

Plasmids captured in *C. metallidurans* CH34: defining the PromA family of broad-host-range plasmids

Géraldine A. Van der Auwera · Jaroslaw E. Król · Haruo Suzuki · Brian Foster · Rob Van Houdt · Celeste J. Brown · Max Mergeay · Eva M. Top

Received: 19 November 2008 / Accepted: 8 February 2009 / Published online: 4 March 2009
© Springer Science+Business Media B.V. 2009

Abstract The self-transmissible, broad-host-range (BHR) plasmid pMOL98 was previously isolated from polluted soil using a triparental plasmid capture approach and shown to possess a replicon similar to that of the BHR plasmids pSB102 and pIPO2. Here, complete sequence analysis and comparative genomics reveal that the 55.5 kb nucleotide sequence of pMOL98 shows extensive sequence similarity and synteny with the BHR plasmid family that now includes pIPO2, pSB102, pTER331, and pMRAD02. They share a plasmid backbone comprising replication, partitioning and conjugative transfer functions. Comparison of the variable accessory regions of these

plasmids shows that the majority of natural transposons, as well as the mini-transposon used to mark the plasmids, are inserted in the *parA* locus. The transposon unique to pMOL98 appears to have inserted from the chromosome of the recipient strain used in the plasmid capture procedure. This demonstrates the necessity for careful screening of plasmids and host chromosomes to avoid mis-interpretation of plasmid genome content. The presence of very similar BHR plasmids with different accessory genes in geographically distinct locations suggests an important role in horizontal gene exchange and bacterial adaptation for this recently defined plasmid group, which we propose to name “PromA”.

Electronic supplementary material The online version of this article (doi:10.1007/s10482-009-9316-9) contains supplementary material, which is available to authorized users.

Keywords Plasmid · Horizontal gene transfer · Broad host range · Transposon

G. A. Van der Auwera
Department of Microbiology and Molecular Genetics,
Harvard Medical School, Boston, MA 02115, USA

G. A. Van der Auwera · J. E. Król · H. Suzuki ·
C. J. Brown · E. M. Top (✉)
Department of Biological Sciences, University of Idaho,
PO Box 443051, Moscow, ID 83844-3051, USA
e-mail: evatop@uidaho.edu

B. Foster
DOE Joint Genome Institute, Walnut Creek, CA, USA

R. Van Houdt · M. Mergeay
Molecular & Cellular Biology, Belgian Center for Nuclear
Energy, SCK.CEN, 2400 MOL, Belgium

Abbreviations

BHR Broad host range
CDS Coding sequence
HGT Horizontal gene transfer
T4SS Type IV secretion system

Introduction

Conjugative gene transfer mediated by plasmids with broad host-ranges is generally believed to be a common and widespread mechanism for the transfer

of genes across a broad phylogenetic range of bacteria (Thomas 2000a). Within the Proteobacteria, broad-host range (BHR) plasmids have been defined as those plasmids that can be introduced and stably maintained in bacterial species from at least two subgroups (e.g., between α - and β -Proteobacteria; Szpirer et al. 1999). The promiscuity of the BHR plasmids seems to be in part dependent on the flexibility of the replicon to interact with host factors of distinct hosts, but the genetic determinants that distinguish BHR from narrow-host-range plasmids are still not exactly known (Espinosa et al. 2000). Moreover, current insights into the diversity of these BHR plasmids are far from complete. Until recently, the known self-transmissible BHR plasmids belonged to the incompatibility groups IncP-1 (IncP in *E. coli* classification), IncN, IncW, and IncU (Mazodier and Davies 1991). Of those, IncP-1 plasmids were the most promiscuous. More recently, complete genome sequence analysis of the plasmids pSB102, pIPO2, and pTER331, and partial sequence of plasmid pMOL98 revealed a new family of BHR plasmids (Schneiker et al. 2001; Tauch et al. 2002; Gstalder et al. 2003; Mela et al. 2008). This and other recent discoveries of novel BHR plasmid groups (Heuer et al. 2009) illustrate the lack of understanding of the genetic diversity of these important mobile genetic elements. Thus, in spite of general agreement on the importance of BHR plasmids in the adaptive evolution of bacteria, surprisingly little is known about the genetic diversity of these plasmids and the traits they encode.

Plasmids typically include two regions (Thomas 2000b): the “plasmid backbone” encodes genes involved in the replication, maintenance, and transfer of the plasmid, while one or more other plasmid sections are comprised of various “accessory” genes that confer specific phenotypic characteristics to the host, such as resistances to antibiotics, heavy metals and UV light, catabolic functions, or production of toxins (Top et al. 2000). The mobility of these accessory genes is evidenced by their widespread association with integrons, transposons, or insertion sequences. However, a large number of traits encoded by plasmids remain cryptic as several open reading frames do not code for any known phenotypes. An extreme example is the recently sequenced plasmid pQBR103, in which no function could be ascribed to 80% of the predicted coding sequences (Tett et al.

2007). These findings underscore our poor understanding of the diversity of functions encoded on these important genetic elements of the horizontal gene pool (Frost et al. 2005).

The goal of this study was to increase our insight into the diversity of BHR plasmids by analyzing the complete sequence of plasmid pMOL98, and comparing it with other members of its group, the “pIPO2-like” plasmid family. Plasmid pMOL98 is a derivative of plasmid pES1, which was isolated from a soil that was heavily polluted by hydrocarbons. Plasmid pES1 was captured from the soil community without retrieving its host, by means of a triparental exogenous plasmid isolation method using a cured derivative of the *C. metallidurans* strain CH34 as recipient (Top et al. 1994). So far no known phenotypes were found to be encoded by this plasmid. Therefore it was marked with a mini-Tn5 transposon in order to test its transferability and host-range (Szpirer et al. 1999; Gstalder et al. 2003), and the resulting plasmid was named pMOL98. Since the plasmid could transfer to and replicate in members of the α -, β -, and γ -Proteobacteria (Szpirer et al. 1999; Gstalder unpublished), it was defined as a BHR plasmid. Gstalder et al. (2003) showed that the oriV region of pMOL98 was very similar to those of plasmids pIPO2 and pSB102, and that their RepA proteins showed some degree of identity with the Rep proteins of IncW plasmids. The complete genome sequence of pMOL98 described in this study shows that the similarity among these three plasmids, the more recently described fourth plasmid family member pTER331, and a newly sequenced plasmid, pMRAD02 extends to the entire backbone regions of the plasmids. Moreover, examination of the accessory regions detected a transposon that may have been acquired from the recipient host during or after the plasmid capture procedure.

Materials and methods

Plasmids

Both pMOL98 and pIPO2T are derivatives of the wild-type plasmids pES1 and pIPO2, respectively (van Elsas et al. 1998; Tauch et al. 2002; Gstalder et al. 2003). In order to monitor their transfer, they were marked with an antibiotic resistance gene using

mini-transposons prior to sequencing. While pIPO2T contains mini-Tn5 *luxAB*, encoding luciferase genes and tetracycline resistance, pMOL98 contains mini-Tn5 Km, encoding kanamycin resistance (de Lorenzo et al. 1990). In contrast, plasmid pSB102, pTER331, and pMRAD02 represent wild-type plasmids (Schneiker et al. 2001; Mela et al. 2008). pIPO2 [GenBank accession: NC_003213] was isolated from wheat rhizosphere in Wageningen, The Netherlands, pSB102 [GenBank accession: NC_003122] from alfalfa rhizosphere in Braunschweig, Germany, pTER331 [GenBank accession: NC_010332] from a dune soil on Wadden Island Terschelling, The Netherlands, and pMOL98 from hydrocarbon-polluted soil in Essen, Germany. pMRAD02 [GenBank accession: NC_010509] was sequenced along with the genome of its host strain, *Methylobacterium radiotolerans* JCM2831, which was isolated from unpolished rice in Japan (Ito and Iizuka 1971).

Sequencing and annotation

The genome sequence of plasmid pMOL98 was determined by shotgun sequencing at the DOE Joint Genome Institute (Walnut Creek, CA). One region of poor sequence quality (mini-Tn5 Km) was resequenced at the University of Idaho by PCR amplification followed by sequencing, using Big Dye Terminator v3.1 Cycle Sequencing Kit and a 3730 DNA Analyzer (Applied Biosystems). Machine annotation of these sequences was performed by the J. Craig Venter Institute Annotation Service and further annotated manually. The complete sequence of pMOL98 has been deposited in the GenBank database [GenBank accession: FJ666348].

Bioinformatic analyses and software

GenBank (<http://ncbi.nlm.nih.gov>) was searched for similar sequences using BLAST (Altschul et al. 1997). Coding sequences and proteins were aligned using ClustalW (Thompson et al. 1994) and manual realignments were unnecessary. Phylogenetic trees were inferred using the neighbor joining algorithm (Saitou and Nei 1987) on protein distances with Poisson correction (MacVector, Accelrys). Plasmid map and alignment figures included in this work were generated using the BioPython, GenomeDiagram, and ReportLab open software packages. Plasmid

alignments were performed using Mauve (Darling et al. 2004) and rendered graphically using PlasmidG: a Python-based sequence map and alignment drawing tool for plasmids (Van der Auwera, in preparation), with identity scoring by the BLAST algorithm bl2seq (Tatusova and Madden 1999).

Results

Basic genome sequence information for plasmid pMOL98

The complete nucleotide sequence of pMOL98, including the 1,845 bp mini-Tn5 tag, was determined to be 55,563 bp long with a G + C content of 59.9% (Fig. 1). Analysis of the coding content and organization revealed few intergenic regions and a total of 67 coding sequences (CDS) (Fig. 1; Supplemental Table 1). The sequence has a 91% coding ratio with an average CDS length of 724 bp. Similarity searches showed that most of the 67 predicted CDSs encoded proteins with similarity to proteins from plasmids of other organisms. While biological functions could be attributed to 44 CDSs, 19 CDSs code for conserved hypothetical proteins, and the remaining four predicted genes do not have any known homologues. The details of all 67 CDSs are available in Supplemental Table 1.

pMOL98 shares a common backbone with pIPO2, pTER331, pSB102, and pMRAD02

The similarity of the replicon of pMOL98 to that of pIPO2 and the other pIPO2-like plasmids, pTER331 and pSB102 has been noted previously (Gstalter et al. 2003; Mela et al. 2008). Similarity searches of GenBank indicate that the sequence similarity and synteny observed for the replicon can be extended to the full length of the plasmid pMOL98 detailed sequence alignments of these sequences as well as a newly sequenced plasmid, pMRAD02 from the soil proteobacterium *Methylobacterium radiotolerans* JCM2831, shows that all five plasmids share a common backbone comprising replication, maintenance/control, and conjugative transfer genes organized in clearly identifiable functional modules (Fig. 2). The fundamental differences between the plasmids reside mainly in discrete regions encoding accessory genes. The genetic content and organization

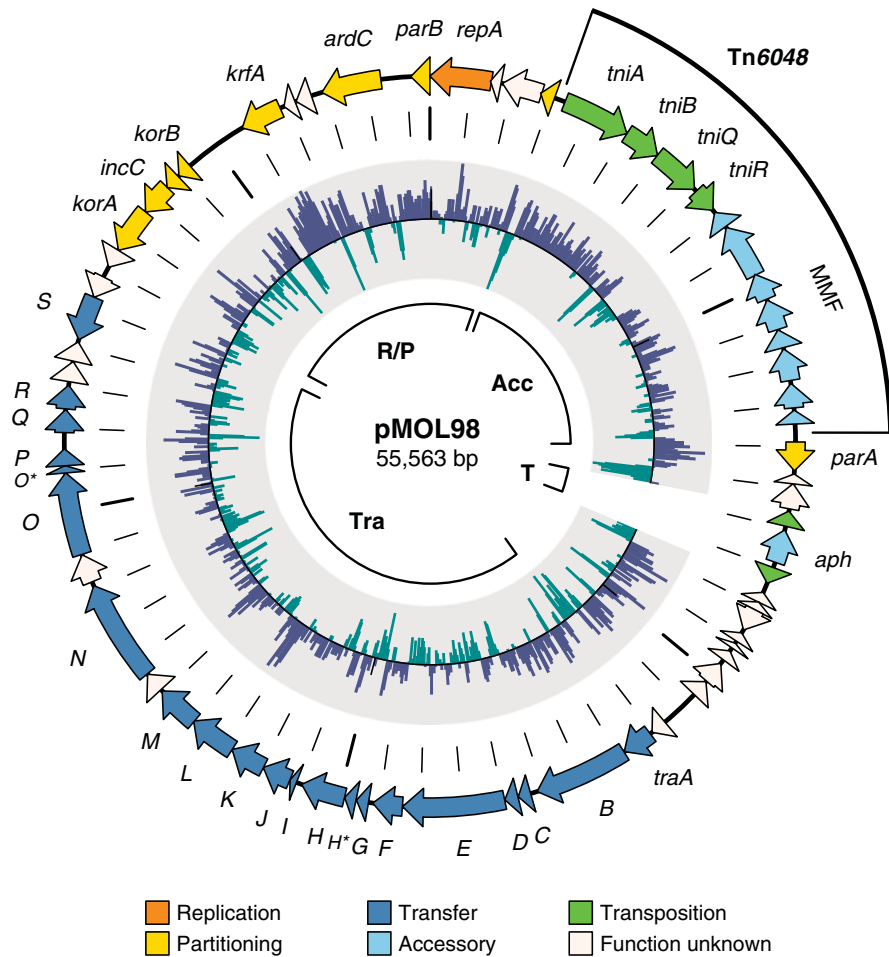


Fig. 1 Circular map of pMOL98. Coding sequences (CDS) are represented by *block arrows* on the *outer circle*. Predicted functions/homologies are indicated by the color key featured below. The *first circle* from the center delineates the major functional modules identified on the plasmid; *Tra* indicates the transfer region, *R/P* indicates the replication and partitioning region, *Acc* indicates the main accessory genes region and *T* indicates the selection marker tag. The *second circle* from the center is a circular bar graph of the G + C composition (100

base window) of the plasmid sequence, with the median value (58.5%) as baseline; values above the *line* are G + C rich (max value 78% G + C) and values below the *line* are A + T rich (min value 39% G + C). Note that the sequence corresponding to the mini-Tn5 Km tag was omitted from these calculations to remove bias, since it is A + T rich. The *third circle* from the center is a graduated size scale with *small tick marks* every 1 kb and *large tick marks* every 10 kb

of the shared backbone has been described extensively most recently for pTER331 (Mela et al. 2008). The high level of conservation of genetic content and synteny observed between the corresponding segments of the five plasmids makes a detailed description of the functional annotation of pMOL98 redundant. Therefore we only point out interesting differences and findings below.

As shown in Fig. 2, the levels of similarity between the five plasmids are not uniform over the entire length of the common backbone. Regardless of

the plasmid pairs considered, the highest identity scores from pairwise alignments of translated DNA were observed for the predicted Tra proteins, TraA to TraM, encoding homologues of the VirB system as defined in the canonical Type IV Secretion System (T4SS) from *Agrobacterium tumefaciens* (Christie and Cascales 2005). In contrast, the Tra region encoding, among others things, the VirD2 and VirD4 homologues of the T4SS, TraN to TraS, display much lower amino acid identity scores, as do the segments encoding the KorA-IncC-KorB, and KrfA elements

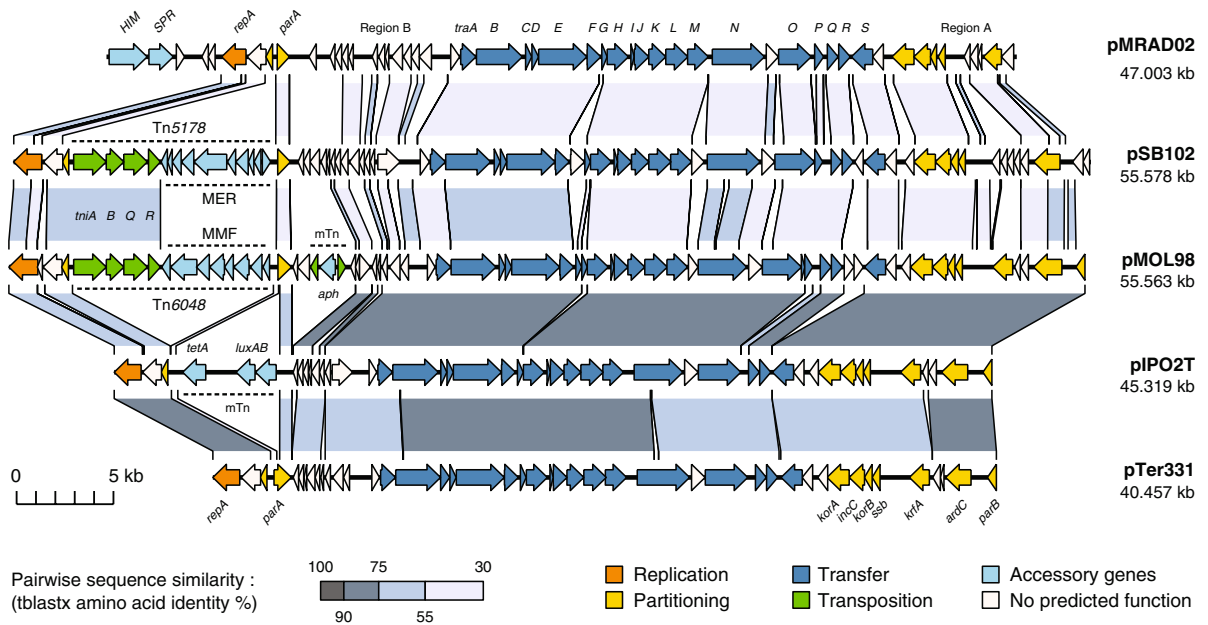


Fig. 2 Linear alignment of pMRAD02, pSB102, pMOL98, pIPO2T, and pTER331. CDSs are represented by *block arrows*. Key genes and functions are annotated above or below the corresponding regions. *HIM* stands for hemagglutinin, *SPR* for serine protease, *MER* for mercury resistance, *MMF* for multiple metal response phenotype and *mTn* for mini-Tn5. See Supplemental Table 1 for full gene annotation details.

Predicted functions/homologies are indicated by the color key featured below the figure. Well-conserved (>30% amino acid identity of translated DNA sequences) segments of the plasmids are paired by *shaded regions*, with local identity percentage ranges indicated by the color key scale featured below the figure. Scale is indicated by the *bar* in the lower left-hand corner

of the maintenance and central control region. One segment in particular, comprising the putative *KrfA* locus on pMOL98, pIPO2T, and pTER331, is almost entirely unrecognizable on pSB102. There, it has been annotated as “Region A” with no assigned putative functions. Furthermore, the corresponding region on pMRAD02 is shorter and also does not contain any predicted functions. A second segment that could not be associated with predicted functions, which has been annotated as “Region B” on pSB102, is poorly conserved across the plasmid group, such that their respective annotations vary widely in terms of predicted CDS in this region. Whether these two segments could truly be considered as part of the “essential” plasmid backbone remains undetermined in the absence of functional information, as they could conceivably correspond to relics of shared ancestral accessory gene modules.

To evaluate the evolutionary relatedness of these five plasmids, a phylogenetic analysis was performed on six concatenated backbone proteins shared by all five plasmids. These proteins were chosen because of

their large size and high level of synteny (*RepA*, *TraB*, *TraE*, *TraN*, *TraO*, and *KorB*). Results suggest that pMOL98 and pIPO2T are most similar to each other and then to pTER331, while pSB102 and pMRAD02 are both more divergent (Fig. 3a).

An additional sequence comparison with other plasmids confirmed that the shared plasmid backbone displays a mosaic structure, as noted previously (Schneiker et al. 2001). Indeed, the different functional backbone modules (involved in replication, maintenance/control and transfer) show phylogenetic relationships to those of plasmids from different known incompatibility groups (data not shown). A broad similarity search showed that no other plasmids known to date share all three backbone regions of these five plasmids. The plasmids pXF51 [*GenBank accession: NC_002490*] and pFBAOT6 [*GenBank accession: NC_006143*] were previously suggested to belong to this plasmid group because they share much sequence similarity with the transfer regions of pIPO2, pSB102, and pTER331 (Marques et al. 2001; Schneiker et al. 2001; Rhodes et al. 2004;

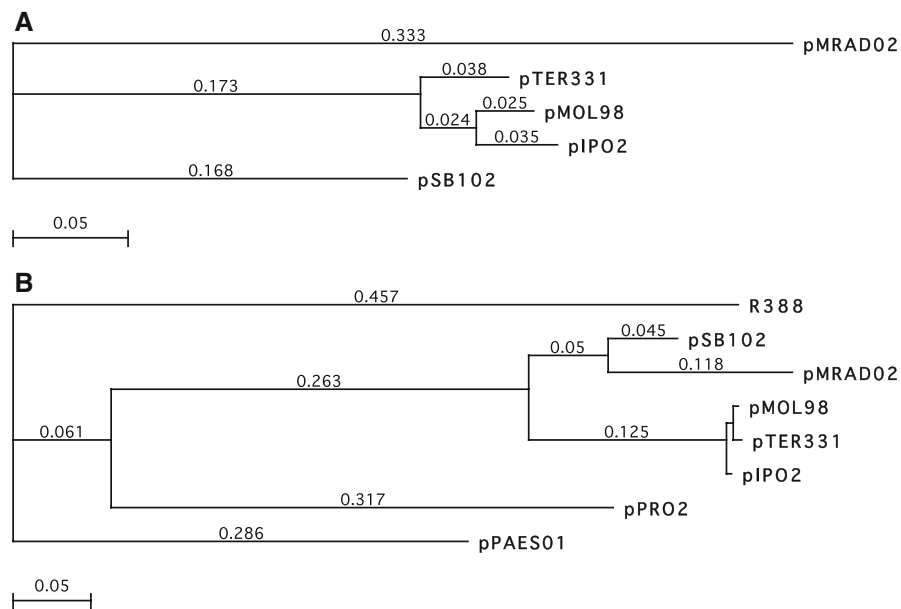


Fig. 3 Phylogenetic trees of (a) six shared backbone proteins and (b) RepA proteins. Phylogenetic relationships were inferred using the neighbor joining algorithm on protein distances with Poisson correction for (a) six concatenated backbone proteins (RepA, TraB, TraE, TraN, TraO, and KorB) and (b) the RepA protein only, from plasmids pMRAD02,

pSB102, pMOL98, pIPO2T, and pTER331. Comparison of RepA also included the IncW plasmids R388 [GenBank accession: BR000038], pPAES01 [GenBank accession: CP001109] and pPRO2 [GenBank accession: NC_008608]. Phylogenetic distance (amino acid difference percentage) is indicated by the length of the tree branches and scale bars

Mela et al. 2008). However, a detailed analysis showed that there is less than 30% amino acid sequence identity between the predicted RepA protein of either pXF51 or pFBAOT6 and that of any of the five plasmids, suggesting that their replication modules are not directly related. Figure 3b shows that based on a phylogenetic analysis of the RepA protein alone, the five plasmids form a cluster related to but distinct from IncW plasmids such as R388 [GenBank accession: BR000038]. Therefore we propose to name this new group of BHR plasmids ‘PromA’, for ‘promiscuous A’. We also suggest that other BHR plasmids like IncN, IncP, IncU, and IncW could then be designated PromN, PromP, PromU, and PromW in this new classification system.

Most unique segments of the PromA plasmids are transposons inserted near or at the *parA* resolvase locus

The unique segments of each plasmid were analyzed in terms of location, genetic content and natural occurrence. Smaller indels (<2 kb with no predicted functions) were disregarded and the following were

identified as major unique regions (Table 1): a 6 kb segment on pMRAD02, a 10 kb segment on pSB102, a 10 kb segment and a 1.8 kb segment on pMOL98, a 5 kb segment on pIPO2T and none on pTER331. These regions may correspond to accessory genes that are not essential for the basic replication, maintenance and transfer functions of the plasmid, but may be important for plasmid success because of the selective advantages they confer to the host. All but one of these were found immediately upstream or downstream of the predicted CDS for the backbone gene *parA*, which is thought to encode a resolvase involved in plasmid maintenance. Sequence comparisons to *parA* on pTER331, which lacks any major unique segment, indicated that the corresponding CDS on the other plasmids are truncated forms interrupted by the presence of the unique segments. In the case of pIPO2T, the insertion seems to have occurred centrally within the CDS, leaving no likely ORF to be annotated for the *parA* gene. The 6 kb segment unique to pMRAD02 constitutes a notable exception as it does not display recognizable transposon features and is not located near the *parA* locus but immediately downstream of the *repA* gene. Interestingly this corresponds

to the location on the PromA backbone of a predicted nuclease gene, *parB*. Thus the *parA* locus could be considered a hot-spot for accessory elements in plasmids of the PromA group.

On pMOL98, a unique 1.8 kb segment located closely downstream of *parA* contains the mini-Tn5 Km transposon that was used to mark the wild-type plasmid. Similarly, on pIPO2T, the 5 kb segment was recognized as mini-Tn5 *luxAB*, used to mark pIPO2 (Table 1). The two remaining unique, larger segments, located immediately upstream of the *parA* locus on pSB102 and pMOL98 were both identified as naturally occurring transposons of the Tn5053-like family. This is a family of transposons characterized by a *tniABQR* transposition module and an accessory module encoding mercury resistance (*mer*) genes (Kholodii et al. 1993), but related/variant forms such as Tn402 and Tn5090 have been shown to contain integrons instead of the *mer* genes (Radstrom et al. 1994). Transposons in this family are typically delineated by 25 bp terminal inverted repeats (IR), generate 5 bp direct repeats upon insertion and show a strong tendency to insert into plasmid resolvases (Minakhina et al. 1999). The 10,413 bp transposon on pSB102 had been identified previously and annotated as Tn5178 (Schneiker et al. 2001), and was found to possess the classic features of the family: nearly perfect 25 bp IRs, perfect 5 bp direct repeats and the *mer* genes (Table 2). The 10,402 bp transposon on pMOL98 displays transposition features that are very similar to those of Tn5178: the *tniABQR* transposition-related genes (87% amino acid identity of predicted products), nearly identical 25 bp IRs, and perfect 5 bp direct repeats (Table 2). The point of insertion of the transposon with respect to the *parA* locus is also almost exactly the same. However, the accessory genes of this transposon do not correspond to known mercury resistance genes, nor do they display integron-like features.

Table 1 Unique segments per plasmid

Plasmid	Content	Segment size (kb)	Associated phenotype
pSB102	Tn5178	10.4	Mercury resistance
pMOL98	Tn6048	10.4	Response to heavy metals
	mini-Tn5 Km	1.8	Kanamycin resistance (marker tag)
pIPO2T	Mini-Tn5 <i>luxAB</i>	5.2	Tetracycline resistance (marker tag)

Based on the results of a nucleotide similarity search against bacterial genomes in GenBank, the 10 kb transposon identified on pMOL98 was found to be 100% identical to a transposon present in three copies in the genome of *C. metallidurans* strain CH34. This transposon was first designated TnCme1 and later renamed Tn6048 (Table 2) (Van Houdt et al. 2009). The accessory genes of Tn6048 have been provisionally annotated as *mmf* for multiple metal response phenotype on the basis of experimental observations (M. Mergeay et al. unpublished data). A transposon similar but not identical to Tn6048 was found on the plasmid pBVIE02 from *Burkholderia vietnamensis* strain G4. The accessory genes within this transposon on pBVIE02 display 97% nucleotide identity to the copies found in *C. metallidurans* CH34 and on pMOL98, and carry a 1,271 bp IS600 insertion sequence interrupting one of the accessory genes. Finally, a similarity search focused on the accessory genes alone of Tn6048 yielded stretches of *Burkholderia cenocepacia*, *Dechloromonas aromatica* and *Ralstonia pickettii* genomes that contain nearly complete sets of homologous genes, with nucleotide identities ranging between 70 and 80%, but are not associated with any transposition-related elements. Thus Tn6048-like transposons seem to be widespread on plasmids and chromosomes of β -Proteobacteria.

Discussion

Based on an earlier characterization of the replicon of pMOL98 (Gstalder et al. 2003) and its inclusion in a subsequent comparative analysis (Mela et al. 2008), pMOL98 was proposed to belong to the same family of BHR plasmids as the fully sequenced plasmids pIPO2T (Tauch et al. 2002), pTER331 (Mela et al. 2008) and pSB102 (Schneiker et al. 2001). The full sequence analysis of pMOL98 has now made it possible to establish the extent of this similarity over its entirety. The 47 kb plasmid pMRAD02 from *Methylobacterium radiotolerans* is a fifth member of this novel BHR plasmid group. These five plasmids were isolated from either rhizosphere or soil in geographically distinct locations in The Netherlands, Germany, and Japan. In spite of these different origins, the five plasmids share a highly conserved backbone sequence comprising replication, maintenance and control, and transfer functions. Since the

Table 2 Features of Tn5178/Tn6048 copies or variants

Transposon and accessory genes			IR sequence	
IR sequences of Tn5053, Tn5178 and Tn6048				
Tn5053 MER			TGTCGTTTTTCAGAAGACGACCGCAC	
Tn5178 MER			TGTCATTTTCAGAAG(AC/GT) ^a GACTGCAC	
Tn6048 MMF			TGTCATTTTCAGAAGACGACTGCAC	
Transposon	Replicon	Locus	Coordinates	DR sequence
Location of the copies of Tn5178 and Tn6048				
Tn5178	pSB102	<i>parA</i>	43700–54113	TTTAC
Tn6048	pMOL98	<i>parA</i>	3206–13608	TTTAG
Tn6048	pAKD16	<i>parA</i>	16925–27327	ATTTG
Tn6048	CH34 chromosome 1	ND	3257335–3267736	GGTCG
Tn6048	CH34 chromosome 2	ND	455407–465809	CAATG
Tn6048	CH34 chromosome 2	ND	1501719–1512121	TCAGT
Tn6048	pBVIE02	ND	190671–206171	GCGCC

^a Mismatch of 2 bp, both strand versions given in parentheses

MER Mercury resistance genes; *MMF* multiple metal response phenotype; *IR* inverted repeat; *ND* Not determined; *DR* direct repeat

entire backbones are not directly related to any single characterized incompatibility group, and the replicons are similar to but distinct from that of IncW plasmids, the plasmids should be classified as a new family, which we propose to name “PromA”.

The variable segments of the PromA plasmids (which consisted mainly of transposons) were found clustered in a specific locus of the shared backbone that corresponds to a predicted resolvase gene. This observation of a possible “hot-spot” for insertion of transposons in this location is in agreement with previous reports (Minakhina et al. 1999; Sota et al. 2007). Factors commonly invoked to explain high frequency insertion of transposons into sites with certain sequence features include the presence of repetitive palindromic elements as well as A/T content and topology of the target DNA (Ason and Reznikoff 2004; Liu et al. 2005; Tobes and Pareja 2006). It has been suggested that transcription levels of genes in the target locus may negatively influence transposon insertion rates (Manna et al. 2004). What remains unknown is whether insertion into plasmid sites might in some cases occur preferentially compared to insertion into the chromosome. It is tempting to speculate that conformational effects of plasmid supercoiling could produce or enhance sites that are especially “attractive” to transposases. Evolutionary pressures on the plasmid and/or on the transposon can

explain preferential insertion of transposons into plasmids. For the transposon, insertion in a plasmid could be a strategy maximizing its overall copy number. Indeed, transposition within a single genome is thought to be self-limited to avoid a “poisoning effect”, i.e., damage due to insertions in essential genes (Reznikoff 2003). Mobile plasmids offer transposons a route to “colonize” other genomes. For the plasmid itself, transposons have the potential of promoting horizontal and vertical spread because they often carry accessory genes encoding host-beneficial functions such as catabolic properties and resistance to drugs or to heavy metals. Nevertheless, insertion sites may be limited because plasmid gene density is typically high and a multiplicity of insertions could lead to plasmids that are less stable, less transferable by conjugation, or confer a higher metabolic cost to their host (Sota et al. 2007). Whatever the mechanism, the tendency of PromA plasmids to “capture” transposons coupled with the ability to transfer to a wide variety of hosts and to retromobilize non-conjugative vectors (Szpirer et al. 1999) makes them effective shuttle vectors for potential host-beneficial genes. Just like the plasmids of the IncP group (Schluter et al. 2007), they may constitute significant agents of bacterial evolution.

It is interesting to examine the nature of plasmid-borne accessory genes in light of the method by

which the plasmid was isolated or identified. For instance, since pSB102 was isolated by selecting for a mercury resistance phenotype, it is not surprising that its accessory genes encode mercury resistance. In contrast, pMRAD02 was not specifically targeted but sequenced along with the complete genome of its host strain in the context of a study on methylotrophic organisms (C. Marx personal communication). The accessory genes identified on pMRAD02 are predicted to encode an adhesin/hemagglutinin and its cognate serine protease. These gene products may be involved in the plant-associated lifestyle of *Methylobacterium*. In other organisms hemagglutinin/protease pairs have been implicated in pathogenesis as invasion factors, and some of these were found on plasmids like pO157 of enterohemorrhagic *Escherichia coli* (Burland et al. 1998). Whatever the true functions of these accessory proteins, they are not of the kind that is easily selected for in a plasmid capture assay. Thus this plasmid would not be known today as a member of the PromA family if its host had not been chosen for genome sequencing.

Plasmid pIPO2T, which was selected for its ability to mobilize a non-conjugative vector, does not possess genetic regions that can be identified as potentially host-beneficial accessory genes at this point. The same is true for plasmid pTER311, which was found in an isolate of *Collimonas fungivorans* and studied for its contribution to the rhizosphere competence of its host (Mela et al. 2008).

Plasmid pMOL98 was isolated based on its ability to mobilize another plasmid in a triparental mating. Like pIPO2T, it did not encode known accessory functions. The large transposable element on pMOL98, Tn6048, is probably an artefact of the isolation procedure, as pMOL98 was captured using a plasmid-cured derivative of *C. metallidurans* strain CH34, which possesses two copies of Tn6048 on its chromosome. The transposon may thus have transposed from the recipient chromosome onto the plasmid during or after the triparental mating procedure in the laboratory. This post-capture transposition hypothesis is supported by the subsequent discovery of another identical copy of the transposon in a recently sequenced IncP-1 plasmid, pAKD16, which was also isolated by triparental mating using the same CH34 derivative as recipient (Drønen et al. 1998; Sen, Rogers, Suzuki, Brown, Top, unpublished data). One cannot completely exclude that a copy of

Tn6048 was already present on pMOL98 before it was captured, given that a similar (but not identical) Tn6048-like element has been found in *B. vietnamsis* and its accessory genes in *B. cenocepacia*, *Dechloromonas* and *Ralstonia pickettii*. However, given the 100% identity with the recipient chromosome and the independent finding of an identical copy on a different plasmid captured with the same host, pMOL98 probably captured Tn6048 in the laboratory, and not in the field.

The wild-type plasmids pMOL98 and pIPO2 (before they captured natural and mini-transposons) may very well be cryptic only because the accessory functions they encode, or the more subtle regulatory effects they may have on the host, are as yet unrecognized. Nevertheless, the existence in nearly all genomes, microbial or otherwise, of small mobile elements that do not encode recognizable beneficial coding sequences seems to support the hypothesis that simply their ability to replicate and spread is sufficient to sustain themselves as so-called “selfish elements” (Koonin and Wolf 2008). There are alternative explanations that invoke the contribution of such elements to genomic plasticity to justify their persistence within genomes, as mobile genetic elements with different structures are likely to be subject to very different balances of evolutionary pressures. As regards BHR plasmids, more data are needed, partly because until now the appreciation of plasmid genome content and diversity has been skewed by the common practice of sequencing elements that were isolated in relation to a specific phenotype (e.g., resistance, degradation or virulence). Hopefully future studies of plasmids isolated without known selective pressure will yield data that can be used to examine this question.

Metagenomics projects may help increase knowledge of natural plasmid diversity to a certain extent by generating large amounts of environmentally relevant data. However, it is not clear at present if “meta” content corresponding to plasmids can always be distinguished from that encoded on chromosomes. Until this practical issue can be resolved, whole genome sequencing of individual strains may remain the best option for characterizing plasmids in their native context. Retrieving plasmids by means of triparental mating procedures that select for mobilization potential only is still a reasonable alternative. However, the observation that two unrelated BHR plasmids that were isolated using this

method, both may have acquired a transposon from the chromosome of the recipient strain raises the concern that the genomic content of these plasmids might be incorrectly interpreted. In the case of plasmids isolated by triparental matings, adequate controls should be used to detect such cases of colonization by the recipient's mobile genetic elements. Therefore, the use of fully sequenced strains for such plasmid captures is strongly recommended.

Finally, although one cannot know for sure the original hosts of plasmids pMOL98, pIPO2T, and pSB102, which were isolated using plasmid capture procedures, a genomic signature comparison method can propose putative long-term plasmid hosts. Indeed, extensive data support the proposal that foreign DNA acquires the host's nucleotide composition during long-term residence (Lawrence and Ochman 1997; Ochman et al. 2000). Thus if the plasmids would have resided in a set of closely related hosts over evolutionary time, their nucleotide composition should be similar to that of these strains. The genomic signature, which is the set of relative abundance values of short oligonucleotides (di-, tri-, and tetra-nucleotides), has been used to compare compositional similarity between bacterial chromosomes and mobile genetic elements such as plasmids and phages (Campbell et al. 1999; Pride et al. 2006; van Passel et al. 2006). Suzuki et al. (2008) applied the Mahalanobis distance method to this approach to predict the putative long-term host(s) of a plasmid with increased confidence. The Mahalanobis distance for each plasmid-chromosome pair was then converted to a *P*-value. *P*-values close to 1 represent high plasmid-chromosome signature similarities, suggesting coevolution of the plasmid with that or closely related hosts. Interestingly, when this method was applied to the pMOL98, pIPO2T, and pSB102, using the trinucleotide relative abundance, the most likely long-term hosts for all three plasmids were the β -proteobacteria *Polaromonas naphthalenivorans* and *Dechloromonas aromatica* ($P = 0.95$ – 0.98 ; results not shown). These organisms are known for their ability to degrade a variety of pollutants including aromatic hydrocarbons (Coates et al. 2001; Jeon et al. 2004; Chakraborty et al. 2005; Mattes et al. 2008). Since pMOL98 was isolated from soil that was heavily polluted by polynuclear aromatic hydrocarbons, the host prediction for this plasmid seems plausible. However, after removing the non-native

transposons Tn6048 and mini-Tn5 Km from the pMOL98 sequence, the signature was still most similar to that of these two strains but the degree of similarity was much lower, with low *P*-values ($P = 0.74$ and 0.65 , respectively) similar to those for plasmid pTER311 (0.79 and 0.72). This may suggest that these plasmids have evolved in bacteria that are not yet sequenced and not very similar to the currently sequenced strains. Alternatively, being promiscuous plasmids they may have moved around so frequently among diverse bacteria that they did not coevolve with any particular host. While this preliminary result requires further study, the approach may help us better understand the ecology and evolution of the environmentally relevant PromA family of BHR plasmids.

Acknowledgments This work was supported by the Microbial Genome Sequencing Program of the National Science Foundation (NSF grant EF-0627988), as well as by a former European Community program BIOTECH (grant BI02-CT92-0491). We are grateful to the DOE Joint Genome Institute (JGI) for providing the DNA sequence of pMOL98 (agreement UA_Top_173_060602). We thank Stacey Poler and Linda Rogers for providing JGI with pMOL98 plasmid DNA and confirming restriction site profiles.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. doi:[10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389)
- Ason B, Reznikoff WS (2004) A high-throughput assay for Tn5 Tnp-induced DNA cleavage. *Nucleic Acids Res* 32:e83. doi:[10.1093/nar/gnh080](https://doi.org/10.1093/nar/gnh080)
- Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR (1998) The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. *Nucleic Acids Res* 26:4196–4204. doi:[10.1093/nar/26.18.4196](https://doi.org/10.1093/nar/26.18.4196)
- Campbell A, Mrazek J, Karlin S (1999) Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proc Natl Acad Sci USA* 96:9184–9189. doi:[10.1073/pnas.96.16.9184](https://doi.org/10.1073/pnas.96.16.9184)
- Chakraborty R, O'Connor SM, Chan E, Coates JD (2005) Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene compounds by *Dechloromonas* strain RCB. *Appl Environ Microbiol* 71:8649–8655. doi:[10.1128/AEM.71.12.8649-8655.2005](https://doi.org/10.1128/AEM.71.12.8649-8655.2005)
- Christie PJ, Cascales E (2005) Structural and dynamic properties of bacterial type IV secretion systems. *Mol Membr Biol* 22:51–61. doi:[10.1080/09687860500063316](https://doi.org/10.1080/09687860500063316)

- Coates JD, Chakraborty R, Lack JG, O'Connor SM, Cole KA, Bender KS, Achenbach LA (2001) Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*. *Nature* 411:1039–1043. doi:[10.1038/35082545](https://doi.org/10.1038/35082545)
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. doi:[10.1101/gr.2289704](https://doi.org/10.1101/gr.2289704)
- de Lorenzo V, Herrero M, Jakubzik U, Timmis KN (1990) Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J Bacteriol* 172:6568–6572
- Drønen AK, Torsvik V, Goksøyr J, Top EM (1998) Effect of mercury addition on plasmid incidence and gene mobilising capacity in bulk soil. *FEMS Microbiol Ecol* 27:381–394. doi:[10.1016/S0168-6496\(98\)00085-3](https://doi.org/10.1016/S0168-6496(98)00085-3)
- Espinosa M, Cohen SN, Couturier M, del Solar G, Diaz-Orejas R, Giraldo R (2000) Plasmid replication and copy number control. In: Thomas CM et al (eds) *The horizontal gene pool: bacterial plasmids and gene spread*. Harwood Academic Publishers, Amsterdam, pp 1–47
- Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* 3:722–732. doi:[10.1038/nrmicro1235](https://doi.org/10.1038/nrmicro1235)
- Gstalter ME, Faelen M, Mine N, Top EM, Mergeay M, Couturier M (2003) Replication functions of new broad host range plasmids isolated from polluted soils. *Res Microbiol* 154:499–509. doi:[10.1016/S0923-2508\(03\)00143-8](https://doi.org/10.1016/S0923-2508(03)00143-8)
- Heuer H, Kopmann C, Binh TT, Top EM, Smalla K (2009) Spreading antibiotic resistance through spread manure: characteristics of a novel plasmid type with low % G + C content. *Environ Microbiol*. doi: [10.1111/j.1462-2920.2008.01819.x](https://doi.org/10.1111/j.1462-2920.2008.01819.x). (Published online)
- Ito H, Iizuka H (1971) Taxonomic studies on a radio-resistant *Pseudomonas*. Part XII. Studies on the microorganisms of cereal grain. *Agric Biol Chem* 35:1566–1571
- Jeon CO, Park W, Ghiorse WC, Madsen EL (2004) *Polaromonas naphthalenivorans* sp. nov., a naphthalene-degrading bacterium from naphthalene-contaminated sediment. *Int J Syst Evol Microbiol* 54:93–97. doi:[10.1099/ijs.0.02636-0](https://doi.org/10.1099/ijs.0.02636-0)
- Kholodii GY, Gorlenko Z, Lomovskaya OL, Mindlin SZ, Yurieva OV, Nikiforov VG (1993) Molecular characterization of an aberrant mercury resistance transposable element from an environmental *Acinetobacter* strain. *Plasmid* 30:303–308. doi:[10.1006/plas.1993.1064](https://doi.org/10.1006/plas.1993.1064)
- Koonin EV, Wolf YI (2008) Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* 36:6688–6719. doi:[10.1093/nar/gkn668](https://doi.org/10.1093/nar/gkn668)
- Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* 44:383–397. doi:[10.1007/PL00006158](https://doi.org/10.1007/PL00006158)
- Liu G, Geurts AM, Yae K, Srinivasan AR, Fahrenkrug SC, Largaespada DA et al (2005) Target-site preferences of sleeping beauty transposons. *J Mol Biol* 346:161–173. doi:[10.1016/j.jmb.2004.09.086](https://doi.org/10.1016/j.jmb.2004.09.086)
- Manna D, Breier AM, Higgins NP (2004) Microarray analysis of transposition targets in *Escherichia coli*: the impact of transcription. *Proc Natl Acad Sci USA* 101:9780–9785. doi:[10.1073/pnas.0400745101](https://doi.org/10.1073/pnas.0400745101)
- Marques MV, da Silva AM, Gomes SL (2001) Genetic organization of plasmid pXF51 from the plant pathogen *Xylella fastidiosa*. *Plasmid* 45:184–199. doi:[10.1006/plas.2000.1514](https://doi.org/10.1006/plas.2000.1514)
- Mattes TE, Alexander AK, Richardson PM, Munk AC, Han CS, Stothard P, Coleman NV (2008) The genome of *Polaromonas* sp. strain JS666: insights into the evolution of a hydrocarbon- and xenobiotic-degrading bacterium, and features of relevance to biotechnology. *Appl Environ Microbiol* 74:6405–6416. doi:[10.1128/AEM.00197-08](https://doi.org/10.1128/AEM.00197-08)
- Mazodier P, Davies J (1991) Gene transfer between distantly related bacteria. *Annu Rev Genet* 25:147–171. doi:[10.1146/annurev.ge.25.120191.001051](https://doi.org/10.1146/annurev.ge.25.120191.001051)
- Mela F, Fritsche K, Boersma H, van Elsas JD, Bartels D, Meyer F et al (2008) Comparative genomics of the pIPO2/pSB102 family of environmental plasmids: sequence, evolution, and ecology of pTer331 isolated from *Collimonas fungivorans* Ter331. *FEMS Microbiol Ecol* 66:45–62. doi:[10.1111/j.1574-6941.2008.00472.x](https://doi.org/10.1111/j.1574-6941.2008.00472.x)
- Minakhina S, Kholodii G, Mindlin S, Yurieva O, Nikiforov V (1999) Tn5053 family transposons are *res* site hunters sensing plasmidal *res* sites occupied by cognate resolvases. *Mol Microbiol* 33:1059–1068. doi:[10.1046/j.1365-2958.1999.01548.x](https://doi.org/10.1046/j.1365-2958.1999.01548.x)
- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304. doi:[10.1038/35012500](https://doi.org/10.1038/35012500)
- Pride DT, Wassenaar TM, Ghose C, Blaser MJ (2006) Evidence of host-virus co-evolution in tetranucleotide usage patterns of bacteriophages and eukaryotic viruses. *BMC Genomics* 7:8. doi:[10.1186/1471-2164-7-8](https://doi.org/10.1186/1471-2164-7-8)
- Radstrom P, Backman A, Qian N, Kraghsbjerg P, Pahlson C, Olcen P (1994) Detection of bacterial DNA in cerebrospinal fluid by an assay for simultaneous detection of *Neisseria meningitidis*, *Haemophilus influenzae*, and streptococci using a seminested PCR strategy. *J Clin Microbiol* 32:2738–2744
- Reznikoff WS (2003) Tn5 as a model for understanding DNA transposition. *Mol Microbiol* 47:1199–1206. doi:[10.1046/j.1365-2958.2003.03382.x](https://doi.org/10.1046/j.1365-2958.2003.03382.x)
- Rhodes G, Parkhill J, Bird C, Ambrose K, Jones MC, Huys G et al (2004) Complete nucleotide sequence of the conjugative tetracycline resistance plasmid pFBAOT6, a member of a group of IncU plasmids with global ubiquity. *Appl Environ Microbiol* 70:7497–7510. doi:[10.1128/AEM.70.12.7497-7510.2004](https://doi.org/10.1128/AEM.70.12.7497-7510.2004)
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schluter A, Szczepanowski R, Puhler A, Top EM (2007) Genomics of IncP-1 antibiotic resistance plasmids isolated from wastewater treatment plants provides evidence for a widely accessible drug resistance gene pool. *FEMS Microbiol Rev* 31:449–477. doi:[10.1111/j.1574-6976.2007.00074.x](https://doi.org/10.1111/j.1574-6976.2007.00074.x)
- Schneider S, Keller M, Droge M, Lanka E, Puhler A, Selbitschka W (2001) The genetic organization and evolution of the broad host range mercury resistance plasmid pSB102 isolated from a microbial population residing in

- the rhizosphere of alfalfa. *Nucleic Acids Res* 29:5169–5181. doi:[10.1093/nar/29.24.5169](https://doi.org/10.1093/nar/29.24.5169)
- Sota M, Tsuda M, Yano H, Suzuki H, Forney LJ, Top EM (2007) Region-specific insertion of transposons in combination with selection for high plasmid transferability and stability accounts for the structural similarity of IncP-1 plasmids. *J Bacteriol* 189:3091–3098. doi:[10.1128/JB.01906-06](https://doi.org/10.1128/JB.01906-06)
- Suzuki H, Sota M, Brown CJ, Top EM (2008) Using Mahalanobis distance to compare genomic signatures between bacterial plasmids and chromosomes. *Nucleic Acids Res* 36:e147. doi:[10.1093/nar/gkn753](https://doi.org/10.1093/nar/gkn753)
- Szipirer C, Top EM, Couturier M, Mergeay M (1999) Retro-transfer or gene capture: a feature of conjugative plasmids, with ecological and evolutionary significance. *Microbiology* 145:3321–3329
- Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247–250. doi:[10.1111/j.1574-6968.1999.tb13575.x](https://doi.org/10.1111/j.1574-6968.1999.tb13575.x)
- Tauch A, Schneiker S, Selbitschka W, Puhler A, van Overbeek LS, Smalla K et al (2002) The complete nucleotide sequence and environmental distribution of the cryptic, conjugative, broad-host-range plasmid pIPO2 isolated from bacteria of the wheat rhizosphere. *Microbiology* 148:1637–1653
- Tett A, Spiers AJ, Crossman LC, Ager D, Ciric L, Dow JM et al (2007) Sequence-based analysis of pQBR103; a representative of a unique, transfer-proficient mega plasmid resident in the microbial community of sugar beet. *ISME J* 1:331–340
- Thomas CM (ed) (2000a) The horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers, Amsterdam
- Thomas CM (2000b) Paradigms of plasmid organization. *Mol Microbiol* 37:485–491. doi:[10.1046/j.1365-2958.2000.02006.x](https://doi.org/10.1046/j.1365-2958.2000.02006.x)
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680. doi:[10.1093/nar/22.22.4673](https://doi.org/10.1093/nar/22.22.4673)
- Tobes R, Pareja E (2006) Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genomics* 7:62. doi:[10.1186/1471-2164-7-62](https://doi.org/10.1186/1471-2164-7-62)
- Top EM, De Smet I, Verstraete W, Dijkmans R, Mergeay M (1994) Exogenous isolation of mobilizing plasmids from polluted soils and sludges. *Appl Environ Microbiol* 60:831–839
- Top EM, Moënne-Loccoz Y, Pembroke T, Thomas CM (2000) Phenotypic traits conferred by plasmids. In: Thomas CM (ed) The horizontal gene pool: bacterial plasmids and gene spread. Harwood Academic Publishers, Amsterdam, pp 249–285
- van Elsas JD, Gardener BB, Wolters AC, Smit E (1998) Isolation, characterization, and transfer of cryptic gen mobilizing plasmids in the wheat rhizosphere. *Appl Environ Microbiol* 64:880–889
- Van Houdt R, Monchy S, Leys N, Mergeay M (2009) New mobile genetic elements in the genome of *C. metallidurans* CH34, their possible roles and their occurrence in other bacteria. *Antonie van Leeuwenhoek* (in press, this issue)
- van Passel MW, Bart A, Luyf AC, van Kampen AH, van der Ende A (2006) Compositional discordance between prokaryotic plasmids and host chromosomes. *BMC Genomics* 7:26. doi:[10.1186/1471-2164-7-26](https://doi.org/10.1186/1471-2164-7-26)