J Mol Evol (2000) 51:607–622
DOI: 10.1007/s002390010124

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2000

The Quality of *merC*, a Module of the *mer* Mosaic

Cynthia A. Liebert, Alice L. Watson, Anne O. Summers

Department of Microbiology, The University of Georgia, Athens, GA 30602-2605, USA

Received: 2 June 2000 / Accepted: 23 October 2000

Abstract. We examined a region of high variability in the mosaic mercury resistance (*mer*) operon of natural bacterial isolates from the primate intestinal microbiota. The region between the *merP* and *merA* genes of nine *mer* loci was sequenced and either the *merC,* the *merF,* or no gene was present. Two novel *merC* genes were identified. Overall nucleotide diversity, π (per 100 sites), of the *merC* gene was greater (49.63) than adjacent *merP* (35.82) and *merA* (32.58) genes. However, the consequences of this variability for the predicted structure of the MerC protein are limited and putative functional elements (metal-binding ligands and transmembrane domains) are strongly conserved. Comparison of codon usage of the *merTP, merC,* and *merA* genes suggests that several *merC* genes are not coeval with their flanking sequences. Although evidence of homologous recombination within the very variable *merC* genes is not apparent, the flanking regions have higher homologies than *merC,* and recombination appears to be driving their overall sequence identities higher. The synonymous codon usage bias (EN_C) values suggest greater variability in expression of the *merC* gene than in flanking genes in six different bacterial hosts. We propose a model for the evolution of MerC as a host-dependent, adventitious module of the *mer* operon.

Key words: Operon organization — Mercury resistance — Horizontal gene transfer — Genome rearrangement — Chi-homologous recombination — Codon usage

Correspondence to: A.O. Summers; Department of Microbiology, 527 Biological Sciences Bldg., The University of Georgia, Athens GA 30602-2605, USA; *e-mail:* summers@arches.uga.edu

Introduction

Bacteria have evolved resistance to both inorganic and organic mercury compounds, and these resistance systems occur world-wide, in many Gram-negative and Gram-positive bacteria from environmental and clinical sources. Early studies on the *mer* operons of Tn*21* (NR1) and Tn*501* (Brown et al. 1986; Gilbert and Summers 1988) led to the impression that the organization of *mer* operons in Gram-negative bacteria was generally conserved. However, now with over 20 sequenced *mer* operons, polymorphism of *mer* operons is well established (Bruce et al. 1995; Liebert et al. 1997).

The *mer* operon has a modular or mosaic structure (Fig. 1). In our previous study (Liebert et al. 1997), the *mer* operons of the Gram-negative fecal isolates from the Primate Amalgam collection (isolated from monkeys) and the Environmental Plasmid Survey (EPS) collection (isolated from humans) were grouped into nine phylogenetic *mer* locus types (Fig. 1) using mapped restriction site polymorphisms (MRSP), restriction fragment length polymorphisms (RFLP) of PCR amplicons, and Southern hybridizations of *mer* genes. Initial grouping of the *mer* loci was based on unique *merA* MRSPs and was further supported by additional tests, including intergenic PCR amplifications to establish gene order. These nine *mer* loci demonstrated considerable polymorphism among themselves and with respect to previously characterized *mer* loci, supporting the view that the *mer* operon is a genetic mosaic and has a predominance of insertions/deletions of functional genes immediately before and after the *merA* gene. In particular, we detected a high degree of variability, both in length and hybridization, in the region 5′ of *merA,* suggesting considerable sequence

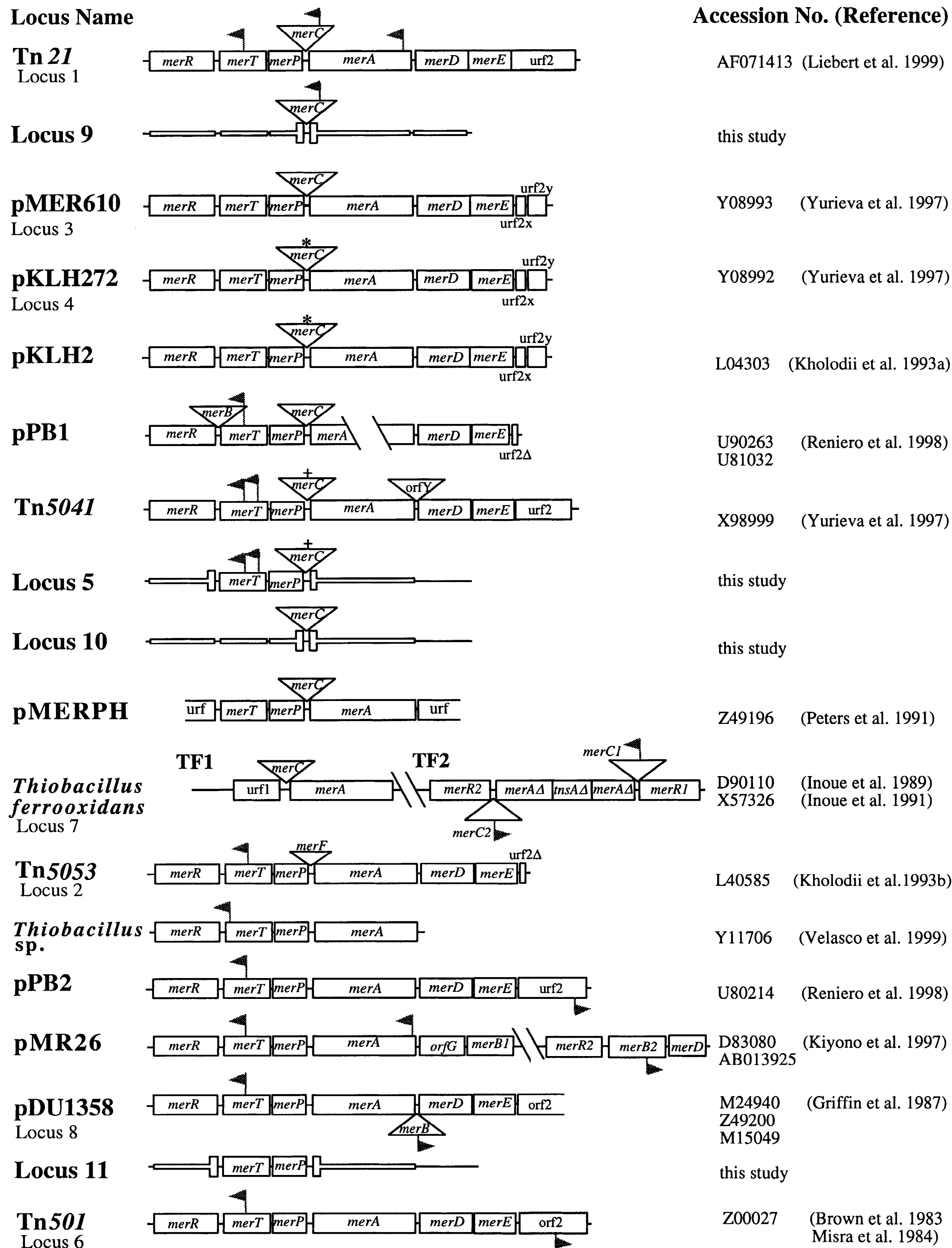


Fig. 1. Unique *mer* operons from Gram-negative bacteria. Arrangement of operons is according to similar gene order and/or inferred phylogenetic relatedness of the *merC* genes. The Tn*21 mer* operon is the first described operon with a *merC* gene. Tn*5053* carries the *merF* gene, and Tn*501* is the first described *mer* operon with no intervening gene between *merP* and *merA*. Those *mer* operons with locus designations were characterized in our previous study (Liebert et al. 1997). Thick boxes indicate regions sequenced and thin boxes indicate Southern hybridizations and/or RFLPs (Liebert et al. 1997) were performed to identify the region. The *merC* genes of pKLH272 and pKLH2 (*) are identical and the *merC* genes of Tn*5041* and Locus 5 (+) are identical, but other regions of these four loci are not identical to each other. Gray flags mark the location and orientation of Chi sites; recombination can be stimulated leftward (top) or rightward (bottom) for up to 8 kb (Smith 1994). Only the ends of the pPB1 *merA* have been sequenced. *mer* operon pMER327 with *merF* and partially sequenced *mer* operon TC97 with *merC* are not shown here.

divergence in this region which we investigate further here.

The essential components of the *mer* operon include *merR,* encoding the regulatory protein (Summers 1992; Zeng Q. et al. 1998), *merT* and *mer P,* together encoding the Hg(II) uptake system (Morby et al. 1995; Steel and Opella 1997), and *merA,* encoding the mercuric ion reductase (Miller et al. 1991; Schiering et al. 1991). In addition to these genes, *mer* loci may also have the genes *merC* or *merF,* encoding inner membrane proteins which function in Hg(II) uptake (Sahlman et al. 1999; Wilson et al. 2000). The *merC* gene is apparently not required for Hg(II) resistance (HgR), if MerP and MerT are present, since there are *mer* operons without *merC* [e.g., those found in transposons Tn*501* (Misra et al. 1984) and Tn*5053* (Kholodii et al. 1993b)]. Moreover, deleting *merC* from the Tn*21 mer* locus had no effect on HgR (Hamlett et al. 1992). The *Thiobacillus ferrooxidans mer* loci have at least three distinct *merC* genes (Inoue et al. 1991), one of which can function alone as a mercury transport system in *E. coli* (Kusano et al. 1990). Sahlman and coworkers have demonstrated that MerC of Tn*21* binds Hg^{2+} (Sahlman et al. 1997) and that the two N-terminal cysteines predicted to lie within the cytoplasmic membrane are involved in Hg(II) uptake (Sahlman et al. 1999). Unlike MerT, uptake of Hg^{2+} by MerC is not enhanced by MerP (Wilson et al. 2000) and the detailed interactions of MerP, MerT, MerC, MerF, and MerA remain to be defined. At the 3′ end of the operon, *merD,* encoding a protein which antagonizes the activation function of *merR* (Mukhopadhyay et al. 1991) and *merE,* an undefined open reading frame (ORF) (Liebert et al. 1999) are generally found. Finally, the *merB* gene, encoding the organomercurial lyase has been found in some *mer* operons located 3′ either of *merA* or of *merR* (Griffin et al. 1987; Liebert et al. 1997; Reniero et al. 1998).

In bacteria, horizontal gene transfer can result in the formation of mosaic loci (Campbell 1994; Maynard Smith 1992). Boyd and coworkers (Boyd et al. 1996) observed that recombination between genes on plasmids takes place at a considerably higher frequency than that observed for chromosomal genes and implicated Chi sites and RecBCD in the mechanism of plasmid borne gene exchange. They suggested that plasmid genes and, by inference, the plasmids themselves are mosaic structures with different regions acquired from different sources. Many *mer* operons are found on large conjugative plasmids and often within transposons (de la Cruz and Grinsted 1982; Yoshifumi et al. 1984). A common recombination event which generates mosaics in transposons is the exchange of transposon modules by recombination at the *res* site (Grinsted et al. 1990; Minakhina et al. 1999). This process is a recurring theme seen in the evolution of transposons carrying HgR loci. Widely disseminated transposons Tn*21* (Liebert et al. 1999), Tn*5041* (Yurieva et al. 1997), and Tn*5053* (Kholodii et al. 1993b) are three examples of such mosaic HgR transposons, all of which also carry genetic elements that can effect site-specific or nonspecific recombination events.

In *mer* operons, mosaicism occurs at two levels: in gene location within the operon and within the genes themselves. The most conserved gene order for the *mer* operon is *merRTPADE.* Of the 20 unique *mer* operons sequenced, 12 carry a *merC* variant(s) (Fig. 1). Two nearly identical *mer* operons carried by transposon Tn*5053* and plasmid pMER327 (Hobman et al. 1994) (not shown) carry *merF* in the *merP–A* region. The *merC* and *merF* genes have only been found 5′ to the *merA* gene. The *mer* operons in Tn*501,* pDU1358, pMR26, pPB2, and *Thiobacillus* sp. do not have any gene present between the *merP* and *merA* genes. Variability also occurs in the region 3′ of *merA,* where *merB* (Griffin et al. 1987; Liebert et al. 1997) or other ORFs (Kiyono et al. 1997; Yurieva et al. 1997) can occur between *merA* and *merD.* In the present study, we focused on the variable region between *merP* and *merA* (P–A), to ask three questions: (1) How diverse are the genes occurring there? (2) What is the consequence of the divergence for the possible structures of the proteins? (3) What might be the agents of the diversity seen in the P–A region?

Materials and Methods

Bacterial Strains. Strains carrying *mer* loci were from the Primate Amalgam collection ($n = 7$) and the EPS collection ($n = 9$) (Table 1). The *mer* loci of these strains were previously described (Liebert et al. 1997) as were the collections (Wireman et al. 1997). The standard mercury-resistant strains used in this study were SK1592(pDU202), SE23, JM83(pKLH2.5), and GD86 and were provided by T. Foster, D. Ritchie, V. Nikiforov, and G. Dougan, respectively. Strains were stored at −70°C.

DNA Isolation. Total cellular DNA was isolated using the IsoQuick Nucleic Acid Extraction Kit (ORCA Research, Inc., Bothell, WA, USA) from 1.5 ml cultures grown in Luria-Bertani (LB) broth at 37°C for 6 h. Two independent DNA samples were prepared for each of two strains of the same *mer* locus type (Table 1).

PCR Primers. Oligonucleotide primers (sequences available on request) were designed using the OLIGO™ Software (National Biosciences, Inc.) and synthesized by the University of Georgia Molecular Genetics Instrumentation Facility (Athens, GA, USA) or BioServe Biotechnologies (Laurel, MD, USA). Priming sites and the regions of the *mer* loci amplified are depicted in Fig. 2. Primary primers T3 and A5 amplified the template DNA used for sequencing (Fig. 2; open rectangle). To distinguish the two *mer* loci (Loci 5 and 11) carried by the same strain (see Results), the primer pairs R1-A5 and R1-A0 were used to generate physically distinct template amplicons for sequencing. Secondary sequencing primers (R1-A0) for these R1-A5 and R1-A0 templates are shown in Fig. 2.

PCR Amplification. Amplification reactions were carried out in a total of 50 μl containing 20 pmol of each primer, 0.2 mM each dNTP, 2.5 U cloned *Pfu* DNA Polymerase (Stratagene, LaJolla, CA, USA) and 1X PCR buffer (20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM

Table 1. The mercury resistance loci described in this study

mer Locus[a]	Source[a]	Strain	Host biotype[b]	GenBank accession number	Closest relative	Identity[c] (%)
1	PA	517H	*Escherichia coli*	AF120960	Tn*21* (NR1)	100
	EPS	1349B	*Escherichia coli* or *Kluyvera*	AF120959		
2	PA	699H	*Citrobacter freundii*	AF120961	Tn*5053*	100
	EPS	1312A	*Escherichia coli*	AF120962		99.8
3	PA	722H	*Klebsiella oxytoca*	AF120975	pMER610	100
	EPS	224	*Citrobacter diversus*	AF120974		
4	PA	742H	*Leclercia adecarboxylata*	AF120965	pKLH272[d]	100
	EPS	209A	*Escherichia coli*	AF120964		
5	PA	660H	*Sphingomonas paucimobilis*	AF120972	Tn*5041*	95
	PA	661H	*Sphingomonas paucimobilis*	AF120973	TC97(5′*merRTP*)[e]	100
6	PA	687H	*Stenotrophomonas maltophilia*	AF120966	Tn*501*	99.7
	EPS	2600	*Ps. aeruginosa or fluorescens*	AF120967		
9	EPS	298B	*Pseudomonas aeruginosa*	AF120971	pMER610	83
	EPS	396	*Pseudomonas* or *Alcaligene*	AF120970		
10	EPS	394B	*Pseudomonas putidia*	AF120968	Tn*5041*	87
	EPS	317B	*Pseudomonas* or *Alcaligenes*	AF120969		
11	PA	660H	*Sphingomonas paucimobilis*	AF120977	pDU1358	89
	PA	661H	*Sphingomonas paucimobilis*	AF120976		

[a] Designation of *mer* operons from our previous study (Liebert et al. 1997). *mer* Loci 5, 9, 10, and 11 are newly described in this study. Loci 7 and 8 were not observed in either the PA (Primate Amalgam) or EPS (Environmental Plasmid Survey) collections (Wireman et al. 1997).
[b] Determined by the API 20E system from bioMerieux Vitek, Inc. (Hazelwood, MO).
[c] Percent identity to closest relative; the entire nucleotide sequence of the segment of *mer* operon sequenced (as indicated in Fig. 2) was compared to GenBank sequences (Fig. 1).
[d] Previous restriction enzyme studies (Liebert et al. 1997) and partial sequencing of 742H *merA* (data not shown) revealed identity to the pKLH272 *mer* operon rather than the pKLH2 *mer* operon.
[e] The portion of *mer* Locus 5 we sequenced is 95% identical overall with the Tn*5041 mer* locus and 100% identical to Tn*5041 merC*. It is also 100% identical with the portion (*merRTP*) of the TC97 *mer* locus that has been sequenced and the TC97 *merC* gene has the same restriction pattern as Tn*5041* (Yurieva et al. 1997).

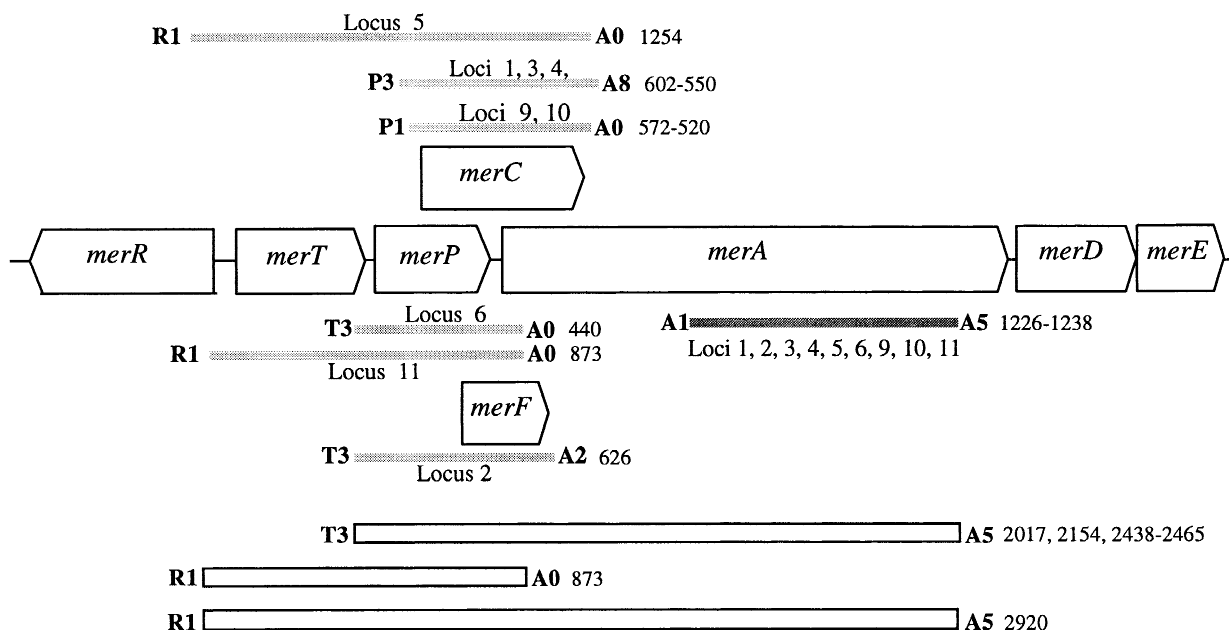


Fig. 2. Locations of the *mer* PCR primers used and regions sequenced in this study. Open boxes indicate the regions amplified (primary amplicons) and used subsequently as DNA templates for sequencing. Light gray boxes represent the regions sequenced. Dark gray box represents the region amplified from the primary amplicon (either T3-A5 or R1-A5) and digested with *BsrF*I to confirm the locus type by MRSP (Liebert et al. 1997). Primer designations (in bold) are on either side of the boxes, the different *mer* loci sequenced are indicated by the locus number above or below the boxes, and the sizes (bp) of the amplicons are to the right of the boxes.

$(NH_4)_2S_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 0.1 mg/ml nuclease free bovine serum albumin). Template DNA (50 ng in 1.0 μl) and a wax bead were added to a 0.5 ml Eppendorf tube and placed on ice. The chilled PCR mixture was added to the tubes, which were immediately placed in the thermal cycler for amplification. Cycles were 97°C for 45 s for one cycle; 97°C for 45 s, the appropriate annealing temperature for 45 s, and 72°C (or 76°C) for the appropriate extension time for 29 cycles; and 72°C (or 76°C) for 10 min for one cycle and finally held at 15°C. PCR annealing and extension temperatures and times are available upon request. To visualize the PCR product, 5 μl of each reaction mix were electrophoresed on 1% agarose gels, in 1X TBE (90 mM Tris base, 90 mM boric acid, 2 mM EDTA (pH 8.0)), stained with 0.1 mg/ml ethidium bromide (EtBr), and photographed. Appropriate positive and negative controls were included in every PCR experiment.

Purification and Characterization of PCR Amplicons. Template amplicons were electrophoresed in 0.7% low melt agarose gels in 0.5 X TAE buffer and then purified with the Wizard PCR Preps DNA Purification System (Promega, Madison, WI, USA). Before sequencing, an aliquot of each purified primary amplicon was used as PCR template to test that the sequencing primers produced an amplicon of the expected size. In addition, each primary amplicon was also used as a template to produce an A1-A5 amplicon. The A1-A5 amplicon (Fig. 2; dark gray rectangle) was digested with the restriction enzyme *BsrF*I (New England Biolabs, Beverly, MA, USA) without prior purification to confirm that the sequenced template was in fact the expected *mer* locus as defined by previous RFLPs studies (Liebert et al. 1997). The reactions were electrophoresed on 15 × 15 cm, 3% MetaPhor gels (FMC BioProducts, Rockland, ME, USA) in 1X TBE, at 100 V for 4.5 h (MetaPhor gels), which were subsequently stained with EtBr and photographed.

DNA Sequences. Sequencing of the template amplicons was performed by the University of Georgia Molecular Genetics Instrumentation Facility (Athens, GA) using an Applied Biosystem model 373A automated DNA sequencing system and a DyeDeoxy termination cycler sequencing protocol (Applied Biosystem, Incorp., Foster City, CA, USA). Both strands of each primary amplicon were sequenced at least twice in the region of interest from independent DNA preparations and amplifications (Fig. 2).

Computer Analyses. The software Sequencher (Gene Codes Corp., Ann Arbor, MI, USA) was used to align, view, and edit sequence data. Sequence analyses were performed using the Wisconsin Package Version 10, Genetics Computer Group (GCG), (Madison, WI, USA) programs: PILEUP, GAP, CODONFREQUENCY, CORRESPOND, and FINDPATTERNS. Codon usage indices were generated by CodonW (http:/bioweb.pasteur.fr/seqanal/interface/codonW/) (Peden 1997) and the GenBank Codon Usage Database (http://www.kazusa.or.jp/codon/) was used to generate the *Escherichia coli, Pseudomonas aeruginosa, Shewanella putrefaciens,* and *T. ferrooxidans* codon frequency tables. The DnaSP program version 3.1 (URL;http://www.bio.ub.es/~julio/DnaSP.html) was used to determine nucleotide diversity (π) (Rozas and Rozas 1997) and the method of Lynch and Crease (Lynch and Crease 1990) was used with the Jukes and Cantor correction (Jukes and Cantor 1969).

Secondary structure predictions were performed using PredictProtein (Rost and Sanders 1994), and PHDsec and PHDtopology (Rost et al. 1995), search for protein domains in PROSITE (Bairoch et al. 1997), and ProDom (Corpet et al. 1998) databases were performed using BLASTP (Altschul et al. 1990). Prediction of transmembrane helices in proteins was performed using TMHMM1.0 (URL;http://www.cbs.dtu.dk/services/TMHMM1.0/) (Sonnhammer et al. 1998) and plotted in CA-Cricket Graph III Version 1.5.3., Computer Associates International (Islandia, NY, USA).

Phylogenetic analyses were performed using the Phylogenetic Analysis Using Parsimony (PAUP) version 4.0 programs (Swofford 1998). Sequences were aligned using the PILEUP program in GCG (Devereux et al. 1984). Calculation of distance values was done with the Kimura 2 parameter or Jukes–Cantor options, and evolutionary trees were constructed with the UPGMA or neighbor-joining options (Saitou and Nei 1987). Bootstrap analysis (100 replicates) was also performed.

Results and Discussion

Is There Unusual Divergence in the Region Between the merP and merA Genes?

Overall Comparison of the P–A Regions

In the nine *mer* loci from wild isolates examined here (Table 1), we observed in the region between *merP* and *merA* either *merC, merF,* or no gene (Fig. 1). The P–A regions of three of these nine primate strain *mer* loci were 100% identical to published sequences we have defined as index examples of *mer* Loci 1, 3, and 4, respectively (Table 1).

Primate strains 660H and 661H carry at least two *mer* operons, only one of which (Locus 5), could we characterize previously (Liebert et al. 1997). Locus 5 contains a *merC* gene, has the operon structure *merRTPCA,* and is 100% identical to the *merRTP* region of the TC97 *mer* operon (see Table 1 legend). Here, using gel purification of amplicons, we have characterized a second *mer* locus, Locus 11, from strains 660H and 661H. Locus 11 lacks *merC,* has the genes *merRTPA,* and is a novel *mer* operon, albeit, with some similarity in its *merRTP* region to the pDU1358 (Locus 8, 89% identical) *mer* operon.

Sequencing of previously defined Loci 9 and 10 revealed that the entire 3′ *merP, merC,* and 5′ *merA* (PCA) region of Locus 9 is only 83% identical to its closest relative, pMER610. The *merC* gene itself in Locus 9 has suffered a 5′ deletion of 42 nt and may be nonfunctional. The PCA region of *mer* Locus 10 is only 87% identical to its closest relative, the *mer* operon of Tn*5041*.

Unlike the Tn*21 mer* operon, Locus 2 has the *merF* gene instead of *merC* lying between *merP* and *merA*. The P–A region of this locus in our primate strain 722H is 100% identical to both published sequences of *merF,* one of which (Tn*5053*) we had defined as the index example of *mer* Locus 2 (Fig. 1, Table 1). The P–A region of strain 1312A differed by one nt from the corresponding published sequences resulting in *merP* Ser27, a small polar amino acid, on the outer surface of the protein (Steel and Opella 1997), being replaced by Phe, a large hydrophobic amino acid.

Unlike the Tn*21* and Tn*5053 mer* operons, the Tn*501* (Locus 6) *mer* operon has no protein-encoding gene but only a 71 bp intergenic region between *merP* and *merA*. The partial sequence of the Locus 6 *mer* operon indicates that it and Tn*501* are the same "sequence type," since they differ by fewer than 0.5% of their nucleotides (Milkman and Bridges 1990). Both *mer* Locus 6 ex-

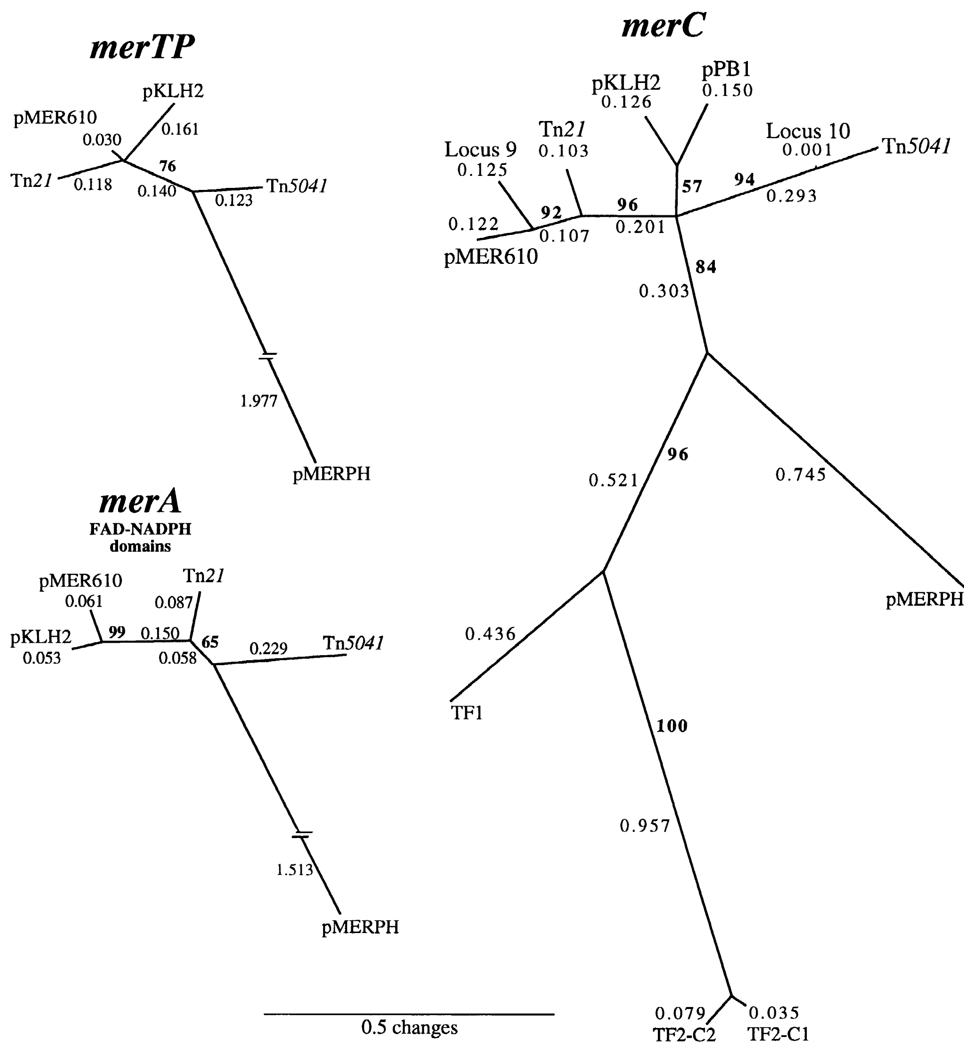


Fig. 3. Individual phylogenies of *mer* operon genes. Distance trees based on the five available *merTP* and *merA* sequences of operons which have *merC,* and on the eleven *merC* gene sequences available; constructed by the neighbor-joining method using PAUP (Swofford 1998). Bootstrap percentages are in bold.

amples here differed by one nucleotide from the published sequence of Tn*501* (Fig. 1, Table 1). The *mer* Locus 7 (*Thiobacillus* TF1 or TF2) and Locus 8 (pDU1358) were not observed in either the EPS or Primate Amalgam collections.

Thus, in this study, we, as others (Fig. 1, Table 1), have observed three gene order types: those with *merC,* those with *merF,* and those with no gene between *merP* and *merA*. The actual sequence divergence in the *merC* genes of Loci 1, 3, 4, 5, 9, and 10 reported in this study confirms the differences detected previously by hybridization (Liebert et al. 1997).

Individual Gene Phylogenies

A phylogenetic tree of all sequenced *merC* genes (Fig. 3) was robust with high bootstrap values. Comparison of different distance and tree drawing algorithms (see Methods) yielded similar branching patterns (data not shown). There are no obvious sharp transitions between homology and nonhomology (i.e., crossover points) to indicate recombination has occurred within any of the *merC* genes (data not shown; map available upon request). *Thiobacillus* and pMERPH *merC*s diverge extensively from the others; this is especially so for *Thiobacillus* TF1 *merC* and *Thiobacillus* TF2 *merC1* (and *merC2*). In a phylogeny based on *merR* genes (Osborn et al. 1995), the *Thiobacillus* TF2 *merR* also diverged from *merRs* of other Gram-negative bacteria. The regions of greatest variation in the *merC* gene corresponded to nt 1–66 (protein residues 1–22), nt 280–309 (protein residues 94–103), and nt 409–420 (protein residues 137–140) in Tn*21 merC* (Fig. 4).

Although there are limits to conclusions which can be drawn from phylogenetic comparisons for highly peripatetic loci such as *mer,* there appear to be at least three lines of descent for *merC* represented by: (1) those in the

MerC

```
                   ....,....1....,....2....,....3....,....4....,....5....,.
TF2 C1        1           MSRILDKTGVLGTVVSGLSCAMCFPALASIGAAVGLGFLSRWEGLFVHTLIPV
TF1           1        MSAITRIIDKIGIVGTIVGSFSCAMCFPAAASLGAAIGLGFLSQWEGLFVQWLIPI
Locus 9       1                                MGCAACFPAIASLGAAIGLGFLQEYEGLFISKLLPL
Tn21          1        MGLMTRIADKTGALGSVVSAMGCAACFPALASFGAAIGLGFLSQYEGLFISRLLPL
pMER610       1        MGLITRIADKAGALGSVVSAMGCAACFPAIASLGAAIGLGFLQEYEGLFISKLLPL
Tn5041        1    MANPLNLFTRIGDKAGSVGVLISAIGCAMCFPALASLGAAFGLGFLSQWEGLFITTLLPL
Locus 10      1    MANSFNLFTRIGDKAGSVGVLVSAIGCASCFPALASLGATIGLGFLSQWEGLFITTLLPL
pKLH2         1        MGLITRIAGKTGALGSVVSAMGCAACFPAIASFGAAIGLGFLSQYEGLFIGILLPM
pPB(broad)    1        MSLITRVTDKAGSLGSVVSAMGCATCFPAIASLAAAIGLGFLSQYEGLFIRILLPT
pMERPH        1      MLIRLLTRFGDKISSLGALVSAMGCVMCFPAIASLGAAVGLGFLSQWEGLFVSTLLPL
consensus     1        m litriadk g lgtvvsamgCa CFPAiASlgAaiGLGFLsqwEGLFis LlPl

               ...6....,....7....,....8....,....9....,....0   ... .,....1..
TF2 C1       54    FLVIALLANGLGWFAHHQWQRSALGCVGPVVALFGDQGMTQHFLPVEMARILFYTGLVIM
TF1          57    FASVALLATLAGWFSHRQWQRTLLGSIGPVLALVGVFGLTHHFLDKDLARVIFYTGLVVM
Locus 9      37    FAVVALLANALGWLSHRQWHRSLLGMIGPAIVFAGTVWLLGNWW...AAH.LVYTGLALM
Tn21         57    FAALAFLANALGWFSHRQWLRSLLGMIGPAIVFAATVWLLGNWW...TAN.LMYVGLALM
pMER610      57    FAVVALLANALGWLSHRQWYRSLLGMVGPAIVLAAMLLFFGNWW...TAN.LLYVGLALM
Tn5041       61    FAGLALLVNALGWLNHRQWRRTALGLIGPTLVLAAVFFMRAYGW...RSGWLLYIGLALM
Locus 10     61    FAGVALLVNALGWFSHRQWRRAAPGLIGPALVLVAVFLMLAHGW...RSGWLLYIGLAVM
pKLH2        57    FAGIALLANAIAWLNHRQWRRTALGTIGPILVLAAVFLMRAYGW...QSGGLLYVGLALM
pPB(broad)   57    FAGIALFANAISWRNHRQWRRTAPGVIGPVLVLAAVFLMRAYGW...RSGWLLYIGLAFM
pMERPH       59    FAWIALVLNGLGWFSHRQWHRSALGVAGPILLLLSLYPFFQYGW....SSYVTYSALGLM
consensus    61    FaavAllanalgWfsHrQWhRsalGmiGPvlvlaavf m g gw       llYvgLalM

                   ..,....2....,....3....,....4
TF2 C1      114    VIIAIWDMFRPAHKHCPVPSGDNQSHC
TF1         117    FLVSIWDMVNPANR.CATDGCETPAPRS
Locus 9      93    VGVSIWDLVSPANRRCGPDGCELPTKQG
Tn21        113    IGVSIWDFVSPAHRRCGPDGCELPAKRL
pMER610     113    IGVSIWDLLSPANRR.....CELPPKHG
Tn5041      118    FGVSIWDLVSPAHRRCGPDACPTPPNQ
Locus 10    118    FGVSIWDFVSPAHRRCSPASCPTPPNQ
pKLH2       114    VGVSVWDFISPAHRRCGPDSCELPEQRG
pPB(broad)  114    VGVSIWDFISPAKPCCGPDSCELPPNAG
pMERPH      115    VAVSLWDIFSPANKRCDDDSCAV
consensus   121    vgvsiWDmvsPAhrrcgpdscelp
```

Fig. 4. Alignment of eleven unique MerC amino acid sequences. Numbers in the top line are the residue position numbers for Tn*21* MerC.

Thiobacillus loci; (2) those in a cluster consisting of Tn*21*, pMER610, pKLH2, pPB1 and Tn*5041;* and (3) the one gene in the pMERPH operon.

Of the *mer* operons which have *merC,* there is sequence information for flanking genes (*merT, merP,* and *merA*) only in Tn*21,* pKLH2, pMER610, Tn*5041,* and pMERPH. We used this information to generate phylogenetic trees of the region including both of the small genes *merT* and *merP* (ca. 627 bp total) and the region encompassing the FAD and NADPH domains of *merA* (675 bp) genes (Fig. 3). Note that here, and in what follows, *merT* and *merP* are considered together because each alone is too small for statistical validity. As in the *merC* tree above, the outlier in the *merTP* and *merA* trees is pMERPH. Surprisingly, the distances of the pMERPH *merTP* and *merA* genes from other *mer* genes are even larger than the distances of the pMERPH *merC* from the other *merC* genes. The shorter branch lengths between *merTP*'s and *merA*'s of pMER610 and pKLH2 likely arise from extensive crossover that occurs in *merT* and *merA* that is not seen in *merC* (Yurieva et al. 1997) (and our data not shown; map available upon request). The branch order for all three genes differs; this could arise if a crossover in *merA* is not always accompanied by crossover in *merTP* but occurs elsewhere. The branch lengths of the three trees indicate close relations among the adjacent *merTP* and *merA* genes in operons containing *merC* with the exception of the pMERPH and the *Thiobacillus* loci and is consistent with two of the three lines of descent noted for *merC* itself.

Codon Utilization

Genes with atypical codon usage compared to flanking sequences are considered to have been acquired recently from distantly related organisms via horizontal transfer (Lawrence and Ochman 1997; Medigue et al. 1991). We compared codon usage of the *merTP, merC,* and *merA* genes using the GCG program, CORRESPOND (Grantham et al. 1981), which compares two Codon Frequency Tables (Devereux et al. 1984) and yields the residuals statistic D^2 (see Table 2 legend). Lower D^2 values imply either similar origins or sufficient elapsed time for amelioration to have occurred. Codon usage comparisons across operons with *merC* indicate significant diversity in the *merC* genes compared to the *merTP* and *merA* regions (Table 2). The general trend of these comparisons is a larger D^2 value (values >3 in boldface) and lower percent DNA sequence identity (data not shown but note Fig. 3) among *merC*s compared to *merTP* or *merA* genes. Moreover, the low D^2 values seen between the TF1 *merA* and *merC* genes and the corresponding Tn*21* genes suggest a similar origin for these genes. Likewise, similar codon usage between pMER610 and TF2 *merC*s and between Tn*5041* and pKLH2 *merC*s suggest common ancestry. Thus, codon usage of most *merC* genes, like many genes found in

Table 2. Codon usage comparison (D^2)[a] among *merTP, merC,* and *merA* genes between operons with *merC*

merTP/TP *merC/C* *merA/A*	Tn*21*	pMER610	pKLH272[b]	pKLH2[b]	pPB1[c]	*Tn5041*[d]	pMERPH	TF1[e]	TF2[e]
Tn*21*	0	<1	<1	<1	1.59	2.30	**3.93**	—[f]	—
	0	**4.71**	**4.21**	**4.21**	**3.72**	**7.20**	**10.2**	2.31	**6.00**
	0	1.00	<1	1.80	—	1.62	**7.81**	2.28	—
pMER610		0	<1	<1	<1	1.76	**3.96**	—	—
		0	**5.24**	**5.24**	**4.72**	**6.79**	**4.34**	**4.93**	2.25
		0	<1	<1	—	<1	**7.66**	1.36	—
pKLH272			0	<1	<1	1.91	**3.86**	—	—
			0	0	**4.12**	2.77	**9.00**	**5.05**	**6.13**
			0	1.18	—	1.33	**7.69**	2.05	—
pKLH2				0	1.40	1.49	**4.79**	—	—
				0	**4.12**	2.77	**9.00**	**5.05**	**6.13**
				0	—	<1	**8.70**	1.25	—
pPB1					0	2.05	**3.44**	—	—
					0	**4.22**	**9.21**	**4.73**	**6.66**
					—	—	—	—	—
Tn*5041*						0	**5.62**	—	—
						0	**7.74**	**5.03**	**8.34**
						0	**9.78**	1.33	—
pMERPH							0	—	—
							0	**6.12**	**4.04**
							0	**7.15**	—
TF1								—	—
								0	**7.25**
								0	—
TF2									—
									0
									—

[a] The D^2 statistic, calculated with the GCG program CORRESPOND, compares the pattern of codon usage generated by the GCG program CODONFREQUENCY (Devereux et al. 1984). The D^2 statistic is the sum over all codons of the residual equation: $D^2 = (\text{frequency}_{\text{codon table 1}} - \text{frequency}_{\text{codon table 2}})^2$. D^2 values of <3.0 are consistent with a common origin for the genes (Grantham et al. 1981). D^2 increases as patterns of codon usage become less similar; D^2 values of >3.0 are in boldface

[b] Due to a recombination event between the pKLH2 and pKLH272 operons (Yurieva et al. 1997) (and our unpublished observations), these operons have identical *merC* sequences

[c] The *merA* gene of the pPB1 *mer* operon has not been sequenced

[d] Tn*5041* and Locus 5 have identical *merC* sequences. Codon usage comparisons of Locus 5 *merTP* to other *merTP* sequences were similar to those of Tn*5041.* The *merA* gene of Locus 5 was not sequenced

[e] Neither *T. ferrooxidans mer* operon has adjacent *merT* or *merP* genes. For the TF2 *mer* operon, the *merC1* sequence was used

[f] Dash indicates one or both sequences not available. The *merTP* and *merA* genes of Loci 9 and 10 were not sequenced. Locus 9 D^2 values for *merC/C* comparisons were similar to those of Tn*21.* Locus 10 D^2 values for *merC/C* comparisons were similar to those of Tn*5041* with the exception of pKLH2 (**4.74**) and TF1 (2.50)

Table 3 Codon usage comparisons (D^2)[a] among *merTP, merC,* and *merA* sequences within the same operon

	Tn*21*	pMER610	pKLH272	pKLH2	pPB1	Tn*5041*	pMERPH	TF1
merTP/A	**3.12**	1.81	2.99	2.28	—	1.45	**3.13**	—
merC/TP	2.82	**8.10**	**7.91**	**8.40**	**4.29**	2.76	**3.19**	—
merC/A	2.45	2.98	**4.53**	**5.76**	—	**9.20**	**3.64**	2.92

[a] See Table 2 legend

integrated cassettes in integrons, is quite dissimilar from other genes found near them (Bissonnette and Roy 1992).

We also compared codon usage among *merTP, merC,* and *merA* sequences within the same operon (Table 3). In five of the eight operons analyzed, codon usage comparisons of *merC* with either *merTP* and/or *merA* yielded high D^2 values. So by D^2 analysis, whether comparisons are made among different operons or between genes in the same operon, *merC* stands out as less related to other genes in the operon than they are to each other.

An additional measure of genetic idiosyncrasy is the synonymous codon usage bias, the effective number of codons (EN_C) (Wright 1990) which quantifies how far the codon usage of a gene departs from equal usage of synonymous codons. This measure indicates the degree of codon preference in a gene. For unicellular organisms, the degree of synonymous codon usage bias is highly

correlated with the level of a gene's expression (Gouy and Gautier 1982; Ikemura 1985). For *E. coli,* highly expressed genes tend to use a subset of 25 preferred codons, whereas weakly expressed genes tend to use a wider variety of the 61 sense codons (Bennetzen and Hall 1982).

Although generally useful distinctions about genetic incursions can be made when the EN_C is plotted against the percent of G+C in the third position of the gene (Wright 1990), we did not observe any trend in the EN_C as a function of G+C content of the *mer* genes. However, distinct trends appeared when we plotted the EN_C against the D^2 values derived from comparisons of *mer* genes with the chromosomal genes of four bacterial species (Fig. 5). *E. coli, P. aeruginosa, S. putrefaciens,* and *T. ferrooxidans* were selected since these species represent bacteria which carry *merC* genes from one of the three deeply divergent branches of the *merC* phylogenic tree (Fig. 3).

With respect to EN_C (Fig. 5), *merA* and *merTP* both have codon usage biases with little dispersion (except for the genes of pMERPH). In contrast, *merC*s' codon biases range widely in their predictions from moderate expression by Tn*21,* to weak expression by pMER610. Thus, *merC* tends to be much more variable in its predicted expressability. Clustering of D^2 values (Fig. 5) for *merA* suggests it is more closely related to the genes of *P. aeruginosa* and *T. ferrooxidans* and very distantly related to the genes of *S. putrefaciens.* D^2 values for *merTP* show some clustering by host but are slightly more dispersed than *merA.* In contrast, *merC* shows much less host-specific clustering. Similar trends were observed when the Codon Adaptation Index (Sharp and Li 1987) was plotted against D^2 values (data not shown).

As with our other comparisons, EN_c and D^2 values of pMERPH *mer* genes (Fig. 5) show obvious differences from the other *mer* genes. The pMERPH genes all appear by EN_C to be weakly expressed. Interestingly, the expression of pMERPH *mer* genes occurs constitutively (Osborn et al. 1996), unlike Tn*21 mer* structural genes which are tightly regulated (Summers 1992). Also, unlike the other *mer* genes, the D^2 values of the pMERPH genes cluster in a manner suggesting a closer relation to chromosomal genes of *E. coli* and to *S. putrefaciens.* The most parsimonious interpretation of the overall trend in the D^2 values suggests that several of these *merC* genes were not coeval with their flanking sequences. Moreover, the EN_C values suggest that *merC* may be expressed at different levels in different host bacteria.

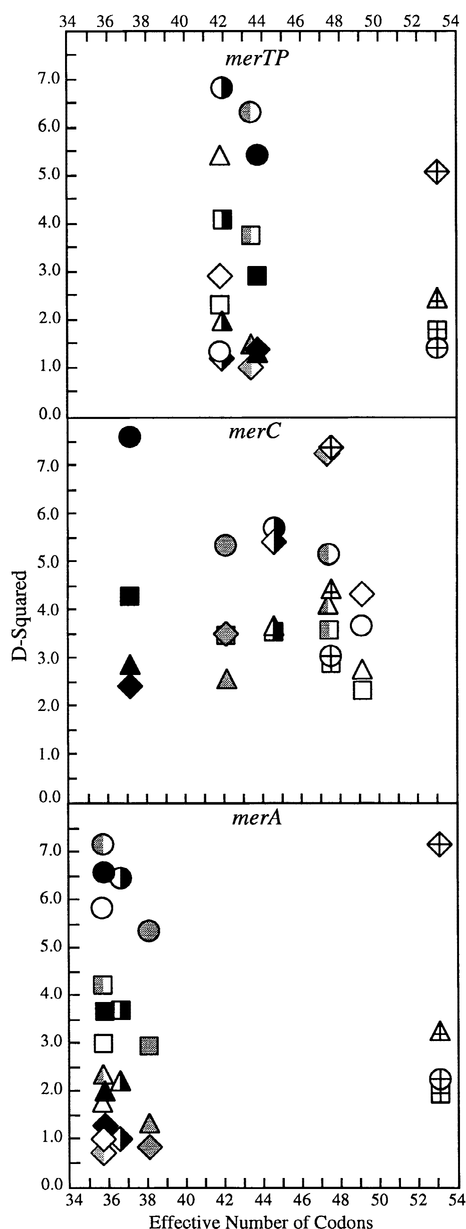


Fig. 5. D^2 values derived from the comparison of codon usage in *merTP, merC,* and *merA* genes of *mer* operons [Tn*21* (solid black); pMER610 (open); pKLH2 (black and white); Tn*5041* (gray and white); pMERPH (crossed); and TF1 (solid gray)] with those for the chromosomal genes of *E. coli* (square); *P. aeruginosa* (diamond); *S. putrefaciens* (circle); and *T. ferrooxidans* (triangle) plotted against the effective number of codons used in *merTP, merC,* and *merA* genes.

Do the merC Genes Diverge More than Adjacent Genes?

The DNA sequence similarity among the *merC* genes is less (with two exceptions) than that among the *merTP* or *merA* genes positioned on either side of them. Excluding pMERPH, which varied greatly from the other loci in every gene, the average percent identity among *merTP* genes was 83.7 ± 5.4 and among the *merA* regions it was 81.0 ± 4.5, whereas among *merC* genes it was 68.4 ± 7.9. These average identities suggest that *merC* is less conserved, although this distinction is only statistically significant at the 68% confidence interval for the limited number of sequences we have for this relatively small gene.

The nucleotide sequences of the *merC* and of the *merR, merP,* and *merA* genes in loci containing a *merC*

Table 4 Nucleotide diversity measure π (%) for the *merR, merP, merC,* and *merA* genes

mer genes[a]	Sites[b]	π[c]
merR	400	39.03
merP	267	35.82
merC	354	49.63
merA	1374	32.58

[a] The *mer* loci used in this analysis were Tn*21,* pMER610, pKLH2, pPB1, Tn*5041,* and *T. ferrooxidans* TF1

[b] Total sites (excluding sites with alignment gaps) analyzed per alignment

[c] Number of nucleotide differences per 100 sites

gene were aligned using the GCG program PILEUP. Alignments of *mer* genes were used to determine overall nucleotide diversity (Table 4), the number of nucleotide differences per site (π) (Rozas and Rozas 1997). These results also indicated that there is less diversity among the flanking *merP* and *merA* genes than among the *merC* genes. The more physically distant *merR* gene, also had a lower π value than *merC.* Thus, diversity in the P–A region consists of the presence or absence of two distinct genes, one of which (*merC*) is itself quite variable by comparison to its immediate neighbors. The variation in *merC* suggests it might not generally be essential or might even be vestigial. To consider this question, we asked whether the variation in the P–A region affects the conserved motifs of the proteins occurring there.

What Would be the Consequences of P–A Region Divergence for the Structure of the Proteins Encoded in This Region?

MerC

The MerC proteins are small, hydrophobic proteins ranging in size from 129–144 aa (Fig. 4). MerC of Tn*21* is in the cytoplasmic membrane (Jackson and Summers 1982). The MerC proteins all conserve two cysteine pairs, except the MerC of pMER610, which lacks the first cysteine of the C-terminal pair. Most MerC proteins are predicted to have four transmembrane helices with the N-terminal cysteine pair lying near the middle of the first transmembrane helix (Fig. 6, top panel) and the C-terminal cysteine pair lying in the cytosol (Rost et al. 1995; Sonnhammer et al. 1998). The exceptions to this predicted topology are the very similar MerC1 (Fig. 6) and MerC2 (not shown) proteins of *Thiobacillus* TF2, which may lack the third transmembrane helix, resulting in their C-terminal cysteine pair probably lying in the periplasm.

All MerC proteins conserve a C̲AXC̲FPA motif in the N-terminal cysteine pair which is distinct from the heavy metal associated (HMA) motif (GMXCXXC) of MerP, the N-terminal region of MerA, and many metal transporting P-type ATPases (Bull and Cox 1994) (Fig. 4). The spacing (C̲XXXXC̲) of the most common C-terminal cysteine pair is also distinct from the HMA motif; note that in the unusual *Thiobacillus* TF2 MerC1 and MerC2, the C-terminal spacing is C̲X_{10}C̲. These latter two MerC variants are identical with the exception of two residues (I/V116, A/V125) and also have a fifth cysteine at position 79 predicted to lie in the cytoplasm (Figs. 4 and 6). Thus, with the exception of *Thiobacillus* TF2 MerC1 and MerC2, the predicted structural variation within MerC itself is quite limited and putative functional elements (metal-binding ligands and transmembrane domains) are strongly conserved.

Comparison of MerC to MerT

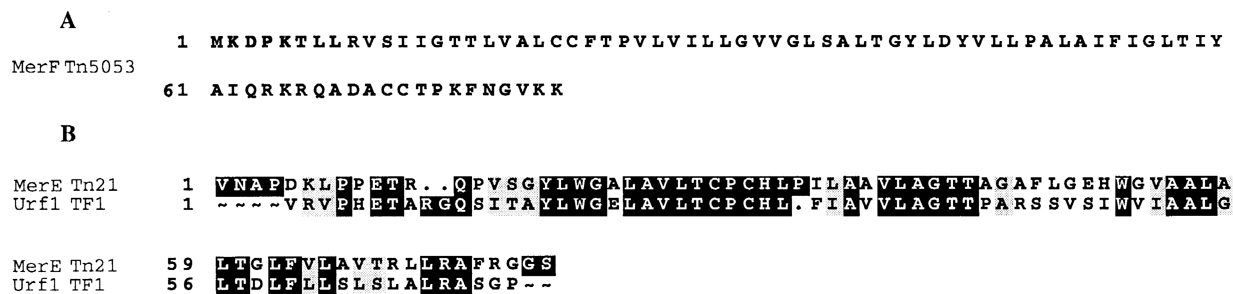
MerT is also an inner membrane protein (Jackson and Summers 1982). The most common MerC protein shares a few similarities with MerT (Fig. 6). Both have a cysteine pair within the first predicted transmembrane helix and a C-terminal cysteine pair predicted to be in the cytosol. However, a subset of MerC proteins, (Tn*21* cluster, Fig. 3) have a C-terminal cysteine pair motif (C̲X-PXXC̲XXP) very similar to that of MerT (C̲XPXXX-C̲XXP). Surprisingly, the predicted topology of *Thiobacillus* TF2 MerC1 and MerC2 resembles that of Tn*21* MerT (Fig. 6), although MerT lacks the last 18 aa of *Thiobacillus* TF2 MerC1 and MerC2 which are predicted to be in the periplasm. Despite the predicted topological similarities, amino acid pairwise comparisons of TF2 MerCs are more similar to other MerCs than to MerTs (data not shown).

MerF and Putative MerE

The 81 aa MerF protein (Fig. 7) is hydrophobic and secondary structure predictions (Rost et al. 1995; Sonnhammer et al. 1998) indicate two transmembrane helices (Fig. 6), similar to MerT. MerF has an N-terminal cysteine pair (C20, C21) predicted to lie near the middle of first helix and its C-terminal cysteine pair (C70, C71) is predicted to lie in the cytsol (Fig. 6). MerF's structure is reminiscent of two other predicted small hydrophobic *mer* proteins, MerE (78 aa) (Liebert et al. 1999) and *Thiobacillus* TF1 Urf1 (Fig. 6) (Inoue et al. 1989). Ten promoter distal, putative *merE* genes and one *Thiobacillus* TF1 urf1 gene, immediately adjacent and upstream of *merC,* have been sequenced (Fig. 1). The predicted MerEs and *Thiobacillus* TF1 Urf1 proteins all have cysteine pairs (CXXC) positioned similarly to the N-terminal cysteine pair of *merF* (Fig. 6), suggesting a possible role in facilitating mercuric mercury transport as recently demonstrated for MerF (Wilson et al. 2000).

Thus, we find that MerC's putative structure is not affected by extensive DNA-level variation. Also, we observe putative structural similarities and conserved cys-

Fig. 6. Topology prediction of Tn*21* MerC, *T. ferrooxidans* TF2 MerC1, Tn*21* MerT, Tn*5053* MerF, Tn*21* MerE, and *T. ferrooxidans* TF1 Urf1. Position of cysteines marked on residue axis. Solid line, transmembrane; dash, cytosol; and dot, periplasm.



A

```
               1  MKDPKTLLRVSIIGTTLVALCCFTPVLVILLGVVGLSALTGYLDYVLLPALAIFIGLTIY
MerF Tn5053
              61  AIQRKRQADACCTPKFNGVKK
```

B

```
MerE Tn21    1  VNAPDKLPPETR..QPVSGYLWGALAVLTCPCHLPILAAVLAGTTAGAFLGEHWGVAALA
Urf1 TF1     1  ~~~~VRVPHETARGQSITAYLWGELAVLTCPCHL.FIAVVLAGTTPARSSVSIWVIAALG

MerE Tn21   59  LTGLFVLAVTRLLRAFRGGS
Urf1 TF1    56  LTDLFLLSLSLALRASGP~~
```

Fig. 7. (A) Sequence of Tn*5053* MerF and (B) alignment of Tn*21* MerE with *T. ferrooxidans* TF1 Urf1.

teine pairs in MerC, MerF, MerT, and MerE. Cysteine pairs in MerT, MerC, and MerF have been shown to be required for transport of mercury across the cell membrane (Hamlett et al. 1992; Morby et al. 1995; Sahlman et al. 1999; Wilson et al. 2000). Thus, with MerT and MerP present, either MerC or MerF could supply additional membrane and cytosolic cysteine pairs to bind mercury, whether this also occurs in MerE has yet to be tested.

What Might Be the Agents of Divergence in the merP-A *Region?*

RecA-Dependent Homologous Recombination

Since many *mer* operons occur in cells which have another *mer* operon (Griffin et al. 1987; Liebert et al. 1997; Reniero et al. 1998), often on the same plasmid, host-mediated homologous recombination might contribute to their reassortment (Lloyd and Low 1996). Crossover events involving sequence exchange greater than 1 kb generally require RecA. Sequence identity-dependent crossovers, in some cases RecA-dependent, can occur between sequences as short as 50- to 100-bases (Shen and Huang 1986; Watt et al. 1985). For shorter sequences, RecA-independent mechanisms can come into play (Lloyd and Low 1996).

Abrupt high/low identities (crossover points) between two sequences (Maynard Smith 1992) signal crossover events and many are observed in *mer* operons (map available upon request). Long regions of sequence identity (>99%) occur between *mer* operons pKLH2 and pKLH272 (1.3 kb), pMER610 and pKLH272 (1.1 kb) and Tn*501* and pPB1 (2.4 kb). Shorter regions of sequence identity, ranging from 50- to 900-bases, occur in many *mer* operons. Interestingly, there are no apparent crossover points in any of the sequenced *mer*Cs. Crossover points are seen between all the *mer* operons in the flanking regions (those with and without *merC* or *merF*) with the exception of the *Thiobacillus* and pMERPH operons. Apparently the higher level of divergence between the *merC* genes does not readily allow homologous recombination within them. However, the flanking regions with higher sequence identities do appear to be recombining, driving their overall sequence identities higher.

The Role of Homologous Recombination and Chi

The RecBCD-dependent recombination activity is stimulated by the Chi DNA sequence and increases homologous recombination for approximately 8 kb to the 5′ side of the sequence (Smith 1994). Chi occurs throughout the *E. coli* chromosome approximately once per 5 kb (Smith 1994). We observed a total of 31 Chi or Chi single-base variants in the total 107.7 kb of sequence containing *mer* operons (1.4 occurrences/5 kb); one-third of these were Chi itself (0.5 Chi/5 kb; Fig. 1, gray flags). The rare Chi variant 5′GCT<u>A</u>GTGG3′ (10% relative activity) (Cheng and Smith 1984) is normally observed twice per 500 kb; we observed it twice per 107.7 kb. Thus, Chi-like sequences are present in and near *mer* operons, but not in greater numbers than expected. Nonetheless, there is a Chi site in T*n21 merA* and one in *merC* that may stimulate recombination near *merC* in this operon.

Other Possible Agents of Variation

Finally, we examined the intergenic sequences for any recurring patterns that might indicate sites of action for yet undefined recombinases. We aligned the intergenic region between *merT* and *merP* (TP) and also the intergenic regions between *merP* and *merC* (PC), and *merC* and *merA* (CA), and the *merP–A* intergenic region of operons lacking *merC* or *merF* (PA). The TP regions (Fig. 8) are very similar with the exception of the pMERPH intergenic region. The TP region is ca. 19 bp (except pMERPH which is 62 bp). Nine of 12 nucleotides immediately upstream from the *merP* start site, containing the ribosome binding site (RBS), are conserved. The largest variability in the TP region occurs immediately 3′ of the *merT* gene.

In contrast to the TP region, the PC region (Fig. 8) ranges from 12 bp (pMERPH) to 35 bp (Tn*21*) and the CA region ranges from 35 bp (Tn*5041*) to 51 bp (Tn*21*). The only conserved regions in the PC and CA regions corresponded to the RBSs for *merC* and *merA,* respectively. The RBSs for *merC* also diverge more from the consensus sequence and from optimal spacing before the first codon than the RBSs of *merP* or *merA.* This divergence is consistent with variability in expression of *merC* as suggested by *merC* EN_C values (Fig. 5).

As yet, only one sequence of the *merF* gene and flanking regions has been described (Hobman et al. 1994; Kholodii et al. 1993b). The intergenic region between *merP* and *merF* consists of only 2 bp and there is no intergenic region between *merF* and *merA* since they overlap by 4 bp (ATGA).

For operons which lack *merC* or *merF* (Tn*501* and pDU1358), the PA regions (Fig. 8; marked with asterisks) are 71 bp each and are 84.5% identical. The PA region of Tn*501* is only 48% identical to the corresponding PC and CA regions of Tn*21* (i.e., flanking the arrow in Fig. 8). The PA region for Locus 11 is identical to that of pDU1358 and the PA region for pPB1 is 97.2% (69/71 bp) identical to that of Tn*501* (data not shown).

Although no obvious patterns in the intergenic regions between *merP* and *merA* were discerned, the high variability seen in *merC* clearly extends beyond the structural gene itself and into the adjacent intergenic regions. If, as previously proposed (Summers 1986), gene duplication and divergence of *merT* led to the formation of *merC* early in operon formation, little evidence of this initial acquisition is likely to be seen. The irregular lengths of the intergenic region flanking *merC* could re-

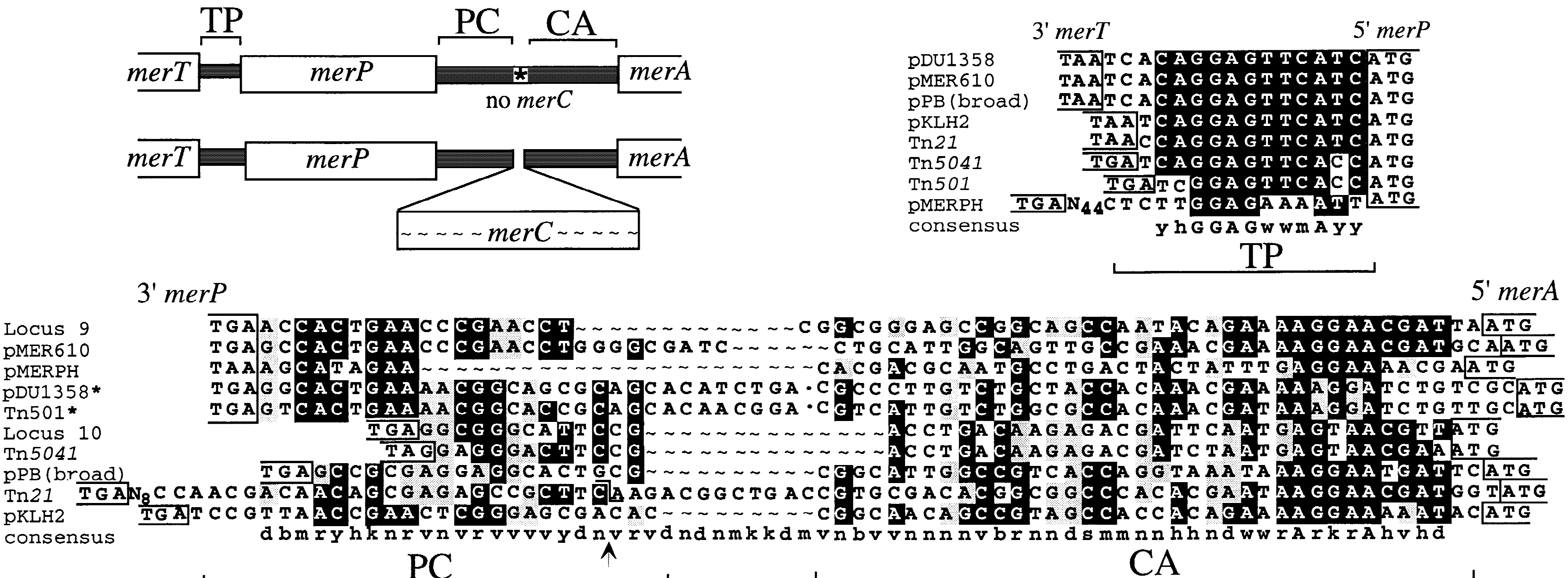


Fig. 8. Intergenic regions between *merT* and *merP* and between *merP* and *merA*. The *mer* operons of Tn*501* and pDU1358 lack *merC* (*). Brackets indicate intergenic regions between *merT* and *merP* (TP), between *merP* and *merC* (PC), or between *merC* and *merA* (CA) of *mer* operons with *merC*. Dot (.) indicates gap and tilde (~) indicates the respective *merC* sequence. Arrow and angle bracket indicate the position of the *merC* in Tn*21*.

flect different evolutionary constraints in various hosts resulting in variable erosion of intervening material.

Thus, we find that with respect to how divergence might arise: (1) Homologous recombination is evident in *mer* operons on either side of *merC* but not within *merC.* (2) *mer* operons have the normal opportunities for Chi-driven homologous recombination, perhaps somewhat increased near *merC* in Tn*21.* (3) There is no evidence of site-specific recombination mechanisms in intergenic flanking regions. Indeed, the spacing variations in *merP–C* and *merC–A* junctions are consistent with very low-fidelity recombination processes.

In summary, the greatest variation in the P–A region is the presence or absence of *merC* or *merF,* and when *merC* occurs, it is even more variable than are flanking genes. Although, homologous recombination mediated through flanking genes can explain the presence or absence of *merC*/*merF,* the greater sequence diversity in *merC* is harder to explain. It is possible that one example of the *mer* operon experienced gene duplication of *merT* and *merP* (Summers 1986) and later exchanged segments with another operon by homologous recombination forming mosaic *mer* operons containing *merC* or *merF.* These events most likely occurred under the constraints defined by the selfish operon model (Lawrence and Roth 1996) which postulates that operon formation is driven by proximal genes, horizontally cotransferred, which maintain a selectable phenotype. In the case of *mer* operons, the *merC* and *merF* genes, wedged between essential genes, would be cotransferred as long as they did not compromise the expressed phenotype in a given host.

So the question arises, is *merC* merely a redundant gene which has yet to be eliminated from some of the *mer* operons? If so, why conserve important metal-binding ligands and transmembrane domains? Perhaps MerC augments mercury resistance under certain conditions or in some bacteria (or niches) and not in others. If MerC confers an advantage, it is more likely to be conserved, but if MerC is nonessential, it will diverge or be deleted altogether. Thus, *merC,* when present, diverges slightly, but measurably more than essential flanking genes. Since *mer* operons are continually undergoing recombination, the more homologous flanking regions become more similar to each other over time, while *merC* continues to drift as the operons variously find themselves in hosts or conditions requiring it or not.

In conclusion, our observations establish quantitatively that the P–A region is highly variable compared to other regions of the *mer* operon. This variability involves alternative small hydrophobic proteins with one or two cysteine pairs and two or four transmembrane helices. Variability within one of these genes (*merC*) is great but does not compromise its more obvious structural elements. Nonetheless, it suggests that MerC may only be sporadically needed. The variation in the P–A region may be mediated by homologous recombination between more conserved adjacent genes as has been shown for other mosaic genetic elements (Campbell 1994). This is the first documentation of such drift in a module of a plasmid-encoded resistance operon. The set of criteria we have applied here may prove useful in the discovery of other such occasionally necessary components of operons which enjoy extensive horizontal transfer.

Acknowledgments. We gratefully acknowledge the cheerful and competent assistance of Lysha A. Cook and Marsha A. Boxx, GCG advice from Michael Weise, comments on the manuscript by Nigel D. F. Grindley, Michael M. Cox, John F. McDonald, and two anonymous reviewers. This project was partially supported by grants from NIH (GM 28211), the Wallace Research Foundation, and the International Academy of Oral Medicine and Toxicology to A.O.S.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bairoch A, Bucher P, Hofmann K (1997) The PROSITE database, its status in 1997. Nucleic Acids Res 25:217–221

Bennetzen JL, Hall BD (1982) Codon selection in yeast. J Biol Chem 257:3026–3031

Bissonnette L, Roy PH (1992) Characterization of In0 of *Pseudomonas aeruginosa* plasmid pVS1, an ancestor of integrons of multiresistant plasmids and transposons of Gram-negative bacteria. J Bacteriol 174:1248–1257

Boyd EF, Hill CW, Rich SM, Hartl DL (1996) Mosaic structure of plasmids from natural populations of *Escherichia coli.* Genetics 143:1091–1100

Brown NL, Ford SJ, Pridmore RD, Fritzinger DC (1983) Nucleotide sequence of a gene from the *Pseudomonas* transposon Tn*501* encoding mercuric reductase. Biochemistry 22:4089–4095

Brown NL, Misra TK, Winnie JN, Schmidt A, Seiff M, Silver S (1986) The nucleotide sequence of the mercuric resistance operons of plasmid R100 and transposon Tn*501*: further evidence for *mer* genes which enhance the activity of the mercuric ion detoxification system. Mol Gen Genet 202:143–151

Bruce KD, Osborn AM, Pearson AJ, Strike P, Ritchie DA (1995) Genetic diversity within *mer* genes directly amplified from communities of noncultivated soil and sediment bacteria. Mol Ecol 4:605–612

Bull PC, Cox DW (1994) Wilson disease and Menkes disease: new handles on heavy-metal transport. Trends Genet 10:246–252

Campbell A (1994) Comparative molecular biology of lambdoid phages. Annu Rev Microbiol 48:193–222

Cheng KC, Smith GR (1984) Recombinational hot spot activity of Chi-like sequences. J Molec Biol 180:371–377

Corpet F, Gouzy J, Kahn D (1998) The ProDom database of protein domain families. Nucleic Acid Res 26:323–326

de la Cruz F, Grinsted J (1982) Genetic and molecular characterization of Tn*21,* a multiple resistance transposon from R100.1. J Bacteriol 151:222–228

Devereux J, Haberti P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Gilbert MP, Summers AO (1988) The distribution and divergence of DNA sequences related to the Tn*21* and Tn*501 mer* operons. Plasmid 20:127–136

Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with expressivity. Nucleic Acids Res 10:7055–7074

Grantham R, Gautier C, Guoy M, Jacobzone M, Mercier R (1981)

Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 9:r43–r74

Griffin HG, Foster TJ, Silver S, Misra TK (1987) Cloning and DNA sequence of the mercuric and organomercurial-resistance determinates of plasmid pDU1358. Proc Natl Acad Sci USA 84:3112–3116

Grinsted J, de la Cruz F, Schmitt R (1990) The Tn*21* subgroup of bacterial transposable elements. Plasmid 24:163–189

Hamlett NV, Landale EC, Davis BH, Summers AO (1992) Roles of the Tn*21 merT, merP,* and *merC* gene products in mercury resistance and mercury binding. J Bacteriol 174:6377–6385

Hobman J, Kholodii G, Nikiforov V, Ritchie DA, Strike P, Yurieva O (1994) The sequence of the *mer* operon of pMER327/419 and transposon ends of pMER327/419, 330 and 05. Gene 277:73–78

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2:13–34

Inoue C, Sugawara K, Kusano T (1991) The *merR* regulatory gene in *Thiobacillus ferrooxidans* is spaced apart from the *mer* structural genes. Mol Microbiol 5:2707–2718

Inoue C, Sugawara K, Kusano T, Kitagawa Y (1989) Nucleotide sequence of the *Thiobacillus ferrooxidans* chromosomal gene encoding mercuric reductase. Gene 84:47–54

Jackson WJ, Summers AO (1982) Biochemical characterization of the $HgCl_2$-inducible polypeptides encoded by the *mer* operon of plasmid R100. J Bacteriol 151:962–970

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein evolution. Academic Press, New York, pp 21–132

Kholodii GY, Gorlenko OL, Lomovskaya SZ, Nikiforov VG (1993a) Molecular characterization of an aberrant mercury resistance transposable element from an environmental *Acinetobacter* strain. Plasmid 30:303–308

Kholodii GY, Yurieva OV, Lomovskaya OL, Gorlenko ZM, Mindlin SZ, Nikiforov VG (1993b) Tn*5053,* a mercury resistance transposon with integron ends. J Mol Biol 230:1103–1107

Kiyono M, Omura T, Inuzuka M, Fujimori H, Pan-Hou H (1997) Nucleotide sequence and expression of the organomercurial-resistance determinants from a Pseudomonas K-62 plasmid pMR26. Gene 189:151–157

Kusano T, Ji G, Inoue C, Silver S (1990) Constitutive synthesis of a transport function encoded by the *Thiobacillus ferrooxidans merC* gene cloned in *Escherichia coli.* J Bacteriol 172:2688–2692

Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 44:387–397

Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–1860

Liebert CA, Hall RM, Summers AO (1999) Transposon Tn*21,* flagship of the floating genome. Microbiol Mol Biol Rev 63:507–522

Liebert CA, Wireman J, Smith T, Summers AO (1997) Phylogeny of mercury resistance (*mer*) operons of Gram-negative bacteria isolated from the fecal flora of primates. Appl Environ Microbiol 63:1066–1076

Lloyd RG, Low KB (1996) Homologous recombination. In: Neidhardt FC (ed) *Escherichia coli* and *Salmonella,* cellular and molecular biology, 2 edn. ASM Press, Washington, DC, pp 2236–2255

Lynch M, Crease TJ (1990) The analysis of population survey data on DNA sequence variation. Mol Evol 34:126–129

Maynard Smith J (1992) Analysing the mosaic structure of genes. J Mol Evol 34:126–129

Medigue C, Rouxel T, Vigier P, Henaut A, Danchin A (1991) Evidence of horizontal gene transfer in *Escherichia coli* speciation. J Mol Biol 222:851–856

Milkman R, Bridges MM (1990) Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics 126:505–517

Miller SM, Massey V, Williams CH, Ballou DP, Walsh CT (1991) Communication between the active sites in dimeric mercuric ion reductase: an alternating sites hypothesis for catalysis. Biochemistry 30:2600–2612

Minakhina S, Kholodii G, Mindlin S, Yurieva O, Nikiforov V (1999) The Tn*5053* family transposons are *res* site hunters sensing plasmidal *res* sites occupied by cognate resolvases. Mol Microbiol 33:1059–1068

Misra TK, Brown NL, Fritzinger DC, Pridmore RD, Barnes WM, Haberstroh L, Silver S (1984) The mercuric-ion resistance operons of plasmid R100 and transposon Tn*501*: the beginning of the operon including the regulatory region and the first two structural genes. Proc Natl Acad Sci USA 81:5975–5979

Morby AP, Hobman JL, Brown NL (1995) The role of cysteine residues in the transport of mercuric ions by the Tn*501* MerT and MerP mercury-resistance proteins. Mol Microbiol 17:25–35

Mukhopadhyay D, Yu H, Nucifora G, Misra TK (1991) Purification and functional characterization of MerD: a coregulator of the mercury resistance operon in Gram-negative bacteria. J Biol Chem 266:18538–18542

Osborn AM, Bruce KD, Ritchie DA, Strike P (1996) The mercury resistance operon of the *IncJ* plasmid pMERPH exhibits structural and regulatory diversity from other Gram-negative *mer* operons. Microbiology 142:337–345

Osborn AM, Bruce KD, Strike KD, Ritchie DA (1995) Sequence conservation between regulatory mercury resistance genes in bacteria from mercury polluted and pristine environments. Syst Appl Microbiol 18:1–6

Peden J (1997) Correspondence of analysis of codon usage, Codon W, University of Nottingham

Peters S, Strike J, Ritchie D (1991) Novel mercury resistance determinants carried by *IncJ* plasmids pMERPH and R391. Mol Gen Genet 228:294–299

Reniero D, Mozzon E, Galli E, Barbieri P (1998) Two aberrant mercury resistance transposons in the *Pseudomonas stutzeri* plasmid pPB. Gene 208:37–42

Rost B, Casadio R, Fariselli P, Sander C (1995) Prediction of helical transmembrane segments at 95% accuracy. Protein Sci 4:521–533

Rost B, Sanders C (1994) Combining evolutionary information and neutral networks to predict protein secondary structure. Proteins 19:55–72

Rozas J, Rozas R (1997) DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput Appl Biosci 13:307–311

Sahlman L, Hagglof EM, Powlowski J (1999) Roles of the four cysteine residues in the function of the integral inner membrane Hg^{2+}-binding protein, MerC. Biochem Biophys Res Commun 255:307–311

Sahlman L, Wong W, Powlowski J (1997) A mercuric ion uptake role for the integral inner membrane protein, MerC, involved in bacterial mercuric ion resistance. J Biol Chem 272:29518–29526

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Schiering N, Kabsch W, Moore MJ, Distefano MD, Walsh CT, Pai EF (1991) Structure of the detoxification catalyst mercuric ion reductase from *Bacillus* sp. strain RC607. Nature 352:168–172

Sharp PM, Li WH (1987) The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15:1281–1295

Shen P, Huang HV (1986) Homologous recombination in *Escherichia coli:* dependence on substrate length and homology. Genetics 112: 441–457

Smith GR (1994) Hotspots of homologous recombination. Experientia 50:234–241

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc of Sixth Int Conf on Intelligent Systems for Molecular Biology, Menlo Park, CA, AAAI Press, pp 175–182

Steel RA, Opella SJ (1997) Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. Biochemistry 36:6885–6895

Summers AO (1986) Organization, expression and evolution of genes for mercury resistance. Annu Rev Microbiol 40:607–634

Summers AO (1992) Untwist and shout: a heavy metal-responsive transcriptional regulator. J Bacteriol 174:3097–3101

Swofford DL (1998) Phylogenetic Analysis Using Parsimony (PAUP), Version 4.0. Computer program distributed by Sinauer Assoc., Inc., Smithsonian Institution, Washington, DC.

Velasco A, Acebo P, Flores N, Perera J (1999) The *mer* operon of the acidophilic bacterium *Thiobacillus* T3.2 diverges from its *Thiobacillus ferrooxidans* counterpart. Extremophiles 3:35–43

Watt VM, Ingles CJ, Urdea MS, Rutter WJ (1985) Homology requirements for recombination in *Escherichia coli.* Proc Natl Acad Sci USA 82:4768–4772

Wilson JR, Leang C, Morby P, Hobman JL, Brown NL (2000) MerF is a mercury transport protein: different structures but a common mechanism for mercuric ion transporters? FEBS Let 472:78–82

Wireman J, Liebert CA, Smith CT, Summers AO (1997) Association of mercury resistance and antibiotic resistance in the Gram negative fecal bacteria of primates. Appl Environ Microbiol 63:4494–4503

Wright F (1990) The effective number of codons' used in a gene. Gene 87:23–29

Yoshifumi I, Watson JM, Haas D, Leisinger T (1984) Genetic and molecular characterization of the *Pseudomonas* plasmid pVS1. Plasmid 11:206–220

Yurieva O, Kholodii G, Minakhin L, Gorlenko Z, Kalyaeva E, Mindlin S, Nikiforov V (1997) Intercontinental spread of promiscuous mercury-resistance transposons in environmental bacteria. Mol Microbiol 24:321–329

Zeng Q., Stålhandske C, Anderson MC, Scott R, Summers AO (1998) The core metal-recognition domain of MerR. Biochemistry 37: 15885–15895