Biometals (2012) 25:75–93
DOI 10.1007/s10534-011-9484-8

# Genome wide identification of *Acidithiobacillus ferrooxidans* (ATCC 23270) transcription factors and comparative analysis of ArsR and MerR metal regulators

**Christian Hödar · Pablo Moreno · Alex di Genova · Mauricio Latorre · Angélica Reyes-Jara · Alejandro Maass · Mauricio González · Verónica Cambiazo**

Received: 24 February 2011 / Accepted: 21 July 2011 / Published online: 10 August 2011


**Abstract** *Acidithiobacillus ferrooxidans* is a chemolithoautotrophic acidophilic bacterium that obtains its energy from the oxidation of ferrous iron, elemental sulfur, or reduced sulfur minerals. This capability makes it of great industrial importance due to its applications in biomining. During the industrial processes, *A. ferrooxidans* survives to stressing circumstances in its environment, such as an extremely acidic pH and high concentration of transition metals. In order to gain insight into the organization of *A. ferrooxidans* regulatory networks and to provide a framework for further studies in bacterial growth under extreme conditions, we applied a genome-wide annotation procedure to identify 87 *A. ferrooxidans* transcription factors. We classified them into 19 families that were conserved among diverse prokaryotic phyla. Our annotation procedure revealed that *A. ferrooxidans* genome contains several members of the ArsR and MerR families, which are involved in metal resistance and detoxification. Analysis of their sequences revealed known and potentially new mechanism to coordinate gene-expression in response to metal availability. *A. ferrooxidans* inhabit some of the most metal-rich environments known, thus transcription factors identified here seem to be good candidates for functional studies in order to determine their physiological roles and to place them into *A. ferrooxidans* transcriptional regulatory networks.

**Keywords** *Acidithiobacillus ferrooxidans* · Transcriptional regulators · Metal resistance · ArsR family · MerR family

C. Hödar and P. Moreno contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10534-011-9484-8) contains supplementary material, which is available to authorized users.

C. Hödar · M. Latorre · A. Reyes-Jara · M. González · V. Cambiazo (✉)
Laboratorio de Bioinformática y Expresión Génica, INTA, Universidad de Chile, El Libano 5524, Santiago, Chile
e-mail: vcambiaz@inta.cl

P. Moreno · A. di Genova · A. Maass
Laboratorio de Bioinformática y Matemática del Genoma, Center for Mathematical Modeling, FCFM, Universidad de Chile, Santiago, Chile

A. Maass · M. González
Center for Genome Regulation, Universidad de Chile, Santiago, Chile

## Introduction

Regulation of gene expression influences almost all cellular processes in an organism, and sequence-specific DNA-binding transcription factors (TFs) are critical to this control. These TFs respond to changes in cellular environment and activate complex circuits

of regulation between them and their target genes. Therefore, the identification of the repertoire of TFs in a genome sequence is an essential step to understand the regulation of gene expression at a global scale and to elucidate gene regulatory networks.

*Acidithiobacillus ferrooxidans* (Pizarro et al. 1996) is a Gram-negative acidophilic bacterium of the $\gamma$-Proteobacteria group that obtains its energy from oxidation of ferrous ion, elemental sulfur or reduced sulfur minerals (Okibe et al. 2003; Rawlings 2002; Rohwerder et al. 2003). *A. ferrooxidans* is one of the microorganisms involved in the bioleaching process, and thus it survives in environmental conditions characterized by an extremely acidic pH and a high concentration of metals, such as copper and iron (Hallberg and Johnson 2001; Pizarro et al. 1996; Rawlings and Johnson 2007). This bacterium represents an attractive target to elucidate and compare the repertoire of TFs because of its interesting physiology and its importance in biomining processes. Furthermore, the availability of the *A. ferrooxidans* (ATCC 23270) genome sequence provides with an excellent opportunity for the identification of the whole repertoire of transcriptional regulators in this bacterium (Quatrini et al. 2005, 2007; Soulere et al. 2008). Some of these TFs might be involved in physiological processes already described in *A. ferrooxidans*, such as cell attachment, electron transport (Bouchal et al. 2006), quorum sensing (Soulere et al. 2008) and phosphate starvation response (Vera et al. 2008). In addition, several plasmids, transposons and insertion sequences have been identified and studied in this microorganism (Rawlings and Tietze 2001).

Here, using annotation and database tools developed in the author‘s laboratory followed by manual curation we identified a set of 87 predicted TFs in *A. ferrooxidans* and evaluated their conservation across prokaryotic species. In addition, we further characterize metalloregulatory proteins of the ArsR/SmtB and MerR families of TFs, which are known regulators of metal efflux and detoxification elements. Our results provide a framework for further studies on *A. ferrooxidans* transcriptional regulatory networks and contribute to extend the annotation of *A. ferrooxidans* genome reinforcing its emerging role as a model to investigate organism life in extreme environments.

## Materials and methods

### Prediction of transcriptional regulators

The general strategy employed to detect and classify transcriptional regulators of *A. ferrooxidans* was a combination of several bioinformatics tools that intended to reduce the number of undetected transcriptional regulators. We designed a pipeline (Online resource 1) using the GenDB API that performs the following steps:

(1) *Bacterial genome annotation*: A local annotation of the proteome was performed using a locally installed version of GenDB (Meyer et al. 2003), which contains standard annotations tools, such as BLAST, SignalP, TMHMM, Helix-turn-Helix, RPSBLAST and databases NR, SWISSPROT, OMNIOME, PDB, KEGG,-COG and Interpro collection.

(2) *Transcription factor features*: Five features were defined in order to identify putative transcriptional regulators: (i) HTH domain, given that most of the known prokaryotic TFs contain the helix–turn–helix (HTH) DNA-binding motif we scanned *A. ferrooxidans* ATCC 23270 genome using the program HelixturnHelix within the EMBOSS software suite (Dodd and Egan 1990) and identified 114 proteins with HTH motifs; (ii) Superfamily Domains, from the Superfamily database we identified a set of 104 family entries associated with TFs, then we scanned the *A. ferrooxidans* proteome using Interpro tool and these family entries to identify proteins that contained the domains of interest. In addition, we scanned a total of 16 Interpro domains associated with TFs that were not included in the Superfamily subset defined above; (iii) Families of TFs, based on the Interpro description of each annotated protein, they were classified as belonging to the TF families Cro/CI, GntR, LacI, RpiR, AraC, TetR, LysR, MerR, MarR, ArgR, DeoR, Fur, ArpU, LuxR, PadR or ArsR; (iv) Match to a regular expression, R1: “regulato” or “transcriptio” followed by any letters and then by the word “factor” (regular expression 1:/(regulato)|(transcriptio).*(factor)/). We searched the regular

expression R1 in the Interpro and BLAST description of each *A. ferrooxidans* protein.

(3) *Filtering of results:* We classified as potential TFs all those proteins that had at least two of the features listed in (2). The remaining proteins were saved for a manual curation step.

(4) *Manual curation:* First, we selected those proteins that exhibited an HTH predicted domain within their sequence. These proteins were sub-divided in two groups: (i) TF candidates exclusively assigned to class K and (ii) TFs assigned to class K plus another class. Proteins from group (i) with the term transcriptional regulation in their annotation sentences were selected and assigned to a family protein based on Interpro or BLAST searches. Proteins from group (ii) were filtered to discard those annotated as having enzymatic activities; proteins that passed this filter were classified into families. TF candidates with HTH domains and COG assignation different to K or *K* where filtered to eliminate those with enzymatic functions into their annotations. Finally, for candidates with HTH and without COG assignation, we discarded those that did not present regulatory domains as determined by Interpro or BLAST analysis. In the case of proteins that lacked HTH domain prediction, a similar procedure was applied. If they did not have a COG assignment, then they were filtered based on any regulatory domains identified in their sequence. Candidates assigned to K class that were or not annotated as involved in transcriptional regulation where classified into families of TFs using Interpro or BLAST information. Proteins that were not assigned into any families where selected for literature examination only if their annotation sentences indicated that they belonged to a TF protein. The group of proteins assigned to *K* class was filtered to exclude chromatin remodelling factors, enzymes or structural components and Interpro information was used to classify them into protein families.

## Validation of TF prediction pipeline

The pipeline described above was employed to predict transcriptional regulators from the complete proteomes of *Escherichia coli* K-12 and *Corynebacterium glutamicum* ATCC 13032 and then, predicted transcriptional regulators were compared with the databases RegulonDB (Salgado et al. 2004) and CoryneRegNet (Baumbach et al. 2006) that contain 172 and 159 transcriptional regulators of *E. coli* and *C. glutamicum*, respectively.

## Proteome comparisons

Predicted proteomes from 957 complete and public genome sequences from microorganisms available at NCBI database (December, 2010) were extracted and used to create a proteome database for each genome. A bidirectional best BLAST hits (BBHs) strategy was implemented using PERL language to determine ortholog pairs. An ortholog pair was defined as bidirectional best BLAST hits between the curated set of *A. ferrooxidans* TFs and the predicted proteome of each genome. Each pair of hits was required to have at least a 70% of coverage and *E*-values below $10^{-5}$ to be considered an ortholog protein. Data were stored in a matrix having TFs as rows and prokaryote species as columns. Microorganism classification was carried out according to the information reported by each Genome Project at NCBI. In order to evaluate the efficiency of our procedure, we determined ortholog proteins among the proteomes of four *Corynebacterium* species, *C. glutamicum*, *C. diphtheria*, *C. efficiens* and *C. jeikeium*. These microorganisms have been extensively studied and their transcriptional regulatory repertoire is well documented (Brune et al. 2005; Baumbach et al. 2009a, b). We used our BBH procedure to identify ortholog pairs between the curated set of 127 *C. glutamicum* TFs (Brune et al. 2005) and the predicted proteomes of *C. diphtheria*, *C. efficiens* and *C. jeikeium*. We were able to identify 48 pairs of orthologs between *C. glutamicum* and *C. diphtheria*, 73 between *C. glutamicum* and *C. efficiens* and 34 between *C. glutamicum* and *C. jeikeium*, which represented 98, 95 and 91% of the ortholog pairs reported in Brune et al. (2005). Moreover, our procedure recovered 86% of the common repertoire of transcriptional regulators present in the four *Corynebacterium* genomes, as reported in Brune et al. (2005).

### Analysis of TF binding sites

A cutoff of 500 bases upstream and downstream of the coding region was used to define the non-coding sequence associated with the orthologs of each member of *A. ferrooxidans* ArsR and MerR families of transcriptional regulators. Upstream and downstream non-coding sequences were obtained using in-house developed scripts. Then, we used MEME program for ab initio detection of potential binding sites, with the following parameters (-nmotifs 100 -maxsize 100000 -p 16 -dna -pal -evt 1e-1 -mod zoops). Predicted binding sites were required to contain a palindromic sequence. Finally the predicted motifs were manually curated using literature search.

### Multiple sequence alignments, phylogenetic reconstruction and protein analysis

For putative members of the ArsR and MerR families of transcriptional regulators further sequence analysis were performed in order to find conserved domains. Briefly, using ortholog sequences identified by BBH we performed a multiple alignment using ClustalW (Thompson et al. 2002) and the MEGA suite (Tamura et al. 2007). Phylogenetic relationships of AFE_2859 orthologs were inferred using Neighbor-Joining method with 500 bootstrap iterations. Domain-based analyses used SMART (http://smart.embl-heidelberg.de) and Interpro (http://www.ebi.ac.uk/interpro). Secondary structure predictions were carried out using Proteus (Montgomerie et al. 2006).

## Results and discussion

### Identification of *A. ferrooxidans* TFs

In order to detect genes encoding transcriptional regulators, we screened the complete genome sequence of *A. ferrooxidans* strain ATCC 23270 using an automatic annotation procedure followed by manual curation as described in online resource 1. In a first step, we applied the automatic annotation procedure to *E. coli K12* and *C. glutamicum* genomes and we recovered 88% and 84% of the TFs previously annotated at RegulonDB and CoryneRegNet databases. When we manually examined *E. coli K12* and *C. glutamicum* proteins that we failed to predict as TFs by our methods, we found that empirical evidence supported their role as transcriptional regulators, although sequence features might have precluded their annotation as TFs. For instance, in the case of *E. coli* GutM, a transcriptional activator that regulates the expression of glucitol operon (Yamada and Saier 1988), our automatic procedure failed to identify any feature in its sequence that would allow as to classify it as a TF. Similarly, using our pipeline we did not recovered the LexA regulator, which is part of the SOS response in both *E. coli* (Little et al. 1981) and *C. glutamicum* (Jochmann et al. 2009). Thus, a manual curation step was thought to be necessary in order to improve our TF annotation. When we analyzed *A. ferrooxidans* genome, a total of 119 putative TFs (Online resource 2) were predicted using our automatic annotation procedure. In the process of refining the set of predicted *A. ferrooxidans* TFs we excluded genes encoding putative histone-like proteins and factors controlling transcription termination such as Rho which have been described as binding to RNA rather than DNA. Then, we applied a set of rules (see "Materials and methods" section and Online resource 1) to manually curate the list of automatically predicted TFs. Finally, we selected a set of 87 TFs that included putative and already characterized regulatory proteins; they represent 3% of the predicted coding sequences in the genome of *A. ferrooxidans*, which is in agreement with previous estimates suggesting that less than 10% of gene products are associated to gene regulation in bacteria (Paulsen et al. 2003; Moreno-Campuzano et al. 2006; Perez-Rueda and Collado-Vides 2000; Charoensawan et al. 2010). Despite of the fact that extremophile bacteria are underrepresented in the databases of sequenced and annotated genomes, the proportion of TFs found in *A. ferrooxidans* appears to be similar to that reported in others bacteria with a comparable genome size (Brune et al. 2005). Several computational analyses of DNA-binding motifs in different genomes indicate that as bacterial genome size increases, the proportion of the genome devoted to regulatory proteins increases as well (Gonzalez et al. 2005; Charoensawan et al. 2010). This tendency appears to be more prominent in free-living bacteria that can growth under variable environmental conditions (Stover et al. 2000; Brune et al. 2005).

TF families in *A. ferrooxidans*

The whole repertoire of *A. ferrooxidans* TFs was grouped into 18 families of varying sizes (Fig. 1, Online resource 3), a number slightly lower than median values reported for bacteria (23 families/species) (Charoensawan et al. 2010). Only six families contained five or more members, they were LysR (13 TFs), Enhancer-Binding Protein (EBP; 13 TFs), OmpR/PhoB (9 TFs), XRE (9 TFs), GTF (6 TFs) and MerR (5 TFs). OmpR and LysR families of transcriptional regulators have been previously described as two of the most abundant families in the prokaryotic kingdom (Perez-Rueda et al. 2004), whereas the number of member of the EBP family seems to be highly variable between bacteria genomes, ranging from one in *C. trachomatis* (Studholme and Dixon 2003) to 52 in *M. xanthus* (von Bodman et al. 2003). Eight of 13 putative *A. ferrooxidans* EBPs possess CheY-like receiver domains (REC domain), however only five of them were found in cognate pairs with a sensor histidine protein kinase for two-component signal transduction. *A. ferrooxidans* EBPs that lack the REC domain have predicted GAF and/or PAS sensor domains within their N-terminal regions as previously described for other members of the EBP family (Studholme and Dixon 2003). Regarding putative OmpR/PhoB family of transcriptional regulators, we detected that six of the members were linked to genes encoding sensor histidine kinases, while three of them seem to be orphan response regulators.

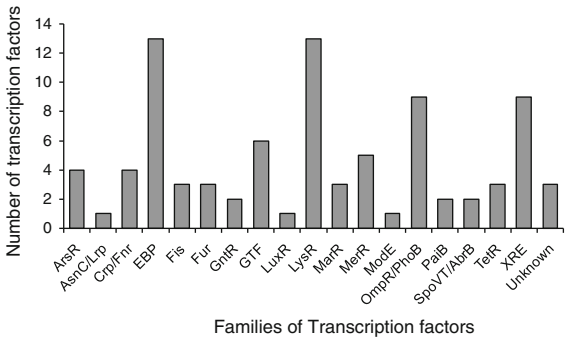


**Fig. 1** Classification of *A. ferrooxidans* TFs into regulatory families. The 18 regulatory families identified in *A. ferrooxidans* are indicated along with the number of members assigned by family

Two families of TFs contained two-members: SpoVT/AbrB, a superfamily of TFs, which includes transition-state regulators, putative regulators of cell wall biosynthesis, regulators of phosphate uptake, and a large number of proteins of unknown activity (Coles et al. 2005; Dong et al. 2004; Hamon et al. 2004), and GntR a large transcriptional regulator family with roles in general metabolism, members of this family possess a conserved N-terminal DNA-binding domain and heterogeneous C-terminal domains that have been used to classify them into sub-families (Rigali et al. 2002). Domain analysis of *A. ferrooxidans* GntR revealed that it contains a C-terminal UbiC transcription regulator-associated (UTRA) domain, which responds to various ligands, such as histidine, fatty acids, sugar and alkylphosphonate (Aravind and Anantharaman 2003). Due to this protein structure prediction these TFs can be classified into the HutC/FarR subfamily of GntRs.

Three members of the Fur family of transcriptional regulators, which control iron uptake and storage, were identified in *A. ferrooxidans* genome. One of them (AFE_0282 in Online resource 1), has been previously characterized by Quatrini et al. (2005) and several of its targets have been identified (Quatrini et al. 2007). *A. ferrooxidans* genome encodes putative members of the MerR (4 TFs) and the ArsR/SmtB (3 TFs) families. ArsR/SmtB and MerR families contain metalloregulatory proteins that control genes whose expression is linked to stress induced by the extracellular increment of metal concentrations. The structural features of these putative metalloregulatory proteins will be discussed below.

Five families of TFs contained only one member per group in the *A. ferrooxidans* genome, among them: LuxR family has been involved in *A. ferrooxidans* quorum sensing. In this microorganism, the classical autoinducer-binding transcriptional regulator LuxR (AFE_1997) and the autoinducer synthesis protein LuxI (AFE_1999) have been functionally characterized (von Bodman et al. 2003), AsnC/Lrp family, represented by the gene AFE_1990, contains global transcriptional regulators widely distributed in both bacteria and archaea (Calvo and Matthews 1994) and the AraC family, represented in *A. ferrooxidans* by gene AFE_3137 gene, seems to group larger numbers of transcriptional regulators in other species of $\gamma$-Proteobacteria (Perez-Rueda et al. 2004). We also identified a gene encoding a protein that consists

of just an HTH_9 domain highly similar to the DNA-binding domain of *modE* (AFE_0723), a regulator of molybdenum metabolism (Grunden et al. 1996). In summary, we found that *A. ferrooxidans* had a slightly lower number of families than median values in bacteria and exhibited differences in the number of members per family in comparison to already well-characterized *B. subtilis* and *E. coli* genomes (Moreno-Campuzano et al. 2006; Perez-Rueda and Collado-Vides 2000). For instance, in *E. coli* LysR is the largest family with 45 TFs, while in *A. ferrooxidans* LysR is also the largest family but with 13 members. In addition, we detected low number of members belonging to otherwise large families, such as, GntR (2 TFs), LuxR (1 TFs). On the contrary, in *A. ferrooxidans* the number of genes encoding EBP, OmpR and Fis response regulators was comparable to that described in *B. subtilis* genome (20 TFs) and *E. faecalis* (13 TFs). This is interesting, since two-component systems are the primary means by which free-living bacteria sense their changing environments, and it has been suggested that the relative increase in transcriptional regulators in more complex bacteria is accompanied by an equal relative increase in sensory systems (Whitworth and Cock 2009).

## Conservation of *A. ferrooxidans* TFs in prokaryotes

In order to examine the conservation of *A. ferrooxidans* TFs in prokaryotic genomes, we compared the 87 family-assigned TFs of *A. ferrooxidans* by aligning the sequence of each of them to predicted proteomes from 957 prokaryotic genomes available at NCBI (Online resource 4). In doing so, we applied the best bidirectional BLAST hits (BBH) method, which is a widely used and accepted procedure to determine ortholog proteins among different proteomes (Hulsen et al. 2006; Altenhoff and Dessimoz 2009; Zhang and Leong 2010) and is practical to explore large amounts of sequence information in comparative genomics studies (Zhang and Leong 2010; Blom et al. 2009; Muller et al. 2010). Here, we operationally defined an ortholog protein based on BBH (Price et al. 2007), using an *E*-value cut-off of $10^{-5}$ and a minimal required alignment coverage of 70%. As we show in Fig. 2a, the selected genomes include diverse prokaryotic phyla with the highest percentage of genomes belonging to the

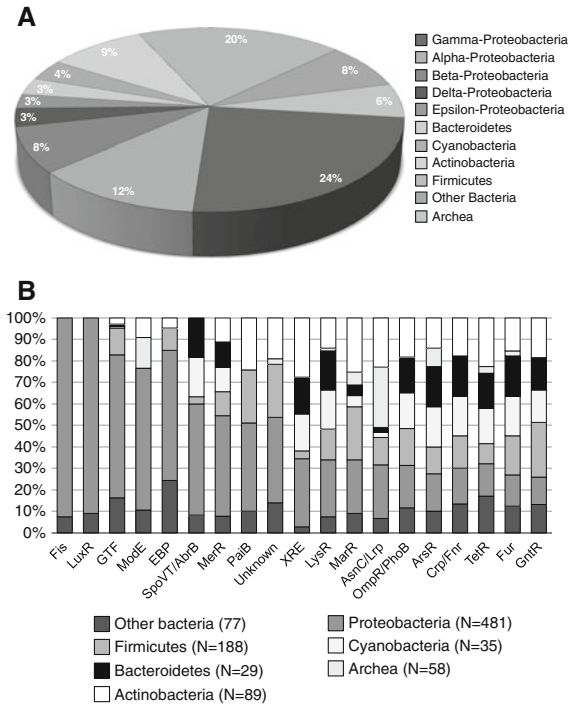


**Fig. 2** Distribution of orthologs of *A. ferrooxidans* TFs in 957 prokaryotic genomes. **a** *Pie chart* showing the percentage of genomes from each phylum (and classes for Proteobacteria). **b** *Bars* represent the cumulative percentage of orthologs of members of *A. ferrooxidans* transcription factors in the 957 proteomes. The *inset* shows the colour assigned to phyla and the number of genomes belonging to each of them

Proteobacteria phylum (50%), followed by Firmicutes (20%). In order to have a better overview of *A. ferrooxidans* TF distribution in the prokaryotes, we examined the proportion of members of each TF family with orthologs in the 957 proteomes (Fig. 2b).

In general the repertoire of transcriptional regulators of *A. ferrooxidans* was identified in different lineage of prokaryotes, however, as expected, a large proportion of *A. ferrooxidans* TFs had orthologs in all Proteobacteria genomes analyzed (Fig. 2b). In particular, orthologs of the Fis and LuxR families were almost exclusively identified in the phylum Proteobacteria. Accordingly, members of the LuxR family, along with LuxI-type protein (the AHL synthase), which are the core of the acylated homoserine lactone driven quorum sensing system (Miller and Bassler 2001) have been only found within Proteobacterial genomes. Even though cases of horizontal gene transfer have occurred, LuxR and LuxI seem to have arisen early in the evolution of the Proteobacteria and

subsequently diverged within each group of organisms as a whole (Case et al. 2008; Gray and Garey 2001; Lerat and Moran 2004). Regarding the members of the Fis family of transcriptional regulators, they have been reported to be unique to bacteria (Aravind et al. 2005; Minezaki et al. 2005) and they seem to be specifically restricted to Proteobacteria (Lozada-Chavez et al. 2006).

*Acidithiobacillus ferrooxidans* orthologs of TFs belonging to the Omp/PhoB, ArsR, Crp/Fnr, TetR, Fur and GntR families were evenly distributed among the different bacteria phyla, although some of them were also present in archaeal genomes (Fig. 2b). In archaea, TFs appear to be more closely related to their eukaryotic counterparts, even though some of them have detectable orthologs in bacteria (Makarova et al. 1999; Minezaki et al. 2005; Perez-Rueda and Janga 2010). In agreement with previous studies, we found that *A. ferrooxidans* orthologs of the AsnC/Lrp, IclR, ArsR, MarR, LysR, TetR, Fur and ModE families were represented in both bacteria and archaea phyla (Fig. 2b). For some of these families (AsnC/Lrp, ArsR, LysR, Fur) it has been proposed that they form part of an ancestral core of transcriptional regulators that have descended from the common ancestor of bacteria and archaea (Brinkman et al. 2003; Peeters and Charlier 2010; Perez-Rueda and Janga 2010). In *A. ferrooxidans*, ModE was the only family of transcriptional regulators that being highly represented among Proteobacteria also showed an important proportion of orthologs in archaea (14%). Therefore, the distribution of orthologs of *A. ferrooxidans* ModE is consistent with previous studies showing that members of ModE family were not detected among Gram-positive bacteria (Studholme and Pau 2003). This family of transcriptional regulators have been found in several species of archaea and also in *Salmonella typhimurium* and *Agrobacterium tumefaciens*. Even though these proteins are not functional for binding molybdate or other solutes, it has been proposed that the HTH domain would be capable of binding DNA and form multimeric protein complexes (Studholme and Pau 2003).

## *A. ferrooxidans* TFs involved in metal homeostasis and metal resistance

The expression of genes implicated in metal homeostasis and resistance is tightly controlled by specific metal-responsive transcriptional regulators, frequently called metalloregulators. Metalloregulator proteins bind metal ions directly and repress, derepress or activate the transcription of genes encoding proteins that function in metal detoxification, sequestration, efflux and uptake (Pennella and Giedroc 2005). Metalloregulatory proteins can be divided into two major groups: those that regulate the expression of genes encoding proteins involved in the uptake of essential metals (Fur, DtxR/MntR, and NikR families) and those that regulate the expression of components of metal efflux and detoxification mechanisms (ArsR/SmtB and MerR families). Here we describe the structural characteristics of known and predicted members of ArsR/SmtB and MerR families of transcriptional regulators identified in *A. ferrooxidans* genome. They encompass seven transcriptional regulators, some of them might be capable of mediating metal resistance in *A. ferrooxidans*, some of these putative regulators map adjacent to the genes that they potentially regulate, while others appear to be isolated in the chromosome.

### MerR family of TFs in *A. ferrooxidans*

Metal resistance systems regulated by the members of the MerR family include, resistance to zinc (ZntR) and copper (CueR and HmrR) excess, as well as detoxification of mercury (MerR), cadmium (CadR) and other toxic metals (Brown et al. 2003). Experimental data indicate that members of this family generally have broad metal specificity, for example CueR reacts with copper, silver and gold, whereas ZntR is mainly regulated by zinc but also responds to cadmium and lead (Brown et al. 2003). All known members of the MerR family of transcriptional regulators are distinguished by its secondary structure folding: $\beta1$–$\alpha1$–$\alpha2$–$\beta2$–$\beta3$–$\alpha3$–$\alpha4$–$\alpha5$–$\alpha6$. They have a conserved N-terminal DNA binding domains with $\alpha$-helix dimerization domains, followed by a long coiled-coil region whereas the C-terminal regions are highly diverse and contain the metal binding domains (MBD) (Brown et al. 2003; Changela et al. 2003). These transcriptional regulators bind to a region of dyad symmetry that lies within an unusually long spacer of 19 bp between the $-10$ and $-35$ motifs of the structural gene promoter. Upon binding of the metal, transcriptional activation involves distortion of the DNA at the centre of the promoter to reduce the

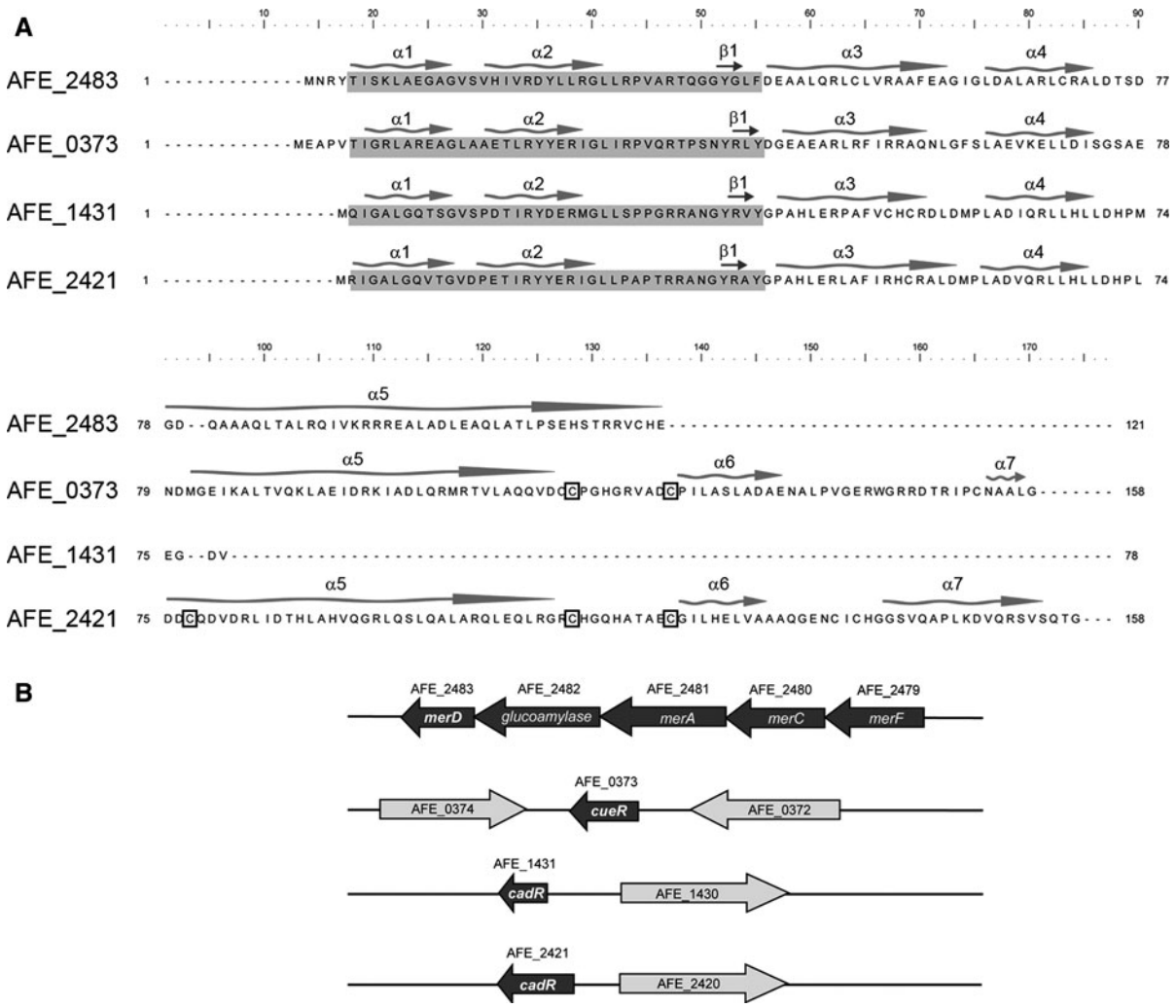


**Fig. 3** Structural features of putative TFs of *A. ferrooxidans* MerR family. **a** Alignment of putative members of *A. ferrooxidans* MerR family showing secondary structure prediction. *Grey box* represents the predicted HTH-DNA binding domain of *A. ferrooxidans* proteins. Secondary structure elements are represented above the sequences: α: α-helices and β: β-strands. Cysteine residues potentially involved in divalent metal coordination are indicated in *boxes*. **b** Genetic map of *A. ferrooxidans mer* operon (AFE_2483–AFE_2479) and locations of the other three putative MerR transcriptional regulators: CueR-like (AFE_0373), CadR-like (AFE_1431 and AFE_2421). AFE_0374 and AFE_0372 genes encode unknown proteins. AFE_1430 and AFE_2420 genes encode putative members of the CDF family

distance between the −35 and −10 sequences, which increases the promoter binding strength (Permina et al. 2006).

In our analysis, four transcriptional regulators were identified as putative members of the MerR family (Fig. 3a). The secondary structure prediction for these proteins revealed a high similarity with the folding pattern previously described for members of this family (Pennella and Giedroc 2005 and Online resource 5). In the putative *A. ferrooxidans* MerR regulators, we detected the presence of tyrosines within the predicted HTH motif at the N-terminal end (black asterisks in Online resource 5, and sequences alignment of Fig. 3a); tyrosine residues probably participate in the direct interaction of these regulators with the DNA (Brown et al. 2003). In particular, AFE_2483 shares high levels (>70%) of sequence identity with MerD proteins encoded in the well characterized Tn501 and Tn21 *mer* operons (Online resource 5) (Liebert et al. 1999). Expression of *mer* operon is primarily regulated by the product of *merR* gene and is inducible by mercury (Brown et al. 2003).

However, it has been proposed that MerD acts as a co-regulator to finely modulate the expression of *mer* operon based on sequence similarities between the N-terminal HTH region of MerD and MerR and the capacity of MerD to bind the *mer* operator region, albeit more weakly than MerR (Mukhopadhyay et al. 1991). Moreover, the expression in trans of *merD* represses the induction of the *mer* operon (Nucifora et al. 1989), suggesting that this protein might act as a transcriptional repressor. In fact, studies of MerD from *Ralstonia metallidurans* indicated that MerD was able to displace Hg-bound MerR from the *mer* operator allowing the synthesis of metal-free MerR in order to shut down the induction of the *mer* genes when the external mercury is exhausted (Champier et al. 2004). Previous studies have determined the composition of chromosomal *mer* operon of *A. ferrooxidans* strain E-15 (Inoue et al. 1991; Inoue et al. 1989; Rawlings and Kusano 1994), it consists of a functional promoter region followed by *merR* (the regulatory gene), *merC* (encoding a mercury transport protein), and *merA* (encoding the enzyme mercuric reductase) (Inoue et al. 1996; Kusano et al. 1990; Inoue et al. 1991; Dietrich et al. 2008), however no orthologs of *merD* were observed in the *mer* operon of this strain. Similarly, no *merD* gene was present in the putative *mer* operons of *A. ferrooxidans* strain ATCC 53993 (NC_011206), which contains orthologs of *merR* (Lferr_0159 and Lferr_0164) and *merC* (Lferr_0160 and Lferr_0163) genes. Variations of chromosomal *mer* operon structure have been also reported in *A. ferroxidans* strain T3.2, including the presence of *merR* and the absence of *merC* and *merD* genes (Velasco et al. 1999). Therefore, the variability of composition and organization of chromosomal *mer* operon of E-15, ATCC 53993, T3.2 and ATCC 23270 strains represents an important polymorphism among strains *of A. ferrooxidans*, as it has been described for other Gram-negative species (Liebert et al. 2000). This variability supports the hypothesis that MerD substituted the regulatory role of MerR in the *mer* operon of *A. ferrooxidans* ATCC 23270.

In a recent study, a high amino acidic sequence identity was described between CueR of *E. coli* and AFE_0373 of *A. ferrooxidans* ATCC 23270 (Navarro et al. 2009). The structure analysis of AFE_0373, revealed the presence of a typical DNA binding domain at its N-terminal region, including a highly conserved tyrosine residue within helix α2 (Online resource 5). In addition, AFE_0373 contains two cysteine residues ($Cys^{116}$ and $Cys^{125}$) that are in the corresponding positions to the $Cys^{112}$ and $Cys^{120}$ residues of *E. coli* CueR (Fig. 3a). In *E. coli*, these two residues are known to be important for metal recognition, since mutation of $Cys^{112}$ or $Cys^{120}$ abolishes regulator response to copper, silver and gold (Brown et al. 2003; Hobman et al. 2005). Moreover, the Cys-Pro motif that is conserved at the end of the metal-binding region in many of the members of MerR family (Changela et al. 2003) was also found in the putative CueR of *A. ferrooxidans*. Mutation of the proline residue diminishes the response to mercury of MerR, supporting a functional role for the Cys-Pro motif (Hamon et al. 2004). As indicated by Permina et al. (2006), the structure of the copper resistance systems seems to be complex, thus CueR might be part of an operon (*E. coli*) or a divergon (*S. typhi* and *S. typhimurium*) or even lies separately and has no candidate binding site (*Vibrio vulnificus*). In the case of *A. ferrooxidans*, AFE_0373 does not seem to be part of an operon (Fig. 3b) and we were not able to predict binding site sequences within its upstream region, thus the structure of this putative *A. ferrooxidans* copper resistance system appears to be more similar to that described in *Vibrio vulnificus*.

In addition, we predicted two putative members of *A. ferrooxidans* CadR, AFE_1431 and AFE_2421. CadR is a cadmium induced TF that is able to control the expression of *cadA* and *CDF* genes (cadmium/zinc transporters) (Permina et al. 2006). A sequence comparison analysis revealed that AFE_1431 encodes a truncated protein of 75 amino acids, which contains the N-terminal DNA binding domain but lacks of the α-helix dimerization domain and the metal-binding region (Fig. 3a). The second ORF, AFE_2421, encodes a 174 amino acid protein, which has >40% of sequence identity with *P. putida* and *P. aeruginosa* CadR regulators and 29% of identity with *E. coli* ZntR (Online resource 5). Differences between ZntR and the CadR transcriptional regulators seem to be restricted to their C-terminal regions where CadR regulators contain a histidine-rich domain (Brocklehurst et al. 2003). Even though, we did not detect a histidine-rich C-terminal region in the putative *A. ferrooxidans* CadR, the analysis of the genome context of AFE_2421 gene provided further

support to classify its product as a CadR regulator (Fig. 3b). A comparative genomic analysis of heavy metal resistance revealed that most cadmium-induced regulators that are located on the chromosome form divergons with *cadA* transporters that belong to the P-type ATPase (P-ATPase) superfamily (Permina et al. 2006), in the case of *A. ferrooxidans* AFE_2421 and AFE_1431 we found that these putative CadR regulators were oriented divergently from a member of cation diffusion facilitator family transporter (CDF; AFE_2420 and AFE_1430). Interestingly, in the same comparative work mentioned above (Permina et al. 2006), the authors reported three cases of plasmid CadR regulators that were divergently arranged to a transporter belonging to the CDF family, suggesting that the *cadR-CDF* systems of *A. ferrooxidans* ATCC 23270 arose from vector acquisition and subsequent insertion in the genome.

A palindromic motif that lies between AFE_2421 and the putative CDF AFE_2420 was identified as conserved among ortholog operons; this DNA binding motif maintains the same configuration described in other *cadR* operons (9–3–9 palindrome) (Permina et al. 2006) (Fig. 4). In the case of AFE_1431, we could not identified orthologos within the different bacteria species analyzed, probably because it is a truncated protein, however, a manual inspection of the region between AFE_1431 and its corresponding CDF transporter, allowed us to find a palindromic sequence that shares a high level of identity (>40%) with both the motif founded in AFE_2421 and the 9–3–9 dyad reported in *P. aeruginosa* CadR (Brocklehurst et al. 2003). An alignment of the regulatory regions of genes encoding *A. ferrooxidans* CDF members and *P. aeruginosa* transporters regulated by CadR illustrates the similarity in the inverted repeat sequences and the complete identity between these motifs (Fig. 4). This observation strongly suggests that AFE_2421 is an *A. ferrooxidans* CadR protein that

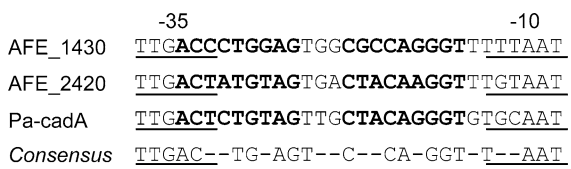


**Fig. 4** Alignment of *A. ferrooxidans* and *P. aeruginosa* CadR promoter regions. −35 and −10 motif sequences are *underlined*, 9 bp inverted repeats are shown in *bold*. Identical bases are shown in the consensus sequence

regulates the expression of the CDF proteins in response to cadmium stimulus. Regarding AFE_1431, the absence of the MBD in the predicted protein suggests that AFE_1431 might be a divergent TF of this family unable to respond to cadmium.

Besides the putative regulators discussed above, a fifth ORF, AFE_2509, was initially identified as a SoxR protein, a redox-sensing transcriptional activator that belongs to the MerR family (Brown et al. 2003). It is known that activation of SoxR requires the presence of two Fe–S clusters, which are coordinated by a cysteine-rich motif (Bradley et al. 1997). The global alignment of AFE_2509 with the well-characterized SoxR of *E. coli* (Brown et al. 2003) revealed the presence of a five cysteine residues in the C-terminal region of the protein, however they do not conform the CI[G/Q]CGC[L/M][S/L]XXXC SoxR-specific cysteine motif required for binding of the 2Fe–2S cluster (Dietrich et al. 2008). Moreover, through several searches of *A. ferrooxidans* ATCC 23270 genome we were not able to find homologous of *soxS*, a second TF regulated by SoxR which in turn control the expression of genes related with oxidative stress (Bradley et al. 1997) (data not shown).

In summary, our bioinformatics analysis permitted us to identify and classify different proteins of the MerR family of transcriptional regulators. The high identity with archetypal regulators described in other species, genome organization and the presence of conserved MBDs, indicated that genes AFE_2483, AFE_0373 and AFE_2421 probably encode the *A. ferrooxidans* ATCC 23270 transcriptional regulators MerR, CueR and CadR, respectively. In particular, the conserved palindromic CadR binding motif located upstream of a putative CDF (AFE_2420) protein suggested that AFE_2421 and AFE_2420 conform a cadmium resistance system in this *A. ferrooxidans* strain. Even though, cadmium tolerance has been reported in *A. ferrooxidans* strains (Dopson et al. 2003) to our knowledge no cadmium resistance systems have been characterized in this bacterium, making the putative *AfcadR-CDF* described here an interesting candidate for functional analyses.

## ArsR/SmtB family of TFs in *A. ferrooxidans*

The members of the ArsR/SmtB family of transcriptional regulators appear to act exclusively as transcriptional repressors of metal-resistance proteins.

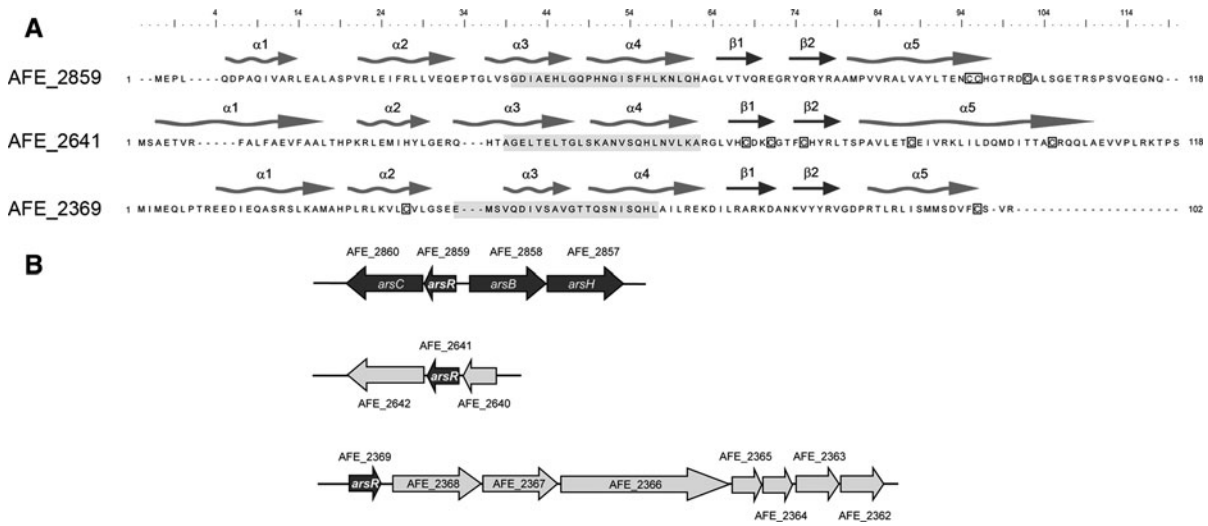


**Fig. 5** Structural features of putative TFs of *A. ferrooxidans* ArsR family. **a** Alignment of putative members of *A. ferrooxidans* ArsR/SmtB family showing secondary structure prediction. *Grey box* represents the predicted HTH-DNA binding domain of *A. ferrooxidans* proteins. Secondary structure elements are represented above the sequences: $\alpha$: $\alpha$-helices and $\beta$: $\beta$-strands. Cysteine residues are indicated in *boxes*. **b** Genetic map of *A. ferrooxidans arsCRBH* operon (AFE_2860–AFE_2857; Butcher et al. 2000) and locations of two putative ArsR transcriptional regulators. Genes in cluster AFE_2642-AFE_2640 encode putative proteins similar to permease of the Major Facilitator Superfamily (AFE_2642), ArsR-like transcriptional regulator (AFE_2641) and beta-lactamase (AFE_2643). Genes from AFE_2369–AFE_2362 encode an ArsR-like (AFE_2369), an outer membrane protein (AFE_2368), a membrane fusion protein (AFE_2367), a cation/multidrug efflux pump (AFE_2366), two rhodanases (AFE_2365 and AFE_2364), a glutaredoxin (AFE_2363) and a thioredoxin (AFE_2362)

They are present in many bacteria and archaea, and sense a variety of di- and multivalent metal ions. Genes encoding ArsR/SmtB repressors have been identified in chromosomal or episomal operons containing genes that encode metal-specific efflux pumps, membrane bound transporters, or intracellular chelators, such as metallothioneins (Busenlehner et al. 2003).

In *A. ferrooxidans* ATCC 23270 genome, one already characterized (AFE_2859) and two putative (AFE_2641 and AFE_2369) members of the ArsR/SmtB family of transcription regulators were identified. A domain analysis of *A. ferrooxidans* ArsR sequences confirmed the presence of ArsR-type helix-turn-helix domain signature (HTH). All of the putative *A. ferrooxidans* members of the ArsR/SmtB family showed the common secondary structure that characterizes this family: $\alpha$1–$\alpha$2–$\alpha$3–$\alpha$4–$\beta$1–$\beta$2–$\alpha$5 (Busenlehner et al. 2003) and contained a DNA-binding motif of the winged helix family ($\alpha$3-turn-$\alpha$4) (Fig. 5a). However, they did not exhibit the conserved metal sensor cysteine residues of ELCV(C/G)D motif located at the start of the DNA binding domain (also called $\alpha$3N metal-binding site) of plasmid R773 and *E. coli* chromosomal ArsR regulators (Shi et al. 1994). In addition, they do not possess a second C-terminal MBD identified at the $\alpha$3 helix ($\alpha$3 metal-binding site) of CadC, SmtB and other members of the ArsR/SmtB family involved in the regulation of metals such as zinc, cobalt and nickel (Rensing 2005).

Among the ArsR regulators identified in the *A. ferrooxidans* genome, we found the already characterized *Af*ArsR (AFE_2859), which has high levels of sequence identity with unknown ArsR from other species as described in Butcher and Rawlings (2002). *Af*ArsR was located within a previously reported *ars* operon (Fig. 5b) (Butcher and Rawlings 2002; Butcher et al. 2000; Wu et al. 2010), comprising genes *arsB*, *arsC* and *arsH* that encode an arsenite permease, an arsenate reductase and a flavoprotein, respectively. *Af*ArsR possesses three cysteine residues, $Cys^{95}$, $Cys^{96}$, and $Cys^{102}$ that form a 3S-coordinate As(III) binding domain within the C-terminal $\alpha$5-helix of each monomer (Fig. 5a) (Qin et al. 2007). It has been proposed that members of the

ArsR/SmtB family contain a conserved DNA-binding region or core domain and a large variety of MBDs, which appear to have arisen by convergent evolution at spatially distinct locations within the ArsR protein (Ordonez et al. 2008; Qin et al. 2007). Thus, a number of divergent binding sites have been functional characterized in several members of the ArsR/SmtB family (Ordonez et al. 2005, 2008; Qin et al. 2007; Wang et al. 2005).

An alignment of the primary sequence of orthologs of *Af*ArsR (AFE_2859) showed structural similarities and conserved residues within the HTH domain. Interestingly the $Cys^{95}$, $Cys^{96}$ and $Cys^{102}$ residues that form a 3-coordinate As(III) MBD are well conserved among the orthologs with the exception of a group of sequences that posses only one or two Cys residues in the MBD (Online resource 6). In agreement with the observations reported by Qin et al. (Qin et al. 2007), we found that *Af*ArsR (AFE_2859) of *A. ferrooxidans* ATCC 23270 was more similar to uncharacterized proteins discovered during sequencing projects of bacteria genomes (for example *A. ferrooxidans* ATCC53993 and *D. desulfuricans* ATCC 27774 in Online resource 7) and no good matches were detected to ArsR transcriptional regulators with known functions.

The secondary structure and domain analysis of *A. ferrooxidans* AFE_2641 protein, indicated that the As(III) binding site of *Af*ArsR and the MDB described in *E. coli* (Qin et al. 2007; Xu et al. 1996) were not conserved in this putative ArsR. However, AFE_2641 possesses five cysteines which might potentially coordinated metals (Fig. 5a). Three of them are located in the predicted $\beta1$–$\beta2$ domain and the others two in the α5 domain. Even though known members of ArsR/SmtB family possess functional metal-binding sites located at the ends of α-helices (Ordonez et al. 2008; Qin et al. 2007; Wang et al. 2005), we can not discard that cysteine residues of AFE_2641 might be part of an unidentified MBD. When we examine the genes immediately upstream and downstream *A. ferrooxidans* AFE_2641 (Fig. 5b), we detected that they encode a predicted member of the Major Facilitator Superfamily (MFS; AFE_2640) and a hydrolase belonging to the metallo-beta-lactamase superfamily (AFE_2642) (Fig. 5b). Interestingly, this pattern of genomic organization was also found in the genomes of *Leptospirillum ferrooxidans* (Coram et al. 2005) and *Leptospirillum sp*. (Lo et al. 2007). An alignment of AFE_2640 and AFE_2642 with the corresponding genes from *L. ferrooxidans* and *Leptospirillum* sp. revealed high levels of sequence identity between the proteins (data not shown). Even though the role of AFE_2641 in metal resistance needs further investigation, our analyses revealed conservation of the operon gene structure in microorganism that share similar environmental condition, suggesting that a positive evolutionary selection might result in the similar organization of this set of genes across these phylogenetically distant genomes.

Analysis of predicted amino acid sequences of the third putative *A. ferrooxidans arsR* gene (AFE_2369) revealed the presence of cysteine residues in helices α2 and α5 (Fig. 5a), which is a characteristic feature of ArsR transcriptional regulators whose binding to promoter regions is not regulated by metals. Known TFs that possess cysteine residues in helices α2 and α5 include HlyU from *V. cholerae*, which positively regulates expression of hemolysin (HlyA) (Saha and Chakrabarti 2006), SoxR from *P. salicylatoxidans*, which negatively regulates expression of the sulphur oxidation (*sox* operon) (Mandal et al. 2007) and EcaR from *E. carotovora* (Campbell et al. 2007). Downstream of AFE_2369 gene (Fig. 5b), we found three proteins: an outer membrane, a membrane fusion protein and a cation efflux pump (AFE_2368–AFE_2366). This structure resembles a multidrug resistance efflux pump known as Mex-type, involved in antibiotic resistance (Fig. 5b) (Nikaido 1996; Poole 2001). Additional genes within this cluster encode two rhodanases (AFE_2365, AFE_2364), a glutaredoxin (AFE_2363) and a thioredoxin (AFE_2362). Similar operon structure was found in *Thiobacillus denitrificans* genome that also shares the highest level of sequence identity with the putative ArsR regulator. The predicted functions of proteins encoded by this gene cluster suggest that they may be involved in the maintenance of redox balance (rhodanases, glutaredoxins, and thioredoxins) and secretion of siderophores (efflux pump). Thus, the product of AFE_2369 gene seems to mediate the transcriptional regulation of an operon that complements the ability of *A. ferrooxidans* to growth in extreme environments.

Analysis of eubacterial ArsR/SmtB regulated promoters have revealed a conserved 12–2–12 palindrome, which is often near the transcriptional start

site and acts as a regulatory DNA binding site (Busenlehner et al. 2003). In Qin et al. (2007), the authors reported that binding of recombinant *Af*ArsR (AFE_2859) protects the sequence ATCCACGAA TATTTCTTGCAGTATTGAC in a DNAse I foot-printing assay. This sequence is located in the middle of the divergent *ars* operon, between −86 and −60 upstream of the arsenite permease ArsB. When we attempted to identify this binding site in the upstream and downstream regions of orthologs of *Af*ArsR, we recovered a putative motif (Fig. 6a) that contained a conserved 10–2–10 inverted repeat. In the case of *Af*ArsR, this motif contained an imperfect inverted repeat that overlapped with the sequence identified in Qin et al. (2007). The same analysis was performed with AFE_2369 and AFE_2641 ArsR/SmtB family members. Within orthologs of AFE_2369, we found a palindromic sequence not well conserved (Fig. 6b) in which we could not recognize any known TF binding site. For AFE_2641 we did not find a conserved motif upstream or downstream to this TF or its orthologs.

Transcriptional regulation in response to As(III) is directly related with the presence of cysteine residues in the MBD and the ability of the TF to release the promoter. In particular, for *Af*ArsR (AFE_2859) it was demonstrated that $Cys^{95}$ and $Cys^{96}$ could bind As(III), but the formed thiolate was not required for DNA binding. However, replacement of C65S, C96S and C102S resulted in a protein that was able to bind DNA but lost the As(III) regulation site, implying that $Cys^{102}$ fulfils the role of creating a stable S3 thiolate in a monomer unit, which is required to release the protein-DNA binding (Qin et al. 2007). In comparison with the MBD of AFE_2859, TFs encoded by AFE_2369 and AFE_2641 posses at the α5 helix domain only one and two cysteines, respectively, suggesting that both TFs might employ different mechanisms for sensing the metal and for DNA binding. Thus, the evolution of MBD might have taken place in parallel with the evolution of DNA binding site of the members of ArsR/SmtB family that require As(III) to regulate gene-expression.

In order to examine this hypothesis we evaluated whether the orthologs of *Af*ArsR (AFE_2859), have differences in the HTH domain and in the MBD located at the α5 helix. Analysis of cysteine residues of the orthologs of *Af*ArsR (Online resource 6) and a phylogenetic reconstruction based on *Af*ArsR primary structure (Online resource 7), allowed us to classify these putative member of ArsR family into four categories: those containing the canonical domain Cys–Cys–$X_{4–6}$–Cys (including AFE_2859), those having a Cys–Cys domain and two other groups that possess one or none cysteine in their α5 helix. For each protein in the first and second categories, we searched the non-coding regions of their genes and detected a consensus sequence containing a 10–2–10 palindrome in which the six nucleotides at the ends of the palindrome were well conserved (Fig. 7a). Moreover, this consensus sequence was similar to the already described DNA binding site motif of *Af*ArsR (Qin et al. 2007). Thus, this result suggests that a

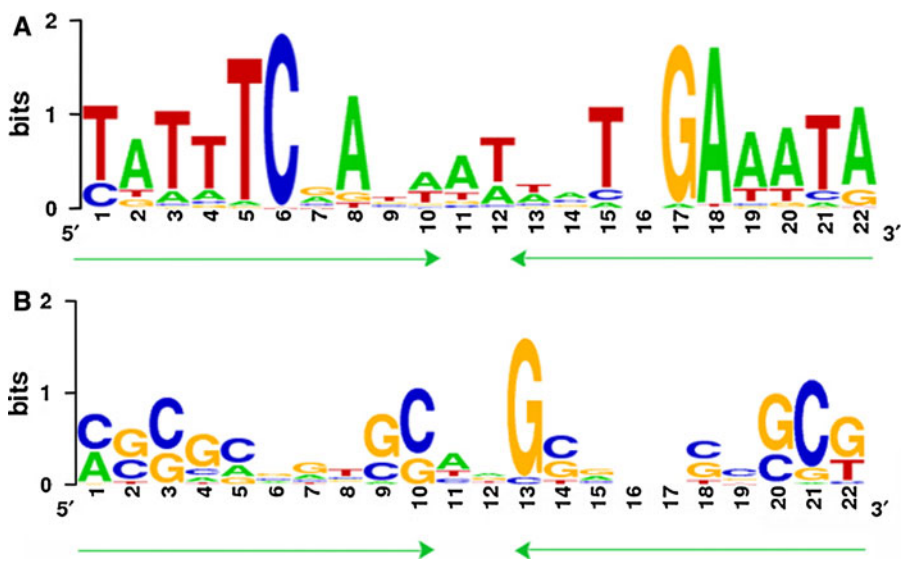


**Fig. 6** Putative DNA binding sites of orthologs of AFE_2859 and AFE_2369. Sequence of motifs that were identified by MEME within the non-coding sequences of orthologs of *A. ferrooxidans* AFE_2859 (**a**) and AFE_2369 (**b**) orthologs were aligned and represented using the WebLogo tool (Crooks et al. 2004). Arrows indicate palindromic sequence

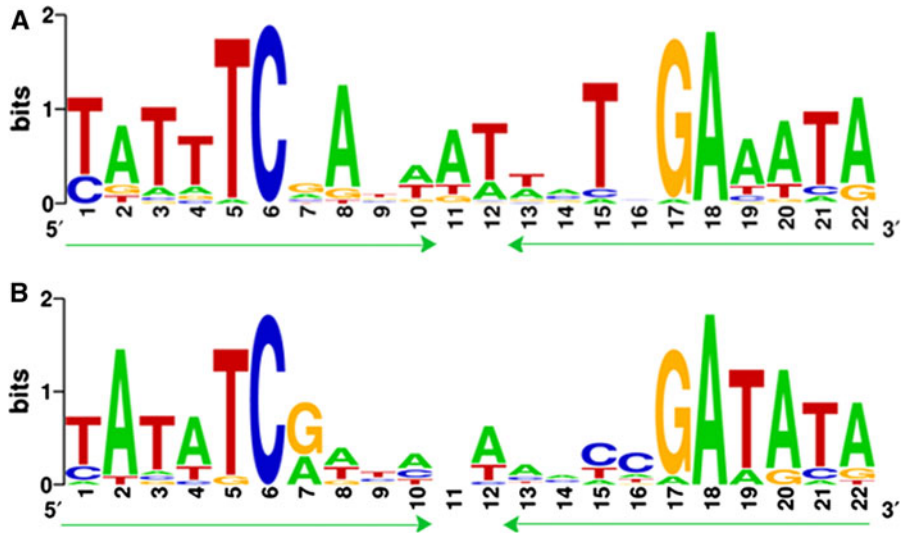


**Fig. 7** Putative DNA binding sites of orthologs of AFE_2859. Sequence of motifs that were identified by MEME within the non-coding sequences of orthologs of *A. ferrooxidans* AFE_2859 were classified according to the numbers of cysteine residues present in their MDB. **a** Cys-Cys-$X_{4-6}$-Cys or Cys-Cys. **b** one Cys or none Cys. *Arrows* indicate palindromic sequence

conserved regulatory motif control the expression of this family of TFs (Fig. 7a). Since proteins in the second category possess only two Cys residues in the MBD, they might employ an alternatively molecular mechanism to sense and coordinate As(III). Alternatively, they might possess affinity for a different metal, as it has been reported for CmtR, a member of ArsR family in *M. tuberculosis* (Banci et al. 2007). In this case, the 3S thiolate of CmtR dimer was formed by two cysteine residues (57 and 61) from one monomer plus the $Cys^{102}$ from the other and the release from DNA was triggered by cadmium (Banci et al. 2007). Finally, within the non-coding regions of genes encoding proteins with one or none Cys residues in the α5 domain (Online resource 7) we detected a second palindromic consensus sequence (Fig. 7b) in which the six nucleotides at the ends of the palindrome were also well conserved. When we compared this consensus sequence with that described above, we found that nucleotides 6, 17 and 18 were well conserved within the entire set of non-coding regions; however the A/T base composition of the six nucleotides at the ends of the palindrome differed between the two consensus sequences. Interestingly, the phylogenetic reconstruction (Online resource 7) indicated that AFE_2641 and AFE_2369 proteins were more closely related to ArsR containing two or one Cys residue in their α5 helices.

Even though members of the ArsR and MerR families of transcriptional regulators have been previously described in other bacteria, our analysis revealed some specific features of members of these families in *A. ferrooxidans* ATCC 23270. Among them, the variable composition and organization of chromosomal *mer* operon of *A. ferrooxidans*, which brought us to suggest that in *A. ferrooxidans* ATCC 23270 *mer* operon is primarily regulated by the product of *merD* gene, whereas in most bacteria including other strains of *A. ferrooxidans* this operon is regulated by MerR. With regards to the members of the *A. ferrooxidans* ArsR/SmtB family, they all share an ArsR-type helix-turn-helix domain signature, however comparison of AFE_2641 and AFE_2369 sequences with that of AFE_2859, the only functionally characterized *A. ferrooxidans* ArsR (Butcher and Rawlings 2002), revealed differences in the number and spacing of cysteine residues that might conform the putative MBDs of AFE_2641 and AFE_2369. The variable composition of MBDs among *A. ferrooxidans* ArsR suggests that a structural diversity of ArsR regulators might be relevant to increase the capacity of *A. ferrooxidans* ATCC 23270 for activating gene expression in response to different sets of metals. It is expected that both specie-specific and strain-specific TFs and their corresponding targets have evolved to deal with the high metal concentration that characterizes *A. ferrooxidans* habitats. Even though further studies will be necessary to identify particular gene regulatory networks that control *A. ferrooxidans* response to metal exposure, the results reported here might be integrated with data

from transcriptomics (Quatrini et al. 2006) and proteomics (Ramirez et al. 2004; Valenzuela et al. 2006) studies in order to advance in the characterization of the transcriptional response of *A. ferrooxidans* to high metal concentrations.

## Conclusions

In summary, we examined the repertoire of TFs in the sequenced genome of *A. ferrooxidans* ATCC 23270 by using a comparative genomics-based approach. The automatic search was followed by manual curation and an extensive literature revision focused on the identification and characterization of genes encoding TFs involved in metal homeostasis and metal resistance. Our data provide valuable information on general *A. ferrooxidans* biology and will increase our understanding of how this microorganism survives under extreme environmental conditions by modulating the expression of relevant genes. Given that the majority of the genes encoding *A. ferrooxidans* TFs have not been studied yet, our data will assist the selection of putative regulators that seems to be good candidates for functional studies in order to determine their physiological roles and to place them into *A. ferrooxidans* transcriptional regulatory networks. Moreover, *A. ferrooxidans* inhabit some of the most metal-rich environments known, thus it is an excellent model system for study of microbial metal resistance. Based on our *in silico* studies is now possible to identify genes associated with metal resistance and initiate the functional characterization of *A. ferrooxidans* metalloregulators, as well as, their regulated metal resistance genes. As an example, we identified a putative cadmium resistance system in *A. ferrooxidans*, which is composite by a CadR transcriptional regulator of the Mer family and a CDF protein forming a divergon unit that might be transcribed from a conserved promoter region. We expect that the *AfcadR-CDF* unit described here will become an interesting candidate for functional analysis of cadmium resistance.

**Acknowledgments** This work was supported by Fondef D04I1257 (AM), Fondecyt 1090211 (VC), 1110427 (MG) and FONDAP-15090007(VC, AM and MG). CH was supported by a doctoral fellowship from Conicyt, a Dr. A. Stekel fellowship and a Postdoctoral grant Fondecyt 311019. The authors thank Marcel Spehr, Carito Vargas, Alvaro Graves and Nicolas Loira for their invaluable efforts on making to work the local GenDB installation. We also thank Burkhard Kinke and Alexander Goesmann at Bielefeld University; they were very helpful every time we reached them.

## References

Altenhoff AM, Dessimoz C (2009) Phylogenetic and functional assessment of orthologs inference projects and methods. PLoS Comput Biol 5(1):e1000262. doi:10.1371/journal.pcbi.1000262

Aravind L, Anantharaman V (2003) HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. FEMS Microbiol Lett 222(1):17–23. doi:10.1016/S0378-1097(03)00242-8

Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev 29(2):231–262. doi:10.1016/j.femsre.2004.12.008

Banci L, Bertini I, Cantini F, Ciofi-Baffoni S, Cavet JS, Dennison C, Graham AI, Harvie DR, Robinson NJ (2007) NMR structural analysis of cadmium sensing by winged helix repressor CmtR. J Biol Chem 282(41):30181–30188. doi:10.1074/jbc.M701119200

Baumbach J, Brinkrolf K, Czaja LF, Rahmann S, Tauch A (2006) CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. BMC Genomics 7:24. doi:10.1186/1471-2164-7-24

Baumbach J, Rahmann S, Tauch A (2009a) Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms. BMC Syst Biol 3:8. doi:10.1186/1752-0509-3-8

Baumbach J, Wittkop T, Kleindt CK, Tauch A (2009b) Integrated analysis and reconstruction of microbial transcriptional gene regulatory networks using CoryneRegNet. Nat Protoc 4(6):992–1005. doi:10.1038/nprot.2009.81

Blom J, Albaum SP, Doppmeier D, Puhler A, Vorholter FJ, Zakrzewski M, Goesmann A (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. BMC Bioinform 10:154. doi:10.1186/1471-2105-10-154

Bouchal P, Zdrahal Z, Helanova S, Janiczek O, Hallberg KB, Mandl M (2006) Proteomic and bioinformatic analysis of iron- and sulfur-oxidizing *Acidithiobacillus ferrooxidans* using immobilized pH gradients and mass spectrometry. Proteomics 6(15):4278–4285. doi:10.1002/pmic.200500719

Bradley TM, Hidalgo E, Leautaud V, Ding H, Demple B (1997) Cysteine-to-alanine replacements in the *Escherichia coli* SoxR protein and the role of the [2Fe-2S] centers in transcriptional activation. Nucleic Acids Res 25(8):1469–1475. doi:10.1093/nar/25.8.1469

Brinkman AB, Ettema TJ, de Vos WM, van der Oost J (2003) The Lrp family of transcriptional regulators. Mol Microbiol 48(2):287–294. doi:10.1155/2010/750457

Brocklehurst KR, Megit SJ, Morby AP (2003) Characterisation of CadR from *Pseudomonas aeruginosa*: a Cd(II)-responsive MerR homologue. Biochem Biophys Res Commun 308(2):234–239. doi:10.1016/S0006-291X(03)01366-4

Brown NL, Stoyanov JV, Kidd SP, Hobman JL (2003) The MerR family of transcriptional regulators. FEMS Microbiol Rev 27(2–3):145–163. doi:10.1016/S0168-6445(03)00051-2

Brune I, Brinkrolf K, Kalinowski J, Puhler A, Tauch A (2005) The individual and common repertoire of DNA-binding transcriptional regulators of *Corynebacterium glutamicum*, *Corynebacterium efficiens*, *Corynebacterium diphtheriae* and *Corynebacterium jeikeium* deduced from the complete genome sequences. BMC Genomics 6(1):86. doi:10.1186/1471-2164-6-86

Busenlehner LS, Pennella MA, Giedroc DP (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. FEMS Microbiol Rev 27(2–3):131–143. doi:10.1016/S0168-6445(03)00054-8

Butcher BG, Rawlings DE (2002) The divergent chromosomal ars operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. Microbiology 148(Pt 12):3983–3992

Butcher BG, Deane SM, Rawlings DE (2000) The chromosomal arsenic resistance genes of *Thiobacillus ferrooxidans* have an unusual arrangement and confer increased arsenic and antimony resistance to *Escherichia coli*. Appl Environ Microbiol 66(5):1826–1833

Calvo JM, Matthews RG (1994) The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol Rev 58(3):466–490

Campbell DR, Chapman KE, Waldron KJ, Tottey S, Kendall S, Cavallaro G, Andreini C, Hinds J, Stoker NG, Robinson NJ, Cavet JS (2007) Mycobacterial cells have dual nickel-cobalt sensors: sequence relationships and metal sites of metal-responsive repressors are not congruent. J Biol Chem 282(44):32298–32310. doi:10.1074/jbc.M703451200

Case RJ, Labbate M, Kjelleberg S (2008) AHL-driven quorum-sensing circuits: their frequency and function among the Proteobacteria. ISME J 2(4):345–349. doi:10.1038/ismej.2008.13

Champier L, Duarte V, Michaud-Soret I, Coves J (2004) Characterization of the MerD protein from *Ralstonia metallidurans* CH34: a possible role in bacterial mercury resistance by switching off the induction of the mer operon. Mol Microbiol 52(5):1475–1485. doi:10.1111/j.1365-2958.2004.04071.x

Changela A, Chen K, Xue Y, Holschen J, Outten CE, O'Halloran TV, Mondragon A (2003) Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. Science 301(5638):1383–1387. doi:10.1126/science.1085950

Charoensawan V, Wilson D, Teichmann SA (2010) Genomic repertoires of DNA-binding transcription factors across the tree of life. Nucleic Acids Res 38(21):7364–7377. doi:10.1093/nar/gkq617

Coles M, Djuranovic S, Soding J, Frickey T, Koretke K, Truffault V, Martin J, Lupas AN (2005) AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-psi beta barrels. Structure 13(6):919–928. doi:10.1016/j.str.2005.03.017

Coram NJ, van Zyl LJ, Rawlings DE (2005) Isolation, sequence analysis, and comparison of two plasmids (28 and 29 kilobases) from the biomining bacterium *Leptospirillum ferrooxidans* ATCC 49879. Appl Environ Microbiol 71(11):7515–7522. doi:10.1128/AEM.71.11.7515-7522.2005

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14(6):1188–1190. doi:10.1101/gr.849004

Dietrich LE, Teal TK, Price-Whelan A, Newman DK (2008) Redox-active antibiotics control gene expression and community behavior in divergent bacteria. Science 321(5893):1203–1206. doi:10.1126/science.1160619

Dodd IB, Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res 18(17):5019–5026

Dong TC, Cutting SM, Lewis RJ (2004) DNA-binding studies on the *Bacillus subtilis* transcriptional regulator and AbrB homologue, SpoVT. FEMS Microbiol Lett 233(2):247–256. doi:10.1016/j.femsle.2004.02.013

Dopson M, Baker-Austin C, Koppineedi PR, Bond PL (2003) Growth in sulfidic mineral environments: metal resistance mechanisms in acidophilic micro-organisms. Microbiology 149(Pt 8):1959–1970

Gonzalez AD, Espinosa V, Vasconcelos AT, Perez-Rueda E, Collado-Vides J (2005) TRACTOR_DB: a database of regulatory networks in gamma-proteobacterial genomes. Nucleic Acids Res 33(Database issue):D98–D102. doi:10.1093/nar/gki054

Gray KM, Garey JR (2001) The evolution of bacterial LuxI and LuxR quorum sensing regulators. Microbiology 147(Pt 8):2379–2387

Grunden AM, Ray RM, Rosentel JK, Healy FG, Shanmugam KT (1996) Repression of the *Escherichia coli* modABCD (molybdate transport) operon by ModE. J Bacteriol 178(3):735–744

Hallberg KB, Johnson DB (2001) Biodiversity of acidophilic prokaryotes. Adv Appl Microbiol 49:37–84

Hamon MA, Stanley NR, Britton RA, Grossman AD, Lazazzera BA (2004) Identification of AbrB-regulated genes involved in biofilm formation by *Bacillus subtilis*. Mol Microbiol 52(3):847–860. doi:10.1111/j.1365-2958.2004.04023.x

Hulsen T, de Vlieg J, Leunissen JA, Groenen PM (2006) Testing statistical significance scores of sequence comparison methods with structure similarity. BMC Bioinform 12(7):444. doi:10.1186/1471-2105-7-444

Hobman JL, Wilkie J, Brown NL (2005) A design for life: prokaryotic metal-binding MerR family regulators. Biometals 18(4):429–436. doi:10.1007/s10534-005-3717-7

Inoue C, Sugawara K, Shiratori T, Kusano T, Kitagawa Y (1989) Nucleotide sequence of the *Thiobacillus ferrooxidans* chromosomal gene encoding mercuric reductase. Gene 84(1):47–54. doi:10.1016/0378-1119(89)90138-8

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(11):2707–2718

Inoue C, Kusano T, Silver S (1996) Mercuric ion uptake by *Escherichia coli* cells producing *Thiobacillus ferrooxidans* merC. Biosci Biotechnol Biochem 60(8):1289–1292

Jochmann N, Kurze AK, Czaja LF, Brinkrolf K, Brune I, Huser AT, Hansmeier N, Puhler A, Borovok I, Tauch A (2009) Genetic makeup of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays. Microbiology 155(Pt 5):1459–1477. doi:10.1099/mic.0.025841-0

Kusano T, Ji GY, 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(5):2688–2692

Lerat E, Moran NA (2004) The evolutionary history of quorum-sensing systems in bacteria. Mol Biol Evol 21(5): 903–913. doi:10.1093/molbev/msh097

Liebert CA, Hall RM, Summers AO (1999) Transposon Tn21, flagship of the floating genome. Microbiol Mol Biol Rev 63(3):507–522.

Liebert CA, Watson AL, Summers AO (2000) The quality of merC, a module of the mer mosaic. J Mol Evol 51(6): 607–622. doi:10.1007/s002390010124

Little JW, Mount DW, Yanisch-Perron CR (1981) Purified lexA protein is a repressor of the recA and lexA genes. Proc Natl Acad Sci USA 78(7):4199–4203

Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, DiBartolo G, Tyson GW, Allen EE, Ram RJ, Detter JC, Richardson P, Thelen MP, Hettich RL, Banfield JF (2007) Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446(7135): 537–541. doi:10.1038/nature05624

Lozada-Chavez I, Janga SC, Collado-Vides J (2006) Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res 34(12):3434–3445

Makarova KS, Aravind L, Galperin MY, Grishin NV, Tatusov RL, Wolf YI, Koonin EV (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. Genome Res 9(7):608–628

Mandal S, Chatterjee S, Dam B, Roy P, Das Gupta SK (2007) The dimeric repressor SoxR binds cooperatively to the promoter(s) regulating expression of the sulfur oxidation (sox) operon of *Pseudaminobacter salicylatoxidans* KCT001. Microbiology 153(Pt 1):80–91. doi:10.1099/mic.0.29197-0

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Puhler A (2003) GenDB—an open source genome annotation system for prokaryote genomes. Nucleic Acids Res 31(8):2187–2195. doi:10.1093/nar/gkg312

Miller MB, Bassler BL (2001) Quorum sensing in bacteria. Annu Rev Microbiol 55:165–199. doi:10.1146/annurev.micro.55.1.165

Minezaki Y, Homma K, Nishikawa K (2005) Genome-wide survey of transcription factors in prokaryotes reveals many bacteria-specific families not found in archaea. DNA Res 12(5):269–280. doi:10.1093/dnares/dsi016

Montgomerie S, Sundararaj S, Gallin WJ, Wishart DS (2006) Improving the accuracy of protein secondary structure prediction using structural alignment. BMC Bioinform 7:301. doi:10.1186/1471-2105-7-301

Moreno-Campuzano S, Janga SC, Perez-Rueda E (2006) Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes—a genomic approach. BMC Genomics 7:147. doi:10.1186/1471-2164-7-147

Mukhopadhyay D, Yu HR, 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(28):18538–18542

Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, Powell S, von Mering C, Doerks T, Jensen LJ, Bork P (2010) eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. Nucleic Acids Res 38(Database):D190–D195. doi:10.1093/nar/gkp951

Navarro CA, Orellana LH, Mauriaca C, Jerez CA (2009) Transcriptional and functional studies of *Acidithiobacillus ferrooxidans* genes related to survival in the presence of copper. Appl Environ Microbiol 75(19):6102–6109. doi:10.1128/AEM.00308-09

Nikaido H (1996) Multidrug efflux pumps of gram-negative bacteria. J Bacteriol 178(20):5853–5859

Nucifora G, Silver S, Misra TK (1989) Down regulation of the mercury resistance operon by the most promoter-distal gene merD. Mol Gen Genet 220(1):69–72

Okibe N, Gericke M, Hallberg KB, Johnson DB (2003) Enumeration and characterization of acidophilic microorganisms isolated from a pilot plant stirred-tank bioleaching operation. Appl Environ Microbiol 69(4):1936–1943

Ordonez E, Letek M, Valbuena N, Gil JA, Mateos LM (2005) Analysis of genes involved in arsenic resistance in *Corynebacterium glutamicum* ATCC 13032. Appl Environ Microbiol 71(10):6206–6215. doi:10.1128/AEM.71.10.6206-6215.2005

Ordonez E, Thiyagarajan S, Cook JD, Stemmler TL, Gil JA, Mateos LM, Rosen BP (2008) Evolution of metal(loid) binding sites in transcriptional regulators. J Biol Chem 283(37):25706–25714. doi:10.1074/jbc.M803209200

Paulsen IT, Banerjei L, Myers GS, Nelson KE, Seshadri R, Read TD, Fouts DE, Eisen JA, Gill SR, Heidelberg JF, Tettelin H, Dodson RJ, Umayam L, Brinkac L, Beanan M, Daugherty S, DeBoy RT, Durkin S, Kolonay J, Madupu R, Nelson W, Vamathevan J, Tran B, Upton J, Hansen T, Shetty J, Khouri H, Utterback T, Radune D, Ketchum KA, Dougherty BA, Fraser CM (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. Science 299(5615):2071–2074. doi:10.1126/science.1080613

Peeters E, Charlier D (2010) The Lrp family of transcription regulators in archaea. Archaea 2010:750457. doi:10.1155/2010/750457

Pennella MA, Giedroc DP (2005) Structural determinants of metal selectivity in prokaryotic metal-responsive transcriptional regulators. Biometals 18(4):413–428. doi:10.1007/s10534-005-3716-8

Perez-Rueda E, Collado-Vides J (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res 28(8):1838–1847. doi:10.1016/j.compbiolchem.2004.09.004

Perez-Rueda E, Janga SC (2010) Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. Mol Biol Evol 27(6):1449–1459. doi:10.1093/molbev/msq033

Perez-Rueda E, Collado-Vides J, Segovia L (2004) Phylogenetic distribution of DNA-binding transcription factors in bacteria and archaea. Comput Biol Chem 28(5–6):341–350

Permina EA, Kazakov AE, Kalinina OV, Gelfand MS (2006) Comparative genomics of regulation of heavy metal resistance in Eubacteria. BMC Microbiol 6:49. doi:10.1186/1471-2180-6-49

Pizarro J, Jedlicki E, Orellana O, Romero J, Espejo RT (1996) Bacterial populations in samples of bioleached copper ore as revealed by analysis of DNA obtained before and after cultivation. Appl Environ Microbiol 62(4):1323–1328

Poole K (2001) Multidrug efflux pumps and antimicrobial resistance in *Pseudomonas aeruginosa* and related organisms. J Mol Microbiol Biotechnol 3(2):255–264

Price MN, Dehal PS, Arkin AP (2007) Orthologous transcription factors in bacteria have different functions and regulate different genes. PLoS Comput Biol 3(9):1739–1750. doi:10.1371/journal.pcbi.0030175

Qin J, Fu HL, Ye J, Bencze KZ, Stemmler TL, Rawlings DE, Rosen BP (2007) Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. J Biol Chem 282(47):34346–34355. doi:10.1074/jbc.M706565200

Quatrini R, Lefimil C, Holmes DS, Jedlicki E (2005) The ferric iron uptake regulator (Fur) from the extreme acidophile *Acidithiobacillus ferrooxidans*. Microbiology 151(Pt 6):2005–2015. doi:10.1099/mic.0.27581-0

Quatrini R, Appia-ayme C, Denis Y, Ratouchniak J, Veloso F, Valdes J, Lefimil C, Silver S, Roberto F, Orellana O, Denizot F, Jedlicki E, Holmes DS, Bonnefoy V (2006) Insights into the iron and sulfur energetic metabolism of *Acidithiobacillus ferrooxidans* by microarray transcriptome profiling. Anglais 83:263–272

Quatrini R, Lefimil C, Veloso FA, Pedroso I, Holmes DS, Jedlicki E (2007) Bioinformatic prediction and experimental verification of Fur-regulated genes in the extreme acidophile *Acidithiobacillus ferrooxidans*. Nucleic Acids Res 35(7):2153–2166. doi:10.1093/nar/gkm068

Ramirez P, Guiliani N, Valenzuela L, Beard S, Jerez CA (2004) Differential protein expression during growth of *Acidithiobacillus ferrooxidans* on ferrous iron, sulfur compounds, or metal sulfides. Appl Environ Microbiol 70(8):4491–4498. doi:10.1128/AEM.70.8.4491-4498.2004

Rawlings DE (2002) Heavy metal mining using microbes. Annu Rev Microbiol 56:65–91. doi:10.1146/annurev.micro.56.012302.161052

Rawlings DE, Johnson DB (2007) The microbiology of biomining: development and optimization of mineral-oxidizing microbial consortia. Microbiology 153(Pt 2):315–324. doi:10.1099/mic.0.2006/001206-0

Rawlings DE, Kusano T (1994) Molecular genetics of *Thiobacillus ferrooxidans*. Microbiol Rev 58(1):39–55

Rawlings DE, Tietze E (2001) Comparative biology of IncQ and IncQ-like plasmids. Microbiol Mol Biol Rev 65(4):481–496. doi:10.1128/MMBR.65.4.481-496.2001

Rensing C (2005) Form and function in metal-dependent transcriptional regulation: dawn of the enlightenment. J Bacteriol 187(12):3909–3912. doi:10.1128/JB.187.12.3909-3912.2005

Rigali S, Derouaux A, Giannotta F, Dusart J (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. J Biol Chem 277(15):12507–12515. doi:10.1074/jbc.M110968200

Rohwerder T, Gehrke T, Kinzler K, Sand W (2003) Bioleaching review, Part A: progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation. Appl Microbiol Biotechnol 63(3):239–248. doi:10.1007/s00253-003-1448-7

Saha RP, Chakrabarti P (2006) Molecular modeling and characterization of *Vibrio cholerae* transcription regulator HlyU. BMC Struct Biol 6:24. doi:10.1186/1472-6807-6-24

Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. Nucleic Acids Res 32 (Database issue):D303–D306. doi:10.1093/nar/gkh140

Shi W, Wu J, Rosen BP (1994) Identification of a putative metal binding site in a new family of metalloregulatory proteins. J Biol Chem 269(31):19826–19829

Soulere L, Guiliani N, Queneau Y, Jerez CA, Doutheau A (2008) Molecular insights into quorum sensing in *Acidithiobacillus ferrooxidans* bacteria via molecular modelling of the transcriptional regulator AfeR and of the binding mode of long-chain acyl homoserine lactones. J Mol Model 14(7):599–606. doi:10.1007/s00894-008-0315-y

Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Nature 406(6799):959–964. doi:10.1038/35023079

Studholme DJ, Dixon R (2003) Domain architectures of sigma54-dependent transcriptional activators. J Bacteriol 185(6):1757–1767. doi10.1128/JB.185.6.1757-1767.2003

Studholme DJ, Pau RN (2003) A DNA element recognised by the molybdenum-responsive transcription factor ModE is conserved in *Proteobacteria*, green sulphur bacteria and Archaea. BMC Microbiol 3:24. doi:10.1186/1471-2180-3-24

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24(8):1596–1599. doi:10.1093/molbev/msm092

Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinform Chapter 2(Unit 2):3. doi:10.1002/0471250953.bi0203s00

Valenzuela L, Chi A, Beard S, Orell A, Guiliani N, Shabanowitz J, Hunt DF, Jerez CA (2006) Genomics, metagenomics and proteomics in biomining microorganisms. Biotechnol Adv 24(2):197–211. doi:10.1016/j.biotechadv.2005.09.004

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(1):35–43

Vera M, Pagliai F, Guiliani N, Jerez CA (2008) The chemolithoautotroph *Acidithiobacillus ferrooxidans* can survive under phosphate-limiting conditions by expressing a C-P lyase operon that allows it to grow on phosphonates. Appl Environ Microbiol 74(6):1829–1835. doi:10.1128/AEM.02101-07

von Bodman SB, Ball JK, Faini MA, Herrera CM, Minogue TD, Urbanowski ML, Stevens AM (2003) The quorum sensing negative regulators EsaR and ExpR(Ecc), homologues within the LuxR family, retain the ability to function as activators of transcription. J Bacteriol 185(23):7001–7007

Wang Y, Hemmingsen L, Giedroc DP (2005) Structural and functional characterization of *Mycobacterium tuberculosis* CmtR, a PbII/CdII-sensing SmtB/ArsR metalloregulatory repressor. Biochemistry 44(25):8976–8988. doi:10.1021/bi050094v

Whitworth DE, Cock PJ (2009) Evolution of prokaryotic two-component systems: insights from comparative genomics. Amino Acids 37(3):459–466. doi:10.1007/s00726-009-0259-2

Wu XL, Miao B, Han JA, Hu Q, Zeng J, Liu YD, Qiu GZ (2010) Purification and enzymatic properties of arsenic resistance protein ArsH from heterogeneous expression in *E. coli* BL21. Trans Nonferr Metal Soc 20(10):1987–1992. doi:10.1016/S1003-6326(09)60406-4

Xu C, Shi W, Rosen BP (1996) The chromosomal arsR gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. J Biol Chem 271(5):2427–2432

Yamada M, Saier MH Jr (1988) Positive and negative regulators for glucitol (gut) operon expression in *Escherichia coli*. J Mol Biol 203(3):569–583. doi:10.1016/0022-2836(88)90193-3

Zhang M, Leong HW (2010) Bidirectional best hit r-window gene clusters. BMC Bioinform 11(Suppl 1):S63. doi:10.1186/1471-2105-11-S1-S63