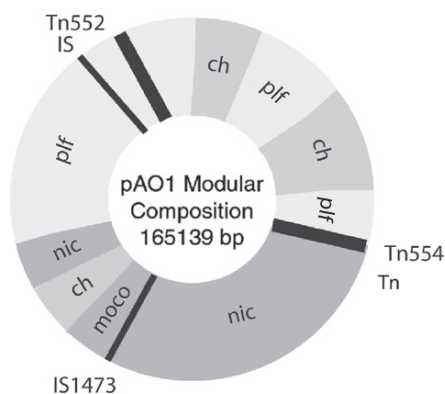


Fig. 1 Modular structure of the megaplasmid pAO1 of *Arthrobacter nicotinovorans*. The genome of the plasmid is composed of DNA segments flanked by putative transposases and IS elements. Regions carrying postulated plasmid and metabolic functions are indicated: *ch* carbohydrate catabolism, *moco* molybdenum cofactor biosynthesis, *nic* nicotine catabolism, *p/f*, plasmid functions

Brandsch 2003). Its G + C content of 56.7% sets it apart from the rest of the pAO1 DNA. Other modules contain putative genes of carbohydrate catabolism and genes related to plasmid functions of replication, partitioning, maintenance and conjugation (Fig. 1). The lower G + C content of the genes involved in nicotine degradation suggest their acquisition by lateral gene transfer from an unknown donor to an ancestral plasmid carrying genes for carbohydrate utilization. *IS1473* and its similar insertion sequence (IS) element on pAO1 represent direct repeats and may define the ends of a large catabolic transposon (Igloi and Brandsch 2003).

A schematic representation of pAO1 open reading frames (ORFs) not involved in carbohydrate- (*ch*), γ -*N*-methylaminobutyrate (*mgaba*), nicotine (*nic*) catabolism or molybdenum cofactor (*moco*) biosynthesis is shown in Fig. 2. Some of these ORFs must encode proteins that allow pAO1 to replicate its DNA, to distribute it at each cell division to the daughter cells, to stably maintain it within the bacterial cell and, since pAO1 was shown to be conjugative (Igloi and Brandsch 2003), to transfer it to other bacteria. These complex biological processes require the function of many of the uncharacterized putative gene products of pAO1.

When a search of the database was performed with the pAO1 ORFs assumed to be involved in the various aspects of plasmid biology (Fig. 2), similarities to ORFs present in the genomes of *A. aurescens* plasmids TC-1, TC-2 and *Arthrobacter sp* strain FB24 plasmids FB24-1, FB24-2 and FB24-3 became evident. There exists a conserved co-linearity of putative genes involved in plasmid functions among the *Arthrobacter* plasmids (not shown), an indication that these plasmids are related

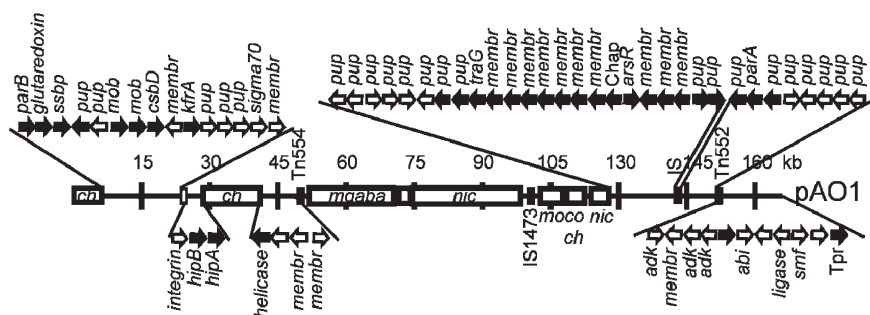


Fig. 2 Putative pAO1 ORFs involved in the functions replication (single strand binding protein *ssbp*, helicase, ligase), partitioning (*parB*, *parA*, *kfrA*), maintenance (high plasmid persistence toxin antitoxin system *hipA/hipB*), conjugation (mobilization *mob*, type IV secretion system *trag*, amidohydrolase/peptidase CHAP-domain protein), natural transformation competence (*smf*), stress response (general phosphate starvation response *csbD*, heavy metal resistance *arsR*), abortive infection phage resistance (*abi*), cell-cell contact (integrin similar FG-GAP repeat protein, tetratricopeptide repeats protein TPR), cell signaling (adenylate kinase *adk*). Many of the ORFs of predicted membrane proteins (*membr*) may participate in conjugation. *Black arrows* indicate ORFs with equivalents on the *A. aurescens* plasmids TC-1, TC-2, and *Arthrobacter sp*. FB24 plasmids FB24-1, FB24-2 and FB24-3; *gray arrows* indicate orfs with equivalents on the chromosome of *A. aureus* TC and *Arthrobacter sp*. FB24; *white arrows* indicate ORFs with no similarities in the database. *Pup*, orfs of putative uncharacterized proteins

and may have a common ancestor (Mongodin et al. 2006). The many ORFs of membrane proteins without function and putative uncharacterized proteins demonstrate that fundamental aspects of *Arthrobacter* plasmid biology are unknown.

3 Predicted Metabolic Capabilities Other than Nicotine Degradation

Besides nicotine catabolism, pAO1 appears to confer to its host the ability to use carbohydrates. Three operons, each controlled by a transcriptional regulator transcribed in opposite direction to the structural genes, and a putative cellulase (*cellul*) appear to serve this function (Fig. 3).

One predicted operon encodes a potential protein similar to lipase G (*lipG*), a putative sucrose hydrolase (*sucrase*), a second putative lipase (*glh*), a putative sugar major facilitator superfamily (MFS) permease and a putative LacI family transcriptional repressor (*lacIR*). Their biochemical function may consist in the catabolism of glycolipids. The putative 1,4- β -D-glucosidase or cellulase (*cellul*) may make available carbohydrates from plant cell wall material.

Also related to carbohydrate metabolism are ORFs and genes that may form one large operon for the oxidation of hexoses controlled by a transcriptional regulator (*pntR*) of the GntR family (Reizer et al. 1991). Many members of this family of transcriptional repressors are regulators of genes of carbohydrate metabolism.

The third hypothetical carbohydrate catabolic operon encodes a LacI repressor protein (*lacIR*), a MSF permease and a putative β -glucosidase (β -*gluc*). The two MFS permeases show only 18% identity in a 510 amino acid overlap, an indication that they transport different carbohydrates.

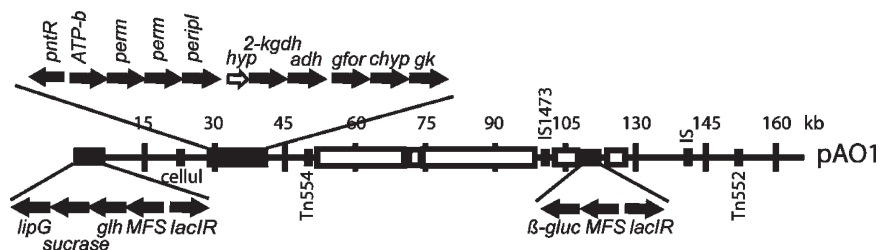


Fig. 3 pAO1 ORFs with putative functions in carbohydrate catabolism. ABC sugar transporter with ATP-binding component (ATP-b), components of the permease (*perm*) and periplasmic sugar binding protein (*peripl*); cellulase (*cellul*); hypothetical protein (*hyp*); putative 2-keto-gluconate dehydrogenase (*2-kgdh*); putative aldehyde dehydrogenase (*adh*); putative glucose/fructose oxidoreductase (*gfor*), conserved hypothetical protein (*chyp*); putative glycerate kinase (*gk*). See text for further details

4 Nicotine Catabolism

At least 40 ORF products of the 165 predicted pAO1 ORFs cooperate in nicotine catabolism. They may be subdivided into three groups: those involved in nicotine breakdown to nicotine blue, those involved in γ -*N*-methylaminobutyrate degradation and those involved in molybdenum cofactor biosynthesis (MoCo).

4.1 From Nicotine to Nicotine Blue

Nicotine catabolism as mediated by pAO1 is presented in Fig. 4. The initial reaction is catalyzed by nicotine dehydrogenase (nicotine:acceptor oxidoreductase, hydroxylating) (NDH) at C6 of the pyridine ring giving 6-hydroxy-nicotine. The heterotrimeric ($\alpha\beta\gamma$)₂ enzyme, which accepts both, L- and D-nicotine as substrates, is composed of a large subunit with a molybdenum cytosine dinucleotide cofactor, a mid-sized flavin adenine dinucleotide (FAD)-subunit and a small subunit carrying two iron–sulfur clusters (Andresen and Fetzner 2002). Further degradation of the stereoisomers 6-hydroxy-L-nicotine and 6-hydroxy-D-nicotine requires stereoselective oxidases, 6-hydroxy-L-nicotine oxidase (6HLNO) with noncovalently bound FAD and 6-hydroxy-D-nicotine oxidase with covalently bound FAD. The crystal structure of this enzyme (Koetter and Schulz 2005) confirmed the autoflavinilation mechanism of FAD attachment to a histidine residue of the protein (Brandsch and Bichler 1991). By spontaneous addition of water to the double bond of 6-hydroxy-*N*-methylmyosmine, *N*-methylaminopropyl-(6-hydroxypyridyl-3)-ketone (6-hydroxy-*pseudo*-oxynicotine) is produced, the substrate of ketone dehydrogenase (KDH). This enzyme hydroxylates the pyridine ring at C2 with the formation of 2,6-dihydroxy-*pseudo*-oxynicotine and consists of three subunits similar to the NDH subunits (Sachelaru et al. 2006). 2,6-dihydroxy-*pseudo*-oxynicotine may be spontaneously dehydrated in vitro to the metabolically inactive methylmyosmine derivative. In vivo 2,6-dihydroxy-*pseudo*-oxynicotine is cleaved by 2,6-dihydroxy-*pseudo*-oxynicotine hydrolase (PONH) (Sachelaru et al. 2005). The crystal structure of this α/β fold hydrolase was solved and a reaction mechanism proposed (Schleberger et al. 2007). 2,6-dihydroxypyridine is hydroxylated by the flavoenzyme 2,6-dihydroxypyridine hydroxylase (DHPH) (Baitsch et al. 2001) to trihydroxypyridine, which in the presence of oxygen rapidly forms a blue pigment known as nicotine blue (4,5,4',5'-tetrahydroxy-3,3'-diazadiphenquinone-(2,2')) (Knackmuss and Beckmann 1973). Alternatively, the hydroxylated pyridine ring may be cleaved, but this pathway has not yet been established.

Nicotine blue is apparently kept in its reduced form inside the bacterial cell by the activity of a NAD(P)H-nicotine blue oxidoreductase, a reaction that may protect *A. nicotinovorans* from oxidative stress (Mihasan et al. 2007). Additional enzyme reactions may take place at the level of trihydroxypyridine and nicotine blue by the products of a predicted Zn-containing polyketide cyclase (*pkc*), a predicted monooxygenase (*mox*), and a putative nitrilase/amidase (*nit*) (Fig. 5).

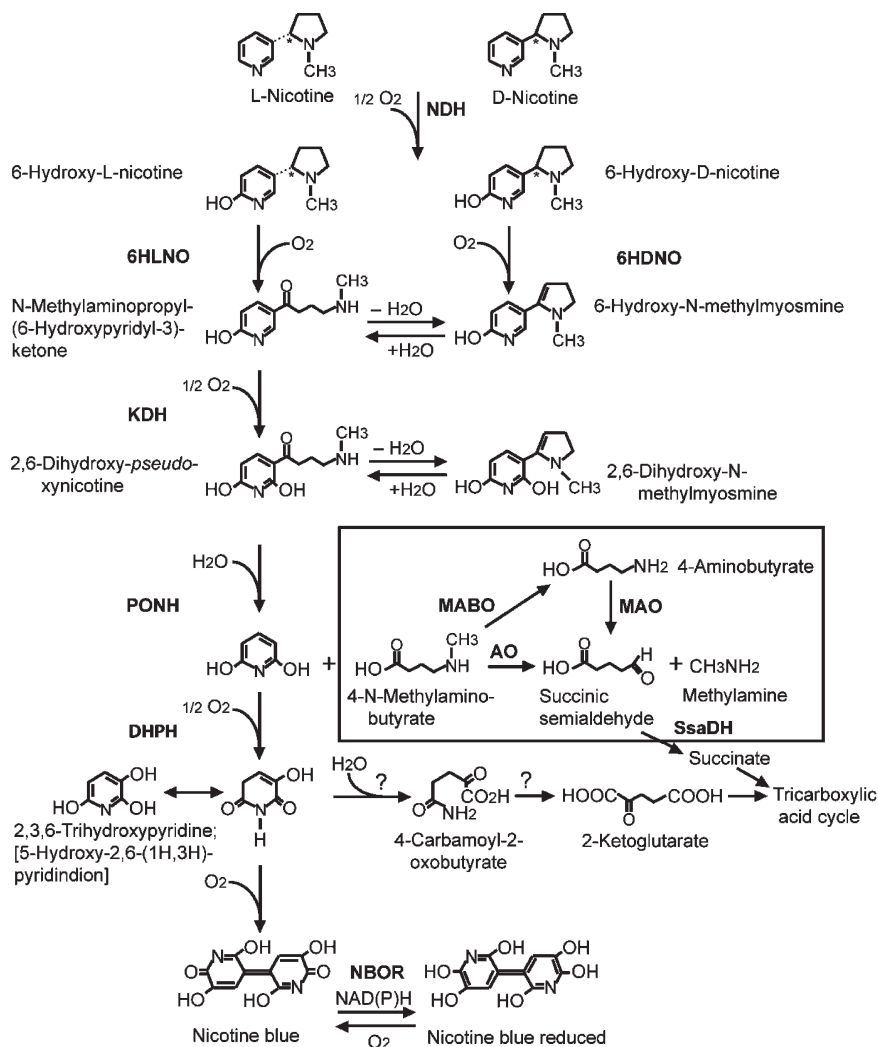


Fig. 4 The pathway of L- and D-nicotine degradation conferred by pAO1. See text for details

There is one large gene cluster for L-nicotine catabolism and a second small one for the D-hydroxynicotine specific oxidase (Fig. 5). The genes are expressed when nicotine is present in the growth medium and form a *nic*-regulon (Brandsch 2006) and may be regulated by the putative transcriptional activators TR1 and TR2. Remarkable is the split arrangement of the *kdh* subunit genes, with *kdhL* as the first gene of one operon and *kdhMS* representing a separate operon. The CoxG-like protein (*coxG*), a putative membrane bound oxidoreductase, is found associated with many molybdenum enzymes (Fuhrmann et al. 2003, Parschat et al. 2007), but

its function is as yet unknown. The gene of the D-hydroxynicotine specific oxidase, which is under the control of HdnR, forms a separate operon and is flanked by the ORFs of two predicted amino acid permeases (Fig. 5).

4.2 A Gene Module for γ -N-Methylaminobutyrate Catabolism

γ -N-Methylaminobutyrate catabolism is achieved by the gene products of two operons defined by transcriptional analysis and a solitary gene (Fig. 6). One operon is formed by *purU*-, *mabO*- and *fold*-like genes and the *nepA* and *nepB* subunit genes of a small multidrug resistance (SMR) pump. The *mabO* gene was shown to encode a sarcosine and dimethylglycine oxidase-like enzyme (Chiribau et al. 2004). The enzyme with covalently bound FAD oxidatively demethylates γ -N-methylaminobutyrate, producing γ -aminobutyrate (Chiribau et al. 2004). The methylenetetrahydrofolate generated in the reaction may be turned over by the bifunctional enzyme methylene-tetrahydrofolate dehydrogenase/cyclohydrolase (FolD) and by the formyl-tetrahydrofolate deformylase (PurU) (Chiribau et al. 2004).

A second operon consists of the gene for a monoamine oxidase (*mao*), the ORF of a hypothetical protein (*hyp*) and the gene for the NAD(P)H-nicotine blue oxidoreductase (*nbor*) (Fig. 6). Transcribed in opposite direction to the *mao* gene is the *ssd* gene of a NADP⁺-dependent succinic semialdehyde dehydrogenase (Chiribau et al. 2006).

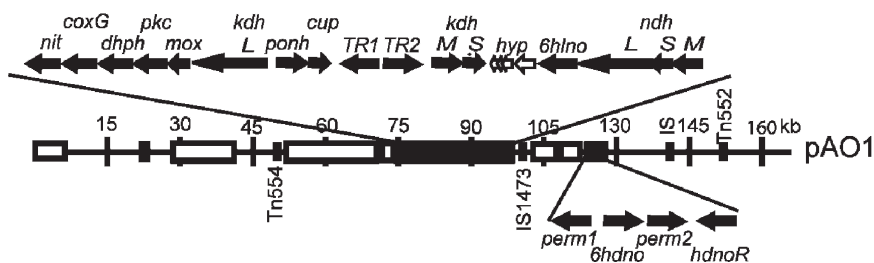


Fig. 5 pAO1 genes and ORFs involved in nicotine degradation. See text for details

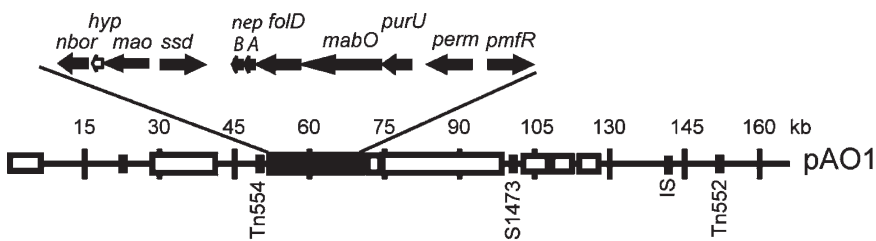


Fig. 6 pAO1 genes and ORFs involved in γ -N-methylaminobutyrate catabolism. See text for details

The unexpected finding that the MAO acts as an amine oxidase (AO) on the secondary amine of γ -*N*-methylaminobutyrate and hence has the same substrate as MABO led to the proposal of two pathways for the catabolism of γ -*N*-methylaminobutyrate (Chiribau et al. 2006) (Fig. 4, framed). γ -*N*-Methylaminobutyrate may be demethylated either by MABO, followed by deamination to succinic semialdehyde by the MAO or by a transaminase provided by the *A. nicotinovorans* chromosome. Alternatively the AO may turn over γ -*N*-methylaminobutyrate directly to methylamine and succinic semialdehyde. The pAO1-encoded succinic semialdehyde dehydrogenase (SsaDH) will lead to the formation of succinate, which enters the citric acid cycle. In vivo the amine oxidase activity of the enzyme predominates over its monoamine oxidase activity since methylamine appears as metabolic end-product in the growth medium (Chiribau et al. 2006; Ganas et al. 2007).

Each of these pathways contains an enzyme of new substrate specificity. The specificity of MABO for γ -*N*-methylaminobutyrate, which contains two C-units more than sarcosine, may have derived from a sarcosine oxidase. AO still shows low monoamine oxidase activity, but has higher specificity for the deamination of the secondary amine of γ -aminobutyrate. Apparently there was selective pressure during the establishment of nicotine catabolism for the evolution of new enzyme specificities starting from enzymes with sarcosine and polyamine oxidase activity.

4.3 A Gene Cluster for Molybdenum Cofactor Biosynthesis and Molybdenum Holoenzyme Assembly

The large subunits of NDH and KDH contain a molybdenum cofactor. The biosynthesis of this cofactor is highly conserved and a similar pathway for its synthesis is found from *Archaea* to humans (Schmitz et al. 2007). A remarkable feature of the pAO1 genome is a gene cluster for the nearly complete pathway of molybdenum cytosine dinucleotide (MCD) cofactor biosynthesis (Fig. 7). Present are the genes *moaA* and *moaC* for precursor Z synthesis from GTP, the genes *moaE* and *moaD*

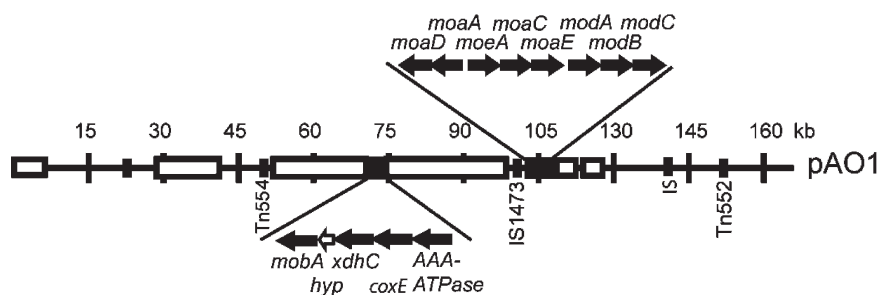


Fig. 7 pAO1 ORFs related to *moco* and MCD biosynthesis genes and insertion of the cofactor into the apoenzymes of molybdenum enzymes. See text for details

for the molybdopterin (MPT) synthase, the genes for a *modABC* transporter for molybdate uptake and a *moeA* homolog for the insertion of molybdenum into MPT and formation of MoCo. Absent on the plasmid are the genes for a MoeB protein, required for the ATP-dependent activation of MoaD, and a MogA homolog, presumably necessary for the formation of MPT-adenylate (Schmitz et al. 2007). Therefore, for the reactivation of the MPT synthase and for the activation of MPT, these adenylyltransferases have to be provided by chromosomal genes. It has been suggested that the pAO1 *mobA*-like gene codes for the MobA protein responsible for the synthesis of the MCD cofactor of NDH and KDH (Sachelaru et al. 2006). This gene is associated with a *xdhC*-like gene (see Fig. 7), which has been shown in *Rhodobacter capsulatus* to be involved in the insertion of the molybdenum cofactor into xanthine dehydrogenase (Neumann et al. 2006). Next to the *mobA*- and the *xdhC*-like genes are situated two ORFs. One of them (*coxE*) predicts a protein with a subtilisin protease domain, the other could encode an AAA-type ATPase, enzymes with multiple functions, including the assembly of multi-protein complexes. It is tempting to speculate that these genes represent a functional unit responsible for the assembly and quality control of the $(\alpha\beta\gamma)_2$ holoenzyme complexes of NDH and KDH. Indeed, it was shown that assembly of NDH and KDH required a functional *mobA* gene (Sachelaru et al. 2006). Ni-chelating chromatography recovered from cell extracts of *A. nicotinovorans* not only the His-tagged KDH-L subunit, but also the KDH-M and KDH-S subunits. In the absence of a functional *mobA* gene, only the His-tagged large subunit was recovered. This suggests that the middle-sized and the small subunits assemble into one sub-complex and the large subunit assembles separately with the MCD cofactor. Only then do the two sub-complexes assemble into the active holoenzyme.

One may speculate that the presence of molybdenum cofactor biosynthetic genes together with molybdenum enzyme genes on catabolic plasmids (Parschat et al. 2007) confers a selective advantage. Since the molybdenum enzyme genes are part of inducible operons, adequate levels of key biosynthesis enzymes are required to provide sufficient amounts of MoCo, which may not be reached by the constitutive expression of chromosomal genes.

4.4 Nicotine Uptake and Extrusion of Nicotine Metabolites

Bacteria have evolved specific uptake systems, including transporters for highly polar compounds and aromatics (Ganas et al. 2007). Uptake of nicotine in *A. nicotinovorans* is inducible and depends on the presence of pAO1 (Igloi and Brandsch 2003). An *A. nicotinovorans* strain devoid of pAO1 was unable to take up nicotine. There are three pAO1 ORFs with similarity to amino acid permeases (see Figs. 5 and 6). The corresponding genes are transcribed in the presence of nicotine, but the role of the predicted permeases, if any, in nicotine uptake by *A. nicotinovorans* is not known.

The NepAB small multidrug resistance (SMR) pump encoded by *nepAB* of the *purUmabOfolDnepAB* operon was shown to work as a metabolic valve responsible

for the proton gradient-dependent extrusion of the end-product of nicotine catabolism, methylamine, and of the metabolic intermediate γ -*N*-methylaminobutyrate, under conditions when it accumulated (Ganas et al. 2007). The position of the pump genes within the operon for γ -*N*-methylaminobutyrate and their co-regulation by PmfR is an example of the functionally efficient organization of the pAO1 genes with respect to nicotine catabolism.

4.5 Pathway Regulators

Early work demonstrated that the nicotine metabolic pathway was subject to catabolite repression (Gloger and Decker 1969). Glucose, but also glycerol, succinate and pyruvate, in the growth medium resulted in inhibition of the synthesis of nicotine catabolic enzymes, especially of nicotine and ketone dehydrogenases (Gloger and Decker, 1996).

HdnoR, a member of the TetR family of transcriptional repressors, was the first nicotine-responsive transcriptional regulator of genes belonging to the pAO1-encoded nicotine regulon that was characterized (Sandu et al. 2003). HdnoR binds with a K_d of 21 nM cooperatively to two operator sites covering the *6hdnO* promoter. Expression of the hypothetical permease1 gene (*perm1*) was not *hdnoR*-regulated and the *6hdnO* and the hypothetical permease2 (*perm2*) genes represent separate transcriptional units (see Fig. 5). 6-hydroxy-D-nicotine and 6-hydroxy-L-nicotine at micromolar concentrations prevented HdnoR binding to its operator sites with the D-enantiomer twice as potent as the L-enantiomer. The poor discrimination of the repressor between 6-hydroxy-D- and 6-hydroxy-L-nicotine may explain the surprising finding that a strictly stereo-specific 6-hydroxy-D-nicotine oxidase is induced by L-nicotine (Sandu et al. 2003).

A second transcriptional regulator that was characterized in greater detail is PmfR, a member of the NtrC/XylR family of transcriptional regulators (Chiribau et al. 2005). It controls expression of the *purUmabOfolDnepAB* and *maOnboR* operons (Fig. 6) integrating them into a functional unit by binding to the 5' upstream DNA region of the operons. Inactivation of its gene by disruption with a chloramphenicol resistance cassette revealed that PmfR acts as a transcriptional activator of the *purUmabOfolDnepAB* operon. The chemical nature of the effector molecule that modulates PmfR activity is not known. The genes for succinate semialdehyde and the putative permease appeared not to be under the control of PmfR (Chiribau et al. 2005).

PmfR may regulate its own transcription as indicated by the presence of a potential PmfR binding sequence in the 5' region of the *pmfR* gene. Upstream of the putative -35 region of the *pmfR* promoter is located a consensus DNA recognition sequence, TGTGA of the catabolite activator protein (CAP) (Chiribau and Brandsch, unpublished). This may indicate a higher order control of *pmfR* expression and thus of the genes it controls, as would be expected from the finding that nicotine catabolism is subject to catabolite repression.

The *nic*-genes seem to be organized in several operons. Transcriptional analysis and gene disruption studies suggested that the genes of *kdhL* to *mobA* form one large operon (Sachelaru et al. 2006). The genes of *ponh* and *orf116*, which are transcribed in opposite direction to the *kdhL-mobA* operon, seem to form another transcriptional unit as do the *kdhMS* genes. The same applies to *ndhMSL* and *6hlnO*. Responsible for regulation of the expression of these and possibly other genes of the nicotine regulon appear to be two putative transcriptional activators, TR1 and TR2 (see Fig. 5) of the LuxR type.

The genes of the MoCo biosynthesis cluster are organized in what appears to be two operons transcribed in opposite directions (Fig. 7). There is no transcriptional regulator predicted in this gene cluster. The nature of the putative regulatory proteins is as yet unknown but one may speculate that TR1 and TR2 may co-regulate expression of the MoCo biosynthesis genes.

References

- Andreesen JR, Fetzner S (2002) The molybdenum-containing hydroxylases of nicotine, isonicotinate, and nicotine. *Metal Ions Biol Sys* 39:405–430
- Armstrong DW, Wang X, Lee J-T, Liu Y-S (1999) Enantiomeric composition of normicotine, anatabine, and anabasine in tobacco. *Chirality* 11:82–84
- Baitsch D, Sandu C, Brandsch R, Igloi GL (2001) Gene cluster on pAO1 of *Arthrobacter nicotinovorans* involved in degradation of the plant alkaloid nicotine: cloning, purification, and characterization of 2,6-dihydroxypyridine 3-hydroxylase. *J Bacteriol* 183:5262–5267
- Brandsch R (2006) Microbiology and biochemistry of nicotine degradation. *Appl Microbiol Biotechnol* 69:493–498
- Brandsch R, Bichler V (1991) Autoflavinylolation of apo-6-hydroxy-D-nicotine oxidase. *J Biol Chem* 266:19056–19062
- Chiribau C-B, Sandu C, Fraaije M, Schiltz E, Brandsch R (2004) A novel γ -N-methylaminobutyrate demethylating oxidase involved in catabolism of the tobacco alkaloid nicotine by *Arthrobacter nicotinovorans* pAO1. *Eur J Biochem* 271:4677–4684
- Chiribau C-B, Sandu C, Igloi GL, Brandsch R (2005) Characterization of PmfR, the transcriptional activator of the pAO1-borne *purU-mabO-fold* operon of *Arthrobacter nicotinovorans*. *J Bacteriol* 187:3062–3070
- Chiribau C-B, Mihasan M, Ganas P, Igloi GL, Artenie V, Brandsch R (2006) Final steps in the catabolism of nicotine. Deamination versus demethylation of γ -N-methylaminobutyrate. *FEBS J* 273:1528–1536
- Fuhrmann S, Ferner M, Jeffke T, Henne A, Gottschalk G, Meyer O (2003) Complete nucleotide sequence of the circular megaplasmid pHCG3 of *Oligotropha carboxidovorans*: function in the chemolithoautotrophic utilization of CO, H₂ and CO₂. *Gene* 322:67–75
- Ganas P, Mihasan M, Igloi GL, Brandsch R (2007) A two-component small multidrug resistance pump functions as a metabolic valve during nicotine catabolism by *Arthrobacter nicotinovorans*. *Microbiol* 153:1546–1555
- Gloger M, Decker K (1969) Zum Mechanismus der Induktion nicotinabbauender Enzyme in *Arthrobacter oxydans*. *Z Naturforsch* 24b:1016–1025
- Igloi GL, Brandsch R (2003) Sequence of the 165-kilobase catabolic plasmid pAO1 from *Arthrobacter nicotinovorans* and identification of a pAO1-dependent nicotine uptake system. *J Bacteriol* 185:1976–1989
- Jones D, Keddie RM (1992) The genus *Arthrobacter*. In: Balows A, Truper AHG, Dworkin M, Harder W, Schleifer KH (ed) *The Prokaryotes*, vol. 2. Springer, Heidelberg, pp 1283–1299

- Kodama Y, Yamamoto H, Amano N, Amchi T (1992) Reclassification of two species of *Arthrobacter oxydans* and proposal of *Arthrobacter nicotinovorans* sp nov. *Int J Syst Bacteriol* 42:234–239
- Knackmuss H-J, Beckmann W (1973) The structure of nicotine blue from *Arthrobacter oxydans*. *Arch Mikrobiol* 90:167–169
- Koetter JWA, Schulz G (2005) Crystal structure of 6-hydroxy-D-nicotine oxidase from *Arthrobacter nicotinovorans*. *J Mol Biol* 352:418–428
- Menéndez C, Igloi GL, Brandsch R (1997) IS 1473, a putative insertion sequence identified in the plasmid pAO1 from *Arthrobacter nicotinovorans*: isolation, characterization and distribution among *Arthrobacter* species. *Plasmid* 37:35–41
- Mihasan M, Chiribau C-B, Friedrich T, Arteni V, Brandsch R (2007) An NAD(P)H-nicotine blue oxidoreductase is part of the nicotine regulon and may protect *Arthrobacter nicotinovorans* from oxidative stress during nicotine catabolism. *Appl Environ Microbiol* 73:2479–2485
- Mongodin EF, Shapir N, Daugherty SC, DeBoy RT, Emerson JB, Shvartzbeyn A, Radune D, Vamathevan J, Riggs F, Grinberg V, Khouri H, Wackett LP, Nelson KE, Sadowsky MJ (2006) Secrets of soil survival revealed by the genome sequence of *Arthrobacter aureescens* TC1. *PLoS Genet* 2:2094–2106
- Neumann M, Schulte M, Jünemann N, Stöcklein W, Leimkühler S (2006) *Rhodobacter capsulatus* XdhC is involved in molybdenum cofactor binding and insertion into xanthine dehydrogenase. *J Biol Chem* 281:15701–15708
- Parschat K, Overhage J, Strittmatter AW, Henne A, Gottschalk G, Fetzner S (2007) Complete nucleotide sequence of the 113-kilobase linear catabolic plasmid pAL1 of *Arthrobacter nitroguajacolicus* Rü61a and transcriptional analysis of genes involved in quinaldine degradation. *J Bacteriol* 189:3855–3867
- Reizer A, Deutscher J, Saier MH, Reizer J (1991) Analysis of the gluconate (*gn*) operon of *Bacillus subtilis*. *Mol Microbiol* 5:1081–1089
- Sachelaru P, Schiltz E, Igloi GL, Brandsch R (2005) An α/β -fold C-C bond hydrolase is involved in a central step of nicotine catabolism by *Arthrobacter nicotinovorans*. *J Bacteriol* 187:8516–8519
- Sachelaru P, Schiltz E, Brandsch R (2006) A functional *mobA* gene for molybdopterin cytosine dinucleotide cofactor biosynthesis is required for activity and holoenzyme assembly of the heterotrimeric nicotine dehydrogenase of *Arthrobacter nicotinovorans*. *Appl Environ Microbiol* 72:5126–5131
- Sandu C, Chiribau C-B, Brandsch R (2003) Characterization of HdnOR, the transcriptional repressor of the 6-hydroxy-D-nicotine oxidase gene of *Arthrobacter nicotinovorans* pAO1, and its DNA-binding activity in response to L- and D- nicotine derivatives. *J Biol Chem* 278:51307–51315
- Schleberger C, Sachelaru P, Brandsch R, Schulz GE (2007) Structure and action of a C–C bond cleaving α/β -hydrolase involved in nicotine degradation. *J Mol Biol* 367:409–418
- Schmitz J, Wuebbens MM, Rajagopalan KV, Leimkühler S (2007) Role of the C-terminal Gly-Gly motif of *Escherichia coli* MoaD, a molybdenum cofactor biosynthesis protein with a ubiquitin fold. *Biochem* 46:909–916

The pMUM Megaplasmid of *Mycobacterium ulcerans* and Closely Related Mycobacteria: A Blueprint for the Synthesis of Mycolactones

Sacha J. Pidot, Nicholas J. Tobias, and Tim Stinear(✉)

Contents

1	Introduction.....	284
2	<i>M. ulcerans</i> and Mycolactone.....	284
3	Mycolactone Synthesis by Polyketide Synthases Encoded by Genes on pMUM Plasmids.....	285
4	pMUM Plasmids in Mycobacteria.....	288
5	PKS Locus Variability between <i>M. ulcerans</i> Strains.....	289
6	Non-PKS Elements of pMUM-like Plasmids.....	290
7	Concluding Remarks.....	292
	References.....	293

Abstract *Mycobacterium ulcerans* is the third most common mycobacterial infection of humans and causes a destructive disease of subcutaneous tissue known as Buruli ulcer. A cytotoxic lipid known as mycolactone mediates the characteristic necrosis seen in Buruli ulcers. A family of highly related mycolactone structural variants have been discovered, all of which are produced by large polyketide synthases (PKSs), encoded by genes on large plasmids harbored by *M. ulcerans* strains. The prototype mycolactone plasmid is pMUM001, a 174-kb circular molecule from a West African epidemic strain of *M. ulcerans*. A striking feature of pMUM001 is that it has one-third of its DNA devoted to three very large genes that encode the mycolactone biosynthetic machinery. The plasmid (pMUM002) from *M. ulcerans* subsp. *liflandii* has recently been completely sequenced, and partial sequence has been obtained for a further pMUM plasmid, pMUM003, from *M. ulcerans* DL240490. Interstrain comparisons of pMUM sequences has revealed the highly mutable nature of the mycolactone PKS genes and given us greater insight into the origins of these large replicons.

S.J. Pidot, N.J. Tobias, and T. Stinear
Department of Microbiology, Monash University, Clayton, VIC 3800, Australia
e-mail: tim.stinear@med.monash.edu.au

1 Introduction

Mycobacterial plasmids were first identified in *Mycobacterium avium-intracellulare* (Crawford and Bates 1979) and then in a number of both slow and rapidly growing mycobacteria (Crawford and Bates 1984; Hull et al. 1984; Meissner and Falkinham 1984). However, it was not until 1988 that the first complete sequence of a mycobacterial plasmid, pAL5000 from *M. fortuitum*, was established (Rauzier et al. 1988). Since that time, extrachromosomal elements ranging in size from 13 to 330 kb have been identified in a range of mycobacterial species, including *M. avium*, *M. scrofulaceum*, *M. peregrinum*, *M. abscessus*, and *M. chelonae* (Jensen et al. 1989; Jucker and Falkinham 1990; Hellyer et al. 1991; Coleman and Spain 2003). While most of the plasmids identified to date have been circular, some species have linear plasmids, with similarity to the linear replicons of the *Actinomycetales* (Picardeau and Vincent 1997; Coleman and Spain 2003). Linear plasmids have been identified in *M. xenopi*, *M. celatum*, and *M. branderi*, but only pCLP from *M. celatum* has been investigated in any detail (Picardeau and Vincent 1997; Le Dantec et al. 2001). A common theme that emerges amongs studies of mycobacterial plasmids is their cryptic nature, with very few mycobacterial plasmids assigned functions or biological roles. This is primarily due to the inherent stability, or perhaps necessity, of these elements, and attempts to cure strains have been unsuccessful (Picardeau and Vincent 1997). Several studies have, however, suggested roles for a number of these elements in hydrocarbon metabolism, heavy-metal resistance, and DNA restriction and modification (Crawford et al. 1981; Meissner and Falkinham 1984; Erardi et al. 1987; Guerin and Jones 1988; Coleman and Spain 2003). One of the largest characterized mycobacterial plasmids is pMUM001 from *M. ulcerans* and it is the only mycobacterial plasmid identified to date to be implicated in pathogenesis (Stinear et al. 2004).

2 *M. ulcerans* and Mycolactone

M. ulcerans is the causative agent of the necrotizing skin disease known as Buruli ulcer. The organism was first identified by researchers investigating six cases of highly unusual skin ulcers in the Bairnsdale region of Victoria, Australia, in 1948 (MacCallum et al. 1948). Since the 1980s, the incidence of Buruli ulcer in western and sub-saharan Africa has steadily increased and in some regions it is now more common than either leprosy or tuberculosis (Debacker et al. 2004; Johnson et al. 2005). Establishing the mode of *M. ulcerans* transmission to humans remains elusive, although close proximity to swampy, stagnant water is a known risk factor (Debacker et al. 2006). The identification of a natural reservoir has also remained obscure, although a number of Australian marsupial species have been observed to harbor the bacterium and recently mosquitoes have been implicated in the transmission of the disease in South East Australia (Portaels et al. 2001; Quek et al. 2007; Johnson et al. 2007). Initial pathological investigations of human ulcers observed necrosis in the absence of bacteria, suggesting that the disease might be toxin-mediated

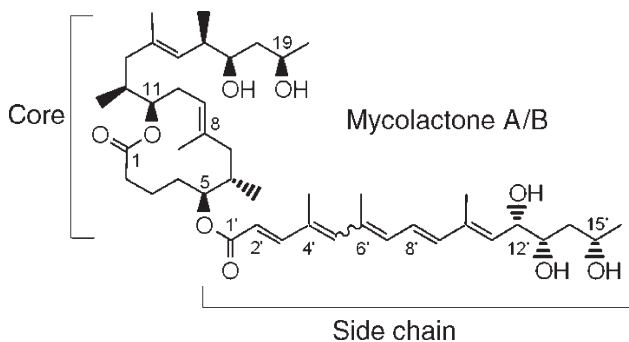


Fig. 1 The structure of mycolactone A/B. The core and side-chain components of the molecule are indicated

(Connor and Lunn 1965). A protein toxin was initially suspected, but several studies aimed at its identification were unsuccessful (Krieg et al. 1974; Read et al. 1974; Hockmeyer et al. 1978). It was not until 1999 that the toxic metabolite, a lipid named *mycolactone*, was discovered. A cytotoxic and immunosuppressive polyketide, mycolactone is the primary virulence factor of *M. ulcerans*, and purified mycolactone injected directly into an animal model is capable of replicating the pathology seen in Buruli ulcers (George et al. 1999). Mycolactone is made up of two polyketide products referred to as the *core* and *side chain* (George et al. 1999) (see Fig. 1). Detailed structural studies have shown that *M. ulcerans* strains from different geographic regions produce structurally distinct mycolactones and a number of minor mycolactone co-metabolites (Hong et al. 2003; Mve-Obiang et al. 2003). There are six known naturally occurring mycolactones (designated mycolactone A/B to F) and one unnatural mycolactone produced by biosynthetic engineering (mycolactone G) (Hong et al. 2007). Comparisons of the biological effects of these variants has shown that mycolactone A/B is the most potent immunosuppressive member of the family, while mycolactones C, E, F, and G showed lower inhibitory effects (Hong et al. 2007; Hong H et al. 2008). All mycolactones identified to date have an identical core and differ only in the structure of the side chain.

How *M. ulcerans* produced these bioactive molecules was revealed during the *M. ulcerans* genome sequencing project (Stinear et al. 2007), when it was discovered that *M. ulcerans* strain Agy99 has a 174-kb plasmid with three large genes encoding type I polyketide synthases (PKSs) (Stinear et al. 2004).

3 Mycolactone Synthesis by Polyketide Synthases Encoded by Genes on pMUM Plasmids

The mycolactone PKSs belong to a large family of type I bacterial PKSs. This broad class of enzymes produces a range of complex and medically important biomolecules, such as the antibiotic erythromycin, the antiparasitic compound avermectin,

and the immunosuppressants FK506 and rapamycin (Weissman and Leadlay 2005). The structure and enzymology of the PKSs are very similar to that of fatty acid synthases (FAS) (Hopwood and Sherman 1990). Modular type I PKSs are composed of a number of modules, which can be further subdivided into functional domains that form a molecular “assembly line” to build the polyketide chain in a processive fashion, where one module catalyses one round of polyketide chain elongation, by the addition of acetyl or propionyl carbon units. Thus, the overall number of PKS modules determines polyketide carbon chain length (Menzella et al. 2005). A minimal module contains ketosynthase (KS), acyl transferase (AT), and acyl carrier protein (ACP) domains, with the specificity of the AT domain for either acetyl-CoA or malonyl-CoA determining which extender unit will be incorporated in the polyketide chain (Jenke-Kodama et al. 2006). Modules may also contain dehydratase (DH), ketoreductase (KR), or enoylreductase (ER) enzymatic domains, constituting a “reductive loop”. As the polyketide chain is passed from module to module, a new 2-carbon or 3-carbon unit, specified by the AT domain, is incorporated and modified by any reductive domains present, until the end of the assembly line is reached, where a terminal thioesterase domain releases the product from the terminal ACP domain of the PKS (Moss et al. 2004).

On pMUM001, the mycolactone PKS locus contains three genes, named *mlsA1* (51 kb), *mlsA2* (7 kb), and *mlsB* (42 kb), which encode the PKSs that synthesize mycolactone (see Fig. 2). Confirmation that these genes were responsible for producing the mycolactone PKS came through investigation of mutants with transposon insertions that mapped to *mlsA1* and *mlsB* (Stinear et al. 2004). Furthermore, naturally occurring strains deficient in parts of the PKS locus are unable to produce mycolactone or cause Buruli ulcer (Stinear et al. 2005a).

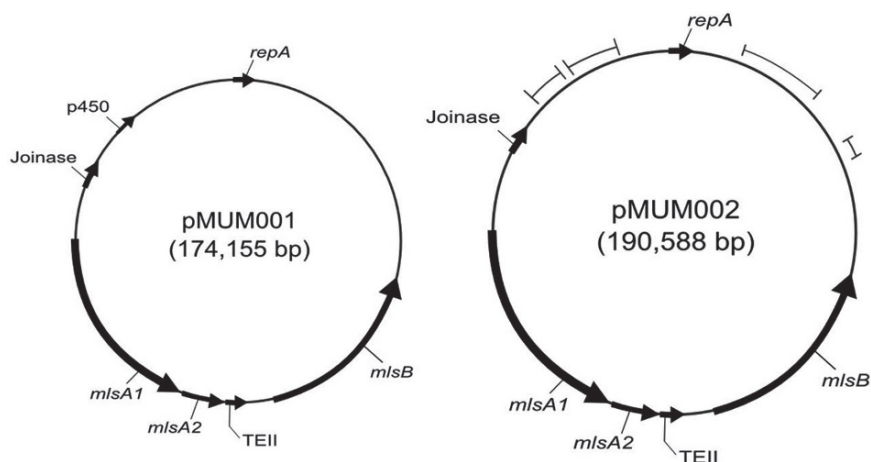


Fig. 2 Circular representation of pMUM001 and pMUM002. The genes involved in mycolactone biosynthesis on pMUM001 and pMUM002 are shown. Brackets in the map of pMUM002 represent regions of DNA >3.5 kb not present in pMUM001

DNA duplication appears to have led to complete identity in the region immediately upstream of both *mlsA1* and *mlsB*. Analysis of this region and investigation of *mls* gene transcription have revealed two transcriptional start points (TSPs) with a powerful sigma-70-like promoter linked to the TSP 491 bp upstream (T-491) of the start of *mlsA1* and *mlsB*. This promoter facilitated constitutive gene expression and exhibited a 20-fold higher level of transcription of the *mls* genes compared with *sigA* in *M. ulcerans* (N.T., S.P., T.S., unpublished observations). In *M. tuberculosis*, *sigA* is constitutively expressed and unaffected by stress, as it is involved in the regulation of housekeeping genes and is an essential element for survival of the bacterium (Ishihama 1988; Lonetto et al. 1992; Hu and Coates 1999). Interestingly, T-491 is located in the 97-bp intergenic region between MUP041 and MUP042 (each encoding putative transposases), suggesting that the *mls* promoter is linked to an insertion sequence. It appears that a single strong promoter is responsible for driving expression right through *mlsA1/mlsA2* and similarly through all of *mlsB* (N.T., S.P., T.S., unpublished observations). Furthermore, this promoter is not only active in *M. ulcerans*, but is constitutively expressed in other mycobacterial species including *M. marinum* and *M. smegmatis*, and is also highly active in *E. coli*.

There are several features of the mycolactone PKS that make them highly unusual. The first of these is the size of the PKS molecules themselves. As a whole, the size of the mycolactone PKS, at over 3.2 MDa, is among the largest known cellular enzyme complexes (Stinear et al. 2004). Recent reports of the formation of PKS domains existing as dimers and the existence of PKS-chaperone protein complexes suggest that the mycolactone PKSs may also be even larger than currently thought (Vivien et al. 2005; Tang et al. 2006). Another striking feature is the very high sequence identity among domains of the same function. PKS clusters typically share 40–70% sequence identity among like domains (Aparicio et al. 1996). The mycolactone PKSs, however, exhibit an unprecedented level of genetic identity (98.7–100% identity) (Stinear et al. 2004).

This high level of DNA homology might be expected to promote frequent genetic rearrangements and make the locus highly unstable. Indeed, this system does appear to be unstable as mycolactone-producing strains have been observed to become mycolactone-negative after laboratory passaging (George et al. 1999; Stinear et al. 2005a). However, despite this laboratory-observed instability, strains isolated from the environment and separated by both time and distance are still capable of producing mycolactone A/B (Mve-Obiang et al. 2003; Stinear et al. 2005a). Furthermore, strains lacking the plasmid (either naturally occurring or experimentally “cured”) have not been identified to date, suggesting that mycolactone plays an essential role in the lifecycle of the bacterium, and is under strong selective pressure. It is possible, therefore, that the cytotoxic effect seen on human cells is not the primary function of mycolactone and that its production enables the bacterium to occupy an environmental niche, wherever that may be.

While the PKSs are predominantly responsible for the production of mycolactone, there are a number of accessory proteins that are also important in mycolactone biosynthesis. The genes encoding these accessory elements include MUP038 (encoding a type II thioesterase), predicted to function as an “editing” enzyme, removing acyl

chains formed by aberrant decarboxylation from the PKS (Heathcote et al. 2001); MUP045 (encoding a putative beta-ketoacyl synthase), thought to be responsible for catalyzing the ester linkage between the core and the side chain; and MUP053 (encoding a cytochrome p450), important for adding a hydroxyl moiety to the final molecule, a modification that also appears to increase the biological activity of mycolactone (Hong et al. 2007). Mycolactone-producing strains that do not possess MUP053 have been identified (Mve-Obiang et al. 2005); however, the other accessory genes appear to be indispensable for mycolactone production, as they are found consistently in all strains. Indeed, in a recent study of the immunological effects of mycolactone, an *M. ulcerans* transposon mutant mapped to the MUP045 gene was used as a mycolactone negative control (Coutanceau et al. 2005). However, the precise role of these enzymes in mycolactone synthesis remains to be determined.

4 pMUM Plasmids in Mycobacteria

With the discovery of pMUM001 and the identification of structurally different mycolactones in a number of human clinical isolates came the realization that genetic differences in pMUM plasmids underlie the production of alternate mycolactones. A study of several clinical isolates of *M. ulcerans* found that they all carried pMUM plasmids of a range of sizes, with none being larger than pMUM001 (Stinear et al. 2005a). Most recently *M. ulcerans*-like mycobacteria, not associated with disease in humans, have also been shown to make mycolactones and contain pMUM plasmids. One of these mycobacterial isolates, designated here as *M. ulcerans* subsp. *liflandii*, was present in a colony of African tropical clawed frogs (*Xenopus tropicalis*) that had developed ulcerative lesions and coelomic effusion (Trott et al. 2004). Other studies have identified mycobacterial isolates from diseased fish from both Chesapeake bay in the United States and the Mediterranean Sea (Rhodes et al. 2005; Ucko and Colorni 2005; Ranger et al. 2006). These strains also produce novel mycolactones (mycolactone E for *M. ulcerans* subsp. *liflandii* and mycolactone F for the fish isolates) and have plasmids that are 20–35 kb larger than pMUM001 (Mve-Obiang et al. 2005; Ranger et al. 2006; Yip et al. 2007) (see Fig. 3). Detailed genetic analyses of all mycolactone-producing mycobacteria have shown that they have evolved from a common *M. marinum* progenitor by acquisition of a pMUM plasmid, followed by divergence into two distinct lineages, one that resembles *M. marinum* in genome size and structure and a second lineage with a reduced genome size as exemplified by *M. ulcerans* clinical isolates from cases of Buruli ulcer in Australia, Southeast Asia, and Africa (Yip et al. 2007).

The complete sequence of the plasmid from *M. ulcerans* subsp. *liflandii* (designated pMUM002) and the sequence of a 105-kb section of non-PKS encoding plasmid DNA (pMUM003) from a fish isolate, *M. ulcerans* DL240490, have recently been completed (Pidot et al. 2008). Conservation of the *mls* locus between pMUM001 and pMUM002 is striking, with the most divergent PKS domains still maintaining 98.6% DNA identity (S.P., T. S., unpublished results). While large regions

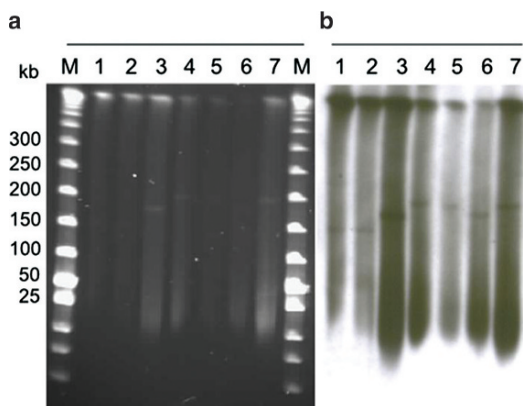


Fig. 3 Pulsed-field gel electrophoresis (a) and Southern hybridization (b) analysis of undigested DNA from a range of *M. ulcerans* strains. Southern hybridization was performed with a probe for IS2404. Lane identification is as follows: **M**, DNA size ladder; **1**, *M. ulcerans* 1615; **2**, *M. ulcerans* 753; **3**, *M. ulcerans* subsp. *Liflandii*; **4**, *M. pseudoshotsii* L15; **5**, *M. ulcerans* DL045; **6**, *M. ulcerans* CC240299; **7**, *M. ulcerans* DL240490

of non-PKS-related DNA are highly similar across the three plasmids, distinct differences, particularly in gene content, can also be observed (Fig. 2).

5 PKS Locus Variability between *M. ulcerans* Strains

The strict enzymatic activities conferred by the various domains of PKS enzymes means that changes in the DNA sequence of a PKS cluster can lead to the production of different metabolites. Within the mycolactone PKS locus, *mlsA1* and *mlsA2* are responsible for the synthesis of the core structure, and *mlsB* encodes the PKS responsible for synthesis of the side chain (see Fig. 4). The production of altered mycolactones has been identified not only in human clinical isolates but also in a number of environmental isolates. In a few instances, the genetic basis of this altered mycolactone production has been investigated. Structural data gained from a Chinese *M. ulcerans* clinical isolate showed that this strain produced a mycolactone with an additional methyl group at C2' of the side chain (Hong et al. 2005) (Fig. 4). Further analysis of this strain showed that the genetic basis behind the production of this altered mycolactone lay in a “domain swap”, where an AT domain in module 7 of *mlsB* was altered from acetate to propionate specificity compared to its counterpart in the sequenced strain *M. ulcerans* Agy99 (Hong et al. 2005). The phenomenon of domain swapping has also been observed in *M. ulcerans* subsp. *liflandii*, which was found to produce mycolactone E (Mve-Obiang et al. 2005) (Pidot et al. 2008) and in *M. marinum* DL240490, which was shown to produce mycolactone F (Pidot et al. 2008). To date, all mycolactone variants differ only in the side-chain portion of the molecule.

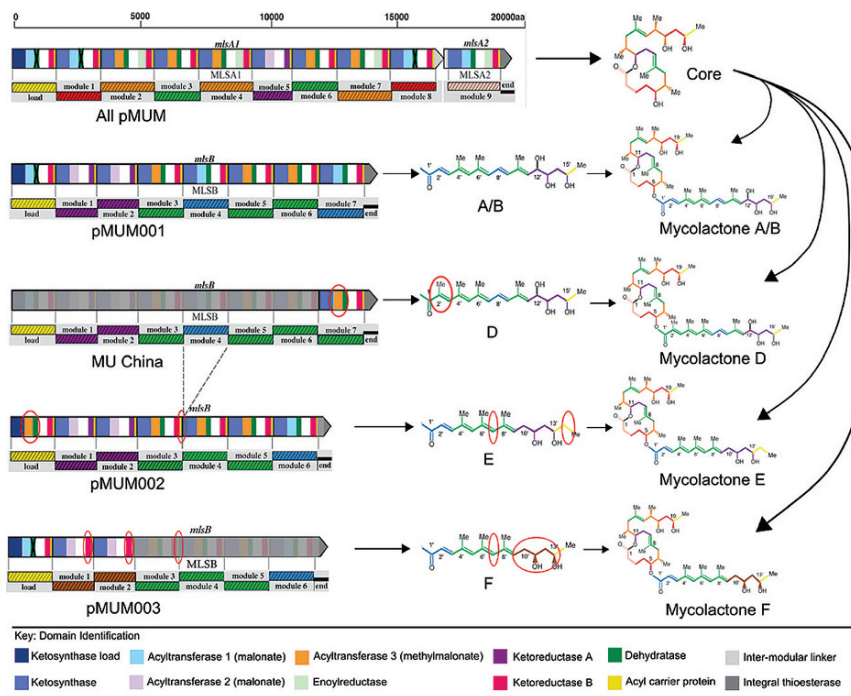


Fig. 4 Genetic organization of mycolactone biosynthetic clusters derived from the complete sequence of pMUM001 and pMUM002 for mycolactones A/B and E, respectively, and predicted from partial *mls* sequences and structural data for mycolactones D and F. Within each of the three genes (*mlsA1*, *mlsA2*, and *mlsB*), predicted enzymatic domain specificity is represented by a colored block, whose designation is outlined in the key. Red ellipses indicate key differences in *mlsB* gene structure and the corresponding alteration of mycolactone side-chain structure. Module arrangements are outlined below each gene, and each hashed box represents a PKS module. The color of each PKS module corresponds directly with the 2- or 3-carbon unit that it selects within the mycolactone molecule. Crosses through dehydratase domains indicate that they are predicted to be inactive on the basis of a mutation in the active site sequence. Grey boxes across a number of domains indicates the DNA sequence has not yet been determined and the corresponding module arrangements are structure-based predictions only

6 Non-PKS Elements of pMUM-like Plasmids

With the non-PKS sequence of three pMUM-like plasmids now unraveled, a comparison of the three elements reveals some interesting findings. Firstly, there is a distinct difference in size between the three plasmids, most of which is due to variations in the non-PKS-related DNA. When a comparison of pMUM001 and pMUM002 is performed, there are nine regions of difference, totaling 31.4 kb (16%), between the two plasmids. There is 22.1 kb of DNA, present in five distinct regions of pMUM003 that is absent from pMUM002, and there are 11 regions of difference, totaling 50.9 kb,

that are present in pMUM003 but not in pMUM001. When the reverse comparison is performed, there is only 2.2 kb of pMUM002 and 5.2 kb of pMUM001 that is not present in pMUM003. Across the three plasmids, this non-PKS sequence harbors a mixture of pseudogenes and coding sequences, representing a number of functional classes, including putative regulatory elements, hypothetical proteins, and insertion sequences (ISs). ISs or fragments of IS elements make up a significant portion of the non-PKS DNA. While multiple copies of the well-defined IS2404 and IS2606 insertion sequences are present in pMUM001 and pMUM003, only one intact copy of each of these IS sequences is present in pMUM002. The insertion sequence IS2404 formed the basis for the first diagnostic PCR used to detect the presence of *M. ulcerans* in clinical samples and has also been used to detect *M. ulcerans* and *M. ulcerans*-like organisms from the environment (Stinear et al. 1999; Stinear et al. 2000). While all three plasmids share a number of non-IS2404, non-IS2606 insertion sequences, pMUM002 and pMUM003 contain a number of elements that are not present on pMUM001. Both pMUM002 and pMUM003 contain an IS sequence (MULP_080 and MUDP_013, respectively) with high homology to an IS sequence from *Rhodococcus erythropolis*. Immediately upstream of this IS sequence in pMUM003 is a putative transposase (MUDP_014) with similarity to a transposase encoded by *R. erythropolis*. The MUDP_014 ortholog in pMUM002, however, is a pseudogene that has been inactivated by DNA deletion and rearrangement.

All three pMUM plasmids encode a large number of hypothetical proteins. There are 26 hypothetical proteins in pMUM001 of which a number are predicted to be membrane associated. As these hypothetical proteins are all unique to *M. ulcerans*-like mycobacteria, and only *M. ulcerans* strains have been identified from human infections, it is possible that they may represent surface-exposed antigens and may be able to be developed further for serodiagnostic applications (Stinear et al. 2005b). The pMUM plasmids have also been found to possess one or more putative serine/threonine signal transduction protein kinases (STPKs). These proteins are components of eukaryotic-like signaling systems and have been found in many bacterial genera to regulate cellular responses to environmental signals (Greenstein et al. 2005). One of these putative STPKs is present on all three plasmids (MUP011, MULP_022, MUDP_078), although MUDP_078 has become a pseudogene because of a frameshift mutation. Both pMUM002 and pMUM003 share another STPK (similar to other mycobacterial STPKs), while pMUM003 also contains a third putative STPK. Several mycobacterial *pkn* genes have been studied in detail, and many have been shown to phosphorylate proteins containing forkhead-associated (FHA) domains (Umeyama et al. 2002; Molle et al. 2003; Curry et al. 2005; Grundner et al. 2005; Villarino et al. 2005). FHA domains contain phosphothreonine-binding motifs and proteins containing this domain have been associated with cellular processes such as signaling DNA damage, vesicular transport, and cell cycle control in eukaryotes (Durocher and Jackson 2002; Hammet et al. 2003). However, the roles of FHA domain-containing proteins in prokaryotes are much less clear, although they have been hypothesized to either mediate the binding of cellular substrates or possibly function as kinase inhibitors (Greenstein et al. 2005). All the pMUM plasmids examined to date contain at least one protein containing a putative FHA domain, although none has a proven association

with any of the STPKs found on the plasmids. The environmental signals that may be “sensed” by these signal transducers also remain to be discovered.

Comparison of the replication region of the three plasmids shows a high degree of conservation of both gene sequence and order extending 6 kb upstream and 10 kb downstream of *repA*. When only pMUM002 and pMUM003 are compared, the extent of this synteny increases to 34.6 kb upstream and 12.5 kb downstream. The region downstream of *repA* includes a putative *par* locus, and upstream DNA comprises a number of well-conserved genes encoding hypothetical proteins. All known plasmid-partitioning loci contain two genes in an operon, *parA* and *parB*, that form a nucleoprotein complex at a centromere-like site and act independently of the replication mechanism of the plasmid (Gerdes et al. 2000). ParA homologs can be easily identified on each of the three plasmids from sequence comparisons; however, although a candidate *parB* exists directly downstream of *parA* in each pMUM plasmid, sequence comparison does not allow definitive identification of the function of these genes. While the high degree of conservation amongst pMUM *rep* regions may suggest that a large replication region is necessary for proper functioning of pMUM plasmids, studies have shown that a subclone consisting of 3.3 kb of pMUM001, encompassing a putative origin of replication (*oriM*), *repA*, and a gene encoding a hypothetical protein downstream of *repA* (MUP002) were sufficient for replication to occur in *M. marinum*, but could not support replication in *M. smegmatis* or *M. fortuitum* (Stinear et al. 2005b). However, stability experiments showed that a plasmid containing this 3.3 kb region was rapidly lost from culture in the absence of antibiotic selection. This may be explained by the absence of a complete *par* locus in this construct, and also suggests that other pMUM001 genes may be needed to support plasmid maintenance (Stinear et al. 2005b). Despite the *rep* region being highly syntenous, RepA from pMUM002 is 44 residues longer than its homolog in either pMUM001 or pMUM003, owing to the deletion of seven nucleotides towards the 5' end of the gene. This apparent increase in size does not appear to have a detrimental effect on the function of RepA, as the plasmid is still capable of replication.

Like other mycobacterial plasmids examined to date, the pMUM plasmids do not possess any form of identifiable conjugative transfer apparatus (Picardeau and Vincent 1998). How the pMUM plasmids have been acquired by *M. ulcerans* is unknown. The high level of DNA similarity among the *mls* genes suggests that they have evolved recently, owing to multiple duplication and recombination events (Stinear et al. 2004). Differences in size and gene content between the pMUM plasmids suggest that the smallest of these plasmids, pMUM001, has reduced the number of nonessential genes, a feature mirrored in the *M. ulcerans* genome (Stinear et al. 2007). This structure is indicative of its host passing through an evolutionary bottleneck and adapting to a new niche environment (Kaser et al. 2007; Stinear et al. 2007).

7 Concluding Remarks

The pMUM plasmids represent a group of homologous replicons whose primary function appears to be the production of the toxin mycolactone. The presence of the highly unstable *mls* genes on these plasmids and the metabolic cost to the bacterium

of maintaining these large, highly repetitive elements suggest that they play an important role in the lifestyle of the bacteria and perhaps give it a selective advantage. Laboratory experiments involving the colonization of the salivary glands of biting water insects by *M. ulcerans* have shown this process to be mycolactone dependant, and whether mycolactone gives a similar advantage to *M. ulcerans* in colonizing this niche in the environment remains to be seen (Marsollier et al. 2005). With the recent discovery of fish and frog pathogenic *M. ulcerans*-like organisms, there remains the possibility that many more pMUM-carrying bacteria remain to be isolated from the environment. Further investigation into this unique family of plasmids may serve to answer fundamental questions about their origin and their importance to the ecology of *M. ulcerans*.

References

- Aparicio JF, Molnar I, Schwecke T, Konig A, Haydock SF, Khaw LE, Staunton J, Leadlay PF (1996) Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* 169:9–16
- Coleman NV, Spain JC (2003) Distribution of the coenzyme M pathway of epoxide metabolism among ethene- and vinyl chloride-degrading *Mycobacterium* strains. *Appl Environ Microbiol* 69:6041–6046
- Connor DH, Lunn HF (1965) *Mycobacterium ulcerans* infection (with comments on pathogenesis). *Int J Lepr* 33 Suppl:698–709
- Coutanceau E, Marsollier L, Brosch R, Perret E, Goossens P, Tanguy M, Cole ST, Small PL, Demangel C (2005) Modulation of the host immune response by a transient intracellular stage of *Mycobacterium ulcerans*: the contribution of endogenous mycolactone toxin. *Cell Microbiol* 7:1187–1196
- Crawford JT, Bates JH (1979) Isolation of plasmids from mycobacteria. *Infect Immun* 24:979–981
- Crawford JT, Bates JH (1984) Restriction endonuclease mapping and cloning of *Mycobacterium intracellulare* plasmid pLR7. *Gene* 27:331–333
- Crawford JT, Cave MD, Bates JH (1981) Evidence for plasmid-mediated restriction-modification in *Mycobacterium avium intracellulare*. *J Gen Microbiol* 127:333–338
- Curry JM, Whalan R, Hunt DM, Gohil K, Strom M, Rickman L, Colston MJ, Smerdon SJ, Buxton RS (2005) An ABC transporter containing a forkhead-associated domain interacts with a serine-threonine protein kinase and is required for growth of *Mycobacterium tuberculosis* in mice. *Infect Immun* 73:4471–4477
- Debacker M, Aguiar J, Steunou C, Zinsou C, Meyers WM, Guedenon A, Scott JT, Dramaix M, Portaels F (2004) *Mycobacterium ulcerans* disease (Buruli ulcer) in rural hospital, Southern Benin, 1997–2001. *Emerg Infect Dis* 10:1391–1398
- Debacker M, Portaels F, Aguiar J, Steunou C, Zinsou C, Meyers W, Dramaix M (2006) Risk factors for Buruli ulcer, Benin. *Emerg Infect Dis* 12:1325–1331
- Durocher D, Jackson SP (2002) The FHA domain. *FEBS Lett* 513:58–66
- Erardi FX, Failla ML, Falkinham JO 3rd (1987) Plasmid-encoded copper resistance and precipitation by *Mycobacterium scrofulaceum*. *Appl Environ Microbiol* 53:1951–1954
- George KM, Chatterjee D, Gunawardana G, Welty D, Hayman J, Lee R, Small PL (1999) Mycolactone: a polyketide toxin from *Mycobacterium ulcerans* required for virulence. *Science* 283:854–857
- Gerdes K, Moller-Jensen J, Bugge Jensen R (2000) Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol* 37:455–466
- Greenstein AE, Grundner C, Echols N, Gay LM, Lombana TN, Miecskowski CA, Pullen KE, Sung PY, Alber T (2005) Structure/function studies of Ser/Thr and Tyr protein phosphorylation in *Mycobacterium tuberculosis*. *J Mol Microbiol Biotechnol* 9:167–181

- Grundner C, Gay LM, Alber T (2005) *Mycobacterium tuberculosis* serine/threonine kinases PknB, PknD, PknE, and PknF phosphorylate multiple FHA domains. *Protein Sci* 14:1918–1921
- Guerin WF, Jones GE (1988) Mineralization of phenanthrene by a *Mycobacterium* sp. *Appl Environ Microbiol* 54:937–944
- Hammet A, Pike BL, McNeese CJ, Conlan LA, Tennis N, Heierhorst J (2003) FHA domains as phospho-threonine binding modules in cell signaling. *IUBMB Life* 55:23–27
- Heathcote ML, Staunton J, Leadlay PF (2001) Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem Biol* 8:207–220
- Hellyer TJ, Brown IN, Dale JW, Easmon CS (1991) Plasmid analysis of *Mycobacterium avium-intracellulare* (MAI) isolated in the United Kingdom from patients with and without AIDS. *J Med Microbiol* 34:225–231
- Hockmeyer WT, Krieg RE, Reich M, Johnson RD (1978) Further characterization of *Mycobacterium ulcerans* toxin. *Infect Immun* 21:124–128
- Hong H, Demangel C, Pidot SJ, Leadlay PF, Stinear TP (2008) Mycolactones: immunosuppressive and cytotoxic polyketides produced by aquatic mycobacteria. *Nat Prod Rep* 25:447–454
- Hong H, Gates PJ, Staunton J, Stinear T, Cole ST, Leadlay PF, Spencer JB (2003) Identification using LC-MSn of co-metabolites in the biosynthesis of the polyketide toxin mycolactone by a clinical isolate of *Mycobacterium ulcerans*. *Chem Commun (Camb)* 22:2822–2823
- Hong H, Spencer JB, Porter JL, Leadlay PF, Stinear T (2005) A novel mycolactone from a clinical isolate of *Mycobacterium ulcerans* provides evidence for additional toxin heterogeneity as a result of specific changes in the modular polyketide synthase. *Chembiochem* 6:643–648
- Hong H, Stinear T, Porter J, Demangel C, Leadlay PF (2007) A novel mycolactone toxin obtained by biosynthetic engineering. *Chembiochem* 8:2043–2047
- Hopwood DA, Sherman DH (1990) Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* 24:37–66
- Hu Y, Coates AR (1999) Transcription of two sigma 70 homologue genes, sigA and sigB, in stationary-phase *Mycobacterium tuberculosis*. *J Bacteriol* 181:469–476
- Hull SI, Wallace RJ Jr, Bobey DG, Price KE, Goodhines RA, Swenson JM, Silcox VA (1984) Presence of aminoglycoside acetyltransferase and plasmids in *Mycobacterium fortuitum*. Lack of correlation with intrinsic aminoglycoside resistance. *Am Rev Respir Dis* 129:614–618
- Ishihama A (1988) Promoter selectivity of prokaryotic RNA polymerases. *Trends Genet* 4:282–286
- Jenke-Kodama H, Borner T, Dittmann E (2006) Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS Comput Biol* 2:e132
- Jensen AG, Bennedsen J, Rosdahl VT (1989) Plasmid profiles of *Mycobacterium avium/intracellulare* isolated from patients with AIDS or cervical lymphadenitis and from environmental samples. *Scand J Infect Dis* 21:645–649
- Johnson PD, Stinear T, Small PL, Pluschke G, Merritt RW, Portaels F, Huygen K, Hayman JA, Asiedu K (2005) Buruli Ulcer (*M. ulcerans* Infection): new insights, new hope for disease control. *PLoS Med* 2:e108
- Johnson PD, Azuolas J, Lavender CJ, Wishart E, Stinear TP, Hayman JA, Brown L, Jenkin GA, Fyfe JA (2007) *Mycobacterium ulcerans* in mosquitoes captured during outbreak of Buruli ulcer, southeastern Australia. *Emerg Infect Dis* 13:1653–1660
- Jucker MT, Falkingham JO 3rd (1990) Epidemiology of infection by nontuberculous mycobacteria IX. Evidence for two DNA homology groups among small plasmids in *Mycobacterium avium*, *Mycobacterium intracellulare*, and *Mycobacterium scrofulaceum*. *Am Rev Respir Dis* 142:858–862
- Kaser M, Rondini S, Naegeli M, Stinear T, Portaels F, Certa U, Pluschke G (2007) Evolution of two distinct phylogenetic lineages of the emerging human pathogen *Mycobacterium ulcerans*. *BMC Evol Biol* 7:177
- Krieg RE, Hockmeyer WT, Connor DH (1974) Toxin of *Mycobacterium ulcerans*. Production and effects in guinea pig skin. *Arch Dermatol* 110:783–788
- Le Dantec C, Winter N, Gicquel B, Vincent V, Picardeau M (2001) Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. *J Bacteriol* 183:2157–2164

- Lonetto M, Gribskov M, Gross CA (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* 174:3843–3849
- MacCallum P, Tolhurst J, Buckle G, Ha S (1948) A new mycobacterial infection in man. *J Pathol Bacteriol* 60:93–122
- Marsollier L, Aubry J, Coutanceau E, Andre JP, Small PL, Milon G, Legras P, Guadagnini S, Carbonnelle B, Cole ST (2005) Colonization of the salivary glands of *Naucoris cimicoides* by *Mycobacterium ulcerans* requires host plasmatocytes and a macrolide toxin, mycolactone. *Cell Microbiol* 7:935–943
- Meissner PS, Falkinham JO 3rd (1984) Plasmid-encoded mercuric reductase in *Mycobacterium scrofulaceum*. *J Bacteriol* 157:669–672
- Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV (2005) Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol* 23:1171–1176
- Molle V, Kremer L, Girard-Blanc C, Besra GS, Cozzone AJ, Prost JF (2003) An FHA phospho-protein recognition domain mediates protein EmrB phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. *Biochemistry* 42:15300–15309
- Moss SJ, Martin CJ, Wilkinson B (2004) Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* 21:575–593
- Mve-Obiang A, Lee RE, Portaels F, Small PL (2003) Heterogeneity of mycolactones produced by clinical isolates of *Mycobacterium ulcerans*: implications for virulence. *Infect Immun* 71:774–783
- Mve-Obiang A, Lee RE, Umstot ES, Trott KA, Grammer TC, Parker JM, Ranger BS, Grainger R, Mahrous EA, Small PL (2005) A newly discovered mycobacterial pathogen isolated from laboratory colonies of *Xenopus* species with lethal infections produces a novel form of mycolactone, the *Mycobacterium ulcerans* macrolide toxin. *Infect Immun* 73:3307–3312
- Picardeau M, Vincent V (1997) Characterization of large linear plasmids in mycobacteria. *J Bacteriol* 179:2753–2756
- Picardeau M, Vincent V (1998) Mycobacterial linear plasmids have an invertron-like structure related to other linear replicons in actinomycetes. *Microbiology* 144:1981–1988
- Pidot SJ, Hong H, Seemann T, Porter JL, Yip MJ, Men A, Johnson M, Wilson P, Davies JK, Leadlay PF, Stinear TP (2008) Deciphering the genetic basis for polyketide variation among mycobacteria producing mycolactones. *BMC Genomics*. Accepted 25 Sep 2008
- Portaels F, Chemlal K, Elsen P, Johnson PD, Hayman JA, Hibble J, Kirkwood R, Meyers WM (2001) *Mycobacterium ulcerans* in wild animals. *Rev Sci Tech* 20:252–264
- Quek TY, Athan E, Henry MJ, Pasco JA, Redden-Hoare J, Hughes A, Johnson PD (2007) Risk factors for *Mycobacterium ulcerans* infection, southeastern Australia. *Emerg Infect Dis* 13:1661–1666
- Ranger BS, Mahrous EA, Mosi L, Adusumilli S, Lee RE, Colorni A, Rhodes M, Small PL (2006) Globally distributed mycobacterial fish pathogens produce a novel plasmid-encoded toxic macrolide, mycolactone F. *Infect Immun* 74:6037–6045
- Rauzier J, Moniz-Pereira J, Gicquel-Sanzey B (1988) Complete nucleotide sequence of pAL5000, a plasmid from *Mycobacterium fortuitum*. *Gene* 71:315–321
- Read JK, Heggie CM, Meyers WM, Connor DH (1974) Cytotoxic activity of *Mycobacterium ulcerans*. *Infect Immun* 9:1114–1122
- Rhodes MW, Kator H, McNabb A, Deshayes C, Reyat JM, Brown-Elliott BA, Wallace R Jr, Trott KA, Parker JM, Lifland B, Osterhout G, Kaattari I, Reece K, Vogelbein W, Ottinger CA (2005) *Mycobacterium pseudoshottsii* sp. nov., a slowly growing chromogenic species isolated from Chesapeake Bay striped bass (*Morone saxatilis*). *Int J Syst Evol Microbiol* 55:1139–1147
- Stinear T, Ross BC, Davies JK, Marino L, Robins-Browne RM, Oppedisano F, Sievers A, Johnson PD (1999) Identification and characterization of IS2404 and IS2606: two distinct repeated sequences for detection of *Mycobacterium ulcerans* by PCR. *J Clin Micro* 37:1018–1023
- Stinear T, Davies JK, Jenkin GA, Hayman JA, Oppedisano F, Johnson PD (2000) Identification of *Mycobacterium ulcerans* in the environment from regions in Southeast Australia in which it is endemic with sequence capture-PCR. *Appl Environ Microbiol* 66:3206–3213

- Stinear TP, Mve-Obiang A, Small PL, Frigui W, Pryor MJ, Brosch R, Jenkin GA, Johnson PD, Davies JK, Lee RE, Adusumilli S, Garnier T, Haydock SF, Leadlay PF, Cole ST (2004) Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. Proc Natl Acad Sci USA 101:1345–1349
- Stinear TP, Hong H, Frigui W, Pryor MJ, Brosch R, Garnier T, Leadlay PF, Cole ST (2005a) Common evolutionary origin for the unstable virulence plasmid pMUM found in geographically diverse strains of *Mycobacterium ulcerans*. J Bact 187:1668–1676
- Stinear TP, Pryor MJ, Porter JL, Cole ST (2005b) Functional analysis and annotation of the virulence plasmid pMUM001 from *Mycobacterium ulcerans*. Microbiology 151:683–692
- Stinear TP, Seemann T, Pidot S, Frigui W, Reyssset G, Garnier T, Meurice G, Simon D, Bouchier C, Ma L, Tichit M, Porter JL, Ryan J, Johnson PD, Davies JK, Jenkin GA, Small PL, Jones LM, Tekaiia F, Laval F, Daffe M, Parkhill J, Cole ST (2007) Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. Genome Res 17:192–200
- Tang Y, Kim CY, Mathews II, Cane DE, Khosla C (2006) The 2.7-Ångstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc Natl Acad Sci U S A 103:11124–11129
- Trott KA, Stacy BA, Lifland BD, Diggs HE, Harland RM, Khokha MK, Grammer TC, Parker JM (2004) Characterization of a *Mycobacterium ulcerans*-like infection in a colony of African tropical clawed frogs (*Xenopus tropicalis*). Comp Med 54:309–317
- Ucko M, Colorni A (2005) *Mycobacterium marinum* infections in fish and humans in Israel. J Clin Microbiol 43:892–895
- Umeyama T, Lee PC, Horinouchi S (2002) Protein serine/threonine kinases in signal transduction for secondary metabolism and morphogenesis in *Streptomyces*. Appl Microbiol Biotechnol 59:419–425
- Villarino A, Duran R, Wehenkel A, Fernandez P, England P, Brodin P, Cole ST, Zimny-Arndt U, Jungblut PR, Cervenansky C, Alzari PM (2005) Proteomic identification of *M. tuberculosis* protein kinase substrates: PknB recruits GarA, a FHA domain-containing protein, through activation loop-mediated interactions. J Mol Biol 350:953–963
- Vivien E, Megessier S, Pieretti I, Cociancich S, Frutos R, Gabriel DW, Rott PC, Royer M (2005) *Xanthomonas albilineans* HtpG is required for biosynthesis of the antibiotic and phytotoxin albicidin. FEMS Microbiol Lett 251:81–89
- Weissman KJ, Leadlay PF (2005) Combinatorial biosynthesis of reduced polyketides. Nat Rev Microbiol 3:925–936
- Yip MJ, Porter JL, Fyfe JA, Lavender CJ, Portaels F, Rhodes M, Kator H, Colorni A, Jenkin GA, Stinear T (2007) Evolution of *Mycobacterium ulcerans* and other mycolactone-producing mycobacteria from a common *Mycobacterium marinum* progenitor. J Bacteriol 189:2021–2029

Megaplasmid pKB1 of the Rubber-Degrading Bacterium *Gordonia westfalica* Strain Kb1

Daniel Bröker and Alexander Steinbüchel(✉)

Contents

1	Introduction: Rubber-Degrading Members of the Interesting Genus <i>Gordonia</i> and Occurrence of Megaplasmids	298
2	Detection and Characterization of Megaplasmid pKB1	299
3	Identification of <i>oriV</i> of Plasmid pKB1 and Construction of <i>E. coli</i> – <i>Gordonia</i> Shuttle Vectors	300
4	Plasmid pKB1-Free Derivative Strains of <i>G. westfalica</i> Kb1	301
5	Possible Involvement of pKB1-Encoded Genes in Rubber Degradation	301
6	The Plasmid pKB1-Encoded Cadmium Resistance Gene <i>cadA</i> ; a Useful pKB1-Specific Selection Marker	303
7	Conjugation Region of Plasmid pKB1	304
8	The Necessity of Modification of Plasmid pKB1 for its Transfer to Cadmium-Resistant Relatives	305
9	<i>G. alkanivorans</i> Strain DSM 44187 Harbors a Megaplasmid Related to pKB1 from <i>G. westfalica</i> Strain Kb1	305
10	Concluding Remarks	306
	References	306

Abstract Sequence analysis of the circular 101,016-bp plasmid pKB1 from the rubber-degrading bacterium *Gordonia westfalica* strain Kb1 revealed 105 open reading frames (ORFs) which could be assigned to three functional groups (a) replication and partitioning, (b) catabolism, and (c) conjugative transfer. Successful conjugative transfer of pKB1 demonstrated the functionality of its conjugative transfer genes. The origin of replication of pKB1 was identified and used for construction of two *Escherichia coli*–*Gordonia* shuttle vectors suitable for several *Gordonia* species and related genera. As expression of the pKB1-encoded *cadA* in *E. coli* mediated resistance to cadmium, *cadA* was used as a pKB1-specific selection marker to monitor transfer of pKB1 by electroporation and conjugation to taxonomically

D. Bröker and A. Steinbüchel

Institut für Molekulare Mikrobiologie und Biotechnologie, Westfälische Wilhelms-Universität
Münster, Corrensstrasse 3, D-48149 Münster, Germany
e-mail: steinbu@uni-muenster.de

related bacteria, mediating cadmium resistance to a maximal concentration of 800 μM to the recombinant pKB1-harboring strains. Plasmid pKB1-free mutants of *G. westfalica* strain Kb1 had lost the ability to use natural rubber (NR) as sole carbon source, thereby suggesting that genes essential for NR degradation are encoded by pKB1. Transcription analysis of pKB1-encoded genes with a putative metabolic function revealed that ORF42 (a putative cytochrome *c* oxidase) and ORF6 (a putative epoxide hydrolase) are induced in cells of *G. westfalica* strain Kb1 during growth on NR but not or only slightly on sodium acetate. As genetic engineering of pKB1 in its host *G. westfalica* strain Kb1 failed due to the lack of an effective gene transfer system for this strain, pKB1 was transferred to genetically approachable strains for engineering and analysis of pKB1-encoded features.

1 Introduction: Rubber-Degrading Members of the Interesting Genus *Gordonia* and Occurrence of Megaplastids

The actinomycete genus *Gordonia* belongs to the suborder *Corynebacterineae*, which includes the families *Corynebacteriaceae*, *Dietziaceae*, *Mycobacteriaceae*, *Nocardiaceae*, *Tsukamurellaceae*, *Williamsiaceae*, and *Gordoniaceae* (the *Gordonia* family) (Stackebrandt et al. 1997). Members of the genus *Gordonia* have attracted much attention in recent years due to a variety of specific abilities comprising transformation, biodegradation, and synthesis of biochemical compounds, which were reviewed recently (Arenskötter et al. 2004). The number of described members of this genus is steadily increasing and at present, the genus *Gordonia* comprises 26 validly published species; some possess special anabolic capabilities, e.g., production of carotenoids by *G. jacobaea* strain MV-1 (De Miguel et al. 2000, 2001), production of gordonan, an acidic polysaccharide with cell aggregation-inducing activity in insect BM-N4 cells, by *Gordonia* sp. strain Y-102 (Kondo et al. 2000), production of exopolysaccharides by *G. polyisoprenivorans* (Fusconi et al. 2002), and transformation of *N*-(2-hexylamino-4-phenylimidazol-1-yl)-acetamide to 1-amino-4-phenylimidazol-2-yl-4-aminobutanoic acid by *G. rubropertincta* and *G. terrae*, respectively (Mikolasch et al. 2003). Furthermore, several members of the genus *Gordonia* exhibit special degradation activities such as dechlorination of atrazine and deethylsimazine as well as deamination of melamine and related compounds by a unique inducible *s*-triazine hydrolase from extracts of *G. rubripertincta* DSM 10347 (formerly *Rhodococcus corallinus*) (Cook and Hütter 1984, 1986; Mulbry 1994), desulfurization of benzothiophene by *G. desulfuricans* (Gilbert et al. 1998; Kim et al. 1999) or dibenzothiophene and diesel oils by *Gordonia* sp. strain CYKS1 (Rhee et al. 1998; Chang et al. 2001) and by *G. amicalis* (Kim et al. 2000), biodegradation of ethyl *t*-butyl ether (ETBE), methyl *t*-butyl ether (MTBE) and *t*-amyl methyl ether (TAME) by *G. terrae* (Hernandez-Perez et al. 2001), degradation of 3-ethylpyridine and 3-methylpyridine by *G. nitida* (Yoon et al. 2000), metabolism of several phthalic acid esters by *Gordonia* sp. strain MTCC 4818 (Chatterjee and Dutta 2003), degradation of car engine base oil by

Gordonia sp. strain NDKY76A (Koma et al. 2003), degradation of propane by *Gordonia* sp. strain TY-5 (Kotani et al. 2003, 2007), and degradation of natural rubber (poly(*cis*-1,4-isoprene), NR) by *G. westfalica* strain Kb1 (Linos et al. 2002) and strains of *G. polyisoprenivorans* (Linos et al. 1999; Arenskötter et al. 2001).

Three strains of *G. polyisoprenivorans* (Linos et al. 1999; Arenskötter et al. 2001) and *G. westfalica* strain Kb1 (Linos et al. 2002) were isolated from the foul water inside a deteriorated car tire from a farmer's field in Münster (Germany). These strains belong to the group of NR-degrading bacteria that shows an adhesive growth with direct contact of the cells to the rubber substrate (Linos et al. 2000). They were able to considerably disintegrate NR, either in the raw state as NR latex concentrate or in the vulcanized state as NR latex glove, as well as raw synthetic isoprene rubber (IR), which allows species of this genus to serve as model organisms for investigation of the biochemical and molecular mechanisms of rubber biodegradation, which are still only poorly understood (Rose and Steinbüchel 2005). Further representatives of this group belong to the genera *Gordonia*, *Mycobacterium* and *Nocardia* (Tsuchii et al. 1996; Linos and Steinbüchel 1998; Linos et al. 1999; Arenskötter et al. 2001; Linos et al. 2002). Members of another group form clear zones on latex-overlay agar plates, indicating an extracellular enzyme activity. They belong to the genera *Actinomadura*, *Actinoplanes*, *Dactylosporangium*, *Micromonospora*, *Microtetraspora*, *Nocardia*, and *Streptomyces* (Jendrossek et al. 1997a; Jendrossek et al. 1997b).

Since factors encoded by linear or circular plasmids are often involved in degradation of complex xenobiotics (Tan 1999; Negoro 2000; Fetzner et al. 2007), and since native plasmids provide the basis for the development of shuttle vectors (Denis-Larose et al. 1998; Nakamura et al. 2001) that may serve as tools to study *Gordonia*, rubber-degrading strains of the genus *Gordonia* were screened for occurrence of extrachromosomal DNA.

2 Detection and Characterization of Megaplasmid pKB1

The rubber-degrading bacteria *G. polyisoprenivorans* strains Kd2^T, VH2, and Y2K as well as *G. westfalica* strain Kb1 were screened for the occurrence of plasmids. Immobilized and linearized total DNA of these strains was separated by pulsed field gel electrophoresis. Only total DNA from *G. westfalica* strain Kb1 displayed a distinct band that was visible exclusively after linearization by S1 treatment. Therefore, this strain harbors a circular megaplasmid, designated pKB1 (Bröker et al. 2004). To obtain more information about pKB1 and its possible metabolic function, it was completely sequenced. The sequence disclosed that plasmid pKB1 is a circular DNA molecule with a size of 101,016 bp and a G + C content of 66 mol% (Table 1). Sequence analysis revealed the presence of 105 putative open reading frames (ORFs), 58 of which are transcribed in one direction and 47 in the other. Sixty-seven ORFs showed significant similarities at the amino acid level to other database entries, of which 46 were homologous to proteins with a putative function (Bröker

Table 1 Key features of plasmid pKB1

Host	<i>G. westfalica</i> strain Kb1
Type	Circular
Plasmid size (bp)	101,016
G + C content	66 mol%
Number of encoded ORFs	105
Associated phenotype	Cadmium resistance
Self-transmissible by conjugation	Yes
Accession number (Reference)	AJ576039 (Bröker et al. 2004)

et al. 2004). On the basis of the putative functions of the hypothetical pKB1-encoded gene products, three functional regions flanked by putative insertion sequence (IS) elements could be identified (a) a replication and putative partitioning region, (b) a putative metabolic region, and (c) a large conjugative transfer region, which is interrupted by an additional putative IS element (Bröker et al. 2004). Consequently, the megaplasmid pKB1 from the rubber-degrading bacterium *G. westfalica* strain Kb1 represents the first extrachromosomal DNA to be described for a member of the genus *Gordonia*.

3 Identification of *oriV* of Plasmid pKB1 and Construction of *E. coli*–*Gordonia* Shuttle Vectors

As there is still a lack of suitable genetic tools to apply genetic manipulation and recombinant DNA techniques to members of the genus *Gordonia*, the origin of replication (*oriV*) of plasmid pKB1 was identified and subsequently used to construct suitable *E. coli*–*Gordonia* shuttle vectors. For identification of *oriV* of pKB1, a plasmid library of pKB1 using the mobilizable suicide vector pBBR1MCS-2 (Kovach et al. 1995) was generated and transferred to *G. polyisoprenivorans* strain VH2 by electroporation (Bröker et al. 2004). Kanamycin-resistant colonies of *G. polyisoprenivorans* strain VH2 contained the suicide vector plus a cloned 7,258-bp fragment of plasmid pKB1, which was further reduced by exonuclease III and endonuclease treatment to identify the genes and *cis*-acting elements essential for replication. The resulting minimal region required for autonomous replication in *G. polyisoprenivorans* strain VH2 consisted of 2,332 bp, containing one ORF (ORF8) and a 39-bp sequence upstream of this ORF (Bröker et al. 2004). This 39-bp sequence was 92% identical at the nucleotide level to a region of plasmid pSOX from *Rhodococcus* sp. strain X309 which is located between the replication genes of this plasmid (Denis-Larose et al. 1998). Neither the 680 amino acid (aa) translational product of ORF8 nor its coding sequence showed similarities to proteins or DNA elements known to be involved in plasmid replication.

The 2,332-bp *oriV*-containing fragment of plasmid pKB1 and the vectors pBBR1MCS-2 (Km^R) and pBBR1MCS-5 (Gm^R) (Kovach et al. 1995), mediating

resistance to kanamycin and gentamycin, respectively, were used for construction of the two *E. coli*–*Gordonia* shuttle vectors pDBMCS-2 and pDBMCS-5 (Bröker et al. 2004). These *E. coli*–*Gordonia* shuttle vectors are suitable for transformation of *G. polyisoprenivorans* strains VH2 and Y2K, *M. smegmatis* strain mc²155, and *Rhodococcus opacus* strain PD630. Recently, it was demonstrated that one of these *E. coli*/*Gordonia* shuttle vectors (pDBMCS-2, Bröker et al. 2004) and an *E. coli*/*Rhodococcus* shuttle vector (pRE-7, Zheng et al. 1997) were suitable for transformation of *G. jacobaea* strain MV-1 by electroporation (Veiga-Crespo et al. 2006). Along with the *E. coli*–*Rhodococcus* shuttle vectors based on pNC903, which are also suitable for transformation of *G. polyisoprenivorans* strains Y2K and VH2 (Arenskötter et al. 2003), the two new *E. coli*–*Gordonia* shuttle vectors pDBMCS-2 and pDBMCS-5 add another functional replication system to the toolbox for *Gordonia*.

4 Plasmid pKB1-Free Derivative Strains of *G. westfalica* Kb1

*G. westfalica*Kb1 was cured of plasmid pKB1 to identify plasmid-encoded functions (Bröker et al. 2004). Southern hybridization and polymerase chain reaction (PCR) analysis of total DNA isolated from the resulting derivative strains revealed the insertion of a 51,527-bp region of plasmid pKB1 flanked by two putative IS elements (ORF1/ORF2 and ORF53, Fig. 1) into the chromosome. The other part of plasmid pKB1, comprising the replication and putative partitioning region, and the putative metabolic region got lost in these derivative strains. Therefore, they were suitable for further investigations with regard to a possible involvement of the putative metabolic region of pKB1 in rubber degradation. Growth experiments in the presence of NR as sole carbon source demonstrated that the plasmid pKB1-free derivative strains of *G. westfalica* strain Kb1 were no longer able to utilize this carbon source. Thus, it was concluded that genes essential for rubber degradation are encoded by the 49,489-bp region of pKB1 that is definitively missing in these derivative strains (Bröker et al. 2004).

5 Possible Involvement of pKB1-Encoded Genes in Rubber Degradation

The biodegradation of rubber in clear zone-forming and adhesively growing bacteria is still only poorly understood (Rose and Steinbüchel 2005). At present, there are two known enzymes, Lcp_{K30} and RoxA, catalyzing the primary poly(*cis*-1,4-isoprene)-cleaving reaction identified in *Streptomyces* sp. strain K30 (Rose et al. 2005) and in *Xanthomonas* sp. strain 35Y (Braaz et al. 2004, 2005), respectively. Furthermore, the involvement of an extracellular superoxide dismutase (SOD) (Arenskötter et al., unpublished data) and of an α -methylacetyl-CoA racemase (Mcr) (Banh et al. 2005)

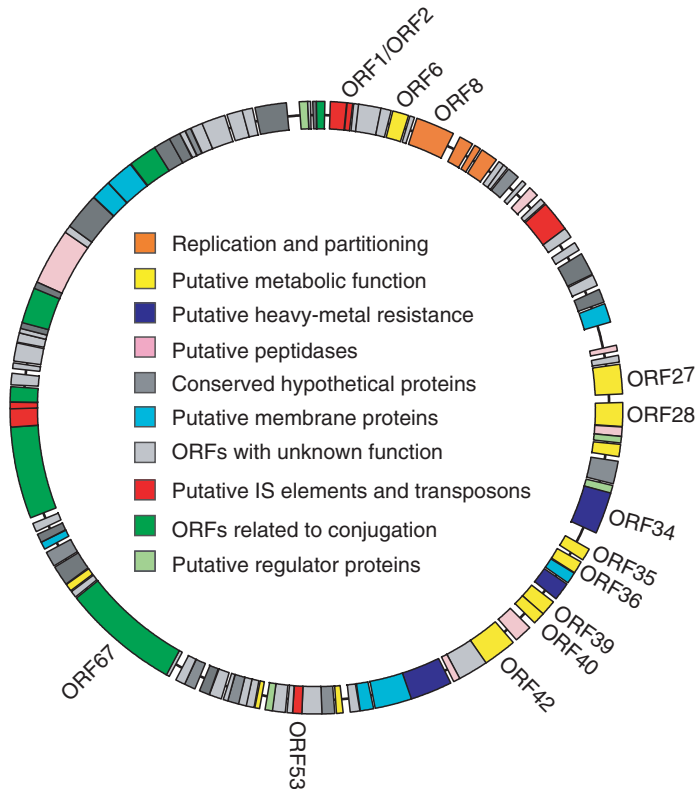


Fig. 1 Schematic representation of the circular 101,016-bp megaplasmid pKB1 from *G. westfalica* Kb1. The predicted 105 ORFs located on plasmid pKB1 are shown as boxes. ORFs mentioned in the text are labeled in the figure. The color code of the boxes explaining the assigned functions of the putative ORF products is located in the figure

in rubber degradation has been experimentally demonstrated. However, none of these proteins related to rubber degradation shares similarities to translational products of plasmid pKB1-encoded genes. Thus, the 49,489-bp region of plasmid pKB1 missing in the pKB1-free derivative strains of *G. westfalica* strain Kb1 seems to encode proteins with a hitherto unknown function in rubber degradation.

As electron transport proteins are often involved in catabolic and anabolic reactions, they might also be involved in rubber degradation. The translational products of ORF27, ORF39, and ORF42 are similar to deduced gene products of some *Rhodococcus* strains (a cytochrome *c* biogenesis protein from *Rhodococcus* sp. strain RHA1 [YP_701998]; 59% identity, a cytochrome *c* biogenesis protein CcdA from *R. erythropolis* strain PR4 [YP_345568]; 65% identity and a cytochrome *c* oxidase from *Rhodococcus* sp. strain RHA1 [YP_706379]; 79% identity, respectively) (Fig. 1). Cytochromes of the *c* type are part of several electron transport

systems and do not only function in aerobic or anaerobic respiration (Freedman and Chan 1984). In vitro experiments revealed that they are also able to perform peroxidase-like reactions in the presence of an electron acceptor like hydrogen peroxide (Vazquez-Duhalt 1999). ORF28 and ORF40 encode a putative divalent cation-transport integral membrane protein (similar to MNTH from *M. bovis* subsp. *bovis* AF2122/97 [CAD93809]; 69% identity) and a possible thioredoxin (similar to a thioredoxin from *Rhodococcus* sp. strain RHA1 [YP_702000], 62% identity), which may play a role in electron transport (Fig. 1). Furthermore, the translational products of ORF6, ORF35, and ORF36 might have putative metabolic functions during rubber degradation (based on similarities to an epoxide hydrolase from *Rhodococcus* sp. strain RHA1 [YP_701500]; 40% identity, an apolipoprotein-*N*-acyltransferase CutE from *R. erythropolis* [NP_898728]; 65% identity and an oxidoreductase from *M. gilvum* strain PYR-GCK [YP_001132180]; 59% identity, respectively) (Fig. 1). Hypothetically, cytochrome *c* may catalyze an epoxidation at the poly(*cis*-1,4-isoprene) molecule (Vazquez-Duhalt 1999). Subsequently, the resulting epoxide may be hydrolyzed to the corresponding diol by an epoxide hydrolase, which is putatively encoded by ORF6. Similar reactions catalyzed by other epoxide hydrolases have been described (Yamada et al. 2000), e.g., in the initiation of isoprene degradation by *Rhodococcus* sp. AD45 (Johan et al. 2000).

Reverse transcriptase-PCR analysis using isolated total RNA from cells of *G. westfalica* strain Kb1 grown on NR or sodium acetate as sole carbon source and oligonucleotides specific for amplification of the plasmid pKB1-encoded genes with a putative metabolic function mentioned above was performed to identify genes induced during rubber degradation (Bröker et al. unpublished data). Transcription of the pKB1-encoded ORF42 was induced in cells of *G. westfalica* strain Kb1 during growth on NR but not on sodium acetate. Furthermore, transcription of ORF6 was also induced in cells of *G. westfalica* strain Kb1 during growth on NR. Only weak transcription of ORF6 was detectable in cells grown on sodium acetate.

These results support the suggestion that plasmid pKB1 and especially the pKB1-encoded ORF6 and ORF42 are related to rubber degradation, and call for further studies involving genetic manipulation techniques. Several attempts to engineer pKB1 in its natural host *G. westfalica* strain Kb1 failed because of the absence of an effective gene transfer system for this strain. Therefore, megaplasmid pKB1 had to be transferred to genetically approachable strains for engineering. For successful transfer of plasmid pKB1, a pKB1-specific selection marker had to be identified.

6 The Plasmid pKB1-Encoded Cadmium Resistance Gene *cadA*; a Useful pKB1-Specific Selection Marker

Microbial heavy-metal resistance proteins have been intensively studied and characterized (Nies and Silver 2007). The translational product of the pKB1-encoded gene *cadA* (ORF34) exhibits the closest similarity (73% identity) to a plasmid-encoded

putative cadmium resistance protein (CadA) from *R. erythropolis* strain BD2 [NP_898704] (Stecker et al. 2003), and is related to a putative cadmium (Cd²⁺)/zinc (Zn²⁺)-transporting P-type ATPase (Fig. 1). Expression of *cadA* in *E. coli* strain RW3110 mediated resistance to cadmium, but not to zinc, and decreased the cellular content of cadmium in this host (Bröker et al. 2004). Furthermore, the generated pKB1-free derivative strains of *G. westfalica* Kb1 exhibited decreased cadmium resistance, thereby indicating that *cadA* confers cadmium resistance to its natural host *G. westfalica* strain Kb1. The suitability of *cadA* as pKB1-specific selection marker to monitor transfer of plasmid pKB1 by electroporation to the taxonomically related cadmium-sensitive strains *G. polyisoprenivorans* strains VH2 and Y2K, and *M. smegmatis* strain mc²155 was demonstrated with pKB1 mediating cadmium resistance to a concentration of 800 µM to the recombinant pKB1-containing strains. Transformation frequencies of plasmid pKB1 were 5.5×10^3 , 1.9×10^3 , and 8.3×10^2 transformants per microgram plasmid DNA for *G. polyisoprenivorans* strains VH2, *G. polyisoprenivorans* strain Y2K and *M. smegmatis* strain mc²155, respectively (Bröker et al. 2008).

7 Conjugation Region of Plasmid pKB1

The large putative conjugative transfer region of plasmid pKB1 comprises several ORFs whose translational products exhibit homologies to proteins involved in conjugational processes. The translational product of ORF67 exhibited 31% identity to TraA encoded by megaplasmid pREA400 of *R. erythropolis* strain AN12 [ABF48485] (Fig. 1), which was recently reported to be essential for conjugation (Yang et al. 2007). The *traA*-encoded DNA relaxases are the main elements in the initiation of conjugative plasmid transfer because they catalyze the cleavage of one plasmid strand by nucleophilic attack at the *nic* site within the *oriT* (for reviews see Byrd and Matson 1997; Grohmann et al. 2003).

The functionality of pKB1-encoded conjugative transfer genes has been experimentally demonstrated. Plasmid pKB1 was successfully transferred from its natural host *G. westfalica* strain Kb1 (donor) to a kanamycin-resistant mutant of *R. opacus* strain PD630 (recipient), designated as strain PD630Km100, by conjugation at mobilization frequencies of approximately 6.2×10^{-8} events per recipient cell. The pKB1-encoded cadmium resistance gene *cadA* was used for selection in this experiment, mediating cadmium resistance to a concentration of 800 µM to the exconjugants (Bröker et al. 2008). Thus, plasmid pKB1 represents the first conjugative plasmid identified in the genus *Gordonia*. The closest relatives known to contain conjugative megaplasmids are strains of the genus *Rhodococcus*, such as *R. erythropolis* strain AN12 harboring the megaplasmids pREA400 (400 kbp) and pREA250 (250 kbp), which were reported to be conjugative for the recipient *R. erythropolis* strain SQ1 at frequencies of approximately 7×10^{-4} events per recipient cell (Yang et al. 2007).

8 The Necessity of Modification of Plasmid pKB1 for its Transfer to Cadmium-Resistant Relatives

Plasmid pKB1 mediated cadmium resistance to *G. polyisoprenivorans* strains VH2 and Y2K as well as to *M. smegmatis* strain mc²155 and *R. opacus* strain PD630Km100, but did not confer rubber utilization to the nonrubber-degrading bacteria *M. smegmatis* strain mc²155 and *R. opacus* strain PD630Km100. As the latter were not able to grow on NR as sole carbon source, and as no cleavage of the residual NR material was detectable by gel permeation chromatography analysis, it was suggested that plasmid pKB1-encoded gene products are not essential for the initial poly(*cis*-1,4-isoprene) cleavage, but might be required for subsequent steps of the catabolic pathway (Bröker et al. 2008). Another suggestion was that *M. smegmatis* strain mc²155 and *R. opacus* strain PD630Km100 do not possess additional chromosomally encoded proteins that might be required for the complex processes of adhesive growth behavior and rubber degradation. To facilitate transfer of plasmid pKB1 into cadmium-sensitive strains of *Gordonia* and related taxa, which might possess the required functions, two additional antibiotic selection markers (kanamycin and gentamycin resistance) were introduced into pKB1 for analysis of plasmid encoded features.

Owing to the lack of an effective DNA transfer system for *G. westfalica* strain Kb1, the genetically approachable recombinant pKB1-containing *G. polyisoprenivorans* strain VH2 and *M. smegmatis* strain mc²155 were used to engineer plasmid pKB1. For this, a kanamycin resistance cassette was inserted into the pKB1-encoded *cadA* gene, ligated to the suicide plasmid pBBR1MCS-5 (Kovach et al. 1995). The resulting plasmid carrying Km^R and Gm^R determinants was electroporated into the pKB1-harboring strains. Homologous recombination between *cadA* on the suicide plasmid and the respective sequence in plasmid pKB1 led to an integration of the suicide plasmid into pKB1 for both recombinant strains. Consequently, plasmid pKB1 was provided with two antibiotic selection markers to monitor plasmid transfer into cadmium resistant strains of *Gordonia* and related taxa, e.g., for analysis of genes essential for rubber degradation (Bröker et al. 2008).

9 *G. alkanivorans* Strain DSM 44187 Harbors a Megaplasmid Related to pKB1 from *G. westfalica* Strain Kb1

The strain *G. nitida* and the two *G. alkanivorans* strains DSM 44187 and DSM 44369, which are the closest relatives of *G. westfalica* strain Kb1 (Arenskötter et al. 2005), were screened for plasmids. Interestingly, in contrast to the nonrubber-degrading strains *G. nitida* and *G. alkanivorans* strain DSM 44369, the rubber-degrading bacterium *G. alkanivorans* strain DSM 44187 harbors plasmid DNA (Baumeister et al. unpublished data).

Further investigations on the plasmid of *G. alkanivorans* strain DSM 44187 indicated a size of approximately 130 kbp by restriction analysis. Partial sequencing

and PCR analysis using pKB1-specific primers revealed nucleotide sequences identical to pKB1 (Bröker et al. unpublished data). Thus, the rubber-degrading *G. alkanivorans* strain DSM 44187 also harbors a megaplasmid, designated as pKB2, which contains regions identical to pKB1.

Efforts to transfer the modified plasmid pKB1 mediating resistance to kanamycin and gentamycin into the nonrubber-degrading, cadmium-resistant strains *G. nitida* and *G. alkanivorans* strain DSM 44369 to restore the ability for rubber degradation failed, suggesting either that plasmid-encoded kanamycin or gentamycin resistance genes were not functionally expressed or that transfer and propagation of foreign DNA was blocked by a highly efficient protection system (Bröker et al. 2008).

10 Concluding Remarks

Investigations on rubber biodegradation in strains of *Gordonia* were hampered because the lack of suitable genetic tools. By the construction of *E. coli*–*Gordonia* shuttle vectors and establishment of transfer and engineering procedures suitable for *Gordonia* and related taxa, a basis for identification of genes involved in rubber degradation was provided. Further investigations will be necessary to reveal the biochemical and molecular mechanisms of rubber biodegradation in the adhesively growing *Gordonia* sp.

Acknowledgements The authors are very grateful to the members of the laboratory, who were engaged in research on degradation of polyisoprenoides in the last 10 years. Research on microbial rubber degradation was in the past and is currently supported by grant provided by the Deutsche Bundesstiftung Umwelt (AZ. 13,072) and the Deutsche Forschungsgemeinschaft (St 386/10–1), respectively.

References

- Arenskötter M, Baumeister D, Berekaa MM, Pötter G, Kroppenstedt RM, Linos A, Steinbüchel A (2001) Taxonomic characterization of two rubber-degrading bacteria belonging to the species *Gordonia polyisoprenivorans* and analysis of hypervariable regions of 16 S rDNA sequences. *FEMS Microbiol Lett* 205:277–282
- Arenskötter M, Baumeister D, Kalscheuer R, Steinbüchel A (2003) Identification and application of plasmids suitable for transfer of foreign DNA to members of the genus *Gordonia*. *Appl Environ Microbiol* 69:4971–4974
- Arenskötter M, Bröker D, Steinbüchel A (2004) Biology of the metabolically diverse genus *Gordonia*. *Appl Environ Microbiol* 70:3195–3204
- Arenskötter M, Linos A, Schumann P, Kroppenstedt RM, Steinbüchel A (2005) *Gordonia nitida* Yoon et al. 2001 is a later synonym of *Gordonia alkanivorans* Kummer et al. 1999. *Int J Syst Evol Microbiol* 55:695–697
- Banh Q, Arenskötter M, Steinbüchel A (2005) Establishment of Tn5096-based transposon mutagenesis in *Gordonia polyisoprenivorans*. *Appl Environ Microbiol* 71:5077–5084
- Braaz R, Fischer P, Jendrosseck D (2004) Novel type of heme-dependent oxygenase catalyses oxidative cleavage of rubber poly(*cis*-1,4-isoprene). *Appl Environ Microbiol* 70:7388–7395

- Braaz R, Armbruster W, Jendrossek D (2005) Heme-dependent rubber oxygenase RoxA of *Xanthomonas* sp. cleaves the carbon backbone of poly(*cis*-1,4-isoprene) by a dioxygenase mechanism. *Appl Environ Microbiol* 71:2473–2478
- Bröker D, Arenskötter M, Legatzki A, Nies DH, Steinbüchel A (2004) Characterization of the 101.016-kbp megaplasmid pKB1 isolated from the rubber degrading bacterium *Gordonia westfalica* Kb1. *J Bacteriol* 186:212–225
- Bröker D, Arenskötter M, Steinbüchel A (2008) Transfer of megaplasmid pKB1 from the rubber-degrading bacterium *Gordonia westfalica* strain Kb1 to related bacteria and its modification. *Appl Microbiol Biotechnol* 77:1317–1327
- Byrd DR, Matson SW (1997) Nicking by transesterification: the reaction catalysed by a relaxase. *Mol Microbiol* 25:1011–1022
- Chang JH, Kim YJ, Lee BH, Cho K-S, Rye HW, Chang YK, Chang HN (2001) Production of a desulfurization biocatalyst by two-stage fermentation and its application for the treatment of model and diesel oils. *Biotechnol Prog* 17:876–880
- Chatterjee S, Dutta TK (2003) Metabolism of butyl benzyl phthalate by *Gordonia* sp. strain MTCC 4818. *Biochem Biophys Res Commun* 254:311–314
- Cook AM, Hütter R (1984) Deethylsimazine: bacterial dechlorination, deamination, and complete degradation. *J Agric Food Chem* 32:581–585
- Cook AM, Hütter R (1986) Ring dechlorination of deethylsimazine by hydrolases from *Rhodococcus corallinus*. *FEMS Microbiol Lett* 34:335–338
- De Miguel T, Sieiro C, Poza M, Villa TG (2000) Isolation and taxonomic study of a new canthaxanthin-containing bacterium, *Gordonia jacobaea* MV-1 sp. nov. *Int Microbiol* 3:107–111
- De Miguel T, Sieiro C, Poza M, Villa TG (2001) Analysis of canthaxanthin and related pigments from *Gordonia jacobaea* mutants. *J Agric Food Chem* 49:1200–1202
- Denis-Larose C, Bergeron H, Labbé D, Greer CW, Hawari J, Grossman MJ, Sankey BM, Lau PCK (1998) Characterization of the basic replicon of *Rhodococcus* plasmid pSOX and development of a *Rhodococcus-Escherichia coli* shuttle vector. *Appl Environ Microbiol* 64:4363–4367
- Fetzner S, Kolkenbrock S, Parschat K (2007) Catabolic linear plasmids. In: A (ed) Steinbüchel *Microbiology monographs*, vol 7. Springer, Berlin, pp 63–98
- Freedman JA, Chan SH (1984) Interactions in cytochrome oxidase: functions and structure. *J Bioenerg Biomembr* 16:75–100
- Fusconi R, Godinho MJL (2002) Screening for exopolysaccharide-producing bacteria from subtropical polluted groundwater. *Braz J Biol* 62:363–369
- Gilbert SC, Morton J, Buchanan S, Oldfield C, McRoberts A (1998) Isolation of a unique benzothiophene-desulfurizing bacterium, *Gordonia* sp. 213E (NCIMB 40816), and characterization of the desulfurization pathway. *Microbiology* 144:2545–2553
- Grohmann E, Muth G, Espinosa M (2003) Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* 67:277–301
- Hernandez-Perez G, Fayolle F, Vandecasteele J-P (2001) Biodegradation of ethyl *t*-butyl ether (ETBE), methyl *t*-butyl ether (MTBE) and *t*-amyl methyl ether (TAME) by *Gordonia terrae*. *Appl Microbiol Biotechnol* 55:117–121
- Jendrossek D, Tomasi G, Kroppenstedt RM (1997a) Bacterial degradation of natural rubber: a privilege of actinomycetes? *FEMS Microbiol Lett* 150:179–188
- Jendrossek D, Tomasi G, Schlegel HG (1997b) Mikrobiologischer Abbau von Kautschuk. *Nachrichten der Akademie der Wissenschaften in Göttingen. II mathematisch-physikalische Klasse. Nr 1*. Vandenhoeck & Ruprecht, Göttingen
- Johan ET, van Vlieg H, Leemhuis H, Spielberg JHL, Janssen DB (2000) Characterization of the gene cluster involved in isoprene metabolism in *Rhodococcus* sp. strain AD45. *J Bacteriol* 182:1956–1963
- Kim SB, Brown R, Oldfield C, Gilbert SC, Goodfellow M (1999) *Gordonia desulfuricans* sp. nov., a benzothiophene-desulfurizing actinomycete. *Int J Syst Bacteriol* 49:1845–1851
- Kim SB, Brown R, Oldfield C, Gilbert SC, Iliarionov S, Goodfellow M (2000) *Gordonia amicalis* sp. nov., a dibenzothiophene-desulfurizing actinomycete. *Int J Syst Bacteriol* 50:2031–2036

- Koma D, Sakashita Y, Kubota K, Fujii Y, Hasumi F, Chung S-Y, Kubo M (2003) Degradation of car engine base oil by *Rhodococcus* sp. NDKK48 and *Gordonia* sp. NDKY76A. *Biosci Biotechnol Biochem* 67:1590–1593
- Kondo T, Yamamoto D, Yokota A, Suzuki A, Nagasawa H, Sakuda S (2000) Gordonan, an acidic polysaccharide with cell aggregation-inducing activity in insect BM-N4 cells, produced by *Gordonia* sp. *Biosci Biotechnol Biochem* 64:2388–2394
- Kotani T, Yamamoto T, Yurimoto H, Sakai Y, Kato N (2003) Propane monoxygenase and NAD⁺-dependent secondary alcohol dehydrogenase in propane metabolism by *Gordonia* sp. strain TY-5. *J Bacteriol* 185:7120–7128
- Kotani T, Yurimoto H, Kato N, Sakai Y (2007) Novel acetone metabolism in a propane-utilizing bacterium, *Gordonia* sp. strain TY-5. *J Bacteriol* 189:886–893
- Kovach ME, Elzer PH, Hill DS, Robertson GT, Farris MA, Roop RM, Peterson KM (1995) Four new derivatives of the broad host range cloning vector pBRR1MCS, carrying different antibiotic resistance cassettes. *Gene* 166:175–176
- Linós A, Steinbüchel A (1998) Microbial degradation of natural and synthetic rubbers by novel bacteria belonging to the genus *Gordonia*. *Kautsch Gummi Kunstst* 51:496–499
- Linós A, Steinbüchel A, Spröer C, Kroppenstedt RM (1999) *Gordonia polyisoprenivorans* sp. nov., a rubber degrading actinomycete isolated from automobile tire. *Int J Syst Bacteriol* 49:1785–1791
- Linós A, Berekaa MM, Reichelt R, Keller U, Schmitt J, Flemming HC, Kroppenstedt RM, Steinbüchel A (2000) Biodegradation of *cis*-1,4-polyisoprene rubbers by distinct actinomycetes: Microbial strategies and detailed surface analysis. *Appl Environ Microbiol* 66:1639–1645
- Linós A, Berekaa MM, Steinbüchel A, Kim KK, Spröer C, Kroppenstedt RM (2002) *Gordonia westfalica* sp. nov., a novel rubber-degrading actinomycete. *Int J Syst Evol Microbiol* 52:1133–1139
- Mikolasch A, Hammer E, Schauer F (2003) Synthesis of imidazol-2-yl amino acids by using cells from alkane-oxidizing bacteria. *Appl Environ Microbiol* 69:1670–1679
- Mulbry WW (1994) Purification and characterization of an inducible *s*-triazine hydrolase from *Rhodococcus corallinus* NRRL B-15444R. *Appl Environ Microbiol* 60:613–618
- Nakamura M, Ogata K, Nagamine T, Tajima K, Matsui H, Benno Y (2001) The replicon of the cryptic plasmid pSBO1 isolated from *Streptococcus bovis* JB1. *Curr Microbiol* 43:11–16
- Negoro S (2000) Biodegradation of nylon oligomers. *Appl Environ Microbiol* 54:461–466
- Nies DH, Silver S (2007) Molecular microbiology of heavy metals. In: A (ed) Steinbüchel *Microbiology monographs*, vol 6. Springer, Berlin
- Rhee SK, Chang JH, Chang HN (1998) Desulfurization of dibenzothiophene and diesel oils by a newly isolated *Gordonia* strain, CYKS1. *Appl Environ Microbiol* 64:2327–2331
- Rose K, Steinbüchel A (2005) Biodegradation of natural rubber and related compounds: recent insights into a hardly understood catabolic capability of microorganisms. *Appl Environ Microbiol* 71:2803–2812
- Rose K, Tenberge KB, Steinbüchel A (2005) Identification and characterization of genes from *Streptomyces* sp. strain K30 responsible for clear zone formation on natural rubber latex and poly(*cis*-1,4-isoprene) rubber degradation. *Biomacromolecules* 6:180–188
- Stackebrandt E, Rainey FA, Ward-Rainey NL (1997) Proposal for a new hierarchical classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* 47:479–491
- Stecker C, Johann A, Herzberg C, Averbhoff B, Gottschalk G (2003) Complete nucleotide sequence and genetic organization of the 210-kilobase linear plasmid of *Rhodococcus erythropolis* BD2. *J Bacteriol* 185:5269–5274
- Tan H-M (1999) Bacterial catabolic transposons. *Appl Microbiol Biotechnol* 51:1–12
- Tsuchii A, Takeda K, Tokiwa Y (1996) Colonization and degradation of rubber pieces by *Nocardia* sp. *Biodegradation* 7:41–48
- Vazquez-Duhalt R (1999) Cytochrome *c* as a biocatalyst. *J Mol Catalysis B: Enzymatic* 7:241–249
- Veiga-Crespo P, Feijoo-Siota L, de Miguel T, Poza M, Villa TG (2006) Proposal of a method for the genetic transformation of *Gordonia jacobae*. *J Appl Microbiol* 100:608–614

- Yamada T, Morisseau C, Maxwell JE, Argiriadi MA, Christianson DW, Hammock BD (2000) Biochemical evidence for the involvement of tyrosine in epoxide activation during the catalytic cycle of epoxide hydrolase. *J Biol Chem* 275:23082–23088
- Yang JC, Lessard PA, Sengupta N, Windsor SD, O'Brien XM, Bramucci M, Tomb J-F, Nagarajan V, Sinsky AJ (2007) TraA is required for megaplasmid conjugation in *Rhodococcus erythropolis* AN12. *Plasmid* 57:55–70
- Yoon J-H, Lee JJ, Kang SS, Takeuchi M, Shin YK, Lee ST, Kang KH, Park YH (2000) *Gordonia nitida* sp. nov., a bacterium that degrades 3-ethylpyridine and 3-methylpyridine. *Int J Syst Evol Microbiol* 50:1203–1210
- Zheng H, Tkachuk-Saad O, Prescott JF (1997) Development of a *Rhodococcus equi*-*Escherichia coli* plasmid shuttle vector. *Plasmid* 38:180–187

Megaplasmid pMP118 of *Lactobacillus salivarius*

Paul W. O'Toole(✉) and Emma J. Raftis

Contents

1	Introduction to Lactobacilli.....	312
1.1	The Biology and Ecology of Lactobacilli.....	312
1.2	<i>Lactobacillus salivarius</i>	312
2	General Genetic Properties of pMP118.....	313
2.1	Discovery of pMP118.....	313
2.2	Replication Region of pMP118.....	313
2.3	GC Content and Coding Capacity.....	314
2.4	Insertion Sequence (IS) Elements and Transposons.....	315
2.5	Pseudogenes.....	315
3	Contribution of pMP118 to Genotype and Phenotype.....	315
3.1	Amino Acid Metabolism.....	316
3.2	Purine and Pyrimidine Anabolism.....	317
3.3	Acquisition of Core Chromosomal Genes by pMP118.....	318
3.4	Role of pMP118 in Carbohydrate Metabolism.....	318
3.5	Sensing and Regulation.....	319
3.6	Host Interaction.....	319
4	Distribution and Diversity of Megaplasmids in <i>L. salivarius</i>	321
4.1	Megaplasmid Distribution in <i>L. salivarius</i> and Other Species.....	321
4.2	Dissemination Mechanism.....	321
5	Concluding Remarks.....	322
	References.....	322

Abstract *Lactobacillus salivarius* UCC118 harbours a multiple-replicon genome, including a 242 kb megaplasmid designated pMP118. pMP118 carries a number of contingency genes, which work in conjunction with chromosomally encoded genes and pathways to broaden the metabolic flexibility of this strain. This increases the potential viability of UCC118 in the competitive environment of the gastrointestinal tract. Annotation and functional studies

P.W. O'Toole and E.J. Raftis
Department of Microbiology & Alimentary Pharmabiotic Centre,
University College Cork, Ireland
e-mail: pwotoole@ucc.ie

have indicated that pMP118 contributes significantly to the probiotic properties of UCC118, encoding a bile salt hydrolase gene and a potent broad-spectrum bacteriocin. pMP118-related megaplastids have been established as a general feature of the species *Lactobacillus salivarius*. Megaplastids have also been identified in six other *Lactobacillus* species of intestinal origin. Dissemination of the pMP118-related megaplastids may have occurred by a conjugation apparatus, which is now non-functional in pMP118. The analysis of pMP118 highlights the contribution of this replicon to the biology and ecology of this commensal species.

1 Introduction to Lactobacilli

1.1 *The Biology and Ecology of Lactobacilli*

Lactobacilli are members of the lactic acid bacteria (LAB) and form the most numerous group in the *Lactobacillaceae*; with the genus *Lactobacillus* containing over 100 species (Felis and Dellaglio 2007). Lactobacilli are members of the gut microbiota (Marteau et al. 2001), are associated with plants and fermented goods (Hammes and Vogel 1995) and have been used for industrial production of foods and beverages (Tannock 2004). Lactobacilli have been employed as probiotics, which are defined as “living micro-organisms which upon ingestion in certain numbers exert health benefits beyond inherent general nutrition” (Guarner and Schaafsma 1998, Klaenhammer 2000). To date, eleven lactobacilli genome sequencing projects have been published, thus increasing our understanding of the genetics, probiotic potential and metabolic capabilities of this important genus.

1.2 *Lactobacillus salivarius*

Lactobacillus salivarius is part of the indigenous microbiota of the gastrointestinal tract (GIT) and oral cavity of humans and hamsters (Molin et al. 1993, Rogosa et al. 1953, Wall et al. 2007). This species has also been isolated from human breast milk (Martín et al. 2006) and the intestinal tract of swine and chickens (Abbas et al. 2007, Casey et al. 2004, Mitsuoka 1969). Our recent molecular analyses have focused on strain *Lactobacillus salivarius* UCC118 (UCC118), which was isolated from the terminal ileum of a healthy human subject undergoing urinary tract reconstruction (Dunne et al. 1999). It has been extensively studied for its probiotic properties in both animal and human models (Corr et al. 2007, Dunne et al. 1999, Sheil et al. 2004).

2 General Genetic Properties of pMP118

2.1 Discovery of pMP118

Smaller plasmids are a common feature of LAB and specifically in the genus *Lactobacillus* (Wang and Lee 1997), with many strains harbouring multiple replicons. The LAB plasmids have been the focus of much study, as plasmid encoded traits have been shown to contribute significantly to the phenotype of industrially important strains of species belonging to the LAB group (Mills et al. 2006, Siezen et al. 2005). Despite this, relatively few large plasmids (>100 kb) had ever been described for the genus *Lactobacillus* (Muriana and Klaenhammer 1987, Roussel et al. 1993) and beyond these initial reports, there has been no further characterization. Megaplasmid pMP118 was the first to be definitively identified and characterised in a probiotic species, and is, to date, the largest plasmid to be sequenced in a LAB (Claesson et al. 2006). The original study by Claesson et al. designated pMP118 as a megaplasmid because (A) it does not encode the unique copy of any gene considered essential for viability of the cell; (B) it encodes neither rRNA nor tRNA genes and (C) it harbours plasmid-related replication genes (Claesson et al. 2006).

The genome of *L. salivarius* comprises four replicons, two of which, plasmids pSF118-20 (20.4 kb) and pSF118-44 (44 kb), had previously been sequenced (Flynn 2001). To further assess the plasmid complement of this strain, genomic DNA was treated with S1 nuclease, which preferentially linearises megaplasmid DNA and converts it into its equivalent unit length (Barton et al. 1995). Pulsed-field gel electrophoresis (PFGE) of the treated DNA resolved a discrete band, below the chromosomal band, which showed consistent migration under various PFGE conditions, relative to a linear DNA marker (Claesson et al. 2006). This electrophoretic mobility was consistent with that of a linearised plasmid. The newly found 242 kb megaplasmid was designated pMP118 (Claesson et al. 2006). Annotation of the pMP118 sequence identified the presence of a plasmid-associated replication gene, *repA*. This gene was experimentally localised to pMP118 using *repA* as a probe for Southern hybridisation analysis of the resolved plasmid profile of *L. salivarius* UCC 118 (Claesson et al. 2006).

2.2 Replication Region of pMP118

Replication of pMP118 is thought to proceed by theta replication, consistent with expectations for a plasmid of this size. Plasmid-encoded Rep proteins are involved in initiating replication and facilitating binding of host proteins (del Solar et al. 1998). The pMP118 *ori* region includes two Rep genes: *repA* and *repE* (Claesson et al. 2006). The replication region also includes the presumptive chromosome partitioning ATPase, ParA, the hypothetical gene *parB* and three pseudogenes. The *repA* gene

product displays 33% identity to the RepA/RepE protein of the *Enterococcus faecalis* theta-replicating plasmid pS86 (Claesson et al. 2006, Martínez-Bueno 2000); also the downstream *repE* gene is weakly related to staphylococcal plasmid replication proteins (Claesson et al. 2006). However, neither RepA boxes nor interons were identified in the replication region of pMP118 and it has not yet been assigned to a plasmid replication family (Claesson et al. 2006). The product of *parA* is 41% identical to a copy number control protein of the *Listeria innocua* plasmid pLI100 (Claesson et al. 2006). The copy number of pMP118 was originally determined relative to the chromosome and was estimated by quantitative PCR to be 4.7 ± 0.6 copies (Claesson et al. 2006). However, more recent comparative genomic hybridisation data (Raftis and O' Toole, unpublished) suggests that pMP118 is in fact present in only one copy per cell, but this remains to be verified by another method.

2.3 GC Content and Coding Capacity

The major features of the *L. salivarius* UCC 118 genome are outlined in Table 1. The 1.83 Mb chromosome is the smallest of the lactobacilli chromosomes published to date. However, including the plasmid content of this strain, the 2.13 Mb genome is of a comparable size to the other sequenced Lactobacilli, with the exceptions of *L. plantarum* (3.3 Mb) and *L. casei* (2.9 Mb). In comparison to the overall GC content of the genome (33.04%), both pSF118-20 and pSF118-44 have elevated GC contents (39.58% and 39.11%, respectively) (Flynn 2001). In contrast, both the chromosome and pMP118 have very similar GC contents (32.9% and 32.09%, respectively), suggesting that these two replicons are likely to have a coordinated evolutionary history, whereas the two smaller plasmids may have been a more recent acquisition.

Table 1 Physical and genetic features of the *L. salivarius* UCC118 chromosome and pMP118

Replicon	<i>L. salivarius</i> UCC118			
	chromosome	pMP118	pSF118-44	pSF118-20
Replicon size (bp)	1,827,111	242,436	44,013	20,417
Topology	Circular	Circular	Circular	Circular
% of total genome size	85.62	11.36	0.02	0.01
G + C content (%)	32.94	32.09	39.58	39.11
Gene density (%)	84.14	75.62	66.89	75.13
Predicted no. functional genes	1,117	222	48	27
Pseudogenes	49	20	4	0
tRNAs	78	0	0	0
rRNA operons	7	0	0	0
IS elements	32	11	1	2

2.4 Insertion Sequence Elements and Transposons

pMP118 harbours 25% percent of the insertion sequence (IS) content of the genome, representing seven different IS families. We identified two pairs of tandem IS elements present on both the chromosome and pMP118, in near identical arrangements, including flanking DNA sequences. On this basis we proposed that IS elements may have contributed to the formation or expansion of pMP118 by an integration of chromosomal genes into an earlier state of the pMP118 replicon (Claesson et al. 2006), as previously suggested for the 191 kb archaeal megaplasmid pNRC100 (Ng et al. 1998).

2.5 Pseudogenes

A recent comparative genomics study of the LAB has shown that gene loss and metabolic simplification are major trends within the evolution of the *Lactobacillales* (Makarova et al. 2006). At 49, the chromosome of strain UCC118 has one of the highest number (equal to *L. brevis*) of pseudogenes of the commensal lactobacilli of a similar genomic size. On a genome-wide scale, strain UCC118 includes 73 pseudogenes, with pMP118 making a substantial contribution (27%) to this number (Claesson et al. 2006). A high number of pseudogenes may be indicative of a replicon in an active state of gene inactivation and elimination, which is often associated with niche adaptation (Makarova et al. 2006), as noted recently for the sequenced strain *L. delbrueckii* ssp. *bulgaricus* (van de Guchte 2006).

Of the 20 pseudogenes identified on pMP118, genes encoding proteins of unknown function are the most prevalent, followed by genes predicted to encode restriction modification systems, transposases, surface proteins and ABC transporters. Frame shifting was found to be the most common form of inactivation of genes associated with both pMP118 and the chromosome (Claesson et al. 2006).

3 Contribution of pMP118 to Genotype and Phenotype

Figure 1 illustrates the COG (clusters of orthologous groups) assignments of the genes harboured by pMP118. Over half of these genes encode hypothetical proteins, and so are not assigned to a COG category. A further 13 are determined to have poor or no COG assignment predictions. Further functional analysis may help to characterise these proteins and determine their functional roles and by extension their role in the biology of UCC18. Of the genes that were assigned to COG categories, those associated with carbohydrate and amino acid transport and metabolism are highly represented in the genes carried by pMP118, as are genes associated with signal transduction and transcription (Fig. 1).

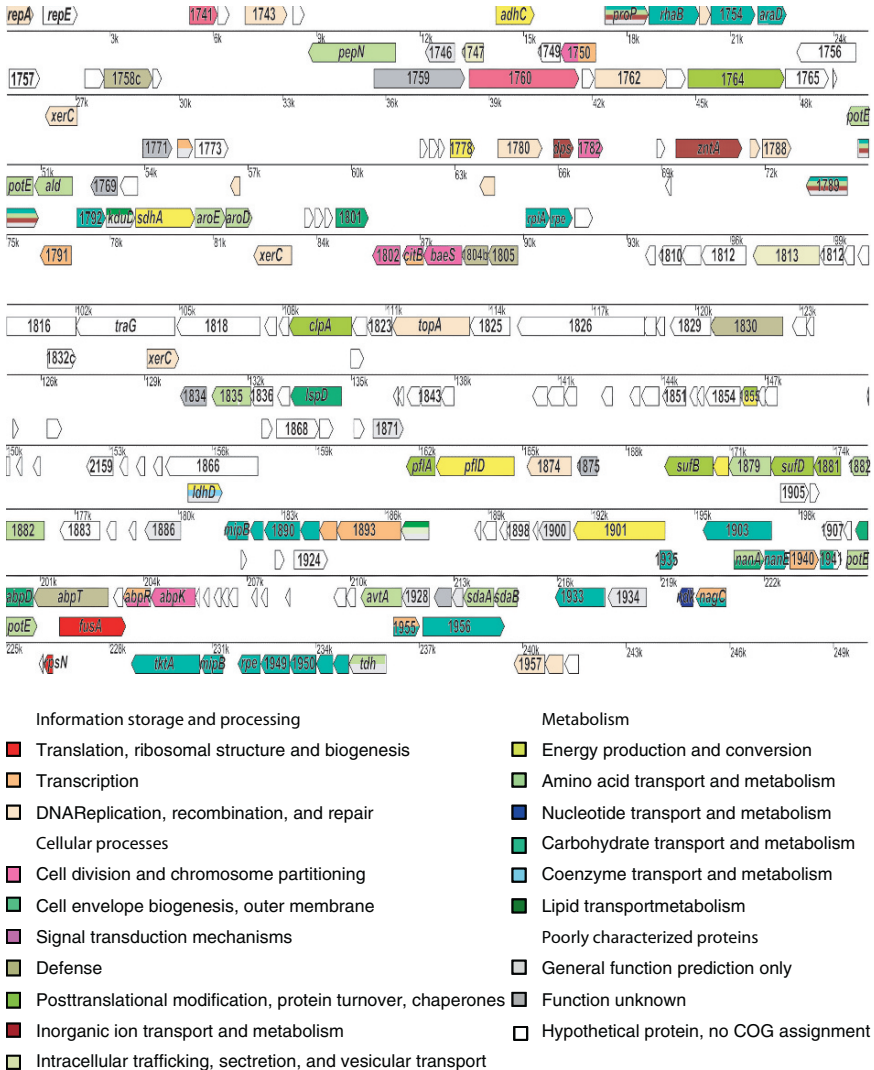


Fig. 1 Linear genome map of pMP118, modified from an image generated by Microbial Genome Viewer v1.0 (Kerhoven et al. 2004), representing the COG assignments of genes harboured by pMP118. Each four digit number corresponds to the locus tag of the represented gene, which includes the prefix LSL_. For clarity of presentation, a number of genes remain unlabeled. Pseudogenes are excluded from COG assignment. The accompanying colour key indicates the COG assignments of each gene.

3.1 Amino Acid Metabolism

pMP118 contributes neither rRNA nor tRNA genes to the strain, and so does not impart the advantages (e.g., rapid response to environmental change) that are associated with these genes (Klappenbach et al. 2000). However, pMP118

encodes a number of genes associated with increased biosynthetic capabilities. Among these (Fig. 1) is that for alanine dehydrogenase (LSL_1768, EC1.4.1.1), which is commonly found in the genus *Bacillus*, but is more rare among the sequenced lactobacilli having only been identified in *L. casei*, *L. brevis* (www.jgi.doe.gov Cited 14/07/2008) and *L. salivarius* (Claesson et al. 2006). This enzyme catalyses the reversible reductive amination of pyruvate into alanine in the presence of NAD⁺. Adjacent to LSL_1768 is a gene encoding a putative alanine permease (LSL_1767).

The chromosome of UCC118 encodes the enzymes *ldhL* and *ldhD*, which are involved in the generation of D-lactate. pMP118 harbours an additional copy of the *ldhD* gene (LSL_1887; Fig. 1), which may increase the efficiency of D-lactate production. This molecule is an important component of cell wall precursors in *L. plantarum* (Goffin et al. 2005)

pMP118 also harbours a paralog (LSL_1927) for one of two enzymes, encoded by the chromosome, that are required for the conversion of pyruvate to L-aspartate (Claesson et al. 2006).

pMP118 encodes both the alpha and the beta subunits of L-serine dehydratase (EC4.3.1.17). This enzyme catalyses the interconversion of pyruvate to serine; serine can subsequently be converted to glycine by a hydroxymethyltransferase, which is chromosomally encoded. Serine may also be thiolated to cysteine by the chromosomally encoded enzyme CysK. Cysteine can then be further converted into methionine, again using chromosomally encoded genes.

pMP118 encodes five genes involved in the biosynthesis of the aromatic amino acids, phenylalanine, tyrosine and tryptophan. Three of these are paralogs of genes encoded by the chromosome and two are uniquely coded for by pMP118, but none of these complete a functional pathway.

In silico analysis of the complete genome predicts that *L. salivarius* UCC118 can synthesise *de novo* or by interconversion nine amino acids, is capable of converting glutamine to three more and is auxotrophic for the remaining eight essential amino acids. By virtue of the unique and contingency genes that pMP118 harbours, *L. salivarius* UCC118 has an elevated level of prototrophy in comparison to other enteric lactobacilli of a similar chromosomal size. This may confer a competitive advantage to strain UCC118 over other lactobacilli in the GIT and broaden the range of intestinal ecological niches in which it can survive, in contrast to other small-genome lactobacilli such as *L. johnsonii* (Pridmore et al. 2004)

3.2 Purine and Pyrimidine Anabolism

The annotated genome suggests that *L. salivarius* UCC118 can synthesise purines and pyrimidines *de novo* from 5-phosphoribosyl-1-pyrophosphate (PRPP). Of note in this context, pMP118 harbours the genes that complete the pentose phosphate pathway (PPP) in strain UCC118 (see later for details), which may be used for the biosynthesis of PRPP. *L. salivarius* is predicted to generate UMP via a chromosomally encoded pathway, which can then be further converted to UTP and CTP by

utilising the megaplasmid encoded gene *dnk* (LSL_1936 diphosphate kinase) (Claesson et al. 2006). It is also possible for these conversions to be catalyzed by the chromosomally encoded gene product pyruvate kinase (LSL_0867) (Kilstrup et al. 2005). The ability to survive independently of host provisions of purines and pyrimidines may allow *L. salivarius* UCC118 to survive in a wider range of environments than those lactobacilli with a more auxotrophic nature.

3.3 Acquisition of Core Chromosomal Genes by pMP118

By comparing the predicted proteomes of five *Lactobacillus* species, Canchaya et al. identified 75 *Lactobacillus* core proteins (Canchaya et al. 2006). Interestingly, there were four genes present in the core that are carried by pMP118. LSL_1901 (Fig. 1) encodes a bi-functional acetaldehyde/alcohol dehydrogenase, which is the only enzyme present in UCC118 that catalyses the formation of ethanol from acetyl-CoA via acetaldehyde, thus providing an additional reductive pathway in UCC118 (Claesson et al. 2006). The three remaining core genes harboured by pMP118 include those for L-serine dehydratase α subunit, ribulose-phosphate 3-epimerase (discussed later) and a paralog of a chromosomally encoded ribose-5-phosphate isomerase gene. Thus the presence of these genes on the megaplasmid may be indicative of mobilization of genes between the chromosome and pMP118 (Canchaya et al. 2006).

3.4 Role of pMP118 in Carbohydrate Metabolism

L. salivarius was originally described as homofermentative (Rogosa et al. 1953). This would infer that like *L. gasserii* (Azcarate-Peril et al. 2008), *L. bulgaricus* (van de Guchte 2006) and *L. acidophilus* (Altermann et al. 2005), sugars would only be fermented via the Embden–Meyerhof–Parnas pathway to ultimately produce lactate as a sole product (Kandler 1983). The genes encoding the complete glycolysis pathway were found to be encoded by the chromosome of strain UCC118 (Claesson et al. 2006). However, we also annotated genes in the genome involved in the pentose phosphate pathway. This suggested the capability of fermenting pentose sugars as a sole carbon source, producing lactate, acetate, ethanol and CO₂ (Kandler 1983), which was subsequently proven experimentally for strain UCC118 (Claesson et al. 2006). pMP118 uniquely encodes two enzymes (Fig. 1), transketolase (LSL_1946) and transaldolase (LSL_1888, LSL_1947), which complete the pentose phosphate pathway that is partially encoded by the chromosome of *L. salivarius* UCC118. The contingency genes encoded by pMP118 are not essential for the biosynthesis of nucleotides when strain UCC118 is grown on glucose. However, the additional copy of a gene encoding ribose-5-phosphate isomerase (EC5.3.1.6) may increase the flexibility and flux of the pentose phosphate pathway (Claesson et al. 2006).

pMP118 was also discovered to harbour a fructose-1,6-bisphosphatase gene (LSL_1903; Fig. 1), which catalyses the formation of fructose-6-phosphate and which completes a gluconeogenesis pathway partly encoded by the chromosome. The presence of a complete gluconeogenesis pathway is unusual among the other sequenced lactobacilli. The presence of a functioning pentose phosphate pathway, together with a complete gluconeogenesis pathway, may be an adaptation to pentose-based growth (Claesson et al. 2006). A functioning pentose phosphate pathway would allow *L. salivarius* to utilise exogenous deoxyriboses, which are likely abundant in the GIT. The gluconeogenesis and the pentose phosphate pathways are prime examples of the synergistic functioning of megaplasmid and chromosomally encoded genes in the central metabolism of *L. salivarius* UCC118.

In silico analysis also predicted that genes encoded by pMP118 would enable strain UCC118 to assimilate sorbitol and rhamnose. The rhamnose fermentation pathway utilises rhamnulokinase, L-rhamnose isomerase and rhamnulose-1-phosphate aldolase, all of which are megaplasmid encoded genes (Claesson et al. 2006). Sorbitol-6-phosphate 2-dehydrogenase is also a megaplasmid encoded gene, necessary for the catabolism of sorbitol in this strain, which is a common component of the human diet. Putative phosphotransferase transporters of sorbitol were also identified upstream of this enzyme.

3.5 Sensing and Regulation

UCC118 harbours nine two-component regulatory systems, two of which are encoded by pMP118, and one of which is the system involved in control of expression of the bacteriocin Abp118 (discussed later), AbpK-AbpR (Fig. 1). Two chromosomally encoded orphan sensors were also identified, one of which (LSL_1454) is a paralog of the megaplasmid located gene LSL_1802 (Claesson et al. 2006). Seven of the 62 transcriptional regulator genes in the genome are present on pMP118 (Claesson et al. 2006), thus emphasizing the potential effect that pMP118 may have on the regulated gene expression and phenotype of this strain.

3.6 Host Interaction

One of the most interesting properties of many *L. salivarius* strains, including UCC118, is their probiotic nature. The contribution of extrachromosomal genes to this phenotype is therefore interesting from biological and evolutionary perspectives.

3.6.1 Bile Salt Hydrolase (BSH) Activity

Enteric species encounter many host defense mechanisms in the GIT. Microbial exclusion via host production of bile is one such mechanism. The ability to tolerate bile

imparts a selective advantage in the GIT environment (Tanaka et al. 1999) and has proven essential for GIT persistence (Dussurget et al. 2002). Many intestinal lactobacilli counteract the damaging effects of bile by encoding bile salt hydrolase enzymes (BSH) (Tanaka et al. 1999). BSH catalyses the breakdown of bile salts, causing the release of taurine or glycine (Tanaka et al. 1999). UCC118 is resistant to both acid and bile (Dunne et al. 1999) and was originally described as having both chromosomally encoded and megaplasmid encoded bile-inactivating enzymes (Claesson et al. 2006). However, recent analysis suggests that pMP118 harbours the sole gene primarily responsible for bile-salt hydrolase activity in this strain (LSL_1801, choloylglycine hydrolase). LSL_1801 shares 53% sequence identity with a BSH gene in *L. monocytogenes*. The survival rate of a *L. salivarius* UCC118 LSL_1801-knock out (KO) strain, grown in porcine bile (0.02%), was shown to be reduced by greater than 2 logs in comparison to the wild type strain (Fang and O' Toole, unpublished). Thus illustrating the potential contribution that the megaplasmid encoded BSH gene plays in the survival and adaptation of UCC118 to the GIT environment.

3.6.2 Surface Proteins

The interaction of bacteria with the intestinal epithelium is thought to modulate immune responses and improve mucosal integrity (Tannock 1999). A number of candidate host interaction proteins have been functionally analyzed in lactobacilli (Buck et al. 2005, Pretzer et al. 2005). The genome of strain UCC118 encodes 119 secreted proteins, 108 of which are encoded by chromosomal genes. Ten of the surface proteins encoded by UCC118 are sortase dependent; meaning they are covalently linked to peptidoglycan by the sortase protein (van Pijkeren et al. 2006). Four of these proteins are encoded by pMP118. Only one of these four genes was annotated as functional (LSL_1838, *lspD*; Fig. 1). However, the LspD gene-product was not found to be a significant component in adherence of UCC118 to Caco C2 cells, whereas the disruption of the chromosomally encoded gene *lspA* resulted in a significant adhesion reduction to this cell line (van Pijkeren et al. 2006). Interestingly, one of the sortase-dependent pseudogenes (LSL_1774b) present on pMP118 was annotated as a caseinolytic cell-surface protease, PrtP (Claesson et al. 2006), and shows homology to a *prtR* gene in *L. rhamnosus* (van Pijkeren et al. 2006). The *prtP* gene is non-functional, and may in part explain the inability of strain UCC118 to grow in milk (Raftis and O'Toole, unpublished). Niche adaptation and selective pressures may have induced the loss of *prtP* function in this GIT associated organism (Makarova et al. 2006, van Pijkeren et al. 2006).

3.6.3 Microbe-Microbe Interactions

A number of plasmids in *Lactobacillus* species are associated with bacteriocin production (Wang and Lee 1997), which is a desirable trait in probiotic strains. *L. salivarius* UCC118 produces Abp118, a broad-spectrum, heat-labile, class II

bacteriocin (Flynn 2001). The genes responsible for the production of Abp118 were originally described as chromosomally encoded (Flynn et al. 2002), but were later localised to pMP118 (Fig. 1) when the genome was sequenced (Claesson et al. 2006). Abp118 is regulated by a quorum-sensing mechanism, governed by the induction of peptide AbpIP (Flynn 2001). The gene conferring immunity to the action of Abp118 is encoded downstream of the genes encoding the Abp118 α and Abp118 β components. Abp118 has been established as the primary mechanism by which protection against *L. monocytogenes* infection is mediated in the mouse model (Corr et al. 2007). Furthermore, Abp118 has broad-spectrum activity (Flynn et al. 2002) and is likely to enhance the competitiveness of UCC118 in the GIT and to potentially modulate the microbiota.

4 Distribution and Diversity of Megaplasms in *L. Salivarius*

Prior to the identification of pMP118, many plasmid profiles of the lactobacilli were determined by methods that pre-dated the application of PFGE for separation of large DNA molecules. The procedure outlined earlier (Sect. 2.1.) that resulted in the discovery of pMP118 was first developed by Barton et al. and was later tailored for examining the complete plasmid profiles of *L. salivarius* and other species of the genus *Lactobacillus* (Barton et al. 1995, Claesson et al. 2006, Li et al. 2007).

4.1 Megaplasmid Distribution in *L. Salivarius* and Other Species

A recent study by Li et al. has revealed that the presence of megaplasms is a general feature of *L. salivarius* (Li et al. 2007). The presence of pMP118-related megaplasms (100–380 kb) was confirmed in 33 strains from diverse sources (Li et al. 2007). These megaplasms all hybridised to the *repA* gene probe derived from pMP118. Preliminary characterization of the strains indicated a diversity of genomic content, phenotypic characteristics, and megaplasmid size (100–380 kb). Subsequent comparative genomic hybridisation analysis of this panel of *L. salivarius* has elucidated a high level of genome plasticity within this species (Raftis and O' Toole, unpublished data). Li et al. (2007) also confirmed the presence of non-*repA* type megaplasms in an additional six *Lactobacillus* species, all of which can be found in the GIT (Li et al. 2007).

4.2 Dissemination Mechanism

The mode of dissemination of megaplasms in *L. salivarius* is as yet unclear. Interestingly, pMP118 was found to harbour a tract of genes (*tra* locus) that show

relatedness to known or suspected conjugation genes in other species, including *E. faecalis* (Claesson et al. 2006) and *L. lactis* (Fang et al. 2008). The *tra* locus of pMP118 spans a 35 kb region and appears to represent a remnant plasmid transfer locus (Claesson et al. 2006, Fang et al. 2008). A recent study from this laboratory showed that pMP118 is incapable of mobilising smaller plasmids and that successful conjugation cannot be achieved with the complement of genes that are present on pMP118 (Fang et al. 2008). Conjugation is a plausible mechanism through which pMP118-related plasmids might have disseminated among *L. salivarius* strains, and the *tra* region encoded by pMP118 may represent a remnant of a previously functioning conjugation pathway.

5 Concluding Remarks

It is evident from the *L. salivarius* genome annotation and functional studies that pMP118 contributes significantly to the biology of its bacterial host cell. pMP118 encodes a variety of proteins that work in conjunction with pathways encoded by the chromosome to increase the metabolic flexibility of UCC118, while also contributing significantly to its probiotic properties. Megaplasms of varying size are a general feature of the species *L. salivarius*. As a genetically well characterised strain-plasmid system, *L. salivarius* UCC118 provides a unique platform for investigating the potential development of megaplasmid-based replicons as cloning vectors for this species, or for cloning large operons in LAB. Further comparative genomic and functional analyses of environmentally diverse strains will help define the impact that size variation and genetic variation have on the genomic content and biological properties of *L. salivarius*.

Acknowledgements Work in PWOT's laboratory is supported by Science Foundation Ireland (CSET and RFP programmes).

References

- Abbas H, Hanan T, Surakka A, Apajalahti J, Saris PEJ (2007) Identification of the most abundant *Lactobacillus* species in the crop of 1- and 5-week-old broiler chickens. *Appl Environ Microbiol* 73:7867–7873
- Altermann E, Russell WM, Azcarate-Peril MA, Barrangou R, Buck BL, McAuliffe O, Souther N, Dobson A, Duong T, Callanan M, Lick S, Hamrick A, Cano R, Klaenhammer TR (2005) Complete genome sequence of the probiotic lactic acid bacterium *Lactobacillus acidophilus* NCFM. *Proc Natl Acad Sci USA* 102:3906–3912
- Azcarate-Peril MA, Altermann E, Goh YJ, Tallon R, Sanozky-Dawes RB, Pfeiler E, O'Flaherty S, Buck BL, Dobson A, Duong T, Miller MJ, Barrangou R, Klaenhammer TR (2008) Analysis of the genome sequence of *Lactobacillus gasseri* ATCC 1 33323 reveals the molecular basis of an autochthonous intestinal organism. *Appl Environ Microbiol* 74:4610–4625

- Barton BM, Harding GP, Zuccarelli AJ (1995) A general method for detecting and sizing large plasmids. *Anal Biochem* 226:235–240
- Buck BL, Altermann E, Svingerud T, Klaenhammer TR (2005) Functional analysis of putative adhesion factors in *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 71:8344–8351
- Canchaya C, Claesson MJ, Fitzgerald GF, van Sinderen D, O'Toole PW (2006) Diversity of the genus *Lactobacillus* revealed by comparative genomics of five species. *Microbiology* 152:3185–3196
- Casey P, Casey G, Gardiner G, Tangney M, Stanton C, Ross R, Hill C, Fitzgerald G (2004) Isolation and characterization of anti-Salmonella lactic acid bacteria from the porcine gastrointestinal tract. *Lett Appl Microbiol* 39:431–438
- Claesson MJ, Li Y, Leahy S, Canchaya C, van Pijkeren JP, Parkhill J, Flynn S, O'Sullivan GC, Higgins D, Shanahan F, Fitzgerald GF, van Sinderen D, O'Toole PW (2006) Multireplicon genome architecture of *Lactobacillus salivarius*. *Proc Natl Acad Sci USA* 103:6718–6723
- Corr SC, Li Y, Riedel CU, O'Toole PW, Hill C, Gahan CGM (2007) Bacteriocin production as a mechanism for the anti-infective activity of *Lactobacillus salivarius* UCC118. *Proc Natl Acad Sci USA* 104:7617–7621
- del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Dí az-Orejas R (1998) Replication and control of circular bacterial plasmids. *Microbiol Mol Biol Rev* 62:434–464
- Dunne C, Murphy L, Flynn S, O'Mahony L, O'Halloran S, Feeney M, Morrissey D, Thornton G, Fitzgerald G, Daly C, Kiely B, Quigley EM, O'Sullivan GC, Shanahan F, Collins JK (1999) Probiotics: from myth to reality. *Demonstration of functionality in animal models of disease and in human clinical trials. Antonie van Leeuwenhoek* 76:279–292
- Dussurget O, Cabanes D, Dehoux P, Lecuit M, Buchrieser C, Glaser P, Cossart P (2002) *Listeria monocytogenes* bile salt hydrolase is a PrfA regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol Microbiol* 45:1095–1106
- Fang F, Flynn S, Li Y, Claesson MJ, van Pijkeren J-P, Collins JK, van Sinderen D, O'Toole PW (2008) Characterization of Endogenous Plasmids from *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 74:3216–3228
- Felis G, Dellaglio F (2007) Taxonomy of Lactobacilli and Bifidobacteria. *Curr Issues Intest Microbiol* 8:44–61
- Flynn S (2001) Molecular characterisation of bacteriocin producing genes and plasmid encoded functions of the probiotic strain *Lactobacillus salivarius* subsp. *salivarius* UCC118. Ph.D Thesis. University College Cork, Cork, Ireland
- Flynn S, van Sinderen D, Thornton GM, Holo H, Nes IF, Collins JK (2002) Characterization of the genetic locus responsible for the production of ABP-118, a novel bacteriocin produced by the probiotic bacterium *Lactobacillus salivarius* subsp. *salivarius* UCC118. *Microbiology* 148:973–984
- Goffin P, Deghorain M, Mainardi JL, Tytgat I, Champomier-Verges MC, Kleerebezem M, Hols P (2005) Lactate racemization as a rescue pathway for supplying d-lactate to the cell wall biosynthesis machinery in *Lactobacillus plantarum*. *J Bacteriol* 187:6750–6761
- Guarner F, Schaafsma GJ (1998) Probiotics. *Int J Food Microbiol* 39:237–238
- Hammes WP, Vogel RF (1995) The genus *Lactobacillus*. In: Wood BJB, Holzapfel WH (ed) *The lactic acid bacteria*, vol 2. Blackie Academic and Professional, London, pp 19–54
- Kandler O (1983) Carbohydrate metabolism in lactic acid bacteria. *Antonie van Leeuwenhoek*:209–224
- Kerkhoven R, van Enckevort FHJ, Boekhorst J, Molenaar D, Siezen RJ (2004) Visualization for genomics: the microbial genome viewer. *Bioinformatics* 20:1812–1814
- Kilstrup M, Hammer K, Ruhdal Jensen P, Martinussen J (2005) Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol Rev* 29:555–590
- Klaenhammer TR (2000) Probiotic bacteria: today and tomorrow. *J Nutr* 130:415–416
- Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* 66:1328–1333
- Li Y, Canchaya C, Fang F, Raftis E, Ryan KA, van Pijkeren J-P, van Sinderen D, O'Toole PW (2007) Distribution of megaplasmids in *Lactobacillus salivarius* and other lactobacilli. *J Bacteriol* 189:6128–6139

- Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Diaz-Muniz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* 103:15611–15616
- Marteau P, Pochart P, Dore J, Bera-Maillet C, Bernalier A, Corthier G (2001) Comparative study of bacterial groups within the human cecal and fecal microbiota. *Appl Environ Microbiol* 67:4939–4942
- Martín R, Jiménez E, Olivares M, Marín ML, Fernández L, Xaus J, Rodríguez JM (2006) *Lactobacillus salivarius* CECT 5713, a potential probiotic strain isolated from infant feces and breast milk of a mother–child pair. *Int J Food Microbiol* 112:35–43
- Martínez-Bueno M, Valdívía E, Gálvez A, Mmqueda M (2000) pS86, A new theta-replicating plasmid from *Enterococcus faecalis*. *Current Microbiology* 41:257–261
- Mills S, McAuliffe OE, Coffey A, Fitzgerald GF, Ross RP (2006) Plasmids of lactococci – genetic accessories or genetic necessities?. *FEMS Microbiol Rev* 30:243–273
- Mitsuoka T (1969) Vergleichende untersuchungen Über die *Laktobazillen* aus den faeces von menschen, schweinen und hühnern. *Bakteriol* 210:32–51
- Molin G, Jeppsson B, Johansson ML, Ahre S, Nobaek S, Stahl M, Bengmark S (1993) Numerical taxonomy of *Lactobacillus* spp. associated with healthy and diseased mucosa of the human intestines. *J Appl Bacteriol*:314–323
- Muriana PM, Klaenhammer TR (1987) Conjugal transfer of plasmid-encoded determinants for bacteriocin production and immunity in *Lactobacillus acidophilus* 88. *Appl Environ Microbiol* 53: 553-560
- Ng WV, Ciuffo SA, Smith TM, Bumgarner RE, Baskin D, Faust J, Hall B, Loretz C, Seto J, Slagel J, Hood L, DasSarma S (1998) Snapshot of a large dynamic replicon in a Halophilic archaeon: megaplasmid or minichromosome? *Genome Res* 8:1131–1141
- Pretzer G, Snel J, Molenaar D, Wiersma A, Bron PA, Lambert J, de Vos WM, van der Meer R, Smits MA, MK (2005) Biodiversity-based identification and functional characterization of the mannose-specific adhesin of *Lactobacillus plantarum*. *J Bacteriol* 187:6128–6136
- Pridmore RD, Berger B, Desiere F, Vilanova D, Barretto C, Pittet AC, Zwahlen MC, Rouvet M, Altermann E, Barrangou R, Mollet B, Mercenier A, Klaenhammer T, Arigoni F, Schell MA (2004) The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. *Proc Natl Acad Sci USA* 101:2512–2517
- Rogosa M, Wiseman RF, Mitchell JA, Disraely MN, Beaman AJ (1953) Species differentiation of oral lactobacilli from man including descriptions of *Lactobacillus salivarius* nov spec and *Lactobacillus cellobiosus* nov spec. *J Bacteriol* 65:681–699
- Roussel Y, Colmin C, Simonet JM, Decaris B (1993) strain characterization, genome size and plasmid content in the *Lactobacillus acidophilus* group (Hansen and Mocolot). *J Appl Bacteriol* 74: 549-556
- Sheil B, McCarthy J, O'Mahony L, Bennett MW, Ryan P, Fitzgibbon JJ, Kiely B, Collins JK, Shanahan F (2004) Is the mucosal route of administration essential for probiotic function? Subcutaneous administration is associated with attenuation of murine colitis and arthritis. *Gut* 53:694–700
- Siezen RJ, Renckens B, van Swam I, Peters S, van Kranenburg R, Kleerebezem M, de Vos WM (2005) complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. *Appl Environ Microbiol* 71:8371–8382
- Tanaka H, Doesburg K, Iwasaki T, Mierau I (1999) Screening of lactic acid bacteria for bile salt hydrolase activity. *J Dairy Sci* 82:2530–2535
- Tannock GW (1999) Analysis of the intestinal microflora: a renaissance. *Antonie van Leeuwenhoek* 76:265–278
- Tannock GW (2004) A special fondness for Lactobacilli. *Appl Environ Microbiol* 70:3189–3194

- van de Guchte M (2006) The complete genome sequence of *Lactobacillus bulgaricus* reveals extensive and ongoing reductive evolution. *Proc Natl Acad Sci USA* 103:9274–9279
- van Pijkeren JP, Canchaya C, Ryan KA, Li Y, Claesson MJ, Sheil B, Steidler L, O'Mahony L, Fitzgerald GF, van Sinderen D, O'Toole PW (2006) Comparative and functional analysis of sortase-dependent proteins in the predicted secretome of *Lactobacillus salivarius* UCC118. *Appl Environ Microbiol* 72:4143–4153
- Wall R, Fitzgerald G, Hussey S, Ryan T, Murphy B, Ross P, Stanton C (2007) Genomic diversity of cultivable *Lactobacillus* populations residing in the neonatal and adult gastrointestinal tract. *FEMS Microbiol Ecol* 59:127–137
- Wang T, Lee BH (1997) Plasmids in *Lactobacillus*. *Critical Rev Microbiol* 17:227–272
- www.jgi.doe.gov (Cited 14/07/2008) Department of Energy Joint Genome Institute. <http://www.jgi.doe.gov>

Index

A

- ATPase(s)
 - AAA-type, 279
 - DnaA, 94
 - P-type, 229, 231, 232, 304
 - partitioning, 17, 19, 25, 74, 313
 - RhcN, 126
- abp/Abp*, 319–321
- abrB/AbrB*, 191, 194, 195
- acc/Acc*, 142
- Acetaldehyde, 39, 62, 318
- Acetaldehyde/alcohol dehydrogenase(s), 318
- Acetaldehyde dehydrogenase(s), 63
- Acetoacetate, 101
- Acetosyringone, 139
- Acetyl-CoA, 39, 45, 286, 318
- Acetyl-CoA acetyltransferase(s), 21
- Achromobacter xylosoxidans*, 43
- Acidovorax facilis*, 240
- Aconitase(s), 22
- acpA/AcpA*, 198, 199
- acs/Acs*, 138
- Actinobacillus actinomycetemcomitans*, 192
- Actinomycetales*, 48
- acv/Acv*, 141, 143
- Acyated homoserine lactones, 121
- Acyl Carrier Protein (ACP) domain, 286
- Addiction systems
 - pHG1, 243
 - pTi-SAKURA, 143
 - Shigella* virulence plasmids, 155, 162, 163
- Adenosine-5'-phosphosulphate (APS) kinase(s), 107
- Adenylyltransferase(s), 279
- Adhesins, 175, 176
- Adhesion
 - of *L. salivarius*, 320
 - of *S. meliloti*, 100
 - of *Y. pestis*, 179
 - of *Y. pseudotuberculosis*, 177
- α -galactoside(s), 101
- agp/Agp*, 101
- Agrobacterium*, 134
- Agrocinopine synthase(s), 138
- Agropine synthase(s), 138
- ags/Ags*, 138
- Alanine, 317
- Alanine dehydrogenase(s), 317
- Alanine permease(s), 317
- Alcaligenes hydrogenophilus* M50, 240
- Aldehyde dehydrogenase(s), 101
- α -methylacyl-CoA racemase(s), 301
- ami/Ami*, 41, 199
- Amidase(s), 21, 22, 199, 275
- Amine oxidase(s) (AOs), 277, 278
- 5-Amino-imidazole-4-carboxamide riboside (AICAR), 109
- amq/Amq*, 49
- Anaerobic class III ribonucleotide reductase(s), 255
- and/And*, 42, 79
- ant/Ant*, 71–74
- Anthracene, 57, 58, 63
- Anthranilate, 42, 48, 49, 73, 79
- Anthranilic acid, 67
- Antimonite, 15
- Antisense RNA, 162
- Antisigma factors, 227
- AO. *See amine oxidases (AOs)*
- Aquaspirillum autotrophicum* SA32, 240
- Arabinose, 101
- arc/Arc*, 15, 109
- Archaea
 - alkaliphilic, 6
 - genome organization, 4
- Arginine, 15, 109, 263

- Arginyl-tRNA synthetase(s), 25
 ARS. *See* autonomously replicating sequence (ARS)
ars/*Ars*, 15
 Arsenate reductase(s), 12, 24
 Arsenic resistance, 12, 15
 Arsenite, 12, 15, 18, 24, 193
 Arsenite(III)-methyltransferase(s), 15
Arthrobacter
 catabolic plasmids in, 35
 degradation of aromatics, 39
Arthrobacter aurescens, 272
Arthrobacter keyseri, 48
Arthrobacter nicotinovorans, 272
Arthrobacter nitroguajacolicus, 48
Arthrobacter sp. strain FB24, 272, 273
asn/*Asn*, 109
 ATP sulfurylase(s), 107
 Atrazine, 67, 298
attI recognition site, 63, 65
atxA/*AtxA*, 190, 191, 195, 196, 199
 Autoflavinylation, 275
 Autolysin, 199
 Autonomously replicating sequence (ARS), 94
 Avermectin, 285
Azohydromonas lata, 240
Azorhizobium caulinodans, 120
azu/*Azu*, 97, 109
- B**
 bacillary dysentery, 152
Bacillus, 57
Bacillus anthracis
 abr gene, 191, 194, 195
 capsule, 188, 190, 197, 198
 germination of spores, 188, 191
 prs genes, 190
 sporulation, 191, 195, 196
 taxonomic position, 188
 toxin production, 188–192, 198, 199
 type IV secretion systems, 192, 199
Bacillus cereus
 abr gene, 194, 195
 as an opportunistic pathogen, 188
 conservation of chromosomal sequences, 202
 toxin production, 194, 196
Bacillus schlegelii, 242
Bacillus subtilis, 198
Bacillus thuringiensis, 188, 197, 199, 201, 202
 Bacteriocins, 34, 179–181, 312, 319–321
bad/*Bad*, 42
bcl/*Bcl*, 195
bed/*Bed*, 41, 43
ben/*Ben*, 42
 Benzaldehyde, 41, 45
 Benzamide, 41
 Benzene, 39, 41, 42, 45
 Benzo[*a*]pyrene, 58
 Benzoate, 39, 41, 42, 45, 255
 Betaine(s), 100, 102, 105
 β -ketoacyl synthase(s), 107, 288
 β -glucans, 103, 104
 β -glucosidase(s), 274
 β -glucosyltransferase(s), 102, 104
bhb/*Bhb*, 101
 Bile salt hydrolase(s) (BSHs), 320
 Binome(s), 227
 Biphenyl, 41, 43, 47, 48, 66
 β -ketoacyl synthase(s), 107, 288
 β -ketoadipate(s), 42, 102
 β -Ketoadipate pathway(s), 42, 67, 73, 79, 102
 β -Ketoaldipyl-CoA, 42
box/*Box*, 42
bph/*Bph*, 41, 43, 47, 48
 6b protein, 138
Bradyrhizobium japonicum, 241, 259
 Bubonic plague, 173
Burkholderia
 degradation of aromatics, 39, 41
 degradation of PAHs, 57
Burkholderia cepacia
 degradation of 2,4-D, 43
Burkholderia glumae, 182
Burkholderia vietnamiensis, 42, 210, 219, 221, 227
Burkholderia xenovorans
 genome size, 34
 degradation of polychlorinated biphenyls, 41, 43
 multiple degradation pathways, 41
 Buruli ulcer, 284–286, 288
- C**
cad/*Cad*, 303–305
 Cadmium resistance
 pKB1, 304, 305
 Cadmium/zinc/cobalt resistance
 pMOL30, 210, 212, 223, 224, 228, 229, 231, 232
caf/*Caf*, 177, 178, 181
 Calvin-Benson-Bassham cycle, 251, 257, 263
 Camphor, 43, 58
 CAM plasmid, 43, 58
cap/*Cap*, 198, 199

- Capsule
B. anthracis, 188, 190, 197, 198
B. cereus, 188
Y. pestis, 177, 178
- Carbamoyl phosphate, 249
- Carbazole, 57, 65, 67, 70, 73, 74, 79
- Carbon monoxide
concentration in the atmosphere, 242
ligand in hydrogenase active site, 247, 249
production via photochemical degradation
of marine Organic matter, 259
- Carbophilus carboxydus*, 242
- Carboxidotrophs, 242
- car/Car*
car-IIKA1 gene cluster, 79
car-IKA1 gene cluster, 79
pCAR1, 70–73
P. resinovorans CA10, 67
Sphingomonas, 79
- Caseinolytic cell-surface protease(s), 320
- Catabolite repression
of *A. nicotinovorans* nicotine catabolism, 280
of the *O. carboxidovorans* OM5 *cox*
genes, 263
of the *R. eutropha* H16 *hox* genes, 251
- cat/Cat*, 42, 73, 79
- Catechol(s), 39, 41, 42, 45, 49, 58, 63, 67, 69, 79, 102
- Catechol 1,2-dioxygenase(s) (C12Os), 39, 255
- Catechol 2,3-dioxygenase(s) (C23Os), 39, 63
- Cation diffusion facilitators (CDFs), 17, 227, 231
- cbb/Cbb*, 243, 251–253, 258, 259, 273
- CcdA/CcdB toxin/antidote system, 151, 162
- C4-dicarboxylate transport system, 101
- Cell septa localization/formation, 96
- Cellulase(s), 274
- Cereulide, 196
- Chaperones
in *A. tumefaciens*, 140
in *B. anthracis*, 190, 192
in *Halobacterium* sp. NRC-1, 20, 21
in *M. ulcerans*, 287
in the maturation of MBH in *R. eutropha*
H16, 248
in *S. meliloti*, 107
in *Yersinia*, 176–178
- Chitooligosaccharide deacetylase(s), 107
- Chlorobenzene(s), 43
- Chlorobenzoate(s), 42, 43
- Chlorocatechol, 42, 43
- Chlorocatechol 1,2-dioxygenase(s), 42
- Chloromuconate cycloisomerase(s), 43
- chr/Chr*, 210, 212, 223, 227, 237, 238
- Chromate resistance, 211, 218, 227, 229
- Chromosome II of *H. marismortui*
gata, 21
genes for penicillin synthesis, 22
megaplasmid status, 21
orc, 21
parA genes, 21
rmB, 21
tm49, 21
- Chrysene, 57
chvE/ChvE, 139
- clc/Clc*, 43
- cnr/Cnr*, 210–212, 223, 227, 233
- Coagulases, 179
- CoA ligase(s), 42
- Cobalt/nickel resistance, 210–213, 227
- CO dehydrogenase(s) (CODHs)
attachment to the inner membrane, 264
[CuSMoO₃], 261, 262
FAD, 261, 262
gas channels, 261
genetic determinants, 259, 261–263
iron-sulfur clusters, 261, 262
maturation and assembly, 262
molybdopterin-cytosin dinucleotide
(MCD), 261, 262
regulation, 263
structure, 261, 262
- Coenzyme A (CoA) thioesters, 42
- Coiled-coil domains, 10, 192
- Colicins, 181
- Collagens, 177
- Comamonas*, 57
- Conjugative transfer
cupriavidus strains, 212
IncP-9 plasmids, 62
mycobacterial plasmids, 292
pAO1, 273
pAW63, 201
pCAR1, 72, 73
pHCG3, 259
pHG1, 243
pHV4, 18
pKB1, 304
pMOL28, 218–220
pMOL30, 222
pMP118, 321, 322
pNG600, 23
pNGR234a, 121
pSymA, 94, 95
pSymB, 94, 95
pXO2, 201
pYpsiP31758.2, 182
Shigella virulence plasmids, 162

Conjugative transfer (*cont.*)
 Ti and Ri plasmids, 143
 Yersinia plasmids, 180
cop/Cop, 14, 18–20, 155, 162, 210, 221, 224,
 229–233
 Copy number control
 Ti plasmids, 142
 Core proteins of lactobacilli, 318
 CO sensing, 263
cox/Cox, 259, 261–263, 279
CpxA/CpxR system, 157
 Cresol(s), 39, 41, 42, 45, 46
 CRISPR sequences, 18, 23
 Cryptic plasmids
 of *A. tumefaciens*, 142
 of mycobacteria, 284
 of *Yersinia species*, 181
 Crystal protein, 188
ctr/Ctr, 96
Cupriavidus
 in clean rooms, 218
 growth on organic solvents, 218
Cupriavidus metallidurans
 lithotrophy, 240
 taxonomical position, 212
 transcriptome, 232, 233
Cupriavidus metallidurans genomic
 islands
 CMGI-2, 222
 CMGI-3, 222
 CMGI-28a, 223, 227
 CMGI-30a, 223–225, 228
 CMGI-28b, 220, 223
 CMGI-30b, 223, 224, 229, 231
 CMGI-28c, 223
Cupriavidus taiwanensis, 219
cyc/Cyc, 109, 263
cyd/Cyd, 8, 12
cys/Cys, 317
 Cysteine, 317
 Cytochrome *a*₁, 263
 cytochrome *b*, 247, 264
 cytochrome *b*₆, 19
 Cytochrome *b*₅₅₈, 264
 Cytochrome *b*₅₆₁, 262, 264
 cytochrome *b*₅₆₃, 264
 Cytochrome *c*, 254, 263, 264, 302, 303
 Cytochrome oxidase(s), 8, 298, 412
 Cytochrome p450, 141, 288
 Cytokine production, 152, 153, 176
 Cytotoxins
 Shigella, 152
czc/Czc, 210, 212, 218, 223, 227–229,
 231, 232

D

2,4-D. *See* 2,4-dichlorophenoxyacetic acid
 (2,4-D)
dct/Dct, 101
 Dead Sea, 4
 Deethylsimazine, 298
 Dehydratase(s), 101, 286, 317, 318
 Denitrification
 R. eutropha H16, 243, 253, 254, 257
 S. meliloti, 97, 109
 De-ubiquinating cysteine protease(s), 176
 DHPH. *See* 2,6-dihydroxypyridine
 hydroxylase(s) (DHPHs)
 D-hydroxynicotine, 276, 277
 Diazotrophs
 H₂ oxidation by, 241, 242
 Dibenzofuran, 67, 80
 Dibenzo-*p*-dioxin, 67
 Dibenzothiophene, 41, 67, 298
 2,4-Dichlorophenoxyacetic acid
 (2,4-D), 35, 43, 81
 Dienelactone hydrolase(s), 43
 1,2-Dihydroxynaphthalene
 dioxygenase(s), 62
 2,6-dihydroxy-*pseudo*-oxynicotine, 275
 2,6-dihydroxy-*pseudo*-oxynicotine
 hydrolase(s) (PONHs), 275
 2,6-dihydroxypyridine, 275
 2,6-dihydroxypyridine hydroxylase(s)
 (DHPHs), 275
 3,4-Dimethylphenol, 39, 42, 43, 58, 62, 63,
 79, 255
 Dioxygenase(s), 39, 42, 43, 58, 62, 63, 79, 255
 Dioxygenation, 58, 67, 79
 DL-mesodiaminopimelic acid, 198
dmp/Dmp, 39, 45, 47
dnaA/DnaA, 94, 162
 DNA polymerase(s), 15, 17, 22, 139
dnk/Dnk, 318
 D-nornicotine, 271
 Domain swapping, 289
dsb/Dsb, 22, 222
 dTDP-rhamnose, 104

E

Ectoine(s), 101, 102
 Edema factor (EF), 190, 195
 EF. *See* edema factor (EF)
ehu/Ehu, 102
 Electrophoresis in Eckhardt gels, 120, 124
 Embden-Meyerhof-Parnas pathway, 318
 Endonuclease(s) III, 21
Enterobacteriaceae, 81, 172

Enteroinvasive *Escherichia coli* (EIEC), 152
envZ/EnvZ, 157
 Epoxidation, 303
 Epoxide hydrolase(s), 298, 303
 EPR spectroscopy
 of *O. carboxidovorans* OM5
 CODH, 261
 of the *R. eutropha* H16 SH, 250
 Erythromycin, 285
 3-Ethylpyridine, 298
 Ethyl t-butyl ether (ETBE), 298
eut/Eut, 102
 EXAFS, 261
 Excinuclease(s), 18
 Exopolysaccharides (EPSs), 92, 101, 103, 110, 120, 298

F

Flavin adenine dinucleotide (FAD), 28, 261, 262, 275, 277
 F1 antigen, 178, 179
 Far east scarlet-like fever (FESLF)
 Y. pseudotuberculosis, 182
 FB24-1, 273
 FB24-2, 273
 FB24-3, 273
 F1 capsular antigen, 177
 F1 capsule anchoring protein, 177
 F1 capsule chaperone protein, 177
 F1 capsule regulatory protein, 177
fdrI/FdrI, 79
fdrII/FdrII, 79
fdx/Fdx, 79, 108, 125
 [Fe(CN⁻)₂CO] complex of the *R. eutropha* H16 MBH, 248, 249
 Fermented goods, 312
 Ferredoxin, 58, 62, 63, 79, 125
 Ferrulic acid, 141
 ΦH, 11
 FHA domains. *See* forkhead-associated (FHA) domains
fhp/Fhp, 254
 Fibronectin, 177
 FISH, 94
fix/Fix, 98, 105, 108, 125, 127
 Flavin mononucleotide (FMN), 250
 Flavoenzyme(s), 275
 Flavohemoprotein(s), 254
 Flavonoid(s), 104, 105, 124–127
 Flavoprotein oxidoreductase(s), 109
 Flavoproteins, 261, 262
 Fluorene, 41, 79

Folding

 of *S. meliloti* Nod proteins, 107
 of *Y. pestis* F1 antigen, 178
fol/Fol, 277
 Folylpoly-γ-glutamate ligase(s), 198
 F1 operon, 178, 279, 280
 Forkhead-associated (FHA) domains, 291
 Formimidoyltetrahydrofolate
 cyclodeaminase(s), 21
 Formyltetrahydrofolate deformylase(s), 277
 Formyltetrahydrofolate synthetase(s), 21
 Fructose-1,6-bisphosphatase(s), 252, 263, 319
 Fructose-1,6-bisphosphate aldolase(s), 252, 263
 Fructose-6-phosphate, 319
fts/Fts, 15, 18, 95, 96
 Fumarate, 101
fur/Fur, 158

G

Galactoglucan (EPS II), 103, 104
 Galactosyltransferase(s), 104
 GALLS gene, 141
 γ-aminobutyrate, 273, 275, 277, 278
 Gastroenteritis, 173
 Gas vesicles, 4, 9, 10, 12
gat/Gat, 21
 Gene 6b. *See* *tml/Tml*
 Genetic colonization, 138
 Genome reduction, 315
 Genomic islands, 34, 47, 56, 72, 221, 222, 227, 233
 Genomic rearrangement, 121, 123, 124, 220
 Gentsiate(s), 39, 42
ger/Ger, 191
glg/Glg, 97
 Gluconeogenesis, 319
 Glucosamine synthase(s), 107
 Glucose-1-phosphate, 103, 104
 Glucose-6-phosphate, 103
 Glucosyltransferase(s), 104
 Glutamine-dependent asparagine synthetase(s), 109
 Glutamyl-transpeptidase(s), 198
 Glyceraldehyde-3-phosphate
 dehydrogenase(s), 252, 263
 Glycogen metabolism, 97
 Glycolipids, 274
 GMP synthase(s), 21
 γ-N-methylaminobutyrate, 275, 277, 278, 280
 γ-N-methylaminobutyrate oxidase(s) (MABOs), 277, 278
 Gordonan, 298

- Gordonia*
 biodegradation, 298, 299
 taxonomy, 298
- Gordonia alkanivorans*, 305
- Gordonia amicalis*, 298
- Gordonia jacobaea*
 production of carotenoids, 298
 transformation of, 301
- Gordonia nitida*, 298, 305
- Gordonia polyisoprenivorans*, 298, 299
 isolation, 299
 transformation of, 300, 301
- Gordonia rubropertincta*, 298
- Gordonia terrae*, 298
- Gordonia westfalica*
 insertion sequences (IS), 300
 isolation, 299
- Great Salt Lake, 4
- groEL*/GroEL, 98, 107
- Group II introns, 99, 257
- gsp*/Gsp, 225
- GTPase-activating proteins, 176
- GTPases, 15, 18, 191
- gtr*/Gtr, 223, 224, 229
- gvp*/Gvp, 12, 15
- H**
- Haloarchaea, 4–6, 12, 25–27
- Haloarcula marismortui*, 4, 20
- Halobacterium* sp. NRC-1
 DNA rearrangements, 4
 gas vesicle-deficient mutants, 9, 10
 gas vesicle genes, 9
 genome organization, 11
 horizontal gene transfer, 26
 megaplasmid evolution, 25
orc genes, 15, 16
 transcription factors, 16
- Haloferax volcanii*, 4, 11, 15, 16, 19
- Haloperoxidase(s), 98
- Haloquadratum walsbyi*, 6, 10, 24
- Helicase(s), 18, 19, 22, 23, 25
- hem*/Hem, 97
- Heme(s), 97, 247, 254, 259
- Heteroaromatics, 67
- Heterologous expression studies, 262
- Hexose oxidation, 274
- 6HLNO. *See* 6-hydroxy-L-nicotine oxidase(s)
 (6HLNOs)
- hmg*/Hmg, 42
- H-NS, 156
 binding sites, 156–158
 MvaT family, 74, 75
- interaction with VirB, 163
 silencing of imported genes by, 161
 structure, 157
 suppression of *Shigella* genes by, 157, 159
- hod*/Hod, 49
- Homogentisate, 42
- Horizontal gene transfer
 in *Burkholderia xenovorans* LB400, 41
 of *car* genes, 67
 between haloarchaeal megaplasms,
 18, 26
 of heavy metal resistance genes, 223, 231
 of *meta*-cleavage pathway genes, 63
 of *nic* genes, 273
 pNGR234a, 122
 of *S. meliloti* genes, 95, 99, 100, 122
 in *S. meliloti*, 95, 101, 125
 of *Shigella* virulence factors, 155, 161, 180
- hox*/Hox, 243, 247–251, 259, 264
- hsn* loci, 125
- htl*/Htl, 12
- hup*/Hup, 264
- Hydratase(s), 63, 101
- Hydrogenophaga pseudoflava*, 242
- Hydrogenotrophs, 241, 242
- Hydrolase(s), 43, 63, 193, 196, 255, 274, 275,
 298, 303
- Hydroperoxidase(s), 98
- Hydroquinone, 39
- Hydroxybenzalpyruvate hydratase-
 aldolase(s), 62
- 2-Hydroxychromene-2-carboxylate
 dehydrogenase(s), 62
- 6-Hydroxy-D-nicotine, 275, 280
- 6-Hydroxy-D-nicotine oxidase(s), 275
- Hydroxylation, 39, 42, 45, 275
- 6-Hydroxy-L-nicotine, 275, 280
- 6-Hydroxy-L-nicotine oxidase(s)
 (6HLNOs), 275, 281
- Hydroxymethyltransferase(s), 317
- Hydroxy muconic semialdehyde
 dehydrogenase(s), 63
- Hydroxy muconic semialdehyde
 hydrolase(s), 63
- Hydroxynitrocarbazole, 67
- 6-Hydroxy-N-methylmyosmine, 275
- 2-Hydroxypenta-2,4-dienoate, 79
- 6-Hydroxy-pseudo-oxinicotine, 275
- hyp*/Hyp, 249, 250, 264, 277
- I**
- iaa*. *See* *tms*/Tms
- ici*/Ici, 42

- ics/Ics*, 159
 IHF. *See* Integration Host Factor (IHF)
 IL-10, 175
 Immunoblotting analysis, 10
incA cis-acting untranslated RNA, 94
 Inc groups. *See* Incompatibility groups
 Incompatibility
 of catabolic plasmids, 35
 determinants on pNGR234*a*, 421
 determinants on pTi-SAKURA, 142
 fusion of incompatible Ti plasmids, 142
 of haloarchaeal plasmids, 26, 57
 of pSymA and pSymB, 94
 Incompatibility groups
 in the genus *Pseudomonas*, 57
 IncFII plasmids, 155, 161, 162
 IncN plasmids, 62
 IncP-1 plasmids, 35, 39, 43, 47, 57, 62, 81
 IncP-2 plasmids, 35, 43, 45, 66
 IncP-7 plasmids, 35, 57, 65, 66, 72, 73
 IncP-9 plasmids, 35, 57, 58, 60–62, 65, 66,
 73, 81
 IncW plasmids, 62
 of *Shigella* virulence plasmids, 154, 155,
 157
 Ti and Riplasmids, 142
 Injectisomes, 175
 Insertion sequences (IS),
 and *mls* expression in *Mycobacterium*
 ulcerans, 287
 of *A. nicotinovorans* pAO1, 273
 of *Burkholderia*, 41
 of *R. metallidurans* genomic islands,
 223, 224
 of catabolic plasmids, 35, 68
 of *G. westfalica* pKB1, 300, 301
 of haloarchaeal plasmids, 4, 7, 9–11, 18,
 24–26
 of *M. ulcerans* pMUM plasmids, 287,
 291, 315
 of *R. eutropha* H16 pHG1, 255, 257
 of *Rhizobium* sp. NGR234
 pNGR234*a*, 127
 of *Rhodococcus* catabolic plasmids, 68
 of *Shigella* virulence plasmids, 154
 of *S. meliloti*, 95, 99
 of *Y. pestis* pMT1, 181
 Integrase(s), 20, 23, 24, 25, 63, 66
 Integration host factor (IHF), 47
 Integrins, 175
 Interferon- γ , 175
 Interposon mutagenesis, 262
int/Int, 66, 223–225, 229
 Intron-encoded proteins (IEPs), 257
 Inversion isomers, 7–9, 25
inv/Inv, 159, 163
ipa/Ipa, 155, 159, 162
ipb/Ipb, 48
ipg/Ipg, 155, 159
ipt/Ipt, 138
 Iron-sulfur cluster(s)
 in CODH, 262
 in nicotine dehydrogenase, 275
 in the periplasmic nitrate reductase
 of *R. eutropha* H16, 254
 in the *R. eutropha* H16 MBH, 247
 in the *R. eutropha* H16 RH, 251
 in the *R. eutropha* H16 SH, 250
 Iron
 regulation of *Shigella* virulence
 genes, 158
 uptake in *S. meliloti*, 102, 103
 IS100, 181
 IS881, 255, 257
 IS1071, 210, 223–225, 227
 IS1086, 224
 IS1473, 273
 IS1627, 190
 IS2404, 291
 IS2606, 291
 IS3 family, 15
 IS5 family, 24
 IS30 family, 15
 ISH2, 9
 ISH3, 9, 19
 ISH4, 9, 19, 24
 ISH6, 24
 ISH8, 9, 19
 ISH9, 24
 ISH11, 19
 Isopropylbenzene, 48
 3-Isopropylcatechol, 48
 ISPre1, 73
 ISPre4, 73
 IS Rm2011-2, 99
 ISRme3, 224
 ISRme10, 224
 ISRtr1, 123
 Iterons, 80
K
 KDH. *See* ketone dehydrogenase(s)
 (KDHs)
kdh/Kdh, 276, 281
 Ketone dehydrogenase(s) (KDHs), 275,
 278–280
kla/Kla, 71

L

- Lactic acid bacteria (LAB), 312
- Lactobacillus*
- core proteins, 318
 - production of BSHs by strains of, 320
 - taxonomy, 312
- Lactobacillus plantarum*, 314, 317
- Lactobacillus salivarius*
- amino acid metabolism, 317
 - genome organization, 313
 - gluconeogenesis, 319
 - homofermentative metabolism, 318
 - isolation of, 312
 - probiotic properties, 319
- Laminins, 177
- Large variable region (VAR), 137, 141
- aspartate, 101, 220, 317
- LcpK30, 301
- lcr/Lcr*, 174–176
- ldh/Ldh*, 317
- Lead resistance, 210, 221, 228, 229
- Leghemoglobin, 259
- Lethal factor (LF)
- binding to protective antigen, 190
 - encoded by pBC218, 196
 - encoded by pBCXO1, 195
 - function, 196
 - lef* gene, 190
- Leucine-rich repeats, 159
- Leucine zippers, 192
- LF. *See* lethal factor (LF)
- 14-3-3-Like proteins, 138
- Linear plasmids, 48, 284
- Linker scanning mutagenesis, 9, 11
- Lipase(s), 274
- lip/Lip*, 274
- Lipochitooligosaccharides, 105
- Lipopolysaccharide(s), 103, 174
- Lithotrophs, 240, 241, 250
- lon/Lon*, 263
- Low calcium response, 175, 176
- Lower pathway(s), 39, 41, 43, 48, 49, 62
- L-rhamnose isomerase(s), 319
- L-serine dehydratase(s), 317, 318
- lsp/Lsp*, 320
- LytTR DNA-binding domain, 263

M

- mab/Mab*, 277, 280, 281
- MABO. *See* γ -N-methylaminobutyrate oxidase(s) (MABOs)
- Major facilitator superfamily (MFS)
- permeases, 19, 229, 254, 274

- Major intrinsic proteins (MIPs), 194, 196
- Malate, 101
- Malonyl-CoA, 286
- Mandelate, 39
- mao/Mao*, 277, 278, 280
- MAO. *See* monoamine oxidases (MAOs)
- mas1/Mas1*, 138
- mas2/Mas2*, 138
- MCD. *See* molybdenum cytosine dinucleotide (MCD)
- M cells, 152
- mcm/Mcm*, 23
- mcr/Mcr*, 301
- Melamine, 298
- Melibiose, 101
- Membrane-bound hydrogenase(s) (MBHs)
- diatomic ligands of the *R. eutropha* H16 enzyme, 247, 249
 - [Fe(CN⁻)₂CO] complex of the *R. eutropha* H16 enzyme, 248, 249
 - gas channels in the *R. eutropha* H16 enzyme, 247
 - iron-sulfur clusters of the *R. eutropha* H16 enzyme, 247
 - maturation and assembly of the *R. eutropha* H16 enzyme, 248, 249
 - [NiFe] center of the *R. eutropha* H16 enzyme, 247, 249
 - of *Oligotropha carboxidovorans* OM5, 259, 264
- Merbromin, 225
- Mercury chloride, 225
- Mercury resistance
- determined by Tn4378 and Tn4380, 225
 - pJJB1, 47
 - pMOL28, 225
 - of Tn5046, 65
- mer/Mer*, 13, 15, 17, 19, 21, 47
- Mesorhizobium loti*, 99
- Meta-cleavage pathway(s)
- NAH7 plasmid, 63
 - pCAR3, 79
 - pEST1026, 39
 - pVII50, 45
 - pWW0, 74
- Methionine, 97, 161, 317
- Methylamine, 278, 280
- Methylase(s), 18
- Methylenetetrahydrofolate, 21, 277
- Methylene-tetrahydrofolate dehydrogenase/cyclohydrolase(s), 277
- 5, 10-Methylenetetrahydrofolate reductase(s), 21
- 4-Methylmuconolactone, 255
- Methylmuconolactone isomerase(s), 255

- Methylmuconolactone methyl isomerase(s), 255
- 3-Methylpyridine, 298
- 2-Methylquinoline, 48
- Methyl t-butyl ether (MTBE), 298
- met*/*Met*, 161
- MFS permeases. *See* Major facilitator superfamily (MFS) permeases
- MHYT domain, 262, 263
- mia*/*Mia*, 158
- Microarray analysis, 73, 228, 232
- min*/*Min*, 96
- MIP proteins. *See* Major intrinsic proteins (MIPs)
- mis*/*Mis*, 138
- Mixotrophy, 241
- mls*/*Mls*, 286–289, 290, 292
- mml*/*Mml*, 255
- mmr*/*Mmr*, 224, 229, 232
- moa*/*Moa*, 276, 278
- mob*/*Mob*, 95, 279, 281
- mod*/*Mod*, 279
- moel*/*Moe*, 279
- Molybdate uptake, 279
- Molybdenum cofactor, 254, 273, 275, 278, 279
- Molybdenum cofactor biosynthesis, 254, 273, 275, 278, 279
- Molybdoenzyme(s), 254, 263
- Molybdoproteins, 261, 262
- Molybdopterin (MPT), 254, 261, 263, 279
- Molybdopterin-cytosin dinucleotide (MCD), 261, 262, 279
- Molybdopterin guanine dinucleotide cofactor (MGD), 254
- Monoamine oxidase(s) (MAOs), 277, 278
- Monoxygenase(s), 39, 42, 275
- moq*/*Moq*, 49
- Mosaic elements (MS), 122
- mox*/*Mox*, 275
- mpf*/*Mpf*, 61, 62
- MPT. *See* Molybdopterin (MPT)
- mrr*/*Mrr*, 22, 24
- mrs*/*Mrs*, 62
- Muconate cycloisomerase(s), 43, 255
- Muconolactone isomerase(s), 255
- Multimer resolvase(s), 25, 243
- MvpT/MvpA toxin/antidote system, 155, 162, 163
- mxi*/*Mxi*, 155, 159, 161
- Mycobacterium*
- degradation of PAHs, 57
 - degradation of rubber, 299
 - Mycobacterium abscessus*, 284
 - Mycobacterium avium*, 284
 - Mycobacterium avium-intracellulare*, 284
 - Mycobacterium branderi*, 284
 - Mycobacterium celatum*, 284
 - Mycobacterium chelonae*, 284
 - Mycobacterium fortuitum*, 284, 292
 - Mycobacterium marinum*, 287, 288, 289, 292
 - Mycobacterium peregrinum*, 284
 - Mycobacterium scrofulaceum*, 284
 - Mycobacterium smegmatis*, 287, 292, 301, 304, 305
 - Mycobacterium ulcerans*
 - distribution of plasmids in, 288
 - genome reduction, 292
 - infection, 284
 - insect hosts, 293
 - insertion sequences (IS), 291
 - transposons, 288
 - in fish hosts, 288
 - Mycobacterium vanbaalenii*, 57, 58
 - Mycobacterium xenopi*, 284
- Mycolactone(s)
- biosynthesis of, 287, 288
 - of *Mycobacterium ulcerans* subsp. *liflandii*, 288
 - properties, 285
 - structure, 285
 - types, 285
- Mycolic acid(s), 48
- Myosin-like repeats, 192
- N**
- N-acetyl-glucosamine, 107, 180
- N-acetylmuramoyl-L-alanine amidase(s), 199
- N-acyl-homoserine lactone, 121, 143
- N-acyltransferase(s), 107, 303
- NAD(P)H:ferredoxin oxidoreductase(s), 79
- NAD(P)H-nicotine blue oxidoreductase(s), 275, 277
- NADP⁺-dependent succinic semialdehyde dehydrogenase(s), 277
- nah*/*Nah*, 41, 60, 61, 62, 63, 66
- NAH7 plasmid, 58
- backbone, 61
 - evolution of, 66
 - incompatibility, 58
 - nah* genes, 62
 - organization, 60
 - origin of, 65
 - Tn4655, 63
- Naphthalene, 35, 39, 41, 57, 58, 60, 62, 63, 65, 66, 71
- Naphthalene *cis*-dihydrodiol dehydrogenase(s), 62
- Naphthalene dioxygenase(s), 62
- Naphthalene 1,2-dioxygenase ferredoxin, 62
- Naphthalene 1,2-dioxygenase reductase(s), 62

- Naphthalene sulfonate, 66
nap/*Nap*, 97, 109, 254
Natronomonas pharaonis, 4, 6
ncc/*Ncc*, 6, 14, 18, 211, 212, 224, 229
ndh/*Ndh*, 275, 278, 279, 281
 NDH. *See* Nicotine dehydrogenase(s) (NDHs)
nep/*Nep*, 277, 279, 280
 N-(2-hexylamino-4-phenylimidazol-1-yl)-acetamide, 298
 Ni-chelating chromatography, 279
nic/*Nic*, 272, 276, 281, 304
 Nicotine
 availability in natural habitats, 271, 272
 catabolism, 275, 276, 277, 278
 extrusion of catabolic end products, 280
 regulation of genes for catabolism, 280, 281
 uptake, 279
 Nicotine blue, 275
 Nicotine dehydrogenase(s) (NDHs), 275, 280
 [NiFe] center(s)
 of the *O. carboxidovorans* OM5 MBH, 264
 of the *R. eutropha* H16 MBH, 247, 249
 of the *R. eutropha* H16 RH, 251
 of the *R. eutropha* H16 SH, 250
nif/*Nif*, 98, 99, 120, 122–124, 144, 254
nir/*Nir*, 109, 254
nit/*Nit*, 275
 Nitrate reductase(s), 19, 97, 109, 253, 254, 257
 Nitric oxide reductase(s), 97, 254
 Nitrilase(s), 275
 Nitrite reductase(s), 97, 254, 257
 Nitrogenase(s), 100, 105, 108, 125, 127
 Nitrogen fixation
 in *Rhizobium* sp. NGR234, 120, 122, 124, 125, 127, 134
 in *S. meliloti*, 92, 96, 99–101, 103, 105, 107–110
 Nitrous oxide reductase(s), 21, 22
 N-methylaminopropyl-(6-hydroxypyridyl-3)-ketone, 275
 N-methylhydantoinase(s), 24
nnr/*Nnr*, 97, 107, 109
Nocardia, 298, 299
noc/*Noc*, 142
nod boxes, 107, 125, 127
 Nod factors, 94, 104, 105, 107, 120, 124, 125
nod/*Nod*, 98, 100, 101, 104, 105, 107, 109, 120, 122, 124, 125, 127
 Nodulation
 Rhizobium sp. NGR234, 120, 122, 124, 125
 S. meliloti, 92, 96, 98–101, 104, 105, 107, 109
noe/*Noe*, 98, 124
nol/*Nol*, 98, 124, 125
 Nopaline synthase(s), 138
nop/*Nop*, 126, 127
 N₂O reductase(s), 109, 257
 NO reductase(s), 109, 257
nor/*Nor*, 97, 109, 254
nos/*Nos*, 21, 97, 109, 138, 254
Novosphingobium, 29, 56, 57, 70
Novosphingobium aromaticivorans, 41, 65, 66
nox/*Nox*, 142
nrd/*Nrd*, 254, 257
nre/*Nre*, 211, 224, 229
nrt/*Nrt*, 97
 NtSIP1, 138
 Nuclear localization signals (NLSs), 140
- O**
 O-antigen, 174
ocs/*Ocs*, 138
 Octane, 43
 Octopine synthase(s), 138
 OCT plasmid, 43
Oligotropha carboxidovorans OM5, 264
omp/*Omp*, 141, 142, 157, 179, 229
 Opines, 135–138, 141–143
orc/*Orc*, 12, 15–17, 19–25
oriM, 292
oriR, 162
oriT (*mob*) sites
 pKB1, 304
 pNGR234a, 241
 pSymA, 94
 pSymB, 94
 Ti and Ri plasmids, 143
oriV
 pAW63, 26
 pCAR1, 71, 72
 pCAR3, 80
 pHCG3
 pKB1, 300
 pMT2, 62
 pMT118, 313
 pNGR234a, 121
 pNRC100, 11
 pNRC200, 11
 pPCP1, 181
 pSymB, 95
 pXO1, 191–193
 pXO2, 201, 202
 Ortho-cleavage pathway(s), 39, 41–43, 49, 67, 255
ots/*Ots*, 96
 Overdrive sequence, 139, 140
 4-Oxalocrotonate decarboxylase(s), 63
 Oxalocrotonate tautomerase(s), 63

Oxidase(s), 4, 8, 12, 109, 264, 275–278, 280, 298, 302
 3-Oxoacid CoA-transferase(s), 24
 3-Oxoadipate enol-lactone hydrolase(s), 255
 2-Oxo-4-hydroxypentanoate aldolase(s), 63
 2-Oxopent-4-enoate hydratase(s), 63

P

PA. *See* Protective antigen (PA)

paal/Paa, 42

pAC27

clcABD operon, 43

pAD1, 199, 201

pagA/PagA, 190, 191

PAGI2-C island, 223

PAHs. *See* Polycyclic aromatic hydrocarbons (PAHs)

pAL1

degradation of anthranilate, 49

pAL5000, 284

pAMβ1, 199, 201

pAOI

carbohydrate catabolism genes, 274

conjugative transfer, 273

γ-N-methylaminobutyrate catabolism genes, 277, 278

hexose oxidation genes, 274

molybdenum cofactor synthesis genes, 273, 275, 278, 279

nep genes, 277, 279, 280

nicotene catabolism genes, 275–277

nicotine uptake, 279

nic regulon regulators, 280, 281

organization, 272

relatives, 273

PAPS. *See* 3'-phosphadenosine-5'-phosphosulphate (PAPS)

Paracoccus pantotrophus, 240

pari/Par

mycobacterial plasmids, 292

pCAR1, 71, 72, 75–77

pCAR3, 80

pHV1, 19

pHV4, 17

pL47, 25

pL131, 24

pMOL28, 219, 220

pMOL30, 222

pMT2, 62

pMP118, 313, 314

pNG400, 23

pNRC100, 12

Shigella virulence plasmids, 163

Partitioning

mycobacterial plasmids, 292

pAO1, 273

pCAR1, 75

pCAR3, 80

pHCG3, 259

pHG1, 243

pHV1, 19

pHV3, 19

pHV4, 17

pKB1, 300, 301

pL47, 25

pMP118, 313

pMT2, 62

pNG200, 15, 23

pNG300, 23

pNG400, 23

pNG600, 22

pNRC100, 11, 12

pSymA, 93

pSymB, 93

Shigella virulence plasmids, 159, 163

pAtC58, 142, 143

Pathogenicity islands (PAIs), 190, 191, 194, 195, 198, 202

P1-ATPase-mediated efflux systems, 229, 231, 232

pAW63, 197, 201

pBBR1MCS-2, 300

pBBR1MCS-5, 300, 305

pBC218, 196, 197

pBc10987, 193–195, 197

pBCXO1, 193, 195, 197

pBD2, 48

pBN6, 66

pbr/Pbr, 210, 212, 223–225, 228, 229, 232

pBT9727, 197, 201, 202

pBVIE01, 42, 221, 222

pBVIE02, 42, 219, 220, 230

pBVIE04, 42

pca/Pca, 42, 73, 102

pCAR1

conjugative transfer, 72, 73, 75

evolutionary relationships, 71, 72

incompatibility, 72

mobile genetic elements, 72, 73

plasmid-chromosome interactions, 73–75

oriV, 72, 74

partitioning, 72, 74–76

regulation of anthranilate degradation genes, 73, 74

replication, 72

tra/Tra, 72

pCAR2, 73

- pCAR3, 66, 70, 77–80
 PCBs. *See* Polychlorinated biphenyl(s) (PCBs)
 pCD1. *See* *pYV*
 pCER270, 193, 196–197
 pCF10, 201
pch/Pch, 41
 pCLP, 284
 pColE1, 181
 pDBMCS-2, 301
 pDBMCS-5, 301
 pDK1, 39
 pDTG1, 61, 63, 65, 66
 p42e, 96
pem/Pem, 243
 penicillin(s), 22
 penicillin acylase(s), 22
 pentose phosphate pathway, 251, 317–319
 peptidyl-prolyl cis-trans isomerase(s), 21
 Periplasmic dissimilatory nitrate reductase(s), 97
 pEST1026, 39
 pEST4011, 43, 47
 pesticin, 179, 180, 181
 pExo, 92
 Peyer's patches, 173
pgsB/PgsB, 198
pha/Pha, 41
 PHB. *See* polyhydroxybutyrate (PHB)
 pHCG3
 cbb genes, 263
 cox accessory genes, 262, 263
 cox genes, 259, 261
 cycH, 263
 gene for a *b*-type cytochrome, 264
 hox genes, 264
 hyp genes, 264
 lon, 263
 megaplasmid-chromosome interactions, 264
 repB-repC intergenic region, 259
 rep genes, 259
 tra/trib genes, 259
 pHCM2, 181
 Phenanthrene, 57, 58, 63
 Phenol(s), 35, 39, 42, 45
 Phenol hydroxylase(s) (PHs), 39
 Phenylacetyl-CoA, 42
 Phenylalkanoate(s), 41
phe/Phe, 39
 pHG1
 cbb genes, 251, 252
 conjugative transfer, 243
 gene for a *b*-type cytochrome, 247
 gene for a MBH-specific C-terminal endopeptidase, 249
 gene for a SH-specific C-terminal endopeptidase, 250
 genes for degradation of aromatics, 255
 genes *hoxA*, *hoxB*, *hoxC* and *hoxJ*, 250, 251
 hyp genes, 249
 IS881, 255, 257
 junkyard region, 255, 257
 MBH gene cluster, 246, 247
 megaplasmid-chromosome interactions, 257, 258
 mobB, 254
 nap genes, 254
 nar genes, 253, 254
 nifM gene, 254
 nor genes, 254
 nos genes, 254
 organization of the *hox* regulon, 243, 248
 parA and *parB* genes, 242
 pemI and *pemK* genes, 242
 repA, 242
 SH gene cluster, 249, 250
 similarity to pMOL28, 242
 similarity to pSym of *Cupriavidus taiwanensis*, 242
 xerC and *xerD* genes, 242
phl/Phl, 39
 pHMT112, 41
pho/Pho, 102
 Phosphoenol pyruvate (PEP), 253
 3'-Phosphoadenosine-5'-phosphosulphate (PAPS), 107, 120
 Phosphatase(s), 176, 191, 252, 263, 319
 Phosphate uptake, 102
 Phosphoglucomutase(s), 103
 Phosphoglycerate kinase(s), 252, 263
 Phosphoglycolate phosphatase(s), 252, 263
 Phospholipase D (PLD), 177, 178
 Phosphorelay signal transduction systems
 B. anthracis, 191
 C. metallidurans, 231
 L. salivarius, 319
 Shigella, 157
 R. eutropha H16, 250, 251
 S. meliloti, 101, 108, 109
 5-Phosphoribosyl-1-pyrophosphate (PRPP), 317
 Phosphoribulokinase(s), 252, 263
 Phosphotransferase system (PTS) transporters, 319
 Phosphotyrosine phosphatase(s) 176
 pH stress, 24
 Phthalate(s), 41, 48, 79
pht/Pht, 48
 pHV1, 5, 16, 19, 20

- pHV2, 6, 11, 16
 pHV3, 5, 16, 18, 19
 pHV4, 5, 16–20, 23
 pIB1. *See* pYV
 pIJB1, 43, 47
pim/Pim, 180
 PIN domain, 143
 pINV A, 154, 155
 pINV B, 154, 155
 pINV-2457T, 154
 pIP501, 199, 201
 pJP4, 43, 47
 pKB1
 analysis of gene expression by
 RT-PCR, 303
 cadmium resistance, 303
 conjugation frequency, 304
 conjugative transfer, 304
 construction of vectors from, 300, 301
 curing of, 301
 derivatives, 305
 genes for c-type cytochromes, 302
 genes for rubber degradation, 302
 identification by pulsed field gel
 electrophoresis (PFGE), 299
 oriV, 300
 tra genes, 304
 transformation frequency, 304
 pKB2, 306
pkc/Pkc, 275
pkn/Pkn, 291
 PKSs. *See* Polyketide synthase(s)
 (PKSs)
 pL23, 6, 24, 25
 pL47, 6, 25
 pL131, 6, 24
pla/Pla, 179–181
 Plasminogen activator, 173, 179, 181
 pLCR. *See* pYV
plcR/PlcR, 195
 PLD. *See* phospholipase D (PLD)
 pM3, 62
pmf/Pmf, 280
 pMOL28
 backbone, 220
 bimA1B1, 227
 chr genes, 211, 223, 227
 cnr genes, 210, 227
 conjugative transfer, 218, 225
 evolution of metal resistance genes, 233
 induction of *cnr* genes, 233
 IS elements, 223
 par genes, 220
 pil genes, 220
 similarity to pHG1, pBVIEO2 and
 pRALTA, 219, 220
 Tn*Cme2* insertion, 223, 225, 227
 trb genes, 220
 pMOL30
 backbone, 221
 bimA2B2, 227
 conjugative transfer, 225
 cop genes, 224, 229, 230
 czc genes, 223, 228
 evolution of metal resistance genes,
 233, 234
 induction of heavy metal resistance
 genes, 231–233
 IS elements, 224
 ncc/nre genes, 211, 224, 229
 pbr genes, 223, 224, 228
 peculiarity of the *copB* gene, 229
 similarity to pBVIE01, 221
 Tn*Cme2* insertion, 225
 pMOL50, 224
 pMP118
 bacteriocin genes, 320, 321
 bile salt hydrolase gene, 319, 320
 copy number, 314
 distribution of pMP118-like plasmids
 in *Lactobacillus* strains, 321
 GC content, 314
 genes for core proteins, 318
 genes for sorbitol and rhamnose utilization, 319
 genes for surface proteins, 320
 genes of amino acid metabolism, 316, 317
 genes of carbohydrate metabolism, 318
 genes of purine and pyrimidine anabolism, 317
 identification of, 313
 IS elements, 315
 ori, 313
 pseudogenes, 315
 repA and *repE*, 313, 314
 tra genes, 321, 322
 GC content, 314
pmr/Pmr, 74–77
 pMT1, 177, 178, 181
 pMT2, 62
 pMUM001
 backbone, 292
 mycolactone locus, 286–288
 non PKS-coding regions, 290, 291
 pMUM002
 backbone, 292
 genes for serine/threonine protein kinase(s)
 (STPKs), 291
 mycolactone locus, 286–288
 non PKS-coding regions, 290, 291

- pMUM003
 backbone, 292
 genes for serine/threonine protein kinase(s) (STPKs), 291
 non PKS-coding regions, 290, 291
- pNC903, 301
- pND6-1, 65, 66, 71, 72
- Pneumonic plague, 173, 180
- pNG100, 20, 23
- pNG200, 20, 23
- pNG300, 20, 23
- pNG400, 18, 20, 23
- pNG500
orc genes, 22
- pNG600, 22, 23
- pNG700, 21
- pNGR234a
 curing of, 120
 G+C content, 121, 122
 genes for a type III secretion system, 126
nif and *fix* genes, 124, 125
nod, *nol* and *noe* genes, 124, 125
oriT, 121
oriV, 121
 rearrangements, 123
 repABC, 121
 replicon fusion, 124
tra genes, 121
tsI, 126
- pNGR234b, 120, 125
- pNL1
bph, *xyl*, *pch*, 41, 80
 incompatibility, 80
nah genes, 66
rep genes, 80
- pNL2, 41
- pNod-Nif, 92
- pNRC100
arsADRC and *arsR2M*, 15
 arsenic-resistance genes, 12
 A+T-rich segment, 11
cydA and *cydB*, 8, 12
 evolution of, 25, 26
gvp genes, 9, 10
 housekeeping genes, 12
hilD, 12
 inverted repeat (IR) sequence, 11, 25
 mobile genetic elements, 4
orc genes, 12
oriV, 10, 11
par genes, 12
repH, 11
ibp genes, 12
trxAB, 8
- pNRC200
arcRABC gene cluster, 15
argS, 15, 25, 26
cydA and *cydB*, 12
ftsZ5, 15
gvp genes, 15
 housekeeping genes, 12
hilD, 12
 inverted repeat (IR) sequence, 11, 12
orc genes, 15
oriV, 11
polB2, 15
repH, 12
ibp genes, 12, 15, 16
trxA and *trxB*, 12
- pnt*/*Pnt*, 274
- pol*/*Pol*, 15, 17, 22
- Polychlorinated biphenyl(s) (PCBs), 41, 48
- Polycyclic aromatic hydrocarbons (PAHs), 48, 57, 58, 65, 80
- Poly- γ -D-glutamate, 188, 198
- Polyhydroxybutyrate (PHB), 101
- Polyketide(s), 285, 286
- Polyketide cyclase(s), 275
- Polyketide synthase(s) (PKSs), 285–287
- Polymorphonuclear leukocytes, 152, 176
- PONH. *See* 2,6-Dihydroxy-pseudo-oxynicotine hydrolase(s) (PONHs)
- Porins, 229
- pP51, 41, 43
- pPCP1, 179, 181
- pPER272, 193, 196–197
- pPGH1, 39
- (p)ppGpp, 45
- pRA500, 41
- pRALTA, 219, 220
- pRE1, 48
- pREA250, 304
- pREA400, 304
- Precursor Z, 278
- Prenyl transferase(s), 141
- pRHL1, 48
- pRHL2, 48
- pRHL3, 48
- pRi1724, 135, 137, 138, 141
- pRi2659, 135, 141
- pRmeSU47a, 92
- pRmeSU47b, 92
- Probiotics, 312, 313, 320, 322
- Propane, 299

- Protective antigen (PA)
 encoded by pBC218, 196
 encoded by pBCXO1, 195
 encoded by pXO1, 188, 189
pagA gene, 190
 proteolytic processing, 190
 secretion of, 190
- Protein phosphorylation, 291
- Proteolytic processing
 of *B. anthracis* protective antigen, 190
 of *R. eutropha* H16 hydrogenase large subunits, 249, 250
 of *Yersinia* pre-Pla protein, 179
- Proteome
 lactobacilli core proteins, 318
- Protocatechuate(s), 39, 41, 42, 48, 58, 79, 102
- prSA/PrsA*, 190
- prt/Prt*, 320
- pSD25, 259
- Pseudomonas resinovorans*, 67, 70, 72, 73
- Pseudoappendicitis, 181
- Pseudogenes, 48, 177, 291, 313, 315, 320
- Pseudomonas* sp. CF600, 45
- Pseudomonas*
 carbazole degradation, 67
car genes, 67, 70
 CO-oxidizing strains, 242
 degradation of aromatics, 39
 degradation of benzene, 41, 43
 degradation of naphthalene, 58, 65, 66, 71
 degradation of PAHs, 57
 incompatibility groups, 57
 IncP-9 plasmids, 58, 61
mvt genes, 75
- Pseudomonas carboxydohydrogena*, 242
- Pseudomonas chlororaphis*, 73
- Pseudomonas fluorescens*, 73
- Pseudomonas putida*
 degradation of naphthalene, 58
 degradation of toluene, 58
 transcriptome, 73–75
 degradation of benzene, 41
 degradation of carbazole, 73
 degradation of 3-chlorobenzoate, 42
 degradation of p-cresol, 41
 degradation of phenols, 39
- Pseudomonas resinovorans*, 67, 70, 72, 73
- Pseudomonas* sp. ND6, 71
- Pseudomonas* sp. strain P51, 43
- Pseudomonas stutzeri*, 73
- Pseudomonas thermocarboxidovorans*, 242
- pSF118-20, 313, 314
- pSF118-44, 313, 314
- pSmeSM11a, 92
- pSmeSM11b, 92
- pSOX, 300
- pst/Pst*, 180
- pSymA, 66
 curing of, 98, 99
 DnaA binding sites, 94
 GC content, 92
 gene expression in bacteroids, 98
 genes for betaine catabolism, 102
 genes for nitrogen fixation, 98
 genes for stress responses, 98
 host-specificity genes, 105
 housekeeping genes, 98
inc α 2, 94
incY, 94
 mobile genetic elements, 99
mob sites, 94
napEFDABC cluster, 109
nirKV, 109
nnrR and *nnrS* genes, 97, 109
nod genes, 98, 104, 105, 107
nol and *noe* genes, 98
nor genes, 97, 109
nos genes, 97
parAB genes, 93
 plant growth inhibitory genes, 98
repABC genes, 93, 94
rhbABCDEF operon, 102, 103
sel genes, 98
 tRNA genes, 97
- pSymB
araABCDEF cluster, 101
 autonomously replicating sequence (ARS), 94
 chromosomal and plasmid features, 95
 coding capacity, 95
dct genes, 101
 dinucleotide signature, 95
 DnaA binding sites, 94
ehu and *eut* genes for ectoine catabolism, 102
 evolutionary history of, 95
exo genes for succinoglycan synthesis, 103, 104
exp genes for galactoglucan synthesis, 104
ftsK2, 96
 GC content, 92, 95
 gene expression, 96
 genes expressed under salt stress, 96, 97
 genes for catabolism of short-chain fatty acids, 101
 genes for PHB degradation, 101, 102
inc α 1, 94
minCDE genes, 95, 96
mob sites, 94

- pSymB (*cont.*)
parAB genes, 93
pcaDCHGB and *pcaIJF* operons, 102
pho genes for phosphate uptake, 102
repABC genes, 93, 94
 synthesis of polysaccharides, 97
 thiamine biosynthesis genes, 101
- pTB1, 259
- pTiBo542, 135, 141, 142
- pTiC58, 135, 141, 142
- pTiS4, 142
- pTi-SAKURA, 135, 137, 141–143
- pTOM, 42
- pTOM9, 211
- pTV1, 43
- Pulse-field gel electrophoresis (PFGE), 8,
 120, 313
- pur*/*Pur*, 277, 279, 280
- Putrescine, 24
- pVI150, 39, 45, 47
- pWR100, 154, 163
- pWR501, 154
- pWW0, 39, 41, 43, 58, 61, 65,
 72, 74
- pWW53, 39, 71
- pXO1, 188–193
abRB, 191, 194
atxA, 190, 191
 backbone, 189
 copy number, 197
ger genes, 191
 organization, 198
oriV, 191, 192
 Rap/Phr phosphatase, 191
 relatives, 193–197
repX, 191, 192
 SLH proteins, 102
 sporulation, 191
topX, 192
 toxin genes, 190
 type IV secretion system, 192
- pXO2
 AcpA and AcpB, 198, 199
 autolysin, 199
atxA gene, 190
 backbone, 197, 198
capBCADE, 198–199
 capsul synthesis genes, 198
 conjugate transfer, 199, 201
 copy number, 197
 PAI, 198
 PgsB, 198
 relatives, 201
 replication, 199
repS, 199
 topoisomerase gene, 199
tra genes, 201
 type IV secretion system, 199
- pXO14, 199
- pYpsiP31758.1, 182
- pYpsiP31758.2, 182
- Pyrene, 57
- Pyruvate, 39, 45, 62, 253, 280, 317
- Pyruvate kinase(s), 318
- pYV
 distribution in *Yersinia* strains, 173
yadA, 176, 177
ysc, 175
- Q**
qox/*Qox*, 49
- R**
 Raffinose, 101
Ralstonia eutropha H16
 methylmuconolactone utilization, 255
 organoautotrophic growth on formate, 252
 synthesis of deoxynucleotides under
 anoxic conditions, 254, 255, 257
Ralstonia eutropha JMP134, 43, 47
Ralstonia oxalatica, 47
Ralstonia pickettii
 Rapamycin, 286
 Rap/Phr system, 191
 Recombinase(s), 122, 220, 223–225, 227, 229,
 233, 234, 255
 Recombination hot spot (RHS) elements, 220
 Reductive pentose phosphate pathway
 in *R. eutropha* H16, 251
 Regulators
 AbpK/AbpR, 319
 AbrB, 191, 194, 195
 Acp, 198, 199
 of anthranilate degradation genes of
 pCAR1, 73, 74
 AraC/XylS family, 74, 102, 157–159
 AtxA, 190, 191, 195, 196, 199
B. anthracis Rap/Phr system, 191
 CafIR, 177, 178
 of capsule synthesis in *B. anthracis*, 198, 199
 of carbohydrate metabolism in
A. nicotinovorans, 274
 CbbR of *O. carboxidovorans* OM5, 263
 CbbR of *R. eutropha* H16, 253, 258
 CopR/CopS two-component system,
 229, 231

- CpxA/CpxR two-component system of *Shigella*, 157
- Crp/Fnr family, 109
- CtrA, 96
- DctB/DctD of *S. meliloti*, 101, 101
- DmpR, 45, 47
- EnvZ/OmpR two-component system of *Shigella*, 157
- ExpG, 104
- Fis, 158
- FixK, 108, 109
- FixL/FixJ, 108, 109
- Fur, 158
- GntR family, 274
- H-NS, 74, 156–159, 161, 163
- HoxA/HoxJ, 250, 251, 264
- IclR family, 102
- LacI family, 274
- LuxR-type, 281
- LysR family, 63, 102, 105, 106, 124, 252, 263
- MarR family, 18, 104
- MpfR, 62
- MvaT family, 74, 75, 77
- MxiE, 155, 159, 161
- NahR, 63
- NarX/NarL of *R. eutropha* H16, 254
- of nicotine catabolism in *A. nicotinovorans*, 280, 281
- NifA of *S. meliloti*, 108, 109
- NifA of *Rhizobium* sp. NGR234, 125, 127
- NikR family, 264
- NnrR/NnrS, 97
- NodD1 of *Rhizobium* sp. NGR234, 124, 125
- NodD2 of *Rhizobium* sp. NGR234, 124
- NodD1 of *S. meliloti*, 105, 107
- NodD2 of *S. meliloti*, 105
- NodD3 of *S. meliloti*, 101, 105, 107
- NorR1, 254
- NorR2, 254
- NtrC/XyIR family, 45, 74, 250, 264, 280
- PbrR, 229
- PcaQ, 102
- PcaR, 102
- PmfR, 280
- Ptr, 74, 75, 77
- quorum-sensing regulators of *Rhizobium* sp. NGR234, 121
- quorum-sensing regulators of Ti/Ri plasmids, 143
- repressor RepB of *Shigella*, 162
- response regulator Spo0F of *B. anthracis*, 191
- RhrA, 102
- of secreted protein genes in *Shigella*, 159
- Spo0A of *H. marismortui*, 23
- SyrA of *S. meliloti*, 101, 109
- SyrM of *S. meliloti*, 101, 105, 107
- SyrM1 of *Rhizobium* sp. NGR234, 124
- SyrM2 of *Rhizobium* sp. NGR234, 124
- TetR family, 280
- TfdR, 43
- of toxin gene expression in *B. anthracis*, 190, 191
- transcriptional activator VirF of *Shigella*, 157–159, 161
- TtsI, 126
- of virulence factor production in *B. cereus*, 195, 196
- of virulence genes in *Shigella*, 157
- of virulence plasmid copy number in *Shigella*, 162
- VirB of *Shigella*, 159
- Xre family, 74
- Regulatory hydrogenase(s) (RHs), 251
- R.e.II*, 257
- Replication
- of *H. volcanii* plasmids, 16, 19, 20
- iterons, 80
- of mycobacterial plasmids, 292
- NAH7, 61
- origin recognition complex, 12, 15, 19–21, 23
- pAO1, 273
- pAW63, 201
- pBT9727, 202
- pCAR1, 71, 72
- pCAR3, 80
- pHCG3, 259
- pHG1, 243
- pIJB1, 47
- pKB1, 300, 301
- pMOL28, 219, 220
- pMOL30, 222
- pMP118, 313, 314
- pMT2, 62
- pNG300, 23
- pNG400, 23
- pNG600, 22
- pNG700, 20
- pNGR234a, 121, 127
- pNL1, 80
- pNRC100, 10–12, 15
- pPCP1, 181
- pSymA, 93
- pSymB, 93, 94
- pXO1, 191
- pXO1-like plasmids, 193, 196

- Replication (*cont.*)
 pXO2, 199
 pXO2, 199
repABC-type replicons, 80, 93–95, 121, 142, 259
 role of DnaA in replication of pSymA and pSymB, 94
 of *Shigella* virulence plasmids, 161, 162
 of Ti plasmids, 136, 142
rep/Rep. *See* Replication
- Resistance nodulation division (RND) efflux systems, 210, 211, 220, 223, 224, 227–231
- Resistance to nickel, cobalt and cadmium, 211, 212, 224, 239
- Resolvase(s), 25, 243
- Restriction/modification systems
 mycobacterial plasmids, 284
 pHV3, 19
 pHV4, 18
 pL47, 25
 pL131, 24
 pMI18, 315
 pNG600, 22
 pNRC100, 12
 pNRC200, 15
- Rhamnose, 319
- Rhamnulokinase(s), 319
- Rhamnulose-1-phosphate aldolase(s), 319
- rhb/Rhb*, 102
- rhc/Rhc*, 126
- Rhizobactin 102, 103
- Rhizobiaceae*
 genomic rearrangements in, 123
 mobile genetic elements, 99
- Rhizobium* sp. NGR234, 120, 125
- Rhizobium etli*, 96, 121, 123
- Rhizobium tropici*, 123, 124
- Rhodobacter capsulatus*, 241, 263, 279
- Rhodococcus*
 degradation of aromatics, 48
 degradation of PAHs, 57
 degradation of rubber, 303
 linear plasmids, 35
 mixotrophy, 34
- Rhodococcus erythropolis*, 48, 291, 304
- Rhodococcus opacus*, 240
- Ribose-5-phosphate isomerase(s), 318
- Ribulose-1,5-bisphosphate carboxylase/oxygenase(s), 252, 263
- Ribulose-phosphate 3-epimerase(s), 252, 263, 318
- Rieske non-heme iron oxygenase(s), 58
- Ri plasmids
 classification, 135
 evolutionary relationships, 137
 GALLS gene, 141
 incompatibility, 142
 genes for opine utilization, 142
 opine type, 135
 organization, 136
rol genes, 138
 TSS, 140
 virulence genes, 139–141
- RmInt1, 99
- RNaseH, 18, 21
- rol/Rol*, 138
- Roseobacter denitrificans* OCh 114, 259
- ros/Ros*, 138, 139
- rox/Rox*, 301
- R1 plasmid, 162
- R100 plasmid, 161–163
- rpoD/RpoD*, 96
- rpoN/RpoN*, 73, 74, 125
- RT-PCR, 72, 79, 232, 257
- Ruegeria* sp. PR1b, 259
- ryh/Ryh*, 158
- S**
- Salicylaldehyde dehydrogenase(s), 62
- Salicylate, 39, 41, 58, 62
- Salicylate hydroxylase(s), 63
- Salmonella enterica*, 157, 161, 181
- SAL plasmid, 58
- Secretion
 of *B. anthracis* toxins, 190, 192, 193
 ESAT-6/WX100 proteins, 190
 of ExpE1 in *S. meliloti*, 104
 of galactoglucan in *S. meliloti*, 104
 of Nops by *Rhizobium* sp. NGR234, 126
 in *L. salivarius*, 320
 of succinoglycan in *S. meliloti*, 104
 type II secretion systems of
C. metallidurans, 225
 type III secretion system of *Rhizobium* sp. NGR234, 126, 127
 type III secretion systems of *Shigella*, 152, 153, 155, 159
 type III secretion systems of *Yersinia*, 174–176, 178, 182
 type IVB secretion systems of
Y. pseudotuberculosis, 182
 type IV secretion systems of *B. anthracis*, 192, 193, 199
- Sedoheptulose-1,7-bisphosphatase(s), 252

- Sedoheptulose-1,7-bisphosphate
aldolase(s), 252
- Selenocysteine synthase(s), 98
- sel/Sel*, 98
- Serine, 317
- Serine/threonine protein kinase(s) (STPKs), 291
- Shigella*
invasion strategy, 152
relationship to EIEC, 152
sub-groups, 152
subversion of host immune response, 153
- Shigella boydii*, 152
- Shigella dysenteriae*, 152
- Shigella flexneri*, 152
- Shigella sonnei*, 152
- Shigella* virulence plasmids
addiction systems, 162
conjugative transfer, 162
control of gene expression, 155
control of *virF* expression by Fis, 158
control of virulence factor expression by
iron, 158
entry region, 155, 157
H-NS and gene regulation, 156–158
incompatibility, 155, 162
initiation protein RepA, 162
instability, 161
integration into host chromosome, 161
interactions between plasmid and
chromosomal genes, 159
mobile genetic elements, 154
mxi and *spa* genes, 155, 159
oriR, 162
partitioning, 163
physiological conditions for expression of
virulence factors, 157
regulation of genes by VirF and VirB, 157
regulation of *virB* by VirF, 159
replication, 161, 162
tra genes, 162
transcriptional slippage, 161
type III secretion system, 152, 153, 155, 159
virB expression and DNA supercoiling, 159
- Shigellosis, 152
- Siderophores
S. meliloti, 102
Yersinia, 180
- sigA/SigA*, 287
- Sigma factors
sigma70_{ECP}, 17, 227
sigma54, 125
- sil/Sil*, 224, 227, 229–231
- Sinorhizobium fredii* NGR234. *See Rhizobium*
sp. NGR234
- Sinorhizobium meliloti*
degradation of aromatic compounds, 102
denitrification, 109
genome organization, 92, 98, 99
host range, 105, 120
mobile genetic elements, 99
plasmids in *S. meliloti*, 92
- Site-specific resolution systems, 63, 66, 96
- S-layer gene(s), 190, 199
- SLH proteins. *See* Surface-layer homology
(SLH) proteins
- Small multidrug resistance (SMR) pumps,
277, 279
- SMR pumps. *See* Small multidrug resistance
(SMR) pumps
- Soluble hydrogenase(s) (SHs), 249, 250
- Solute transport systems, 96
- Sorbitol, 319
- Sorbitol-6-phosphate 2-dehydrogenase(s), 319
- Sortases, 192, 320
- Southern hybridization, 8, 9, 72, 124, 301, 313
- spa/Spa*, 155, 159
- Spermidine synthase(s), 24
- Sphingobium*, 57, 66
- Sphingomonas*
carbazole degradation, 67, 70, 78, 79
car genes, 67, 79
degradation of aromatics, 39
- Sphingomonas xenophaga*, 66
- Sphingomonas yanoikuyae*, 58
- spo0F/Spo0F*, 191
- Spores
germination of *B. anthracis* spores, 191
germination of *B. cereus* spores, 196
infectious state of *B. anthracis*, 188
inhibition of sporulation by PlcR and
AtxA, 195
sporulation of *B. anthracis*, 188, 191, 195
sporulation of *B. cereus*, 196
- 5S rRNA, 21
- SsaDH. *See* Succinate-semialdehyde
dehydrogenase(s) (SsaDHs)
- ssd/Ssd*, 277
- Staphylococcus*, 57
- Staphylococcus aureus*, 210, 272
- stb/Stb*, 163
- Sterne strains, 189, 192
- STPKs. *See* serin/threonine protein kinase(s)
(STPKs)
- Streptomyces* sp. 299, 301
- Streptomyces thermoautotrophicus*, 242
- S-triazine hydrolase(s), 298
- Subtilisin protease domain, 279
- Succinate, 73, 74, 98, 101, 278, 280

- Succinate-semialdehyde dehydrogenase(s),
21, 278
- Succinoglycan (EPS I), 103, 104
- Succinyl-CoA, 39
- Sucrose hydrolase(s), 274
- Superoxide dismutase(s) (SODs), 18, 227, 301
- Surface-layer homology (SLH) proteins, 192, 199
- Surface proteins, 315, 320
- syc/Syc*, 176
- Symbiotic islands, 120
- sym* loci, 99
- syr/Syr*, 101, 105, 107, 109, 124
- T**
- t-amyl methyl ether (TAME), 298
- TATA binding proteins (TBPs), 12, 15, 16, 19
- tbp/Tbp*, 12, 15, 16, 19
- TC-1, 273
- TC-2, 273
- TCA cycle. *See* Tricarboxylic acid (TCA) cycle
- tcb/Tcb*, 43
- T cell priming, 176
- T-complex, 139
- T-DNA, 137
 - covalent linkage to VirD2, 139
 - entry into plant nucleus, 140
 - formation of T-complex, 139
 - genes for synthesis of opines, 135, 138
 - genes for synthesis of phytohormones, 138
 - of pRi1724, 138
 - of pRiA4b, 137
 - processing, 139
 - rol* genes, 138
 - transfer by the VirB apparatus, 140
- T-DNA transfer stimulator sequence (TSS), 140
- Tellurite resistance, 71
- tfb/Tfb*, 15, 16, 19, 21, 22
- tfd/Tfd*, 43, 47
- tgt/Tgt*, 158
- Thioesterase(s), 286, 287
- Thiomersal, 225
- Thioredoxin, 4, 8, 12, 21, 22, 303
- Thioredoxin reductase(s), 4, 8, 12, 21
- thu/Thu*, 97
- tiorf24/Tiorf24*, 142, 143
- tiorf25/Tiorf25*, 142, 143
- Ti plasmids
 - classification, 135
 - copy number, 142
 - curing of, 143
 - genes for opine utilization, 142
 - incompatibility, 142
 - organization, 136
 - rep* genes, 142
 - tra* genes, 143
 - vir* genes, 139–141
- tml/Tml*, 138
- tmr/Tmr*, 138
- tms/Tms*, 138
- Tn501, 154, 225
- Tn552, 272
- Tn554, 225, 272
- Tn917, 199
- Tn4371, 47, 222
- Tn4378, 223, 225
- Tn4380, 223, 225, 227
- Tn4651, 72
- Tn4653, 65, 72
- Tn4655, 63, 65, 66
- Tn4676, 72
- Tn5046, 65
- Tn5280, 43
- Tn5530, 47
- Tn5542, 41
- TnBvie1, 227
- TnCme2, 223–225, 227, 229
- TnCme3, 15
- Tn501 family, 15
- tnm/Tnm*, 225
- tnp/Tnp*, 63–66, 72, 73, 223, 225
- tod/Tod*, 43
- TOL plasmid, 58, 72
- Toluene, 41–43, 58, 63, 71–74, 81
- Toluene ortho-monooxygenase(s), 42
- tom/Tom*, 42
- Topoisomerases, 192, 199
- topX/TopX*, 192
- Toxins
 - B. anthracis*, 188–192, 194, 196
 - BCEA0165 and BCEA0203, 194
 - B. cereus*, 187–192, 194, 196, 198, 202
 - B. thuringiensis*, 201, 202
 - Shigella*, 152
 - Yersinia*, 177, 178, 181
- Toxoflavin, 182
- TR1, 276, 281
- TR2, 276, 281
- tra*-box, 142
- Transaldolase(s), 318
- Transaminase(s), 278
- Transcriptional slippage, 161
- Transcription factors (TFBs), 15, 16, 19, 21, 22
- Transcriptome
 - Cupriavidus metallidurans*, 232, 233
 - pNGR234a, 127
 - Pseudomonas putida*, 73–75
- Transketolase(s), 252, 263, 318
- Translational coupling, 162

- Translation initiation factor IF-2, 19
- Transporters
- for arabinose in *s. meliloti*, 101
 - for arsenite efflux in *H. volcanii*, 18
 - for arsenite efflux in *N. pharaonis*, 24
 - for C₄-dicarboxylates in *s. meliloti*, 101
 - for cobalt and nickel in *C. metallidurans*, 227
 - for ectoine in *s. meliloti*, 102
 - for α -galactosides in *s. meliloti*, 101
 - for heavy metals in *N. pharaonis*, 24
 - for lead in *C. metallidurans*, 229
 - for 4-methylmuconolactone in *R. eutropha* H16, 23
- MFS permease for nitrate/nitrite in *R. eutropha* H16, 255
- MFS permease in *A. nicotinovorans*, 274
- for molybdenum in *A. nicotinovorans*, 279
 - for nickel in *O. carboxidovorans*, 264
 - for polyglutamate in *B. anthracis*, 148
 - for phosphate in *s. meliloti*, 102
 - peptide transporters in *H. volcanii*, 19
 - proteinase export in *s. meliloti*, 104
 - solute transporters in *S. meliloti*, 96
 - for sorbitol in *L. salivarius*, 319
- Transposase(s), 16, 18–20, 22–24, 48, 63, 72, 73, 99, 122, 190, 225, 227, 233, 234, 255, 257, 287, 291, 315
- Transposon mutagenesis, 263
- Transposons, 34, 35, 43, 47, 56, 63, 65, 66, 72, 98, 210, 225, 226, 229, 272, 273
- tra*/Tra, 11, 12, 23, 27, 121, 162, 321
- AW63, 201
 - InCP-9 plasmids, 62
 - Ti and Ri plasmids, 143
- trb*/Trb, 47, 136, 137, 143, 163, 220, 222, 259
- Trehalose, 96
- Trehalose 6-phosphate synthase(s), 96
- Trehalose synthase(s), 96
- Tricarboxylic acid (TCA) cycle, 45, 58, 79
- tRNA(s), 21, 95, 97, 98, 158, 316
- trx*/Trx, 8, 12
- tryptophanyl-tRNA synthetase(s), 18
- tts*/Tts, 126
- Tumour necrosis factor- α , 175
- tye*/Tye, 175
- Type IV pilus, 100
- Tyrosinase(s), 194
- tzs*/Tzs, 141
- U**
- UDP-galactose, 103
 - UDP-glucosamine, 107
 - UDP-glucose, 103
 - UDP-glucose-4-epimerase(s), 103
- Upper pathway(s), 39, 41, 43, 48, 62
- ura3*-based gene knockout, 15
- V**
- V antigen. *See* *lcr*/*Lcr*
 - vap*/*Vap*, 143
 - Variovorax paradoxus* H-4, 240
 - vir* boxes, 139
 - VirE2-interacting proteins (VIPs), 140
 - vir*/*Vir*, 163, 175
 - Agrobacterium*, 136, 137, 139, 140, 141, 144
 - B. thuringiensis*, 201
 - Shigella*, 157–159, 161–163
 - Yersinia*, 175
- X**
- Xanthine dehydrogenase(s), 263, 279
 - Xanthomonas*, 301
 - xdh*/*Xdh*, 263, 279
 - xer*/*Xer*, 243
 - Xylene(s), 35, 39, 41, 43, 66, 72
 - xyl*/*Xyl*, 39, 41, 43, 63, 73, 74
- Y**
- yad*/*Yad*, 176, 177
 - YD motifs, 220
 - Yersinia*, 172, 173
 - Yersiniabactin, 180
 - Yersinia enterocolitica*, 176,
 - biogroups, 173
 - contribution of plasmids to virulence, 174, 182
 - pathogenesis, 173
 - role of *YadA* in virulence, 177
 - Yersinia murine* toxin, 177
 - Yersinia* outer membrane proteins (Yops), 174–176
 - Yersinia pestis*
 - acquisition of plasmids, 180
 - anti-phagocytic activity, 179
 - capsule, 178, 179
 - colonization of flea digestive tract, 178
 - contribution of plasmids to virulence, 181, 182
 - lack of O-antigen, 174
 - life cycle, 173
 - mobile genetic elements, 181
 - Pla and survival in the lungs, 180
 - plasminogen activator, 179
 - role of *Ymt* in virulence, 177, 178
 - Yops, 174, 175

Yersinia pseudotuberculosis

- biotypes, 173
- contribution of plasmids to virulence,
 - 181, 182
- role of YadA in virulence,
 - 176, 177
- symptoms, 181
- ymt*/*Ymt*, 177, 178, 181
- yop*/*Yop*, 174, 176

ypk/*Ypk*, 174–177, 179*ysc*/*Ysc*, 175*y4xI*, 126*y4yS*, 126**Z***Zavarzinia compransoris*, 242*zntA*/*ZntA*, 231