

Insights into the Transposable Mobilome of *Paracoccus* spp. (*Alphaproteobacteria*)

Lukasz Dziewit, Jadwiga Baj, Magdalena Szuplewska, Anna Maj, Mateusz Tabin, Anna Czyzkowska, Grazyna Skrzypczyk, Marcin Adamczuk, Tomasz Sitarek, Piotr Stawinski, Agnieszka Tudek, Katarzyna Wanasz, Ewa Wardal, Ewa Piechucka, Dariusz Bartosik*

Department of Bacterial Genetics, Faculty of Biology, Institute of Microbiology, University of Warsaw, Warsaw, Poland

Abstract

Several trap plasmids (enabling positive selection of transposition events) were used to identify a pool of functional transposable elements (TEs) residing in bacteria of the genus *Paracoccus* (*Alphaproteobacteria*). Complex analysis of 25 strains representing 20 species of this genus led to the capture and characterization of (i) 37 insertion sequences (ISs) representing 9 IS families (IS3, IS5, IS6, IS21, IS66, IS256, IS1182, IS1380 and IS1634), (ii) a composite transposon Tn6097 generated by two copies of the ISPfe2 (IS1634 family) containing two predicted genetic modules, involved in the arginine deiminase pathway and daunorubicin/doxorubicin resistance, (iii) 3 non-composite transposons of the Tn3 family, including Tn5393 carrying streptomycin resistance and (iv) a transposable genomic island TnPpa1 (45 kb). Some of the elements (e.g. Tn5393, Tn6097 and ISs of the IS903 group of the IS5 family) were shown to contain strong promoters able to drive transcription of genes placed downstream of the target site of transposition. Through the application of trap plasmid pCM132TC, containing a promoterless tetracycline resistance reporter gene, we identified five ways in which transposition can supply promoters to transcriptionally silent genes. Besides highlighting the diversity and specific features of several TEs, the analyses performed in this study have provided novel and interesting information on (i) the dynamics of the process of transposition (e.g. the unusually high frequency of transposition of TnPpa1) and (ii) structural changes in DNA mediated by transposition (e.g. the generation of large deletions in the recipient molecule upon transposition of ISPve1 of the IS21 family). We also demonstrated the great potential of TEs and transposition in the generation of diverse phenotypes as well as in the natural amplification and dissemination of genetic information (of adaptive value) by horizontal gene transfer, which is considered the driving force of bacterial evolution.

Citation: Dziewit L, Baj J, Szuplewska M, Maj A, Tabin M, et al. (2012) Insights into the Transposable Mobilome of *Paracoccus* spp. (*Alphaproteobacteria*). PLoS ONE 7(2): e32277. doi:10.1371/journal.pone.0032277

Editor: Laurence Van Melderen, Universite Libre de Bruxelles, Belgium

Received: October 7, 2011; **Accepted:** January 24, 2012; **Published:** February 16, 2012

Copyright: © 2012 Dziewit et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Ministry of Science and Higher Education, Poland (grant PBZ-MNiSW-04/I/2007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bartosik@biol.uw.edu.pl

Introduction

Transposable elements (TEs) are components of nearly all prokaryotic genomes. They are mobilized to change their location in DNA by the action of a transposase (TPase), an enzyme which catalyses transposition, i.e. recombination that does not require sequence homology between the TE and the target site. The presence and activity of TEs may lead to structural changes in both the size and composition of a genome. TEs can generate (i) various mutations, such as insertions, deletions, duplications, inversions and translocations of even large DNA fragments, and (ii) replicon fusions [1]. Some elements are able to activate silent, promoterless genes, when they are located in close proximity to the target site of transposition [2,3]. Moreover, TEs can mobilize chromosomal genes for transposition, which results in shuffling of genetic information among various replicons present in a bacterial cell [4]. Thus, these highly recombinogenic elements may be considered as major architects responsible for shaping the structure of prokaryotic genomes. Their activity significantly enhances genome variability and consequently the adaptive and evolutionary capacities of their hosts. In addition, TEs very

frequently occur on mobile plasmids and bacteriophages, which enhances the propagation of TEs by horizontal gene transfer (HGT) between various bacteria. This results in significant enrichment of the bacterial mobilome, i.e. a pool of mobile DNA involved in HGT, which is considered to be the driving force of bacterial evolution.

The simplest TEs are insertion sequences (ISs), which are highly abundant in bacteria. They carry only genetic information that is necessary and sufficient for transposition and its regulation. The majority of ISs are composed of a single open reading frame (ORF), encoding a transposase, flanked by two inverted repeat sequences (IRs) that are recognized by the TPase during transposition. In most cases, the transposition of ISs causes duplication of the target site, so that the inserted element is bordered by two short (2–15 bp) identical direct repeat sequences (DRs) [5].

ISs are also able to form more complex TEs such as (i) composite transposons consisting of random segments of genomic DNA (core region) bordered by a pair of ISs, or (ii) transposable modules (TMs) and ISCR elements composed of one IS copy, which can mobilize an adjacent DNA segment for transposition

[3,6]. ISs are also responsible for the mobility of non-autonomous TEs (i.e. not encoding their own transposase), such as (i) mobile insertion cassettes (MICs), identified in *Bacillus cereus* [7] and (ii) miniature inverted repeat transposable elements (MITEs), which may constitute up to 2% of a bacterial genome [8].

There is another widely distributed group of TEs called the non-composite transposons, which is divided into the Tn3 and Tn7 families. Besides a transposase, the Tn3 elements contain a site-specific recombination module encoding a resolvase (an enzyme responsible for resolution of cointegrates, which are intermediate forms in replicative transposition), while the Tn7 elements are known for their complex transposition machinery (they encode five transposition proteins) [9,10]. These transposons often encode beneficial functions such as drug or heavy metal resistance, pathogenicity or the ability to utilize different carbon sources [11,12].

Large-scale genome sequencing projects have led to an explosion in the number of annotated prokaryotic and eukaryotic TEs (for example, more than 3800 IS elements have now been isolated from over 295 prokaryotic species; ISfinder database [13]). However, in most cases their activity has not been experimentally confirmed.

In 2009 we initiated a project aimed at identifying functional transposable elements harbored by bacteria belonging to the genus *Paracoccus* (*Alphaproteobacteria*). This genus currently comprises 31 species, which have been isolated from many geographical locations and from different environments. *Paracoccus* spp. are physiologically among the most versatile bacteria, and are able to perform a number of different growth modes. Many are aerobic chemoorganoheterotrophs utilizing a wide variety of organic compounds, including potential pollutants like acetone, *N,N*-dimethylformamide and methylamine [14]. Some strains of *Paracoccus* spp. are facultative chemolithoautotrophs utilizing reduced sulfur compounds, molecular hydrogen and Fe(II) as energy sources [14,15]. Because of their versatile metabolism, these bacteria can play an important role in the cycling of chemical elements in the environment. In addition, these physiological properties raise the possibility of employing paracocci in bioremediation systems, particularly since many species can use nitrate as an alternative electron acceptor.

Although there is no direct evidence linking the presence of TEs with important phenotypic traits of *Paracoccus* spp., it is tempting to speculate that their physiological heterogeneity might result from various transposition events.

In this study, TEs residing in strains representing 20 *Paracoccus* species were analyzed. For the identification of functional TEs we used trap plasmids, which are convenient tools enabling the direct identification of even phenotypically silent elements (reviewed by Solyga and Bartosik) [16]. In preliminary studies, TEs residing in several strains of four species were examined: *P. pantotrophus* [17,18], *P. solventivorans* [19], *P. marcusii* [20], and *P. methylutens* [3]. This allowed identification of (i) twelve insertion sequences – *ISPme2*, *ISPpa2*, *ISPpa3*, *ISPpa4*, *ISPso2*, *ISPso3* (IS5 family), *ISPpa5* (IS66 family), *ISPpa1*, *ISPso1* (IS256 family) and *IS1247a*, *ISPme1* (IS1380 family), *ISpmar4* (ISAs1 family) [3,17,19,20], (ii) two closely related transposons of the Tn3 family – cryptic Tn3434 and streptomycin resistant Tn5393 [17], (iii) a transposable genomic island TnPpa1, whose transposition is mediated by two copies of Tn3434 [18], and (iv) a group of TMs, which are composed of a single copy of an IS1380-family element and adjacent fragments of genomic DNA of various lengths [3,20]. Thus, through this complex analysis we have defined a transposable mobilome in these *Paracoccus* spp. strains. We also analyzed the structure, frequency of transposition, specific features

and distribution of the captured TEs, which allowed us to draw more general conclusions concerning their evolutionary impact, as well as the direction and frequency of horizontal gene transfer in this group of bacteria.

Materials and Methods

Bacterial Strains and Culture Conditions

For identification of transposable elements rifampicin-resistant (Rif^r) derivatives of the wild-type strains of *Paracoccus* spp. (listed in Table 1) were used. *Escherichia coli* TG1 [21] was used for plasmid construction. The majority of the strains were grown in lysogeny construction broth (LB) medium (Sigma) at 30°C (*Paracoccus* spp.) or 37°C (*E. coli*). *P. homiensis* DSM 17862 was cultivated in Marine Broth (Difco). Where necessary, the medium was supplemented with sucrose (10%) and with antibiotics at the following concentrations (µg·ml⁻¹): kanamycin, 50; rifampicin, 50; tetracycline, 0.3–35 (depending on the strain).

Table 1. *Paracoccus* spp. strains used in this study.

<i>Paracoccus</i> spp. strain	References
<i>P. aestuarii</i> DSM 19484	[59]
<i>P. alcaliphilus</i> JCM 7364	[60]
<i>P. alkenifer</i> DSM 11593	[61]
<i>P. aminophilus</i> JCM 7686	[62]
<i>P. aminovorans</i> JCM 7685	[62]
<i>P. bengalensis</i> DSM 17099	[63]
<i>P. denitrificans</i> LMD 22.21	[64]
<i>P. ferrooxidans</i> NCCB 1300066	[15]
<i>P. haeundaensis</i> LMG P-21903	[65]
<i>P. halophilus</i> JCM 14014	[66]
<i>P. homiensis</i> DSM 17862	[67]
<i>P. kondratievae</i> NCIMB13773 ^T	[68]
<i>P. koreensis</i> JCM 21670	[69]
<i>P. marcusii</i> DSM 11574	[70]
<i>P. marcusii</i> OS22	[71]
<i>P. methylutens</i> DM12	[72]
<i>P. pantotrophus</i> DSM 11072	[73]
<i>P. pantotrophus</i> DSM 11073	[73]
<i>P. pantotrophus</i> DSM 65	[74]
<i>P. pantotrophus</i> LMD 82.5	[64]
<i>P. seriniphilus</i> DSM 14827	[75]
<i>P. solventivorans</i> DSM 6637	[76]
<i>P. solventivorans</i> DSM 11592	[76]
<i>P. sulfuroxidans</i> JCM 14013	[77]
<i>P. thiocyanatus</i> JCM 20756	[78]
<i>P. versutus</i> UW400	[79]
<i>P. yeei</i> CCUG 32052	[80]
<i>P. yeei</i> CCUG 32053	[80]
<i>P. yeei</i> CCUG 46822	[80]
<i>P. zeaxanthinifaciens</i> ATCC21588 ¹	[81]

doi:10.1371/journal.pone.0032277.t001

Plasmids Used in This Study

For the identification of TEs, the following trap plasmids were used: pMEC1 [17], pMAT1 [20], pMMB2 [18], pEBB10 and pCM132TC [3]. The shuttle *E. coli*-*Paracoccus* spp. trap plasmids pMEC1 and pMMB2 contain the *cI-tetA* cassette, composed of (i) a silent tetracycline resistance gene *tetA* under the control of the bacteriophage lambda pR promoter, and (ii) the gene encoding the lambda cI repressor. Inactivation of the repressor gene (e.g. through insertion of an IS), results in constitutive expression of tetracycline resistance [22]. The broad host range (BHR) trap plasmid pMAT1 or its Tc^r version pEBB10, contains the *sacB* gene of *Bacillus subtilis*, coding for levan sucrose – an enzyme that catalyzes sucrose hydrolysis and levan extension. The products of this reaction are toxic for gram-negative bacteria. Therefore, cells carrying the functional *sacB* gene are sucrose sensitive (Suc^s) and their cultivation in medium containing sucrose results in cell lysis [23]. This allows direct selection of *sacB* mutants (Suc^r) (e.g. carrying inserted TEs), whose growth is not affected under these conditions, thus enabling positive selection of transposition mutants. The BHR trap plasmid pCM132TC, contains a promoterless tetracycline resistance gene *tetA*. In this case, the transposition of TEs containing strong promoters upstream of the *tetA* gene can initiate its expression, resulting in tetracycline resistance in cells carrying such mutated plasmids [3]. Other plasmids, used for analysis of the frequency of transposition of Tn*Ppa1* were (i) pBBR1MCS-3 [24], (ii) pDG12 (*oriV* of pMAR4 of *P. marcusii* DSM 11574; *oriT* RK2; Km^r), (iii) pDIY703 [25], (iv) pMAO-oriT (*oriV* RA3; *oriT* of RK2; Km^r), (v) pMAO-MS (*oriV* RA3; *oriT* of pIGMS31 of *Klebsiella pneumoniae* 287-w; Km^r) [26], and (vi) pMAO-RK (*oriV* RA3; *oriT* of pIGRK of *K. pneumoniae* 287-w; Km^r) [26].

Introduction of DNA into Bacterial Cells

DNA was introduced into *Paracoccus* spp. strains by triparental mating as previously described [27]. Chemical transformation of *E. coli* cells was performed according to the method of Kushner [28].

DNA Manipulation and PCR Conditions

Plasmid DNA was isolated using a standard alkaline lysis procedure [29] and when required, purified by CsCl-ethidium bromide density gradient centrifugation. Total DNA from *Paracoccus* spp. was isolated by the procedure described by Chen and Kuo [30]. Common DNA manipulation methods were performed as described by Sambrook and Russell [31]. DNA amplification by PCR was performed in a Mastercycler (Eppendorf) using synthetic oligonucleotides (listed in Table S1), HiFi or Taq DNA polymerase (Qiagen; with supplied buffer), dNTPs and appropriate template DNAs. PCR products were analyzed by electrophoresis on 0.8% or 2% agarose gels and where necessary, purified using a Gel Out Kit (A&A Biotechnology).

Identification and analysis of a pool of TEs

Trap plasmids were introduced into the Rif^r strains of *Paracoccus* spp. The overnight cultures of the transconjugants were spread on plates with solidified LB medium supplemented with tetracycline (in the case of pMEC1, pMMB2 or pCM132TC) or sucrose (pMAT1 or pEBB10). Appropriate dilutions of the cultures were also spread on LB medium in order to determine the frequency of transposition. For each strain, the plasmids of at least 100 Tc^r or Suc^r clones were analyzed. Three classes of plasmids were distinguished carrying (i) potential insertion sequences (inserts <3 kb), (ii) putative transposons (inserts >3 kb) or (iii) point

mutations (replicons of the same size as the trap plasmid). The insertion sites of the elements were localized by performing PCRs with the trap plasmid insertion derivatives (as template DNA) and previously described sets of cassette-specific primers [17,20]. Unique TEs identified in each strain (selected on the basis of restriction and DNA hybridization analyses) were further sequenced.

Determination of the Genomic Location of Tn6097 by IPCR

Templates for inverse PCR (IPCR) were prepared from total DNA of *Paracoccus ferrooxidans* NCCB 1300066 digested with restriction endonucleases NcoI or XhoI. The digested DNA preparations, purified using a Clean Up kit (A&A Biotechnology), were then self-ligated overnight at 16°C with T4 DNA ligase (Fermentas). The ligation mixture was used as the template for amplification by PCR of the DNA flanking Tn6097 in the host genome. The synthetic oligonucleotides used in PCR were LIPCRPF2 and RIPCRPFE (Table S1).

DNA-DNA Hybridization

Dot blot analysis was performed using a Bio-Dot apparatus (Bio-Rad) according to the manufacturer's instructions. To determine copy number of TEs in parental strains Southern blotting was carried out as described by Sambrook and Russell [31]. Probe DNA fragments of the transposase gene were amplified by PCR using specific oligonucleotide primer pairs listed in Table S1. The fragments were gel-purified and labeled with digoxigenin (Roche). Hybridization and visualization of bound digoxigenin-labeled probes were carried out as recommended by the supplier (Roche). For each of the tested elements, a specific probe was hybridized with appropriately digested total DNA and plasmid DNA of the host strain. The restriction enzymes for DNA digestion were carefully selected for each of the tested elements to avoid multiple hybridization signals derived from a single copy of a given IS. The number of DNA bands hybridizing with the probe was therefore equivalent to the minimum number of copies of a given element within the genome.

Localization of Promoters Enabling Activation of Transcriptionally Silent Genes

A pool of derivatives of the trap plasmid pCM132TC (containing inserted TEs able to drive transcription of a promoterless *tetA* gene in this vector) were analyzed to localize the TE-derived promoters. Deletion analysis was performed for two of the analyzed elements (Tn5393 and Tn6097), by removal of different parts of the inserted TE by digestion with appropriate restriction enzymes, followed by religation and characterization of the resistance phenotype. The constructed mutants contained different parts of the 3'-end of the transposon, adjacent to the promoterless *tetA* gene. For the insertion sequences IS*Pam1*, IS*Pkr1* and IS*Paes3*, PCR was applied to amplify (i) the terminal 3'-end DNA fragments of the elements (identification of possible terminal, outwardly oriented promoters) and (ii) analogous fragments with an attached short DNA region of the trap plasmid, adjacent to the target site of transposition (identification of possible hybrid promoters). The amplified DNA fragments were cloned into pCM132TC (Tc^r) upstream of the promoterless *tetA* gene or into pCM132 upstream of promoter-less *lacZ* reporter gene. The presence of a promoter (tested in *Paracoccus* spp. hosts) resulted in a Tc^r phenotype (pCM132TC) or expression of β-galactosidase (pCM132). β-galactosidase activity was measured as previously described [32]. The same strategy was used to localize promoters

with a pCM132TC::pKLW1 co-integrate. Primers used for PCR amplification are listed in Table S1.

DNA Sequencing, Sequence Analyses and Annotation

Nucleotide sequences of TEs were determined using a dye terminator sequencing kit and an automatic sequencer (ABI 377 Perkin Elmer). A combination of vector-derived primers and primer walking was used to obtain the entire nucleotide sequences. Similarity searches were performed using the ISfinder [33] and BLAST programs [34] provided by the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). G+C plots were created using the program Artemis [35] with a window setting of 70 nucleotides. The numbering of Tn6097 and Tn6122 was assigned by the Tn Number Registry website (UCL Eastman Dental Institute; <http://www.ucl.ac.uk/eastman/tn/>) according to the generally accepted nomenclature [36].

Nucleotide Sequence Accession Numbers

The nucleotide sequences of the elements identified in this study have been submitted to the ISfinder and GenBank databases. Accession numbers of the sequences are given in Table 2.

Results

Identification of a pool of functional TEs in *Paracoccus* spp.

To identify functional TEs of *Paracoccus* spp. we used mobilizable trap plasmids pMEC1, pMMB2, pMAT1, pEBB10 and pCM132TC carrying three types of cassette enabling positive selection of transposition events (see Materials and Methods). The plasmids pMEC1, pMMB2, pMAT1 and pEBB10 (containing the *cl-tetA* or *sacB* cassettes) enable the capture of all types of TEs, irrespective of their structure, specific features and genetic load. In contrast, pCM132TC only permits the identification of elements containing strong outwardly oriented promoters or those able to generate hybrid promoters, which drive the transcription of nearby genes.

The trap plasmids were introduced by conjugation into 25 strains of *Paracoccus* spp. and pools of clones carrying mutated plasmids were obtained and analyzed as described in Materials and Methods. Using this strategy, we captured (i) 37 insertion sequences, (ii) one composite transposon and (iii) 3 non-composite transposons belonging to the Tn3 family (Figures 1 and 2). TEs were identified in the majority of tested strains, with the exception of *P. aminovorans* JCM 7685, *P. homiensis* DSM 17862, *P. sulfuroxidans* JCM 14013 and *P. yeii* (strains CCUG 46822 and CCUG 32052).

Insertion sequences

The nucleotide sequences of the captured ISs were subjected to detailed analysis. All contained a transposase gene(s) and terminally placed inverted repeat sequences (IRs). Moreover, all were able to generate DRs at the target site of transposition (Table 2). Since they were “typical” ISs, their nucleotide sequences as well as amino acid (aa) sequences of their predicted transposases (TPases), were subjected to comparative analyses using the ISfinder (gathers the nucleotide sequences of well defined ISs) and NCBI databases (BLAST program).

Most bacterial transposases possess three domains (designated N2, N3 and C1), which contain three conserved residues (two aspartate and one glutamate) constituting the DDE motif (so-called catalytic triad), that is highly conserved in the majority of bacterial TPases [5] (see Figure S1). The spacing between these residues as well as the presence of other conserved residues within the

domains varies between different IS families or groups [5]. Since the number of described ISs has increased significantly in recent years (800 known ISs in 2002 and >3800 in 2011), we performed a multiple alignment of the amino acid sequences of the transposases of ISs present in the ISfinder database with all the *Paracoccus* spp. ISs identified in this study. Analogous analysis was also used to compare the nucleotide sequences of the terminal IRs of the ISs (Figure S1).

Preliminary analysis revealed that the vast majority (32) of the identified ISs were novel elements. In five cases, we captured ISs that were highly related to known elements (since their transposases showed >95% identity, at the amino acid level, they were considered different isoforms of the same element) – IS1248f, ISPko1a, ISPpa3a, ISPpa5a and ISPve1a. In only four cases we found completely identical elements residing in different bacterial strains – ISPpa2, IS1247, IS1248 and ISPpa8. The identified ISs and their characteristics are presented in Table 2.

Based on the results of these comparative analyses, the ISs of *Paracoccus* spp. could be classified into the following families: (i) IS3 family (IS407 group) – 4 captured elements, (ii) IS5 family (IS5, IS427 and IS903 groups) – 22 elements, (iii) IS6 family – 2, (iv) IS21 family – 3, (v) IS66 family – 2, (vi) IS256 family – 1, (vii) IS1182 family – 1, (viii) IS1380 family – 1, and (ix) IS1634 family – 1.

As shown in Figure 1A, the majority of the ISs contain a single ORF encoding a transposase. However, some of them (belonging to the IS407 group of the IS3 family and IS427 group of the IS5 family) carry two overlapping ORFs and possess a conserved frameshift motif (Figure 1B), which is likely to promote the generation of a fusion protein (ORF1+ORF2) as a result of programmed translational frameshifting [37]. In each case, only the predicted fusion protein contains the complete DDE motif characteristic of functional transposases. Since the generation of fusion proteins is not a frequent phenomenon (as shown for IS1 of *E. coli*) [37], it is apparent that ribosomal frameshifting may participate in negative regulation of the transposition of these elements. So far, the production of such a transframe protein has been reported exclusively for members of the IS1 and IS3 families [37,38].

Interestingly, we found that one of the members of the IS427 group of the IS5 family (ISPha1 of *P. halophilus* JCM 14014) is a natural point mutant, because, in contrast to other elements of this group, it carries a single large ORF encoding a functional DDE transposase.

The third class of the identified ISs consists of elements that contain additional ORF(s) besides the transposase gene(s). In most cases, the specific role of the predicted proteins encoded by these ORFs is unknown, although it is highly probable that they are involved in the regulation of the process of transposition. Such multi-ORF structures are typical for members of the IS21 and IS66 families (Figure 1CDE).

We also performed a detailed inspection of the nucleotide sequences of the captured ISs in order to predict any DNA sequences that may constitute a site of interaction with host-encoded factors. This analysis revealed that the majority of the elements of the IS427 group of the IS5 family (8 ISs) contain, within their left IR (IRL; located at the 5' end of the element), a sequence partially matching the consensus sequence of the integration host factor (IHF) binding site (5'-WAT-CAANNNTTTR-3') [39]. IHF is a DNA-bending protein, which influences genome architecture. It may also act as a positive factor in the transposition of some TEs, e.g. IS10 [40]. It is probable that it also modulates the transposition of elements of the IS427 group.

Table 2. Transposable elements of *Paracoccus* spp. identified in this study.

TE	TE family/group	Length (bp)	IR (bp) ^a	DR (bp)	Acc. no. or reference	Host strain (copy number; location) ^b
IS1247	IS1380	1672	16/13	5	[82]	<i>P. halophilus</i> JCM 14014 (2; 1 in pHAL1)
IS1248	IS5/IS427	832	14/13	2	[83]	<i>P. pantotrophus</i> DSM 11073 (2)
IS1248f	IS5/IS427	832	14/13	2	GQ871939	<i>P. pantotrophus</i> LMD 82.5 (6)
ISPaes1	IS5/IS427	851	18/16	2	GU826197	<i>P. aestuarii</i> DSM 19484 (6)
ISPaes2	IS5/IS427	841	21/18	2	GU826198	<i>P. aestuarii</i> DSM 19484 (2)
ISPaes3	IS256	1393	39/27	8	GU826199	<i>P. aestuarii</i> DSM 19484 (2)
ISPaK1	IS5/IS427	853	15/13	2	GQ871938	<i>P. alkenifer</i> DSM 11593 (1)
ISPam1	IS5/IS903	1050	15/15	9	GQ468940	<i>P. aminophilus</i> JCM 7686 (3)
ISPam2	IS5/IS903	1054	19/18	9	GQ468941	<i>P. aminophilus</i> JCM 7686 (2)
ISPam3	IS3/IS407	1197	39/30	4	[32]	<i>P. aminophilus</i> JCM 7686 (1 in pAMI2)
ISPam4	IS5/IS427	865	15/13	2	[32]	<i>P. aminophilus</i> JCM 7686 (7; 1 in pAMI2)
ISPbe1	IS3/IS407	1198	12/12	2	GQ871936	<i>P. bengalensis</i> DSM 17099 (6)
ISPbe2	IS5/IS427	851	18/15	2	HQ384167	<i>P. bengalensis</i> DSM 17099 (4)
ISPfe1	IS3/IS407	1248	15/15	4	HQ384168	<i>P. ferrooxidans</i> NCCB 1300066 (7)
ISPfe2	IS1634	1884	13/13	6	HQ384169	<i>P. ferrooxidans</i> NCCB 1300066 (2)
ISPha1	IS5/IS427	842	13/12	4	GU826204	<i>P. halophilus</i> JCM 14014 (6)
ISPha2	IS5/IS5	1387	16/15	4	GU826205	<i>P. halophilus</i> JCM 14014 (1 in pHAL1)
ISPhae1	IS5/IS5	1485	16/13	6	GQ868754	<i>P. haeundaensis</i> LMG P-21903 (2)
SPko1	IS5/IS903	1050	18/18	9	GQ868751	<i>P. kondratievae</i> NCIMB 13773 ^T (4 in pKON1)
ISPKo1a	IS5/IS903	1050	18/18	9	GU826201	<i>P. koreensis</i> JCM 21670 (2)
ISPKr1	IS21	2469	34/26	8	GU826200	<i>P. koreensis</i> JCM 21670 (2)
ISPlc1	IS5/IS903	1050	27/23	9	GQ864160	<i>P. alcaliphilus</i> JCM 7364 (8)
ISPmar1	IS6	821	21/18	9	GU997095	<i>P. marcusii</i> DSM 11574 (2)
ISPmar2	IS3/IS407	1269	25/18	4	GU997096	<i>P. marcusii</i> DSM 11574 (1)
ISPmar3	IS5/IS5	1282	18/14	4	GU826203	<i>P. marcusii</i> OS22 (2; 1 in pMOS8)
ISPPa2	IS5/IS427	832	14/13	2	[17]	<i>P. methylutens</i> DM12 (3) <i>P. ferrooxidans</i> NCCB 1300066 (2)
ISPPa3a	IS5/IS903	1054	21/19	9	GQ864161	<i>P. alcaliphilus</i> JCM 7364 (6)
ISPPa5a	IS66	2829	22/20	8	GQ871937	<i>P. bengalensis</i> DSM 17099 (4)
ISPPa6	IS5/IS427	850	15/12	2	EU909900	<i>P. pantotrophus</i> DSM 11073 (8)
ISPPa7	IS66	2731	35/26	8	EU909901	<i>P. pantotrophus</i> DSM 65 (7)
ISPPa8	IS5/IS903	1050	17/16	9	EU909902	<i>P. pantotrophus</i> DSM 65 (9) <i>P. pantotrophus</i> DSM 11073 (3)
ISPPa9	IS6	812	15/15	8	EU909903	<i>P. pantotrophus</i> DSM 65 (1)
ISPse1	IS1182	1653	16/16	4	GQ868753	<i>P. seriniphilus</i> DSM 14827 (1)
ISPth1	IS5/IS903	1054	18/15	9	GQ868752	<i>P. thiocyanatus</i> JCM 20756 (2)
ISPve1	IS21	2503	19/13	8	HQ384165	<i>P. versutus</i> UW400 (1)
ISPve1a	IS21	2503	29/19	7	HQ384166	<i>P. bengalensis</i> DSM 17099 (6)
ISPze1	IS5/IS427	851	18/13	2	GU826202	<i>P. zeaxanthinifaciens</i> ATCC 21588 ^T (6)
Tn3434a	Tn3	3695	38/38	6	[25]	<i>P. aminophilus</i> JCM 7686 (4; 1 in pAMI1, 1 in pAMI7 and 2 in pAMI8)
Tn5393	Tn3	5470	81/77	6	[17]	<i>P. pantotrophus</i> LMD 82.5 (1)
Tn6097		17,759	13/13	6	JN122276	<i>P. ferrooxidans</i> NCCB 1300066 (1)
Tn6122	Tn3	3792	39/39	5	JN127372	<i>P. halophilus</i> JCM 14014 (1 in pHAL1)
TnPPa1	Tn3	44,386	35/35	5	[18]	<i>P. pantotrophus</i> DSM 11072 (1)

^aThe length of the IRs/the number of identical residues.^bLocation in plasmids, if applicable.

doi:10.1371/journal.pone.0032277.t002

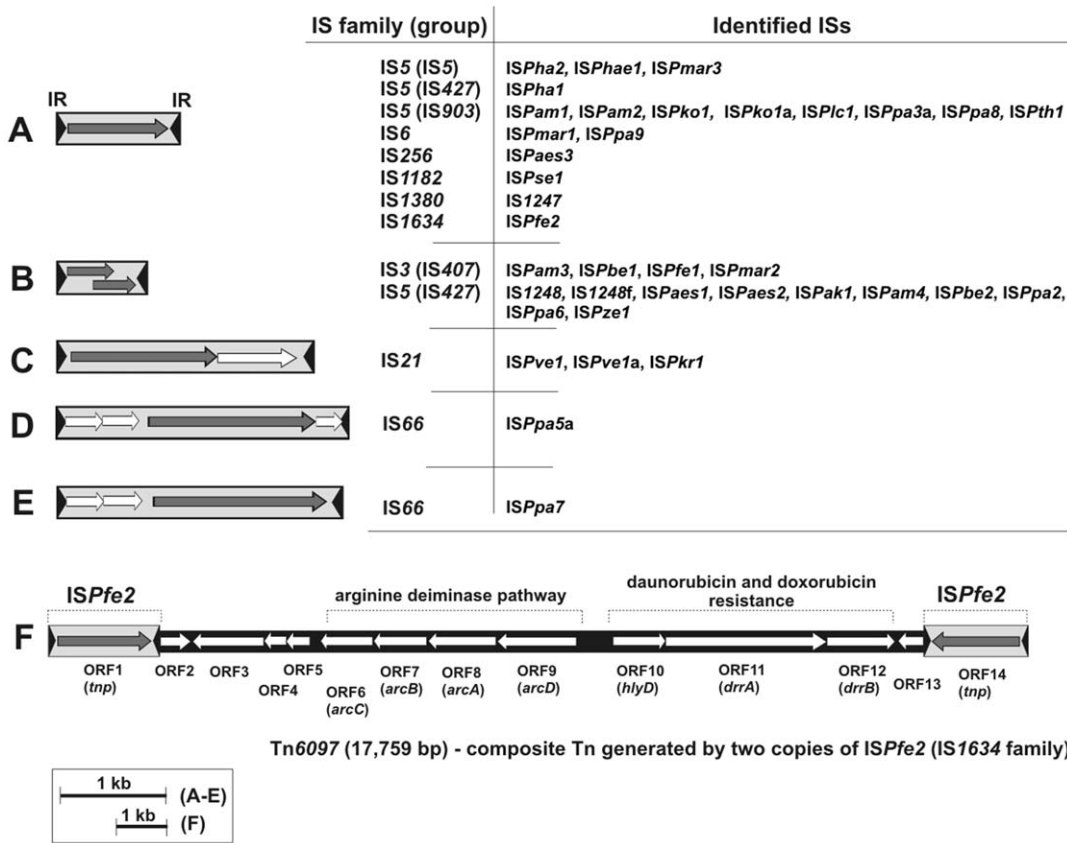


Figure 1. Genetic organization of the insertion sequences (A, B, C, D, E) and composite transposon Tn6097 (F) of *Paracoccus* spp. identified in this study. The families and names of the identified ISs are shown on the panel. Inverted repeats (IRL – left IR; IRR – right IR) flanking ISs are marked by black arrowheads. Predicted coding regions are represented by arrows indicating the direction of transcription. Gray arrows indicate TPase genes and white arrows indicate additional ORFs present within the ISs and within the core of Tn6097 (ORF2 and ORF13) are truncated (the disruptions are most probably remnants of the ancient ISPfe2 transposition events that led to the formation of Tn6097 in its native host). The predicted genetic modules of the transposon are indicated.
doi:10.1371/journal.pone.0032277.g001

Composite transposon Tn6097

The only composite transposon identified in this study was a novel element, designated Tn6097 (17,759 bp), which was captured in *P. ferrooxidans* NCCB 1300066 using the trap plasmid pCM132TC. A summary of the predicted ORFs of Tn6097, including their position, the size of the encoded proteins, and their closest homologs is presented in Table S2.

Transposition of Tn6097 into the selective cassette of pCM132TC generated 6-bp DRs. This transposon contains two identical, convergently oriented copies of ISPfe2 (1884 bp), that are responsible for its mobility (Figure 1F). Using trap plasmid pMAT1, we also demonstrated the transposition of a single copy of ISPfe2, which generated DRs of the same length as Tn6097 upon insertion into the *sacB* gene. ISPfe2 has been classified within the IS1634 family. It is the only member of this family identified so far in *Paracoccus* spp. The TPase of ISPfe2 and that encoded by the most closely related element, ISThsp7 of *Thiomonas* sp. 3As (accession no FP475956) (*Betaproteobacteria*), share 51% aa sequence identity.

The two copies of ISPfe2 flank the large core region of Tn6097, comprised of 13,991 bp. The average G+C content of this region is 63.6 mol%, which is lower than that of the total DNA of the native host *P. ferrooxidans* NCCB 1300066 (67 mol%). The core region contains 12 ORFs whose putative products show similarity to proteins conserved in many bacteria, including some involved in the metabolism of nitrogen compounds and resistance to chemother-

apeutics. Based on the predicted functions of the ORFs, they may be divided into three clusters, which can also be differentiated by the G+C content of their nucleotide sequences (data not shown).

The first cluster contains four ORFs (ORF2-ORF5) (Figure 1F) encoding putative proteins involved in nitrogen metabolism: (i) a FtrB/NarB-like regulator of transcription with significant aa sequence similarity (51%) to a positive transcription regulator of a respiratory nitrate reductase gene from a strain of *P. pantotrophus* [41], (ii) proteins involved in the response to NO [42] and (iii) a protein responsible for the transformation of 5-formyltetrahydrofolate and ATP into 5,10-methenyltetrahydrofolate, ADP and phosphate, which results in the creation of carbon-nitrogen bonds due to the cyclo-ligase activity [43].

The second cluster of Tn6097 contains four ORFs (ORF6-ORF9) (Figure 1F) with significant similarity to *arcABCD* genes, encoding the arginine deiminase pathway. This pathway permits bacteria to grow anaerobically with arginine as the substrate by catalyzing its conversion to ornithine, CO₂, and NH₃, with the generation of ATP [44]. The *arcABCD* genes encode (i) arginine/ornithine antiporter (ArcD), (ii) arginine deiminase (ArcA), (iii) ornithine carbamoyltransferase (ArcB) and (iv) carbamate kinase (ArcC), respectively. The *arcABCD* operon of Tn6097 exhibit the highest nucleotide sequence similarity (76%) and synteny to homologous genes in plasmid pOANT03 of *Ochrobactrum anthropi* ATCC 49188 (accession no. NC_009671).

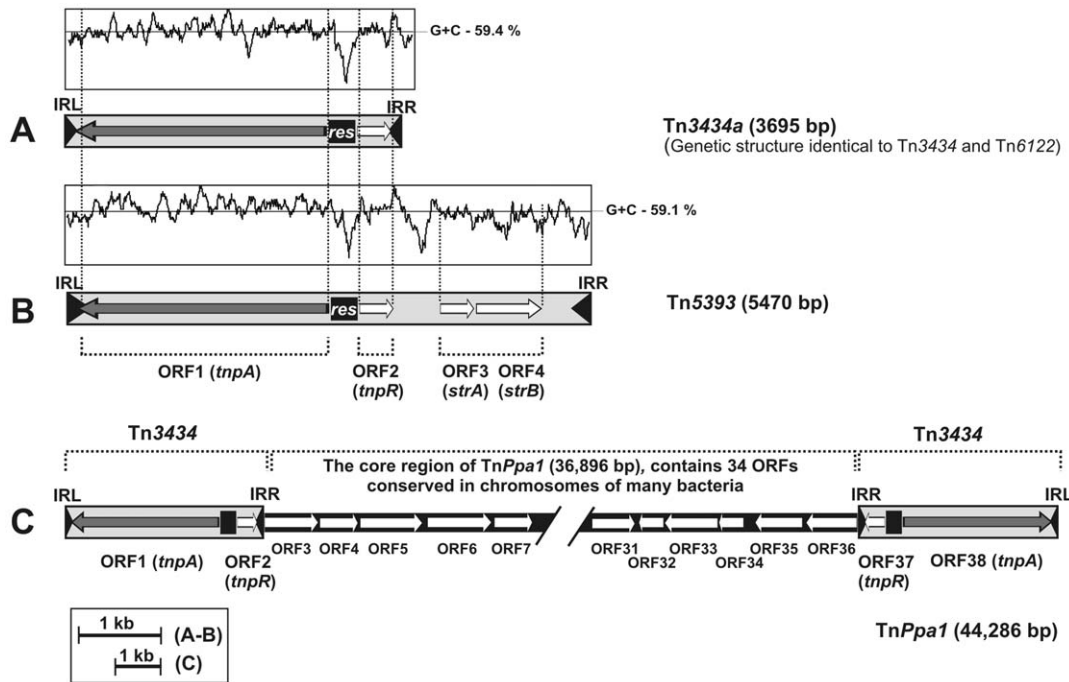


Figure 2. Genetic organization of the Tn3 family transposons captured in *Paracoccus* spp. using trap plasmids. The cryptic transposon Tn3434a (A), Tn5393 carrying streptomycin resistance genes (B) and the composite element TnPpa1 flanked by two identical copies of Tn3434 (C). Predicted coding regions are represented by arrows indicating the direction of transcription. Gray arrows indicate TPase genes (*tnpA*) and additional genes (including resolvase genes – *tnpR*) are indicated by white arrows. Inverted repeats (IRs) flanking transposons are marked by black arrowheads (IRL – left IR; IRR – right IR) and *res* sites are shown by black squares. Plots of the G+C content of the Tn3434a and Tn5393 sequences are shown in panels A and B (the average value is given on the right). doi:10.1371/journal.pone.0032277.g002

The third putative genetic module of Tn6097 consists of three ORFs (ORF10-ORF12) (Figure 1F) encoding (i) a putative secretion protein of the HlyD family, which groups efflux pumps involved in resistance to various antibiotics and heavy metals, and (ii) two putative ABC-type transporters with similarities to ATP-binding proteins DrrA and DrrB, which are components of a bacterial exporter complex. DrrA and DrrB confer resistance to daunorubicin and doxorubicin (chemotherapeutics of the anthracycline family) [45]. The antibiotics are active exclusively against gram-positive bacteria (they do not enter the cells of gram-negative strains). Tn6097 also carries a truncated ORF (ORF13; lacking its N-terminal coding part) adjacent to the right-hand copy of *ISPfe2*, which encodes a putative NAD-dependent epimerase/dehydratase that is most similar (85% aa identity) to a protein of *P. denitrificans* PD1222 (Figure 1F).

DNA hybridization analysis revealed that Tn6097 is present in one copy in the *P. ferrooxidans* NCCB 1300066 genome (data not shown). Using inverse PCR (IPCR) we determined the nucleotide sequence of its original target site in the host genome (see Materials and Methods for details). This revealed that Tn6097 is located within a gene (disrupted upon transposition), whose predicted product shows similarity to a number of bacterial proteases. The highest identity (67%) was observed with a *htpX* gene of *Polymorphum gilvum* SL003B-26A1 (accession no YP_004302375). The transposon was flanked by 6-bp long DRs (5'-GGCTCG-3'), confirming that it had been incorporated into the host genome by means of transposition.

Transposons of the Tn3 family

A common feature of all captured non-composite transposons (three elements) was the presence of two putative genetic modules

(responsible for transposition and resolution of cointegrates resulting from replicative transposition), which are highly conserved in members of the Tn3 family of transposons. Two of the identified transposons (Tn6122 and Tn3434a) were identified as cryptic elements, solely encoding genetic information required for transposition (Figure 2 and Table 2).

Tn6122 (3792 bp) was captured in *P. halophilus* JCM 14014 using the trap plasmid pMEC1. This element is composed of (i) identical 39-bp IRs, (ii–iii) two divergently oriented ORFs encoding a large transposase protein (TnpA; 967 aa) and a resolvase (TnpR) as well as (iv) an AT-rich putative recombination site (*res*) separating the ORFs (this region is involved in cointegrate resolution and regulation of expression of the *tnpA* and *tnpR* genes) [11]. Comparative sequence analysis revealed that homologous cryptic transposons occur within the genomes of three strains belonging to the *Alphaproteobacteria*: (i) *Sulfitobacter* sp. NAS-14.1, (ii) *Ruegeria* sp. PR1b and (iii) *Roseovarius* sp. TM1035. These transposons encode highly related transposases and resolvases, and are bordered by IRs identical to those of Tn6122 (data not shown).

The second trapped cryptic transposon, Tn3434a (3695 bp), identified using pEBB10 in *P. aminophilus* JCM 7686, is an isoform of the transposon Tn3434, that we trapped previously in *P. pantotrophus* DSM 11072 [17]. The newly identified isoform shares 98% nucleotide sequence identity with Tn3434 (36 mismatches). The resolvases encoded by these elements are 100% identical, while their transposases differ in 30 aa (data not shown).

Tn3434a is also related to Tn6122. As shown in Figure 2, both elements have an identical genetic organization. They encode highly related transposases (73% aa identity), but their resolvases are more divergent (28% identity).

In a previous study using pMEC1, we showed that *P. pantotrophus* LMD 82.5 contains a non-composite transposon Tn5393, which is another member of the Tn3 family [17]. This element, in contrast to the closely related Tn3434a and Tn6122, carries two streptomycin resistance genes (*strA* and *strB*) placed downstream of the *tnpR* gene (Figure 2). Interestingly, in the present study we observed transposition of this element (in the strain LMD 82.5) into the selection cassette of the trap plasmid pCM132TC, which revealed that this transposon can drive the transcription of genes placed downstream of the target site of transposition.

All three trapped non-composite transposons contained related IRs. The IRs of Tn3434a and Tn6122 differ slightly in length (38 bp and 39 bp, respectively) and their sequence identity is only 50% (14 mismatches). In contrast, the IRs of Tn5393 are much longer (81 bp with 4 mismatches between IRL and IRR), but their termini show significant similarity to those of Tn3434a, Tn6122 and other members of the Tn3 family (Figure S1). Despite their differences, the transposition of all these elements produced a 6-bp duplication of an AT-rich target sequence (Table 2).

During analysis carried out in *P. pantotrophus* DSM 11072, we captured another Tn3 element, TnPpa1 (driven by Tn3434), which was previously identified in this strain [18]. However, further analyses performed in the present study revealed an unusual feature of this element, which is described below.

Specific features of the identified TEs

Analyses performed using trap plasmids not only enabled the capture of many functional TEs but also identified specific features of several elements, including their ability to (i) activate transcriptionally silent genes, (ii) generate deletions within the target sites of transposition and (iii) transpose at unusually high frequency.

TEs enabling transcription of downstream genes

Using the trap plasmid pCM132TC (contains a promoterless tetracycline resistance gene as the selection cassette) we identified several TEs that are able to drive the transcription of promoterless genes placed downstream of the target site of transposition. The majority of the elements identified in this way were insertion sequences from four different groups: (i) the IS903 group of the IS5 family (*ISPam1* and *ISPpa8*), (ii) the IS407 group of the IS3 family (*ISPam3* and *ISPfe1*), (iii) the IS21 family (*ISPrk1*, *ISPve1*) and (iv) the IS256 family (*ISPaes3*). We also captured two transposons: (i) Tn5393 of the Tn3 family and (ii) the novel composite transposon Tn6097 bordered by two copies of *ISPfe2* (IS1634 family).

In order to localize the promoters responsible for the resistance phenotype, we analyzed the TEs trapped within pCM132TC in the Tc^r clones, as described in Materials and Methods. The results of these experiments (summarized in Figure 3) revealed that there are several possible ways in which transposition can deliver promoters to transcriptionally silent genes. The transcription of nearby genes can be driven from (i) outwardly oriented promoters located in the terminal parts of the ISs – this is the case for all tested members of the IS5 (IS903 group), IS3 (IS407 group) and IS21 families (Figure 3A), (ii) a hybrid promoter, most probably composed of a –35 hexamer present in the terminal part of the TE and a –10 hexamer located in close proximity to the target site of transposition (suggested for the IS256 member) (Figure 3B), (iii) the promoter of the transposase gene (Tn5393 of the Tn3 family) (Figure 3C), or (iv) a promoter present in a core region of the composite transposon Tn6097 (Figure 3D).

An unexpected phenomenon was observed upon selection of Tc^r clones in *P. pantotrophus* DSM 11073 containing pCM132TC. All of the obtained Tc^r mutants lacked the autonomous form of the trap plasmid, and DNA hybridization analysis revealed that in

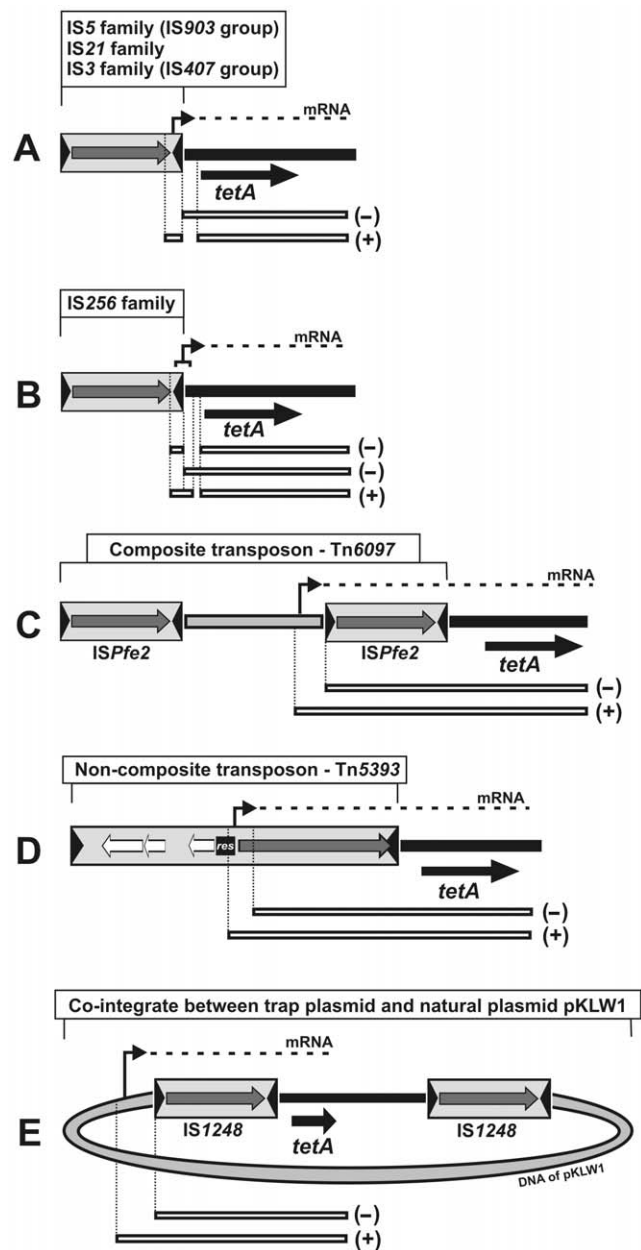


Figure 3. Schematic overview of five ways in which transposition can deliver promoters to the transcriptionally silent *tetA* (tetracycline resistance) gene of the trap plasmid pCM132TC. The location of promoters in the plasmids pCM132TC::TE, conferring a Tc^r phenotype, are appropriately indicated: an outwardly oriented promoter in the terminal parts of a TE (A), a hybrid promoter composed of a –35 hexamer (delivered by the TE) and a –10 hexamer located in close proximity to the target site of transposition (B), the promoter of a TPase gene (C), a promoter present in the core region of a composite transposon (D), and a promoter derived from another plasmid delivered by the generation of transient co-integrates resulting from replicative transposition (E). DNA fragments used in the localization of the promoters are shown as open thin boxes below each panel. The activity of promoters (tested in *Paracoccus* spp.) accompanying the presence of the DNA fragments is indicated on the right: (+) promoter activity, (–) lack of promoter activity.

doi:10.1371/journal.pone.0032277.g003

every case, the plasmid was present as part of a co-integrate generated between pCM132TC and plasmid pKLW1 (approx. 100 kb) which naturally occurs in strain DSM 11073.

DNA sequencing revealed that the recombinational event that led to the formation of these co-integrates occurred within the selection cassette of pCM132TC (upstream of the *tetA* gene). The inserted trap plasmid was bordered by two copies of the insertion sequence IS1248 (IS427 group of the IS5 family), which strongly suggests that co-integrates were generated upon replicative transposition of this IS (Figure 3E).

The co-integrates therefore represent intermediate forms of transposition, which should be resolved by homologous recombination into individual, separate plasmids. After approximately 100 generations of growth of the strains under non-selective conditions we observed (by DNA electrophoresis; data not shown) the appearance of a DNA band corresponding to the resolved trap plasmid. In contrast, we were never able to detect individual replicons when plasmid DNA was isolated from bacteria grown in medium supplemented with tetracycline, which strongly suggests that this antibiotic provides selection for the maintenance of the co-integrates.

DNA sequencing of the selection cassette of the resolved trap plasmid confirmed the presence of an inserted copy of IS1248. When introduced into the strain DSM 11073, this plasmid (pCM132TC::IS1248), did not confer the Tc^r phenotype, which precludes the possibility that a promoter of the IS may drive transcription of a nearby gene.

These observations indicate that expression of the *tetA* gene in the co-integrate is driven by a promoter located within pKLW1. This conclusion is supported by the DNA sequence of regions flanking the original insertion site of IS1248 within pKLW1 and the detection of activity of the upstream promoter (Figure 3E).

TEs generating deletions

The transposition of the vast majority of TEs identified in this study resulted in the simple insertion of the elements into the trap plasmids and the generation of DRs. However, we found that the insertion of two ISs (IS*Pmar2* of *P. marcusii* DSM 11574 and IS*Pve1* of *P. versutus* UW400) into selection cassettes could be accompanied by deletions at the transposition target site.

In the case of IS*Pmar2* (IS407 group of the IS3 family) we identified only one deletion mutant (out of five analyzed pMEC1-derivatives containing inserted IS*Pmar2*). The deletion (15 bp) occurred at the 3'-end of the inserted IS, and therefore the intact element was not bordered by the DRs, in contrast to the other clones tested.

More interesting was the case of IS*Pve1* (IS21 family), which frequently generated much larger deletions within the selection cassettes of various trap plasmids. Initial analysis of a pool of Tc^r clones of strain UW400 (obtained with pMEC1) revealed the presence of two types of pMEC1 mutants representing either (i) simple insertion of IS*Pve1* (7% of Tc^r clones), or (ii) insertion of IS*Pve1* associated with deletions within the trap plasmid (16%). The remaining Tc^r clones did not contain any TEs.

In the case of simple insertions, transposition of IS*Pve1* resulted in the generation of 8-bp long DRs. Moreover, in both the insertion and deletion mutants, the element was always placed in the same orientation within the *cI* gene of pMEC1 (Figure 4). We further analyzed the IS*Pve1* insertion sites that underwent structural alterations and found that the deletions were always unidirectional, and they comprised different segments of the trap plasmid adjacent to the 5' end of the IS (Figure 4). The deletions ranged in size from 0.5 to 4.5 kb. (Figure 4). Analogous deletions associated with the transposition of IS*Pve1* were also observed with other trap plasmids (e.g. pMMB2 and pMAT1) and in each case, the location of the replication system was the factor limiting the

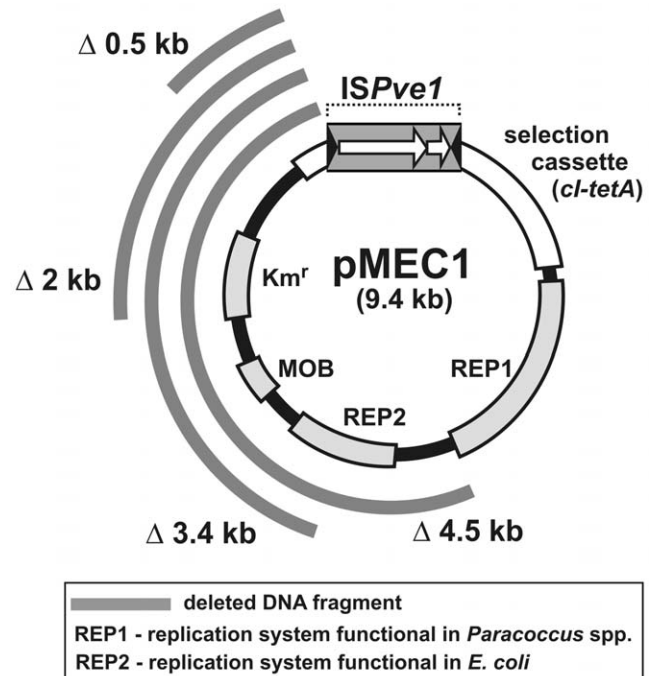


Figure 4. Deletions within the trap plasmid pMEC1 generated in *P. versutus* UW400 upon transposition of IS*Pve1*. The range of the deletion within four individual derivatives of pMEC1 is shown by curved gray lines.

doi:10.1371/journal.pone.0032277.g004

range of deletions generated within the plasmid (Figure 4 and data not shown).

Frequently transposing TEs

In most cases, the frequency of transposition of the captured TEs ranged from 10^{-8} to 10^{-6} depending on the element and the bacterial strain (see Table S3). The lowest frequency (10^{-10}) was observed for IS*Paes3* of *P. aestuarii* DSM 19484. In contrast, the element most dynamic in transposition was Tn5393 (*P. pantotrophus* LMD 82.5), which transposed into the selection cassette of the trap plasmid pCM132TC at a very high frequency (10^{-3}). This is in agreement with our previous observations performed in the strain LMD 82.5 with pMEC1 [17].

An interesting phenomenon was observed after conjugal transfer of the trap plasmid pMEC1 into *P. pantotrophus* DSM 11072. Surprisingly, in this case, the vast majority of the transconjugants (92%) did not contain pMEC1 (9.4 kb) but they carried a very large replicon (>50 kb), that was not originally present in DSM 11072. The remaining transconjugants (8%) contained intact pMEC1. DNA hybridization followed by DNA sequencing revealed that the large nascent plasmids are in fact insertion mutants of pMEC1. They were generated by transposition into the trap plasmid of a large previously described composite transposon Tn*Ppa1* (44,286 bp) [18]. This transposon (considered to represent a transposable genomic island) is composed of two divergently oriented copies of the Tn3 family transposon Tn3434, which border the core region of the element containing a set of genes conserved in the chromosomes of many bacteria (Figure 2C). The core region contains a large set of putative genes, whose products show similarity to enzymes involved in central intermediary metabolism (e.g. tricarboxylic acid cycle or 2-methylcitrate cycle), transporters, transcriptional regulators and conserved proteins of unknown function [18].

Tn*Ppa1* has previously been captured in strain DSM 11072R using pMMB2, although its transposition into the selection cassette of the trap plasmid was a relatively rare event (10^{-6}) [18]. In the present study, we observed massive transposition of Tn*Ppa1* into pMEC1. Intriguingly, all the transposon insertions were detected in randomly tested Km^r transconjugants, i.e. without applying positive (tetracycline) selection for the identification of transposition events. DNA sequencing of several insertion derivatives of pMEC1 confirmed that transposition of Tn*Ppa1* had targeted random parts of the trap plasmid. In each case the inserted element was bordered by DRs, of a length typical for other members of the Tn3 family (5 bp) (data not shown).

We also tested whether Tn*Ppa1* is able to transpose as frequently into other replicons. For this purpose, several mobilizable plasmids were transferred into strain DSM 11072R. The plasmids contained different replication systems derived from (A) three broad host range replicons – (i) RA3 (pMAO-oriT, pMAO-MS and pMAO-RK which contain different systems enabling mobilization for conjugal transfer), (ii) RK2 (pCM132TC) and (iii) pBBR1 (pMAT1, pBBR1MCS-3), as well as from (B) three plasmids originating from *Paracoccus* spp. – (i) pMAR4 of *P. marcusii* DSM 11574 (pDG12), (ii) pAMI7 of *P. aminophilus* JCM 7686 (pDIY703) and (iii) pTAV1 of *P. versutus* UW1 (pMMB2).

Analysis of the plasmid content of the obtained transconjugants revealed massive transposition of Tn*Ppa1* into plasmids containing replication systems of pMAR4 and pAMI7 as well as into all derivatives of RA3 (20–95% of tested clones carried Tn*Ppa1*, depending on the plasmid used), but not into the other tested plasmids (data not shown). This strongly suggested that the observed transposition phenomenon is dependent on the nature of the incoming replicon.

Copy number, location and distribution of TEs in the genomes of *Paracoccus* spp.

Southern blotting and DNA hybridization analysis were performed to examine the copy number and genomic location (plasmid/chromosome) of the identified TEs in their natural hosts.

The analysis revealed that most of the identified insertion sequences were present in multiple copies in *Paracoccus* spp., which most probably reflects the replicative transposition of some of these elements. As shown in Table 2, the highest copy numbers (4–9 copies) were observed for most members of (i) the IS427 group of the IS5 family (IS1248f, IS*Paes1*, IS*Pbe2*, IS*Pha1*, IS*Ppa2*, IS*Ppa6*, IS*Pze1*), (ii) the IS903 group of the IS5 family (IS*Pko1*, IS*Plc1*, IS*Ppa3a*, IS*Ppa8*), (iii) the IS407 group of the IS3 family (IS*Pbe1*, IS*Pfe1*), and individual members of (iv) the IS21 family (IS*Pve1a*) and (v) the IS66 family (IS*Ppa7*). Some of the detected IS copies displayed different signal intensities in the hybridization analysis, which most probably reflects divergence in the nucleotide sequences of closely related elements or the presence of truncated copies of the same element (data not shown).

In contrast, a low copy number (1–2 copies) was observed for all members of (i) the IS5 group of the IS5 family (IS*Pha2*, IS*Phae1*, IS*Pmar3*), (ii) the IS6 family (IS*Pmar1*, IS*Ppa9*) and (iii) the IS256 family (IS*Paes3*), as well as for the identified transposons (we detected single copies of the non-composite transposons Tn5393 and Tn6122, and of the composite elements, Tn*Ppa1* and Tn6097).

Hybridization analysis also revealed that only a few of the TEs identified in this study are located within natural plasmids of *P. aminophilus* JCM 7686, *P. halophilus* JCM 14014, *P. kondratievae* NCIMB 13773^T, and *P. marcusii* OS22 (Table 2). Thus, the vast majority of the TEs seem to be of chromosomal origin. However, it is worth mentioning that all of the tested paracoccal strains carry

megaplasmids, which cannot be purified using a standard alkaline lysis procedure. The question of whether the identified transposable elements reside in these high-molecular weight plasmids remains open.

In the final part of this study we investigated whether the captured TEs are specific only for the host strains, or are widespread among *Paracoccus* spp. (the analysis was extended to all so-far identified TEs of *Paracoccus* spp.; Figure 5). Dot blot hybridization analysis was performed using the DNA probes described above and total DNAs of all tested strains of *Paracoccus* spp. (see Materials and Methods for details). As shown in Figure 5, this analysis revealed that TEs are very widely distributed in *Paracoccus* spp. and that each strain contains several elements from different families. The most common are members of the IS5 and IS3 families. In contrast, the elements IS1182 and IS*As1* appear to be less abundant, which suggests that they may have been acquired by HGT quite recently (Figure 5).

Discussion

In this study we have identified and analyzed transposable elements of bacteria of the genus *Paracoccus*. For the identification of TEs we used trap plasmids, enabling the positive selection of transposition events. All of the elements were, therefore, defined by their ability to transpose into the selective cassette of these plasmids, which unequivocally confirmed their mobility. This is the first such complex analysis aimed at defining a functional transposable part of the mobilome of a number of strains representing a significant part of one bacterial genus.

The genus *Paracoccus* currently comprises 31 species. In this study we analyzed 25 strains of 20 species isolated from various environments. These represent almost all species of this genus (excluding patented strains) that were available in 2009, when this project started. In total, we tested approximately 7000 individual clones to identify trap plasmid insertion derivatives. However, despite the complexity of this analysis, it is obvious that not all the TEs residing in *Paracoccus* spp. were defined; we only identified those that are the most dynamic in transposition, i.e. elements of highest evolutionary impact. As a result, we captured 41 elements representing (i) insertion sequences, (ii) an IS-driven composite transposon and (iii) non-composite transposons of the Tn3 family.

In total, together with our previous studies [3,17–20], we have described over 50 novel TEs of *Paracoccus* spp. As expected, the vast majority of the captured elements were insertion sequences. We identified 47 elements, representing 10 IS families: IS3, IS5, IS6, IS21, IS66, IS256, IS1182, IS1380, IS1634 and IS*As1*. Interestingly we did not encounter elements of the IS1, IS4, IS91 or IS110 families, which are widely distributed among other bacteria [46,47].

The most numerous in *Paracoccus* spp. were members of the IS5 family (26 elements), while only single elements from the IS*As1*, IS1182 and IS1634 families were captured. However, dot blot DNA hybridization revealed that TEs are much more abundant in *Paracoccus* spp. than we had expected. These results suggest that each tested strain carries elements belonging to several different families (Figure 5), the most abundant of which are members of the IS3, IS5, IS21, IS66 and IS256 families. In contrast, elements of the IS*As1* and IS1182 families seem to be rare in *Paracoccus* spp. (Figure 5).

The obtained data are consistent with the results of a survey of the ISfinder database to distinguish all ISs identified so far in *Alphaproteobacteria* (Figure 6). This revealed that members of the IS5 and IS3 families predominate in this group of bacteria, while ISs of other families are unevenly distributed. In the light of this analysis,

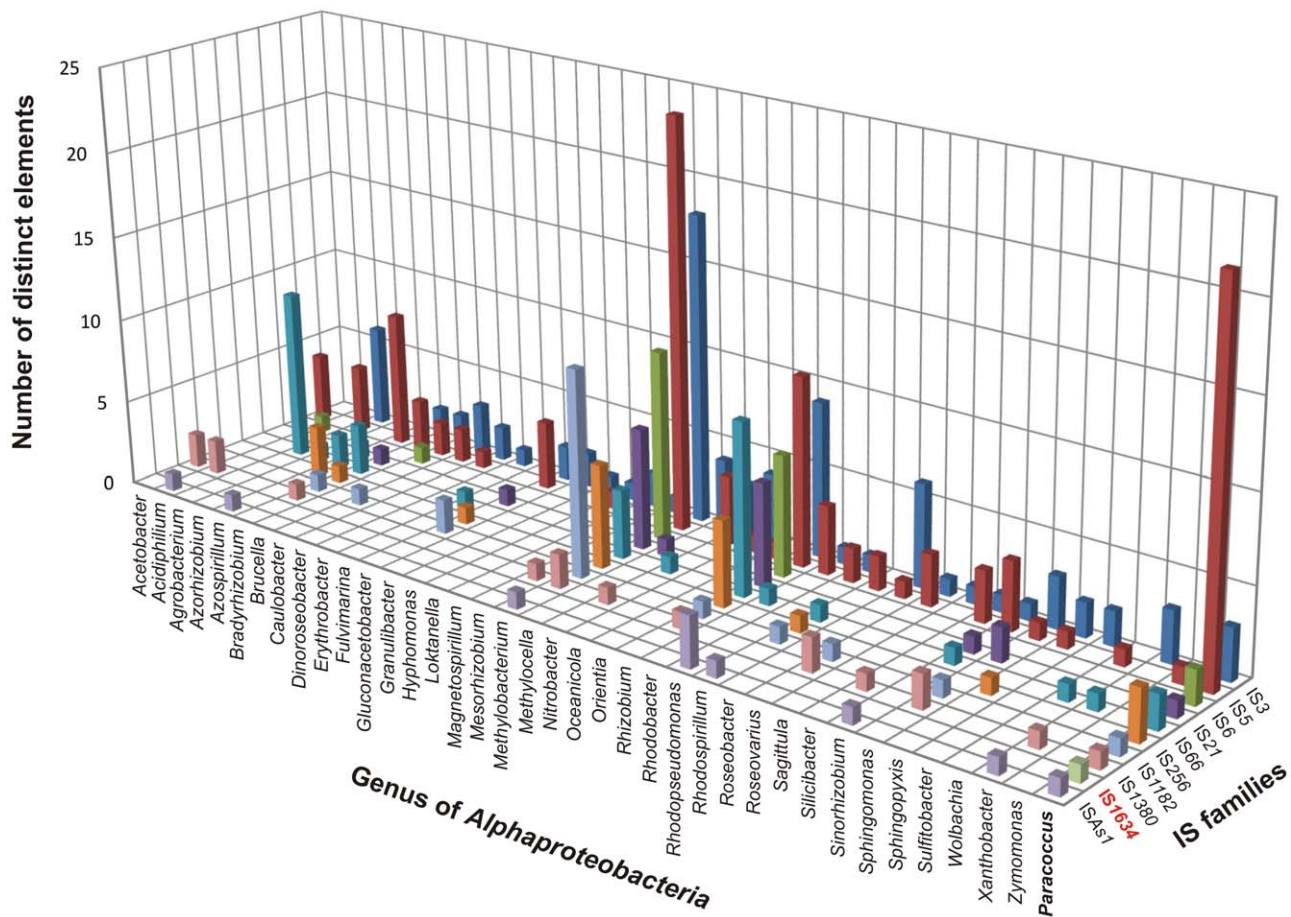


Figure 6. Distribution of IS-families identified in *Paracoccus* spp., among other genera of the class Alphaproteobacteria (ISfinder database). The *Paracoccus* spp. elements are members of 10 IS families (IS3, IS5, IS6, IS21, IS66, IS256, IS1182, IS1380, IS1634, ISAs1). Each IS-family is represented by a different color on the three dimensional histogram. The family IS1634 unique to the genus *Paracoccus* is shown in red. doi:10.1371/journal.pone.0032277.g006

These chemotherapeutics are active exclusively against gram-positive bacteria, since they do not enter the cells of gram-negative strains [48]. The second module contains putative genes involved in the arginine deiminase (ADI) pathway [45]. The ADI pathway comprises reactions catalyzed by arginine deiminase, ornithine transcarbamylase and carbamate kinase (encoded by the *arcABC* genes), which enable bacteria to grow anaerobically in the presence of arginine. The enzymes mediate the conversion of arginine to ornithine, NH_3 , and CO_2 , with the generation of ATP. It has been shown that this pathway constitutes a major source of energy for several microorganisms, e.g. strains of *Mycoplasma* spp., *Pseudomonas* spp. and *Bacillus* spp. [49].

However, we have observed that *P. ferrooxidans* NCCB 1300066 (host strain of Tn6097) is unable to grow anaerobically in the presence of arginine (data not shown), which suggests that the *arcABCD* operon may be not functional. It has been previously demonstrated that anaerobic expression of the closely related operon of *Pseudomonas aeruginosa*, requires a transcriptional activator [50], which is not encoded within Tn6097. It is therefore possible that the ADI module of Tn6097 may be functional exclusively in the strains encoding a compatible transcription regulator.

We have also identified several non-composite TEs of the Tn3 family, which represent different stages in the evolution of this group of transposons. It is believed that the Tn3 transposons originate from progenitor IS elements (several such ISs coding for

the Tn3-like transposase have been identified; ISfinder), which acquired an additional genetic module involved in site-specific resolution of cointegrates (intermediates of replicative transposition) as well as other genes whose products can directly affect the phenotype of their host [51].

In *Paracoccus* spp. we identified two cryptic transposons (Tn6122 and Tn3434a) that contain the conserved module for site-specific recombination. We also captured the transposon Tn5393 (closely related to Tn3434a), which apart from the resolvase module, carries two additional streptomycin resistance genes (Figure 2). Interestingly, analysis of the G+C profiles of these transposon nucleotide sequences revealed that the individual genetic modules of the elements are bordered by short stretches of AT-rich sequence (Figure 2), which presumably constitute traces of consecutive recombinational events that led to incorporation of the different portions of genetic information into the transposons.

The next step in the evolution of Tn3 elements is represented by Tn*Ppa1* (44,286 bp) of *P. pantotrophus* DSM 11072, which can be considered a composite transposon generated by two copies of the non-composite Tn3434 [18]. The identification of Tn*Ppa1* provides evidence that the Tn3-family transposons (similarly to ISs) are able to mobilize large segments of genomic DNA for transposition. The generation of such composite elements seems to be very rare because of transposition immunity (a phenomenon that is thought to apply to members of the Tn3 family), which

precludes transposition of more than one copy of the element into a single replicon [11]. Therefore Tn*Ppa1*, composed of two identical copies of Tn*3434*, has a unique structure and, to our knowledge, is the only element of this type identified so far.

In this study, we found that Tn*Ppa1* is able to transpose into random sites of several mobilizable plasmids with unusually high frequency, not previously observed for any other bacterial TE. Our observations strongly suggest that the massive transposition of Tn*Ppa1* might result from the specific features of the transferred plasmids. Irrespective of the precise mechanism of Tn*Ppa1* transposition, it is clear that targeting of mobilizable or conjugal plasmids contributes to the dispersal of this element among bacterial populations. Within its core region, Tn*Ppa1* carries a number of genes (encoding putative enzymes involved in central intermediary metabolism, membrane transporters, transcription regulators and proteins of unknown function), whose presence may potentially improve the ecological fitness of the host cells [18]. Transposition of such a large portion of genetic information (containing putative housekeeping genes) into co-residing plasmids significantly enriches the pool of mobile DNA, which may then be spread by HGT among even phylogenetically-distinct bacterial hosts, to produce a variety of phenotypic effects [18]. The transposition of TEs containing housekeeping genes into bacterial plasmids may also have a significant impact on the structure and evolution of bacterial genomes, e.g. it may stimulate the formation of multi-chromosome genomes, which are very common in *Alphaproteobacteria* (the presence of two chromosomes has also been demonstrated for two paracoccal strains, whose genomes have been fully sequenced) [52,53].

In this study, we also observed very frequent transposition (10^{-3}) of another member of the Tn3 family: the streptomycin resistance Tn*5393* (*P. pantotrophus* LMD 82.5) – far higher than that observed for other non-composite members of this family (Table S3) (e.g. the transposition frequency of Tn*3434* was 10^{-6}) [17]. Such dynamic transposition of Tn*5393* explains the wide dissemination of this transposon in many bacterial isolates [54,55] and is evidence that transposition may be a key mechanism responsible for the natural amplification of antibiotic resistance genes in environmental bacterial strains.

We have found that the Tn*5393* TPase gene carries a strong promoter, which is able to drive the transcription of genes placed downstream of the target sites of transposition. It is therefore highly probable that the frequent transposition of Tn*5393* is directly linked with the strength of this promoter.

Using the trap plasmid pCM132TC containing a promoterless tetracycline resistance gene, we identified a pool of TEs whose promoters are able to drive transcription of genes placed downstream of the target site of transposition. Such elements therefore represent portable expression systems that enable activation of single promoterless genes or operons as well as foreign genes introduced by HGT whose native promoters are non-functional in their new hosts. Thus transposition of these elements may result in the immediate appearance of novel phenotypes of adaptive value [13].

We showed that activation of nearby genes is a typical feature of insertion sequences of the IS3, IS21 and IS256 families, which corroborates the findings of previous studies [56,57]. Moreover, for the first time, we have demonstrated such ability for (i) elements of the IS903 group of the IS5 family (contain an outwardly oriented promoter in the terminal part of the elements), and (ii) a transposon of the Tn3 family (the aforementioned Tn*5393*).

The individual transposition events which led to the Tc^r phenotype (activation of the promoterless *tetA* gene of pCM132TC) were examined in detail. This analysis revealed that transposition-

mediated delivery of promoters to transcriptionally silent genes can occur in at least five ways (Figure 3). The most intriguing was observation that transcription can be driven by promoters derived from natural plasmids, which form transient co-integrates with other co-residing replicons as a result of replicative transposition. We observed that such co-integrates can be quite stably maintained even in the absence of selective pressure (tetracycline), which supports their role as efficient expression systems.

Another interesting phenomenon observed in this study was connected with the insertion sequence IS*Pve1* (IS21 family), which was prone to generate deletions within target DNA molecules (Figure 4). The transposition in a cell of IS*Pve1* therefore results in the generation of a mixed population of plasmid molecules. The resulting diminished forms of the replicons may potentially differ in their specific properties (e.g. stability and copy number). This indicates that transposition may significantly shape plasmid genomes and promote plasmid diversity, which could have significant evolutionary implications.

Interestingly, this “deletion” phenomenon was not observed for other members of the IS21 family identified in this study, including IS*Pve1a*, an isoform of IS*Pve1* from *P. bengalensis* DSM 17099 (99% identity at the nucleotide sequence level). The high frequency of deletions generated by IS*Pve1* (in comparison with “pure” insertion events) might therefore be a specific feature of this element. Alternatively, it might result from the presence of host factors, which stimulate such deletions. The latter hypothesis is supported by data recently published by Kusumoto *et al.* [58], which provide evidence that deletions of ISs or DNA regions adjacent to them might be promoted (in an IS transposase-dependent manner) by a chromosomally-encoded protein IEE (IS-excision enhancer). Further studies on IS*Pve1* and the strain *P. versutus* UW400 are necessary to determine the likelihood of these two possibilities.

To conclude, the analysis performed in this study together with our previous work concerning the identification of diverse TMs generated by a single copy of IS*Pme1* [3], have highlighted the diversity and wide distribution of TEs in *Paracoccus* spp. It should also be borne in mind that many TEs (including composite transposons and TMs) cannot be distinguished in bacterial genomes by classical *in silico* sequence analysis, unless they have been inserted into a highly conserved genetic context (e.g. a trap plasmid selection cassette), to enable precise definition of their termini. For this reason, the identification of functional TEs using trap plasmids may produce many interesting and surprising findings even when carried out in bacteria whose genomes have been fully sequenced.

Our analysis of the transposable part of the mobilome of *Paracoccus* spp. has also generated considerable interesting and unique data concerning the dynamics of the process of transposition (Table S3). The results also indicate the powerful role of transposition in the dissemination of diverse genetic information (possibly of adaptive value) by HGT. Considering their status as one of the most recombinogenic factors in bacterial genomes, TEs can be considered as the major driving force in the evolution of prokaryotes.

Supporting Information

Figure S1 Multiple alignment of nucleotide sequences of the terminal inverted repeats (IRL, IRR) of IS-families identified in *Paracoccus* spp. (IS3, IS5, IS6, IS21, IS66, IS256, IS1182, IS1380, IS1634, ISAs1). The sequences used for the comparisons were obtained from the ISfinder database (<http://www-is.biotoul.fr/is.html>). Sequence alignments were performed with the Pictogram program (WE-

BLOGO; <http://weblogo.berkeley.edu/logo.cgi>). The size of the letter in Pictogram represents the frequency of the occurrence of that nucleotide at each position. In the consensus sequences, shown below each Pictogram, uppercase letters indicate conservation within the IS family, lowercase letters indicate the predominant nucleotides, and dots indicate non-conserved residues. The family and group specific DDE motifs determined by multiple alignment of the amino acid sequences of the transposases of ISs present in the ISfinder database has been shown in the middle of the figure ([5] and this study). Residues forming the DDE motif are indicated with a black background. The N2, N3 and C1 domains are enclosed in boxes and labeled. The numbers in parentheses are the distances (in amino acids – aa) between the residues forming the DDE motif. The consensus IR and DDE motif sequences for each IS-family is compared with all sequences identified so far in *Paracoccus* spp.

(TIF)

Table S1 Oligonucleotides used in this study.

(DOC)

References

- Kidwell MG, Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* 55: 1–24.
- Hübner A, Hendrickson W (1997) A fusion promoter created by a new insertion sequence, IS1490, activates transcription of 2,4,5-trichlorophenoxyacetic acid catabolic genes in *Burkholderia cepacia* AC1100. *J Bacteriol* 179: 2717–2723.
- Bartosik D, Putyrski M, Dziejewicz L, Malewska E, Szymanik M, et al. (2008) Transposable modules generated by a single copy of insertion sequence IS*Pme1* and their influence on structure and evolution of natural plasmids of *Paracoccus methylotens* DM12. *J Bacteriol* 190: 3306–3313.
- Mahillon J (1998) Transposons as gene haulers. *APMIS Suppl* 84: 29–36.
- Chandler M, Mahillon J (2002) Insertion sequences revisited. In: Craig NL, Craigie M, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington (DC): American Society for Microbiology. pp 305–366.
- Toleman MA, Bennett PM, Walsh TR (2006) ISCR elements: novel gene-capturing systems of the 21st century? *Microbiol Mol Biol Rev* 70: 296–316.
- De Palmenaer D, Vermeiren C, Mahillon J (2004) IS237-MIC231 elements from *Bacillus cereus* sensu lato are modular. *Mol Microbiol* 53: 457–467.
- Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, et al. (2000) The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature* 403: 665–668.
- Peters JE, Craig NL (2001) Tn7: smarter than we thought. *Nat Rev Mol Cell Biol* 2: 806–814.
- Peters JE, Craig NL (2001) Tn7 recognizes transposition target structures associated with DNA replication using the DNA-binding protein TnsE. *Genes Dev* 15: 737–747.
- Grindley NDF (2002) The movement of Tn3-like elements: transposition and cointegrate resolution. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington (DC): American Society for Microbiology. pp 272–302.
- Wyndham RC, Cashore AE, Nakatsu CH, Peel MC (1994) Catabolic Transposons. *Biodegradation* 5: 323–342.
- Siguié P, Filée J, Chandler M (2006) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* 9: 526–531.
- Kelly DP, Euzéby JP, Goodhew CF, Wood AP (2006) Redefining *Paracoccus denitrificans* and *Paracoccus pantotrophus* and the case for a reassessment of the strains held by international culture collections. *Int J Syst Evol Microbiol* 56: 2495–2500.
- Kumaraswamy R, Sjollem K, Kuenen G, van Loosdrecht M, Muyzer G (2006) Nitrate-dependent [Fe(II)EDTA]2- oxidation by *Paracoccus ferroxidans* sp. nov., isolated from a denitrifying bioreactor. *Syst Appl Microbiol* 29: 276–286.
- Solyga A, Bartosik D (2004) Entrapment vectors - how to capture a functional transposable element. *Pol J Microbiol* 53: 139–144.
- Bartosik D, Sochacka M, Baj J (2003) Identification and characterization of transposable elements of *Paracoccus pantotrophus*. *J Bacteriol* 185: 3753–3763.
- Mikosa M, Sochacka-Pietal M, Baj J, Bartosik D (2006) Identification of a transposable genomic island of *Paracoccus pantotrophus* DSM 11072 by its transposition to a novel entrapment vector pMMB2. *Microbiology* 152: 1063–1073.
- Bartosik D, Szymanik M, Baj J (2003) Identification and distribution of insertion sequences of *Paracoccus solventivorans*. *Appl Environ Microbiol* 69: 7002–7008.
- Szuplewska M, Bartosik D (2009) Identification of a mosaic transposable element of *Paracoccus marcusii* composed of insertion sequence IS*Pnar4* (IS*A1* family) and an IS1247a-driven transposable module (TMO). *FEMS Microbiol Lett* 292: 216–221.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Schneider D, Faure D, Noirclerc-Savoye M, Barriere AC, Coursange E, et al. (2000) A broad-host-range plasmid for isolating mobile genetic elements in gram-negative bacteria. *Plasmid* 44: 201–207.
- Gay P, Le Coq D, Steinmetz M, Berkelman T, Kado CI (1985) Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. *J Bacteriol* 164: 918–921.
- Kovach ME, Phillips RW, Elzer PH, Roop RM, Peterson KM (1994) pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* 16: 800–802.
- Dziejewicz L, Adamczuk M, Szuplewska M, Bartosik D (2011) DIY series of genetic cassettes useful in construction of versatile vectors specific for *Alphaproteobacteria*. *J Microbiol Methods* 86: 166–174.
- Smorawska M, Szuplewska M, Zaleski P, Wawrzyniak P, Maj A, Plucienniczak A, Bartosik D (2012) Mobilizable narrow host range plasmids as natural suicide vectors enabling horizontal gene transfer among distantly related bacterial species. *FEMS Microbiol Lett* 326: 76–82.
- Bartosik D, Szymanik M, Wysocka E (2001) Identification of the partitioning site within the repABC-type replicon of the composite *Paracoccus versutus* plasmid pTAV1. *J Bacteriol* 183: 6234–6243.
- Kushner SR (1978) An improved method for transformation of *E. coli* with ColE1-derived plasmids. In: Boyer HB, Nicosia S, eds. *Genetic Engineering*. Amsterdam: Elsevier/North-Holland. pp 17–23.
- Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7: 1513–1523.
- Chen W, Kuo T (1993) A simple and rapid method for the preparation of gram-negative bacterial genomic DNA. *Nucleic Acids Res* 21: 2260.
- Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Dziejewicz L, Dmowski M, Baj J, Bartosik D (2010) Plasmid pAM12 of *Paracoccus aminophilus* JCM 7686 carries N,N-dimethylformamide degradation-related genes whose expression is activated by a LuxR family regulator. *Appl Environ Microbiol* 76: 1861–1869.
- Siguié P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34: 32–36.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944–945.
- Campbell A, Starlinger P, Berg DE, Botstein D, Lederberg EM, et al. (1979) Nomenclature of transposable elements in prokaryotes. *Plasmid* 2: 466–473.
- Sekine Y, Ohtsubo E (1989) Frameshifting is required for production of the transposase encoded by insertion sequence 1. *Proc Natl Acad Sci U S A* 86: 4609–4613.
- Zheng J, McIntosh MA (1995) Characterization of IS1221 from *Mycoplasma hyorhinis*: expression of its putative transposase in *Escherichia coli* incorporates a ribosomal frameshift mechanism. *Mol Microbiol* 16: 669–685.
- Friedman DI (1988) Integration host factor: a protein for all reasons. *Cell* 55: 545–554.
- Chalmers R, Guhathakurta A, Benjamin H, Kleckner N (1998) IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* 93: 897–908.
- Wood NJ, Alizadeh T, Bennett S, Pearce J, Ferguson SJ, et al. (2001) Maximal expression of membrane-bound nitrate reductase in *Paracoccus* is induced by nitrate via a third FNR-like regulator named NarR. *J Bacteriol* 183: 3606–3613.

Table S2 ORFs located within Tn6097 of *Paracoccus ferroxidans* NCCB 1300066.

(DOC)

Table S3 Rate of transposition of the TEs of *Paracoccus* spp. identified in this study.

(DOCX)

Acknowledgments

We acknowledge (i) J.-W. Bae for providing *P. aestuarii* DSM 19484, (ii) L. Drewniak and A. Sklodowska for strain *P. marcusii* OS22, (iii) J. Hirschberg for *P. marcusii* DSM 11574 and (iv) A. Lipski for *P. alkenifer* DSM 11593.

Author Contributions

Conceived and designed the experiments: DB LD. Performed the experiments: LD JB MS AM MT AC GS MA TS PS AT KW EW EP. Analyzed the data: DB LD MS. Contributed reagents/materials/analysis tools: DB. Wrote the paper: DB LD.

42. Kwiatkowski AV, Laratta WP, Toffanin A, Shapleigh JP (1997) Analysis of the role of the *mrrR* gene product in the response of *Rhodobacter sphaeroides* 2.4.1 to exogenous nitric oxide. *J Bacteriol* 179: 5618–5620.
43. Ogwang S, Nguyen HT, Sherman M, Bajaksouzian S, Jacobs MR, et al. (2011) Bacterial conversion of folinic acid is required for antifolate resistance. *J Biol Chem* 286: 15377–15390.
44. Maghnoij A, de Sousa Cabral TF, Stalon V, Vander Wauwen C (1998) The *arcABDC* gene cluster, encoding the arginine deiminase pathway of *Bacillus licheniformis*, and its activation by the arginine repressor *argR*. *J Bacteriol* 180: 6468–6475.
45. Malla S, Niraula NP, Singh B, Liou K, Sohng JK (2010) Limitations in doxorubicin production from *Streptomyces peucetius*. *Microbiol Res* 165: 427–435.
46. De Palmenaer D, Siguier P, Mahillon J (2008) IS4 family goes genomic. *BMC Evol Biol* 23: 8–18.
47. Siguier P, Gagnevin L, Chandler M (2009) The new IS1595 family, its relation to IS1 and the frontier between insertion sequences and transposons. *Res Microbiol* 160: 232–241.
48. Gumpert J, Dornberger K, Smith TH (1982) Antimicrobial activities of daunorubicin and adriamycin derivatives on bacterial and protoplast type L- from cells of *Bacillus subtilis* 170, *Escherichia coli* B, and *Proteus mirabilis* VI. Structure, activity relationship. *Z Allg Mikrobiol* 22: 687–692.
49. Zúñiga M, Pérez G, González-Candelas F (2002) Evolution of arginine deiminase (ADI) pathway genes. *Mol Phylogenet Evol* 25: 429–444.
50. Gamper M, Zimmermann A, Haas D (1991) Anaerobic regulation of transcription initiation in the *arcDABC* operon of *Pseudomonas aeruginosa*. *J Bacteriol* 173: 4742–4750.
51. Nakatsu C, Ng J, Singh R, Straus N, Wyndham C (1991) Chlorobenzoate catabolic transposon Tn5271 is a composite class I element with flanking class II insertion sequences. *Proc Nat Acad Sci U S A* 88: 8312–8316.
52. Li K, Wang S, Shi Y, Qu J, Zhai Y, et al. (2011) Genome sequence of *Paracoccus* sp. Strain TRP, a Chloropyrifos Biodegrader. *J Bacteriol* 193: 1786–1787.
53. Winterstein C, Ludwig B (1998) Genes coding for respiratory complexes map on all three chromosomes of the *Paracoccus denitrificans* genome. *Arch Microbiol* 169: 275–281.
54. L'Abée-Lund TM, Sørum H (2000) Functional Tn5393-like transposon in the R plasmid pRAS2 from the fish pathogen *Aeromonas salmonicida* subspecies *salmonicida* isolated in Norway. *Appl Environ Microbiol* 66: 5533–5535.
55. Sundin GW, Bender CL (1995) Expression of the *strA-strB* streptomycin resistance genes in *Pseudomonas syringae* and *Xanthomonas campestris* and characterization of IS6100 in *X. campestris*. *Appl Environ Microbiol* 61: 2891–2897.
56. Rogers MB, Bennett TK, Payne CM, Smith CJ (1994) Insertional activation of *cepA* leads to high-level beta-lactamase expression in *Bacteroides fragilis* clinical isolates. *J Bacteriol* 176: 4376–4384.
57. Ziebuhr W, Krimmer V, Rachid S, Lössner I, Götz F, et al. (1999) A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol Microbiol* 32: 345–356.
58. Kusumoto M, Ooka T, Nishiya Y, Ogura Y, Saito T, et al. (2011) Insertion sequence-excision enhancer removes transposable elements from bacterial genomes and induces various genomic deletions. *Nat Commun* 2: 152.
59. Roh SW, Nam YD, Chang HW, Kim KH, Kim MS, et al. (2009) *Paracoccus aestuarii* sp. nov., isolated from tidal flat sediment. *Int J Syst Evol Microbiol* 59: 790–794.
60. Urakami T, Tamaoka J, Suzuki K, Komagata K (1989) *Paracoccus alcaliphilus* sp. nov., an alkaliphilic and facultatively methylotrophic bacterium. *Int J Syst Bacteriol* 39: 116–121.
61. Lipski A, Reichert K, Reuter B, Spröer C, Altendorf K (1998) Identification of bacterial isolates from biofilters as *Paracoccus alkenifer* sp. nov. and *Paracoccus solventivorans* with emended description of *Paracoccus solventivorans*. *Int J Syst Bacteriol* 48: 529–536.
62. Urakami T, Araki H, Oyanagi H, Suzuki KI, Komagata K (1990) *Paracoccus aminophilus* sp. nov. and *Paracoccus aminovorans* sp. nov. which utilize *N,N*-dimethylformamide. *Int J Syst Bacteriol* 40: 287–291.
63. Ghosh W, Mandal S, Roy P (2006) *Paracoccus bengalensis* sp. nov., a novel sulfur-oxidizing chemolithoautotroph from the rhizospheric soil of an Indian tropical leguminous plant. *Syst Appl Microbiol* 29: 396–403.
64. Rainey FA, Kelly DP, Stackebrandt E, Burghardt J, Hiraishi A, Katayama Y, et al. (1999) A re-evaluation of the taxonomy of *Paracoccus denitrificans* and a proposal for the combination *Paracoccus pantotrophus* comb. nov. *Int J Syst Bacteriol* 49: 645–651.
65. Lee JH, Kim YS, Choi TJ, Lee WJ, Kim YT (2004) *Paracoccus haundaensis* sp. nov., a Gram-negative, halophilic, astaxanthin-producing bacterium. *Int J Syst Evol Microbiol* 54: 1699–1702.
66. Liu ZP, Wang BJ, Liu XY, Dai X, Liu YH, et al. (2008) *Paracoccus halophilus* sp. nov., isolated from marine sediment of the South China Sea, China, and emended description of genus *Paracoccus* Davis 1969. *Int J Syst Evol Microbiol* 58: 257–261.
67. Kim BY, Weon HY, Yoo SH, Kwon SW, Cho YH, et al. (2006) *Paracoccus homiensis* sp. nov., isolated from a sea-sand sample. *Int J Syst Evol Microbiol* 56: 2387–2390.
68. Doronina NV, Trotsenko YA (2000) A novel plant-associated thermotolerant alkaliphilic methylotroph of the genus *Paracoccus*. *Microbiology* 69: 593–598.
69. La HJ, Im WT, Ten LN, Kang MS, Shin DY, et al. (2005) *Paracoccus korensis* sp. nov., isolated from anaerobic granules in an upflow anaerobic sludge blanket (UASB) reactor. *Int J Syst Evol Microbiol* 55: 1657–1660.
70. Harker M, Hirschberg J, Oren A (1998) *Paracoccus marcusii* sp. nov., an orange gram-negative coccus. *Int J Syst Bacteriol* 2: 543–548.
71. Drewniak L, Styczek A, Majder-Lopatka M, Skłodowska A (2008) Bacteria, hypotolerant to arsenic in the rocks of an ancient gold mine, and their potential role in dissemination of arsenic pollution. *Environ Pollut* 156: 1069–1074.
72. Doronina NV, Trotsenko YA, Krausova VI, Suzina NE (1998) *Paracoccus methylutus* sp. nov. - a new aerobic facultatively methylotrophic bacterium utilizing dichloromethane. *Syst Appl Microbiol* 21: 230–236.
73. Jordan SL, McDonald IR, Kraczkiewicz-Dowjat AJ, Kelly DP, Rainey FA, et al. (1997) Autotrophic growth on carbon disulfide is a property of novel strains of *Paracoccus denitrificans*. *Arch Microbiol* 168: 225–236.
74. Gerstenberg C, Friedrich B, Schlegel HG (1982) Physical evidence for plasmids in autotrophic, especially hydrogen-oxidizing bacteria. *Arch Microbiol* 133: 90–96.
75. Pukall R, Laroche M, Kroppenstedt RM, Schumann P, Stackebrandt E, et al. (2003) *Paracoccus seriniphilus* sp. nov., an L-serine-dehydratase-producing coccus isolated from the marine bryozoan *Bugula plumosa*. *Int J Syst Evol Microbiol* 53: 443–447.
76. Siller H, Rainey FA, Stackebrandt E, Winter J (1996) Isolation and characterization of a new gram-negative, acetone-degrading, nitrate-reducing bacterium from soil, *Paracoccus solventivorans* sp. nov. *Int J Syst Bacteriol* 46: 1125–1130.
77. Liu XY, Wang BJ, Jiang CY, Liu SJ (2006) *Paracoccus sulfuroxidans* sp. nov., a sulfur oxidizer from activated sludge. *Int J Syst Evol Microbiol* 56: 2693–2695.
78. Katayama Y, Hiraishi A, Kuraish H (1995) *Paracoccus thiocyanatus* sp. nov., a new species of thiocyanate-utilizing facultative chemolithotroph, and transfer of *Thiobacillus versutus* to the genus *Paracoccus* as *Paracoccus versutus* comb. nov. with emendation of the genus. *Microbiology* 141: 1469–1477.
79. Bartosik D, Baj J, Sochacka M, Piechucka E, Włodarczyk M (2002) Molecular characterization of functional modules of plasmid pWKS1 of *Paracoccus pantotrophus* DSM 11072. *Microbiology* 148: 2847–2856.
80. Daneshvar MI, Hollis DG, Weyant RS, Steigerwalt AG, Whitney AM, et al. (2003) *Paracoccus yeeii* sp. nov. (formerly CDC group EO-2), a novel bacterial species associated with human infection. *J Clin Microbiol* 41: 1289–1294.
81. Berry A, Janssens D, Humbelin M, Jore JP, Hoste B, et al. (2003) *Paracoccus zeaxanthinifaciens* sp. nov., a zeaxanthin-producing bacterium. *Int J Syst Evol Microbiol* 53: 231–238.
82. Van der Ploeg J, Willemsen M, van Hall G, Janssen DB (1995) Adaptation of *Xanthobacter autotrophicus* GJ10 to bromoacetate due to activation and mobilization of the haloacetate dehalogenase gene by insertion element IS1247. *J Bacteriol* 177: 1348–1356.
83. Van Spanning RJ, Reijnders WN, Stouthamer AH (1995) Integration of heterologous DNA into the genome of *Paracoccus denitrificans* is mediated by a family of IS1248-related elements and a second type of integrative recombination event. *J Bacteriol* 177: 4772–4778.