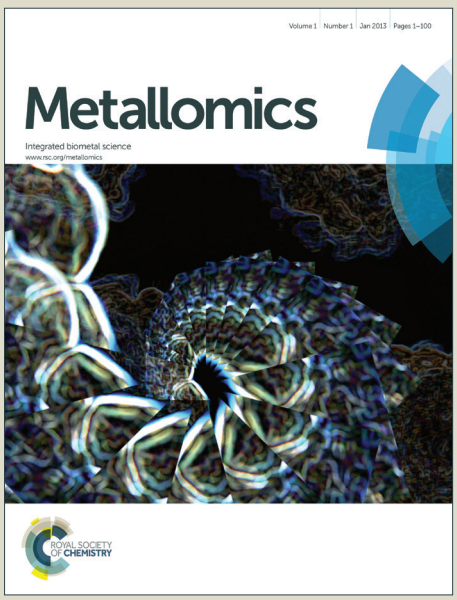
Metallomics

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/metallomics

The cation diffusion facilitator protein EmfA of *Rhizobium etli* belongs to a novel subfamily of Mn^{2+}/ Fe^{2+} transporters conserved in α-proteobacteria

Ciro Cubillas[1], Pablo Vinuesa[1], Maria Luisa Tabche[2], Araceli Davalos[1], Alejandra Vázquez[2], Ismael Hernandez-Lucas[2], David Romero[1] and Alejandro García-de los Santos[1*]

[1] *Programa de Ingeniería Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Apdo. Postal 565-A. Cuernavaca, Morelos, México. Fax: +52 777 317 5581; Tel: +52 5622 7691; E-mail: alex@ccg.unam.mx*

[2] *Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. Fax: +52 777 317 2388; Tel: +53 777 329 1600.*

**Corresponding author.*

ABSTRACT

Manganese (Mn^{2+}) plays a key role in important cellular functions such as oxidative stress response and bacterial virulence. The mechanisms of Mn^{2+} homeostasis are not fully understood, there are few data regarding the functional and taxonomic diversity of Mn^{2+} exporters. Our recent phylogeny of the cation diffusion facilitator (CDF) family of transporters classified the bacterial Mn^{2+}-CDF transporters characterized to date: *Streptococcus pneumoniae* MntE and *Deinococcus radiodurans* DR1236 in two monophyletic groups. DR1236 was shown to belong to the highly-diverse metal specificity clade VI, together with TtCzrB, a Zn^{2+}/Cd^{2+} transporter from *Thermus Thermophilus*, the Fe^{2+} transporter Sll1263 from *Synechocystis* sp and eight uncharacterized homologs whose potential $Mn^{2+}/Zn^{2+}/Cd^{2+}/Fe^{2+}$ specificities could not be accurately inferred because only eleven proteins were grouped in this clade. A new phylogeny inferred from the alignment of 197 clade VI homologs revealed three novel subfamilies of uncharacterized proteins. Remarkably, one of them contained 91 uncharacterized α-proteobacteria transporters (46% of the protein data set) grouped into a single subfamily. The Mn^{2+}/Fe^{2+} specificity of this subfamily was proposed through the functional characterization of *Rhizobium etli RHE_CH03072* gene. This gene was upregulated by Mn^{2+}, Zn^{2+}, Cd^{2+} and Fe^{2+} but conferred only Mn^{2+} resistance to *R. etli.* The expression of RHE_CH03072 gene in an *E. coli mntP/zitB/zntA* mutant did not relieved either Zn^{2+} or Mn^{2+} stress but slightly increased its Fe^{2+} resistance. These results indicate that the *RHE_CH03072* gene, now designated as *emfA*, encodes for a bacterial Mn^{2+}/Fe^{2+} resistance CDF protein, having orthologs in more than 60 α-proteobacterial species.

Introduction

Manganese (Mn^{2+}) is an essential micronutrient for all living cells. In bacteria, it plays a key role in the defense against oxidative stress and in the regulation of bacterial virulence [1,2,3,4]. Due to its relevance, there is an increased interest in understanding the mechanisms that maintain its homeostasis. Mn^{2+} importer proteins have been studied in several bacterial models including the symbiotic nitrogen-fixing α-proteobacteria *Sinorhizobium meliloti*, *Rhizobium leguminosarum* and *Bradyrhizobium japonicum*. In these rhizobia, the Mn^{2+} uptake systems have been shown to be dependent of *sitABC* [5,6] and *mntH* [7] , respectively; whose expression is regulated by Mur[6, 8] and Fur[7] under low Mn^{2+} concentrations. *Rhizobium leguminosarum* also contains uncharacterized *Escherichia coli* MntH orthologs [6].

To maintain the Mn^{2+} homeostasis and avoid toxicity, cells need to efflux the excess of Mn^{2+}. In this regard, two families of efflux proteins have been involved in bacterial Mn^{2+} resistance: 1) The DUF204 family (Pfam2659) which includes the characterized YebN homologues from *Xanthomonas oryzae* [2], *Neisseria meningitidis* [3], and *Escherichia coli* [9] efflux proteins; 2) The CDF family which includes MntE [4] and DR1236 [10] proteins from *Streptococcus pneumoniae* and *Deinococcus radiodurans*, respectively.

Recently, we reported the functional classification of 318 non-redundant CDF transporters using phylogenomic inference [11]. According to this study, all known Mn^{2+}-CDF proteins are distributed in three monophyletic clades: I, IV and VI [11]. Clades I and VI contain CDF proteins from bacterial origin while clade IV contains the eukaryotic ones. MntE is the only characterized member of bacterial Mn^{2+}-CDF grouped in clade I, whereas DR1236 is located in the highly-diverse metal specificity clade VI, together with the characterized exporter proteins of Zn^{2+}/Cd^{2+} TtCzrB from *Thermus Thermophilus* and the Fe^{2+} transporter Sll1263 from *Synechocystis* sp.

Among the uncharacterized members of this clade are *Rhizobium etli* CFN42 RHE_CH03072 and its rhizobial orthologs. *R. etli* CFN42 is a symbiotic nitrogen-fixing bacterium whose metal transportome is being elucidated in our laboratory, in order to understand how this soil bacterium deals with fluctuating metal concentrations in its environment. The diverse metal specificities of characterized clade VI proteins difficult an accurate prediction of potential Mn^{2+} transporters included among the uncharacterized homologs. The reduced numbers of homologs currently grouped in clade VI (eleven proteins) additionally restrict an accurate prediction of clade-specific motifs involved in metal selectivity. To overcome these limitations, in this study we report an improved maximum-likelihood (ML) phylogeny of clade VI inferred from 197 protein homologs. This new phylogeny suggested the existence of substrate-defined subfamilies for Mn^{2+}, Zn^{2+}/Cd^{2+} and Fe^{2+} and also uncovered three novel subfamilies of uncharacterized proteins. One of them contained 91 proteins, exclusively from α proteobacteria, including the putative transporter RHE_CH03072 from *R. etli* CFN42. The Mn^{2+}/Fe^{2+}-specificities of this subfamily were proposed from three independent evidences: 1) The Mn^{2+}-sensitive phenotype of a *R. etli* mutant, 2) The induction of *RHE_CH03072* gene expression by Mn^{+2} and Fe^{+2} and 3) the heterologous expression of *RHE_CH03072* relieving Fe^{2+} stress in *E. coli*. This study allowed us to predict the existence of different residues involved in metal selectivity, other than the canonical A, B, and C metal binding sites (MBS). We propose *emfA* (efflux of Mn^{2+}/Fe^{2+}) as a new denomination for *RHE_CH03072* and we use this name throughout this study.

RESULTS

The metal specificity of EmfA cannot be inferred from amino acid sequence data. For clade VI proteins in general [11] and for RHE_CH03072 in particular, substrate prediction was difficult to infer from sequence comparisons, since the characterized TtCzrB (Zn^{2+}/ Cd^{2+})[12,13] DR1236 (Mn^{2+}) [10] and Sll1263 (Fe^{2+}) [14] proteins share very close identity percentages (%I) when alignments are done relative to each other: RHE_CH03072/DR1236 (55% I, 89% alignment coverage), RHE_CH03072/Sll1263 (44% I, 90% alignment coverage), and RHE_CH03072/TtCzrB (51% I, 95% alignment coverage).

To get a broader insight into the metal specificity of EmfA, the putative metal binding sites (MBS) A, B or C were examined through multiple sequence alignments (MSA) of clade VI proteins with the Zn^{2+}/Cd^{2+} transporters YiiP (also known as FieF) and TtCzrB whose MBS A, B and C sites have been deduced from crystal structures [13, 15]. The metal discrimination of YiiP is known to reside in the DD-HD residues of site A [16]. In this alignment we also included characterized CDFs with identical metal specificities but classified in different clades according to [11]. Examination of MSA revealed that all proteins grouped in clade VI have identical MBS, in spite of their differences in metal recognition, suggesting that metal specificity does not reside in MBS (Fig. 1, Fig S1). Additionally, the presence of *Saccharomyces cerevisiae* ScMMT1 (Fe^{2+}) or *Stylosanthes hamata* ShMTP1 (Mn^{2+}) in the alignment, and its comparison relative to clade VI proteins, suggests that more than one MBS A composition can be used for the recognition of the same metal.

The phylogenetic reconstruction of clade VI groups EmfA and its homologues in a new subfamily of uncharacterized CDF proteins. Due to the absence of variations in MBS A, B, and C, and the close BLASTP similarity values observed among clade VI proteins, we decided to investigate their functional divergence using a phylogenetic approach. This new phylogeny, inferred from 197 CDF homologues (see methods), classified the proteins into six monophyletic subfamilies with Shimodaira-Hasegawa-like *p*-values ≥ 0.90 supporting the substrate-defined bipartitions (Fig. 2). The putative substrate of three subfamilies was inferred by the presence of Mn^{2+}, Zn^{2+}/Cd^{2+} and Fe^{2+} CDF-transporters previously characterized. The other three are novel subfamilies of uncharacterized proteins; one of them comprised 91 proteins exclusively from α-proteobacteria, including the putative transporter EmfA from *R. etli* CFN42. As we describe below, the functional characterization of EmfA suggested that Mn^{2+} and Fe^{2+} may be the preferred substrates of this group. Interestingly, the uncharacterized subfamily A contains 16 putative cation transporters from β, γ and δ proteobacteria whereas B subfamily contains 32 putative transporters exclusively from Actinobacteria.

The *emfA* gene is essential for Mn^{2+} resistance in *R. etli*. To determine the spectrum of metals to which *emfA* confers resistance, growth of wild type strain and an *emfA* mutant (constructed as described in Materials and Methods) was assessed in minimal medium (MM) plates containing minimal inhibitory concentrations of Mn^{2+}, Zn^{2+}, Cd^{2+}, Cu^{2+}, Ni^{2+} or Co^{2+} for the wild type strain, as previously reported [11]. These assays indicated that growth of the *emfA* mutant was inhibited only by the presence of Mn^{2+}. The mutant recovered its ability to grow in the presence of toxic concentrations of Mn^{2+} upon introduction of the wild type *emfA* gene (Fig. 3).

The e*mfA* gene is upregulated in presence of divalent metals. The metal-dependent expression of *emfA* gene was measured by qRT-PCR in samples of *R. etli* cells grown in MM and exposed for 30 min to Zn^{2+}, Cd^{2+}, Fe^{2+} or Mn^{2+}. The relative expression of the *emfA* gene increased in all tested metals as follows: $Zn^{2+} > Fe^{2+} > Mn^{2+} > Cd^{2+}$, with Zn^{2+} being the best inducer (Fig. 4). The presence of Zn induced the expression of *emfA* 2.5-fold and 3.1-fold compared to the presence of Fe, and Mn, respectively. The expression of *emfA* in Fe^{2+} was almost identical to that observed for Mn^{2+} (Fig. 4). These results foster the possibility that Zn^{2+}, Cd^{2+} and Fe^{2+} could also be substrates of EmfA.

Expression of the *emfA* gene in *E. coli* relieves Fe^{2+} but not Cd^{2+}, Zn^{2+}or Mn^{2+} stress. The upregulation of *emfA* by Zn^{2+}, Cd^{2+}, or Fe^{2+}, in addition to Mn^{2+}, suggests that *emfA* may encode a protein with the ability to confer resistance to these metals; this phenotype may be concealed in the *R. etli emfA* mutant, due to functional redundancy of EmfA with other transporters able to efflux these metals. To test this hypothesis, the *R. etli emfA* gene was introduced by transformation into an *E. coli* Δ*zitB*/*zntA*::Km/Δ*mntP*::Cm triple mutant (strain CC49) and exposed to MICs of Mn^{2+}, Cd^{2+}, Zn^{2+} and Fe^{2+}. Fig. 5 shows that the presence of the wild type *emfA* gene increases 10-fold the number of CFUs found in the presence of Fe^{2+} as compared to cells carrying the empty vector. In contrast, Mn^{2+}, Zn^{2+} and Cd^{2+} resistances were not significantly increased under comparable conditions, despite the fact that *emfA* complemented the Mn^{2+} sensitivity of the *R. etli emfA*- mutant.

The increased tolerance to Fe^{2+} in the *E. coli* background but not in the *R. etli emfA*- mutant supports the proposition that *R. etli* could have additional proteins involved in Fe^{2+} efflux that mask the role of *emfA* for Fe^{2+} tolerance. Conversely, lack of increase in Mn^{2+} tolerance in *E.*

coli by *emfA,* suggests that additional factors present in *R. etli*, but absent in *E. coli*, limit Mn^{2+} recognition.

The Mn^{2+}, Zn^{2+}, Cd^{2+}, and Fe^{2+} recognition by clade VI subfamilies is not dependent on MBS A variations or His-rich stretches. An important challenge in the characterization of metal transporter proteins is the identification of conserved amino acids involved in metal binding and metal selectivity. The metal discrimination of *E. coli* YiiP protein is known to reside in the DD-HD residues of site A. To discern the composition of putative MBS A for proteins from the 4 subfamilies of clade VI, we performed a ConSurf analysis [17] which estimates the evolutionary conservation of amino acid positions in an alignment. This analysis revealed four highly conserved amino acids, EN-HD, which forms part of the putative MBS A from all these subgroups (Fig. S2, ESI). In addition, the full sequence alignment of clade VI proteins showed that they lack identifiable His-rich tracts, similar to those observed in Ni^{2+}/Co^{2+} and Co^{2+} transporters NepA and CepA from clades III, and XII (Fig. S1, ESI). The absence of either MBS A variations or His-rich stretches hampers the identification of residues critical for metal selectivity in clade VI proteins.

It has been suggested that plant Mn-CDF proteins use a DD-DD motif in their MBS A sites for Mn^{2+} recognition [18]; thus, the DXXXD motif located in the putative transmembrane domain (TMD) 5 is commonly used for functional predictions. Previous studies [10] described that DR1236 and MntE shared a DXXXD motif. The authors argued that Mn^{2+} recognition by these proteins is related to this motif. However, the alignments presented in Figs. 1 and S1, show that the MBS A-associated DXXXD motif is absent in clade VI proteins. Also, it is noteworthy that the assigned DXXXD motif in DR1236 [10] is close to TM 6 and contains the D residue involved in dimer

formation (Fig. 1), it does not form part of MBS A; consequently, it is not suitable for prediction of Mn^{2+} as a substrate.

DISCUSSION

The relevance of Mn^{2+} in diverse bacterial physiological processes is exemplified not only by its key role in the defense against oxidative stress, in the regulation of bacterial virulence and as cofactor of critical metabolic enzymes; but also by the redundancy of Mn^{2+} importers present in a cell. At least two high-affinity systems (the ABC transporter SitABCD and the NRAMP homologue MntH) have been characterized in diverse bacteria; but the residual Mn^{2+} uptake activity of such mutants suggests the existence of unspecific Mn^{2+} transporters [19]. The genome of the Mn^{2+} hyper-accumulator *Lactobacillus plantarum* encodes five acquisition systems, three of them specifically upregulated by Mn^{2+} limitation [20]. This versatility to acquire Mn^{2+} should be accompanied by Mn^{+2} efflux pumps able to deal with increased intracellular Mn^{2+} levels, in order to avoid a deleterious Mn^{2+} imbalance.

The phylogenetic reconstruction of clade VI and the characterization of *emfA* gene described in this study represent first steps to unravel the functional and taxonomic diversity of Mn^{2+} exporters. Based in the limited experimental information available for some partially characterized proteins we propose two independent subgroups for Mn^{2+} transporters: the Mn^{2+} subgroup that includes potential Mn^{2+} transporters from the order *Deinococcales* whose only characterized member is *D. radiodurans* DR1236; and the Mn^{2+}/Fe^{2+}, whose first characterized member would be *R. etli* CFN42 EmfA. This subgroup is the most diverse, containing 91 proteins (46% of the protein dataset) from 69 species, all of them belonging to class α-proteobacteria (Table S1, ESI). The Mn^{2+}/Fe^{2+} subgroup, the Mn^{2+} subgroup and the TtCzrB

subgroup are monophyletic to each other (*p*-value=1). This observation suggests that Mn^{2+} and Fe^{2+} transport probably represent an ancestral feature that has persisted in the EmfA subgroup of clade VI proteins. Whether Mn^{2+} is a substrate of TtCzrB [12, 13], Zn^{2+} a substrate of DR1236 [10] or Mn^{2+} can induce Sll1263 [14] remains to be elucidated. The results of this study complement the taxonomic diversity of bacterial Mn^{2+}-CDFs inferred in our previous phylogeny [11] where we reported that clade I groups Mn^{2+} transporters from the phylum firmicutes, with *S. pneumoniae* MntE being the only characterized transporter; and that clade IV exclusively contains Mn^{2+} transporters from plants and fungi [11]. These data, in addition to Mn^{2+} exporters from *Neisseria meningitidis* MntX [3], *Xanthomonas oryzae* YebN [2] and *Escherichia coli* MntP efflux protein [9], belonging to the DUF204 family (Pfam02659), depict a broader view of the functional and taxonomic diversity of Mn^{2+} exporters. However, it is important emphasize that the metal specificities proposed in this study for each group needs to be reinforced with biochemical evidence through *in vitro* transport assays as well as with the characterization of additional members of each clade.

The inability to detect Fe^{2+}, Zn^{2+} and Cd^{2+} sensitivity in *R. etli emfA* mutant could be due to a functional redundancy of EmfA with RHE_CH03719 a putative P_{IB}-type ATPase homolog of *Sinorhizobium meliloti* 2011 SMc04128 (67% I, 99% of alignment coverage) that confers Zn^{2+} and Cd^{2+} resistances [21]. It is currently unknown which protein(s) is involved in Fe^{2+} efflux in rhizobiaceae. Conversely, the increased Fe^{2+} but not Mn^{2+}, Zn^{2+} and Cd^{2+}resistance in *E. coli*, suggest that efflux of these metals is dependent of additional factors absent in this bacterium but present in *R. etli,* such as one still unidentified metallochaperone which might load these metals directly onto EmfA MBSs.

Other way to explain iron resistance of *E. coli* cells harboring *emfA* gene may be presuming that a low Mn^{2+} efflux activity conferred by EmfA to *E. coli*, undetectable by our Mn^{2+} resistance assay, increases Mn^{2+} efflux over influx producing an imbalance in the Mn:Fe ratio. The disproportion between both ions might increase Fe^{2+} pro-oxidant property over Mn^{2+} anti-oxidant activity leading to intracellular accumulation of reactive oxygen species (ROS)[22]. Thus, the iron resistance observed in *E. coli* harboring *emfA* gene may be part of the cell response to control iron-induced ROS instead of an iron efflux ability conferred by EmfA. Perhaps, both iron resistance mechanisms may coexist. Additional evidences as measurements of intracellular content of iron and manganese as well as metal transport assay in vesicles are required to confirm the iron specificity of EmfA.

Regarding metal selectivity, biochemical studies of Zn^{2+} transporters ZnT5 and YiiP have led to the proposal that metal specificity of CDF proteins resides in highly conserved coordinating charged residues located in the MBS A [16]. However, further complexity of metal selectivity is exemplified in our study. At present, it is not clear how discrimination of Mn^{2+}, Mn^{2+}/Fe^{2+}, Zn^{2+}/Cd^{2+} and Fe^{2+} is achieved by the clade VI proteins, as they have identical MBS A, B or C and lack His-rich tracts. Likewise, the fact that clade VI proteins differ in their MBS compositions when compared to other bacterial Mn-, Zn/Cd and Fe-CDFs, such as MntE, YiiP or ScMMT1 is intriguing, it may indicate the existence of different coordination environments to recognize the same metal (Fig. 1). Phylogenetic and MSA analyses, coupled to biochemical and genetic approaches to examine every kind of variations in this protein family, could help in determining the molecular basis of metal selectivity in CDF proteins.

This study is an important contribution to the comprehension of rhizobial Mn^{2+} homeostasis, which had focused mainly on metal uptake. Since EmfA is highly conserved in almost all

analyzed rhizobial genomes, with exception of *B. japonicum*, its characterization offers new insights regarding the Mn^{2+}/Fe^{2+} efflux mechanism in this bacterial group.

EXPERIMENTAL

Data set, multiple sequence alignment and phylogenetic analysis of clade VI CDF proteins and their homologs. The data set analyzed in this study included the previously characterized clade VI- associated [11] CDF proteins TtCzrB [12, 13], Sll1263 [14] and DR1236 [10] as well as EmfA (formerly RHE_CH03072) and their close homologous proteins present in the CDF family available at Pfam (accession PF01545, 12806 sequences). To collect the closest homologues, the 12806 CDF sequences were downloaded from the Pfam data base and filtered to remove redundant (100% identity cut-off) sequences by using CD-hit [23]. The resultant dataset (~8800 sequences) was used to generate a local BLAST database. This database was searched for close BLASTP homologs (80 sequences) for each characterized CDF protein, resulting in a 197 non-redundant sequences dataset (Table S1, ESI). An alignment of the more conserved regions in the 197 sequences dataset was obtained with hmmalign from the HMMER 3.0 package [24] in combination with the HMMER3/b cation efflux family HMM model PF01545.16, downloaded from Pfam [25] and the resultant alignment was used for the phylogenetic analysis. The Maximum-likelihood (ML)-tree strategy was identical to the one recently applied to CDF proteins [11] and included 100 random seed trees in addition to a BioNJ tree to start 101 searches. Tree searching under the Maximum-likelihood criterion was performed with PhyML v3.0 [26] using the LG + G + f model as the substitution matrix with gamma-correction of among-site rate variation [11]. The best tree, shown in Fig. 2A had the highest log-likelihood score from these 101 searches.

The hmmer alignment of 197 sequences was used with ConSurf to estimate the position-specific evolutionary rate of amino acids present in the putative site A, B and C and their neighborhood for clade VI subgroups proteins and compared them with known Zn^{2+}-CDF transporters from clade V as reported [11].

Bacterial strains, media and conditions. The strains are listed in Table S2 (ESI). *Rhizobium etli* strains were grown at 30°C in rich PY medium and minimal medium (MM), while *E. coli* strains were grown in Luria-Bertani medium at 37°C. Antibiotics for *R. etli* were added at the following concentrations (μg/mL): nalidixic acid, 20; streptomycin, 100; gentamicin, 5; spectinomycin, 100; and tetracycline, 3. For *E. coli* the antibiotic concentrations were (μg/mL.): kanamycin, 15; chloramphenicol 15; gentamicin and tetracycline 10.

Genetic Manipulations. The *R. etli emfA* mutant was generated by recombination-based vector integration mutagenesis. An internal 0.3 kbp DNA fragment of *emfA* gene was amplified by PCR with primers A/B, (Table S3, ESI), cloned into suicide plasmid pPDGm [27] and mobilized into *R. etli* CFN42 by triparental mating. Disruption of target gene by single crossover was confirmed by Southern blot hybridization [28]. The *emfA* mutant was complemented with a 955 bp fragment amplified by PCR using primers C/D, first cloned into TOPO TA, subcloned as a *KpnI-XbaI* fragment into pBBR1MMCS-3 and mobilized by triparental mating into the *emfA*- mutant.

The *E. coli* ZitB/ZntA/MntP triple mutant was constructed by using the *E. coli* strain GG48 (Δ*zitB*::Cm, *zntA*::Km) as starting background to add a mutation in the *mntP* (formerly *yebN*) gene by the Datsenko and Wanner procedure [29]. Briefly, the GG48 strain was transformed with the pCP20 plasmid to remove the Cm resistance cassette inserted into the Δ*zitB* gene, and Cm sensitive colonies were obtained. PCR products of 1.1 kb contained the Cm resistance cassette as well as 42 bp flanking sequence of the *mntP* gene were obtained with primers E/F. The PCR products were transformed into the GG48 Cm sensitive strain harboring the red recombinase (pKD46) to obtain a Δ*yebN*::Cm mutant. Then the pKD46 plasmid was removed at 42°C. Finally, the correct insertion of the Cm resistance cassette into the *mntP* gene was verified by PCR and sequencing with primers G/H.

Cation sensitivity assays. Metal sensitivity was determined using a plate assay as follows: 50 mM stock solutions of Fe, Mn, Zn and Cd chloride salts (Sigma-Aldrich, St Louis, MO) were prepared in milli-Q water, filter sterilized and added at increasing concentrations to solid (1.5% wt/vol agar) MM. The *R. etli* overnight cultures were adjusted to OD_{620} = 0.7, washed twice with $MgSO_4$ 10 mM, serially diluted (10^{-1}-10^{-4}) and spotted (20 µl) on solid MM with metal ions as indicated or without them as controls. Rhizobial growth was recorded after 7 days of incubation at 30ºC. For *E. coli* metal sensitivity assays, chemically competent *E. coli* cells were transformed with the respective plasmids, 12 h cultures were grown from single colonies and adjusted them to OD_{550} = 1, washed twice with $MgSO_4$ 10 mM, serially diluted (10^{-1}-10^{-4}) and spotted (15 µl) on solid LB with metal ions as indicated or without them as controls. Growth was recorded after 24 h of incubation at 37ºC

Transcriptional response of *emfA* gene to metals. To avoid potential signal noise produced by rich PY medium, overnight *R. etli* CFN42 cultures were grown in MM supplemented with thiamin (1 mg l^{-1}), biotin (1 mg l^{-1}) and 0.2% casaminoacids, adjusted at DO_{620}= 0.05 and allowed to grow up to OD_{620} = 0.45 - 0.55 in the same medium. Then, 10 ml of these cells were exposed during 30 min to 5 mM of $FeCl_3$, $ZnCl_2$, $CdCl_2$ and $MnCl_2$. Control cells unexposed to metals were also included. After this time, mRNA was extracted by using TriPure isolation reagent (Roche). The total RNA (DNA free) was reverse transcribed to cDNA by using ReverAid H minus FirstStrand cDNA Synthesis (Fermentas). Quantitative real-time PCR was performed on PCR System 3700 (Applied Biosystems) using Maxima Syber Green/ROX qPCR master Mix (Fermentas). The *emfA* and *hisCd* genes were amplified by using I/J and Y/Z primers (Table S3, ESI), respectively. Their expression levels in the presence and absence of metals were normalized to the expression level of housekeeping *hisCd* gene. The data represent the average

of four independent experiments with three technical replicates each. The fold change in gene expression was calculated using the $\Delta\Delta C_T$ method as reported [11].

CONCLUSIONS

In this work, the RHE_CH3072 gene of *R. etli* encoding for a cation diffusion facilitator protein was shown to be responsible of Mn^{2+} tolerance in this organism. Thus, RHE_CH3072, now denominated EmfA, is the first member characterized of a novel subfamily of α-proteobacteria Mn^{2+}/Fe^{2+} CDF transporters. This claim is supported by genetic, phylogenetic and gene expression analyses in a number of conditions and genetic backgrounds.

The Mn^{2+} resistance determined by EmfA is to be present in at least 60 α-proteobacterial species, including those from rhizobiaceae; providing insights about how this group of soil organisms manages Mn^{2+}.

The sequence analyses we carried out comparing proteins belonging to clade VI subgroups revealed no differences in their MBS, implying the existence of other still unknown selectivity determinants responsible of their observed functional divergency.

ACKNOWLEDGMENTS

We gratefully acknowledge Professor D.H. Nies for sending us his E. coli strains. We also thank Susana Brom for critically reviewing the manuscript, Laura Cervantes and Josue Ocelotl for their skillful technical assistance. We thank Programa de Apoyo a Proyectos de Investigación e Inovación Tecnológica (PAPIIT Grant number IN201112 to A.G.S),Consejo Nacional de Ciencia y Tecnologia (Mexico) for the PhD scholarship given to C.C. (CVU 269108) and the Ciencia Basica Grant (179133) to P.V. for financial support of this research.

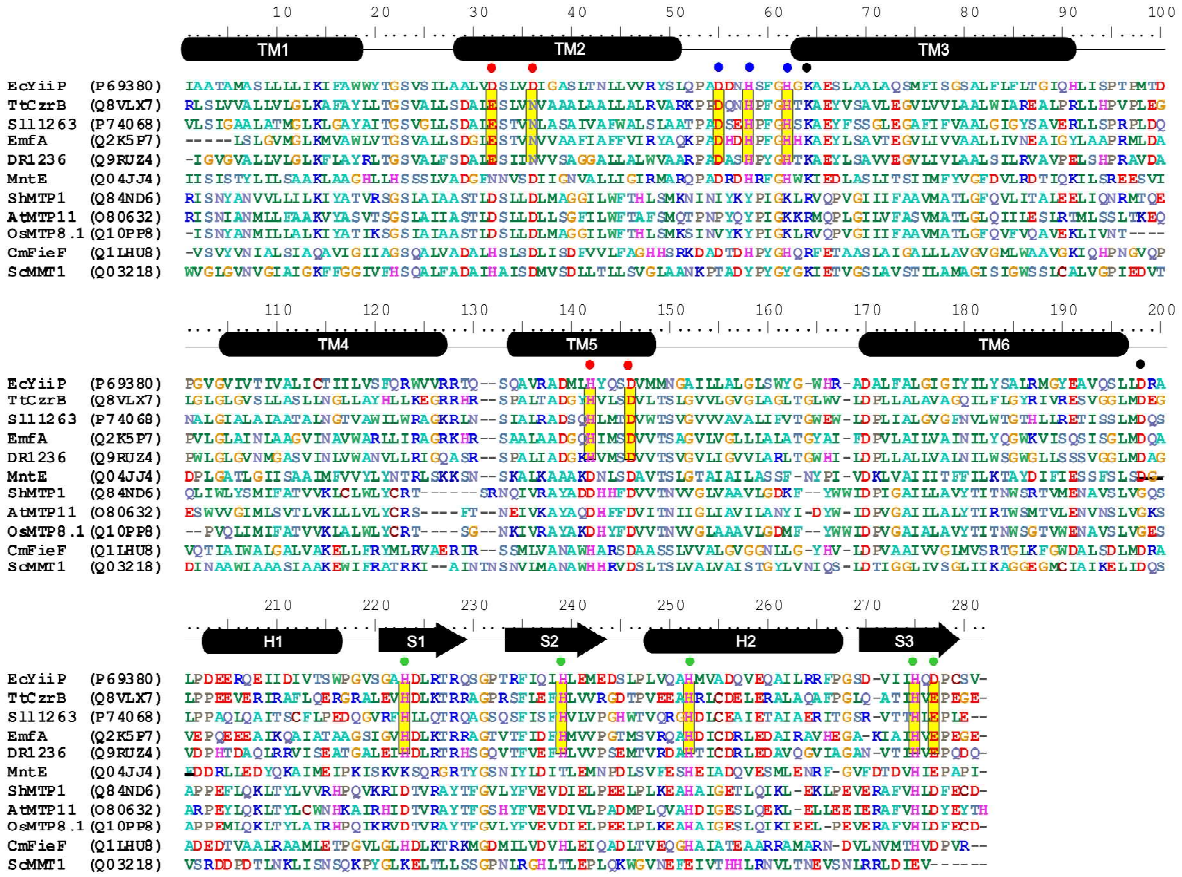


Fig. 1. Multiple sequence alignment of clade VI proteins shows identical MBS A, B and C in spite of their different metal substrates. Amino acid sequences from *R. etli* EmfA, clade VI characterized proteins (*D. radiodurans* DR1236, *T. termophilus* CzrB and *Synechocystis sp.* Sll1263) and characterized transporters having similar substrates but from other clades [11] such as *E. coli* YiiP (Fe, Zn/Cd), *Stylosanthes hamata* MTP1 (Mn), *Arabidopsis thaliana* MTP11 (Mn), *Oryza sativa* MTP8.1 (Mn), *Cupriavidus metallidurans* FieF (Fe, Cd, Zn, Ni, Co), and *Saccharomyces cerevisiae* MMT1 (Fe) were aligned with hmmalign (HMMER). Only conserved domains of CDF proteins obtained by comparison with the CDF family consensus (CDF-HMM profile) are shown. The red, blue, green and black dots indicate the MBS A, B, C and the interlocked Lys^{77}-Asp^{207} salt bridge respectively, as reported for *E. coli* YiiP (PDB 3H90) and *T. termphilus* CzrB (PDB 3BYR) Zn-bound protein structures. Conserved residues in putative MBS

A, B and C of clade VI proteins are highlighted. The erroneous DXXXD motif common to MntE and DR1236 previously reported [10] and assumed to be part of MBS A is underlined.

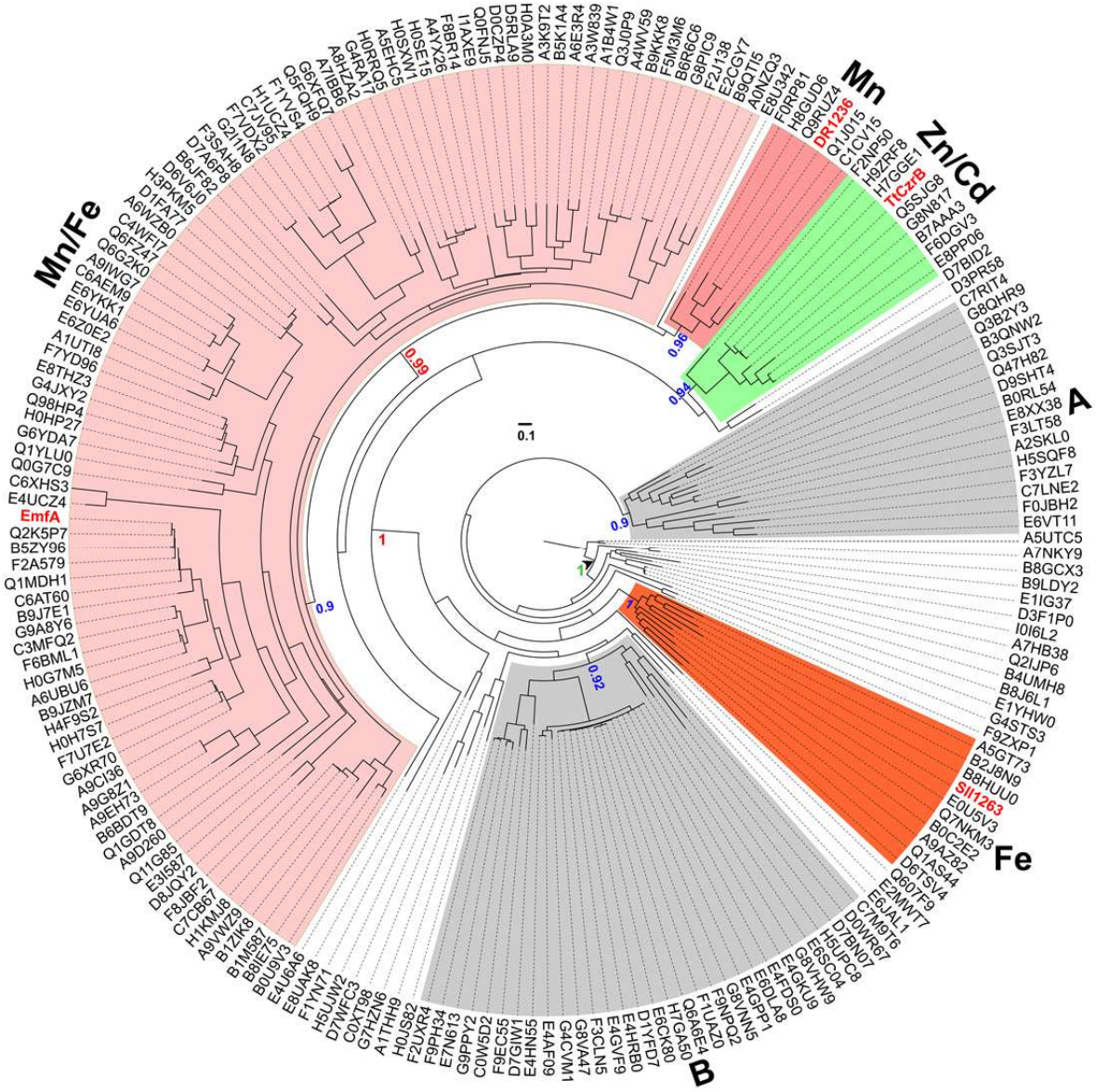


Fig. 2. The Maximum-likelihood phylogenetic tree of 197 CDF proteins, homologous of those grouped into clade VI, classifies them into six monophyletic substrate-defined subfamilies. Subfamilies of uncharacterized proteins are indicated with A and B. Green *p*-value (Shimodira-Hasegawa-like approximate LRT test) indicates a monophyletic origin for all clade VI proteins. Red *p*-value supports the monophyletic origin of Mn^{2+}/Fe^{2+}, Mn^{2+}, Zn^{2+}/Cd^{2+} protein subfamilies. Blue *p*-values support bipartitions for substrate-defined clades (*p*-values ≥0.9). The scale bar indicates the expected number of amino acid substitutions per site under the LG model.

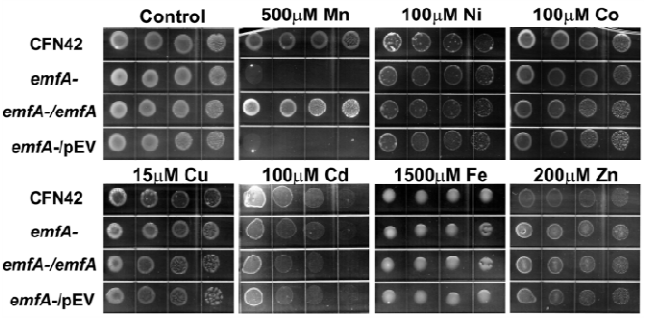


Fig. 3. The *emfA* gene is essential for *R. etli* Mn^{2+} resistance. Cation sensitivity plate assay showing the growth of serial dilutions of wild type (CFN42), mutant (*emfA*-), mutant complemented with the wild type gene *(emfA-/emfA*) and mutant with the empty vector (*emfA-/p*EV) in the presence or absence (control) of metals, as indicated. Growth inhibition of *emfA*- and *emfA-/p*EV mutants in the presence of 500 μM $MnCl_2$; and full growth of complemented mutant *emfA-/emfA* shows that *emfA* gene is essential for Mn^{2+} resistance. Serial dilutions 10^{-1}-10^{-4} (left to right) are shown.

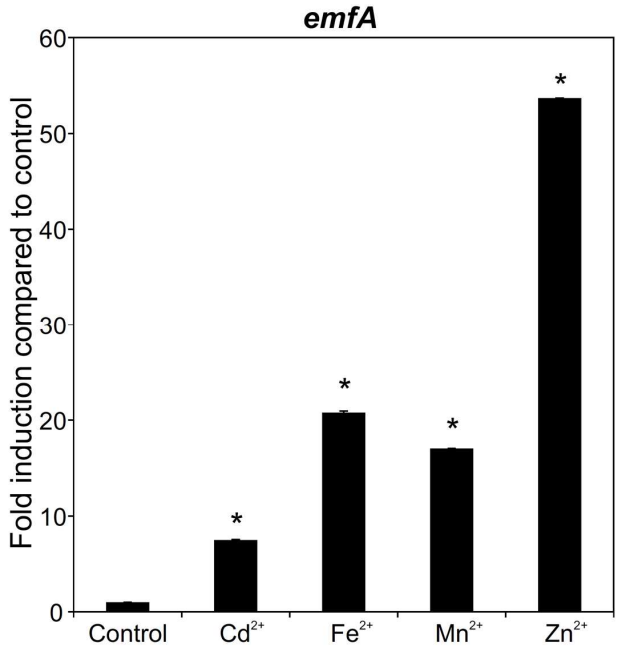


Fig. 4. Transcription of *R. etli emfA* gene is upregulated by Zn^{2+}, Mn^{2+}, Cd^{2+} and Fe^{2+}. Data represent the fold induction, defined as the ratio between the mRNA levels of *emfA* in the presence of 5 mM of $FeCl_3$, $ZnCl_2$, $CdCl_2$ and $MnCl_2$ and mRNA levels in the absence of metals (control). Both were normalized to *hisCd* mRNA levels as published [11] (mean ± s.d., n = 3) (*t*-test, **P≤0.01).

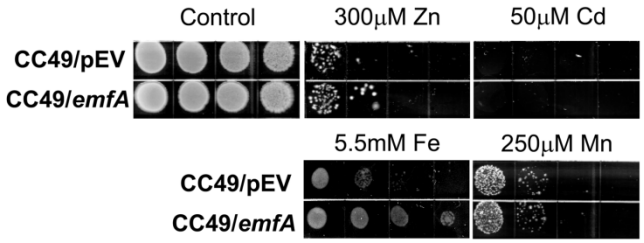


Fig. 5. The expression of *R. etli emfA* gene in *E. coli* CC49 mutant (Δ*zitB*/*zntA*::Km/Δ*mntP*::Cm) partially relieves Fe^{2+} but not Mn^{2+}, Zn^{2+}, or Cd^{2+} stress. Cation sensitivity plate assay showing the growth of serial dilutions (10^{-1} to 10^{-4} left to right) of *E. coli* CC49 mutant transformed with the empty pBBR-MSC3 (pEV) plasmid or with the same plasmid containing the wild type *emfA* gene.

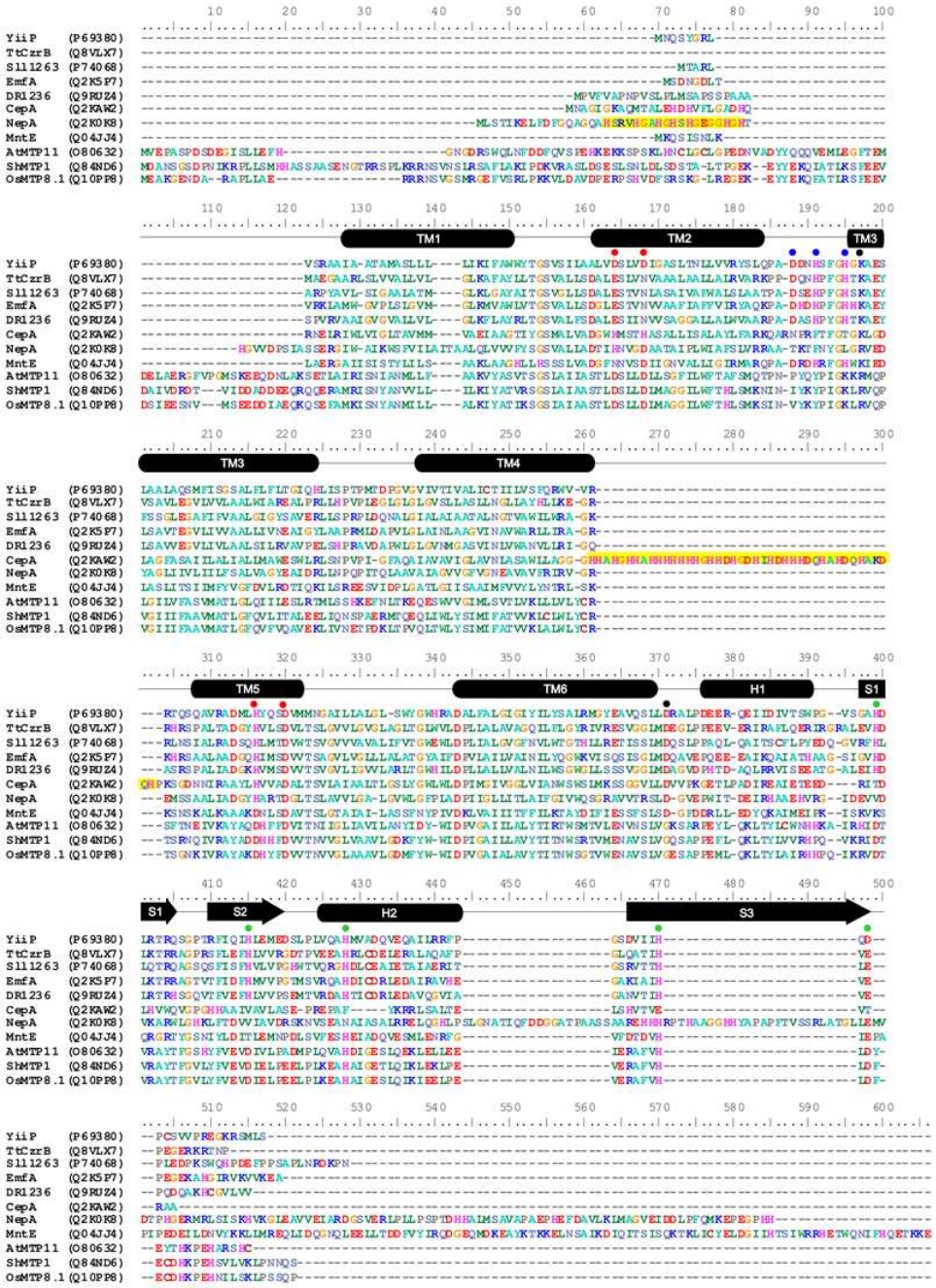


Fig. Suppl. 1. Absence of His-rich tracts in clade VI proteins. Full-length amino acid sequences of clade VI proteins, Mn^{2+}- and Fe^{2+}-CDFs from diverse origins and NepA and CepA as representative Ni^{2+}/Co^{2+} CDF with His-rich tracts, as previously reported [11], were aligned with MUSCLE. The His-rich regions exclusively present in NepA and CepA CDF proteins are highlighted.

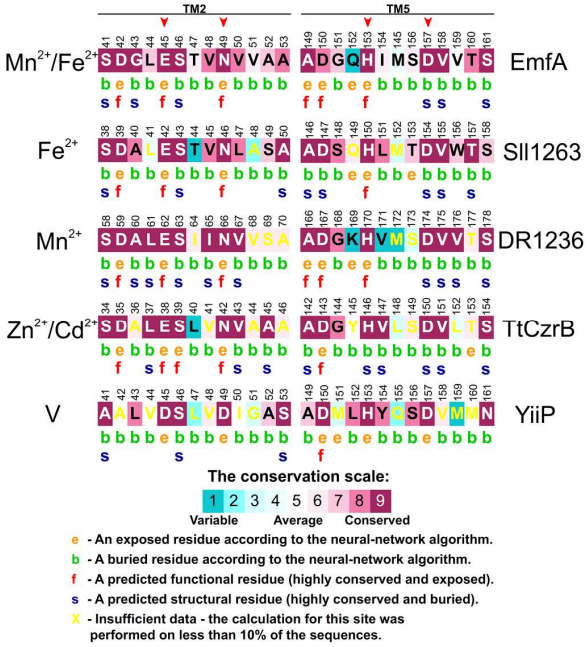


Fig. Suppl. 2. Conserved residues of putative MBS A are identical among proteins of clade VI subfamilies. Proteins from each subfamily were collected and aligned with hmmalign. The degree of evolutionary conservation of amino acid positions in related sequences was analyzed with ConSurf (see methods); such conservation degree is reflected by the color scale. For reference Zn/Cd-CDF proteins from the YiiP (V) group [11] were also included. Numbering correspond to proteins indicated on right.

References

1. A. Anjem, S. Varghese and J. A. Imlay, *Mol Microbiol*, 2009, 72, 844-858.
2. C. Li, J. Tao, D. Mao and C. He, *PLoS One*, 2011, 6, e21983.
3. F. J. Veyrier, I. G. Boneca, M. F. Cellier and M. K. Taha, *PLoS Pathog*, 2011, 7, e1002261.
4. J. W. Rosch, G. Gao, G. Ridout, Y. D. Wang and E. I. Tuomanen, *Mol Microbiol*, 2009, 72, 12-25.
5. B. W. Davies and G. C. Walker, *J Bacteriol*, 2007, 189, 2101-2109.
6. E. Diaz-Mireles, M. Wexler, G. Sawers, D. Bellini, J. D. Todd and A. W. Johnston, *Microbiology*, 2004, 150, 1447-1456.
7. T. H. Hohle and M. R. O'Brian, *Mol Microbiol*, 2009, 72, 399-409.
8. T. C. Chao, A. Becker, J. Buhrmester, A. Puhler and S. Weidner, *J Bacteriol*, 2004, 186, 3609-3620.
9. L. S. Waters, M. Sandoval and G. Storz, *J Bacteriol*, 2011, 193, 5887-5897.
10. H. Sun, G. Xu, H. Zhan, H. Chen, Z. Sun, B. Tian and Y. Hua, *BMC Microbiol*, 2010, 10, 319.
11. C. Cubillas, P. Vinuesa, M. L. Tabche and A. Garcia-de Los Santos, *Metallomics*, 2013, 5, 1634-1643.
12. S. Spada, J. T. Pembroke and J. G. Wall, *Extremophiles*, 2002, 6, 301-308.
13. V. Cherezov, N. Hofer, D. M. Szebenyi, O. Kolaj, J. G. Wall, R. Gillilan, V. Srinivasan, C. P. Jaroniec and M. Caffrey, *Structure*, 2008, 16, 1378-1388.
14. H. B. Jiang, W. J. Lou, H. Y. Du, N. M. Price and B. S. Qiu, *Plant Cell Physiol*, 2012, 53, 1404-1417.
15. M. Lu, J. Chai and D. Fu, *Nat Struct Mol Biol*, 2009, 16, 1063-1067.
16. E. Hoch, W. Lin, J. Chai, M. Hershfinkel, D. Fu and I. Sekler, *Proc Natl Acad Sci U S A*, 2012, 109, 7202-7207.
17. F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz and N. Ben-Tal, *Bioinformatics*, 2003, 19, 163-164.
18. B. Montanini, D. Blaudez, S. Jeandroz, D. Sanders and M. Chalot, *BMC Genomics*, 2007, 8, 107.
19. D. G. Kehres and M. E. Maguire, *FEMS Microbiol Rev*, 2003, 27, 263-290.
20. M. N. Groot, E. Klaassens, W. M. de Vos, J. Delcour, P. Hols and M. Kleerebezem, *Microbiology*, 2005, 151, 1229-1238.
21. S. Rossbach, D. J. Mai, E. L. Carter, L. Sauviac, D. Capela, C. Bruand and F. J. de Bruijn, *Appl Environ Microbiol*, 2008, 74, 4218-4221.
22. J. P. Lisher and D. P. Giedroc, *Frontiers in cellular and infection microbiology*, 2013, 3, 91.
23. W. Li and A. Godzik, *Bioinformatics*, 2006, 22, 1658-1659.
24. S. R. Eddy, *Genome Inform.*, 2009, 23, 205-211.
25. R. D. Finn, J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman, *Nucleic Acids Res*, 2008, 36, D281-288.
26. S. Guindon and O. Gascuel, *Syst Biol*, 2003, 52, 696-704.
27. J. M. Martinez-Salazar and D. Romero, *Gene*, 2000, 243, 125-131.
28. M. Flores, V. Gonzalez, S. Brom, E. Martinez, D. Pinero, D. Romero, G. Davila and R. Palacios, *J Bacteriol*, 1987, 169, 5782-5788.

29. K. A. Datsenko and B. L. Wanner, *Proc Natl Acad Sci U S A*, 2000, 97, 6640-6645.