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

Susanne Schneiker, Mathias Keller, Marcus Dröge, Erich Lanka¹, Alfred Pühler* and Werner Selbitschka

Fakultät für Biologie, Lehrstuhl für Genetik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany and ¹Max-Planck-Institut für Molekulare Genetik, Dahlem, D-14195 Berlin, Germany

Received June 28, 2001; Revised October 8, 2001; Accepted October 22, 2001 DDBJ/EMBL/GenBank accession no. AJ304453

ABSTRACT

Employing the biparental exogenous plasmid isolation method, conjugative plasmids conferring mercury resistance were isolated from the microbial community of the rhizosphere of field grown alfalfa plants. Five different plasmids were identified, designated pSB101–pSB105. One of the plasmids, pSB102, displayed broad host range (bhr) properties for plasmid replication and transfer unrelated to the known incompatibility (Inc) groups of bhr plasmids IncP-1, IncW, IncN and IncA/C. Nucleotide sequence analysis of plasmid pSB102 revealed a size of 55 578 bp. The transfer region of pSB102 was predicted on the basis of sequence similarity to those of other plasmids and included a putative mating pair formation apparatus most closely related to the type IV secretion system encoded on the chromosome of the mammalian pathogen Brucella sp. The region encoding replication and maintenance functions comprised genes exhibiting different degrees of similarity to RepA, KorA, IncC and KorB of bhr plasmids pSa (IncW), pM3 (IncP-9), R751 (IncP-1β) and RK2 (IncP-1α), respectively. The mercury resistance determinants were located on a transposable element of the Tn5053 family designated Tn5718. No putative functions could be assigned to a quarter of the coding capacity of pSB102 on the basis of comparisons with database entries. The genetic organization of the pSB102 transfer region revealed striking similarities to plasmid pXF51 of the plant pathogen Xylella fastidiosa.

INTRODUCTION

The deliberate release of genetically engineered microorganisms (GEMs) into the environment has raised concern about the potential negative ecological impact of the organisms and the nucleic acid they carry. One concern is related to the genetic stability of released GEMs. The acquisition of genes from the indigenous bacterial community by released GEMs could alter their properties and may modulate the potential ecological effects exerted by GEMs. To assess the nature of such events the acquisition of mercury resistance by deliberately released strains has been studied.

Genetic determinants conferring resistance to mercury (Hg²⁺) are ubiquitous in bacterial species of various natural environments (1). Mercury resistance genes (mer) are often located on mobile genetic elements such as transposons or conjugative plasmids. Self-transmissible plasmids, some of which have broad host range (bhr) properties, have been isolated from bulk soil, rhizosphere soils and the phyllosphere of plants employing Pseudomonas spp. as recipients (2–4). During a deliberate release experiment with a Pseudomonas strain tagged with the lacZY genes, acquisition of mercury resistance plasmids in the phyllosphere of sugar beet was readily observed (5). The acquisition of such plasmids was correlated with a specific stage in sugar beet development and was shown to temporarily increase the ecological fitness of the GEM compared to its plasmid-free parent strain (6).

In the context of a joint project the first deliberate release of GEMs was performed in Germany (7). The GEMs were derivatives of Sinorhizobium meliloti strain 2011 genetically tagged with the firefly luciferase gene (lac) mediating bioluminescence (8). These bacteria were released in 1995 at field plots of the Federal Agricultural Research Center (FAL, Braunschweig), mainly for monitoring purposes. The question arose whether self-transmissible plasmids conferring mercury resistance present at the release site were able to transfer to and be
maintained in *S. meliloti* and whether these were related to plasmids already characterized.

Thus the main objective of this work was the identification and characterization of conjugative mercury resistance plasmids residing in the microbial population of the alfalfa rhizosphere. Self-transmissible plasmids conferring Hg resistance were exogenously isolated employing a *S. meliloti* recipient. One environmental mercury resistance plasmid, designated pSB102, which exhibited bhr properties was selected for sequence analysis. The molecular structure of this plasmid pSB102 as well as its relatedness to other plasmids and chromosomal DNA fragments are reported.

**MATERIALS AND METHODS**

**Bacterial strains and plasmids**

Bacterial strains and plasmids used in this study are listed in Table 1. *Escherichia coli* strains and *Pseudomonas* sp. B131 and its derivatives were grown in Luria–Bertani (LB) medium at 37°C and 30°C, respectively. *Sinorhizobium meliloti* strains were grown at 30°C in Tryptone yeast medium. The final concentrations of antibiotics for selective growth were 200 µg ml⁻¹ streptomycin and 100 µg ml⁻¹ rifampicin. Bacterial strains were stored at –20°C in 50% glycerol.

**Construction of the *S. meliloti* strain FP2 tagged with green fluorescent protein (GFP)**

The gfp marker gene was introduced into *S. meliloti* strain M4 using suicide plasmid pAG308-1 (12). Plasmid pAG308-1 contains a Tn5-derivated mini transposon, carrying a promoterless gfp gene and a kanamycin resistance gene. pAG308-1 was mobilized from *E. coli* strain S17-1 pir to *S. meliloti* M4 in filter matings. Neomycin-resistant transconjugants were selected on TY agar containing 50 µg ml⁻¹ neomycin and 7 µg ml⁻¹ nalidixic acid. Transconjugants arose at a frequency of ~4.4 × 10⁸ per recipient cell. A selected colony designated FP2 exhibited strong fluorescence, indicating that the promoterless gfp gene was under the control of a strong indigenous *S. meliloti* promoter (data not shown).

**Exogenous plasmid isolation from the bacterial community of the alfalfa rhizosphere**

 Conjugal plasmids of the indigenous bacterial community inhabiting the root surface of field grown alfalfa plants were isolated using gfp-tagged *S. meliloti* strain FP2 as recipient in filter crosses. For each mating 20 g root material from two different field plots at the FAL in Braunschweig was collected. The roots were cut into pieces and transferred to 200 ml Erlenmeyer flasks. An aliquot of 100 ml of PBS buffer (2) was added and the flasks were shaken for 20 min at 250 r.p.m. The supernatant was centrifuged at 5 min at 1000 g to remove soil particles. Bacterial cells were pelleted by 10 min centrifugation of the supernatant at 8000 g. The pellet was resuspended in 2 ml of PBS. In the first mating 0.6 ml of rhizosphere suspension, corresponding to 3.5 g root material, was mixed o/n (overnight) with 0.3 ml of cultures of *S. meliloti* FP2. In the second mating 1 ml of rhizosphere suspension, corresponding to 5 g root material, was mixed o/n with 0.5 ml of culture of *S. meliloti* FP2. The mating mixture as well as the appropriate controls of 1 ml of rhizosphere suspension and 1 ml of recipient culture were concentrated by 30 s centrifugation at 14 000 g and were spotted on nitrocellulose filters (Sartorius AG, Göttingen, Germany). Filters were incubated at 30°C for 24 h on R2A agar plates (Difco Laboratories, Detroit, MI) supplemented with 100 µg ml⁻¹ cycloheximide. The titer of rhizosphere bacteria was determined on R2A supplemented with 100 µg ml⁻¹ cycloheximide. Mercury resistance served as the selection marker for plasmid transfer and streptomycin was used to counterselect against the bacterial community. Transconjugants were selected on TY medium containing 100 µg ml⁻¹ cycloheximide, 200 µg ml⁻¹ streptomycin and 20 µg ml⁻¹ mercury(II) chloride (HgCl₂). *Sinorhizobium meliloti* transconjugants were identified on the basis of their green fluorescence phenotype under UV irradiation.

**Plasmid conjugation tests by filter matings**

An aliquot of 1 ml of a donor culture was mixed o/n with 0.5 ml of a culture of the recipient strain and concentrated by 30 s centrifugation at 14 000 g. Matings were performed as described above, using filters placed on TY or LB medium. Serial dilutions of mating mixtures were plated onto TY or LB plates supplemented with the appropriate antibiotics and heavy metals.

**Analysis of plasmids for antibiotic and heavy metal resistances**

Five milliliters of TY medium containing various concentrations of the selective agent under investigation were inoculated with 50 µl of a late logarithmic culture of each plasmid-harboring strain. After incubation o/n the OD₅₆₀ values of the cultures were determined and compared with OD₅₆₀ values of the control cultures of *S. meliloti* FP2 lacking the corresponding plasmid. The following compounds were used in resistance tests: 0–10 µg ml⁻¹ tetracycline, 0–36 µg ml⁻¹ amikacin, 0–15 µg ml⁻¹ gentamycin, 0–70 µg ml⁻¹ spectinomycin, 0–120 µg ml⁻¹ chloramphenicol, 0–150 µg ml⁻¹ novobiocin, 0–10 µg ml⁻¹ nourseothricin, 0–5 µg ml⁻¹ norfloxacin, 0–30 µg ml⁻¹ rifampicin, 0–40 µg ml⁻¹ HgCl₂, 0–20 µg ml⁻¹ phenylmercuric acetate (PMA, C₅H₅HgOCOCH₃), 0–0.5 mM cadmium sulfate (CdSO₄), 0–3 mM cobalt chloride (CoCl₂), 0–10 mM copper sulfate (CuSO₄), 0–20 mM ferric chloride (FeCl₃), 0–3 mM nickel sulfate (NiSO₄) and 0–5 mM zinc sulfate (ZnSO₄).

**Subcloning, sequencing and annotation of plasmid pSB102**

To determine the DNA sequence of pSB102, plasmid DNA was digested with EcoRI and ligated with EcoRI-linearized pBluescript II KS(+) vector. Plasmid DNA was isolated using a modified HB lysis method (13). The protocol was modified to include purification by CsCl–ethidium bromide buoyant density centrifugation (14). The preparation of competent *E. coli* cells and the transformation procedures were performed according to Sambrook *et al*. (15). Ten EcoRI fragments ranging in size from 0.5 to 11 kb were cloned in vector pBlue- script II KS(+). The ends of the EcoRI fragments were sequenced with M13 standard forward and reverse primers. Using this sequence information, two types of specific walking primers were designed. One type was used for walking on the cloned fragments. The second type, outwardly directed from the EcoRI fragments, was employed for walking on pSB102 DNA to connect the fragments. For sequencing purposes...
plasmid DNA was isolated using the Nucleobond AX 100 kit (Macherey-Nagel, Düren, Germany) according to the protocol provided by the manufacturer. DNA was sequenced by the dideoxy chain termination method (16) using PCR-based sequencing with a Taq Dye Deoxy Terminator Cycle-Sequencing Kit on an ABI Prism 377 DNA Sequencer (Applied Biosystems, Perkin Elmer) by IIT Biotech/Bioservice GmbH (Bielefeld, Germany). Base calling was done using PHRED. Walking primers were designed with PRIDE from the Staden package. Sequence assembly was done with the GAP4 sequence assembly tool from the Staden package (17). Annotation was performed using the automated genome interpretation system GenDB (Center for Genome Research, Bielefeld, Germany). Open reading frames (ORFs) were predicted by GLIMMER 2.0 (18). Each identified ORF was the subject of BLAST database similarity searches (19). The complete sequence was searched for $\sigma^{70}$-dependent promoters using Neural Network Promoter Prediction (NNPP) (20). Repeat regions within the pSB102 sequence were analyzed with the computer program REPuter (21). Prediction of transmembrane helices was performed by the dense alignment surface (DAS) method using the DAS transmembrane prediction server.
(22). Amino acid sequence similarities were calculated with the ALIGN computer program (23). Prediction of signal peptides was performed with the neural network systems SignalP (24). The sequence of plasmid pSB102 is deposited in the EMBL nucleotide sequence database under accession no. AJ304453.

Classification of plasmids by PCR analyses

Template DNA was obtained from single colonies of *S. meliloti*. For this purpose bacterial colonies were suspended in 50 μl of lysis buffer (0.5 M NaOH, 0.25% SDS) and boiled for 5 min. An aliquot of 1.5 μl of a 10-fold dilution of the boiled sample was used in PCR. PCR was carried out in a volume of 25 μl, containing 1.6 μM both forward and reverse primers, 0.2 mM each dNTP, 1 U *Taq* polymerase (Qiagen, Hilden, Germany) and 1× PCR buffer (Qiagen). After an initial step of 5 min at 94°C, PCR amplification involved 30 cycles of 1 min at 94°C for denaturation, 1 min at the annealing temperature recommended for the respective primer system and 1 min polymerisation at 72°C. The final primer extension reaction was done for 10 min at 72°C. PCR products were analyzed in 1.5% agarose gels with TA buffer (40 mM Tris, 10 mM sodium acetate, 1 mM EDTA). Primers employed for PCR-based classification were specific for incompatibility (Inc) groups A/C (25), N, P and W (26). PCR analyses were performed using plasmids RA1, RN3, R702 and pSa (Table 1) as positive controls.

RESULTS AND DISCUSSION

Isolation and characterization of conjugal mercury resistance plasmids from the microbial community of the alfalfa rhizosphere

To assess the presence of conjugal mercury resistance plasmids at a field site where *S. meliloti* GEMS tagged with the firefly luc gene had been released, the exogenous plasmid isolation method was employed. Using the alfalfa bacterial rhizosphere communities, sampled twice during November 1998 and July 1999 from two different field plots, as donors in filter matrices with *gfp*-tagged *S. meliloti* strain FP2 as recipient, mercury-resistant colonies exhibiting the GFP phenotype were readily recovered. According to their *Eco*RI restriction pattern a total of five different plasmids designated pSB101–pSB105 were identified (Table 2). During the first mating in November 1998 two plasmids were isolated, which differed in their *Eco*RI restriction fragment patterns. Transconjugants obtained from the plot I sample carried plasmid pSB101 and transconjugants obtained from the plot II sample carried plasmid pSB102. In the second mating performed in July 1999 five distinct plasmids were recovered. Transconjugants obtained from the plot I sample carried plasmids pSB102 and pSB104 and transconjugants obtained from the plot II sample carried plasmids pSB101, pSB103 and pSB105. The fact that plasmids pSB101 and pSB102 were repeatedly recovered over time at a spatial distance of at least 3 m may indicate that these two plasmids were abundant in bacterial communities of the field site.

As a first step in the characterization of the newly identified plasmids, their sizes were determined on the basis of their restriction fragment patterns in agarose gels. These analyses revealed sizes ranging from 52 to 75 kb (Table 2). To address the question of whether additional resistances to antibiotics or heavy metals were located on the various plasmids, resistance tests were performed using the antibiotics and heavy metals listed in Materials and Methods. All plasmids conferred resistance to inorganic (HgCl₂) and organic (PMA) mercury compounds. Plasmids pSB103 and pSB104 were resistant to 10 µg ml⁻¹ HgCl₂ but were unable to grow at 20 µg ml⁻¹ HgCl₂, whereas all other plasmids conferred resistance to 20 µg ml⁻¹ HgCl₂. Plasmids pSB101 and pSB102 showed resistance to 10 µg ml⁻¹ PMA, while all the other plasmids were only able to grow at 4 µg ml⁻¹ PMA. Plasmid-free *S. meliloti* FP2 was unable to grow in liquid medium at 5 µg ml⁻¹ HgCl₂ or 2 µg ml⁻¹ PMA. None of these five plasmids displayed resistance against the antibiotics tetracycline, amikacin, gentamicin, spectinomycin, chloramphenicol, novobiocin, norfloxacin, nurse-othricin and rifampicin or heavy metal ions of cadmium, cobalt, copper, iron, nickel and zinc.

To characterize the host range of the newly isolated plasmids, mating experiments were performed. *Sinorhizobium melloti* FP2 served as the donor strain and three different bacterial strains, *S. meliloti* L331 belonging to the α-Proteobacteria and *Pseudomonas* sp. strain B131 and *E.coli* strain HBR101 belonging to the γ-Proteobacteria, were used as recipients. As expected, all plasmids were self-transmissible. The transfer frequencies were in the range 10⁻³–10⁻⁵ (Table 2). Plasmid pSB102 was transferred to all recipients used and thus displayed bhr properties, whereas the host range of all other plasmids was restricted to *S. melloti* (Table 2). Employing a Tn5-tagged derivative of plasmid pSB102, transfer from *E.coli* to *Ralstonia eutropha* strain 11098, which belongs to the β-Proteobacteria, was also observed (unpublished results). PCR replicon typing employing primers specific for Inc groups IncP-1, IncW, IncN and IncA/C, respectively, yielded no PCR products, suggesting that plasmid pSB102 belongs to an as yet undescribed group of bhr plasmids.

Sequence analysis of bhr mercury resistance plasmid pSB102 employing an approach which minimizes the number of sequencing reactions

The 55 578 bp sequence of plasmid pSB102 was determined by primer walking. Employing this strategy the entire

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Size (in kb)</th>
<th>S. meliloti</th>
<th>Pseudomonas</th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSB101</td>
<td>56</td>
<td>10⁻⁷</td>
<td>n.d. (&lt;10⁵)</td>
<td>n.d. (&lt;10⁵)</td>
</tr>
<tr>
<td>pSB102</td>
<td>56</td>
<td>10⁻⁷</td>
<td>10⁻¹</td>
<td>10⁻⁴</td>
</tr>
<tr>
<td>pSB103</td>
<td>72</td>
<td>10⁻⁴</td>
<td>n.d. (&lt;10⁵)</td>
<td>n.d. (&lt;10⁵)</td>
</tr>
<tr>
<td>pSB104</td>
<td>52</td>
<td>10⁻⁴</td>
<td>n.d. (&lt;10⁵)</td>
<td>n.d. (&lt;10⁵)</td>
</tr>
<tr>
<td>pSB105</td>
<td>75</td>
<td>10⁻⁷</td>
<td>n.d. (&lt;10⁵)</td>
<td>n.d. (&lt;10⁵)</td>
</tr>
</tbody>
</table>

*The frequencies of conjugative transfers are given per recipient cell. *S. meliloti* strain FP2 served as the donor and *S. meliloti* strain L331, *Pseudomonas* sp. strain B131 and *E.coli* strain HBR101 served as recipients in the respective mating experiments. Transconjugants were selected on LB or TY medium containing 20 µg ml⁻¹ HgCl₂ and 100 µg ml⁻¹ rifampicin. The values given in parentheses indicate the limit of detection. n.d., not detected.*
double-stranded DNA sequence was established with a minimal number of 260 sequencing reactions, leading to a 2.5-fold coverage of the genome. The average GC content of the plasmid is 59.8%. No obvious significant deviations in the GC content were observed. Potential ORFs were identified with an interpolated Markov model using GLIMMER2.0 (18). GLIMMER2.0 predicted a total of 101 frequently overlapping ORFs. The annotation of plasmid pSB102 was performed using the automated genome interpretation system GenDB. Taking the results of database searches and the existence of putative Shine–Dalgarno sequences (Table 3) into account, the number of ORFs was reduced from 101 to 66. BLASTX searches with short sequence segments of pSB102 revealed the existence of three further ORFs (ssb, traC and traD), which were not predicted by GLIMMER2.0.

Significant similarities to database entries for 46 of a total of 69 hypothetical ORF-encoded proteins were identified. Putative functions could be assigned to 39 ORFs of pSB102. These genetic functions comprise genes for: (i) mating pair formation (Mpf); (ii) DNA replication and replication (Dtr); (iii) pilus formation (Pil); and (iv) a mercury resistance transposon. The relevant information about putative ORFs of pSB102, such as GC content of coding regions, number of amino acid residues and molecular weight, as well as the closest relatives among protein sequences stored in the databases, are summarized in Table 3. The genetic map of pSB102 is shown in Figure 1. The complete sequence was searched for σ70-dependent promoters using the NNPP program (20). The NNPP program predicted a total of 110 putative promoters using a threshold score of 0.8, which recognizes 60% of all promoters, with a rate of false positive predictions of 0.4%. Finally, 14 putative pSB102 promoters were selected from the NNPP predicted promoters by the logic of gene organization (Table 4 and Fig. 1). One promoter, P\(_{\text{psec}}\), was chosen due to conformity with published promoter/operator sequences of mercury resistance operons.

**Putative mating pair formation genes of plasmid pSB102**

show highest similarity to *Brucella* sp. virB genes encoding a type IV secretion system

Transfer of conjugative plasmids requires the interaction of a number of cell surface components, to form a mating bridge between donor and recipient strains. The putative Mpf genes of pSB102, designated traA to traM, show highest similarity to the virB gene cluster of *Brucella suis* (27) and *Brucella abortus* (28; Fig. 2). Moreover, significant similarity was noted to the tra genes of plasmid pXF51 (29), transfer (tra) genes of bhr plasmid pKM101 and the virB gene cluster of the *Agrobacterium tumefaciens* plasmid pTiA6.

**traA**, **traC**. The predicted pSB102 TraA protein contains the amino acid motifs 2\(_8\)AVHVES\(_{75}\), 9\(_9\)GAQV\(_{73}\) and 12\(_{12}\)YSSG\(_{30}\) related to those of blocks A, B and C associated with the active site of the *E.coli* soluble lytic transglycosylase (30). The core region of propilin-like proteins, such as TrbC of IncP-1 plasmids, was found to be highly conserved, including a LepB cleavage site of *E.coli* leader peptides I and two transmembrane helices in the central part (31). The pSB102 TraC protein contains an N-terminal peptide cleavage site (\(_{58}\)AFA\(_{63}\)GQ\(_{80}\)) at positions 33–34 and two predicted transmembrane helices in the remaining part of the protein (positions 49–66 and 84–91). Hence, pSB102 TraC may be post-translationally processed at the N-terminus, by removal of a 33 amino acid leader peptide and cyclization. IncP-1-like C-terminal maturation of the pSB102 prophilin by a TraF-like peptidase might occur, because a cyclization recognition site (\(_{58}\)AE\(_L_{59}\)A) is conserved near the C-terminus (31).

**traE**, **traF**, **traH**. The pSB102 TraE protein contains a Walker A nucleoside triphosphate (NTP)-binding domain 49\(_{49}\)GQSGAGKTV\(_{65}\) (32) which is conserved in VirB4 homologs. The *Agrobacterium* VirB4 protein is tightly associated with the cytoplasmic membrane and is proposed to form dimers (33). The traF encoded protein of pSB102 displays a putative signal peptide cleavage site at residues 20–21 (\(_{19}\)ANA\(_{20}\) QE\(_{21}\)), indicating that it may be exported. The TraH protein of plasmid pSB102 contains five putative transmembrane domains as predicted with the DAS transmembrane prediction server (22).

**traI**. TraI of pSB102 contains a lipoprotein consensus (\(_{12}\)LAGC\(_{17}\)) sequence, suggesting that, like *Agrobacterium* VirB7, it is anchored to the outer membrane. In *Agrobacterium* the periplasmic VirB9 protein interacts via disulfide bridges with the VirB7 lipoprotein. The pSB102 TraF protein contains a single cysteine residue which is located in the putative signal sequence, whereas the plasmid pSB102 VirB9-like protein, TraK, lacks any cysteine residues. Therefore, if Tra and TraK of pSB102 form a complex, this interaction is not mediated by a disulfide bond. Using a yeast two-hybrid system circumstantial evidence has been presented that *Agrobacterium* VirB7 and VirB9 could also form a heterodimer without disulfide bond formation (34).

**traM**. Key players of the type IV secretion system are energy generators, probably in the form of ATPases. The pSB102 TraM protein contains four highly conserved motifs (35), which are common among the members of the VirB11 family: a nucleotide-binding site type A (Walker A Box 183\(_{183}\)AG268), an aspartate box (190\(_{190}\)RQRTIEDVPEI\(_{215}\)), a nucleotide-binding site type B (Walker B Box 251\(_{251}\)LMRKPITRLLAE\(_{260}\)) and a histidine box (253\(_{253}\)SHGGSITSLH\(_{268}\)). Using the GeNomenclature classification system for hexameric NTPases (36; http://cpcmnet.columbia.edu/dept/figurski/genomenclature/), which allows the online submission of VirB11 homologous proteins, the pSB102 TraM protein grouped to the VirB11 subfamily of NTPases. Interestingly, the VirB11 subfamily exclusively comprises NTPases implicated in the delivery of virulence factors, such as pertussis toxin, or the *A.tumefaciens*-mediated delivery of oncogenic T-DNA to plants.
Table 3. Location of the coding regions of pSB102 and the closest relatives of the deduced proteins

<table>
<thead>
<tr>
<th>ORF</th>
<th>Gene</th>
<th>Position</th>
<th>putative Shine-Dalgarno sequence</th>
<th>GC content (%)</th>
<th>Protein length (aa)</th>
<th>Protein size (kDa)</th>
<th>Closest relationship</th>
<th>Accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF1</td>
<td>repI</td>
<td>44 1498</td>
<td>ccttAAGGcagggcgagcATGcctt</td>
<td>61.1</td>
<td>464</td>
<td>54.0</td>
<td>replication initiation protein Skhigella flavina pSB124 (24%)</td>
<td>Q04599</td>
</tr>
<tr>
<td>ORF2</td>
<td>sop</td>
<td>1512 1835</td>
<td>tttgcccAGCattacaAGTGGgta</td>
<td>54.0</td>
<td>113</td>
<td>12.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF3</td>
<td>hsp</td>
<td>1846 2397</td>
<td>cagcaggagagacagcagcagcagcagcagc</td>
<td>62.6</td>
<td>452</td>
<td>50.7</td>
<td>Yersiella fastidiosa pXPF51 (39%)</td>
<td>Q9PH6</td>
</tr>
<tr>
<td>ORF4</td>
<td>hsp</td>
<td>3043 4401</td>
<td>ccacgTAGGgagacagcagcagcagcagcagc</td>
<td>59.9</td>
<td>374</td>
<td>15.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF5</td>
<td>hsp</td>
<td>4674 5078</td>
<td>aagcrGGAAGCAcatcctggaAGTGGgta</td>
<td>59.9</td>
<td>99</td>
<td>10.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF6</td>
<td>hsp</td>
<td>5109 5408</td>
<td>ccttgcGGAAGGcaggcagcagcagcagcagc</td>
<td>65.0</td>
<td>109</td>
<td>12.2</td>
<td>X. fastidiosa (50%)</td>
<td>Q9PB5</td>
</tr>
<tr>
<td>ORF7</td>
<td>hsp</td>
<td>5436 5765</td>
<td>cagcagcGGAAGGcaggcagcagcagcagcagc</td>
<td>58.5</td>
<td>139</td>
<td>14.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF8</td>
<td>hsp</td>
<td>5762 6079</td>
<td>cagcagcGGAAGGcaggcagcagcagcagcagc</td>
<td>56.8</td>
<td>118</td>
<td>13.4</td>
<td>single stranded binding protein E. aerogenes argR (29%)</td>
<td>Q03237</td>
</tr>
<tr>
<td>ORF9</td>
<td>hsp</td>
<td>6153 6572</td>
<td>cagcagcGGAAGGcaggcagcagcagcagcagc</td>
<td>58.3</td>
<td>139</td>
<td>14.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ORF10</td>
<td>sb</td>
<td>7900 8286</td>
<td>ccgcagcAGAAGGcaggcagcagcagcagcagc</td>
<td>58.3</td>
<td>139</td>
<td>14.4</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Raw text:

...and a predicted signal sequence cleavage site at positions 14–15 within this motif and one predicted transmembrane helix (positions 99–105). TraR shows homology to IncP-1α plasmid RP4 (Q03537) and IncP-1β plasmid R751 (P71188) encoded protein TrbM. pSB102 TraR contains an export signal sequence cleavage site between amino acids 22 and 23 (20AQA↓ED24), indicating that the TraR protein may be exported. Downstream of trbM a putative ρ-independent transcription terminator is located (positions 13660–13685).

Figure 1. Genetic map of plasmid pSB102. The 69 ORFs identified in the nucleotide sequence of plasmid pSB102 (Table 3) are located on a circular map. ORFs are shown as arrows. The numbers of the ORFs are given as the inner circle, the names of the putative genes are given as the outer circle. Putative promoters are indicated as black triangles on the inner circle. Different colors of the arrows indicate transfer genes with mating pair formation (Mpf) and DNA processing/transfer (Dtr) functions (light and deep green, respectively), genes encoding plasmid replication and maintenance (Rep and Mai) (yellow), Tn5718 encoded genes involved in mercury resistance and transposition (red and blue, respectively) as well as genes to which no similarities to database entries were detected (gray). Long DRs are indicated as light and dark brown rectangles, respectively. The putative origin of transfer (oriT) is shown as a filled circle.

...traB. In contrast to the virB gene cluster of B. suis, the 11 virB homologous transfer genes of pSB102 are intermixed with virB unrelated genes (Fig. 2). TraB shows similarity to type I and type III topoisomerases of various organisms, such as plasmid pXF51 encoded putative TopA (Table 3), TopA of E. coli K12 (P06612*), VC1730 of Vibrio cholerae (Q9KRB2*), To of Pseudomonas aeruginosa (AAG06399) and TopA of Zymomonas mobilis (Q9X3X7).

traG, ORF32. The traG gene encoding a putative entry exclusion function of pSB102 was found to be related to the Xylella fastidiosa plasmid pXF51 (29) encoded putative protein XFa0038 (Q9P9Q7), a hypothetical 8.5 kDa protein of IncW plasmid R388 (O50332), lipoprotein MagB05 (AAG24432) of Actinobacillus actinomycetemcomitans plasmid pVT745 (37) and the IncN plasmid pKM101 entry exclusion protein (Q46700) (38). The predicted pSB102 TraG protein contains a putative conserved lipoprotein modification/processing site (↓17LSG↓C20). The hypothetical ORF32 encoded protein contains four putative transmembrane domains and a putative signal peptide cleavage site (↓18LLA↓AR61) between amino acids 59 and 60. The ORF32 encoded protein may therefore function in the transmembrane transport machinery.

c, complementary DNA strand.

*The putative Shine–Dalgarno (SD) sequences of plasmid pSB102 are indicated by bold capital letters. Start codons are given in capitals. Overlaps of start codons with previous stop codons are underlined.

*Identities were overall identities at the amino acid sequence level.

Accession numbers are given in SPTREMBL format or, as indicated by *, in SWISSPROT format.
Putative genes for the processing of DNA for transfer and establishment of the plasmid in the recipient cell are scattered over plasmid pSB102

The Dtr genes of plasmid pSB102 (traB, traN, traO, traP and traS) were identified by comparison with the putative Dtr genes of plasmids pXF51 (29) and RK2/RP4 (39). They were found to be scattered over one half of plasmid pSB102 and are intermixed with Mpf genes (Figs 1 and 3). This arrangement contrasts with other conjugative bhr plasmids, such as IncP-1 or IncN, where Dtr genes are clustered in a single region (40).

oriT, ORF15, traS. The pSB102 encoded putative relaxase TraS is most closely related to the X. fastidiosa plasmid pXF51 encoded relaxase MobB (Table 3). Next nearest relatives are the A. actinomycetemcomitans plasmid pVT745 (AAG24403) encoded relaxase, TaxC of IncX plasmid R6K (P17232) and Agrobacterium encoded relaxase VirD2 (P66686*), which are all grouped into the IncP-1-type family of DNA relaxases. The pSB102 TraS protein possesses three conserved motifs of IncP-1-type relaxases (41). Upstream of the relaxase gene traS protein possesses three conserved motifs of all grouped into the IncP-1-type family of DNA relaxases. The encoded relaxase VirD2 (P06668*), which are encoded relaxase, TaxC of IncX plasmid R6K (P17232) and the encoded relaxase MobB (Table 3). Next nearest relatives are X. fastidiosa traS is most closely related to the oriT, ORF15, traS.

The pSB102 encoded putative relaxase TraS is most closely related to the X. fastidiosa plasmid pXF51 encoded relaxase MobB (Table 3). Next nearest relatives are the A. actinomycetemcomitans plasmid pVT745 (AAG24403) encoded relaxase, TaxC of IncX plasmid R6K (P17232) and Agrobacterium encoded relaxase VirD2 (P66686*), which are all grouped into the IncP-1-type family of DNA relaxases. The pSB102 TraS protein possesses three conserved motifs of IncP-1-type relaxases (41). Upstream of the relaxase gene traS protein possesses three conserved motifs of IncP-1-type relaxases (41).

The putative replication gene of plasmid pSB102 is related to the replication gene of the broad host range plasmid pSa repA. The pSB102 encoded RepA protein shows highest similarity to RepA of IncW plasmid pSa (Table 3). The pSB102 RepA protein is unusually large (484 amino acids), compared to the pSa RepA protein (333 amino acids). An alignment of these two sequences indicated an addition at the N-terminus of the pSB102 encoded RepA. The C-terminal

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>7853</td>
<td>7893</td>
<td></td>
</tr>
<tr>
<td>ORF15</td>
<td>11401</td>
<td></td>
</tr>
<tr>
<td>ORF17</td>
<td>13136</td>
<td></td>
</tr>
<tr>
<td>traQ</td>
<td>14934</td>
<td></td>
</tr>
<tr>
<td>ORF22</td>
<td>15506</td>
<td></td>
</tr>
<tr>
<td>traN</td>
<td>21419</td>
<td></td>
</tr>
<tr>
<td>traI</td>
<td>26161</td>
<td></td>
</tr>
<tr>
<td>ORF39</td>
<td>36107</td>
<td></td>
</tr>
<tr>
<td>ORF40</td>
<td>38362</td>
<td></td>
</tr>
<tr>
<td>ORF41</td>
<td>38341</td>
<td></td>
</tr>
<tr>
<td>ORF53</td>
<td>43672</td>
<td></td>
</tr>
<tr>
<td>morR</td>
<td>44295</td>
<td></td>
</tr>
<tr>
<td>morT</td>
<td>44231</td>
<td></td>
</tr>
<tr>
<td>morI</td>
<td>50174</td>
<td></td>
</tr>
<tr>
<td>ORF68</td>
<td>54142</td>
<td></td>
</tr>
</tbody>
</table>


ORF2, ORF3, ORF68, ORF69. Hypothetical ORF2 and ORF3 are located 13 bp downstream of the pSB102 repA gene. ORF2 overlaps the start of ORF3 by 8 bp (ATGCTAA) (Table 3). The putative proteins encoded by ORF2 and ORF3 do not display any significant similarity to database entries. Hypothetical ORF68 and ORF69, preceded by putative promoter P ORFSH, are located upstream of the pSB102 repA gene.

The putative maintenance region of plasmid pSB102 shows similarities to the central control region of IncP-1 plasmids

Putative pSB102 Mai genes are packed close together and are preceded upstream by putative promoter P ssb (Table 4). ORF11, ORF12 and ORF13 are arranged in the same order in pSB102 as the related genes korA, incC and korB in the central control region of IncP-1 β plasmids RK2 and RP4 (Fig. 4).

ssb. The first gene in the pSB102 Mai region, the ssb gene, encodes a single-stranded DNA-binding protein. Nearest relatives of the putative pSB102 SSB protein are the SSB proteins of RK2 (Q02327) and R751 (Q56467). Plasmid-encoded SSB proteins presumably reduce SSB starvation in cells receiving single-stranded DNA during conjugation (40).

ORF11, ORF12, ORF13. The ORF11 encoded protein shows weak similarity to the KorA protein of Pseudomonas putida.
IncP-9 plasmid pM3, whereas the ORF12 encoded protein is most closely related to KorB of IncP-1α plasmid RK2 (Table 3). The ORF13 encoded protein carries three conserved ATP-binding motifs, residues 10 KGGVGK17, 29 LRRNKVLVIDF39 and 115 FDICIDTNP124, related to the respective motifs of short IncC proteins of RK2/RP4 and R751. Both plasmids encode two in-frame IncC products which differ in their length. The short IncC protein of IncP-1 plasmids was shown to support partitioning of plasmids (47).

**ORF53.** Plasmid pSB102 ORF53 encodes another plasmid-stabilizing protein, a DNA resolvase, which may be involved in multimer resolution. The pSB102 ORF53 encoded protein shows 72% identity to ParA of RP4 (Table 3), which is part of the parABCDE multimer resolution system. Plasmid pSB102 specifies a distinct res site to RP4, as the deduced consensus sequence of the RP4 res site (48) was not found upstream of the pSB102 parA gene. A mercury resistance encoding transposon grouping to the Tn5053/Tn402 family, whose members are known as res site hunters (49), is located 265 bp upstream of ORF53.

**The mercury resistance transposon Tn5718 belongs to the Tn5053 family of transposons**

Plasmid pSB102 carries a mercury resistance transposon, designated Tn5718, which is 10 414 bp in size. It is delimited by 25 bp inverted repeats (IRs) and flanked by 5 bp direct repeats (DRs). Tn5718 contains 14 putative ORFs, nine of which were identified as mer genes and five of these genes, named tniA, tniB, tniQ, tniR and tniM (Fig. 1), are putatively involved in Tn5718 transposition. The 25 bp IRs of Tn5718 show significant similarity to integron repeats (50), which are also found as IRs at the ends of mercury resistance transposon Tn5090/Tn402 (52). Transposons Tn5053 and Tn5090 have been shown to display an unusual strict insertion preference for the non-transcribed spacer of the par locus of IncP-1 plasmid RP1 (49). A 38 nt Tn21-like IR element is fused to merR in the same way as the mer-proximal IR in Tn501 and Tn21. The fusion of
a TnI-like relic IR was reported for other mercury resistance operons (51), such as Tn5053, Tn5041, pKLH2 and pMR26 (53).

The Tn5718 mercury resistance gene cluster contains many of the known mer genes identified to date. Compared to the arrangement of genes in representative mer operons, the structure of the Tn5718 mer operons is one of the most complex reported so far. Figure 5 shows an alignment of the mer operons of Pseudomonas sp. K62 plasmid pMR26, Pseudomonas sp. ED23-33 transposon Tn5058 (GenBank accession no. Y17897), Serratia marcescens plasmid pDU1358, Acinetobacter calcoaceticus plasmid pKLH2, Ralstonia sp. plasmid pMER610 and Shigella flexneri plasmid R100 transposon Tn21, which are closely related to individually encoded Tn5718 mer genes.

merR. The deduced protein encoded by the merR gene is 100% identical to MerR2 of the organomercurial resistance operon 2 from plasmid pMR26 and Pseudomonas sp. ED23-33 Tn5058 (Table 3). The merR gene and the merTPGCABDE genes are transcribed divergently from a compact promoter/operator region, in which their potential −10 RNA polymerase recognition hexamers overlap by 4 bp (Pmer 44265 TAAGGT44266 and Pmer 44268 TAAACCT44283).

merT, merP, merC, merG. Peculiar to Tn5718, the gene merG, which is supposed to reduce cellular permeability to phenylmercury (54), is inserted between merP and merC. The MerT and MerC proteins are integral membrane proteins (55). Their postulated function is the transport of Hg^{2+} across the cytoplasmic membrane from the periplasmic Hg^{2+}-binding protein MerP to cytoplasmic mercuric reductase MerA. MerP of Tn5718 contains a consensus heavy metal-binding motif GMXXCXXC26, which has been correlated with the specific binding of Hg^{2+} (56). merT, merP and merC genes are frequently found in various mer operons related to Tn21, whereas the merG gene is only found in the broad spectrum mer operon of plasmid PMR26 of Pseudomonas strain sp. K-62 (53). The deduced Tn5718 MerG protein carries a signal peptide cleavage site (32ALA141YD23), like MerG from pMR26. Deletion of merG in pMR26 had no effect on inorganic mercury resistance, but rendered the bacterium more sensitive to organic mercury (54).

merD, merE. MerD, which is co-transcribed with the mer structural genes, was described as a second regulatory protein. In pDU1358 MerD down-regulates expression of the mer operon by specifically binding to the mer promoter/operator sequence (57). The function of the most promoter-distal gene, merE, which is frequently found in other mer operons, is still unknown. Tn5718 encoded merE overlaps the 3′-end of merD by 4 bp at a 5′-ATGA-3′ motif (Table 3), indicating translational coupling as a regulatory mechanism. The putative Tn5718 MerE protein contains a characteristic cysteine pair (Cys28 and Cys30) and adjacent residues (LTCPCHL) which are conserved in all MerE polypeptides, therefore MerE may play a role in mercury transport (58).

The transposition module of Tn5718 contains five putative transposition genes, designated mtiA, mtiB, mtiQ, mtiR and mtiM. With the exception of mtiB, the deduced amino acid sequences of the four remaining transposition genes show highest similarity to TniA, TniQ, TniR and Urf2 of Pseudomonas sp. ED23-33 mercury resistance transposon Tn5058 (Table 3). mtiA, mtiB, mtiQ. The Tn5718 encoded TniA protein contains the D,D(35)E motif, which is characteristic of transposases of members of the IS3/repeton superfamily (59). The predicted TniB protein of Tn5718 contains two sites, 6GPTNNNGK82 and 148MLVID152, which are potentially involved in mercury detoxification. The Tn5718 merA gene encodes a cytoplasmic redox enzyme of 561 amino acids belonging to the class I pyridine nucleotide-disulfide oxidoreductases. The gene merB is located 14 bp downstream of merA in Tn5718. The MerB encoded enzyme, organomercurial lyase, cleaves the covalent C–Hg bond of certain organomercurials.
involved in binding and hydrolysis of NTPs (52). The GTG start codon of the following gene \textit{tniQ} overlaps the TGA stop codon of \textit{tniB} (Table 3), suggesting that translational coupling may be involved in \textit{tniQ} expression. Although the function of TniQ in transposition is unknown, deletion mutation analysis with \textit{tniA}, \textit{tniB} and \textit{tniQ} of Tn5053 have shown that these genes were essential for transposition (60).

\textit{tniR}, \textit{tniM}. The fourth gene of Tn5718, \textit{tniR}, encodes a recombinase of the resolvasen invertase family. The potential resolvase (\textit{res}) region of Tn5718 is located upstream of the \textit{tniR} gene in the center of a pair of 14 bp imperfect inverted repeats (IR1, bp 50113–50126; IR2, bp 50114–50129). The sequence of the Tn5718 \textit{res} region is identical to the \textit{res} region of \textit{Pseudomonas} sp. ED23-33 transposon Tn5058 and \textit{Xylella campestris} transposon Tn5053 (60). Thus, Tn5718 may transverse via a two-step replicative pathway including co-integrate formation and resolution at the \textit{res} region. The fifth gene of the transposition module, designated \textit{tniM}, shows highest similarity to Urf2 of \textit{Pseudomonas} sp. ED23-33 mercury resistance transposon Tn5058 (Table 3), which is closely related to modulator proteins ORF2 (Q48362) of Tn501 and TnpM (P04162*) of Tn21. TnpM has been proposed to enhance transposition and to modulate co-integrate resolution in Tn21 (61).

Two clusters of hypothetical ORFs of unknown function exist on plasmid pSB102

Plasmid pSB102 contains two regions, designated regions A and B, where hypothetical ORFs of unknown function are clustered. These regions are flanked by two nearly identical pairs of 39 bp IRs (Fig. 6A). The first 39 bp IR (bp 2903–2941 and 6690–6728) is imperfect (one mismatch), whereas the second 39 bp IR (bp 36998–37036 and 42533–42571) is perfectly conserved. These two 39 bp IRs are part of two different pairs of long DRs, which are 460 and 301 bp in size, respectively (Fig. 6B). The imperfect 460 bp DR (DR1) contains 24 mismatches, while the 301 bp DR (DR2) is perfectly conserved (Fig. 6B). The function of these direct and inverted repeats remains an enigma.

Hypothetical ORFs in region A. Region A contains six putative ORFs of unknown function. The first of these hypothetical ORFs, ORF4, encodes a putative protein of 452 amino acids. The C-terminal part of the ORF4 protein is related to \textit{X. fastidiosa} plasmid pXF51 encoded hypothetical protein XFa0062 (Q9PHE6), with a size of 284 amino acids, whereas residues 78–135 of ORF4 are similar to residues 73–130 of the 297 amino acid anti-restriction protein ArcD (Q9R2K1) of IncW plasmid pSa. The downstream hypothetical ORFs, ORF5–ORF9, encode small proteins, with sizes ranging from 99 to 138 amino acids. ORF5, ORF6, ORF8 and ORF9 show no similarities to any database entries. ORF7 displays similarity to hypothetical proteins XF2067 (Table 3) and XFa0020 (Q9PH15) encoded by \textit{X. fastidiosa} genomic DNA and plasmid pXF51, respectively.

Hypothetical ORFs in region B. Region B is 5224 bp in size and contains 13 hypothetical ORFs of unknown function. The first ORF of this region, ORF40 (383 amino acids), is located on the complementary strand, whereas the remaining 12 ORFs are divergently oriented and encode proteins of rather small sizes, ranging from 49 to 166 amino acids. The deduced proteins of 11 of these ORFs, ORF40–ORF45, ORF47–ORF49, ORF51 and ORF52 did not reveal any significant similarities to database entries.

The striking clustering of genes in regions A and B encoding proteins of unknown function which display weak or no similarities to database entries rises two questions: (i) what functions are encoded by these genes; and (ii) why do the genes accumulate in these two regions of the plasmid? At present, transposon mutagenesis is under way to assess whether some of these putative proteins are involved in plasmid stability, replication or transfer.

Plasmid pSB102 defines a new family of environmental \textit{bhr} plasmids

As shown in the previous sections, plasmid pSB102 is closely related to \textit{X. fastidiosa} plasmid pXF51 (29), whose sequence was established in the course of the \textit{X. fastidiosa} genome sequencing project (62). With the exception of the replication region, both plasmids share structural characteristics (Fig. 7). The common features are as follows. (i) The presence of a putative topoisomerase encoding gene between Mpf genes coding for a putative lytic transglycosylase and putative propilin, respectively. This arrangement contrasts with the location of the topoisomerase encoding gene \textit{traE} in IncP-\(\alpha\) plasmid RP4, where \textit{traE} is located between \textit{traF} and \textit{traD} in the Dtr gene region (see Fig. 3). (ii) The orientation of the relaxase gene is unique in that transcription is oriented towards the other Dtr genes. In those \textit{bhr} plasmids characterized at the nucleotide sequence level to date, the relaxase genes are oriented in the same direction as the other Dtr genes (Fig. 3). (iii) The organization of the Mai region is nearly the same on both plasmids. Homologs of the IncP-1-related korA, \textit{incC} and \textit{korB} genes (Fig. 4) are also encoded by pXF51. Interestingly, this common set of survival genes is preceded by the gene \textit{xsb} on both plasmids. This finding may be indicative of
the importance of single-stranded DNA-binding proteins in the maintenance of or establishment of plasmids in their bacterial hosts. (iv) Both plasmids carry at least one pair of unrelated DRs spanning several hundred base pairs. Whether these repeated sequences are structural components of plasmids involved in transfer, replication or maintenance or whether they confer some mobility to the ORFs they border remains an open question.

In conclusion, plasmid pSB102, together with pXF51, is representative of a new class of environmental plasmids. Surprisingly, both plasmids reside in plant-associated bacteria, either in bacterial communities inhabiting the plant rhizosphere (pSB102) or in a plant pathogenic organism (pXF51). This may indicate that both plasmids are variants of an archetypical plasmid class associated with phytosphere bacteria. Therefore, such plasmids may well contribute to the genomic plasticity of plant-associated bacteria.

ACKNOWLEDGEMENTS

This work was supported by grants from the BMBF (BE 03111202 and BEO 0311952) and EU Bio4 CT98-0483 (‘TRAFFIC’ project). Support of the EU for EU-BIOTECH Concerted Action Bio4-CT-099 ‘MECBAD’, which provided opportunities for constructive discussions, is gratefully acknowledged.

REFERENCES