Identification and Characterization of Transposable Elements of *Paracoccus pantotrophus*

Dariusz Bartosik, Marta Sochacka and Jadwiga Baj


Updated information and services can be found at: http://jb.asm.org/content/185/13/3753

**REFERENCES**

This article cites 49 articles, 26 of which can be accessed free at: http://jb.asm.org/content/185/13/3753#ref-list-1

**CONTENT ALERTS**

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
Identification and Characterization of Transposable Elements of Paracoccus pantotrophus

Dariusz Bartosik,* Marta Sochacka, and Jadwiga Baj

Warsaw University, Institute of Microbiology, Department of Bacterial Genetics, 02-096 Warsaw, Poland

Received 10 March 2003/Accepted 18 April 2003

We studied diversity and distribution of transposable elements residing in different strains (DSM 11072, DSM 11073, DSM 65, and LMD 82.5) of a soil bacterium Paracoccus pantotrophus (α-Proteobacteria). With application of a shuttle entrapment vector pMEC1, several novel insertion sequences (ISs) and transposons (Tns) have been identified. They were sequenced and subjected to detailed comparative analysis, which allowed their characterization (i.e., identification of transposase genes, terminal inverted repeats, as well as target sequences) and classification into the appropriate IS or Tn families. The frequency of transposition of these elements varied and ranged from $10^{-6}$ to $10^{-2}$ depending on the strain. The copy number, localization (plasmid or chromosome), and distribution of these elements in the Paracoccus species P. pantotrophus, P. denitrificans, P. methylutens, P. solvovorans, and P. versutus were analyzed. This allowed us to distinguish elements that are common in paracocci (ISPpa2, ISPpa3—both of the IS5 family—and ISPpa5 of IS66 family) as well as strain-specific ones (ISPpa1 of the IS256 family, ISPpa4 of the IS5 family, and Tn3 of the Tn3 family), acquired by lateral transfer events. These elements will be of a great value in the design of new genetic tools for paracocci, since only one element (IS1248 of P. denitrificans) has been described so far in this genus.

Insertion sequences (ISs), which are components of nearly all bacterial genomes, are the simplest forms of transposable elements (for reviews, see references 10 and 38). Although ISs do not encode genes that directly affect the phenotype of the bacterial host, they are not, however, eliminated in the course of evolution. It thus seems little likely that they are only “selfish” DNA that escapes natural selection due to more active replication related to transposition. The transposition of these elements usually involves structural changes in DNA that lead to the formation of mutations: insertions, deletions, inversions, and translocations (of even large DNA fragments), which may result in varied phenotypic effects (19). Insertion of ISs can lead to gene disruption or activation of adjacent cryptic genes by formation of upstream promoters (29). ISs are therefore the most recombinogenic factor of bacterial genomes. Their presence to a large extent determines that these genomes are not static, monolithic structures, but are subject to many structural rearrangements. These elements thus play the role of a factor that significantly enhances variability and consequently the adaptive and evolutionary capacities of their hosts.

In addition, ISs frequently occur on plasmids and bacteriophages, which can propagate them by lateral transfer between various bacterial populations. Some ISs are able to form composite transposons (Tns), which can carry various phenotypic traits (e.g., resistance to antibiotics or heavy metals and the ability to utilize different carbon sources) (1, 40). To date, more than 800 IS elements have been isolated from more than 200 prokaryotic species of both bacteria and archaea (http://www-is.biotoul.fr/is.html). The majority of transposable elements were identified from various sequencing projects. However, in most cases, their activity was not experimentally confirmed.

We initiated studies aimed at the identification and characterization of functional transposable elements of paracocci (α-Proteobacteria), which are physiologically among the most versatile bacteria performing a number of different growth modes. During the past decade, the genus Paracoccus has undergone serious taxonomic changes. Several new species were isolated, and the status of others was re-evaluated (2, 32). The genus currently embraces 17 species. Thus far, the only known transposable element of these bacteria is IS1248 of P. denitrificans PdX13, classified as part of the IS5 family (3). There is a much greater wealth of information on ISs in the related genus Rhizobium (http://www.is.biotoul.fr/is.html).

In this paper, we present an analysis of the transposable elements of P. pantotrophus, which is a facultative chemolithoautotroph. It is able to grow with molecular hydrogen, sulfide, or thiosulfate as an electron donor under aerobic conditions. All three substrates can be used for mixotrophic growth. A large variety of organic compounds support aerobic and anaerobic (denitrifying) growth of these bacteria (32). P. pantotrophus strains show a great physiological heterogeneity, which is manifested, for instance, in the capacity of only some of them for methylotrophic growth (41). The high variability and plasticity of the species are indicated, for example, by the observation of the appearance within strain ATCC 35512 (unable to utilize methanol) of spontaneous mutants that were able to grow with methanol as the sole source of carbon and energy (17). Although there is no evidence linking the presence of transposable elements with important phenotypic characteristics of paracocci, however, that their physiological heterogeneity might result from various transposition events cannot be ruled out.
of an  

region of pWKS1 (5) was cloned into the corresponding multiple cloning site 

Eco 

tle vector, the 1.8-kb 

were purchased from either Promega or Fermentas. 

standard procedures as described by Sambrook and Russel (43). All enzymes 

ligations, and agarose gel electrophoresis were conducted in accordance with 

gradient centrifugation. Cloning experiments, digestion with restriction enzymes, 

fi 

Birnboim and Doly (8) and, if required, puri 

ced by CsCl-ethidium bromide 

Paracoccus 
° 

C( 

Bertani (LB) medium (43) at 30 

plasmids used in this study are listed in Table 1. Bacteria were grown in Luria-

/H9262 

kanamycin, 50 

When necessary, the medium was supplemented with antibiotics as follows: 

E. coli 

/H9262 

P. denitrificans DSM 413 

contains pHG16-a (~70 kb) and pHG16-b (~400 kb) 7, 22 

P. versutus UW1 

contains pTAV1 (107 kb) 6, 50 

TABLE 1. Bacterial strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strain or plasmid</th>
<th>Characteristics</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P. pantotrophus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSM 11072</td>
<td>Contains pWKS1 (2.7 kb) and pWKS3 (&gt;400 kb)</td>
<td>31</td>
</tr>
<tr>
<td>UWPI</td>
<td>Rif(^r) derivative of DSM 11072 deprived of pWKS1</td>
<td>5</td>
</tr>
<tr>
<td>DSM 11073</td>
<td>Contains pKLW1 (~100 kb) and pKLW2 (~400 kb)</td>
<td>7, 31</td>
</tr>
<tr>
<td>DSM 11073R</td>
<td>Rif(^r) derivative of DSM 11073</td>
<td>7</td>
</tr>
<tr>
<td>DSM 65</td>
<td>Contains pHG16-a (~70 kb) and pHG16-b (~400 kb)</td>
<td>7, 22</td>
</tr>
<tr>
<td>DSM 65R</td>
<td>Rif(^r) derivative of DSM 65</td>
<td>5</td>
</tr>
<tr>
<td>LMD 82.5</td>
<td>Contains pPAN1 (<del>110 kb) and pPAN2 (</del>&gt;400 kb)</td>
<td>7</td>
</tr>
<tr>
<td>LMD 82.5R</td>
<td>Rif(^r) derivative of LMD 82.5</td>
<td>5</td>
</tr>
<tr>
<td><strong>P. denitrificans DSM 413</strong></td>
<td>Contains pHG16 (~400 kb)</td>
<td>22, 41</td>
</tr>
<tr>
<td><strong>P. methyluten DM12</strong></td>
<td>Contains pMTH4 (~23 kb), pMTH1 (<del>40 kb), pMTH2 (100 kb) and pMTH3 (</del>&gt;400 kb)</td>
<td>2, 16</td>
</tr>
<tr>
<td><strong>P. solventivorans DSM 6637</strong></td>
<td>Contains pSOS1 (~70 kb)</td>
<td>2, 47</td>
</tr>
<tr>
<td><strong>P. versutus UW1</strong></td>
<td>Contains pTAV1 (107 kb)</td>
<td>6, 50</td>
</tr>
<tr>
<td><strong>E. coli</strong></td>
<td>Host strain for plasmids</td>
<td>43</td>
</tr>
<tr>
<td>TG1</td>
<td>Host strain for pRK2013</td>
<td>43</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pABW1</td>
<td>Km(^r); mobilizable cloning vector; ColE1 origin; oriT RK2</td>
<td>4</td>
</tr>
<tr>
<td>pGBG1</td>
<td>Km(^r); contains a selective cartridge composed of cl gene of bacteriophage λ and tetacycline resistance gene under control of pR promoter</td>
<td>45</td>
</tr>
<tr>
<td>pWKS1</td>
<td>2.7 kb; natural plasmid of <em>P. pantotrophus</em> DSM 11072</td>
<td>5</td>
</tr>
<tr>
<td>pRKR2013</td>
<td>Km(^r); helper plasmid carrying RK2 tra genes</td>
<td>15</td>
</tr>
<tr>
<td>pMEC1</td>
<td>Km(^r); shuttle vector composed of pABW1, replicator region of pWKS1, and selective cartridge of pGBG1</td>
<td>This study</td>
</tr>
<tr>
<td>pMEC1△</td>
<td>Km(^r) Te(^c); deletion derivative of pMEC1 deprived of functional cl gene</td>
<td>This study</td>
</tr>
<tr>
<td>pMEC100-199</td>
<td>Te(^c) derivatives of pMEC1 obtained in <em>P. pantotrophus</em> DSM 11072</td>
<td>This study</td>
</tr>
<tr>
<td>pMEC200-299</td>
<td>Te(^c) derivatives of pMEC1 obtained in <em>P. pantotrophus</em> DSM 11073</td>
<td>This study</td>
</tr>
<tr>
<td>pMEC300-399</td>
<td>Te(^c) derivatives of pMEC1 obtained in <em>P. pantotrophus</em> DSM 65</td>
<td>This study</td>
</tr>
<tr>
<td>pMEC400-499</td>
<td>Te(^c) derivatives of pMEC1 obtained in <em>P. pantotrophus</em> LMD 82.5</td>
<td>This study</td>
</tr>
</tbody>
</table>

With application of the entrapment vector pMEC1 (which carries a replicon of plasmid pWKS1 that is functional in para-
cocci) (5), we identified several novel transposable elements (some of which were most probably gained by lateral gene transfer). We showed that ISs and Tns are common in *P. pantotrophus*, which points to the high potential of these elements in determining internal genetic rearrangements of their hosts’ genomes.

**MATERIALS AND METHODS**

**Bacterial strains, plasmids, and culture conditions.** All bacterial strains and plasmids used in this study are listed in Table 1. Bacteria were grown in Luria-Bertani (LB) medium (43) at 30°C (*Paracoccus* spp.) or 37°C (*Escherichia coli*). When necessary, the medium was supplemented with antibiotics as follows: kanamycin, 50 μg/ml; streptomycin, 50 μg/ml; rifampin, 50 μg/ml; and tetracycline, 0.5 to 3 μg/ml for different strains of *P. pantotrophus* (Table 2) and 20 μg/ml for *E. coli*.

**DNA manipulations.** Plasmid DNA was isolated according to the method of Birnboim and Doly (8) and, if required, purified by CsCl-ethidium bromide gradient centrifugation. Cloning experiments, digestion with restriction enzymes, ligations, and agarose gel electrophoresis were conducted in accordance with standard procedures as described by Sambrook and Russel (43). All enzymes were purchased from either Promega or Fermentas.

**Entrapment vector construction.** For the construction of the entrapment shut-
tle vector, the 1.8-kb EcoRI-PstI restriction fragment carrying the replicator region of pWKS1 (5) was cloned into the corresponding multiple cloning site (MCS) of an *E. coli* specific mobilizable vector, pABW1 (4). The hybrid plasmid was linearized with PstI and ligated with a 2.9-kb entrapment cartridge (obtained from plasmid pGBG1 (45) composed of (i) a silent tetracycline resistance gene (*tetA*) under control of the pR promoter of bacteriophage λ and (ii) a gene coding for the λ CI repressor. Inactivation of a repressor gene or operator (e.g., through insertion of a mobile genetic element) results in constitutive expression of the tetracycline resistance. The resulting plasmid was designated pMEC1. A deletion derivative of pMEC1 (pMEC1△) was constructed by deletion of a 0.5-kb HindIII fragment coding for the terminal part of the cl gene. Disruption of this gene resulted in constitutive expression of tetracycline resistance.

**TABLE 2. Analysis of Te\(^c\) mutants isolated from different *P. pantotrophus* strains by using pMEC1**

<table>
<thead>
<tr>
<th><em>P. pantotrophus</em> strain</th>
<th>Tetracycline concn (μg/ml)</th>
<th>Frequency of Te(^c) mutants(^a)</th>
<th>Frequency of mutation (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSM 11072</td>
<td>3.0</td>
<td>5.0 × 10^-6</td>
<td>21</td>
</tr>
<tr>
<td>DSM 65</td>
<td>0.5</td>
<td>8.9 × 10^-5</td>
<td>80</td>
</tr>
<tr>
<td>DSM 11073</td>
<td>1.0</td>
<td>1.1 × 10^-3</td>
<td>0</td>
</tr>
<tr>
<td>LMD 82.5</td>
<td>1.0</td>
<td>1.2 × 10^-3</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\) Calculated as the ratio between the number of Te\(^c\) clones obtained on selective medium and the total number of bacteria.

\(^b\) One hundred Te\(^c\) clones were analyzed.

\(^c\) All the inserts were identical in size.
Transformation. Competent cells of E. coli TG1 were prepared and transformed as described by Kushner (33). Transformants were selected on solidified LB medium supplemented with the appropriate antibiotic.

Triparental mating. Overnight cultures (spun down and washed to remove antibiotics) of the donor strain E. coli TG1 carrying a mobilizable vector, P. pantotrophus as the recipient, and E. coli DH5α carrying the helper plasmid pRK203 were mixed 1:2:1. An aliquot of 100 μl of such mixture was spread on a plate with solidified LB medium. After overnight incubation at 30°C, the bacteria were washed off the plate, and suitable dilutions were plated on selective media containing rifampin (selective marker of the recipient strain) and kanamycin to select transconjugants. Spontaneous resistance of the recipient strains to kanamycin was undetectable under these experimental conditions.

Isolation of insertion mutants. The entrapment vector pMEC1 was introduced into recipient P. pantotrophus strains by triparental mating. The overnight culture of the Km" transconjugant, carrying pMEC1, was spread on plates with solidified LB medium supplemented with tetracycline. Appropriate dilutions of the culture were spread on tetracycline-free LB medium in order to determine the frequency of transposition. One hundred Tc" colonies of each strain were further analyzed for plasmid content and restriction pattern. Spontaneous resistance of each strain to tetracycline was undetectable under these experimental conditions.

PCR amplification. For amplification of transposable elements, the following primer pairs of forward and reverse primers (based on the sequence of a selective cartridge) were used: ALIS (5'-TTGTAGTACCGTGTCCGC-3') and ARIS (5'-TCTGCTGTGTTGGTGAAGG-3'), BLIS (5'-TGGTGGCGTGTTGCTGTA-3') and BRIS (5'-GTACGTGACCGGATTACCG-3'), CLIS (5'-TCCGCTGCTACATGGAAGA-3') and CRIS (5'-GGGCATATCAAAATTTGCA-3') and ELIS (5'-GCAAGACTGGCATGATAAGG-3'), and DLIS (5'-GGAAGGGTACGTTATGAGA-3'). For amplification of the internal fragments of the transposable elements (used as probes in hybridization), the following primer pairs were used: LPPA1 (5'-GGCGCATACGACGCGGTGT-3') and RPPA1 (5'-TCTTGACCTGGCAGGTGGC-3') (positions 35 to 1276 of ISpPa1 sequence), LPPA2 (5'-TGATCAGTCGTGCGACAG-3') and RPPA2 (5'-TGGCCGTGAGGCAAACCGCT-3') (positions 75 to 787 of ISpPa2 sequence), LPPA3 (5'-TGAGGTCTCAGTCCAGTGC-3') and RPPA3 (5'-ATGAGATCGCTGCCAAGG-3') (positions 107 to 899 of ISpPa3 sequence), LPPA5 (5'-TCGAGGATCTCATGATCGCA-3') and RPPA5 (5'-ACCGTCGTTCTGCTCACGG-3') (positions 1106 to 1865 of ISpPa5 sequence), L3343 (5'-ATCCGCAAGCAGCGATGCC-3') and R3343 (5'-CAGGTTGACTCAGCTGAG-3') (positions 1 to 2592 of Tn3343 sequence), and L3345 (5'-ACCGCTTCTGCAGAAGG-3') and R3345 (5'-AGATAAGCTCAGTTATGAGA-3') (positions 4483 to 5212 of Tn3393 sequence). Amplification was performed in a Mastercycler (Eppendorf) with the synthetic oligonucleotides described above, Thermus aquaticus polymerase (Qiagen), and appropriate template DNAs. PCR products were separated by 0.8% agarose gel electrophoresis and if necessary purified with the Gel Out kit (DNA Glandik II).

DNA sequencing and analysis. The nucleotide sequence was determined with a terminator sequencing kit and a Perkin-Elmer automatic sequencer (ABI 377) and deposited in GenBank database under accession nos. AY177680, AY177508, AY177681, AY177682, AW225410, and AY232820, respectively.

RESULTS

Construction of entrapment vector pMEC1. In the first stage of our studies, we constructed a novel shuttle entrapment vector, pMEC1, which enables direct in vivo capture of transposable mobile elements in paracocci (Fig. 1). The vector is composed of (i) a replicator region of a small cryptic plasmid, pWKS1 (2.7 kb), of P. pantotrophus DSM 11072 (which ensures stable maintenance in all paracocci and is compatible with their natural plasmids) (5); (ii) E. coli-specific Km" mobilizable vector pABW1 (4); and (iii) an entrapment cartridge carrying a silent tetracycline resistance (tetA) gene (see Materials and Methods for details).

Both kanamycin and tetracycline markers (present in pMEC1) were suitable for selection in Paracoccus spp. However, the tetracycline resistance used for paracocci depends on the strain and copy number of the Tc" plasmid (unpublished results). Therefore, we constructed a deletion derivative of pMEC1 constitutively expressing resistance to tetracycline (pMEC1Δ; see Materials and Methods for details), which allowed direct detection of appropriate tetracycline concentrations for the strains studied (Table 2).

Analysis of a pool of Tc" mutants of P. pantotrophus strains. Plasmid pMEC1 (Km") was introduced into different strains of P. pantotrophus (DSM 11072, DSM 11073, DSM 65, and LMD 82.5), and then Tc" mutants were selected as described in Materials and Methods. As shown in Table 2, Tc" clones appeared with frequencies from 10⁻⁴ to 10⁻⁴, depending on the strain. We analyzed the plasmid pattern of 100 Tc" clones of each of the Paracoccus strains studied. For DSM 11072 and DSM 65, the size of the inserts (roughly estimated by restriction analysis) was differentiated (Table 2). In both cases, preliminary classes of plasmids carrying potential (i) ISs (below 3 kb), (ii) Tns (above 3 kb), and (iii) point mutations (plasmids of the size of pMEC1), were distinguished. In the case of two other strains tested, all of the Tc" mutants carried one kind of insert of the sizes 2.8 kb (DSM 11073) and 5.5 kb (LMD 82.5).

Hybridization analysis. Initially we focused on the analysis of the insertion sequences “captured” in DSM 11072. For determination of their accurate size, we designed (on the basis of the sequence of the selective cartridge) five pairs of nested primers (Fig. 1) (nucleotide sequences are given in Materials and Methods). The primers together with pMEC1-derived plasmids (as template DNA) were used in PCRs. This additionally enabled precise localization of the transposable ele-
ments in the cassette and the choice of the appropriate primers for sequencing of the entire element and its target site. We amplified all of the inserts, which confirmed that the insertions took place within the cI gene. On the basis of the sizes determined, four IS classes, of approximately 0.8, 1.3, 2.8 kb, were distinguished. Randomly selected representatives of each class (present in pMEC135, pMEC181, pMEC114, and pMEC156, respectively) were used for further analysis. To check whether the ISs classified in the individual classes are homologous, hybridization analysis was performed. For this purpose, the digoxigenin (DIG)-labeled internal fragments of the ISs (present in the four pMEC1 derivatives mentioned above) were probed, respectively, against the collection of previously PCR-amplified paracoccal transposable elements. The hybridization results confirmed that each of the distinguished IS classes contained highly homologous (probably identical) elements, which do not cross-hybridize with other elements tested (data not shown). The same probes were used for hybridization with PCR-amplified ISs of DSM 65 and DSM 11073. It appeared that the 1-kb-long IS of DSM 65 (present in pMEC382) did not hybridize with any of the probes. Each of the remaining three IS elements “captured” in this strain (present in pMEC306, pMEC355, and pMEC399) was 0.8 kb and gave weak hybridization signals with the IS-based probe of pMEC135, which indicates that they are slightly divergent (data not shown). On the other hand, all of the elements of DSM 11073 hybridized with the IS-based probe of pMEC156 (data not shown).

**IS nucleotide sequences.** The representative elements of each of the distinguished class (present in pMEC114, pMEC135, pMEC181, and pMEC156 of DSM 11072 as well as pMEC382 and pMEC399 of DSM 65) were sequenced. Comparison of the nucleotide sequences with those in databases revealed that only one of them (present in pMEC399) had been described earlier, as ISJ248a of *P. denitrificans* PDx13 (52). The others are novel elements. Detailed analysis of the nucleotide sequences of the captured elements allowed to determine their structure, that is to identify (i) the transposase (Tnp) gene or genes, (ii) terminal inverted repeats (IRs) being the sites for Tnp binding and action, and (iii) target sequence (DR), which in most cases is duplicated upon insertion. Since all of the features mentioned (summarized in Table 3) were present, the captured elements were designated ISP*pa*1 (from pMEC114), ISP*pa*2 (from pMEC135), ISP*pa*3 (from pMEC181), ISP*pa*4 (from pMEC382), and ISP*pa*5 (from pMEC156). The G+C content of the sequences was in the range 61 to 63.5 mol% (Table 3), which is slightly lower than previously determined for *P. pantotrophus* total DNA (64 to 68 mol%) (41).

(i) **Characterization of ISP*pa*1.** The nucleotide sequence of ISP*pa*1 did not show any significant similarity to the sequences deposited in databases. Analysis of sequence data revealed the presence of one major open reading frame (ORF1), spanning 91% of the element. The presumptive ribosome binding site (RBS) is located within a purine-rich region (5'-GAGAAAG GGA-3'), 9 bp upstream of the initiation codon (ATG) of ORF1. No sequences homologous to the consensus of the E. coli promoters could be identified upstream of ORF1. ORF1 (nucleotides 96 to 1346) encodes a putative peptide of 416 amino acids (aa) with a predicted molecular mass of 47.3 kDa. The theoretical pI calculated from the deduced amino acid sequence of the putative protein is 9.76, which is typical for IS-encoded transposases (10). A comparison of this hypothetical protein with those in databases revealed the highest degree of homology with the product of *Magnetospirillum magnetotacticum* gene *Magn5889* (accession no. ZP 00053133), putative transposases of ISJ490 and ISJ413 of *Burkholderia cepacia* (36), and the product of *Bradyrhizobium japonicum* chromosomal gene *idI45* (accession no. AF322013). Significant similarities were also observed with transposases of ISJ632 of *B. japonicum* (30), ISJ6120 of *Mycobacterium smegmatis* (25), ISJ116 of *Rhodococcus zopfii* (13), IST2 of *Thiobacillus ferrooxidans* (54), and IS666 of *Mycobacterium avium* (44). All of the elements listed above are classified as members of a distinct closely homologous subclass within the family IS256; however, for the majority of them, the transposition activity has not been experimentally determined (http://www-is.biotoul.fr/is.html). Of particular importance was the presence within the ORF1 product of three conserved domains (N2, N3, and C1) that encompass the DDE motif (in boldface; lowercase letters represent residues that are not highly conserved) (DG [65 aa] DG- -GF [106 aa] I-TSN- -E - - - -VR), which has been shown to be conserved among members of the IS256 family (da/g [67 aa] Dg/a- -g/f [112 aa] I/l-t/st/nN- -E - - - -v/l/k) (38). This motif, composed of three acidic residues, was found in the catalytic domains of transposases of many bacterial mobile elements and integrases of retroviruses (10).

ISP*pa*1 has imperfect terminal IR sequences (IR1; Fig. 2A) of 39 bp (with 10 mismatches), which show similarities to corresponding regions of several members of IS256 family (the best matches are shown in Fig. 2B). The sequences adjacent to IR1 of ISP*pa*1 encode additional two stretches of conserved inverted sequences IR2 (10 bp, with three mismatches) and

---

**TABLE 3. Characteristic features of novel transposable elements of *P. pantotrophus***

<table>
<thead>
<tr>
<th>Transposable element</th>
<th>Length (bp)</th>
<th>G+C content (mol%)</th>
<th>IR length (bp)</th>
<th>DR length in bp (sequence)</th>
<th>IS (or Tn) with highest level of homology</th>
<th>% Tnp identity/similarity</th>
<th>Special features</th>
<th>IS (or Tn) family/group</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP<em>pa</em>1</td>
<td>1,376</td>
<td>61</td>
<td>39</td>
<td>9 (CCAAAAAAG)</td>
<td>ISJ490 (<em>R. capsulatus</em>)</td>
<td>55/65</td>
<td>IHT binding site</td>
<td>IS256</td>
</tr>
<tr>
<td>ISP<em>pa</em>2</td>
<td>832</td>
<td>60</td>
<td>14</td>
<td>2 (TA)</td>
<td>ISJ248 (<em>P. denitrificans</em>)</td>
<td>92/96 (ORF1), 90/90 (ORF2)</td>
<td>IHT binding site; framesshifting</td>
<td>IS5/IS427</td>
</tr>
<tr>
<td>ISP<em>pa</em>3</td>
<td>1,055</td>
<td>61</td>
<td>24</td>
<td>9 (CTCTACTCT)</td>
<td>ISAv2 (<em>A. avenue</em>)</td>
<td>55/70</td>
<td>Chi site</td>
<td>IS5/IS903</td>
</tr>
<tr>
<td>ISP<em>pa</em>4</td>
<td>1,053</td>
<td>61</td>
<td>15</td>
<td>9 (AAGGTGATG)</td>
<td>ISAv2 (<em>A. avenue</em>)</td>
<td>56/63</td>
<td>IS5/IS903</td>
<td>IS5/IS903</td>
</tr>
<tr>
<td>ISP<em>pa</em>5</td>
<td>2,829</td>
<td>63.5</td>
<td>32</td>
<td>8 (GAACATCC)</td>
<td>ISm15 (<em>S. meliloti</em>)</td>
<td>56/65 (ORF1), 88/90 (ORF2), 71/76 (ORF3)</td>
<td>Translational coupling</td>
<td>IS66</td>
</tr>
<tr>
<td>Tn3434</td>
<td>3,659</td>
<td>59</td>
<td>38</td>
<td>6 (AGGGTA)</td>
<td>Tn5393</td>
<td>63/77</td>
<td>Cryptic</td>
<td>Tn3</td>
</tr>
</tbody>
</table>

---
IR3 (5 bp) separated by a 13-bp spacer (Fig. 2A). These sequences might correspond to the internal IRs (being putative host factor binding sites), which have been previously reported for, e.g., IS256 of Staphylococcus aureus and the related sequence, IS191 of Streptococcus thermophilus, but not for all members of IS256 family (24). The IR1 located at the 5′ end of the element (IR1L) contains, within a conserved region, a sequence (5′-AAGTTGGTGAT-3′) similar to the consensus sequence of the integrase host factor (IHF) binding site (5′-AAAG-3′) (21). Interestingly, within the cartridge, downstream of ISpPa1, we localized additional sequence (5′-AAGCCAGTGAT-3′), partially located within the target site, matching the putative IHF binding site of the left-terminal IR (IRL) of ISpPa1. A similar observation has been made for IS1, which has a preferred insertion site close to IHF binding site of pBR322 (21). IHF thus seems to be an important factor, which can participate in transposition of some transposable elements.

It has been shown that several members of the IS256 family are able to create, after excision, circular forms, which are the intermediates in transposition (37). We were unable, however, to amplify by PCR (using outward-oriented primers based on the ISpPa1 sequence and total DNA of Paracoccus pantotrophus DSM 11072) a DNA fragment with the possible junction between IR1L and IR1R of ISpPa1 (data not shown).

(ii) Characterization of ISpPa2. ISpPa2 is the shortest element identified in P. pantotrophus (Table 3). Similarly to the case in ISpPa1, the IRL of ISpPa2 contains a putative IHF binding site (one mismatch) (Fig. 3B). However, no analogous sequences were identified in the proximity of the target site. ISpPa2 encodes two overlapping ORFs (159-bp overlap) (Fig. 3A). ORF1 (nucleotides 65 to 412) has an ATG initiation codon and encodes a potential protein of 115 aa with a predicted molecular mass of 13.1 kDa (pI 10.18). A putative RBS (5′-GGAG-3′) is 7 bp upstream of ORF1. ORF2 (positions 253 to 822), with a GTG initiation codon and TGA stop codon placed within the right-terminal IR (IRR), has capacity to encode a 189-aa polypeptide of 21.8 kDa (pI 10.43). A potential frameshift motif (5′-AAAAAG-5′; position 389) located within a loop of a predicted mRNA stem-loop structure is present within the overlapping region (data not shown). It is probable that this motif (as shown for the IS1 and IS3 family members) (18, 26, 55) is able to promote a programmed translational frameshifting, which leads to the formation of a functional fusion transposase. Translational slippage in this case might result in the formation of a fusion protein (ORF1 plus ORF2) of 252 aa and a predicted molecular mass of 29.1 kDa (pI 10.62). The N-terminal part of ORF1 product (as well as of the putative fusion protein) carries a putative DNA binding helix-turn-helix (HTH) motif (data not shown). Additionally the fusion protein (but not the single ORFs) has a DDE motif (in boldface V-IDAT [76 aa] LGD-YD [44 aa] YK-R-IE-F-RLK) similar to transposases of IS427 group of IS5 family (v-Idst/s [76 aa] laD-YD [45 aa] Yk/r-R-ive-F-k/RK) (38). The overall sequences of the ORF1 and ORF2 products showed the highest level of identity with putative transposases encoded by ORF1 and ORF2 of IS248b (52, 53). Significant similarity was also observed with two corresponding ORFs of IS422 of Ralstonia solanacearum (accession no. BAA97980), as well as of IS869 residing in plasmid pTIA2 of Agrobacterium tumefaciens (39), IS1301 of Neisseria meningitidis (27), and IS711 of Brucella ovis (12)—all of which belong to the IS427 group. The conservation of the IR sequences (Fig. 3B)
of ISPpa2 and the ISs listed above additionally confirmed the place of ISPpa2 within the subgroup of family IS5 mentioned.

(iii) Characterization of ISPpa3 and ISPpa4. Within the sequence of ISPpa3 of DSM 11072, we identified one long ORF (positions 68 to 1000), encoding a putative protein of 310 aa with a calculated molecular mass of 35.3 kDa (pI 10.81). No significant similarities were observed between the nucleotide sequence of ISPpa3 and sequences deposited in databases. However, the predicted translation product of ORF1 was similar to several transposases of the IS903 group of the IS5 family and carried the DDE (boldface) (LLV-DGAFDTR [67 aa] K- - -GYH-RSRI [67 aa]) motif (boldface) (LLV-DGAFDTR [67 aa] K- - -gyh-RSRI [67 aa]) highly conserved within this family (IaID-TG [71 aa] S-DGAYDTr [67 aa] K- - -gYh-RSRI [67 aa]). The most significant similarities were observed with putative transposases of plasmid pSD25 of Ruegeria sp. (accession no. AAN05200), ISAav2 of Acidovorax avenue subsp. citrulli (accession no. AF086815), IS17 of the pAT484 plasmid of Acinetobacter hae-

molyticus (42), ISAva2 of plasmid pJM1 of Vibrio anguillarum (51), and IS1420 of Ralstonia solanacearum (accession no. AB028897). Interestingly, ISPpa3 encodes a sequence 5'-GCT GGTGG-3' (position 421) identical to the Chi sites, which specify the place of action for RecBCD nuclease. This sequence of ISPpa3 might thus serve as a portable enhancer of recombination (19).

The nucleotide sequence of ISPpa4 of DSM 65 (1,053 bp) appeared to be highly homologous to ISPpa3 (73% of identity); however, no Chi-related sequence could be distinguished within this element. ISPpa4 contains one large ORF, encoding a putative peptide of 310 aa (pI 10.81) with 77% identity and 83% similarity to the putative transposase of ISPpa3. The DDE motifs as well as the IRs of these two ISs are also highly conserved (data not shown and Fig. 4, respectively). Taking into account results of the comparative analysis, we can state that ISPpa3 and ISPpa4 can be classified as newly closely related elements of the IS903 group of the IS5 family.

(iv) Characterization of ISPpa5. ISPpa5 of DSM 11072, the longest IS identified in P. pantotrophus (Table 3), carries four similarly oriented ORFs (Fig. 5A) encoding putative proteins with predicted molecular masses of 14.1 kDa (ORF1, positions 100 to 483), 12.6 kDa (ORF2, positions 480 to 827), 60.3 kDa (ORF3, positions 890 to 2518), and 11 kDa (ORF4, positions 2518 to 2814), respectively. Database searches and comparison of their predicted amino acid sequences as well as comparison of the IR sequences (Fig. 5B) revealed that ISPpa5 belongs to the IS66 family. Members of this family have been previously reported exclusively in bacteria belonging to the family Rhizobiaceae (genera Agrobacterium and Rhizobium) (38). More detailed analysis performed by Han et al. (28) demonstrated, however, that they are also widely distributed in other genera of gram-negative bacteria, including Escherichia, Pseudomo-

nas, and Vibrio. Homology searches of the databases revealed that ISPpa5 is the most similar to several related ISs residing in the chromosome and symbiotic plasmid pSymA of Sinorhizobium meliloti (20). The highest similarities were observed with ISRm15 (72% identity at the nucleotide sequence level), which has been found to be inserted between the nodI and nodQ genes (nodulation) of S. meliloti (accession no. AF357834).

The majority of IS66-like elements encode three ORFs, corresponding to ORF1, ORF2, and ORF3 of ISPpa5. Some of these ISs encode an additional ORF or ORFs, which are not, however, homologous to ORF4 of ISPpa5 (data not shown). In the case of ISPpa5 (and other IS66 family members), the TGA termination codon of ORF1 overlaps the initiation codon of ORF2 (located in the −1 frame with respect to ORF1) within the 5'-ATTAGA-3' sequence. The intergenic region between ORF2 and ORF3 of ISPpa5 is 61 bp long (usually 20 bp in the IS66 family) (28). In contrast, the stop codon of ORF3 partially overlaps the start codon of ORF4 (located in the −1 frame with respect to ORF3) within the sequence 5'-TGATG-3'.

Database searches and comparison of the ORF3 product of ISPpa5 with over 40 homologous sequences revealed the presence of four conserved DDEE residues, which probably encompass a transposase-specific DDE catalytic triad (data not shown). These conserved residues are indicated in a PILEUP alignment of ORF3 and its closest relatives from ISRm15, ISpre3 of Pseudomonas resinovorans (plasmid pCAR1; accession no. NC 004444), ISEC8 of E. coli EDL933 (associated with
a pathogenicity island) (46), and two unnamed ISs of Mesorhizobium loti R7A (associated with a symbiosis island) (48) and Ruegeria sp. strain PR1b (plasmid pSD25; accession no. AF416331) (Fig. 6).

Identification of transposons of \textit{P. pantotrophus}. As previously mentioned, we distinguished in \textit{P. pantotrophus} a class of putative transposons (Table 2). None of these elements was a composite Tn flanked by any of the IS elements described above (as confirmed by hybridization analysis [data not shown]). We analyzed two transposons in detail. One of them was the only transposon caught by the entrapment vector in strain DSM 11072 (as was judged from the restriction analysis and then confirmed by DNA hybridization [data not shown]). The second one, isolated from LMD 82.5, seemed interesting because of its extremely high frequency of transposition (Table 2). The comparative analysis of the nucleotide sequence of the transposon of LMD 82.5 showed that it is identical to Tn5393 (accession no. AJ431260) of the Tn3 family. Transposition of this element generated a 6-bp duplication of the target sequence (5′-H11032-TGAACA-3′/H11032). Tn5393, besides genetic information required for transposition (transposase [\textit{tnpA}] and resolvase [\textit{tnpR}] genes, resolution sequence [\textit{res}], and terminal IRs) carries two streptomycin resistance genes (\textit{strA} and \textit{strB}) placed downstream of the \textit{tnpR} gene (Fig. 7A).

Analysis of the nucleotide sequence of the transposon of DSM 11072, designated Tn3434 (Table 3), revealed a structure similar to that of Tn5393. Tn3434 does not, however, contain the \textit{str} module (Fig. 7A). The 38-bp IRs of Tn3434 (with one mismatch) were of the length typically found in Tn3-type transposons (23). In contrast, IRs of Tn5393 were much longer (81 bp with four mismatches), but their termini were homologous to those of Tn3434 and other members of Tn3 family (Fig. 7B).

FIG. 6. Alignment of deduced amino acid sequences of the ORF3 translational product of IS\textit{Ppa5}, IS\textit{Rm15} of \textit{S. meliloti}, IS\textit{Pre3} of \textit{P. resinovorans}, IS\textit{Eco8} of \textit{E. coli}, and two unnamed IS66-like elements of \textit{Rueg.} and \textit{M. loti}. Dots indicate gaps introduced to maximize the alignment. Amino acids identical to the protein of IS\textit{Ppa5} are boldface and shaded. The four conserved DDEE residues (encompassing the putative DDE motif), distinguished on the basis of comparative studies of 42 related sequences, are shown on a black background. Amino acid numbering is shown to the right of the sequence lines.

VOL. 185, 2003 TRANSPOSABLE ELEMENTS OF \textit{PARACOCCUS PANTOTROPHUS} 3759
Distribution, copy number, and localization of the analyzed transposable elements. To investigate whether the identified transposable elements are specific only for the host strains or are widely spread among various P. pantotrophus isolates (or other paracoccal species), hybridization analysis was performed. For this purpose, PCR-amplified internal fragments of the elements (nucleotide sequences of the primers are given in Materials and Methods) were probed against total DNAs of P. pantotrophus strains and four other representatives of the genus Paracoccus: P. versatilis UW1, P. denitrificans DSM 413, P. solventivorans DSM 6637, and P. methylutens DM12 (Fig. 8). The restriction enzymes for DNA digestion were chosen individually for each of the elements tested to avoid multiple hybridization signals derived from a single copy of a given IS. The number of hybridized DNA fragments was therefore equivalent to the minimum number of copies of a given element within the genome.

We detected a single copy of ISPpa1 only in one strain of P. pantotrophus: DSM 11072 (hybridization with one XbaI-BglII fragment of 9 kb) (data not shown). On the other hand, ISPpa2 is present in multiple copies in all strains of P. pantotrophus studied as well as in P. methylutens DM12 and P. denitrificans DSM 413 (Fig. 8A). The apparent copy number ranged from 5 to 14. The hybridization patterns of the strains tested were different. Some IS copies displayed different signal intensities, which reflects the presence of other closely related elements homologous to ISPpa2 (e.g., IS1248). At least seven copies of ISPpa3 could be detected in P. pantotrophus DSM 11072, and one copy could be detected in strain LMD 82.5 as well as in P. versatilis UW1 (Fig. 8B). The related element ISPpa4 was shown to be present in one copy in parental DSM 65 (hybridization with one 11-kb PstI fragment) (data not shown). In ISPpa5, positive signals were detected in three strains of P. pantotrophus (two copies in the host strain DSM 11072, three copies in DSM 11073, and two copies in LMD 82.5) as well as one copy in P. denitrificans DSM 413 (Fig. 8C). The cryptic Tn Tn3434 is present in two copies in parental strain DSM 11072 (Fig. 8D), while streptomycin-resistant Tn5393 resides in a single copy only in LMD 82.5 (as judged from hybridization of the strAB-specific probe with a 6-kb BamHI-HindIII restriction fragment of the total DNA of this strain) (data not shown). This is in agreement with our observation that of all paracoccal strains tested, only LMD 82.5 was streptomycin-resistant (data not shown).

In order to localize the transposable elements within the hosts’ genomes (plasmid or chromosome), additional hybridization analysis against DNAs of pHG16-a, pKLW1, pPAN1, pTAV1, pSOS1, and pMTH4 (harbored by the paracoccal strains tested) was performed. We found that pHG16-a of DSM 65 carries single copies of ISPpa2 (Fig. 9B, lane 2) and ISPpa4 (Fig. 9B, lane 3), while pKLW1 of DSM 11073 carries two copies of ISPpa2 (Fig. 9B, lane 1). The other elements were not present in plasmids occurring in these strains (data not shown).

DISCUSSION

We have identified and characterized several novel transposable elements residing in different strains of a soil bacterium, P. pantotrophus (Table 3). These are the first elements to
be identified in this species and in the entire genus Paracoccus (with the exception of IS2428 of P. denitrificans) (53). All of the elements have been identified by their integration into a selective cartridge of a novel vector, pMEC1, which proves that they are functional in transposition. These ISs and Tns may thus serve as a basis for the construction of novel tools for transposon mutagenesis, which may be very useful for the genetic analysis of this interesting group of bacteria.

In general, we can state that the transposable elements are abundant in this species and they represent four distinct families: IS256 (ISPpa1), IS5 (ISPpa2, ISPpa3, and ISPpa4), IS66 (ISPpa5), and Tn3 (Tn3434, Tn5393). The genetic organization of ISPpa1, ISPpa3, and ISPpa4 is typical for the majority of known ISs. They encode a single ORF for a transposase protein with a family-specific DDE motif. The structure of ISPpa2 (IS247 group of IS5 family) differs, since it encodes two overlapping ORFs as well as the conserved frameshift motif, which is likely to promote the generation of a fusion protein (ORF1 plus ORF2) as a result of programmed translational frameshifting (9). The putative fusion protein of ISPpa2 encodes the transposase-specific DDE motif, which strongly suggests that ribosomal frameshifting may participate in the regulation of the transposition of this element. So far, the production of a transframe protein has been reported exclusively for the members of the IS1 (18) and IS3 families (26, 55).

A completely different mechanism of regulation of the expression of transposase genes seems to function in ISPpa5. This element encodes two sets of slightly overlapping ORFs: ORF1 and ORF2 (analogous overlap was observed in all members of IS66 family) and ORF3 and ORF4. ISPpa5 (and its relatives) does not encode the frameshifting motif, which precludes the possibility of generating transframe fusion transposases (9). It is thus probable that ORF1 and ORF2 of all these elements (as well as ORF3 and ORF4 of ISPpa5) may be produced by a translational coupling mechanism (28). ORF3 of ISPpa5 (although it carries a putative DDE motif), unlike other bacterial transposases, has a relatively low isoelectric point (pI 6.75) and does not encode an HTH DNA binding motif. Interestingly, ORF1 (pI 8.77) encodes an HTH motif (data not shown), while ORF2 (the most conserved protein of IS66 family) encodes a basic protein (pI 10.41). We speculate that the proteins encoded by IS66-like elements may form multiprotein complexes with transposase activity. This is in agreement with the results of mutational analysis of IS679 (identified in plasmid pB171 of enteropathogenic E. coli B171) (28), which proved that all three IS679-encoded proteins (homologous to ORF1 to ORF3 of ISPpa5) are necessary for transposition. ORF4 of ISPpa5 is not conserved in other members of the IS66 family: its role needs to be investigated.

Besides the ISs, we identified in P. pantotrophus two Tns, which are members of the Tn3 family. These are a novel cryptic Tn designated Tn3434 and Tn5393 (homologous to the former one), which carry the streptomycin resistance module. Previous analysis demonstrated that Tn5393 is widespread among many gram-negative bacteria isolated from agricultural habitats where streptomycin was used (34). Our observations of the high rate of transposition of Tn5393 into the entrapment vector (Table 2) may explain the reason for its wide dissemination in many environmental isolates (11, 49). This points to the importance of soil bacteria as reservoirs of antibiotic resistance determinants in the environment.

In order to study the distribution and copy number of the identified elements, we performed hybridization analysis. It should be kept in mind that such analysis does not necessarily imply that all copies of a particular IS (or Tn) element detected are identical and functional. The only strain-specific transposable elements were ISPpa1, Tn3434 of P. pantotrophus DSM 11072, and Tn5393 of LMD 82.5, as well as ISPpa4 of DSM 65 (residing on plasmid pHG16-a), a finding that suggests they might have been acquired by recent lateral transfer events.

Among the elements tested, ISPpa2 (closely related to IS2428) appeared to be the most widely distributed in paracocci. Since ISPpa2 and IS2428 cross-hybridize, we were unable to distinguish these elements by hybridization (they are 87% identical at the nucleotide sequence level). It has been previously shown that IS2428 is present in multiple copies in all strains of P. denitrificans tested (53). The presence of ISPpa2-like sequences in plasmids of DSM 65 (pHG16-a) and DSM 11073 (pKLW1) suggests the possibility of their dissemination by lateral transfer. Members of the IS5 family preferentially integrate within the YTAR sequence (Y = C or T; R = A or G) (often 5'-CTAG-3'). In the case of ISPpa2, the sequence was slightly divergent: RTAR (5'-GTAA-3'). It is probable that preferential insertion of these ISs within the putative stop codons (TAG and TAA) may help to minimize the frequency of transpositional disruption of DNA coding regions, thus making possible the potentially broader dissemination of these elements (35).

ISPpa3 (IS903 group of IS5 family) and ISPpa5 homologues are unevenly distributed in the strains tested. Interestingly, hybridization and partial sequencing revealed that the only element “caught” by the entrapment vector in DSM 11073 was ISPpa5. The reason for the unusually high frequency of transposition of this element in this strain (10^3 higher than in DSM 11072; Table 2) remains unclear. It cannot be excluded that its transposition might be stimulated by a host-encoded factor or factors or DNA sequence or sequences adjacent to the insertion site (e.g., transcriptional activation of ISPpa5-encoded genes by a foreign promoter or promoters). We have initiated studies aimed at revealing the cause of this phenomenon. The high sequence similarities of ISPpa5 to the elements residing in chromosome and plasmid pSymA of Sinorhizobium melloti suggest that the exchange of genetic material between members of the family Rhizobiaceae and Paracoccus spp. occurs frequently. This assumption can be supported by the results of our previous studies, in which we have shown that several paracoccal plasmids carry repABC replicons, homologous to that of pSymA and many other rhizobial megaplasmids (7).

The majority of the identified transposable elements seem to be of chromosomal origin. It is worth mentioning, however, that all of the paracoccal strains tested carry megaplasmids (2), which cannot be purified by the standard alkaline lysis procedure. The question of whether the transposable elements identified reside in one of these high-molecular-weight plasmids remains open.

The results presented in this paper are a preliminary part of our complex studies on the distribution and diversity of transposable elements of all species of the genus Paracoccus. We believe that the results obtained will serve to draw conclusions
of a more generalized nature regarding the frequency and directions of lateral transfer in this group of soil bacteria.

ACKNOWLEDGMENTS

We acknowledge M. Mikosa and A. Solyga for excellent technical assistance.

This work was supported by the State Committee for Scientific Research, Poland (grant no. 6 P04A 048 21).

REFERENCES


