Tn5041-like transposons: molecular diversity, evolutionary relationships and distribution of distinct variants in environmental bacteria

G. Kholodii,1 Zh. Gorlenko,1 S. Mindlin,1 J. Hobman2 and V. Nikiforov1

Author for correspondence: G. Kholodii. Tel: +7 095 196 0015. Fax: +7 095 196 0015.
e-mail: kholodii@img.ras.ru

1 Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia
2 School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

A detailed study on the geographic distribution, molecular diversity and evolutionary relationships of 24 closely related variants of the Tn5041 transposon found among 182 mercury resistant environmental Gram-negative strains from the IMG-Hg Reference Collection is reported here. RFLP analysis, followed by the determination of partial DNA sequences, identified 14 distinct types of these transposons, which differed from each other by 1–7 single-event DNA polymorphisms. No polymorphisms were detected at the right arm of the transposons except an insertion of a new mobile DNA element carrying a mer operon (named the mer2 cassette) within the Tn5041 mer operon. According to the model presented here, the insertion occurred via homologous recombination with a circular form of the mer2 cassette. A total of 8 point mutations, 1 internal deletion, 2 end-involving deletions, 3 mosaic regions and 2 insertions were detected at the left arm of the transposons. The insertions were a transposon closely related to Tn21 but lacking the integron and a new group II intron (named INT5041C). Inspection of the geographic distribution of the Tn5041 variants suggested that at least three long-distance waves of dissemination of these variants had occurred, accompanied by homologous recombination between different Tn5041 lineages. Movements of circular DNAs by homologous recombination as a source of mosaic genes and new mer genes, and formation of unusual mosaics ending or beginning at the Tn5041 att site are discussed.

Keywords: mer gene acquisition, group II intron, mosaic genes, homologous recombination, att site

INTRODUCTION

Mercury resistance (mer) transposons have been recognized for their extraordinary success in colonizing various bacteria all over the world. This makes them an ideal model for studying gene transfer in bacterial populations (Osborn et al., 1997; Hobman & Brown, 1997). The success of the Tn21 subgroup of mer transposons in medically important bacteria has been attributed to the acquisition of an integron containing an integrase system (Grinsted et al., 1990; Liebert et al., 1999; Partridge et al., 2001). This system has promoted the integration of various genes, especially those for antibiotic resistance, into these transposons (Stokes et al., 1997; Partridge et al., 2000), thus conferring on them a selective advantage. Our studies of bacteria isolated from non-medical sources have suggested that transposons of the Tn21 subgroup are not the sole transposon vehicles responsible for the broad dissemination of mercury resistance in the natural environment (Kholodii et al., 1993a, 1997, 2000a; Hobman et al., 1994; Mindlin et al., 2001; and our unpublished data). One of the “new” mer transposons, Tn5041, was identified in a Pseudomonas strain isolated from a mercury mine in Central Asia (Kholodii et al., 1997). It is 14907 bp in length and contains a typical Tn3 family transposase gene (tnpA) that is distantly related to tnpA genes found in the Tn21 subgroup transposons, and a mer operon at the right arm of the transposon that is

Abbreviations: AC, accession number; PMA, phenylmercuric acetate.
The accession numbers for the nucleotide sequences reported in this work are given in the legends for Figs 1 and 2.
A comparison of the sequence of INT5041C with other proteins is available as supplementary data on Microbiology Online (http://mic.sgmjournals.org).

0002-4883 © 2002 SGM
Fig. 1. Tn5041 variants. The genetic/restriction map of Tn5041A together with the Southern hybridization probes 1 to 5 (described in Methods) are at the top. E, EcoRI; P, PstI; EV, EcoRV; Sc, SacII; orfP, the ORF encoding a porin-like protein; R', an incomplete merR gene; κγ, the minitransposon with terminal inverted repeats closely related to those of Tn5044 and Tn5046 (Kholodii et al., 2000a; Mindlin et al., 2001; ACs Y17691 and Y18360); 1 and 2, the urf-1 (merE) and urf-2 genes. Other Tn5041 determinants are described in Introduction and Results. The thin lines represent unsequenced DNA; sequenced regions are shown by black bars (Tn5041 core sequence type), white boxes (regions of high divergence) and intermittent thick lines (documented by Yurieva et al., 1997; ACs Y09209 and Y09210). Shown on the right are the host strains and (in parentheses) the types of transposons. The point mutations are as follows: in position 811, G → A eliminating the Tn5041 SacII site; 952, G → A; 1240, deletion of G; 1241, C → T; 1265, C → T; 2875, A → G eliminating the Tn5041 EcoRV site; 3773, G → A eliminating the Tn5041 PstI site. EV, P' and Sc' are the EcoRV, PstI and SacII mutant sites. The insertions (symbolized by triangles) are located between nt 5439 and 5440 (all introns, INT), 6984 and 6985 (all mer2 cassettes), and 5900 and 5901 (Tn21 In2 transposons). The LS47-4 deletion (~ 100 bp) shown by // lies between nt 2394 and 3772. The orientation of INT5041 and the homologous segment of INT5041C are shown by arrows. The different sequenced regions are marked by lower-case letters. The examples, each representing identical
70% identical to Tn21 subgroup mer operons. The left arm of Tn5041 contains the orfI gene which encodes an integrase-like recombinase with ~30% similarity to its most closely related enzymes, XerC and XerD from Escherichia coli [accession nos (ACs) P22885 and M54884] and IntI1 from class 1 integrons (AC J02967). It also contains a predicted attachment site (att5041) showing striking similarity to the E. coli chromosomal dif site (Kholodii et al., 1997), the divergently transcribed gene orfQ of unknown function and a 4.5 kb region of DNA that encodes a number of apparently defective genes and mobile elements (Fig. 1, top). Recent studies have shown that the tnpA gene from Tn5041 and unidentified host-specific cell factor(s) control the transposition intermediate (cointegrate) formation, but no efficient resolution system in the transposon was detected (Kholodii et al., 2000b). Other studies conducted with a Tn5041 relative, Tn452 (AC AF151431), suggest the possibility that a novel gene (tnpC) also controls the transposition of Tn5041 (Hörak & Kivisaar, 1999).

Previously, we cloned and partially sequenced the mer operon-containing fragments related to the right arm of Tn5041 from DNA elements isolated from four strains from Eurasia. No sequence differences from the prototype Tn5041 (Tn5041A) were found in these elements except an insertion of a Tn21-related element in one strain and a mosaic at the beginning of the mer operon in another strain (Yurieva et al., 1997). The detection of such signature markers in otherwise non-polymorphic sequences suggested that Tn5041-like elements could provide a model system for the detailed epidemiological study of mercury resistance gene flow in environmental bacteria. To explore this possibility, we screened a larger collection of mercury resistant (HgR) environmental strains of various geographical origins for the presence of full-size Tn5041-like elements.

**METHODS**

**Bacterial strains and plasmids.** The environmental HgR strains screened in this study were from the IMG-Hg Reference Collection partly described previously (Khesin & Karasyova, 1984; Bogdanova et al., 1988; Yurieva et al., 1997; Mindlin et al., 2001). The strains characterized in this study are listed in Table 1; their identification was performed as described (Khesin & Karasyova, 1984; Bogdanova et al., 1988). E. coli K-12 strains JM83 and AB1456, and Pseudomonas aeruginosa PAO-R are described in our previous studies (Kholodii et al., 2000a, b). *Pseudomonas fluorescens* P22-1-2 is a Hg derivative of KHP22 (Khesin & Karasyova, 1984). pUC19 (Yanisch-Perron et al., 1985), pGEM-5Zf(+) and pGEM-7Zf(+) (Promega) were used as cloning/sequencing vectors and were propagated in E. coli JM83. RP1 (Pansegrau et al., 1994) was used in conjugation experiments, and pBR322 (Watson, 1988), pBR322::Tn5041 and pBR322::Tn5041D (Kholodii et al., 2000b) were used in the phenylmercuric acetate (PMA) resistance assay; pVS982 (pBR322::Tn21) (Grinstein et al., 1982) was used as the source of probe 8.

**DNA manipulations and sequence analysis.** Routine methods for DNA manipulation (Sambrook et al., 1989) and DNA analysis programs (Kholodii et al., 1997) were used. Sequencing was performed manually on both DNA strands of cloned fragments or purified PCR amplicons by the dyeoxy chain-termination method (Sanger et al., 1977). At least two independent DNA amplifications were sequenced in the region of interest. Full details of sequencing and structure of all primers used are available upon request.

**Origin of the probes used in RFLP analysis.** All probes are shown in Figs 1 and 2. Probe 1 is the 0.45 kb Sau3AI fragment from RP1::Tn5041, which includes the flanking sequence of RP1. Probes 2 and 3 are the 1.7 and 2.45 kb PstI fragments from Tn5041, respectively. Probe 3 is the 0.73 kb NcoI fragment from the merA gene of Tn5041. Probe 5 is the 2.2 kb PstI fragment from RP1::Tn5041, which includes the flanking sequence of RP1 (shown not to scale in Fig. 1). Probe 6 is the 1.02 kb AgeI-PstI fragment from INT5041. Probe 7 is the 0.93 kb BglII fragment from the mer2 cassette insertion of Tn5041D. Probe 8 is the 1.2 kb EcoRI fragment from the tnpA gene of Tn21.

**Localization of mer transposons.** In nine strains, the localization of the mer transposon to a plasmid was identified by its ability to be transferred by conjugation to *P. aeruginosa* PAO-R or *P. fluorescens* P22-1-2 using mercury resistance as a selective marker (Tra+ plasmids in Table 1). The transfer of Tn5041-related elements was confirmed by hybridization of total DNA from the transconjugants with probes 3 and 4, originating from the Tn5041 merA and tnpA genes, respectively (Fig. 1). The four strains that did not demonstrate an ability to transfer (Tra- plasmids in Table 1) were investigated further by agarose gel electrophoresis of crude lysates prepared according to Eckhardt (1978) and by hybridization of the plasmid DNA (transferred from the gel to Hybond-N filters) with probes 3 and 4. For those strains in which we did not detect HgR plasmids by either method, we hypothesize that these transposons are located on the chromosome (Table 1).

**PMA resistance.** Tn5041D-conferred resistance to organomercurials was determined by the method of bacterial culture titration on LB agar containing different concentrations of PMA. Inhibitory concentrations of PMA were determined according to complete growth inhibition. The PMA resistant strain, E. coli JM83(pBR322::Tn5041D), grew at 7.5 µg PMA ml⁻¹, whereas the susceptible strains, E. coli JM83(pBR322) and E. coli JM83(pBR322::Tn5041), grew only at 2.5 µg PMA ml⁻¹. Each experiment was repeated 3-4 times.

**RESULTS**

**Screening and RFLP typing of Tn5041-related elements**

A total of 182 Gram-negative HgR strains (together with KHP41 harbouring Tn5041) isolated from soil, fresh water and mercury ores from different parts of the world...
Table 1. Strains containing Tn5041-like elements

<table>
<thead>
<tr>
<th>Pseudomonas strain</th>
<th>Origin</th>
<th>RFLP type</th>
<th>Transposon</th>
<th>Name</th>
<th>Activity*</th>
<th>Location†</th>
<th>Terminal fragments (kb)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>KHP41 (non-fluorescent)</td>
<td>Kyrgyzstan, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>+</td>
<td>Chromosome</td>
<td>1.8 ± 1.50</td>
</tr>
<tr>
<td><em>P. fluorescens</em> NC16-2</td>
<td>Caucasus, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>2.4 ± 2.7</td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC23-2</td>
<td>Carpathians, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>1.9 ± 3.0</td>
</tr>
<tr>
<td><em>P. fluorescens</em> LS26-1</td>
<td>Moscow River, Russia</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>2.8 ± 3.8</td>
</tr>
<tr>
<td><em>P. fluorescens</em> FA7-8</td>
<td>Kamchatka Peninsula, soil</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>9.3 ± 3.3</td>
</tr>
<tr>
<td><em>P. fluorescens</em> NC2-3</td>
<td>Caucasus, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>2.9 ± 4.0</td>
</tr>
<tr>
<td>NC2-4 (non-fluorescent)</td>
<td>Caucasus, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>2.9 ± 4.0</td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC23-1</td>
<td>Carpathians, mercury mine</td>
<td>A</td>
<td>Tn5041</td>
<td>Tn5041A</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>1.8 ± 2.1</td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC29-5</td>
<td>Carpathians, mercury mine</td>
<td>B</td>
<td>Tn5041</td>
<td>Tn5041B</td>
<td>+</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC37-6</td>
<td>Carpathians, mercury mine</td>
<td>B</td>
<td>Tn5041</td>
<td>Tn5041B</td>
<td>+</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>TC40-2 (fluorescent)</td>
<td>Carpathians, mercury mine</td>
<td>B</td>
<td>Tn5041</td>
<td>Tn5041B</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>2.7 ± 4.4</td>
</tr>
<tr>
<td><em>P. fluorescens</em> LS53-4</td>
<td>European Russia, soil</td>
<td>B</td>
<td>Tn5041</td>
<td>Tn5041B</td>
<td>rs</td>
<td>7.5 ± 4.8</td>
<td></td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC30-1</td>
<td>Carpathians, mercury mine</td>
<td>H</td>
<td>Tn5041</td>
<td>Tn5041B+</td>
<td>−</td>
<td>Plasmid (Tra+)</td>
<td>2.0 ± 12.0$</td>
</tr>
<tr>
<td><em>P. putida</em> TC29-1</td>
<td>Carpathians, mercury mine</td>
<td>C</td>
<td>Tn5041</td>
<td>Tn5041C</td>
<td>+</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>TC36-1 (non-fluorescent)</td>
<td>Carpathians, mercury mine</td>
<td>C</td>
<td>Tn5041</td>
<td>Tn5041C</td>
<td>+</td>
<td>Chromosome</td>
<td>3.8 ± 3.5, 4.8 ± 1.5 + 3.5</td>
</tr>
<tr>
<td><em>P. putida</em> MU10-2</td>
<td>East Lansing, USA, soil</td>
<td>D</td>
<td>Tn5041</td>
<td>Tn5041D</td>
<td>+</td>
<td>Plasmid (Tra+)</td>
<td>2.7 ± 4.5</td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC97</td>
<td>Carpathians, mercury mine</td>
<td>D</td>
<td>Tn5041</td>
<td>Tn5041D</td>
<td>+</td>
<td>Chromosome</td>
<td>3.0 ± 4.5 $|$ 1.5 ± 4.5$|$</td>
</tr>
<tr>
<td>LS47-4 (fluorescent)</td>
<td>European Russia, water</td>
<td>E</td>
<td>Tn5041</td>
<td>Tn5041E</td>
<td>−</td>
<td>5.1 ± 1.2</td>
<td></td>
</tr>
<tr>
<td>TC39-4 (fluorescent)</td>
<td>Carpathians, mercury mine</td>
<td>F</td>
<td>Tn5041</td>
<td>Tn5041F</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>2.7 ± 2.3</td>
</tr>
<tr>
<td><em>P. fluorescens</em> TC24-3</td>
<td>Carpathians, mercury mine</td>
<td>F</td>
<td>Tn5041</td>
<td>Tn5041F</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>2.7 ± 2.3</td>
</tr>
<tr>
<td><em>P. fluorescens</em> KHP22</td>
<td>Kyrgyzstan, mercury mine</td>
<td>G</td>
<td>Tn5041</td>
<td>Tn5041G</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>2.5 ± 6.8</td>
</tr>
<tr>
<td><em>P. fluorescens</em> KHP25</td>
<td>Kyrgyzstan, mercury mine</td>
<td>G</td>
<td>Tn5041</td>
<td>Tn5041G</td>
<td>rs</td>
<td>Plasmid (Tra+)</td>
<td>2.5 ± 6.8</td>
</tr>
</tbody>
</table>

* +, Transposition frequency approximately $10^{-5}$ for KHP41, $10^{-6}$ for MU10-2 and $10^{-7}$ for other strains; −, transposition frequency was $<10^{-8}$; rs, refractory to introduction of RP1.
† Tra’, conjugative; Tra+, non-conjugative (reflecting the observed phenotypes, see Methods).
‡ Size of the PstI (left-terminal, LTF) and EcoRI (right-terminal, RTF) junction fragments showing Southern hybridization with probes 1 and 5, respectively. Besides transposon sequences, these probes contained adjacent target (RP1) sequences (Fig. 1 and Methods). Control Southern hybridization, applied to each of the strains in this table, showed no homology to the total DNA from the strains tested with the RP1 probe. Some searches for the terminal sequences (in TC29-5, TC37-6 and TC29-1) were omitted because the termini were sequenced (Fig. 1).
§ See Results section ‘Tn5041 variants containing a group II intron’.
|| These additional fragments imply the existence of an additional transposon. Our data indicate that the two transposons residing in TC36-1 are identical. First, RFLP analysis detected no additional internal fragments (Table 2, type C). Second, no heterogeneity was observed upon sequencing of the TC36-1 PCR amplicons from the left arm, which is indicative of the transposon type (Fig. 1).
$ The presence of an additional, truncated copy of Tn5041D1 (Tn5041Da; Fig. 1) can be concluded from these data, together with the data on the absence of heterogeneity upon sequencing of PCR amplicons and the data presented in Table 2 (type D). Apparently, its precursor has lost the orf1-containing terminus; more extended deletion [covering the sequences hybridizing with probe 2, 3 or 4 (Fig. 1)] implies the appearance of an additional fragment that was not observed (Table 2, type D).

(see Methods) were screened by colony hybridization to a Tn5041 tnpA gene probe (Fig. 1, probe 4). Twenty-four strains [all Pseudomonas] hybridized to the probe in this test. These strains were: 3 of 19 examined from Central Asia (Kyrgyzstan), 11 of 25 from the Carpathians (Ukraine), 4 of 15 from the North Caucasus, 1 of 13 from the Kamchatka Peninsula (Far East of Russia), 3 of 77 from Central Russia, 1 of 7 from New Zealand and 1 of 26 from the USA (East Lansing, Bethesda, Chicago, New York). All strains that hybridized to probe 4 were further screened for the presence of active transposons by introduction of the broad-host-range plasmid RP1 and mating out to E. coli AB1456 as previously described (Kholodii et al., 2000a). Eight strains were refractory to RP1 introduction. In the remaining 16 strains, transposition of mercury resistance

3572
into RP1 was observed in 5 strains (Table 1) including the Tn5041 host.

Total DNA preparations isolated from all strains that contained Tn5041-related elements were digested with PstI and subjected to Southern hybridization using probe 4. All but two strains contained identical PstI fragments that hybridized to this probe (Table 2). The two strains that displayed a different hybridization pattern were found to contain DNA elements only distantly related to Tn5041 (data not shown) and were excluded from further analysis. On the basis of RFLP patterns observed with the left- and right-arm internal Tn5041 probes, all strains were categorized into eight types, A–H (Table 2, probe 2, PstI digestion; probe 4, EcoRI digestion). The prototype transposon Tn5041 (renamed hereafter as Tn5041A) belonged to RFLP type A (Table 1). Four new mobilizable transposons belonged to RFLP types B, C and D (Table 1). For all strains, the fragments that hybridized with mer probe 3 were identical in size to the fragments that hybridized with tnpA probe 4 (Table 2, EcoRI digestion), indicating the physical linkage of the mer and tnp determinants. Additional hybridizations were designed to detect different insertions within these elements (Table 2, probes 6, 7 and 8). The terminal sequences fused with host target sequences (either plasmid or chromosomal, see Table 1 and Methods) were characterized with the help of probes 1 and 5 (Table 1, last two columns). In most cases the sizes of junction fragments were unique, suggesting insertion into unique targets. Only for some pairs of strains from the same locations (NC2-3 and NC2-4; TC24-3 and TC39-4; and KHP22 and KHP25) did our data indicate insertion into unique targets. Only for some pairs of strains from the same locations (NC2-3 and NC2-4; TC24-3 and TC39-4; and KHP22 and KHP25) did our data indicate the insertion into the same target (plasmids) (Table 1). In three strains (TC30-1, TC36-1 and TC97) additional junction fragments were observed, indicating the presence of additional full-sized or truncated Tn5041-like transposons (Table 1 and the footnotes).

**Detailed characterization of Tn5041-like elements**

To characterize the nature of the diversity of Tn5041-like elements recognized by RFLP typing, we analysed their DNA (cloned or amplified) by fine restriction enzyme site mapping and DNA sequencing of the segments marked in Fig. 1. As seen from Fig. 1, the mer transposons studied were very closely related to the prototype transposon Tn5041A. Consequently, we did not assign new numbers to these transposons but considered them to be variants of Tn5041 and marked them by capital letters referring to their types (Table 1 and Fig. 1). No sequence variations in the right arm were detected over ~1.8 kb of sequenced DNA. At the same time, 8 single-base mutations, 1 deletion (in Tn5041A sequence) in the 47 bp tnpA gene. ACs for the sequences determined are: AJ426646 and AJ426648 (TC36-1); AJ426647 (TC30-1); Y18999 (TC29-1); AJ318529 (MU10-2); AJ426033 and AJ426034 (TC97); AJ426035 and AJ426036 (LS47-4); AJ426037 (TC39-4); AJ426038 (TC24-3); AJ422224 (KHP22); AJ422225 and AJ422226 (KHP25). The TC39-4 and TC24-3 merA sequences [identical to AJ426035 (LS47-4)] were not deposited in any database.
Table 2. RFLP typing of Tn5041-like transposons

<table>
<thead>
<tr>
<th>RFLP type</th>
<th>Size of fragment showing Southern hybridization (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$PstI$</td>
</tr>
<tr>
<td></td>
<td>2 (LH)</td>
</tr>
<tr>
<td>A</td>
<td>1-7</td>
</tr>
<tr>
<td>B</td>
<td>9±3</td>
</tr>
<tr>
<td>C</td>
<td>4-5</td>
</tr>
<tr>
<td>D</td>
<td>1-7</td>
</tr>
<tr>
<td>E</td>
<td>1-6</td>
</tr>
<tr>
<td>F</td>
<td>7-7</td>
</tr>
<tr>
<td>G</td>
<td>7-6</td>
</tr>
<tr>
<td>H</td>
<td>9±3+10±4</td>
</tr>
</tbody>
</table>

*The DNA probes are shown in Figs 1 and 2, and described in Methods. A dash indicates the absence of hybridization with a probe; LH, left-hand arm; RH, right-hand arm; INT5041C and ‘mer2’, a group II intron and the mer2 cassette, respectively, identified in this study. Shown in bold are the data that were sufficiently conclusive for the RFLP typing of the transposon.
† The additional fragment(s) corresponding to the right terminal fragment(s) (Table 1) was seen in each hybridization.
‡ The additional fragment, 29 kb in length, corresponding to that expected due to the mer2 cassette insertion was apparent.
§ The additional fragment, 4-3 kb in length, corresponding to that expected due to the Tn214Im2 insertion was apparent.

an extended version of the $\rho$ sequence: it started at the same point as $\rho$ and was 100% identical to the sequence of $\rho$ in the overlapping region. The $\rho$1 substitution covered most of the orfQ gene, but did not disrupt its reading frame. One more variant region ($\psi$), with a divergence level of 16% relative to the same region in Tn5041A, was detected in Tn5041A3. It started within the unsequenced region (before nt 612 in Tn5041A) and ended at the left end of att5041. The observed diversity, including that resulting from the end-involving deletions (Table 1 and the footnotes) and the insertions described below, showed that there were 7 additional Tn5041 types which were named A1, A2, A3, C1, C1A, D1 and D1A (Table 1 and Fig. 1). From our data, most of the RFLP types observed (A–G; Table 2) were due to the elimination of the Tn5041 $PstI$ site (nt 3773, G→A) and/or the acquisition of additional $PstI$ or $EcoRI$ sites due to one of the insertions described below.

Tn5041 variants containing a group II intron

Restriction analysis detected a 2-2 kb insertion relative to Tn5041A at the left arm of the transposon Tn5041C mobilized by RPl from the Carpathian strain TC29-1. DNA sequencing demonstrated that the 2206 bp insertion, named INT5041C, interrupted the left terminal inverted repeat of a small Tn3/Tn5044 subgroup mobile element $\kappa$ (Figs 1 and 2) and showed features characteristic of group II self-splicing introns that have been found in both organelles and bacteria (Martínez-Abarca & Toro, 2000; Dai & Zimmerly, 2002). INT5041C carried an ORF, orf494 (Fig. 2), encoding a 494 aa polypeptide showing significant similarity to proteins encoded by cyanobacterial and chloroplast group II introns. The sequence alignment is available as supplementary data at http://mic.sgmjournals.org. This protein contained a clearly recognizable maturase (nucleic acid binding) domain X (Mohr et al., 1993) and reverse transcriptase subdomains 1 to 7 (Xiong & Eickbush, 1990). As is characteristic of most bacterial group II introns (Michel et al., 1993; Martinez-Abarca & Toro, 2000), the region downstream of orf494 was well conserved (Fig. 3a) and the predicted INT5041C RNA sequence of this region can be folded to produce a secondary structure typical of group II introns (Michel et al., 1989). This structure (Fig. 3b) contained a well conserved domain V with a bulging dinucleotide and a purine-rich terminal loop, and a domain VI with a bulging adenine, 7 bases away from the splice site which constitutes a 2¢ to 5¢ lariat branch point in spliced introns. The sequences at the 5¢ and 3¢ ends of INT5041C (5¢-GTGCG-3¢ and 5¢-GAC-3¢, respectively) conformed to the consensus sequences 5¢-GTGYG-3¢ and 5¢-RAY-3¢ deduced for group II introns (Michel et al., 1989). No target DNA duplication was found at the site of insertion of INT5041C and, as in the case of many other bacterial group II introns (Dai & Zimmerly, 2002), this element did not interrupt any gene with proven or suggested function.

Southern hybridization experiments with the INT5041C-specific probe 6 revealed the presence of the intron sequences in two other Carpathian strains: TC36-1 and TC30-1 belonging to RFLP types C and H, respectively (Table 2). RFLP analysis and DNA sequencing demonstrated that TC36-1 contained two copies of a Tn5041 variant, Tn5041C1, with a complete INT5041C inserted into the same target site as in
Tn5041-like transposons

Re-examination of the Tn5041A sequence showed that it contained a 195 bp DNA fragment (nt 4224–4418) with 88% identity to the 3' end of INT5041C (Fig. 3a). The undamaged end of the element (named INTA5041) was oriented towards orfP (Fig. 1), whereas the damaged one was found to be fused to a damaged end of σλ, a relic of an IS2-related element (Kholodii et al., 1997). Restriction analysis of the mobilized Tn5041-like transposons demonstrated that the NeoI site within INTA5041 was conserved, thus indicating the presence of this element in all these transposons.

Tn5041 variants containing the mer2 cassette

The 7120 bp insertion (named the mer2 cassette) was completely sequenced in Tn5041D (Fig. 2). The insertion was bounded by rearranged regions of a mer operon (named mer2; white bars in Fig. 4b) which was distinct from the mer operon (named mer1; hatched bars in Fig. 4b) common to all full-sized Tn5041-like transposons. It occurred within the merT gene of a mer1 operon and resulted in the formation of two functional tandem operons, named mer1x2 and mer2x1, separated by a 2.35 kb spacer containing a truncated copy of a novel insertion sequence, named IS1015A (Fig. 4b, bottom). This IS element encoded a protein 57% and 50%, respectively, identical to transposases from the IS490 (AC U80795) and Tn4502 (AC U60645) elements belonging to the IS256 family (Mahillon & Chandler, 1998). Both the order of mer genes from mer1 and mer2 found in the tandem (Fig. 4b, bottom) and the structure of the junction fragments (where the corresponding merT sequences from the target and insertion transferred to each other (Fig. 4a)) were strongly indicative of the integration of the circularized mer2-containing intermediate within mer1 via homologous recombination (Fig. 4b). As in other cases of homologous recombination involving non-identical partners (Doherty et al., 1983; Rossignol et al., 1984), DNA sequence correction (gene conversion) was likely to accompany this recombination (see Fig. 4a legend).

The mer2 operon contained a standard set of mer genes (merRTPA) followed by the merG and merB genes responsible for broad-spectrum (organomercurial) resistance. The Tn5041D merB gene was functional when expressed in E. coli K-12 conferring, together with the other mer genes, PMA resistance (see Methods). It encoded an organomercurial lyase that is 75% and 83% identical to MerB proteins from pMR26 (Kiyono et al., 1997; AC D83080) and Tn5058 (AC Y17897), respect-
Fig. 4. (a) The Tn5041D sequences near the ends of the mer2 cassette insertion. The boundary sequences between the mer1 and mer2 operons are aligned with respect to each other and the prototype sequences to show the transfer of the (Tn5041) target sequence into the (pDU1358) incoming one, and vice versa, respectively. Dots, identical bases; underlined, Chi single-base-substitution variants. If the homologous recombination were reciprocal, i.e., occurred strictly within the identical sequences (5'-CTCGCCTCGGCTTGCTG-3'), this sequence at the right junction would have been followed by the Tn5041 sequence, but not by the pDU1358 one. The observed lengthening of the pDU1358 sequence on the right, up to the boxed 47 bp, was likely to result from a migration of the chiasma to the right to form the 47 bp heteroduplexes and the conversion of at least one of them to a homoduplex by a mismatch-repair mechanism. (b) Proposed mechanism of the acquisition of the 7120 base mer2 cassette. Hatched bars, Tn5041 core DNA; white bars, mer2 operon DNA; other designations are the same as in Figs 1 and 2. The two homologous recombination-mediated events – excision of the mer2 cassette from the donor (via IS1015<IS1015∆), supplying the circular intermediate, and its integration into the ancestor – could have led to Tn5041D. For IL, D1, D2 and D3 which show the probable rearrangements of the tandem formed, and the Greek letters which mark the segments for different (homologous) mer sequence types, see Results. Shown in the frame are the Chi and Chi-like octamers identified [from the known active octamers (Cheng & Smith, 1984)], their symbols and relative activities (Cheng & Smith, 1994); recombination can be stimulated leftward (top flags) or rightward (bottom flags) for up to 8 kb (Smith, 1994).

The mer2 operon proved to be a complex mosaic. Here, the DNA sequences belonging to the five sequence types,
denoted by Greek letters in Fig. 4(b) were identified, which subdivided the Tn5041D mer2 operon into eight segments. The 474 bp sequence at the beginning of this operon belonged to the pKHL2 sequence type (λ), with 1 and 2 mismatches found with respect to the end sequences of Tn5036 (Yurieva et al., 1997; AC Y09025) and Tn5041D isolated from pKHL2 (Khododii et al., 1993b; AC AF213017). In this sequence, a Tn21 subgroup terminal repeat (black arrowhead in Figs 2 and 4b) and the merR gene sequence were identified. Because no tnp and tnp-end sequences were detected in the Tn5041D insertion, it was concluded that the mer2 operon forms part of a one-ended transposon (named the mer2 transposon). The pKHL2-type sequence (177 bp) was encountered once more in the mer2 operon. Here, it covered the merT gene 3’ end and was 100% identical to the equivalent sequence from pKHL2. The DNA fragments of two further sequence types, π and δ, came between the λ segments in the presumed donor of the mer2 operon (Fig. 4b, top). The π sequence (78 bp), first detected in the Tn5041D1-containing strain TC97 (Yurieva et al., 1997; AC Y09220), was most closely related (91% identity) to the equivalent sequences from pKHL2 and pDU1358; and the δ sequence (194 bp) belonged to the inferred crossover fragment that was split into two subfragments during the integration of the cassette and was indistinguishable from the equivalent sequence of pDU1358 (Griffin et al., 1987; AC M29400). A 133 bp DNA fragment of a previously unknown sequence type (σ), differing from the equivalent fragments of pKHL2 and Tn5041 by 7% and 22-6% base substitutions, respectively, covered the merP gene 5’ end. After this fragment, the continuation of the pDU1358-type sequence with 2 mismatches over 1045 bp was found, which was followed by the DNA sequence (1986 bp) of the extended version of the β-type sequence first detected in pKHL272/Tn5036 (Yurieva et al., 1997; ACs Y08992 and Y09025). In the merA region, the sequence of β was ~90% identical to that of pMR26 (Kiyono et al., 1997; AC D83080) and Tn5058 (AC Y17897), and in the merG–B–D region, it was 80-7% identical to that of pMR26. Hybridization to the Tn5041D mer2 cassette-specific probe 7 was found in 4 strains isolated from remote geographical areas of Eurasia and belonging to RFLP types D1, E and F (Tables 1 and 2). Using appropriate primers, we demonstrated that all these strains contained transposons with the same insertion at the same site as in Tn5041D isolated from a USA strain (Figs 1 and 2). DNA database searches showed that two Hg²⁺-resistant Spingomonas paucimobilis strains isolated from monkey’s faeces (USA; Liebert et al., 2000) contained the mer1x2 and mer2x1 mosaic sequences, of 842 and 1255 bp (ACs AF120972, AF120973, AF120976 and AF120977), identical to the Tn5041D sequences at the boundaries of the mer2 cassette and target (Fig. 4a). One could imagine repeated events of cassette integration at the same point, but in this case it is impossible to explain why the accompanying gene conversion (Fig. 4b and the legend) was the same.

**Tn5041 variants containing a close relative of Tn21 lacking the In2 integron**

Two different strains of *P. fluorescens* (KHP22 and KHP25) from the same location, which were identified as RFLP type G, hybridized with the Tn21-specific probe 8 (Table 2) and, according to all our tests, contained indistinguishable transposons. Using primers designed for the Tn5041 and Tn21 sequences, the left and right junction fragments from KHP22 and KHP25, of ~450 and 610 bp, were amplified and sequenced. These fragments contained (with one mismatch) the merR- and tnpA-proximal terminal sequences of Tn21, respectively (Fig. 2). The adjacent Tn5041 core sequences formed parts of a single sequence which was interrupted by the insertion of a Tn21-like element (named Tn21Aln2); and the target duplication of 5 bp characteristic of the Tn21 subgroup of Tn3 transposons (Grinstead et al., 1990) was found. It should be noted that the right-arm junction fragment from KHP22 was cloned and partly sequenced previously (Yurieva et al., 1997; Figs 1 and 2, the intermittent thick lines). Addition of the sizes of the ‘left-hand-arm’ and ‘Tn21’ fragments from the G type transposons (Table 2, probes 2 and 8, after PsI) showed that the insertion in these elements was 8-7 kb in length, i.e., it corresponded to the size of Tn21 lacking the In2 integron. The absence of the In2 integron sequences in strains KHP22 and KHP25 was confirmed by Southern hybridization (data not shown). Although only part of the Tn21Aln2 mer operon was sequenced, the presence of an additional mer fragment expected due to the Tn21Aln2 insertion (Table 2, the footnote to probe 3) indicated the presence of the other mer genes. Like other Tn21-subgroup transposons (Yurieva et al., 1997), Tn21Aln2 seems to be recombinant due to a cross-over in the res site, since it differs from Tn21 in its left arm (by the presence of the EcoRI and PsI sites and a base substitution) but not in its right arm (Fig. 2).

**DISCUSSION**

Even a single-event DNA change may give rise to a successful Tn5041 clone

To facilitate the understanding of the mechanisms of response by bacterial communities to the impact of industrial pollution, and long-term trends of dissemination of mobile genetic elements, we have carried out a detailed study of the molecular diversity and evolutionary relationships of mer transposons recovered from environmental bacteria of widespread geographical origin. In this work, which is a larger-scale continuation of previous studies on Tn5041-like elements (Yurieva et al., 1997), we have extended the studies to the level of full-size transposons. Screening of an enlarged collection of environmental strains for the presence of these transposons has allowed eight RFLP types to be recognized (Table 2), with the A type being referred to as Tn5041, and the H type (strain TC30-1) being due to the presence of two different Tn5041-like
Fig. 5. (a) Hypothetical evolutionary relationships for the various Tn5041 types derived from the archetypal transposon type A1. The different Tn5041 types are depicted by circles. The widely distributed types are shown by bold circles. For the point mutations (numbers), insertions and segments \( \phi \), \( \rho \), and \( \rho_1 \) ‘creating’ the mosaics, see Fig. 1. A1* and A1** are the unidentified derivatives of Tn5041A1; B* is the unidentified derivative of Tn5041B; hatched boxes, the unidentified genetic elements; X, Xₐ, X₂, X⁵, and X⁶, the homologous recombination crossovers; \( \Delta LH \) and \( \Delta RH \), the left- and right-end-involving deletions. (b), (c), (d) and (e) show the recombining transposons (drawn schematically) from X₂, Xₐ, X⁵ and X⁶ crosses, respectively; resulting products are shown by the dotted lines.
advantages to a new variant, as in the case of the merB gene.

Recombinations at the left arm of Tn5041-like elements

DNA polymorphisms found in Tn5041-like elements are reminiscent of DNA polymorphisms found in certain transposons from clinically important bacteria such as the Tn1546-like elements from vancomycin-resistant enterococci (Willems et al., 1999). Both groups of elements represent very young lineages composed of closely related types of transposons, displaying a very small number of single-base mutations, punctuated by characteristic insertions of large DNA fragments and, in the case of Tn5041-like elements, by DNA fragments creating mosaics. The only significant difference lies in the fact that no contribution of homologous recombination is seen in the phylogeny of Tn1546-like elements proposed by Willems et al. (1999), whereas this pathway must have played a noticeable role in the phylogeny of Tn5041-like elements, especially as a mechanism that reassorted the variations (Fig. 5).

The fact that one of the ends of the divergent segments creating the mosaic genes (the left end of ϕ, and the right end of ρ and ρ1) was found to be present at the corresponding end of the predicted Tn5041 att site (att5041) suggested that the site-specific recombination system was involved in this process. The proposed dif-like structure for att5041 (Kholodii et al., 1997; Fig. 6a, b) delineates the probable overlap region, where the strand cleavage and exchange by a XerC/XerD-like recombinase is expected to occur, in the centre of att5041 and predicts (Blakely et al., 2000) an asymmetric location of the inserted DNA, from the sequence asymmetry in this region. If this structure is accepted, it is impossible to explain the location of the substitutions, ϕ and ρ/ρ1, on either side of the att site; and it is difficult to explain why the diversity associated with ϕ and ρ/ρ1 does not start just after the overlap region. Re-examination of the DNA sequence of att5041 showed a structure that avoids these discrepancies (Fig. 6c). In each half of the strong palindrome forming the potential att site, a structure was identified that was similar to the structure (known as the simple site) of the att site from class 1 integrons, attI1 (Partridge et al., 2000). Moreover, the sequences of the boundaries in the recombinants containing the ϕ and ρ/ρ1 segments were attI1 core-site-like, i.e., similar to the boundaries of inserted integron cassettes (Stokes et al., 1997), so that the putative crossover position on either side of att5041 coincided with that inferred for ϕ and ρ/ρ1 (Fig. 6c). These data suggested that the segments under consideration could be captured by a site-specific mechanism similar to the class 1 integron IntI1 integrase-dependent mechanism that acquires mobile gene cassettes (Stokes et al., 1997). Inspection of the att-distal ends of ρ and ρ1 (the corresponding end of ϕ was not sequenced) showed no palindrome related to the consensus sequence inferred for the att site (59-base element) of the mobile gene cassette (Stokes et al., 1997). From these data, a most probable mechanism underlying the formation of ϕ and ρ/ρ1 might consist of the fusion between two partly homologous (orfI-att5041-orfQ-containing) replicons at the att5041 sites, mediated by an IntI1-like recombinase, followed by resolution of the co-integrate via homologous recombination in the region adjacent to att5041.

Fig. 6. (a) The E. coli dif site (Blakely et al., 2000). (b), (c) Versions of att5041. Asterisks mark the bp that are common for dif and att5041. Shown in bold are the imperfect inverted repeats shown or expected to be involved in recognition of the att site with the recombinase. 1 and 2 are the inverted repeats related to the 7 bp core site consensus (5'-GTTRRRY-3') and the 7 bp inverse core site consensus (5'-RRYYAAC-3'), respectively, inferred for the attI1 site of class 1 integrons and the 59-base element (Stokes et al., 1997; Partridge et al., 2000). The vertical arrows mark the observed or inferred cross-over points. The overlap region and its analogues are boxed with continuous and dotted lines, respectively. The termini of ϕ and ρ/ρ1 are shown in square brackets beneath the att5041 sequences that these termini substitute in the recombinants (AC042216, Y19000 and Y19001). (d) The proposed mechanism of formation of the ϕ, ρ and ρ1 substitutions (exemplified by the ρ1 formation). Large circles indicate the replicons; small circles, the att sites; thick segments, the homologous DNAs. 1 and 1′, the alleles of orfI; Q and Q′, the alleles of orfQ; Rec, the homologous recombination.
(Fig. 6d). Note that the formation of ρ could follow either this pathway or via formation of ρ1 according to the mentioned mechanism (Fig. 6d) and subsequent shortening of ρ1 (ρ1 → ρ in Fig. 5a) via the allelic exchange (Fig. 5b).

The left-arm recombination events suggest that the partners must have encountered each other in the same cell. Three strains (TC30-1, TC36-1, TC97; Table 1) carried an additional closely related transposon, indicating that this important event does occur relatively frequently in the wild.

Movement of circular DNAs by homologous recombination as a source of mosaic and new genes

The view of the mer operon from Gram-negative bacteria which has emerged (Liebert et al., 1997, 2000) is of a dynamic mosaic locus with relatively constant (essential) genes such as merR, merT, merP and merA interspersed with incoming (accessory) genes of known (merB, merC, merF), partly defined [merD, merG, merE (urf-1)] or unknown functions (urf-2). For the functions of the genes, see Mukhopadhyay et al. (1991), Hobman & Brown (1997), Wilson et al. (2000) and Liebert et al. (2000). Although the non-random location of the accessory genes seems to have been identified, with the hotspots on the immediate 5' and 3' proximal regions of merA (Liebert et al., 1997, 2000), little data concerning the mechanistic basis of the gains and/or losses of the mer genes (termed as unusual ‘plug and play’ recombination events) has been presented so far. There is a fundamental difference between how an operon originally gained an additional gene in the hotspot regions flanking merA and how these arrangements subsequently become distributed into other operons. The primary acquisition implies the involvement of site-specific recombination, whereas the distribution (re-assortment, or secondary acquisition of genes) implies participation of homologous recombination. So far, no consideration has been given in the literature to the cases of acquisition of mer genes via specific sites, e.g., the secondary sites described for the integron DNA integrase (Francia et al., 1993; Recchia et al., 1994) or res-like sites permitting a relatively stable integration of circular DNA (Kholodii, 2001). At the same time, there is indirect evidence for the secondary acquisition of mer genes due to homologous recombination between more conserved adjacent genes (Liebert et al., 2000), like the process involved in module exchange in lambdoid phages (Campbell, 1994). Clear evidence for the operation of such a mechanism in the mer system has come from the sequence of a variant of the pKLH2 mer operon (AC AJ251272) which has pKLH2 sequences at the flanks and a long divergent segment with a merB gene (absent in pKLH2) in the centre (G. Kholodii, unpublished data). In Tn5041D and its microderivatives, not only did we find one more example of the insertion of a mer determinant at a hotspot (5' of merA), but also obtained data strongly indicative of another secondary mechanism to acquire new mer gene(s). The specific features of the junction sequences (Fig. 4a) clearly provide one more example of fusion of circular DNAs by homologous recombination but, for the first time, operating within mer loci. The integration of mer-containing circular DNA, like the one proposed for Tn5041D (Fig. 4b), probably occurred in other cases too, where the head-to-tail homologous HgR loci were found in the same replicon, e.g., in pDU1358 (Griffin et al., 1987), pMR26 (ACs D83080 and AB013925), Tn5056 (Mindlin et al., 2001) and Tn5058 (AC Y17897).

The tandem of two mer operons, similar to the one found in Tn5041D, should be prone to further rearrangements resulting from deletion/excision of one copy due to illegitimate or homologous recombination. With illegitimate recombination, deletion of one of the tandem operons may lead to the emergence of a locus with a new set of mer genes. For instance, the deletion of mer2x1 (IL in Fig. 4b, bottom) could have led to the substitution of a narrow-spectrum HgR locus for a broad-spectrum one. (What is important in this case is that one of the junctions indicative of the participation of homologous recombination in the acquisition of mer genes would have been lost as well.) With homologous recombination, excision of one of the tandem operons would result in a series of both circular mosaic DNAs (D1, D2 and D3 in Fig. 4b, bottom) and single mosaic operons with overlapping mosaic segments. Such a process may provide an explanation of the fact that the mer2 cassette itself is a genetic mosaic. The same process may account for the origin of different-length mosaic segments of sequence type α that were revealed in the pKLH2-type mer operons (Yurieva et al., 1997).

It seems that it is not by chance that almost all the regions where we have observed or proposed homologous recombination were regions where hotspots for such a pathway (RecBCD), known as Chi or Chi-like sites, were found. They included the sites shown in Fig. 4(b) and a Chi-like site, 5'-ACTGGTG-3', occupying nt 1394–1401 of Tn5041. The data concerning the Tn5041D tandem of the two different mer operons (Fig. 4b) are especially demonstrative: the mer region contains five occurrences of the hotspots in 5 kb versus 14 occurrences in 5 kb inferred from 20 sequenced mer operons (Liebert et al., 2000).

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basic Research Grants 00-15-97751 and 02-04-48327, and National Institute of Health FIRCA grant TW00396-02.

REFERENCES


Received 22 April 2002; revised 9 July 2002; accepted 16 July 2002.