Two Large, Related, Cryptic Plasmids from Geographically Distinct Isolates of *Sulfobacillus thermotolerans*\(^\dagger\)

S. M. Deane* and D. E. Rawlings

Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

Received 7 July 2011/Accepted 3 September 2011

Two large cryptic plasmids (59.2 and 65.9 kb) from isolates of *Sulfobacillus thermotolerans* from Yellowstone National Park (United States) and the Caribbean island of Montserrat were isolated and sequenced. This analysis revealed a common “backbone” region coding for a potential plasmid stability system plus a non-pheromone conjugation system containing homologues of both type IV and type II (tight adherence, or Tad-like) secretion systems.

Bacteria of the Gram-positive genus *Sulfobacillus* are typically endospore-forming, moderately thermophilic (40 to 60°C), acidophilic, and iron and sulfur oxidizing (4, 16, 30). Members of the genus *Sulfobacillus* form an important part of the microbial consortia responsible for the biooxidation of sulfide ores in the biomining industry (13, 33, 37). To date, five species have been differentiated phenotypically as well as by 16S rRNA gene sequence analysis. *Sulfobacillus thermosulfidooxidans* (16) appears to be the most active in oxidation of *S. thermotolerans* strains L15 and Y0017 was prepared using a gentle lysis protocol followed by CsCl gradient centrifugation as previously described (44). Plasmids pL15 and pY0017 were “captured” and propagated in *E. coli* EC100D pir\(^+\) cells using the EZ::TN in vitro transposition system (EZ::TN<R6K<ori/KAN-2-> transposon; Epicentre, Madison, WI) as previously described (44). A transposon containing a replicon capable of replication in *E. coli* pir\(^+\) cells and a kanamycin resistance gene was randomly inserted, and kanamycin-resistant *E. coli* transformants were screened using colony hybridizations with total digoxigenin-labeled pL15 as the probe (Roche). The transposon-captured constructs were designated pL15::EZTn and pY0017::EZTn, respectively, and plasmid DNA was prepared using a Nucleobond AX100 kit (Macherey-Nagel, Düren, Germany). A single insertion into each plasmid was confirmed by sequencing. No DNA rearrangements appeared to have taken place after the plasmids were propagated in *E. coli*, as comparisons to the native pL15 and pY0017 isolated directly from *S. thermotolerans* revealed identical SaI and NotI restriction enzyme fragments, apart from the inserted EZ::TN DNA fragments (data not shown).

**General features and comparison of pL15 and pY0017.** Plasmids pL15::EZTn and pY0017::EZTn were fully sequenced on both strands by means of the Roche GS FLX system (Inqaba Biotechnical Industries [Pty.] Ltd., Pretoria, South Africa), generating total data of 1.7 Mb (7,646 reads) and 1.6 Mb (7,266 reads), respectively. Sequences were analyzed using the Glimmer 3 (www.tigr.org/softlab) (11) and DNA-MAN (LynnOn Biosoft) programs. Comparison searches were performed with the gapped BLAST program (1) and the Conserved Domain Database (24) at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Protein function and structure were analyzed using the Center for Biological Sequence Analysis online prediction server at http://www.cbs.dtu.dk. Sequence analysis of plasmid pL15 indicated that it is 65,903 bp long and contains 65 complete open reading frames (ORFs) and 3 disrupted or incomplete ORFs. pY0017 is 59,212 bp long and contains 59 complete ORFs and 1 disrupted ORF. Approximately 70% of the ORFs encode potential proteins with homology to

* Corresponding author. Mailing address: Department of Microbiology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa. Phone: 27-21-808 5803. Fax: 27-21-808 5846. E-mail: smd@sun.ac.za.
† Supplemental material for this article may be found at http://aem.asm.org/.
\\(\dagger\) Published ahead of print on 16 September 2011.
sequences in databases (Fig. 1; also, see Tables S1 and S2 in the supplemental material). The overall GC content of pL15 and pY0017 calculated from the nucleotide (nt) sequence is 57.5 and 58.2 mol%, respectively, which is considerably higher than the 48.2 ± 0.5 mol% GC for the *S. thermotolerans* type strain DSM17362 (4). Overall the two plasmids have 63.5% sequence identity, with certain regions showing identity as high as 92% (Fig. 2). Plasmid pL15 has approx-

![FIG. 1. Genetic maps of *Sulfobacillus* plasmids pL15 and pY0017. The relative location and size of ORFs are indicated by gray arrows (denoting direction of transcription) or bars (in the case of small ORFs). White arrows indicate the presence of insertion sequences. Black bars depict the 37-bp inverted repeats of the composite transposon. An asterisk marks the position of insertion of the EZ::Tn used to capture the plasmids.](http://aem.asm.org/)

![FIG. 2. Diagrammatic representation of nucleotide similarity (% identity) between *Sulfobacillus* plasmids pL15 (top line; 65,903 bp) and pY0017 (bottom line; 59,212 bp). Gray blocks represent regions of more than 90% similarity and white blocks represent dissimilar regions. Insertion features or elements (∆) and their respective nucleotide positions are indicated above and below the lines for pL15 and pY0017, respectively. The alignment was performed with the Vector NTI Advance 11 trial version package.](http://aem.asm.org/)
approximately 9.6 kb of additional DNA, bordered by the 37-bp inverted repeats of a class II transposon.

**Genes likely to be associated with conjugation.** Plasmids pY0017 and pL15 each contain a region of more than 23 kb encoding homologues of the type IV secretion system (T4SS) components VirB1, VirB4, VirB11, and VirD4, as well as homologues of conjugation genes from, almost exclusively, Gram-positive species. Approximately 22 putative ORFs occur within this region, of which approximately 14 can be assigned putative conjugation functions (Table 1). Significantly, all the ORFs predicted within this region are transcribed in the same direction, suggesting an operon-style scheme consistently found in conjugative systems (21).

Relaxases are common to all transmissible plasmids and can be divided into six families, some of which partially overlap (14, 40). A phylogenetic tree (see Fig. S1 in the supplemental material), including the relaxases of the *Sulfobacillus* conjugative region, was prepared by M. Garcillán-Barcia as previously described (14) and indicates that the relaxases are members of the MOBP-type family but do not belong to any known branch and, therefore, may represent the first members of a new MOBP branch closely associated with the MOBPu clade (Fernando de la Cruz, personal communication). In keeping with the coevolution of relaxases and their often closely associated type IV coupling proteins (T4CPs) (14), the pL15 and pY0017 T4CPs lie on a phylogenetic branch that includes the coupling proteins from the MOBPu plasmids (Fernando de la Cruz, personal communication) (data not shown).

A number of other macromolecular transport homologues are potentially coded for by the *Sulfobacillus* conjugative region. An ORF coding for a putative Flp pilus assembly protein (CpaF-like) is closely followed by ORFs coding for type II secretion system (T2SS) TadB- and TadC-like proteins (Table 1). The *A. actinomycetemcomitans* *tadB* and *tadC* genes may be

<table>
<thead>
<tr>
<th>ORF&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Putative homologue(s)</th>
<th>Proposed function, comments, and reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L17/Y16 MinD</td>
<td>Possible role similar to that of TadZ, which is also a member of the MinD/ParA superfamily of ATPases</td>
<td></td>
</tr>
<tr>
<td>L23/Y22 CpaF/TadA</td>
<td>Flp pilus assembly ATPase-like proteins are thought to play a role in providing the energy needed to power the DNA transport process, although they are also known to function in capacities other than as constituents of T4SSs (6).</td>
<td></td>
</tr>
<tr>
<td>L24/Y23 TadB L25/Y24 TadC</td>
<td>TadB and TadC are putative integral inner membrane proteins, postulated to form a heteromultimer that allows the passage of the other Tad components across the inner membrane (9, 10).</td>
<td></td>
</tr>
<tr>
<td>L30/Y30 CpaB</td>
<td>ORFs L30 and Y30 show weak identity (30% over 175 aa) to the Flp pilus assembly CpaB protein of <em>Moorella thermoacetica</em> and are similarly sized. Like <em>Moorella</em> <em>cpaB</em>, the genes are also immediately adjacent to a prepilin peptidase gene.</td>
<td></td>
</tr>
<tr>
<td>L31/Y31 CpaA/TadV</td>
<td>Prepilin peptidases are responsible for the processing of pilins in type IV pilus system and T2SS as well as in Tad systems (reviewed in reference 41).</td>
<td></td>
</tr>
<tr>
<td>L32/Y32 StbA</td>
<td>StbA-like proteins belong to the family of actin-like ATPases (5, 15). The <em>stbABC</em> locus of the conjugative IncN plasmid pKM101 appears to prevent plasmid instability in a recombination-proficient host and has been postulated to resolve plasmid dimers for efficient partitioning to daughter cells (34).</td>
<td></td>
</tr>
<tr>
<td>L36/Y36 Relaxase</td>
<td>DNA substrate recruitment for translocation during conjugation requires the action of a relaxase as well as a coupling protein.</td>
<td></td>
</tr>
<tr>
<td>L37/Y37 TraG</td>
<td>TraG-like type IV coupling proteins are VirD4 homologues required to link the DNA-protein (relaxosome) substrate to the transmembrane transfer complex (17). L37 is interrupted by a group II intron and would require exact splicing to reconstitute a functional gene product.</td>
<td></td>
</tr>
<tr>
<td>L39/Y38 LtrC</td>
<td>Primase (LtrC-like protein)</td>
<td></td>
</tr>
<tr>
<td>L40 L41/Y39 TrsB TrsL</td>
<td>Homology to hypothetical proteins of pBMB67 (8), similar to the transfer complex proteins TrsB/TraB and TrsL/TraL, of <em>Staphylococcus aureus</em> conjugative plasmid pG01 (29)</td>
<td></td>
</tr>
<tr>
<td>L44/Y41 VirB4</td>
<td>ATPase-like protein thought to play a role in providing the energy needed to power the DNA transport process (6)</td>
<td></td>
</tr>
<tr>
<td>L45/Y42&lt;sup&gt;b&lt;/sup&gt; Lytic transglycosylase</td>
<td>L45 and Y42 show homology to lytic transglycosylases in their C-terminal regions and contain a conserved domain for binding to N-acetyl-D-glucosamine, as well as the catalytic residue for localized lysis of the peptidoglycan cell wall (24).</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> ORFs with “L” and “Y” designations are found in pL15 and pY0017, respectively. Detailed descriptions of the ORFs are given in Tables S1 and S2 in the supplemental material.

<sup>b</sup> Y42 contains a frameshift after the first approximately one-third of the ORF that may be the result of a mutation during propagation of the captured pY0017 plasmid in *E. coli*. 

TABLE 1. ORFs putatively involved in conjugation of plasmids pL15 and pY0017
paralogues, as their protein products show low but significant identity (35), as does TadB with TadC in both plasmids pL15 and pY0017 (approximately 23% amino acid [aa] identity in both cases). CpaF and TadA are related to the VirB11 family of putative secretion ATPases and appear to have analogous functions in Caulobacter (the cpa locus) and Aggregatibacter (the tad locus), (36, 39). The tight-adherence (Tad) secretion system, dedicated to the assembly and export of Flp pili, is found in a wide variety of bacteria and archaea. Although some distance upstream of the other tad-like genes, the Sulfolobus plasmid ORFs L17 (in pL15) and Y16 (in pY0017) may code for a TadZ-like homologue, since the predicted product shows a close relationship to the cell division inhibitor MinD of Thermococcus sibiricus (Table 1; also see Tables S1 and S2 in the supplemental material) and sequence comparisons place TadZ in the MinD/ParA superfamily of ATPases. The Sulfolobus plasmid ORFs L31 and Y31 code for potential homologues of TadV, the prepilin peptidase coded for by the tad locus (Table 1). There is a striking similarity in the gene order and predicted function of the components of the Caulobacter pilA-cpa locus and the A. actinomycetemcomitans flp-rcp-tad locus (39, 42) and both of the Sulfolobus plasmids’ putative tad-like loci (see Fig. S2 in the supplemental material). In Gram-positive bacteria, the tad locus is shorter than that in Gram-negative bacteria and encodes only the TadZ, TadA, TadB, and TadC products and none of the products known to localize to the outer membrane of Gram-negative bacteria, for example, the secretin homologues RepC and CpaC (9, 42). The presence of tad-like genes on the putatively conjugable Sulfolobus plasmids suggests a function other than that of mere macromolecular transport. Although Tomich and coworkers classify the Tad system as a T2SS, the presence of certain genes (among them, repC, a cpaB-like homologue) is unique to the tad-like operons (42). Only one plasmid-borne tad locus, on the pSym plasmid in Sinorhizobium meliloti, has been identified (42), and it has an arrangement different from that of a similar tad-related locus on the chromosome of the same organism. A potential T4SS encoded on a plasmid in one strain and on the chromosome of another strain of A. actinomycetemcomitans prompted Novak and coworkers to propose that the macromolecular transport system may have been adapted to serve dual functions—macromolecular transport in the case of the chromosomal locus and conjugation in the plasmid locus (31). Similarly, multiple tad-related loci with different arrangements are found in Gram-negative and Gram-positive bacteria as well as archaea, suggesting a multitude of functions (42).

Genes involved in plasmid stability. ParA-like (ORFs L6 and Y5) and ParB-like (ORFs L7 and Y6) proteins are coded for by both Sulfolobus plasmids. Downstream of the parB-like gene is a region of approximately 650 bp (in pL15) and 830 bp (in pY0017) with no obvious ORFs. This region contains four 16-bp palindromes with the consensus sequence TGGTC AACGTGAAACA (see Tables S1 and S2 in the supplemental material). A fifth palindromic repeat with a single mismatch to the consensus sequence at either end is found in pY0017. In both plasmids, the first of the repeats overlaps the stop codon of parB. In Bacillus subtilis, the parA and parB homologues (soj and spoOJ, respectively) play a role in sporulation and chromosomal partitioning (reviewed in reference 3). SpoOJ binds to specific sites called parS found on either side of the site of chromosomal replication, oriC (22). The parS sites are 16-bp palindromic sequences (TGGTCACGTGAAACA), one of which is located within spoOJ. Yamaichi and Niki (45) showed that an unstable mini-F plasmid (sojABC deleted) could be stabilized in E. coli by the addition of a fragment carrying soj, spoOJ, and parS, suggesting that the basic mechanism for chromosomal and plasmid partitioning is conserved among bacteria.

A second putative stability region is found immediately upstream of the par region on pL15 and pY0017. Two small ORFs (L3 and L4 in pL15 and Y3 and Y4 in pY0017) code for putative proteins with conserved domains homologous to the YcF superfamily (HicA-like) and the uncharacterized protein UPF0150 superfamily (HicB-like), while a third small ORF (L5 and Y5) has very weak homology to a similarly sized CopG-like DNA-binding protein of Bartonella grahamii (see Tables S1 and S2 in the supplemental material). A toxin-antitoxin system consisting of a HicA cassette with a possible RNA binding and cleavage mechanism has been described (23).

A third region on plasmids pL15 and pY0017 coding for a StbA-like protein may also serve as a stability locus (Table 1).

Genes involved in plasmid replication. No clear homologue of a RepA protein is evident among the ORFs of either pY0017 or pL15. No replication in E. coli, independent of the R6K ori, was achieved for these captured plasmids, so it was not possible to test for the minimal replicon using that host.

Accessory genes. An ORF (L49/Y46) coding for a putative phosphatidylycerine/phosphatidylglycerophosphate/cardioliopin synthase-like protein occurs downstream of the putative conjugative region and within an area of approximately 92% nt identity between the two plasmids. None of the other seven putative ORFs in this ~6-kb region appears to be likely to confer an advantage to the host. Directly following this is a region of only 45% nt identity between the two plasmids (Fig. 2), containing seven ORFs and a transposon in pL15 and 10 ORFs in pY0017, nine of which read in the complementary direction. In pY0017, ORFs Y54, Y53, and Y52 potentially encode for an arginosuccinate synthase-like protein, a clavaminate synthase-like protein, and a possible permease of the drug/metabolite transporter superfamily, respectively. In pL15, ORF L54 codes for an apparently prematurely terminated protein that belongs to the major facilitator superfamily. ORFs L63, L64, and L65 encode a putative alanine tRNA synthetase (C terminus region only), a phosphoribosylglycinamide synthetase-like protein, and a possible permease of the ATP-grasp superfamily, respectively. The only ORF conserved between pL15 and pY0017 within this region is ORF L68/Y56, which potentially codes for an invertase/recombinase-like protein.

Transposon-contained region. The transposon-carried DNA of pL15 contains a gene for a transposase belonging to the transposase 7 family (pfam01526), a resolvase gene, and six putative ORFs conferring no obvious advantage to the host. Southern hybridization, using a probe made to the pL15 1.6-kb EcoRV fragment (nt position 49318 to 50922) that extends from approximately 200 bp inside the invert repeat of the transposon to include ORFs L55 and L56, indicated that the transposon also occurs in a single copy on the chromosome of S. thermotolerans strain L15 but not Y0017 (data not shown). This implies that the putative genes carried by the transposon are not an essential part of the S. thermotolerans chromosome.
Insertion sequences and phage-related ORFs. Three DNA insertions (one in pL15 and two in pY0017) are found in a region of otherwise >90% nt similarity between the two plasmids (Fig. 2). A 1,902-bp group II intron (nt position 22804 to 24705 in pY0017) has interrupted the putative ORF Y27 and contains a coding sequence for a putative protein with 48% identity to a prophage LambdaSa1 transcriptase/maturase protein. This protein also has 43% identity, and is closer in size and start position, to a reverse transcriptase (RT) of a group II intron of Pelotomaculum thermophorum (see Table S2 in the supplemental material). In pL15, a 2,517-bp group II intron (nt position 30689 to 32305) interrupts L37 and contains an ORF with 66% identity over the entire putative 603-aa protein to an RT of Desulfitalbactrium hafniense. There is also strong identity (47%) to the group II intron-encoded protein (IEP) LtrA of Bacillus cereus (see Table S1). Group II introns are large catalytic RNA molecules, found in both eukaryotes and prokaryotes, which act as mobile genetic elements (25). By comparison of the pL15 and pY0017 sequences, it is easy to identify the exact sites at which the two introns have inserted, and the 5’ end (GTGGCG for both pL15 and pY0017) and the 3’ end (AGCCCTAC for pL15 and ACTCTAC for pY0017) agree well with the consensus sequences for the 5’ end (GU GGY) and the 3’ end [AXX(X)RAY] described by Martínez-Abarca and Toro (25). Most of the size difference between the pL15 and pY0017 group II introns is attributable to the difference in size of their IEPs. The IEP of pY0017 shows only 26% aa identity to the IEP of pL15 and is more than 200 aa smaller. Group II introns, such as those found on pL15 and pY0017, are commonly found within genes involved in DNA mobility, suggesting that they use horizontal gene transfer as an additional means of dissemination (25, 27). The pL15 group II intron interrupts a putative traG gene, and exact removal of the intron sequence at the splice site junction would potentially reconstitute a 2,223-bp mRNA for a 740-aa protein (with 42% identity to a TraG family protein from Thermoanaerobacterium thermosaccharolyticum [see Table S1]). Intron splicing to reconstitute an interrupted conjugative release gene has been demonstrated for plasmid pRS01 in lactococci (27). In some cases, both the splicing activity (46) and the insertion of group II introns into target DNA (28) have been shown to be influenced by temperature. The possibility that the group II introns of pL15 and pY0017 are similarly temperature sensitive (which would likely affect the conjugational ability of the plasmids of these moderate thermophiles) is the subject of further study in this laboratory.

A second insertion sequence of 1,650 bp (position 10910 to 12559) is found in pY0017. It codes for a putative ORF (OrfY15) of 415 aa with 37% identity to a transposase of the ISChy2 element in Carboxydothermus hydrogenoformans (see Table S2). Immediately upstream of the insertion sequence containing the ISChy2-like transposase is a complete ORF (Y14) coding for a potential helix-turn-helix-containing protein of the excisionase family. An excisionase, or Xis protein, is a small protein that is not enzymatically active but binds and promotes excise recombination (32). A similar ORF (L16) is also present in pL15 and at other loci in both plasmids (Y19/Y48 and L20/L51). Other possibly phage-derived genes are L12/Y9 and L13/Y10 (putative bacteriophage immunity repressor) and L52/Y49 (gp57).

In conclusion, we have determined the complete nucleotide sequences of two S. thermotolerans plasmids and shown that they contain a common region comprising conjugation genes, a B. subtilis-type par putative stability system, and a number of insertion elements, including group II introns. No obvious or previously described replication genes or origin of replication was identified. Using the essential relaxase and conjugative T4CPs as classification tools, we can place pL15 and pY0017 within the MOBp family of plasmids. In comparison to Gram-negative systems, relatively little is known about nonpheromone conjugation in Gram-positive bacteria, including the mechanism of establishing physical contact between donor and recipient cells. This prompted us to do a detailed investigation of this region of pL15 and pY0017. We speculate that the pL15 and pY0017 tad-like operon might be involved in some form of close cell contact in order to facilitate conjugal transfer of the plasmids, as opposed to the production of pili for adherence to a substrate. The finding that backbone of plasmids isolated from different strains of S. thermotolerans from different geographical locations are related but that the accessory genes are different appears to be a recurring theme in iron- and sulfur-oxidizing bacteria. Species-specific plasmid backbones have been reported in Acidithiobacillus ferrooxidans (7, 12, 43), and Acidithiobacillus caldus (44). In A. caldus and S. thermotolerans, these plasmids appear to be either mobilizable or conjugable and may be involved in the movement of the accessory genes within isolates of the same species and possibly also between species.

Nucleotide sequence accession numbers. The annotated nucleotide sequences of pL15 and pY0017 are available under the GenBank accession numbers JN119829 and JN119830, respectively.

Sulfobacillus thermotolerans strains L15 and pY0017 were a kind gift from Barrie Johnson. The phylogenetic trees of the relaxase and T4CPs of the Sulfobacillus plasmids were prepared by María García-Basciá, together with helpful comments by Fernando de la Cruz.

This work was supported by grants from the University of Stellenbosch, the National Research Foundation (Preatoria), the BioMinE project 500329 of the EU framework 6, and the BHP-Billiton Johannesburg Technology Centre.

REFERENCES


