

New Findings on Evolution of Metal Homeostasis Genes: Evidence from Comparative Genome Analysis of Bacteria and Archaea†

J. M. Coombs* and T. Barkay

Department of Biochemistry and Microbiology, Cook College, Rutgers University, New Brunswick, New Jersey 08901

Received 30 December 2004/Accepted 22 June 2005

In order to examine the natural history of metal homeostasis genes in prokaryotes, open reading frames with homology to characterized P_{IB} -type ATPases from the genomes of 188 bacteria and 22 archaea were investigated. Major findings were as follows. First, a high diversity in N-terminal metal binding motifs was observed. These motifs were distributed throughout bacterial and archaeal lineages, suggesting multiple loss and acquisition events. Second, the CopA locus separated into two distinct phylogenetic clusters, CopA1, which contained ATPases with documented Cu(I) influx activity, and CopA2, which contained both efflux and influx transporters and spanned the entire diversity of the bacterial domain, suggesting that CopA2 is the ancestral locus. Finally, phylogenetic incongruences between 16S rRNA and P_{IB} -type ATPase gene trees identified at least 14 instances of lateral gene transfer (LGT) that had occurred among diverse microbes. Results from bootstrapped supported nodes indicated that (i) a majority of the transfers occurred among proteobacteria, most likely due to the phylogenetic relatedness of these organisms, and (ii) gram-positive bacteria with low moles percent G+C were often involved in instances of LGT. These results, together with our earlier work on the occurrence of LGT in subsurface bacteria (J. M. Coombs and T. Barkay, *Appl. Environ. Microbiol.* 70:1698–1707, 2004), indicate that LGT has had a minor role in the evolution of P_{IB} -type ATPases, unlike other genes that specify survival in metal-stressed environments. This study demonstrates how examination of a specific locus across microbial genomes can contribute to the understanding of phenotypes that are critical to the interactions of microbes with their environment.

The dissemination of toxic heavy metals through the environment is a natural part of geochemical cycling on earth. Microbes from many different environments have had to adapt to soils containing heavy metal ores, to metal-rich plumes from hydrothermal vent fluids (15, 54), and to environments affected by atmospheric deposition of metals resulting from geothermal activities such as volcanic eruptions (19, 52). This has resulted in the development or acquisition of genetic systems to transform, sequester, or remove these harmful elements to mitigate their toxic effects. Some of these metals, such as Zn(II) and Cu(I), are toxic at high levels but are required at low levels for cell function (12, 38), necessitating strict regulatory mechanisms to maintain metal homeostasis in the cell cytoplasm.

The evolution of metal homeostasis genes has significant implications for understanding microbial adaptations in environments that are subject to change. Many factors influence the survival of organisms exposed to toxic levels of heavy metals, including lateral gene transfer (LGT) for the dissemination of resistance phenotypes throughout microbial communities (10, 35) and changes in active-site residues that affect substrate specificity of metal homeostasis proteins (53). Examining patterns of mutation and dissemination within a collection of microorganisms not only reveals insights into the history of metal resistance genes but also provides the background for predictions about how their gene products mediate microbial

survival in metal-affected environments. In the case of sites contaminated with metals from mining or industrial wastes, such studies may lead to more informed bioremediation strategies and better environmental management.

P-type ATPases are a large family of transmembrane transporters that are responsible for carrying ions and small organic molecules into or out of the cell. Phylogenetically, they separate into distinguishable families on the basis of their substrate specificity (2). The subfamily that regulates the transport of heavy metals, the P_{IB} -type ATPases, export those which are toxic and regulate the concentration of those which are essential through influx to, or efflux from, the cell cytoplasm. Proteins in this family may be specific for monovalent [Cu(I)/Ag(I)] or divalent [Zn(II)/Cd(II)/Pb(II)/Co(II)] ions or, in the case of *Oscillatoria brevis*, for both (53). However, a single sequence motif that governs specificity has not yet been identified. Therefore, although these molecules tend to separate into two phylogenetically related groups (42), the lack of a single hallmark motif identifying them as belonging to one group or the other means that the true transport activities of homologues should be identified through additional biochemical testing. P_{IB} -type ATPase activities have been characterized in several organisms of environmental relevance, such as bacilli (48), pseudomonads (26), *Ralstonia* spp. (8, 31), and cyanobacteria (49). Genes encoding P_{IB} -type ATPases are found in the majority of sequenced bacterial and archaeal genomes (this study).

One of the most significant insights emerging from prokaryotic genomes is the degree of LGT that has occurred during prokaryotic evolution (17). Although it has been known for several years that, in the reality of low frequencies of recombination, LGT plays a significant role in promoting and main-

* Corresponding author. Present address: Department of Biology, Adelphi University, 110 Science Building, 1 South Ave., Garden City, NY 11530. Phone: (516) 877-4196. Fax: (516) 877-4209. E-mail: coombs@adelphi.edu.

† Supplemental material for this article may be found at <http://aem.asm.org/>.

taining diversity among prokaryotes (27), the scope of transfer is only now becoming clear. While much is known about laterally transferred pathogenicity islands such as those characterized in *Vibrio cholerae* and other pathogenic organisms (20), and while extensive attention has been paid to LGT of elements associated with antibiotic resistance determinants (13), little is known about the impact of LGT on metal resistance genes in full genome sequences. The availability of sequenced genomes provides an opportunity for the comprehensive examination of the role of LGT in the evolution of metal resistance among prokaryotes.

We have examined 465 deduced amino acid sequences of hypothetical P_{IB} -type ATPase genes from the annotated genomes of 188 bacteria and 22 archaea. This analysis has revealed instances of LGT involved in the evolution of this locus, highlighted patterns of diversification in conserved functional domains, and suggested possible functions for specific clades within the P_{IB} -type ATPase phylogeny.

MATERIALS AND METHODS

Querying of completed genome databases. P_{IB} -type ATPases from *Escherichia coli* (ZntA; accession no. P37617), *Staphylococcus aureus* (CadA from plasmid pI258; accession no. AAB59154), and *Ralstonia metallidurans* (PbrA; accession no. CAI11271) were used in BLAST searches of 220 completed and annotated microbial genomes from the collection available at NCBI (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). Sequences of the resulting BLAST hits were scanned for the presence of P-type ATPase-specific signatures: the phosphatase motif (TGES), the phosphorylation motif (DKTGT), the ATP binding motif (GDGIND), and the transmembrane metal binding motif that distinguishes P_{IB} -type ATPases from other P-type ATPases (CPCAL) (2). Molecules that contained synonymous or nonsynonymous mutations in these sites were also selected, and a total of 433 bacterial and 32 archaeal molecules were downloaded into a local database. The small-subunit 16S rRNA gene sequences were obtained from the structural RNA list under the feature tables of each genome, or by a BLASTN search of specific genomes for which this feature was not available (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). A list of the P_{IB} -type ATPases and 16S loci that were analyzed in this study is provided in Table S1 in the supplemental material.

Phylogenetic and sequence analysis. Sequence data from the P_{IB} -type ATPases were examined for functionally conserved motifs whose activity had been documented (the phosphatase domain TGES [5], TGEP [41], SGEs [22], SGEA [4], and TGET [37] and the transmembrane metal binding motif [CPCAL [5], CPCGL [53], SPCAL [45], and CPHAL [47]). When P_{IB} -type ATPases from two or more related organisms were >95% similar, a single representative sequence was selected for tree building. Sequence data from 16S rRNA genes and the deduced amino acid sequences of putative P_{IB} -type ATPases were aligned in ClustalX (50) using default program settings, followed by verification of the alignment by eye. Construction of unrooted trees was performed using the parsimony and distance functions of PAUP* v4.0b10 (Sinaur and Associates, Sunderland, MA) for the construction of heuristic and neighbor-joining trees, respectively. Bootstrap analyses were performed in PAUP* v4.0b10. A subset of 296 P_{IB} -type ATPase sequences were aligned and used to create a phylogenetic tree, (see Fig. S1 in the supplemental material). Smaller trees indicating relevant transfer events were also created (see Results).

The moles percent G+C content of each P_{IB} -type ATPase was calculated using the nucleotide composition feature of BioEdit (21). The moles percent G+C contents of individual P_{IB} -type ATPase sequences were then compared to the moles percent G+C contents of the genomes from which the sequence originated (HGT-DB [16]). Standard deviations from the mean moles percent G+C content were also obtained from this source. For genomes for which this information was not available, moles percent G+C content was obtained from the Codon Usage Database (31), and the ranges were verified against values published in the online manual "The Prokaryotes" (<http://141.150.157.117:8080/prokPUB/index.html>) when available. Analysis of insertion/deletion events (indels) was performed visually by observing ClustalW (51) alignments created in BioEdit (21). Only indels of shared length present in more than one organism which were bracketed by conserved sequences on either side were accepted in the analysis.

Analysis of flanking regulatory genes. Genes encoding regulatory proteins within 5 kilobases of P_{IB} -type ATPases, which therefore might be cotranscribed, were identified through BLAST searches using either an ArsR-like sequence (CadC from *S. aureus* [33]) or two MerR-like sequences (ZntR and PbrR from *E. coli* [8] and *R. metallidurans* [7], respectively) from the GenBank database (<http://www.ncbi.nlm.nih.gov:80/BLAST/>). These were downloaded into a local database (Vector NTI; Invitrogen Corporation), analyzed in AlignX (Vector NTI; Invitrogen Corporation), and subjected to phylogenetic analysis as described above (see Fig. S2 in the supplemental material). Regulatory gene phylogenies were compared to the expected 16S phylogeny as well as to the phylogeny obtained with P_{IB} -type ATPases by manual observation.

RESULTS

Sequence analysis and motif identification of P_{IB} -type ATPases. A molecular approach was used to elucidate the evolutionary history of P_{IB} -type ATPases from completed and annotated genomes. Of 188 bacterial and 22 archaeal genomes surveyed for the presence of P_{IB} -type ATPases, 40 bacterial and four archaeal genomes lacked open reading frames for this locus. The bacterial genomes lacking P_{IB} -type ATPases all belonged to parasites of animals or plants, whose genomes tend to be more streamlined than those of free-living organisms, while the archaeal genomes lacking this locus included representatives from diverse groups. The other genomes examined in this study contained as few as one (e.g., *S. aureus*, accession no. NC_003923) or as many as seven (e.g., *Sinorhizobium meliloti*, accession no. NC_003047) different open reading frames with homology to P_{IB} -type ATPases. Sequences shared from 95% to 9.7% pairwise sequence identity, indicating that this locus is subject to a large amount of genetic variation.

Close examination of the N terminus, a region of the sequence that contained a high degree of variability, revealed diversity and a lack of phylogenetic coherence in metal binding motifs (Fig. 1). The best characterized of these (consensus sequence CXXC) was found in one or more copies in the N termini of 60% of the bacterial P_{IB} -type ATPases analyzed (Fig. 1A). Thirteen percent of the sequences had a histidine-rich metal binding domain (Fig. 1B), 4% of which contained both a histidine-rich region and a CXXC motif. An obvious N-terminal metal binding motif was not recognized in 27% of the cases (Fig. 1C). In the crenarchaeota, five sequences contained the CXXC motif, while one had no N-terminal metal binding domain. The euryarchaeota, like the bacteria, varied greatly in the phylogenetic distribution of different N-terminal motifs, with some sequences containing examples of more than one type. We suggest that this variability in N-terminal motifs and their random distribution among bacteria and archaea are evidence for a mosaic nature of the gene, indicating that this metal binding domain has been independently lost and regained over time in several lineages.

Multiple-sequence alignments in preparation for phylogenetic analysis revealed that several of the P_{IB} -type ATPase homologues contained mutations in consensus active-site sequences such as the phosphorylation (TGES) motif and the metal translocation (CPCAL) motif. In archaeal sequences, synonymous mutations were common and often involved substitutions of other nonpolar amino acids in the fourth and fifth positions of the CPCAL motif. Another common substitution was the replacement of the second cysteine residue in the motif with a histidine. A P_{IB} -type ATPase from *Enterococcus hirae*

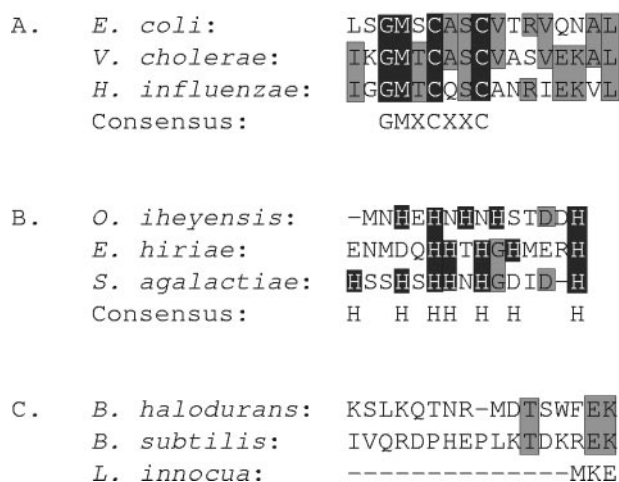


FIG. 1. Amino acid sequence alignments showing the different conserved regions of the N termini of P_{IB}-type ATPases. Amino acids specific to the motif are boxed in black, while amino acids that align but are not specific to the motif are boxed in gray. (A) Alignment of sequences from selected *Gammaproteobacteria* demonstrating the consensus metal binding motif (GMXCXXC). (B) Alignment of sequences from selected gram-positive bacteria demonstrating the histidine-rich metal binding motif. (C) Alignment of sequences from selected gram-positive bacteria showing lack of any known metal binding motif.

(B45995) that has this motif was demonstrated to export copper from the cell (34); therefore, this substitution is unlikely to disrupt the metal-transporting function of the protein. Interestingly, five euryarchaeal sequences that contained a CPHAL motif all had a histidine-rich N terminus (see above). This correspondence between modifications of the metal-translocating and binding motifs was observed for only five of the histidine-rich bacterial sequences in our data set.

ATPase sequences from bacterial genomes contained 110 instances, and archaeal sequences contained 8 instances, where the CPCAL metal translocation motif was altered by nonsynonymous amino acid replacements. The most common was a change in the first cysteine residue to a serine. A P_{IB}-type ATPase with this substitution from *Synechocystis* sp. strain PCC6803 3 (accession no. NP_442633) has been implicated in cobalt resistance (45), indicating that substituted motifs retain transport activity. Two other motifs, the phosphorylation domain (DKTGT) and the ATP binding domain (GDGXND), were highly conserved in all sequences.

The substrate specificity of P_{IB}-type ATPase sequences is reflected in their phylogeny. In agreement with previous reports of the coherence between the substrate specificity and phylogeny of P-type ATPases (2), a preliminary phylogenetic analysis distinguished those sequences that clustered with CopA or ZntA proteins, which had previously been characterized as mono- and divalent cation transporters, respectively (see Fig. S1 in the supplemental material). These groups were used for the construction of ZntA-like (Fig. 2) and CopA-like (Fig. 3) P_{IB}-type ATPase phylogenies. Archaeal P_{IB}-type ATPases have not yet been characterized with respect to their substrate specificity but are labeled (Fig. 4A) on the basis of similarity to bacterial sequences of known specificity (data not shown).

Phylogenetic analysis of the CopA-like sequences revealed two clusters, CopA1 and CopA2, whose separation was supported by bootstrap analysis (Fig. 3). The smaller of the two groups, the CopA1 group, contains predominantly proteobacterial sequences. The larger of the groups, the CopA2 group, contains genes spanning the entire diversity of the bacterial domain, including those from deep-branching bacteria such as *Thermotoga*.

Evidence for LGT among bacterial P_{IB}-type ATPases. Phylogenetic incongruence, which is suggested when the phylogeny of a gene of interest differs from the phylogeny of a marker gene such as the small-subunit 16S rRNA gene, is a common method for implicating LGT during gene evolution (14, 39). It is often used as a stand-alone method for detecting LGT, but it can be used in conjunction with other markers such as atypical sequence composition (moles percent G+C content [30]), patterns of insertion/deletion events (shared indels [18]), and the presence of adjacent genes on an operon that demonstrate similar patterns of phylogenetic incongruence (such as those found on pathogenicity islands [20]). We first used phylogenetic incongruence to look for evidence of LGT of the P_{IB}-type ATPase locus (Fig. 2 to 4). However, since all methods of detecting LGT by molecular signature contain uncertainties, and therefore some degree of error, we then tested the results with other approaches (Table 1). Instances in which phylogenetic incongruence and one or more of these alternative methods indicate gene transfer are considered strong evidence in support of an LGT event.

Examination of the trees in Fig. 2 and 3 reveals 11 instances of possible LGT at bootstrap-supported nodes. In the ZntA-like tree (Fig. 2), four of the five instances of phylogenetic incongruence are indicative of transfer involving *Gammaproteobacteria*. The position of two of these sequences deep within a betaproteobacterial clade, *Azotobacter vinelandii* 1 (accession no. ZP_00088927) and *Pseudomonas aeruginosa* 3 (NP_252380), suggests that the host genomes recently acquired these genes from *Betaproteobacteria*. These transfers are supported by complementary data (Table 1). The *Gamma*- and *Betaproteobacteria* are the most recently diverged of the bacterial lineages and therefore are the groups most likely to participate in gene exchange. The other two transfers involving *Gammaproteobacteria* suggest exchange across a greater phylogenetic distance, as the sequences for both *Legionella pneumophila* 2 (YP_095043) and *Stenotrophomonas maltophilia* (AJ251015 [1]) cluster with high bootstrap support with the low-moles-percent G+C gram-positive clade. The *S. maltophilia* transfer is supported by both the low moles percent G+C content of the gene and evidence of transfer of a proximal regulatory gene, encoding a CadC-like regulator (Table 1). Since the phylogeny of the regulatory gene demonstrates the same pattern of phylogenetic incongruence as the P_{IB}-type ATPase, it indirectly supports the hypothesis that a transfer event involving this gene cluster has occurred. The regulator itself also demonstrated an unusual moles percent G+C content and was most similar to the regulatory proteins of organisms with low-moles-percent G+C content, such as *Bacillus halodurans* (accession no. NP_244904). Further support for transfer of the *L. pneumophila* sequence was not detected (Table 1).

Phylogenetic evidence presented in the CopA tree (Fig. 3) indicates that LGT of genes from *Gammaproteobacteria* may have played a significant role in the evolution of the CopA2

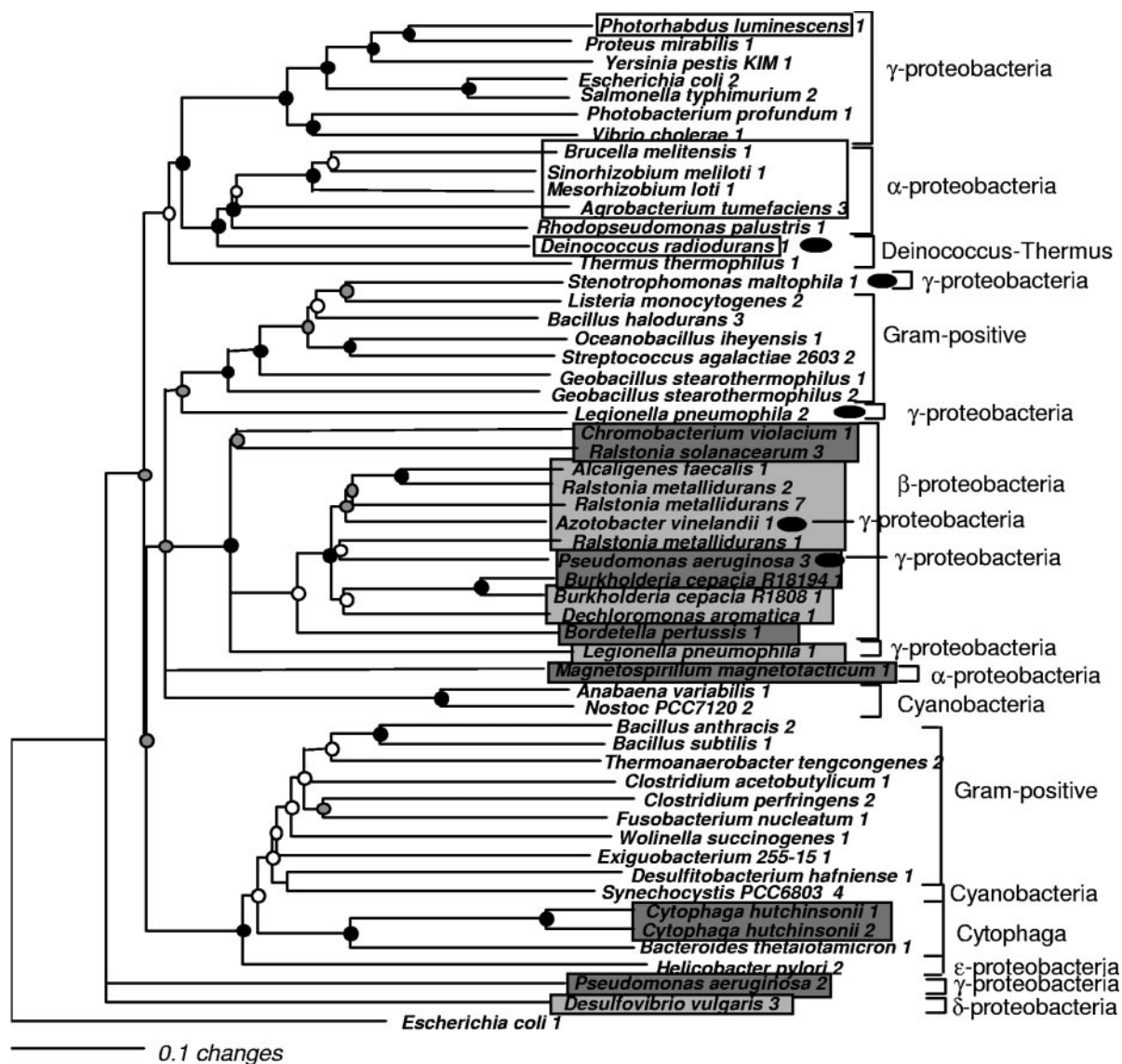


FIG. 2. Phylogeny of ZntA-like sequences from bacterial genomes. Neighbor-joining analysis of sequences from completed bacterial genomes that were identified by their homology to ZntA-like P_{IB} -type ATPases is shown. Circles at each node indicate the level of bootstrap support obtained with both parsimony and distance methods: black, >80%; gray, >50%; white, supported at >50% by one method only. Dark and light gray boxes indicate the presence of a histidine-rich metal binding motif or no metal binding motif at the N terminus, respectively. White boxes indicate both CXXC and histidine-rich motifs. Black ovals mark sequences that demonstrate phylogenetic incongruence compared to the 16S rRNA phylogeny.

locus as well. The CopA/SilA-like P_{IB} -type ATPase 4 (accession no. NP_252380) from the gammaproteobacterium *P. aeruginosa* and the ATPase from *Microbulbifer degredans* 3 (ZP_00317424) cluster with a high degree of significance with a group of sequences from the betaproteobacterial clade. Neither of these transfer events is supported by additional evidence (Table 1). A different LGT event, represented on the tree by two genera of *Gammaproteobacteria* clustering with high bootstrap support within a clade of *Alphaproteobacteria*, appears to have occurred prior to the divergence of *Salmonella enterica* serovar Typhimurium and *A. vinelandii*. The LGT of the *A. vinelandii* gene is supported by the presence of a shared indel (Table 1). For *S. enterica* serovar Typhimurium, gene

transfer is supported by the proximity of a gene coding for an HmrR-like regulatory protein, which is similar in sequence to an HmrR protein from *Sinorhizobium meliloti* that has been shown to regulate transcription of the CopA/SilA-like locus ActP (40). In a phylogenetic tree of HmrR-like sequences that are located upstream of CopA/SilA sequences on the genomes of *Alpha*- and *Gammaproteobacteria*, the HmrR-like sequence from *S. enterica* serovar Typhimurium clusters with the *Alphaproteobacteria* (see Fig. S2 in the supplemental material), suggesting acquisition of both HmrR-like and CopA2-like genes in the same LGT event. If so, the HmrR regulator must have been lost in the *A. vinelandii* genome. Interestingly, the absence of a metal binding motif in the N terminus of



FIG. 3. Phylogeny of CopA-like sequences from bacterial genomes. These sequences were identified on the basis of sequence similarity to CopA-like P_{IB}-type ATPases. See the Fig. 2 legend for details.

A. vinelandii 5 (accession no. ZP_00088950) and the presence of the GMXCXXC motif in *S. enterica* serovar Typhimurium 3 (NP_459348) suggest alterations in this motif following the transfer event, supporting our conclusion of multiple loss and acquisition events in the evolution of the N-terminal domain during the evolution of P_{IB}-type ATPases (see above). Finally, a CopA2 sequence from a gram-positive organism, *Ureaplasma parvum* (NP_078035), clusters within a clade of *Epsilonproteobacteria* (no support by complementary data [Table 1]). This provides further evidence of the involvement of gram-positive organisms in LGT.

In *Deinococcus radiodurans*, LGT of P_{IB}-type ATPases may have occurred on at least two separate occasions, for the

D. radiodurans ZntA sequence 1 (NP_285396) (Fig. 2) and Cop2A sequence 3 (NP_296173) (Fig. 3), both of which appear to involve genes from *Alphaproteobacteria*. These two genes are not in close proximity to each other in the genomes of either of the analyzed *Alphaproteobacteria* or *D. radiodurans*, which might have indicated cotransfer of these loci. The *zntA* locus has a G+C content similar to that of the *Deinococcus* host (67%), while the *copA*-like gene has a lower moles percent G+C content, closer to the mean values for the genomes of the *Alphaproteobacteria* *Mesorhizobium loti* and *S. meliloti* (62.7% and 62.7%, respectively), supporting evidence for two separate LGT events, although movement of the genes within the genome combined with differing rates of amelioration cannot be ruled out at this time.

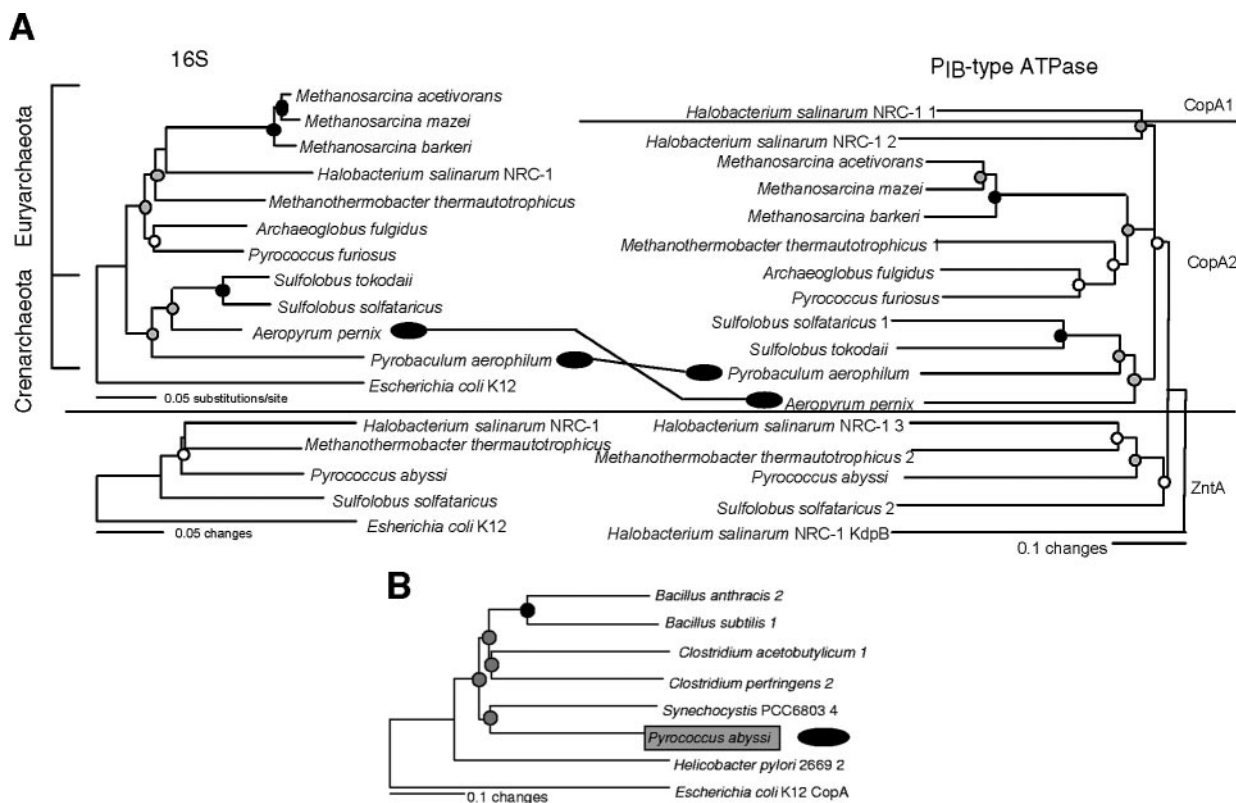


FIG. 4. Phylogeny of ZntA-like and CopA-like sequences from archaeal genomes. (A) Neighbor-joining analysis of 16S rRNA gene (left) and P_{1B} -type ATPase (right) sequences. Circles at each node indicate the level of bootstrap support obtained when analysis was by both parsimony and distance methods (see the legend to Fig. 2). Instances of incongruence between the two trees are indicated by black ovals. (B) Neighbor-joining analysis of select ZntA-like P_{1B} -type ATPases belonging to both the bacterial and archaeal domains. Evidence supporting a single interdomain gene transfer is boxed and marked with a black oval.

Evidence for LGT among archaea and between domains. Phylogenetic incongruence between two crenarchaeota, *Aeropyrum pernix* 1 (accession no. NP_147955) and *Pyrobaculum aerophilum* (NP_560235), was identified (Fig. 4A). LGT involv-

ing *P. aerophilum* is supported by evidence from shared indels and moles percent G+C analysis, evidence which is not present for the *A. pernix* gene (Table 1). When analyzing sequences from both archaea and bacteria, evidence for interdomain transfer emerged (Fig. 4B). The ZntA-like locus from *Pyrococcus abyssi* (NP_126615) clustered with a cyanobacterial sequence, *Synechocystis* sp. strain PCC6803 sequence 4 (NP_442636), at a bootstrap-supported node on the ZntA tree. This sequence is not shared by its close relatives *Pyrococcus furiosus* and *Pyrococcus horikoshii*, indicating either that acquisition occurred after speciation or that it was lost in the other two lineages. Lateral transfer of this gene is supported (Table 1) by the presence of a shared indel and an adjacent sequence that clusters phylogenetically with SmtB-like regulatory sequences. SmtB is a regulator of *ziaA*, a *zntA*-like gene in cyanobacteria (49), and its proximity to *zntA* in the *P. abyssi* genome suggests cotransfer of the two from a cyanobacterial donor. The fact that only one transfer between bacteria and archaea was detected in this study may be due to the small size of the archaeal data set, because interdomain transfer between microorganisms appears to be a common event (24, 32).

DISCUSSION

The large-scale sequencing of microbial genomes is one of the most dramatic developments in microbiology in recent years. Data emerging from these projects provide information on the genetic potential of microorganisms and their metabolic

TABLE 1. Supporting molecular evidence for LGT of ZntA-like and CopA-like sequences

Protein	Organism and sequence	LGT supported by:		
		G+C content ^a	Proximal regulatory gene ^b	Shared indel
ZntA	<i>Deinococcus radiodurans</i> 1	No	No	No
	<i>Stenotrophomonas maltophilia</i> 1	Yes	Yes (<i>cadC</i> -like)	No
	<i>Legionella pneumophila</i> 2	No	No	No
	<i>Azotobacter vinelandii</i> 1	No	No	No
	<i>Pseudomonas aeruginosa</i> 3	No	No	No
	<i>Pyrococcus abyssi</i>	No	Yes (<i>smtB</i> -like)	Yes
CopA	<i>Salmonella enterica</i> serovar Typhimurium 3	No	Yes (<i>hmrR</i> -like)	No
	<i>Deinococcus radiodurans</i> 3	Yes	No	No
	<i>Pseudomonas aeruginosa</i> 4	No	No	No
	<i>Ureaplasma parvum</i>	No	No	No
	<i>Azotobacter vinelandii</i> 3	No	No	Yes
	<i>Microbulbifer flagellatus</i> 1	No	No	No
	<i>Pyrobaculum aerophilum</i>	Yes	No	Yes
	<i>Aeropyrum pernix</i> 1	No	No	No

^a Unusual moles percent G+C content is defined here as that which is greater than one standard deviation from the Mean for the genome.

^b Proximal is defined here as a gene that is within 5 kilobases of a gene encoding a P_{1B} -type ATPase.

capacities, biochemical pathways, regulatory circuits, and means for surviving environmental stress. A comparison of information from multiple genomes allows analysis of gene structure and organization that would have been difficult or impossible in the pregenomic era. This information enables us to set up testable hypotheses for how environmentally important genes contribute to microbial survival, enrich community diversity, and affect microbial evolution.

Our analyses indicate that P_{IB}-type ATPases are widespread but not ubiquitous in microbial genomes. This is an expected finding, since others have looked at the entire P-type ATPase gene family and found it to be present in most but not all microbial genomes (<http://biobase.dk/~axe/Pathbase.html>). Microbes often differed with respect to the number of P_{IB}-type ATPases per genome, with the highest numbers generally found among actinobacteria, cyanobacteria, and *Alpha*- or *Beta*-*proteobacteria*. Most P_{IB}-type ATPases from the same genome were phylogenetically distinct, lacking the high degree of sequence identity that would indicate recent gene duplication. Sequence comparison with groups of biochemically characterized metal transporters indicated that at least some of these gene variants may have novel substrate specificities (see below).

The N-terminal domains of P_{IB}-type ATPases are diverse in structure, suggesting that multiple mechanisms for metal binding exist. Evidence from deletion studies with an *E. coli* ZntA mutant (29) and a *Listeria monocytogenes* CadA (3) mutant indicate that N-terminal metal binding domains are not essential for metal-translocating activities, and thus, divergence from a consensus sequence might have been tolerated during evolution of this locus. Further analysis as to how these N-terminal motifs affect the activities of P_{IB}-type ATPases may be expanded upon in the future through chimera construction and site-directed mutagenesis studies.

Variations on the N-terminal metal binding motif corresponded in many cases to substitutions within the metal translocation motif. In all archaeal sequences, proteins with histidine-rich N termini also contained a CPHAL rather than the consensus CPCAL translocation motif on transmembrane helix six. This suggests that there are some constraints on how the two domains interact. Where the N-terminal metal binding domain is absent, the transmembrane metal binding motif deviated from the classical CPCAL motif 71% of the time.

The criteria for inclusion in the P_{IB}-type ATPase data set included phosphatase (TGES) and metal translocation (CPCAL) motifs. The resulting trees contained clades separating sequences classified as ZntA-like (Fig. 2) from those classified as CopA-like (Fig. 3). The CopA-like group was split into two separate bootstrap-supported clusters, CopA1 and CopA2 (Fig. 3). The CopA1 clade contained sequences from *Alpha*-, *Beta*-, and *Gammaproteobacteria*. One of the loci belonging to this group, FixI, supplies Cu(I) for the biosynthesis of a copper-heme oxidase (cytochrome *c* oxidase) in *Bradyrhizobium japonicum* (37) and therefore serves as an influx rather than an efflux pump. With the exception of *Neisseria meningitidis* (accession no. NP_274076) and *chromobacterium violaceum* (accession no. NP_900935), *copA1* genes in all bacterial genomes are located adjacent to genes encoding the subunits of copper-heme oxidases. Thus, loci belonging to the CopA1 cluster may represent P_{IB}-type ATPases that function in ensuring the sup-

ply of monovalent cations that is essential for biosynthesis of metal-dependent enzymes. Members of the CopA2 group may be either influx (*Synechocystis* sp. strain PCC6803 CtaA [NP_441938] and PacS [NP_440588] and *E. hirae* CopB [B45995]) or efflux (*Helicobacter pylori* copA [4], *E. hirae* CopA [47], and *E. coli* CopA [41]) transporters. The CopA2 cluster contains genes from a much broader diversity of microorganisms than the CopA1 group (Fig. 3). Metal homeostasis must have been functioning early in microbial evolution, especially if life arose in the metal-rich waters of geothermal environments (43). Therefore, the CopA2 group may reflect the protein's ancestral state, perhaps coevolving with known metal influx proteins such as the Nramp homologue MntH and the ABC transporter ZnuABC (6). If so, the subsequent evolution of CopA1 may represent a specialization in Cu(I) influx.

Phylogenetic analysis of chromosomally located P_{IB}-type ATPases reveals 14 instances of incongruence with respect to the 16S rRNA gene among bacteria and archaea, indicating evolution by LGT (Fig. 2 to 4). Of these, six instances are supported by additional evidence (Table 1). Although LGT has played a definite role in the evolution of this locus, it is rare rather than rampant, and most inheritance in microorganisms appears to be vertical. This is surprising in light of the large scope of LGT in the evolution of other genes that facilitate survival in hostile environments, such as antibiotic (13) and mercury (35) resistance genes. Two explanations may account for this discrepancy. First, most P_{IB}-type ATPase gene products serve as "housekeeping" genes for metal homeostasis and therefore may be less prone to transfer (44). Second, the *copA/zntA* gene products must be successfully folded and inserted into the cell membrane in order to function. According to the complexity hypothesis (23), the more interactions a protein participates in, the less likely the gene encoding this protein is to be successfully integrated into and maintained in the genome of a heterologous host. Thus, the interactions required for the activity of the metal transporter may favor vertical over horizontal inheritance of *zntA/copA* loci.

Of the transfers that have occurred, many involve *Gammaproteobacteria* exchanging genes with other *Proteobacteria*. This may be due in part to bias in the analyzed data set, since the initial filtering based on known active consensus sequences resulted in *Proteobacteria* comprising one of the largest single groups (152 sequences out of 296). Recent evidence shows that organisms that are closely related phylogenetically are more likely to exchange genes than less related organisms, due to the ease of homologous recombination (28, 46) and to structural characteristics of the genome (25). However, these transfers are the least likely to be detected, due to the high overall similarity of molecules from closely related organisms, suggesting that we may have underestimated the frequency of LGT. Transfer also appears to have played a role in the evolution of this locus in low-G+C gram-positive organisms. Mobile elements often have a low moles percent G+C content, which may aid homologous recombination and thus contribute to the mobilization of these genes (11).

An interdomain transfer event, where *P. abyssi* acquired a *zntA*-like gene from a cyanobacterium, is supported by an apparent cotransfer of an adjacent regulatory gene for this locus (Table 1). Operon structure is less conserved in archaea than in bacteria, and therefore the conservation of gene order in

this segment of DNA may further support a bacterial origin. Interdomain transfer is known to occur and has been detected in several instances concerning other genes, such as *gltB* (32). However, photosynthetic cyanobacteria and the hyperthermophile *P. abyssi* occupy mutually exclusive ecological niches, suggesting that transfer may have occurred between niche-sharing ancestors or via a third population that could have occupied both habitats. Several species of thermophilic cyanobacteria (36) and mesophilic crenarchaeota (9) have been found, providing evidence that relatives may have shared a habitat in the past.

Analyzing metal homeostasis loci through the "latitudinal mining" of genomes enables examination of patterns of gene occurrence and base substitutions in conserved domains and the creation of complex phylogenies, where gene history and influences on the evolution of a specific locus are observed. Results of the analyses allow the formulation of hypotheses that can then be tested in the laboratory through biochemical characterization and in the environment through community analyses and field studies, to examine how a specific gene contributes to the activity and fitness of organisms and entire communities in challenging environments.

ACKNOWLEDGMENTS

We thank Christopher Rensing and three anonymous reviewers, whose comments and suggestions improved this paper.

This research was funded by the Natural and Accelerated Bioremediation Research (NABIR) program, Biological and Environmental Research (BER), U.S. Department of Energy (grant DE-FG02-99ER62864).

REFERENCES

- Alonso, A., P. Sanchez, and J. L. Martinez. 2000. *Stenotrophomonas maltophilia* D457R contains a cluster of genes from gram-positive bacteria involved in antibiotic and heavy metal resistance. *Antimicrob. Agents Chemother.* **44**:1778–1782.
- Axelsen, K. B., and M. G. Palmgren. 1998. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol.* **46**:84–101.
- Bal, N., E. Mintz, F. Guillaumin, and P. Catty. 2001. A possible regulatory role for the metal-binding domain of CadA, the *Listeria monocytogenes* Cd²⁺-ATPase. *FEBS Lett.* **506**:249–252.
- Bayle, D., S. Wängler, T. Weitzenegger, W. Steinhilber, J. Volz, M. Przybylski, K. P. Schäfer, G. Sachs, and K. Melchers. 1998. Properties of the P-type ATPases encoded by the *copAP* operons of *Helicobacter pylori* and *Helicobacter felis*. *J. Bacteriol.* **180**:317–329.
- Beard, S. J., R. Hashim, J. Membrillo-Hernandez, M. N. Hughes, and R. K. Poole. 1997. Zinc(II) tolerance in *Escherichia coli* K-12: evidence that the *zntA* gene (*o732*) encodes a cation transport ATPase. *Mol. Microbiol.* **25**:883–891.
- Blencowe, D. K., and A. P. Morby. 2003. Zn(II) metabolism in prokaryotes. *FEMS Microbiol. Rev.* **27**:291–311.
- Borremans, B., J. L. Hobman, A. Provoost, N. L. Brown, and D. van der Lelie. 2001. Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. *J. Bacteriol.* **183**:5651–5658.
- Brocklehurst, K. R., J. L. Hobman, B. Lawley, L. Blank, S. J. Marshall, N. L. Brown, and A. P. Morby. 1999. ZntR is a Zn(II)-responsive MerR-like transcriptional regulator of *zntA* in *Escherichia coli*. *Mol. Microbiol.* **31**:893–902.
- Burns, B. P., F. Goh, M. Allen, and B. A. Neilan. 2004. Microbial diversity of extant stromatolites in the hypersaline marine environment of Shark Bay, Australia. *Environ. Microbiol.* **6**:1096–1101.
- Coombs, J. M., and T. Barkay. 2004. Molecular evidence for the evolution of metal homeostasis genes by lateral gene transfer in bacteria from the deep terrestrial subsurface. *Appl. Environ. Microbiol.* **70**:1698–1707.
- Daubin, V., E. Lerat, and G. Perriere. 2003. The source of laterally transferred genes in bacterial genomes. *Genome Biol.* **4**:R57.1–R57.12.
- Duxbury, T. 1986. Microbes and heavy metals: an ecological overview. *Microbiol. Sci.* **3**:330–333.
- Dzidic, S., and V. Bedekovic. 2003. Horizontal gene transfer-emerging multidrug resistance in hospital bacteria. *Acta Pharmacol. Sin.* **24**:519–526.
- Eisen, J. A. 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* **8**:163–167.
- Elderfield, H., and A. Schultz. 1996. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**:191–224.
- Garcia-Valve, S., E. Guzman, M. A. Montero, and A. Romeu. 2003. HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res.* **31**:187–189.
- Gogarten, J. P., W. F. Doolittle, and J. G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**:2226–2238.
- Gupta, R. S. 2001. The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *Int. Microbiol.* **4**:187–202.
- Gustin, M. S. 2003. Are mercury emissions from geologic sources significant? A status report. *Sci. Total Environ.* **304**:153–167.
- Hacker, J., G. Blum-Oehler, B. Hochhut, and U. Dobrindt. 2003. The molecular basis of infectious diseases: pathogenicity islands and other mobile genetic elements. *Acta Microbiol. Immunol. Hung.* **50**:321–330.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**:95–98.
- Herrmann, L., D. Schwan, R. Garner, H. L. Mobley, R. Haas, K. P. Schafer, and K. Melchers. 1999. *Helicobacter pylori* *cadA* encodes an essential Cd(II)-Zn(II)-Co(II) resistance factor influencing urease activity. *Mol. Microbiol.* **33**:524–536.
- Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Natl. Acad. Sci. USA* **96**:3801–3806.
- Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake. 2002. Horizontal gene transfer in microbial genome evolution. *Theor. Pop. Biol.* **61**:489–495.
- Lawrence, J. G., and H. Hendrickson. 2003. Lateral gene transfer: when will the adolescence end? *Mol. Microbiol.* **50**:739–749.
- Lee, S.-W., E. Glickmann, and D. A. Cooksey. 2001. Chromosomal locus for cadmium resistance in *Pseudomonas putida* consisting of a cadmium-transporting ATPase and a MerR family response regulator. *Appl. Environ. Microbiol.* **67**:1437–1444.
- Levin, B. R., and C. T. Bergstrom. 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl. Acad. Sci. USA* **97**:6981–6985.
- Majewski, J., and F. M. Cohan. 1999. DNA sequence similarity requirements for interspecific recombination in *Bacillus*. *Genetics* **153**:1525–1533.
- Mitra, B., and R. Sharma. 2001. The cysteine-rich amino-terminal domain of ZntA, a Pb(II)/Zn(II)/Cd(II)-translocating ATPase from *Escherichia coli*, is not essential for its function. *Biochemistry* **40**:7694–7699.
- Muto, A., and S. Osawa. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc. Natl. Acad. Sci. USA* **84**:166–169.
- Nakamura, Y., T. Gojohori, and T. Ikemura. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**:292.
- Nesbo, C. L., S. L'Haridon, K. O. Stetter, and W. F. Doolittle. 2001. Phylogenetic analyses of two "Archaeal" genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Mol. Biol. Evol.* **18**:362–375.
- Nucifora, G., L. Chu, T. K. Misra, and S. Silver. 1989. Cadmium resistance from *Staphylococcus aureus* plasmid p1258 *cadA* gene results from a cadmium-efflux ATPase. *Proc. Natl. Acad. Sci. USA* **86**:3544–3548.
- Odermatt, A., H. Suter, R. Krampf, and M. Solioz. 1993. Primary structure of two P-type ATPases involved in copper homeostasis in *Enterococcus hirae*. *J. Biol. Chem.* **268**:12775–12779.
- Osborn, A. M., K. D. Bruce, P. Strike, and D. A. Ritchie. 1997. Distribution, diversity, and evolution of the bacterial mercury resistance (*mer*) operon. *FEMS Microbiol. Rev.* **19**:239–262.
- Papke, R. T., N. B. Ramsing, M. M. Bateson, and D. M. Ward. 2003. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**:650–659.
- Preisig, O., R. Zufferey, and H. Hennecke. 1996. The *Bradyrhizobium jaboricum* *fixGHIS* genes are required for the formation of the high-affinity *cbb3*-type cytochrome oxidase. *Arch. Microbiol.* **165**:297–305.
- Puig, S., and D. J. Thiele. 2002. Molecular mechanisms of copper uptake and distribution. *Curr. Opin. Chem. Biol.* **6**:171–180.
- Ragan, M. A. 2001. Detection of lateral gene transfer among microbial genomes. *Curr. Opin. Genet. Dev.* **11**:620–626.
- Reeve, W. G., R. P. Tiwari, N. B. Kale, M. J. Dilworth, and A. R. Glenn. 2002. ActP controls copper homeostasis in *Rhizobium leguminosarum* bv. *viciae* and *Sinorhizobium meliloti* preventing low pH-induced copper toxicity. *Mol. Microbiol.* **43**:981–991.
- Rensing, C., B. Fan, R. Sharma, B. Mitra, and B. P. Rosen. 2000. CopA: an *Escherichia coli* Cu(I)-translocating ATPase. *Proc. Natl. Acad. Sci. USA* **97**:652–656.
- Rensing, C., M. Ghosh, and B. P. Rosen. 1999. Families of soft-metal ion transporting ATPases. *J. Bacteriol.* **181**:5891–5897.
- Rensing, C., Y. Sun, B. Mitra, and B. P. Rosen. 1998. Pb(II) translocating P-type ATPases. *J. Biol. Chem.* **273**:32614–32617.
- Rivera, M. C., R. Jain, J. E. Moore, and J. A. Lake. 1998. Genomic evidence

- for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* **95**: 6239–6244.
45. **Rutherford, J. C., J. S. Cavet, and N. J. Robinson.** 1999. Cobalt-dependent transcriptional switching by a dual-effector MerR-like protein regulates a cobalt-exporting variant CPx-type ATPase. *J. Biol. Chem.* **274**: 25827–25832.
46. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* **112**: 441–457.
47. **Solioz, M., and A. Odermatt.** 1995. Copper and silver transport by CopB-ATPase in membrane vesicles of *Enterococcus hirae*. *J. Biol. Chem.* **270**: 9217–9221.
48. **Solovieva, I. M., and K.-D. Entian.** 2002. Investigation of the *yjgW* *Bacillus subtilis* chromosomal gene involved in Cd²⁺ ion resistance. *FEMS Microbiol. Lett.* **208**:105–109.
49. **Thelwell, C., N. J. Robinson, and J. S. Turner-Cavet.** 1998. An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc. Natl. Acad. Sci. USA* **95**:10728–10733.
50. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
51. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
52. **Tomiyasu, T., M. Okada, R. Imura, and H. Sakamoto.** 2003. Vertical variations in the concentration of mercury in soils around Sakurajima Volcano, Southern Kyushu, Japan. *Sci. Total Environ.* **304**:221–230.
53. **Tong, L., S. Nakashima, M. Shibasaki, M. Katsuhara, and K. Kasamo.** 2002. Novel histidine-rich CPx-ATPase from the filamentous cyanobacterium *Oscillatoria brevis* related to multiple-heavy-metal cotolerance. *J. Bacteriol.* **184**:5027–5035.
54. **Von Damm, K. L., S. E. Oosting, R. Kozlowski, L. G. Buttermore, D. C. Colodner, H. N. Edmonds, J. M. Edmond, and J. M. Grebmeier.** 1995. Evolution of East Pacific Rise hydrothermal vent fluids following a volcanic eruption. *Nature* **375**:47–50.