LINKING GENE WITH FUNCTION: ANALYSIS AND INVESTIGATION OF NOVEL PLAYERS IN BACTERIAL ZINC HOMEOSTASIS

By

CRYSTEN ELIZABETH BLABY-HAAS

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To my family

ACKNOWLEDGMENTS

I would like to thank Dr. Valérie de Crécy-Lagard for giving me the opportunity and support to work on this project, which has sparked my interest in microbial metal metabolism. I would also like to thank the members of my doctoral committee both past and present, Dr. Keelnatham T. Shanmugam, Dr. James F. Preston, Dr. Graciela L. Lorca, Dr. Robert J. Cousins and Dr. Thomas J. Lyons, for their invaluable advice, guidance and time. I would like to thank the faculty and staff of the Microbiology and Cell Science Department, specifically John Rice, Chris Gardner and Fernando Pagliai. I would like to thank present and past graduate students from MicroCell, particularly Dr. Franz St. John, who was my original research mentor and taught me almost everything I know. I would like to thank Dr. Phi Min Do for assistance with hydrogenase and glyoxalase assays. I would like to thank Dr. Shou-Mei Chang and Dr. Charles Guo. I would like to thank Drs. Richard Gourse, Herbert Schweizer, Shouguang Jin and Chris Rensing for strains and/or plasmids. I would like to thank Dr. Zhonglin Mou for use of his electroporator and UV-crosslinker. I would like to thank Dr. Eric Triplett for using his restriction enzymes from time to time. I would like to thank Dr. Nemat Keyhani for allowing me to borrow amino acids. I would like to thank the staff at the ICBR sequencing and mass spec core as well as Dr. George Kamenov for ICP-MS analysis. I would like to thank Dr. Maupin for supervising me during my PhD rotation and her lab members, especially Dr. Katie Rawles and Dr. Matthew Humbard for advice. I am grateful for the advice and support of my lab members, Gabriella Philips, Patrick Thiaville, Dr. Marc Bailley, Kevin Gulig and Dr. Basma El Yacoubi. I would like to thank my many collaborators, with whom I have been most privileged to publish with, Dr. Dmitry Rodionov, Dr. Sabeeha Merchant, Dr. Janette Kropat, Dr. Davin Malasarn, Dr.

Irina Artsimovitch, Ran Furman, Dr. Deborah Zamble and Jessica Flood. I am very grateful to Dr. Nigel Robinson and Dr. Kevin Waldron for allowing me to be a squatter in their lab and showing me how to fractionate *E. coli* for determination of metal speciation and analyzing samples by ICP-MS.I would like to thank my family for constant support and sympathy. I would especially like to thank my part-time lab partner/ part-time husband, Dr. Ian Blaby, whose continued support and guidance I could not have done without.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

%	percent
(I)	oxidation state of 1+
(II)	oxidation state of 2+
(III)	oxidation state of 3+
mAmp	mili Amplitude
BLAST	Basic Local Alignment Search Tool
°C	Degrees Celsius
Co	Cobalt
CO_2	Carbon dioxide
COG	Cluster of Orthologous Groups
Cu	Copper
Da	Dalton (atomic mass unit)
DTT	Dithiothreitol
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
Fe	Iron
FLP	Flippase
FRT	Flippase Recognition Target
g	gram
GTP	Guanosine 5’-triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hg	Mercury
hrs	hours
ICP-MS	Inductively-Coupled Plasma-Mass Spectrometry (Spectrometeter)

IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
L	Liter
M	Molar
min	minute
mg	milligram
Mn	Manganese
Mo	Molybdenum
N	Normal
NaOH	Sodium hydroxide
Ni	Nickel
NTP	Nucleoside triphosphate
PAGE	Polyacrylamide gel electrophoresis
Pb	Lead
RNAP	RNA-polymerase
SDS	Sodium dodecyl sulfate
TPEN	*N*,*N*,*N'*,*N'*-Tetrakis(2-pyridylmethyl)-ethylenediamine
UV	Ultraviolet
V	Volt
v/v	volume per volume
WT	Wild-type
w/v	weight per volume
Zn	Zinc

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

LINKING GENE WITH FUNCTION: ANALYSIS AND INVESTIGATION OF NOVEL PLAYERS IN BACTERIAL ZINC HOMEOSTASIS

By

Crysten Elizabeth Blaby-Haas

May 2011

Chair: Valérie de Crécy-Lagard
Major: Microbiology and Cell Science

When deprived of the essential trace nutrient zinc, bacteria are known to adapt by increasing the intracellular concentration of zinc through transport or by lowering the zinc quota of the cell through the use of zinc-independent back-up proteins. These mechanisms were discovered by functional and computational analyses of Zur regulons. A recent computational analysis of Zur regulons from newly sequenced genomes suggested the existence of genes in addition to those presently known that may be involved in the adaptive response to zinc-depletion. Of these novel genes the role of COG0523 genes and paralogs of zinc-dependent proteins are investigated and discussed.

COG0523 proteins are, like the nickel chaperones of the UreG and HypB families, part of the G3E family of GTPases, strongly suggesting a link to metallocenter biosynthesis. Even though the first COG0523-encoding gene, *cobW*, was identified almost 20 years ago, little is known concerning the function of this ubiquitous family. Therefore, a phylogenomic study of the COG0523 family was performed, leading to the separation of the family into fifteen subgroups. To further probe the function of the COG0523 family, *yeiR* from *Escherichia coli* was chosen to test the apparent link

between some COG0523 genes and zinc homeostasis. This previously uncharacterized gene was found to be involved in survival during EDTA or cadmium challenge and was linked to optimal growth during zinc-depletion.

In addition to the canonical DksA protein, the *Pseudomonas aeruginosa* genome encodes a closely-related paralog DksA2 that lacks the Zn-finger motif and whose gene was predicted to be regulated by Zur. A study on the role of *dksA2* and verification of its regulation by zinc and Zur was performed. *dksA2* was found to be able to functionally substitute for the canonical *dksA*, *in vivo*. Expression was repressed by the presence of exogenous zinc, deletion of Zur resulted in constitutive expression, and Zur bound specifically to the upstream region of *dksA2*. This data suggests that DksA2 plays a role in zinc homeostasis and serves as a back-up copy of the canonical zinc-dependent DksA in zinc-poor environments.

CHAPTER 1
LITERATURE REVIEW

Introduction

It is estimated that twenty elements are required for life and that these elements were likely required by primitive life (Williams, 1997). Traditionally, biology students are taught the importance of the “organic” elements, carbon, hydrogen, nitrogen, oxygen, phosphorous and sulfur, with little mention of the “inorganic” elements, especially transition metals that are also of great importance. With the exception of iron, these metals are often found in trace amounts, causing their identification and characterization in cells difficult. For instance, the requirement of nickel by a protein was not recognized until 1975 (Dixon *et al.*, 1975) and it wasn’t until 2010 that researchers found that at least one-third of metal-bound proteins in the cell are uncharacterized or mischaracterized (Cvetkovic *et al.*, 2010).

Because of the breadth and importance of metals in biology and implications on health, the study of these protein cofactors has called upon multiple disciplines ranging from physical to biological sciences. Inorganic chemists, biochemists, biologists and bioinformaticians have all added their particular expertise and personal perspective on the use and evolution of these elements in biology. Organic chemists seek understanding through the creation of synthetic analogs of biomolecule metal sites. Biochemists strive to understand reaction mechanisms. Biologists aspire to understand homeostasis and how the geological history of earth has shaped life. In the last decade, the benefit of bioinformatics to the study of metals in biology has also become apparent. By drawing upon accumulated experimental and genomic data, the field of bioinformatics has generated new avenues of research.

Trace elements and their bioavailability across geological time have undoubtedly had a large impact on the evolution of life. Sometimes complicated and intricate machinies have had to evolve to carefully control their concentration and location within the cell. Many of these metals are essential but at the same time high concentrations can be toxic; maintenance of their concentration within a defined window is essential. The details concerning these homeostatic systems are only now being addressed and many questions remain to be answered. In addition, mechanisms are being discovered that are causing researchers to reevaluate some of the assumed interactions between metals and life processes.

In the last decade, researchers discovered that like the more toxic metals iron and copper, the intracellular concentration of zinc in bacteria is highly regulated and a “free pool” of zinc does not exist. This conclusion indicates that once imported zinc is immediately bound to proteins or small molecules. Therefore, an open question with regard to zinc metabolism is how are newly synthesized proteins able to acquire zinc? This question has spurred an investigation into how zinc is trafficked within the cell and if a hierarchy of metal acquisition exists under conditions where zinc is limiting.

The discovery of metal trafficking chaperones at the end of the 20th century has spurred a renewed interest in metallocenter biosynthesis. Are all metals trafficked within the cell by metal-specific chaperones? Do all metal-dependent enzymes have an affiliated chaperone to aid in metal acquisition? Simple questions such as to the intracellular concentration of metals and the abundance of metal-dependent proteins in the cell are just now being addressed.

The biology of essential and non-essential trace mineral elements with special attention on zinc is the topic of this review. The current understanding on why the activity of some proteins is absolutely dependent on the presence of a metal ion and how proteins acquire that specific metal to the exclusion of others will be covered. Lastly, the contribution of bioinformatics to the understanding of gene function and how that relates specifically to metal metabolism will be discussed. The ability of bioinformatics to look at all genomes simultaneously has had and will continue to have an impact on uncovering the answers to open questions concerning metal metabolism in prokaryotes. In particular, a combination of bioinformatics with more traditional molecular biology approaches has opened the door to a revolution in the understanding of metals in biology.

Transition Metals in Biology

The period four transition metals cobalt, copper, iron, manganese, nickel and zinc play important roles in chemical reactions that are requisite for known life. A recent and popular estimate suggests that at least one-third of all proteins are metalloproteins and as such require a metal cofactor for either maintenance of structure, catalysis or regulation (Tottey *et al.*, 2007). This and similar estimates are based mainly on surveys of three-dimensional structures that take into account the presence of the non-transition metals, calcium and magnesium (Andreini *et al.*, 2008b). A recent bioinformatic analysis suggests that zinc, non-heme iron and copper proteins constitute roughly 10% of bacterial and eukaryotic proteomes and 13% of archaeal proteomes (Andreini *et al.*, 2006, Andreini *et al.*, 2007, Andreini *et al.*, 2008a). All six Enzyme Commission classes contain proteins that are strictly-dependent on the presence of a transition metal cofactor (Andreini *et al.*, 2008b).

The prevalence of each metal in catalysis differs among the enzyme classes and can also vary from one organism to the next. Currently, iron and zinc are the most prevalent transition metal cofactors in nature (Andreini *et al.*, 2008b). Only a handful of nickel-dependent enzymes are known (Dixon *et al.*, 1975, Drake *et al.*, 1980, Ellefson *et al.*, 1982, Moura *et al.*, 1984, Youn *et al.*, 1996, Clugston *et al.*, 1998). Specifically, higher organisms such as plants encode only one known nickel-dependent enzyme, urease (Gerendas *et al.*, 1999); a nickel-dependent protein is yet to be discovered in vertebrates (Zhang *et al.*, 2009). Roughly twenty manganese-dependent enzymes and several corrin, non-corrin and porphyrin-bound cobalt dependent enzymes are also known (da Silva and Williams, 2001). A comprehensive understanding on the extent to which metals are used in biology will certainly increase as techniques are developed to directly survey the proteome and thereby lead to the discovery of novel metalloproteins.

Identification of Metal-Dependent Proteins

The seminal studies on the metalloproteomes of *Synechocystis* PCC 6803 and *Pyrococcus furiosis* highlight the feasibility and richness of such analyses (Tottey *et al.*, 2008, Cvetkovic *et al.*, 2010). Under non-denaturing conditions, the proteome of a cell can be fractionated by liquid chromatography and subsequently analyzed for metals by inductively coupled plasma mass spectrometry (ICP-MS). By coupling this technique to high-throughput tandem mass spectrometry (HT-MS/MS), 158 potentially novel metalloproteins were identified in *P. furiosis* (Cvetkovic *et al.*, 2010). Additionally, metals were found that were previously not known to be incorporated by this organism, or the metal bound to specific proteins was different from the metal predicted based on sequence similarity.

These ICP-MS techniques have the benefit of identifying the native metal cofactor, which is often not obvious. Two hurdles currently exist in identifying metalloproteins: cambialism and reliability on sequence homology to characterized metalloproteins to predict metal-dependence. Several enzymes are cambialistic (Martin *et al.*, 1986), which means that *in vitro* they display comparable activity with more than one metal, leading to ambiguity as to the native *in vivo* cofactor (Neu, 1967, Rajagopalan *et al.*, 1997, D'souza and Holz, 1999, Proudfoot *et al.*, 2004, Zheng *et al.*, 2005). This ambiguity is compounded as in some cases, cambialism is a true mechanism to ensure enzyme activity under metal-deficient growth conditions (Yee and Morel, 1996). In other cases, an enzyme purified under non-native conditions will co-purify with the "wrong" metal and/or *in vitro* experiments will show that multiple metals can activate the enzyme.

A prime example is the identification of the native cofactor of the γ-class carbonic anhydrases. The prototype of this class is from the anaerobic prokaryote *Methanosarcina thermophila*. The X-ray structure of this enzyme, purified aerobically in *Escherichia coli*, revealed the presence of a zinc ion (Kisker *et al.*, 1996) and dependence on zinc was also verified *in vitro* (Alber *et al.*, 1999). These results were expected as carbonic anhydrases from the other known classes are classic zinc-dependent enzymes (Christianson and Fierke, 1996). However, when the enzyme was reconstituted with iron, activity exceeded that of the zinc-form (Tripp *et al.*, 2004) and when purified natively from *M. thermophila* it contained iron, not zinc (Macauley *et al.*, 2009). Therefore, the γ -class of carbonic anhydrases is most certainly composed of iron-dependent enzymes. As the vast majority of metalloenzymes are over-expressed

and purified under non-native conditions, usually in *E. coli*, identification of the “wrong” metal is a real concern (Jacob *et al.*, 1998).

The second hurdle to cofactor identification is that sequence similarity is not necessarily reliable for assigning a specific metal to a metalloprotein, although attempts are frequently made (Zhang and Gladyshev, 2009, Zhang and Gladyshev, 2010). The ArsR/SmtB family of metal proteins is a key example of the unreliability of sequence homology. This family is composed of metal-sensing transcription factors that affect the transcription of target genes based on the concentration of a specific metal in the cell (reviewed in (Busenlehner *et al.*, 2003)). ArsR/SmtB homologs form eight distinct groups based on overall sequence similarity, however, membership in a group does not correlate with the specific metal sensed (Campbell *et al.*, 2007). For instance, the cadmium-sensing regulators from Firmicutes have higher overall similarity to the arsenic-sensing regulators of Firmicutes than to cadmium-sensing regulators from other organisms. The only way to correctly predict which metal is bound is by the detection of specific sensory motifs (Campbell *et al.*, 2007). Motif searching is not a standard protocol in the annotation of genomes. Therefore, it is easy to see how misannotations in databases arise.

However in some cases, even motifs are not sufficient in predicting the *in vivo* cofactor. Two different enzymes can have the same metal-binding motif but the native metal cofactor is different, as is the case for MncA and CucA in *Synechocystis* PCC 6803. The metal-binding ligands are exactly the same in each protein but manganese is the *in vivo* cofactor of MncA and copper is the *in vivo* cofactor of CucA (Tottey *et al.*, 2008).

To a certain extent, assignment of metal cofactors by sequence similarity and motifs can be strengthened by other comparative genomic techniques such as co-occurrence with the necessary metal-transporters or by the presence of regulatory sites upstream of the corresponding gene (Zhang *et al.*, 2009, Rodionov *et al.*, 2006b). If a protein is predicted to bind cobalamin based on sequence similarity but both known cobalamin transporters and biosynthetic pathways are missing from that organism, then most likely the protein is not dependent on cobalamin. Targeted experimental verification of this discrepancy can led to either identification of a novel cofactor or novel transporters. A better understanding of how metal-allocation is dictated should translate into improved predictive tools.

Based on the fundamental issues in the identification of metals in biology, metalloproteomic approaches such as those discussed above will inevitably advance the community's understanding of metal speciation (i.e. identity and abundance of metal complexes) by providing vital clues as to the identity of metalloproteins and their native cofactors. Unfortunately, like most high-throughput techniques, metalloproteomic approaches give only a snapshot of the cell's constituents. The abundance and identity of a cell's metalloproteins are dynamic and in constant flux. Microorganisms are continually faced with varying environmental conditions and the cell's response to each situation can be drastically different, resulting in induction and/or repression of different gene sets. These gene sets may be composed of different types of metal-dependent proteins. Additionally, introduction to a novel environment can result in the need to adapt to a different bioavailbility of metal ions.

The Dynamic Metallome

The metabolic state of the cell and metal availability contribute to a dynamic view of the metallome (element composition of the cell). The specific growth conditions of the cell can have a large impact on the abundance of metalloproteins and consequently the abundance of metals in the cell; the needs of the cell will determine the presence and abundance of certain metal-dependent proteins. In particular, the absence of oxygen appears to have a major impact on the metallome of *E. coli* leading to increases in copper and nickel accumulation (Outten *et al.*, 2001b, Rowe *et al.*, 2005). Anaerobic growth of *E. coli* leads to production of [NiFe]-hydrogenase (Ballantine and Boxer, 1985) and induction of nickel import (Rowe *et al.*, 2005). Previous work has also shown that anaerobic *E. coli* cells accumulate 20-fold higher copper than aerobically grown cells (Outten *et al.*, 2001b). Perhaps, this phenomenon is due to reduction of copper to Cu(I) under anaerobic conditions, which can more easily diffuse through the membrane (Rensing and Grass, 2003).

The availability of a specific metal in the environment can also lead to a change in the metallome beyond a decrease or increase of that metal. The activity of multiple metalloproteins (many of which are essential proteins) is strictly dependent on the presence of a metal and in most cases a specific metal. Mechanisms have evolved that allow the cell to switch between metal cofactors (by switching metalloprotein use) to carry out a specific reaction depending on the availability of that metal in the environment. The genomes of some organisms encode at least two proteins that can catalyze the same reaction but employ different metals (Merchant *et al.*, 2006, Panina *et al.*, 2003). During iron-limitation, manganese-dependent superoxide dismutase (MnSOD) can take the place of iron-dependent superoxide dismutase (FeSOD)

(Niederhoffer *et al.*, 1990). During copper-limitation, iron- dependent cytochrome c_6 can replace copper-dependent plastocyanin (Merchant and Bogorad, 1986b). Backup enzymes have also been described for reactions that are dependent on cobalt (Rodionov *et al.*, 2003), nickel (Kim *et al.*, 2000), molybdenum (Jacobitz and Bishop, 1992) and zinc (Gabriel and Helmann, 2009, Sankaran *et al.*, 2009). Therefore, limitation of one metal can lead to a drastic difference in the profile of metals in the cell.

Comparative metallomics is a recently coined term that refers to an emerging field focused on the qualitative and quantitative measurement of a cell's metallome under different growth conditions (Szpunar, 2004). This approach can also be applied to assaying the difference between a mutant and the isogenic parent. Currently, due to technical limitations these studies have focused on measuring total metal content. Currently, the scientific breakthroughs are in simply discovering the identity of metalloproteins and the native cofactor under standard laboratory conditions. As the field progresses, comparative metallomics will likely become a feasible and routine approach in the same way the use of DNA microarrays and comparative transcriptomics have become routine for many research groups.

Allocation of Metals within the Cell

The existence of metallloprotein back-up copies underscores the redundancy of metal-catalyzed chemistry; two metals can catalyze the same reaction. Except for the characterized cases on cambialism, metalloproteins, however, are generally specific to one metal or the other. Even when a protein is genuinely cambialistic, not all metal ions are sufficient for activity and a spectrum of activity is observed from one metal to another. The carbonic anhydrase from *M. thermophila* has the highest activity with iron, followed by cobalt then zinc; the presence of copper, manganese, nickel and cadmium

result in less than 10% activity (Tripp *et al.*, 2004). Superoxide dismutase has two well characterized isoforms, MnSOD and FeSOD; manganese-substituted FeSOD (Yamakura, 1978, Yamakura and Suzuki, 1980) and iron-substituted MnSOD (Ose and Fridovich, 1979, Yamakura *et al.*, 1995) retain little or no enzyme activity.

Why proteins have evolved to use a particular metal cofactor appears to be governed by three criteria: 1) functional potential, 2) abundance, and 3) availability. The most important is functional potential as certain metals will only catalyze certain reactions. As catalytic redundancy among various metals does exist, speciation of metals throughout life appears to have been prejudiced by the abundance and availability of one metal over an analogous metal in the environment. Abundance and availability are not synonymous. A metal ion may be highly abundant in an environment but not available to an organism. A difference in oxidation state can render a metal ion water-soluble or insoluble and therefore available or not as a protein cofactor.

Functional potential of transition metals

Generally, metal ions that can easily change oxidation states are important in electron transfer and redox reactions (cobalt, copper, iron, manganese and nickel), while zinc is commonly found serving as a Lewis acid or structural cofactor. Zinc with a full *d* orbital does not readily accept or donate electrons and as such only one oxidation state is found in nature. Indeed according to the International Union of Pure and Applied Chemistry definition, zinc is not considered a transition metal (McNaught and Wilkinson, 1997). The uniqueness of zinc among the *d*-block elements is epitomized by the century old debate of where zinc belongs in the periodic table (Jensen, 2003).

As a catalytic cofactor, zinc is generally only found in enzymes that catalyze non redox reactions such as hydrolases. Zinc commonly activates water or substrates

leading to an increase in acidity, nucleophilicity or nucleophilicity and electrophilicity (Vallee and Auld, 1990, Andreini *et al.*, 2008b). In zinc-dependent carbonic anhydrase, zinc facilitates the formation of a zinc-bound hydroxide ion which is able to attack carbon dioxide ultimately converting it to bicarbonate (Lindskog and Coleman, 1973). In the majority of cases, the functional group is a zinc-bound hydroxide but zinc has also been found to activate thiols for nucleophilic attack (Myers *et al.*, 1993, Hightower and Fierke, 1999, Peariso *et al.*, 1998). In both cobalamin-dependent and independent methionine synthase, a methyl group from methyltetrahydrofolate is transferred to homocysteine to form methionine. The zinc in both enzymes binds to the sulfur group of homocysteine increasing the nucleophilicty of the thiolate group, which then abstracts a methyl group from the methyl donor resulting in methionine (Peariso *et al.*, 1998, Koutmos *et al.*, 2008).

Copper and nickel are comparable Lewis acids to zinc but zinc predominates as the Lewis acid of choice for most proteins (Andreini *et al.*, 2008b). The preference for zinc over copper or nickel could be due to the propensity of the later metal ions to change oxidation states and cause the generation of free radicals (Gutteridge and Wilkins, 1983, Torreilles and Guérin, 1990). Most likely due to a relative nontoxic nature, zinc was found to be one of the most abundant metals in enzymes annotated as such in public databases (second only to magnesium) (Andreini *et al.*, 2008b). The absence of redox chemistry could also explain the predominance of zinc in structural sites, which enable otherwise thermodynamically unfavorable protein folds such as the zinc-finger (Berg and Shi, 1996, Reddi and Gibney, 2007).

Cobalt, copper, iron, manganese and nickel have partially filled *d* electron orbitals. This characteristic permits energetically favorable oxidation or reduction of substrates; they are also occasionally found serving as Lewis acids. Iron and copper are commonly used in biological systems that require the transfer of electrons. A “wire” of three Fe-S clusters present in [NiFe]-hydrogenase is responsible for the transfer of electrons between an electron donor (or acceptor) and the active site of the protein (Volbeda *et al.*, 1995). The active site where oxidation of H_2 (or reduction of H^+) occurs is deeply buried within the protein (Volbeda *et al.*, 1995). Three Fe-S clusters are proposed to enable sequential transfer of electrons between the active site and the protein surface (Volbeda *et al.*, 1995). At the protein’s surface electrons are transferred to a redox partner (Yagi, 1970, Fritz *et al.*, 2001).

While cobalt, copper, iron, manganese and nickel can all participate in redox reactions, these metals do differ in redox potential and vary in the types of group transfer reactions usually catalyzed. The redox potential of the Cu(II)/ Cu(I) pairs in enzymes is between 0.25 and 0.75 V, while Fe(III)/ Fe(II) is -0.5 to 0.6 V (Crichton and Pierre, 2001). Manganese and iron are preferred in the transfer of oxygen atoms, while cobalt and nickel are more suited to transfer methyl groups (da Silva and Williams, 2001).

The choice of metal ion is also determined by other chemical characteristics such as lability. Cobalt is rarely found in biology outside of a corrin ring (referred to as cobalamin or B_{12}). Oxidation of Co(II) to Co(III) can be deleterious to protein activity; Co(III) is kinetically inert and therefore does not readily bind or release groups (e.g. substrate) (Werner, 1913). The corrin ring of cobalamin, however, is able to perturb the

chemical properties of Co(III) and ligand exchange is relatively fast (Marques and Knapton, 1997). Incorporation of cobalt (outside of a corrin ring) into an active site is rarely observed in nature, especially when other metals can catalyze the same reaction without the risk of protein inactivation. An interesting exception in the cobalt-type of nitrile hydratase that has a non-corrin Co(III) in the active site (Brennan *et al.*, 1996). As with the corrin ring, perhaps the polypeptide chain of nitrile hydratase is able to advantageously perturb the chemical properties of Co(III) (Shearer *et al.*, 2001).

Abundance and availability of transition metals

Copper is also commonly found as a cofactor in electron-transfer proteins, however, iron and not copper is the cofactor used in the characterized hydrogenases. In general, iron is one of the most prevalent metal cofactors. This observation is due to the second and third metal preference criteria: abundance and availability. The abundance and availability of metal ions is thought to have had a large impact on metal speciation throughout the domains of life (Williams, 1997).

Iron is the most common redox-active metal cofactor found in proteins (Andreini *et al.*, 2008b) because it was proposed that early life evolved under anaerobic conditions (Oparin, 1938). Water-soluble Fe(II) was plentiful but the majority of copper was found in the water-insoluble Cu(I) form (Crichton and Pierre, 2001). Therefore, life evolved with iron and not copper. As atmospheric dioxygen content increased, Fe(II) was gradually oxidized to the water-insoluble Fe(III) form and Cu(I) was oxidized to the water-soluble Cu(II) form. Therefore, copper-containing proteins are thought to have evolved only after the emergence of dioxygen (Egami, 1975).

These two metals can catalyze the same reactions but, although iron is currently less bioavailable, copper cannot simply replace iron in proteins. Due to the unfortunate

predominance and dependence on iron as a protein cofactor, life had to compensate. A popular strategy for microorganisms is the use of siderophores, which successfully compete with hydroxyl ion for Fe(III) (Neilands, 1995).

Under some circumstances, copper proteins eventually evolved and replaced iron-dependent analogs, such as found for diatoms (Peers and Price, 2006, Strzepek and Harrison, 2004). The availability for copper is higher relative to iron in open oceanic waters (Bruland, 1980). The low zinc concentration in these waters has also led to unusual metal-dependencies such as the use of cobalt or cadmium instead of zinc in carbonic anhydrase (Yee and Morel, 1996, Xu *et al.*, 2008) and calcium instead of zinc in alkaline phosphatase (Kathuria and Martiny, 2011).

Nickel and cobalt-dependent enzymes are relatively rare. The modern day existence of these enzymes is thought to be a remnant from the Archaean era, which was dominated by methanogens and preceded the catastrophic oxidation event roughly 2.4 Gyr (gigayear; 10^9 years) ago (Bekker *et al.*, 2004, da Silva and Williams, 2001). At the end of the Archaean era, nickel abundance diminished presumably due to reduced volcanic activity, and this event is proposed to have precipitated a decline of methanogens and the rise of atmospheric oxygen (Konhauser *et al.*, 2009). Cobalt is mainly found bound in the form of B_{12}, biosynthesis of which is thought to have evolved 2.7 – 3.5 Gyr ago (Scott, 1993). In several systems, dependence on these metals has become redundant by the more recent evolution of analogous enzymes with different metal cofactors. For instance, B_{12}-dependent enzymes appear to have been lost by all land plants, most unicellular eukaryotes and insects (Zhang *et al.*, 2009). For

methionine synthesis, these organisms encode a cobabalin-independent methionine synthase (Ravanel *et al.*, 1998, Matthews *et al.*, 2003, Gophna *et al.*, 2005).

In general, selection of a particular metal cofactor has been based on the principle of “economical utilization of resources” as described by da Silva and Williams (da Silva and Williams, 2001). This principle describes that a metal, which can perform the necessary function, is chosen based on cost, in terms of energy, required for uptake from the environment. The distribution of metals today is the result of billions of years of evolution under various and changing selective conditions that reflect the changing geochemistry of Earth.

Relationship between the Metal Ion and the Protein

Metalloproteins have evolved to specifically bind and harness the chemical reactivity of a particular metal ion. Several theories govern the selectivity and specificity of amino acid ligands for a metal ion. The hard and soft acids and bases principle from Parr and Pearson describes how soft Lewis bases and acids will form complexes, while hard Lewis bases will pair with hard acids (Parr and Pearson, 1983). For metal-ligand interactions, metal ions are Lewis acids with hard (Fe(III) and Co(III)), intermediate (Fe(II), Co(II), Ni(II), Cu(II) and Zn(II)) or soft (Cu(I)) characteristics. Accordingly, the side groups of several amino acids can act as Lewis bases and are commonly found bound to metal ions: the imidazole of histidine, the thiolate of cysteine, the carboxylate of aspartic acid and glutamic acid, the phenoxide of tyrosine, and the thioether of methionine. The intermediate Lewis acid imidazole of histidine is the most common metal-binding residue followed by the soft base side group of cysteine (Rulísek and Vondrásek, 1998).

Metal ions can be divided into two classes; 1) the imidazole class includes manganese, cobalt and iron, and 2) the sulfur class includes nickel, zinc and copper (Zheng *et al.*, 2008). Both groups are found to interact mainly with oxygen, nitrogen and sulfur side groups of amino acids with a preference for the imidazole group of histidine (Zheng *et al.*, 2008). The sulfur class in addition to preferring imidazole, also shows a preference for thiol side groups, and the imidazole class also shows preference for the side groups of aspartic and glutamic acid (Zheng *et al.*, 2008).

In addition to an observed preference for a type of ligand, metal ions also prefer to bind to a particular number of ligands (determined by the coordination number, CN) and for those ligands to bind to the metal with a specific orientation (referred to as coordination geometry) (Figure 1-1). Surveys of metalloprotein X-ray structures have led to insights into these partialities (Harding, 1999, Harding, 2004, Dudev *et al.*, 2006, Zheng *et al.*, 2008). Co(II) and Ni(II) are commonly found in an octahedral arrangement (CN=6); Zn(II) ions are found in tetrahedral arrangements (CN=4) while Cu(II) ions are commonly found in square planar arrangements (CN=4) (Rulísek and Vondrásek, 1998).

The protein does not simply provide a scaffold to which a metal can bind. The chemical properties of a metal ion can be influenced and fine-tuned by the bound protein. The polypeptide provides a more or less rigid structure where the metal-binding residues are fixed in space and the coordination geometry can be distorted relative to what is preferred by the metal. The classic example of distorted coordination geometry is plastocyanin from plants (Katoh and Takamiya, 1961, Katoh *et al.*, 1961). These proteins are often referred to as blue copper proteins because when the copper ion is

oxidized, it is an intense blue color (Katoh, 1960). Plastocyanin is an electron-transfer protein involved in photosynthesis; to function the copper ion switches between oxidation states (Katoh and Takamiya, 1961, Fork and Urbach, 1965).

Researchers suspected early on that the intense blue color was the result of an unusual copper site since the properties of copper observed in these proteins were not known in inorganic chemistry (Siiman *et al.*, 1974). A high resolution crystal structure of plactocyanin from poplar was solved and revealed that the copper was bound in a very distorted geometry (Guss and Freeman, 1983). The coordination geometry was neither tetrahedral (the preferred geometry of Cu(I)) nor square-planar (the preferred geometry of Cu(II)) but intermediate between the two (Guss and Freeman, 1983). This suggests that the copper geometry in plastocyanin is in an entatic state: the copper site resembles the transition state (Vallee and Williams, 1968). Instead of decreasing the activation energy, the energy of the metal ion is increased (Vallee and Williams, 1968). The overall effect is lowering of the energy barrier. Because of the intermediate geometry imposed by the protein, reorganization energy of Cu(II) to Cu(I) to Cu(II) is minimal.

Other features of the protein can affect reactivity of the metal ion. It is essential that the iron in cytochrome b transfers electrons but it is critical that iron does not transfer electrons during transport of oxygen by hemoglobin (Misra and Fridovich, 1972). Oxidation and reduction of iron in cytochrome b enables electron transfer. However, oxidation of Fe(II) to Fe(III) inactivates hemoglobin, as Fe(III) has almost no affinity for dioxygen (Conant, 1923). The iron in both proteins is bound to protoporphyrin IX so it is not the porphyrin ring that imparts this duality of function (Gribble and

Schwartz, 1965). Instead the protein is responsible for the functional constraints that cause iron to transfer electrons in one case but not in the other. The globin chains of hemoglobin are thought to stabilize the heme-oxygen complex thereby discouraging formation of Fe(III) (Percy *et al.*, 2005). Indeed, mutation of key residues in the heme pocket were found to stabilize Fe(III) (Gerald and Efron, 1961, Hayashi *et al.*, 1980, Priest *et al.*, 1989, Brennan and Matthews, 1997, Hojas-Bernal *et al.*, 1999). In the light of these examples, it is clear that these proteins have specifically evolved to optimize the functional potential of iron.

Due to the relationship that has evolved between the protein and the metal ion, the activity of metalloproteins is, consequently, strictly dependent on the presence of a metal and in most cases a specific metal. Metallation by the wrong metal can have detrimental effects such as an inactive or mis-folded protein (Coleman, 1967, Cooper *et al.*, 1997, Predki and Sarkar, 1994).

Correct Allocation of Metals within the Cell

It was initially assumed that pools of uncomplexed metal ions were available within the cell, such that a nascent polypeptide would acquire its cofactor solely through the metal-affinity of the chelating ligands. However, more recently it has become clear that this picture of metal metabolism is oversimplified. Metal-binding ligands in proteins are not sufficiently selective to ensure that the proper cofactor is loaded and “free” metals do not appear to exist within the cell but are chelated by proteins and small molecular ligands (Outten and O'Halloran, 2001, Waldron and Robinson, 2009). The protein ligands that bind the metal and the binding-pocket of a folded protein can favor the correct metal over an incorrect one. As discussed in the previous section, metals prefer to bind to certain chemical groups and prefer those groups to bind with a particular

geometry (Rulísek and Vondrásek, 1998, Harding, 1999). However, the influence of the protein in excluding incorrect metals is insufficient in a biological setting. The Irving-Williams series dictates that the affinity between a ligand and a metal follows an order of preference with Mg(II) and Ca(II) forming the weakest bonds and Cu(II) and Zn(II) forming the strongest bonds (Irving and Williams, 1948). If metal-protein speciation was dependent solely on the protein then all metal proteins regardless of the metal-ligands would contain copper (Tottey *et al.*, 2007).

However, proteins successfully acquire the correct metal from the cellular milieu. The mechanisms responsible are just now being elucidated (O'Halloran and Culotta, 2000, Finney and O'Halloran, 2003, Tottey *et al.*, 2005, Leach and Zamble, 2007, Tottey *et al.*, 2007, Zhou *et al.*, 2008, Tottey *et al.*, 2008, Waldron and Robinson, 2009, Robinson and Winge, 2010). The availability of the metal ion is an important aspect and the cell actively regulates the intracellular concentration of metals through the use of multiple metal-responsive transcription factors, transporters, and storage proteins (Brocklehurst *et al.*, 1999, Outten *et al.*, 2000, Robinson *et al.*, 2001, Liu *et al.*, 2004, Iwig *et al.*, 2008, Leitch *et al.*, 2007, Lee and Helmann, 2007). The folding location of a protein, cytoplasm versus periplasm, can also help ensure that a protein binds the correct metal. Two proteins from *Synechocystis* PCC 6803 with identical metal-binding residues were found to contain manganese or copper depending on the fact that one protein was allowed to fold in the cytoplasm (where manganese is available) and the other was not allowed to fold until in the periplasm (where copper is available) (Tottey *et al.*, 2008). As expected from the Irving-Williams series, when the Mn-protein folds *in vitro* in the presence of equimolar copper and manganese, it specifically acquires

copper (Tottey *et al.*, 2008). Therefore, availability of metal ions in the cell is an important aspect in regards to metalloprotein maturation.

As copper is the preferred partner based on affinity for protein ligands, several organisms encode special trafficking proteins that essentially make copper unavailable in the cell. Copper is also delivered to specifically to copper-dependent enzymes by these specialized proteins, which are referred to as metallochaperones. Since the discovery of the first copper metallochaperone, Atx1 (Pufahl *et al.*, 1997), numerous accessory factors involved in ensuring target proteins are loaded with the correct metal have also been characterized. These accessory factors can be divided into two main types: metallochaperones and insertases.

In general, metallochaperones are responsible for storage and delivery of a metal cofactor to a target metalloprotein (Rae *et al.*, 1999, Shi *et al.*, 2008). These proteins are particularly important in the trafficking of toxic metals and as such are frequently found trafficking copper (Glerum *et al.*, 1996, Pufahl *et al.*, 1997, Culotta *et al.*, 1997) but different metallochaperones have been discovered that specifically traffic and deliver iron (Bou-Abdallah *et al.*, 2004, Shi *et al.*, 2008), cobalt (in the form of B_{12}) (Padovani *et al.*, 2008), manganese (Luk *et al.*, 2003) or nickel (Musiani *et al.*, 2004, Zhang *et al.*, 2005) to target apoenzymes. In contrast to these other metals, a cytoplasmic zinc chaperone that targets a single zinc-dependent protein has not been identified.

While metallochaperones are responsible for storage and delivery, insertases facilitate incorporation of the metal ion in an energy-dependent manner into the target protein's catalytic site (Jeon *et al.*, 2001). In studied systems, the metal is specifically

inserted into the folded protein. The copper chaperone CCS (copper chaperone for SOD1) appears to insert copper into the fully folded superoxide dismutase homodimer (Schmidt *et al.*, 2000). A key characteristic for some of these insertases is the coupling of NTP hydrolysis with metal insertion (Loke and Lindahl, 2003, Ba *et al.*, 2009). For several metalloproteins, such as [NiFe]-hydrogenase (Volbeda *et al.*, 1995), the metal site is buried within the protein. This organization creates a kinetic trap and the metal ion is not replaced by more competitive metal ions (Dudev and Lim, 2008). Commonly, maturation of these proteins requires the coupling of NTP hydrolysis with metal insertion, as will be discussed in the next section. This observation could be evidence of the structural rearrangements that must occur for metal insertion.

The G3E Family

Studies involving the maturation of Ni-urease and [NiFe]-hydrogenase have provided an extensive picture of how metal-dependent catalytic centers are assembled. These two nickel-containing proteins require a suite of accessory proteins to properly insert nickel into the catalytic site (only one exception has been found to date; *Bacillus subtilis* encodes a functional urease in the absence of the canonical accessory proteins (Kim *et al.*, 2005)). In both cases, a GTPase (UreG for urease or HypB for hydrogenase) is proposed to be involved in the incorporation of the nickel ion. Deletion of *hypB* abolishes hydrogenase activity, which can be partially restored with the addition of exogenous nickel (Waugh and Boxer, 1986, Jacobi *et al.*, 1992). Deletion of *ureG* abolishes urease activity and purified urease from cells lacking UreG is nearly devoid of nickel (Lee *et al.*, 1992). The exact mechanism of how these proteins assist in maturation is still unclear. Perhaps they enable structural changes that allow cofactor insertion (acting as insertases) or regulate association between the other accessory

factors also required for maturation (Soriano and Hausinger, 1999, Leach and Zamble, 2007).

The identification of these GTPases and the proposal for their involvement as maturation factors were aided by gene clustering. The genes encoding HypB and UreG are commonly found in gene clusters that contain genes encoding the target metalloenzyme and/or accessory factors also involved in maturation. Synthesis of the [NiFe]-hydrogenase active site requires the involvement of at least six proteins (Jacobi *et al.*, 1992). In characterized organisms, these genes form gene clusters with each other and the structural genes of hydrogenase (Vignais *et al.*, 2001). The *E. coli* genome encodes four [NiFe]-hydrogenases (Andrews *et al.*, 1997). HypB is required for the activity of each (Lutz *et al.*, 1991), but the corresponding gene is found in an operon divergently transcribed from an operon containing most of the structural genes for the formate hydrogenlyase complex (hydrogenase-3) (Figure 1-2A). Maturation of the urease active site requires at least four accessory proteins (Lee *et al.*, 1992). The corresponding genes commonly form gene clusters with one another and the three structural genes for urease (Carter *et al.*, 2009). In *Klebsiella aerogenes*, *ureG* is found at the end of a seven gene urease cluster (Figure 1-2B).

Clustering on the chromosome between accessory factors and the target metalloenzyme is a common occurrence for systems in addition to [NiFe]-hydrogenase and urease. Gene clusters including chaperones and target metalloezymes are found for nitrogenase (Jacobson *et al.*, 1989, Hong *et al.*, 1996), nitrile hydratase (Hashimoto *et al.*, 1994, Wu *et al.*, 1997), trimethylamine oxide reductase (Ilbert *et al.*, 2003), tyrosinase (Lopez-Serrano *et al.*, 2004), nitrate reductase (Blasco *et al.*, 1998), acetyl-

CoA synthase (Loke and Lindahl, 2003) and carbon monoxide dehydrogenase (Kerby *et al.*, 1992). As novel metalloenzymes are discovered and metallocenter biosynthesis is further characterized, this list will most certainly increase.

HypB and UreG are closely related and form two subfamilies of the G3E family of phosphate-binding loop (P-loop) GTPases (G3E family) as defined by Leipe and colleagues (Figure 1-3) (Leipe *et al.*, 2002). The G3E family is separated from the rest of the SIMIBI class of GTPases (for signal recognition GTPases, MinD superfamily, and BioD superfamily) by a glutamate residue in the Walker B motif and an intact G4 motif (Leipe *et al.*, 2002). The P-loop NTPase fold (also referred to as the mononucleotide-binding fold) is one of the most common folds and comprises roughly 15% of all gene products in some organisms (Koonin *et al.*, 2000). This fold is characterized by an N-terminal Walker A motif with the consensus sequence GXXXXGK[TS], which functions in the positioning of the triphosphate group of the nucleotide (la Cour *et al.*, 1985, Saraste *et al.*, 1990). The Walker B motif, typically represented by DXXG, coordinates a water-bridge magnesium ion, which binds to the nucleotide (Pai *et al.*, 1990). In HypB, the glutamine (present instead of aspartic acid) in the Walker B motif is, in contrast, a direct ligand to magnesium (Gasper *et al.*, 2006). The G4 motif (NKXD) provides specificity to GTP (Hwang and Miller, 1987).

GTPases are often referred to as molecular switches since binding and hydrolysis of GTP cause conformational changes effectively turning on or off interactions with other macromolecules (Vetter and Wittinghofer, 2001). As can be predicted by the commonality of the P-loop fold, a myriad of functions are carried out by GTPases. GTPases have been found to be involved in translation (Hauryliuk, 2006, Rodnina *et al.*,

2000), signal transduction (Narumiya, 1996), cell motility (Wittmann and Waterman-Storer, 2001), intracellular trafficking (Segev, 2010), cell cycle regulation (Meier *et al.*, 2000), and membrane transport (Molendijk *et al.*, 2004).

Like other characterized GTPases, HypB and UreG have low intrinsic GTPase activity (Fu *et al.*, 1995, Zambelli *et al.*, 2005). Hydrolysis of GTP by these proteins was found to be essential for nickel-insertion into their target (Maier *et al.*, 1995, Soriano and Hausinger, 1999, Moncrief and Hausinger, 1997). Mutation of the P-loop domain of UreG abolishes formation of the urease maturation complex and subsequently urease activity (Moncrief and Hausinger, 1997). Upon interaction with a GTPase-activating protein (GAP) or molecular target such as RNA, GTPase activity is stimulated (Meier *et al.*, 2000, Salomone-Stagni *et al.*, 2007). Based on structure modeling and predicted similarity to other GAPs, the urease maturation factor UreF was proposed to perform the role of a GAP for UreG (Soriano *et al.*, 2000, Salomone-Stagni *et al.*, 2007). The recent X-ray structure, however, displayed little resemblance to the predicted structure model and as a consequence little if any similarity to characterized GAPs is actually observed (Lam *et al.*, 2010). Therefore UreF may not function as a GAP (Lam *et al.*, 2010). A GAP for HypB has not been predicted.

Functionally, HypB and UreG belong to an expanding group of bacterial P-loop NTPases proposed to be involved in synthesis of enzyme metal-sites. The GTPases CooC and AcsF are proposed to insert nickel into carbon monoxide dehydrogenase (Jeon *et al.*, 2001) and acetyl-CoA synthase (Loke and Lindahl, 2003), respectively. Several proteins involved in Fe-S cluster assembly are NTPases. HydF has GTPase activity and is proposed to be a scaffold protein involved in synthesis and transfer of the

[2Fe-2S] center into [FeFe]-hydrogenase (Shepard *et al.*, 2010). ApbC was found to hydrolyze ATP and is responsible for insertion of a [4Fe-4S] cluster into isopropylmalate isomerase (Boyd *et al.*, 2008). The P-loop GTPase MobB is proposed to be involved in Mo-cofactor synthesis or transfer (Eaves *et al.*, 1997). Having a molecular switch participate in maturation of metalloenzymes could serve to ensure that the correct metal ion or metal cofactor is being inserted.

In addition to HypB and UreG, the G3E family is subdivided into two other subfamilies: MeaB, which is required for the activation of methylmalonyl-CoA mutase (an adenosylcobalamin-dependent enzyme) (Padovani and Banerjee, 2006), and COG0523, a large and diverse subfamily of proteins with poorly defined function (Figure 1-3). In general, characterized G3E family proteins have been found to function as either metal-insertases or as a dual function metallochaperone/insertase.

MeaB appears to fulfill the role of an adenosylcobalamin insertase, facilitating the insertion of B_{12} into methylmalonyl-CoA mutase (MCM) (Figure 1-3B) (Padovani *et al.*, 2006). Adenosyltransferase serves as the chaperone that delivers B_{12} to the MeaB-MCM complex (Padovani *et al.*, 2008). MeaB has been described as a gatekeeper and is able to distinguish active from inactive cofactor and elicit insertion or release of the cofactor as necessary (Korotkova and Lidstrom, 2004, Padovani and Banerjee, 2006, Padovani and Banerjee, 2009, Takahashi-Íñiguez *et al.*, 2010). MeaB has low GTPase activity, which is stimulated roughly 100-fold by interaction with its target (Padovani *et al.*, 2006). A large structural rearrangement occurs upon interaction between MeaB and its target (Padovani *et al.*, 2006), suggesting that MeaB may be responsible for the structural changes that must occur for cofactor incorporation.

Unlike MeaB, HypB is thought to carry out both the insertase and metallochaperone roles in a subset of organisms. The metallochaperone role is proposed based on a histidine stretch located at the N-terminus of some HypB proteins which presumably serves to store nickel ions for subsequent transfer to hydrogenase (Olson *et al.*, 1997). In *Bradyrhizobium japonicum*, this histidine stretch was found to bind 18-nickel ions per dimer (Fu *et al.*, 1995). In *E. coli*, this His-stretch is missing from HypB, and SlyD is presumed to be the metallochaperone component that delivers nickel to the assembly complex (Zhang *et al.*, 2005, Leach *et al.*, 2007). In contrast to the HypB proteins, all characterized UreG proteins lack His-stretches (except some plant orthhologs (Witte *et al.*, 2001)). Accordingly, UreG is proposed to serve solely as an insertase providing the GTPase activity required for urease maturation and a separate protein, UreE, is proposed to serve as the metallochaperone component delivering nickel to the maturation complex (Remaut *et al.*, 2001, Benoit and Maier, 2003, Musiani *et al.*, 2004).

Little is known regarding the fourth G3E subfamily, COG0523. Currently, spurious and inconsistent reports have linked members of this subfamily to cobalamin biosynthesis (Crouzet *et al.*, 1991), nitrile hydratase activation (Hashimoto *et al.*, 1994) and zinc homeostasis (Gaballa and Helmann, 1998).

Zinc Homeostasis

Like all essential transition metals, a dichotomy exists that is defined by the essential but potentially toxic nature of zinc. As multiple essential proteins are reliant on zinc for activity, such as RNAP (Scrutton *et al.*, 1971) and tRNA synthetases (Mayaux *et al.*, 1982, Xu *et al.*, 1994, Sood *et al.*, 1999, Sankaranarayanan *et al.*, 2000, Zhang *et al.*, 2003b, Bilokapic *et al.*, 2006, Sokabe *et al.*, 2009), the presence of zinc to the cell is

essential. Because of implications in human health, an environment of particular interest is the vertebrate host, where access to zinc by a pathogen is thought to be limited. In 1973, Kochan introduced the concept of nutritional immunity as a defense strategy against invading pathogens (Kochan, 1973). The host organism actively deprives metals from the invader inducing both hypoferremia and hypozincemia (deficiency of iron and zinc, respectively, in the blood) as part of the acute inflammatory response (Weinberg, 1975, Liuzzi *et al.*, 2005, Motley *et al.*, 2004, Kehl-Fie and Skaar, 2010). Therefore, the mechanisms that enable a pathogen to overcome this host-induced Zn-starvation are considered essential to a pathogen's ability to cause infection (Panina *et al.*, 2003, Kim *et al.*, 2004, Pasquali *et al.*, 2008).

Excess zinc, however, can be toxic. In contrast to vertebrates, plants may utilize this aspect of zinc as a defense strategy (Poschenrieder *et al.*, 2006). Whereas for some pathogens high-affinity zinc uptake is essential for virulence, in *Agrobacterium tumefaciens*, that machinery was down-regulated in response to plant signals (Yuan *et al.*, 2008) and repression of that machinery was required for full virulence of the plant pathogens, *Xanthomonas campestris* and *Xanthomonas oryzae* (Tang *et al.*, 2005, Huang *et al.*, 2008, Yang *et al.*, 2007). The exact mechanism by which zinc becomes cytotoxic is not known, but most likely toxicity arises from the misincorporation of zinc into proteins, thereby displacing the native cofactor leading to reduced activity or inactivation. This mechanism could explain inhibition of the respiratory electron transport chain in bacteria by zinc (Kasahara and Anraku, 1972, Kasahara and Anraku, 1974, Kleiner, 1978, Singh and Bragg, 1974, Beard *et al.*, 1995). Zinc toxicity may also be mediated through binding to nascent polypeptides disrupting proper folding or

binding to folded proteins generating protein aggregates through zinc-mediated protein-protein interactions (Maret and Li, 2009).

Because zinc-depletion and zinc excess can be detrimental to growth and survival, the intracellular concentration of zinc is tightly controlled through the use of zinc-sensing transcriptional regulators (Westin and Schaffner, 1988, Patzer and Hantke, 1998, Brocklehurst *et al.*, 1999, Zhao and Eide, 1997), import and export machineries (Palmiter and Findley, 1995, Zhao and Eide, 1996, Rensing *et al.*, 1997, Grotz *et al.*, 1998, Patzer and Hantke, 1998, McMahon and Cousins, 1998), compartmentalization (MacDiarmid *et al.*, 2000, Devirgiliis *et al.*, 2004) and storage proteins (Richards and Cousins, 1975, Blindauer *et al.*, 2002). Using both narrow and broad-specificity import systems, *E. coli* can actively accumulate zinc to a level of 200,000 atoms/cell (Outten and O'Halloran, 2001), a 1,000-fold excess over the typical zinc concentration in the medium. However, biochemical measurements indicate that there is essentially no free zinc in an *E. coli* cell (Outten and O'Halloran, 2001), suggesting that, once imported, zinc becomes sequestered by proteins and small molecules. It has been proposed therefore that the only way for a newly synthesized protein to obtain zinc is via ligand-exchange with zinc-loaded macromolecules (Heinz *et al.*, 2005). Whether specific proteins, such as a zinc-specific chaperone, are involved in mediating this process is currently not known.

Zinc-binding proteins account for 5% of the average proteome (Andreini *et al.*, 2006), and ribosomes most likely constitute the largest zinc reservoir. A rapidly growing *E. coli* cell contains as many as 50,000 ribosomes (Bremer and Dennis, 2008), each with possibly 3-4 bound zinc ions (Katayama *et al.*, 2002, Gabriel and Helmann, 2009),

thereby tying up most of the zinc present. Other abundant proteins must sequester the remaining zinc pool; RNAP contains two bound zinc ions (Giedroc and Coleman, 1986) and is present at ~2000 copies/cell (Lewis *et al.*, 2000).

Zinc Transport and Regulators

With the exception of a few non-transport mechanisms that will be discussed, bacteria appear to rely heavily on influx and efflux to maintain zinc homeostasis. Multiple families of transporters are currently known to specifically transport zinc in prokaryotes. These transporters can be divided into two main groups: low-affinity and high-affinity. Low-affinity transporters tend to have a broader spectrum as to which divalent metal cations they will transport, while the high-affinity transporters are specifically regulated in response to the intracellular concentration of zinc and are highly specific in the transport of zinc. For high-affinity transport, specificity is critical to avoid poisoning the cell with a metal previously in sufficient supply. How zinc is transported across the inner membrane of gram-negative bacteria is fairly well characterized, but how zinc is transported across the outer membrane is only known in a couple of organisms.

Zinc efflux

When the intracellular zinc level becomes high, zinc-specific efflux pumps are employed to reestablishment of the intracellular zinc concentration. These transporters include low-affinity CDF (cation diffusion facilitator) proteins, RND (resistance, nodulation, division) efflux transporters, CorA family transporters and high-affinity P-type ATPases/CPx-ATPases.

In *E.* coli, a high concentration of zinc activates expression of *zntA* and energy-dependent efflux of zinc is induced (Rensing *et al.*, 1997). ZntA is specifically a

Zn(II)/Pb(II)/Cd(II)-translocating P-type ATPase (Liu *et al.*, 2006) and expression of *zntA* is specifically induced by Zn(II), Pb(II) or Cd(II) (Binet and Poole, 2000). The transcriptional regulator responsible for expression is ZntR, a member of the MerR family of regulators (Brocklehurst *et al.*, 1999). Like other MerR regulators, apo-ZntR (metal-free) binds to specific promoter DNA repressing expression (Outten *et al.*, 1999). When bound to zinc, ZntR distorts the promoter, repositioning the -10 and -35 sequences, leading to a more favorable interaction with RNAP and expression of *zntA* in induced (Ansari *et al.*, 1995, Outten *et al.*, 1999). ZntR homologs have also been characterized in *Bordetella pertussis* (Kidd and Brown, 2003).

In the gram-positive bacterium *Staphylococcus aureus* and several cyanobacteria, zinc efflux is regulated by a member of the ArsR family of regulators (Endo and Silver, 1995, Thelwell *et al.*, 1998, Liu *et al.*, 2004). ArsR family proteins bind to target DNA in the absence of metal; in the presence of metal, a conformational change causes the protein to dissociate and transcription can occur (Eicken *et al.*, 2003, Busenlehner *et al.*, 2003).

In *Salmonella enterica,* the efflux transporter ZntB is related to the CorA family of magnesium transporters (Worlock and Smith, 2002). CorA family proteins were originally characterized in the transport of magnesium catalyzing both import and export. ZntB has been shown to specifically export zinc (Worlock and Smith, 2002).

In *Cupriavidus metallidurans* and *P. aeruginosa* CMG103, a member of the RND multi-drug efflux transport family, CzcABC, has been shown to be involved in zinc efflux as well as cobalt and cadmium (Mergeay *et al.*, 1985, Hassan *et al.*, 1999). This

transport system is a cation/proton antiporter and is the only machinery known to span both membranes (Nies, 1995).

CDF proteins have been shown to be involved in resistance to zinc in *C. metallidurans* (Nies, 1992), *B. subtilis* (Guffanti *et al.*, 2002), *S. aureus* (Xiong and Jayaswal, 1998) and *E. coli* (Grass *et al.*, 2001, Grass *et al.*, 2005b). These proteins also transport zinc with an antiporter mechanism; efflux of zinc is catalyzed in exchange for potassium or hydrogen cations (Guffanti *et al.*, 2002). Structural studies of the CDF protein YiiP from *E. coli* suggest that these efflux proteins are autoregulated by cytoplasmic zinc and thereby serve as first-responders to higher than normal zinc concentrations (Lu and Fu, 2007, Lu *et al.*, 2009). Efflux by ZntA (discussed above), in contrast, is reliant on zinc-sensing by ZntR and subsequent gene expression, translation and insertion into the membrane.

Zinc uptake

Several transport families have been identified and characterized among bacterial organisms that are responsible for uptake of zinc. Low-affinity ZIP (ZRT/IRT-like proteins) family transporters and high-affinity ABC (ATP-binding cassette) transporters appear mainly responsible for zinc uptake. However, zinc can enter the cell through the use of multiple broad spectrum metal ion importers and/or porins (Blencowe and Morby, 2003). As in eukaryotes, ZIP family zinc-transporters have been implicated in bacterial zinc uptake. The first characterized bacterial ZIP transporter was ZupT in *E. coli* (Grass *et al.*, 2002). ZupT appears to be expressed constitutively at a low level and is also capable of transporting other divalent cations (Grass *et al.*, 2005a). Other characterized secondary zinc transporters include the PIT (inorganic phosphate transport) system in

E. coli (Beard *et al.*, 2000) and the citrate-metal cotransporter CitM in *B. subtilis* (Krom *et al.*, 2000).

When transport by these systems is inadequate, a high-affinity zinc transporter ZnuABC is specifically induced (Patzer and Hantke, 1998, Gaballa and Helmann, 1998). These systems are generally composed of a membrane-spanning permease (ZnuB), an ATP-hydrolyzing component (ZnuC) that energizes transport and a periplasmic chaperone (ZnuA) that delivers zinc to the permease (Claverys, 2001). In addition to these principal components, some organisms express an auxiliary component ZinT that was recently shown to facilitate recruitment of zinc in the periplasm (Graham *et al.*, 2009, Petrarca *et al.*, 2010). Zinc-specificity is proposed to be due to ZnuA conformational changes specifically induced by zinc (Yatsunyk *et al.*, 2008). Other metal ions do not stabilize ZnuA to the extent zinc binding does (Yatsunyk *et al.*, 2008).

Zur

Expression of the genes encoding the components of ZnuABC is dependent on the zinc-sensing regulator Zur, which was concurrently discovered in *E. coli* and *B. subtilis* (Patzer and Hantke, 1998, Gaballa and Helmann, 1998). Zur was shortly thereafter described for *Listeria monocytogenes* (Dalet *et al.*, 1999) and *S. aureus* (Lindsay and Foster, 2001). Zur is a member of the Fur family of metal-sensing regulators. The founding member Fur regulates gene expression in response to iron (Hantke, 1981). Other members of this family include Mur (manganese-sensing) and Nur (nickel-sensing) (Lee and Helmann, 2007). Zur orthologs are present throughout Bacteria (Lee and Helmann, 2007) with the exception of *Lactococcus lactis* and *Streptococcus* spp. which appear to use a MarR family regulator AdcR to repress genes

involved in the response to zinc deficiency (Llull and Poquet, 2004, Reyes-Caballero *et al.*, 2010).

Through a metal-induced repression mechanism, zinc occupancy in a dose-dependent manner directly controls affinity between Zur and DNA (Outten *et al.*, 2001a). In the presence of zinc, Zur binds to operator sequences upstream of target genes, preventing binding of RNAP and thus transcription initiation (Outten *et al.*, 2001a). Conversely, upon zinc-depletion, repression by Zur is lifted and expression of target genes is increased (Gaballa and Helmann, 1998). Zur proteins are found as dimers in solution (Gaballa and Helmann, 1998, Patzer and Hantke, 2000, Lucarelli *et al.*, 2007, Shin *et al.*, 2007, Feng *et al.*, 2008) and are therefore proposed to bind as homodimers to palindromic DNA sequences (Patzer and Hantke, 2000, Maciag *et al.*, 2007). Zinc chelators abolish Zur-DNA interactions and zinc enhances those interactions (Gaballa and Helmann, 1998, Maciag *et al.*, 2007, Huang *et al.*, 2008, Li *et al.*, 2009). For the *E. coli* and *Mycobacterium tuberculosis* Zur proteins, two distinct metal-binding sites are present per monomer (Outten *et al.*, 2001a, Lucarelli *et al.*, 2007). With only one zinc ion per monomer the protein does not bind DNA. This zinc site is proposed to be a structural site (Outten *et al.*, 2001a) and is conserved in both Zur and Fur homologs (Patzer and Hantke, 2000). Compared to the structural site, where the zinc is not readily exchangeable, the second zinc site is comparatively readily exchangeable and proposed to be the regulatory site (Outten *et al.*, 2001a, Lucarelli *et al.*, 2007).

In the crystal structure of the *M. tuberculosis* Zur, the structural zinc site is composed of two CXXC motifs. One motif is located in the N-terminus DNA binding domain and the other in the C-terminus dimerization domain (Lucarelli *et al.*, 2007). The

proposed regulatory site is located at the hinge region between the two domains (Lucarelli *et al.*, 2007). Additionally, a third zinc ion was found in the crystal structure, however, the biological role of this zinc or whether it is an artifact of crystallization is not known (Lucarelli *et al.*, 2007).

Bioinformatics and Gene Function Discovery

Over 1,000 genomes have now been published and another 6,000 are currently underway (Liolios *et al.*, 2010). The current release of the Reference Sequence (RefSeq) collection (as of November 10, 2010) hosted by the National Center for Biotechnology Information (NCBI) includes a set of 11,652,892 non-redundant protein sequences from 11,354 organisms (Pruitt *et al.*, 2007). From these sequences, a large volume of information about metabolic potential, differential responses to stimuli, niche adaptation and evolutionary history is available. However, in order to access that wealth of information, sense must be made of the countless strings of A's, T's, G's and C's. It is the field of bioinformatics and offshoots thereof that gives meaning and discerns information from the seemingly random order of an organism's chromosome. As a most basic definition, bioinformatics is the field of interpreting sequence information but also involves interpretation of any large data set of biological information such as gene expression, 3D protein structures, or literature (Luscombe *et al.*, 2001).

When the first genome of a free-living organism was published in 1995 (Fleischmann *et al.*, 1995), the field of bioinformatics had already existed in one form or the other for almost thirty years. One could imagine that the field had its foundation in the first use of molecular sequences for evolutionary studies (Zuckerkandl and Pauling, 1965), the first collection of protein sequences (Dayhoff, 1968) or the first use of dynamic programming for sequence comparison (Needleman and Wunsch, 1970). The

central tenet to bioinformatics is that a greater understanding of biological systems can be ascertained by comparison. Since those pioneering works, the field of bioinformatics has expanded to include larger and larger datasets including computational studies of data from genome-wide experiments on transcript or protein abundance (Luscombe *et al.*, 2001). Advancement in the bioinformatics field is best exemplified by current progress with genome-scale metabolic reconstructions (Reed *et al.*, 2006b, Reed *et al.*, 2006a, Feist and Palsson, 2008, Suthers *et al.*, 2009).

Metabolic network reconstruction is the *in silico* modeling of an organism's metabolism that is converted to a mathematical format. Instead of studying the individual components of the cell, which has been the main focus of biology during the last half of the 20th century, metabolic reconstructions focus on integration and viewing the cell as a system (Palsson, 2006). A genome-level model begins with a parts list which is provided by the genome sequence (Edwards and Palsson, 2000). The increasing wealth of genome-wide high-throughput data sets plus published reports provide information about functional, physical and temporal interactions between those components (Covert *et al.*, 2004). Central to these models is the genotype-phenotype relationship (Edwards *et al.*, 2001). The model can be used to predict experimental outcomes such as the deletion of gene A will lead to phenotype1 and the outcomes of these experiments are used to improve the model in an iterative manner (Palsson, 2006). Without the tools, datasets and concepts provided by the field of bioinformatics, systems biology would not be conceivable.

The Functional Annotation Dilemma

For the majority of genomes, annotations continue to be determined by sequence similarity to characterized or partially characterized genes in other organisms, mainly

model organisms such as *E. coli* and *Saccharomyces cerevisiae* (Salzberg, 2007). The term annotation as used here is defined as the functional description of a gene or gene product. With the avalanche of genome sequence data and automated transfer of annotations between those genomes, the definition of function has become increasingly vague.

Traditionally, gene function discovery and/or verification have largely been achieved one gene at a time by bench scientists. As of 2008, 59.3% of the genes found in the *E. coli* genome are affiliated with some type of experimental data (Keseler *et al.*, 2009). *E. coli*, however, is the exception rather than the rule, because experimentally characterizing the tens of millions of genes currently sequenced is an unrealistic goal. As a consequence, the vast majority of annotations are sequence-derived predictions. A few of these annotations are based on a combination of bioinformatic evidence, as will be discussed in the next section, but, in most cases, annotations are based solely on sequence similarity to a gene, most likely also annotated in the same way. It is often difficult to find the experimentally validated progenitor gene and how the annotation was originally acquired (Poptsova and Gogarten, 2010).

Unfortunately, current sequence-based approaches cannot predict a function for one-third of sequenced genes; moreover, for some gene families at least 60% of the gene predictions are wrong (Schnoes *et al.*, 2009). The unreliability of gene annotations will inevitably become more apparent as newly sequenced genomes emerge containing genes of an ever-increasing phylogenetic distance from those experimentally characterized. Attempts are being made to address the need for reliable and accurate gene annotations, such as developing Gold Standard datasets of experimentally verified

annotations by the COMputational Bridge to Experiments (Galperin and Koonin, 2010); nevertheless, this functional annotation dilemma is one of the largest challenges faced in the post-genomic era and threatens to undermine efforts to extract knowledge from genome sequencing efforts.

Comparative Genomics and the Gene Family Approach

The unreliability of assigning gene function based on sequence similarity soon became apparent after the first wave of genome sequences (Eisen *et al.*, 1997, Galperin and Koonin, 1998, Brenner, 1999, Bork, 2000, Attwood, 2000, Devos and Valencia, 2001). As a solution, non-homology based methods have been developed and employed to link genes with function. In combination with homology based methods, novel discoveries in all fields of biology have been made.

Increasingly sophisticated bioinformatics techniques are available to aid in the discovery of gene function, which is no longer limited to studying one gene in isolation but as a member of an ever evolving genomic community. Entire gene families from diverse organisms can be studied simultaneously. The assumption is that a gene family is an evolutionary unit that represents maintenance and divergence of function. The function of uncharacterized members can be inferred by comparison to characterized members through the use of phylogenomics and by cross-species comparison through the use of comparative genomics (Eisen, 1998b, Galperin and Koonin, 2000, Overbeek *et al.*, 2005), as described in the following sections.

Phylogenomics

Phylogenomics is the approach put forth by Eisen in the late 1990's as a strategy to improve functional predictions (Figure 1-4) (Eisen *et al.*, 1997, Eisen, 1998a). Instead of predicting the function of a gene based only on sequence similarity to a characterized

gene, the evolutionary history of those genes is also taken into account (Eisen, 1998b). To infer function from this type of study, the assumption must be made that gene functions change as a consequence of evolution and, by reconstructing the evolutionary history of a gene family, function for an uncharacterized gene can be inferred based on its relatedness to characterized family members (Eisen, 1998b). The first step is to build a phylogenetic tree of the gene family of interest; members with any functional information are specifically included. Next, that functional information is overlaid on the tree. The last step is interpretation of the functionally annotated tree. The history of functional changes is used to predict functions for uncharacterized members.

This approach was first applied to the MutS family of mismatch repair proteins (Eisen *et al.*, 1997, Eisen, 1998a). MutS is the protein responsible for recognizing and binding to the site of a mismatch or a loop in double-stranded DNA (Su and Modrich, 1986, Levinson and Gutman, 1987). Some MutS proteins do not function in DNA repair but are involved in meiotic crossover and chromosome segregation (Ross-Macdonald and Roeder, 1994, Hollingsworth *et al.*, 1995). It became apparent that sequence similarity was not able to properly distinguish between these functions; as a consequence, *mutS* family genes were being erroneously annotated in published genomes (Eisen *et al.*, 1997). By building a phylogenetic tree with available MutS protein sequences and overlaying that tree with functional information from previously published studies, the function of uncharacterized MutS homologs was predicted with more confidence than by sequence similarity alone. If the uncharacterized protein belonged to a subfamily involved in mismatch repair, then that protein was more likely to be involved in mismatch repair than in chromosome segregation (Eisen, 1998a). As

more genes sequences have become available, this analysis has been refined, and new MutS subfamilies with novel functions have been proposed (Lin *et al.*, 2007, Sachadyn, 2010).

In addition to phylogenomic analysis, which is concerned principally with the evolutionary history of gene families, other types of comparative genomic information can be used to infer gene function. The function of an uncharacterized gene can be deduced from the function of genes surrounding it on the chromosome (Overbeek *et al.*, 1999a), co-occurrence patterns with genes of known function (Pellegrini *et al.*, 1999) and shared regulatory sites with genes of known function (Figure 1-5) (Gelfand *et al.*, 2000).

Gene neighborhoods

Thirty-six years before the first bacterial genome was sequenced, Demerec and Hartman, in 1959, noted that *loci* controlling related functions tend to cluster physically on the chromosome and that only the force of selective adaptation could account for this phenomena (Demerec and Hartman, 1959). The advantage of gene neighborhoods is presumably co-transcription of genes that interact either physically and/or are in the same metabolic or response pathway (Dandekar *et al.*, 1998, Price *et al.*, 2005). Since transcription and translation occur simultaneously in *E. coli,* having two genes in the same transcript reduces the delay in complex formation (Shapiro and Losick, 1997, Pal and Hurst, 2004). Other theories for the co-localization of functionally-related genes in genomes have emerged such as the selfish operon paradigm. In this theory, gene clusters predominate because maintenance of horizontally transferred genes is more likely to occur if all the genes necessary for a particular function are co-transferred (Lawrence and Roth, 1996). Regardless of the reason for functionally-related genes to

be clustered in genomes, this observation has provided a useful tool in the prediction of gene function (Overbeek *et al.*, 1999a, Overbeek *et al.*, 1999b).

Overbeek *et al.* were the first to develop an algorithm to find and score gene neighbors (Overbeek *et al.*, 1999a). They first detected “pairs of close bidirectional best hits” in sequenced genomes. A PCBBH is basically two genes that are neighbors in one genome with the homologs of those two genes that are also neighbors in another genome. The presence of gene neighbors in closely related genomes is not necessarily significant. Genes could be neighbors due to chance and not enough evolutionary time may have passed in closely related genomes to rearrange those genes. A score can be given to gene neighbors, which takes into account the number of genomes and the phylogenetic distance between those genomes where the gene pairs occur.

Gene clustering is also found in eukaryotes (Blumenthal, 1998, Lee and Sonnhammer, 2003) but the use of this method has had much more success in prokaryotes. In prokaryotes, functionally coupled genes often share the same operon or are transcribed divergently from the same promoter (Beck and Warren, 1988). However, functionally coupled genes can be neighbors without being co-transcribed (Yanai *et al.*, 2002).

Co-occurrence

Phylogenetic co-occurence profiles infer functional coupling by assuming that during evolution functionally related proteins are maintained or eliminated in a correlated manner (Pellegrini *et al.*, 1999, Osterman and Overbeek, 2003). Also, analysis of co-occurrence can identify non-orthologous gene displacement; distribution of non-orthologous genes that functionally replace one another will be negatively correlated (Morett *et al.*, 2003, Makarova and Koonin, 2003). Co-occurrence profiles are

useful in identifying missing genes in pathways whether a gene is locally missing (the gene for a pathway step is found in some genomes but missing from others) or globally missing (the gene encoding a particular pathway step has not been identified in any genome) (Osterman and Overbeek, 2003).

Shared regulatory sites

The genes encoding proteins that belong to the same pathway or proteins that respond to the same environmental stimulus are usually part of the same regulon. In most cases, membership in a regulon is dictated by the binding of a specific regulatory protein upstream of a gene (Epstein and Beckwith, 1968). More recently, mechanisms of gene regulation has expanded to include mRNA-binding proteins (Babitzke and Yanofsky, 1993, Liu and Romeo, 1997), regulatory RNA (Liu *et al.*, 1997, Nahvi *et al.*, 2002, Wagner *et al.*, 2002) and RNAP-binding proteins (Stepanova *et al.*, 2009). Co-regulation between a gene of known function and a gene of unknown function can by inference lead to the prediction of a general function for the uncharacterized gene.

Multiple comparative genomic tools are now available for the robust identification of co-regulated genes based on genome sequence. These tools rely on the computational identification of transcription factor binding sites or riboswitches upstream of genes (Rodionov, 2007). Since these sites are rarely identical, two main methods are available for the representation of binding sites or riboswtiches.

A consensus sequence is a representative sequence that matches all known sites but through the use of an extended language that properly describes degeneracy (Stormo, 2000). For instance, the presence of a C or T at a single position would be represented as a Y (for pyrimidine) (Cornishbowden, 1985). Building upon the idea of consensus sequences, sequence logos are a way to display sequence motifs by giving

the relative frequencies of each nucleotide at each position (Schneider and Stephens, 1990).

A more informative representation of a degenerate motif is the PWM (positional weight matrices). PWMs take into account the observation that some positions are a site are more important than others (Stormo *et al.*, 1982). Most methods for searching for potential regulatory sites and therefore regulation of novel genes use PWM (Rodionov, 2007). The training set composed of regulatory regions can be from several sources such as experimentally-determined sites or DNA microarray experiments.

Comparative genomic approaches can also be used to determine novel regulatory sites with no *a priori* knowledge of the regulatory protein or the corresponding binding sites. For these *in silico* analyses, the upstream regions of potentially co-regulated genes, such as genes from the same metabolic pathway or involved in responding to the same stress, are used (McGuire *et al.*, 2000). There are two main comparative approaches to identification of novel regulatory elements. The consistency check approach assumes that sets of co-regulated genes are conserved in genomes that encode orthologous transcription factors (Mironov *et al.*, 1999, Tan *et al.*, 2001). Confidence in the site predictions is strengthened if it is found upstream of the same gene in several genomes. The phylogenetic footprinting approach makes the same assumption as the consistency check but it also assumes that regulatory sites diverge slower than non-regulatory sites (Blanchette and Tompa, 2002). This approach searches the upstream regions of orthologous genes for conserved regions. An important consideration in use of this approach is the selection of genomes; if the genomes are too closely related the alignment of upstream regions will be informative,

whereas sites are not sufficiently conserved in genomes too distantly related (Rodionov, 2007).

These techniques have led to the identification and validation of novel regulons including fatty acid biosynthesis (McCue *et al.*, 2001, Zhang *et al.*, 2002), ribonucleotide reductases (Qin *et al.*, 2003, Torrents *et al.*, 2007), N-acetylglucosamine utilization (Yang *et al.*, 2006a) and NAD metabolism (Rodionov *et al.*, 2008).

The SEED database and the subsystem-based approach

In 2005, the Fellowship for Interpretation of Genomes (FIG) released the SEED database, an annotation platform that puts genome annotation into the hands of experts (Overbeek *et al.*, 2005). The SEED is built around the idea of subsystems which are basically spreadsheets of gene sets from sequenced genomes. The user builds a subsystem around a particular metabolic pathway such as NAD biosynthesis (Gerdes *et al.*, 2006) or a process such as metal homeostasis (Rodionov *et al.*, 2006b) by adding functional roles that are known or predicted to be involved in that pathway or process. For a NAD biosynthesis subsystem some example functional roles would be ADP-ribose pyrophosphatase and nicotinamide phosphoribosyltransferase. For manganese homeostasis, the functional roles would be manganese-sensing regulators and manganese transporters. Once the functional roles are defined the subsystem is populated with genes that belong to those functional roles. Each row in the subsystem represents a genome and each column represents a functional role (Figure 1-6). The genes are found within the cells of the spreadsheet.

Once the subsystem is built, the user can then refine it by detecting mis-annotations, over-annotations or unannotated genes. Missing genes become immediately evident. The subsystem enables visual inspection of the presence of

proposed gene functions in genomes. If an organism is missing every gene in a given pathway but one, that one gene may be mis-annotated. If all genes for a given pathway are present but one, that one gene may be mis-annotated or locally missing. The SEED platform provides comparative genomic tools to finding these missing genes.

For instance, using the subsystem approach, novel enzymes required for the oxidation of D- and L- lactate to pyruvate were discovered. *Shewanella* spp. are able to use lactate as a sole carbon and energy source (Venkateswaran *et al.*, 1999), however, homology searches failed to detect orthologs of known lactate oxidizing enzymes (Serres and Riley, 2006). To solve this discrepancy, Pinchuk *et al.* built a “Lactate utilization” subsystem that contained all known functional roles involved in the utilization of lactate (Pinchuk *et al.*, 2009). Genes conserved and missing in *Shewanella* spp. and other organisms were determined. Using the tools available through SEED, a chromosomal cluster of uncharacterized genes was predicted to encode the missing D- and L-lactate dehydrogenases (Pinchuk *et al.*, 2009).

The comparative genomic evidence leading to this prediction is provided. First, these uncharacterized genes clustered with one of the characterized genes in the pathway. This cluster was found to be missing from the one *Shewanella* sp. that lacked the ability to grow on lactate. Phylogenetic co-occurrence analysis revealed a negative correlation between one of these uncharacterized genes and the gene encoding the canonical D-lactate dehydrogenase. The three other uncharacterized genes always formed a putative operon. There was also a negative correlation between these genes and the gene encoding the canonical L-lactate dehydrogenase in some organisms. The prediction that one of the uncharacterized genes encoded a novel D-lactate

dehydrogenase was tested and confirmed as was the prediction that the other three uncharacterized genes encoded a novel L-lactate dehydrogenase (Pinchuk *et al.*, 2009).

Comparative Genomics and Metals

Simulating metal-depleted environments in the laboratory has proven difficult because common metal chelators exhibit broad specificity that precludes targeted depletion of one specific metal from the culture medium. To observe growth defects linked to the deletion of genes involved in zinc homeostasis, in most cases, the absence of the high-affinity zinc transporter is required (Petrarca *et al.*, 2010, Gabriel and Helmann, 2009). By growing *E. coli* under continuous culture conditions in a specially designed metal-free chemostat, sufficient zinc-depletion was achieved to reveal growth defects in the wild-type *znuABC* background (Graham *et al.*, 2009). However, this approach is labor-intensive and not amenable to a broad study of zinc homeostasis mechanisms across the bacterial kingdom.

Fortunately, identification of putative Zur-binding sites has proven a productive way to identify novel proteins involved in the adaptation to zinc-limited environments. By using experimentally determined Zur-binding motifs and phylogenetic footprinting, PWMs were built and used to search available bacterial genomes (Panina *et al.*, 2003). The ZnuABC transporter imports zinc in an ATP-dependent manner and is thought to be the main target for Zur regulation (Patzer and Hantke, 2000). However, this analysis revealed that the Zur regulon in several bacteria is not limited to zinc transporters, and other mechanisms that allow adaptation to zinc-depleted environments were found.

As ribosomes most likely constitute the largest reservoir of zinc within the bacterial cell, it is no surprise that a system has evolved to utilize this stockpile when needed. A

comparative genomic analysis originally performed by Makarova, *et al.* led to the identification of four ribosomal protein paralog pairs: one copy contained a Cys4 zinc-ribbon motif (and are thus called C^+) and the other copy lacked two or more of those residues (and are thus called C^-) (Makarova *et al.*, 2001). At the time, it was not understood why some organisms retained both copies. The Zur regulon analysis by Panina, *et al.* provided an explanation. In genomes where both C^+ and C^- copies were found, the C^- copy was downstream of a Zur binding site (Panina *et al.*, 2003), leading to the hypothesis that when zinc levels dropped, the C^- ribosomal paralog would replace the C^+ copy in ribosomes liberating zinc for use by other zinc-requiring proteins (Panina *et al.*, 2003). This mechanism was subsequently experimentally confirmed in *B. subtilis* and *Streptomyces coelicolor* (Nanamiya *et al.*, 2004, Natori *et al.*, 2007, Shin *et al.*, 2007, Gabriel and Helmann, 2009).

Several other mechanisms were discovered by this comparative genomic analysis. ZinT, whose gene was fused to *znuA* in some genomes, was proposed to encode an additional component of the zinc ABC transporter (Panina *et al.*, 2003). This prediction was recently confirmed (Petrarca *et al.*, 2010). The presence of uncharacterized genes belonging to COG1469 and COG0523 were also found in the Zur regulons of several bacteria.

COG1469 was later shown to encode a GTP cyclohydrolase I, an example of non-orthologous gene displacement. In sequenced genomes, a negative correlation was found between the presence of the canonical GTP cyclohydrolase I and COG1469, which was subsequently shown to also have GTP cyclohydrolase I activity (El Yacoubi *et al.*, 2006). Regulation of COG1469 by zinc was later rationalized. Some genomes

contain both a gene for the canonical GTP cylohydrolase I and COG1469. It is in those genomes where COG1469 is predicted to be regulated by Zur. A model where COG1469 (a zinc-independent enzyme) replaces the canonical GTP cylohydrolase I (a zinc-dependent enzyme) was put forth and validated (Sankaran *et al.*, 2009). The function of the COG0523 genes or what their role in zinc homeostasis may be is still unknown.

Comparative genomic approaches have also been used to analyze the regulation of genes by iron, heavy metals, manganese, nickel and cobalt. A comparative genomic analysis of Fur regulons led to the identification of novel iron acquisition mechanisms in γ-proteobacteria, including novel siderophores and transporters (Panina *et al.*, 2001). In Eubacteria, heavy metal resistance is commonly regulated by transcription factors from the MerR family of regulators. Individual members are responsible for mercury, copper, cadmium, lead or zinc resistance. Using a comparative genomic approach that included phylogenetic analyses and identification of regulatory sites, the functions of uncharacterized MerR proteins could be determined and novel resistance genes were discovered (Permina *et al.*, 2006). Analysis of reconstructed manganese-responsive MntR and Mur regulons in α-proteobacteria revealed novel manganese transporters (Rodionov *et al.*, 2006a). Analysis of nickel-responsive NikR binding sites and B_{12}-responsive riboswtiches led to the identification of novel ABC-type transporters (Rodionov *et al.*, 2006b). These *in silico* studies reveal the advantages of using comparative genomics to indentify metal homeostatic mechanisms. A plethora of predictions in a large breadth of organisms are made with relative ease than can be used for targeted experimental validation.

The relative disadvantage of using traditional approaches to discover novel metal-related mechanisms is epitomized by efforts to study the response of *E. coli* to zinc-depletion. Studies on the effect of zinc-depletion on *E. coli* started at the end of the 20th century. These were made possible by the identification and characterization of high-affinity zinc transport and the corresponding zinc-sensing regulator (Patzer and Hantke, 1998, Patzer and Hantke, 2000). Five years later, Panina and colleagues built a putative Zur regulon in *E. coli* and over twenty other bacterial genomes. The first attempt to study the *in vivo* response of gene regulation to zinc-depletion employed assistance from the metal chelator TPEN. Over 100 genes were found to be differentially expressed in the presence and absence of TPEN (Sigdel *et al.*, 2006) compared to the five predicted by Panina. TPEN, however, is not specific to zinc and other divalent metal cations can be chelated. As a consequence, genes involved in iron and copper metabolism were detected in addition to any genes specifically responding to zinc-depletion. This experiment highlights the difficultly accompanying high-throughput studies of regulation. They often reveal both direct and indirect responses which cannot be easily teased apart.

To overcome the inherent problems with using metal chelators, Graham, *et al.* devised a method to specifically deplete cells of zinc through the use of continuous culture. Ten years after the discovery of ZnuABC and Zur, they were finally able to achieve specific depletion of zinc to see growth defects due to known homeostatic mechanisms. After specific zinc-depletion was achieved (85-fold reduction over normal minimal medium), a microarray analysis of cells grown in zinc-replete and zinc-deplete

conditions was performed. In stark contrast to the TPEN experiment, only nine genes were identified to be differentially expressed (Graham *et al.*, 2009).

In comparison to the *in silico* Panina study six years earlier, three genes were in common: *zinT*, *znuA*, and a C^{-} paralog. The Graham studied failed to detect induction of the permease and ATPase component of the high-affinity transporter, which were members of the Panina Zur regulon. Six genes in the Graham study were not found in the Panina study. None of these genes have previously been shown to be directly regulated by Zur and could be indirect effects of zinc-depletion.

Project Rationale, Design and Objectives

Discovery of novel proteins and an understanding of these mechanisms can be accomplished by the combination of comparative genomic predictions and targeted experimental studies. The main objective of this project was to gain insight into novel mechanisms that allow bacteria to grow in environments poor in available zinc.

The first goal of this project was to analyze the COG0523 protein family, which is composed of uncharacterized proteins and poorly characterized proteins with putative links to metal-related processes. As mentioned on page 43, COG0523 is a subfamily of the G3E family of GTPases, characterized members of which are involved in the maturation of metal-dependent enzymes. The hypothesis was formed that COG0523 proteins may also be chaperones for yet unknown enzymes. Since the characterized members of GE3 cluster physically on the chromosome with their targets, COG0523 genes may cluster with their target.

A comparative genomic approach, specifically a combination of phylogenomics and gene clustering, was chosen to examine this family in all domains of life. Gene clustering of COG0523 genes in prokaryotic chromosomes was determined in the

SEED database. From this analysis, representative homologs were chosen to reconstruct the phylogeny of this family. To infer the gene function of uncharacterized members, COG0523 proteins found in the literature were included. Known and novel roles were examined by overlaying the phylogenetic tree with gene clustering data, functional information from the literature, and regulatory site predictions. Prior to the start of this project, an update to the Panina Zur regulon analysis was performed and has been made publically available through the RegPrecise database (Novichkov *et al.*, 2010). This database includes regulon predictions for more recently sequenced genomes from proteobacteria, firmicutes and cyanobacteria. COG0523 homologs were frequently found in these putative regulons. A second group of COG0523 homologs were previously identified downstream of B_{12}-dependent riboswitches (Rodionov *et al.*, 2003). From the phylogenomic analysis, general and specific predictions were made. One of these specific predictions was chosen and experimentally tested in *E. coli* genetically.

The second goal of this project was to investigate the existence of paralogs for zinc-dependent enzymes found in the putative Zur reglons. For targeted experimental verification, the DksA paralog was chosen. *P. aeruginosa* was used to test the hypothesis that a DksA paralog missing an essential motif can functionally replace the characterized DksA protein and expression of the paralog is regulated by zinc. *P. aeruginosa* is an ideal organism to study how bacteria have evolved to thrive in growth conditions where access to zinc is relatively restricted. Such an environment is encountered in the human host. Therefore, an understanding of how pathogens like *P.*

aeruginosa can persist under zinc-depleted conditions can give valuable insight into infection.

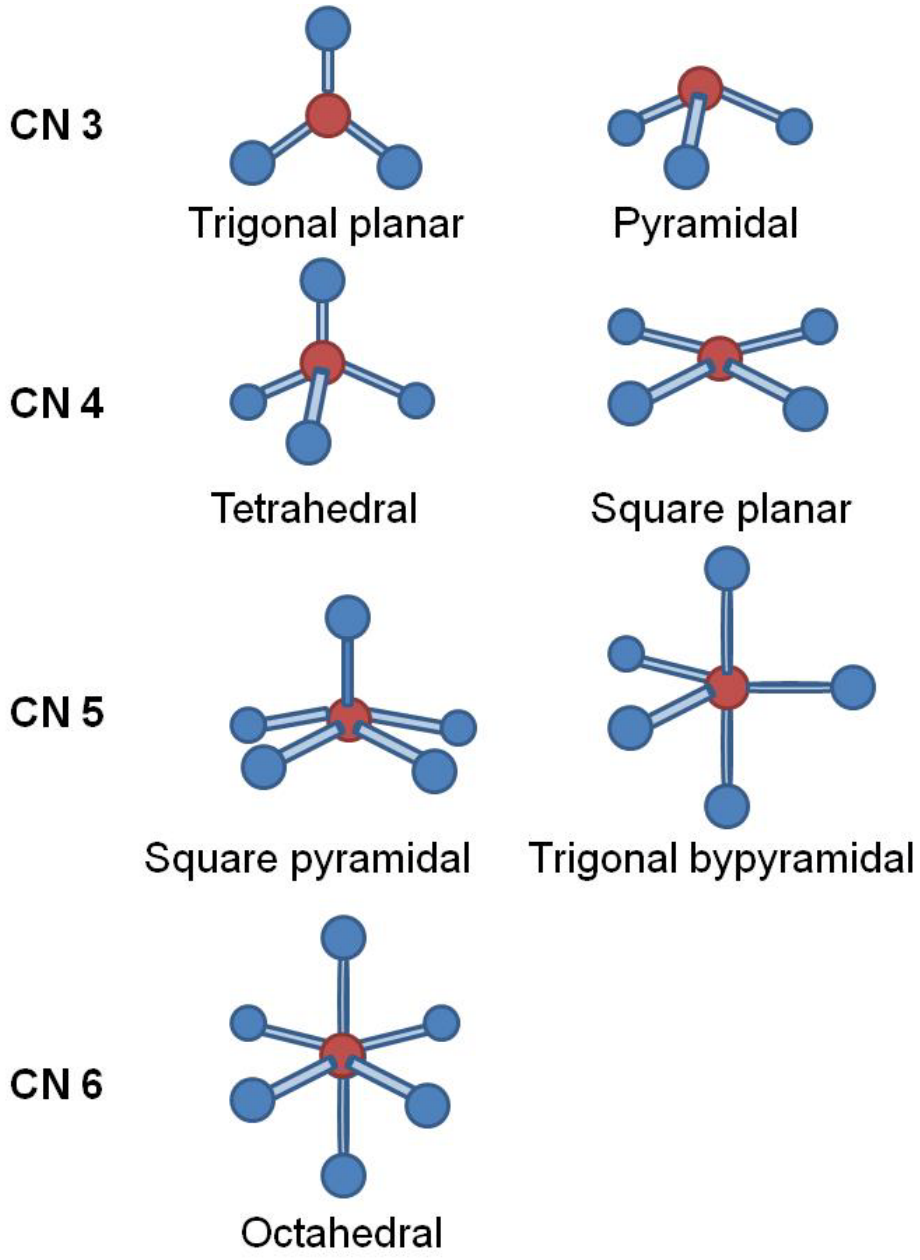
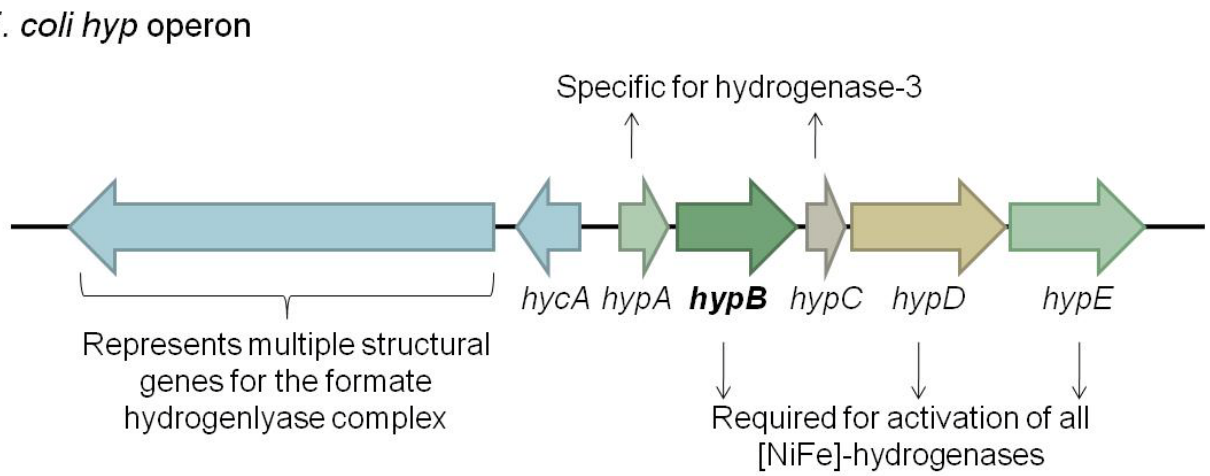


Figure 1-1. Typical coordination geometry involving metal ions. Metal ions are represented by red spheres. Metal-binding ligands and bonds to the metal are represented by blue spheres and blue sticks, respectively. CN, coordination number.

A



B

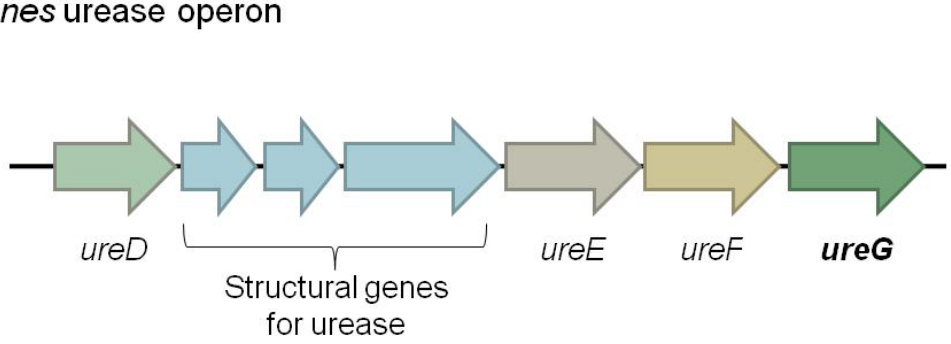


Figure 1-2. Organization of the *hyp* and *ure* gene clusters. A) Gene cluster from *E. coli* containing *hypB*. B) Gene cluster from *K. aerogenes* containing *ureG.*

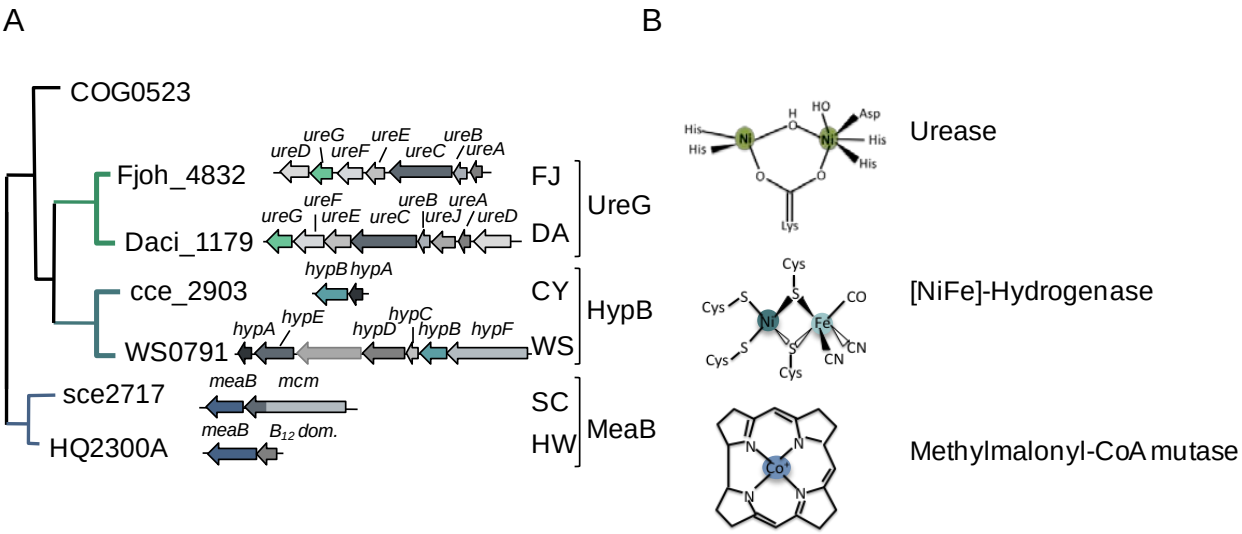


Figure 1-3. The G3E family of GTPases. A) A neighbor-joining tree of the G3E family divided into the four subfamilies, COG0523, UreG, HypB and MeaB. Representative gene clusters for *ureG*, *hypB* and *meaB* are given. With the exception of "COG0523", the leaves are protein names from the given organisms: FJ, *Flavobacterium johnsonia*; DA, *Delftia acidovorans*; CY, *Cyanothece* sp. ATCC 51142; WS, *Wolinella succinogenes*; SC, *Sorangium cellulosum*; HW, *Haloquadratum walsbyi*. Gene abbreviations: *ureA*, urease gamma subunit (EC 3.5.1.5); *ureB*, urease beta subunit (EC 3.5.1.5); *ureC*, urease alpha subunit (EC 3.5.1.5); *ureD*, urease accessory protein; *ureE*, urease accessory protein; *ureF*, urease accessory protein; *ureG*, urease accessory protein; *ureJ*, HupE-UreJ family metal transporter; *hypA*, [NiFe]-hydrogenase nickel incorporation protein; *hypB*, [NiFe] hydrogenase nickel incorporation-associated protein; *hypC*, [NiFe] hydrogenase metallocenter assembly protein; *hypD*, [NiFe] hydrogenase metallocenter assembly protein; *hypE*, [NiFe] hydrogenase metallocenter assembly protein; *hypF*, [NiFe] hydrogenase metallocenter assembly protein; *meaB*, methylmalonyl-CoA mutase auxillary protein; *mcm*, methylmalonyl-CoA mutase; *B_{12} dom.*, B_{12}-binding domain with homology to B_{12}- binding domain of methylmalonyl-CoA mutase. B) Cartoons of the metal-centers from the target metalloenzymes of UreG (urease), HypB ([NiFe]-hydrogenase) and MeaB (methylmalonyl-CoA mutase).

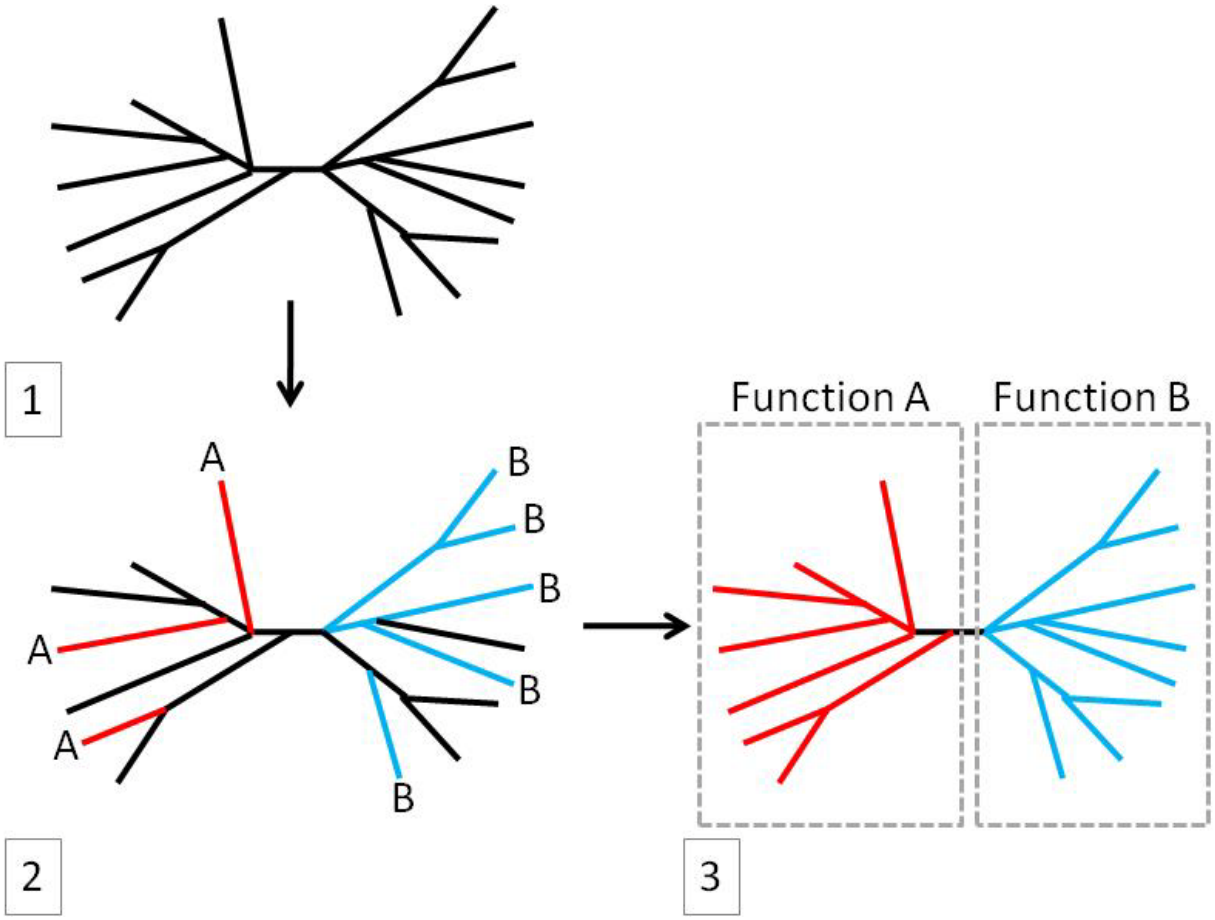


Figure 1-4. General schematic of phylogenomic analysis. In step 1, a phylogenetic tree composed of members from a protein family is constructed. In step 2, functional information from the literature is overlaid on the tree. A red line signifies that a protein has the characterized function “A”. A blue line signifies that a protein has the characterized function “B”. In step 3, function is inferred to uncharacterized members of the protein family.

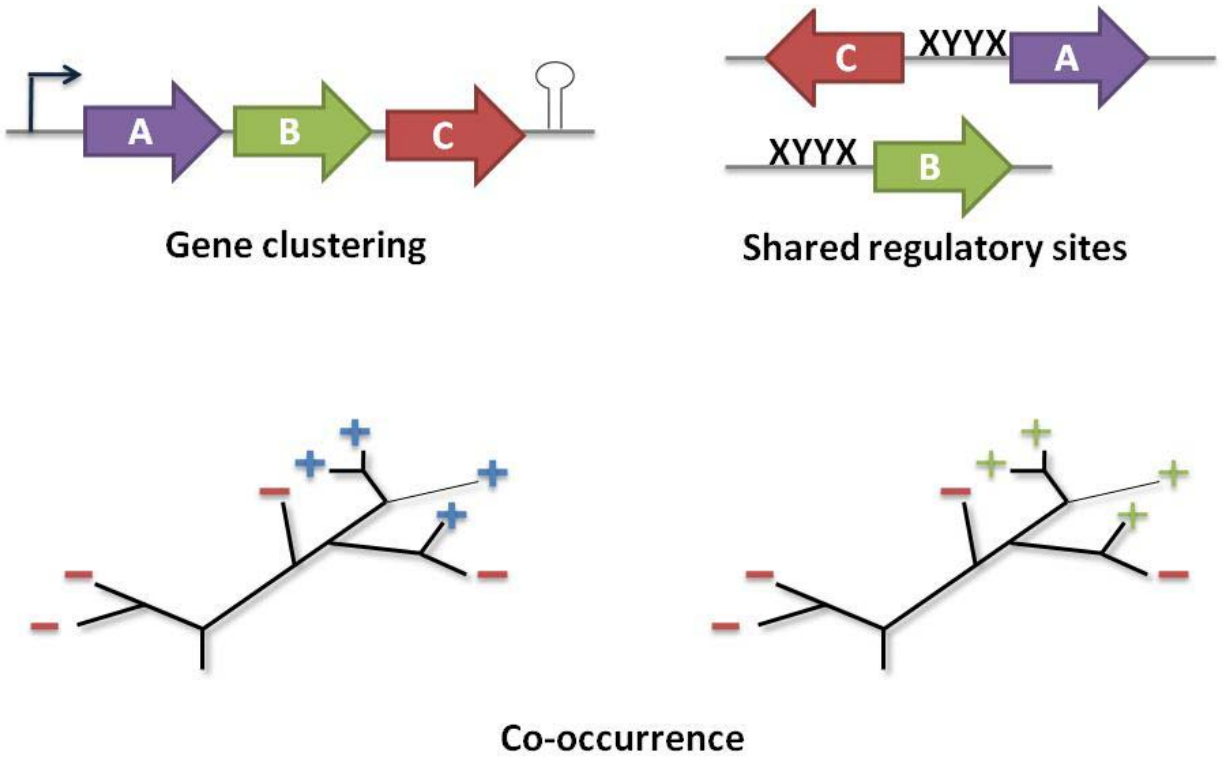
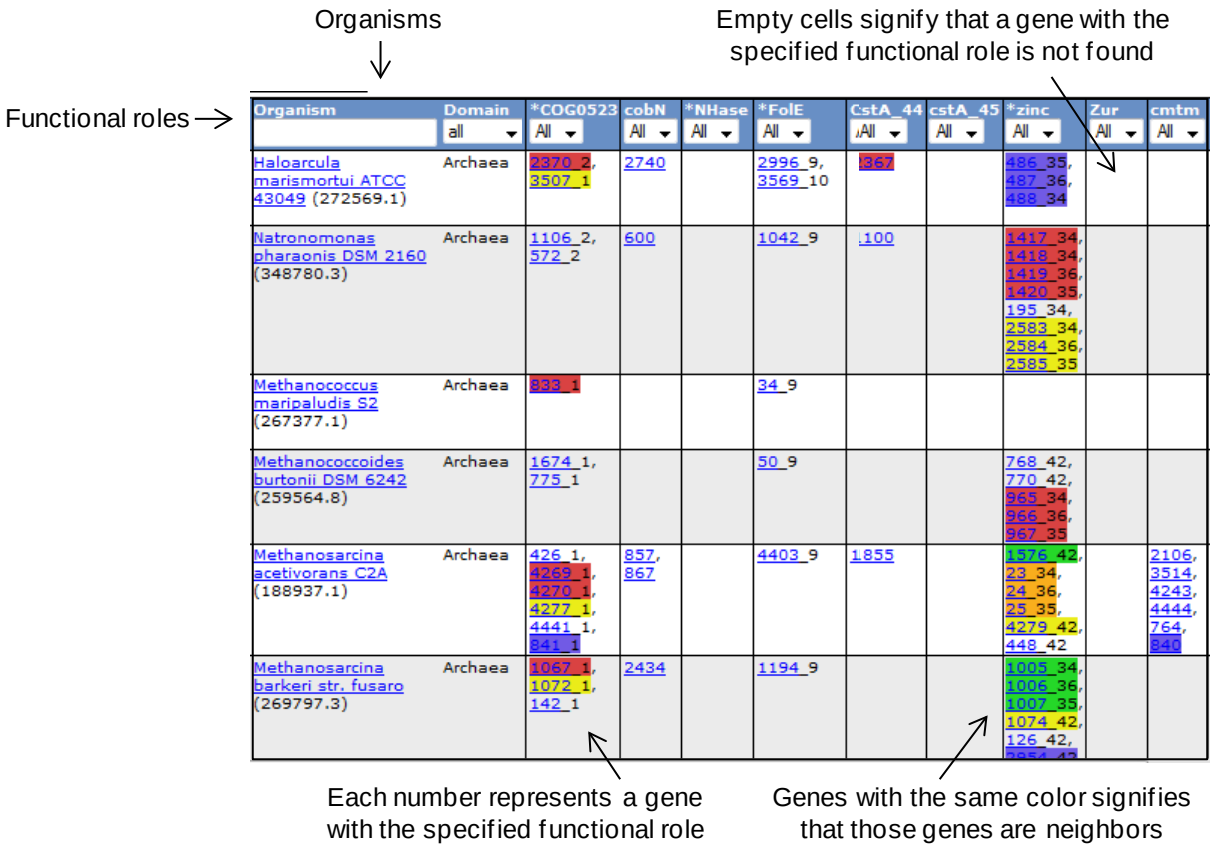


Figure 1-5. Schematic of common comparative genomic approaches to prediction gene function. The function of an unknown gene can be inferred by clustering or co-occurrence with genes of known function. In phylogenetic co-occurrence, the occurrence of genes in the same genomes is analyzed. Correlations in the phylogenetic distribution of genes relative to one another can be informative.



Organism	Domain	*COG0523	cobN	*NHase	*FolE	CstA_44	cstA_45	*zinc	Zur	cmtm
	all	All	All	All	All	All	All	All	All	All
Haloarcula marismortui ATCC 43049 (272569.1)	Archaea	2370_2, 3507_1	2740		2996_9, 3569_10	367		486_35, 487_36, 488_34		
Natronomonas pharaonis DSM 2160 (348780.3)	Archaea	1106_2, 572_2	600		1042_9	100		1417_34, 1418_34, 1419_36, 1420_35, 195_34, 2583_34, 2584_36, 2585_35		
Methanococcus maripaludis S2 (267377.1)	Archaea	833_1			34_9					
Methanococcoides burtonii DSM 6242 (259564.8)	Archaea	1674_1, 775_1			50_9			768_42, 770_42, 965_34, 966_36, 967_35		
Methanosarcina acetivorans C2A (188937.1)	Archaea	426_1, 4269_1, 4270_1, 4277_1, 4441_1, 841_1	857, 867		4403_9	1855		1576_42, 23_34, 24_36, 25_35, 4279_42, 448_42		2106, 3514, 4243, 4444, 764, 840
Methanosarcina barkeri str. fusaro (269797.3)	Archaea	1067_1, 1072_1, 142_1	2434		1194_9			1005_34, 1006_36, 1007_35, 1074_42, 126_42,		

Figure 1-6. Screenshot of a subsystem from SEED (http://theseed.uchicago.edu/FIG/). Terms discussed in the text are shown and descriptions for several elements are given. The term “Neighbors” is used to two or more genes that occur in a run as calculated in (Overbeek *et al.*, 1999a).

CHAPTER 2
MATERIALS AND METHODS

Chemicals and Strains

Materials

Restriction endonucleases, T4 DNA ligase, prestained Protein Marker broad range (7-175 kDa), DNA ladders, Taq DNA polymerase, Phusion® high-fidelity polymerase and Phototope®-Star Detection Kit were from New England BioLabs (Beverly, MA). Desalted and biotin-labeled oligonucleotides were from Integrated DNA Technologies (Coralville, IA). Organic and inorganic analytical grade chemicals were from Sigma-Aldrich (St. Louis, MO) and Fisher Scientific (Atlanta, GA). Metal salts ($ZnSO_4$, $CoCl_2$, $CuCl_2$, $MnCl_2$, $FeCl_3$ and $NiCl_2$) were purchased from Alfa Aesar (Ward Hill, MA) and were puratronic grade. Quantitative real-time polymerase chain reaction (qRT-PCR) reagents and materials, gel-filtration standard, 40% arylamide/bis-acrylamide solutions, Quick Start™ Bradford Protein Assay kit were purchased from Bio-Rad (Hercules, CA). TRIzol® LS reagent and One Shot® Top 10 chemically competent *E. coli* were purchased from Life Technologies (Carlsbad, CA). RNase*Zap*® and TURBO™ DNase were from Ambion (Austin, TX). 2X Laemmli sample buffer, polyvinyl difluoride membranes (0.45 µm pore size), Tris-buffered saline with 0.3% nonfat milk, alkaline phosphatase conjugate antibody developed in goat were from Sigma-Aldrich. Ni Sepharose™ High Performance resin was from GE healthcare (Waukesha, WI). QIAprep® plasmid purification kit, QIAquick® PCR and Gel extraction kits and RNeasy® mini kits were purchased from QIAGEN (Valencia, CA). For gel-shift assay, positively charged nylon membranes were purchased from Roche (Indianapolis, IN).

Strains, Plasmids and Oligonucleotides

Strains, plasmids and oligonucleotides used in each study are listed in Appendix F. *E. coli* K-12 MG1655 and *P. aeruginosa* PAO1 were used as WT strains. *E. coli* Top10 (Life Technologies) was used for routine cloning and *E. coli* BL21 DE3 was used for overexpression of recombinant protein.

General Growth Conditions

Media

E. coli K-12 MG1655, *P. aeruginosa* PAO1 and derivatives were routinely grown at 37°C in LB-Lennox medium (LB) or in minimal medium containing 1X M9 salts (Sambrook and Russell, 2001), 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 3 mg L^{-1} $FeSO_4 \cdot 7H_20$ and 0.2% (w/v) glycerol as the carbon source. Incubation of liquid cultures was routinely performed in a Multitron® INFORS shaker/incubator at 200 rpm. For *E. coli* EDTA and cadmium sensitivity assays, cells were grown in a low-phosphate (LP) minimal medium (Mergeay *et al.*, 1985) supplemented with 0.3% casamino acids and 0.2% glycerol. The iron citrate in this medium was replaced with 3 mg L^{-1} $FeSO_4 \cdot 7H_20$. When required for induction of P_{BAD}, arabinose was added at a concentration of 0.0002% - 0.2% (w/v) as specified. Liquid browth was solidified with 15 g of agar/liter.

For plasmid maintenance in *E. coli*, the medium was supplemented with 100 µg ml^{-1} ampicillin (Amp), 10 µg ml^{-1} tetracycline (Tet), or 15 µg ml^{-1} gentamicin (Gm).For marker selection, 50 µg ml^{-1} kanamycin (Kan) or 30 µg ml^{-1} chloramphenicol (Cam) was used for the appropriate antibiotic resistance gene. For *P. aeruginosa*, 30 µg ml^{-1} Gm was used for marker selection and 200 µg ml^{-1} Amp was used for plasmid maintenance.

Growth Curves

Overnight cultures of strains inoculated with a single colony were grown in 5 mL LB, washed once with the appropriate minimal medium, and then normalized to an optical density at 600 nm of 1.0 for *E. coli* and optical density at 660 nm of 1.0 for *P. aeruginosa*. Optical density was measured with a 1 ml disposable cuvette in a BioSpec-mini spectrometer. Normalized cultures were then diluted 1/500 into fresh minimal medium. Growth curves were generated with a Bioscreen C™ (Growth Curves USA, Piscataway, NJ) (37°C, intensive shaking, 600 nm wavelength) in triplicate unless otherwise specificed. All growth experiments were repeated independently at least three times.

Plate Assays

Overnight cultures grown in 5 mL LB were washed with the appropriate minimal medium and normalized to an optical density of 1.0 at 600 nm for *E. coli* and optical density of 1.0 at 660 nm for *P. aeruginosa*. Cells were washed once with an equal volume of the appropriate minimal medium, normalized to the same optical density, then serially diluted and 10 µL of each dilution were plated onto appropriate solid media. Normalization of cell number was confirmed by plating the various strains on minimal medium without selection. Growth was imaged at 24 hours after plating for *E. coli* and 36 or 67 hours for *P. aeruginosa.* All experiments were repeated independently at least three times.

Bioinformatic Techniques

Multiple Sequence Alignments

Protein sequences were downloaded from the SEED (Overbeek *et al.*, 2005, 2009) or Genbank (Benson *et al.*, 2009) databases. Sequences were aligned using the

ClustalW2 or MUSCLE algorithm with default parameters (Thompson *et al.*, 2002, Edgar, 2004). Alignments were edited using Jalview and viewed using ESPript 2.2 (Waterhouse *et al.*, 2009, Gouet *et al.*, 1999).

Subsystems

Subsystems were built in the SEED database (Overbeek *et al.*, 2005). Subsystems in their original format are available on the public SEED server at www.theseed.org.

Sequence Analysis

Identification of protein motifs was performed with Fuzzpro from the EMBOSS software package (Rice *et al.*, 2000). Sequence identity was determined with the Needleman-Wunsch Global Sequence Alignment Tool available at NCBI (Needleman and Wunsch, 1970).

Phylogenetic Tree Reconstruction

Phylogenetic analyses were carried out by employing the Phylip 3.67 program package (Felsenstein, 1997). Distance-based matrices were generated between all pairs of sequences using the Jones-Taylor-Thornton matrix as employed in Protdist (Phylip). Phylogenetic trees were generated from these matrices using the neighbor-joining method as implemented in Neighbor (Phylip). Reliability of branches was determined with the bootstrap method of 1000 replicates using Bootseq (Phylip). Tree visualization was performed with Treedyn (Chevenet *et al.*, 2006).

DNA Techniques

DNA Amplification

DNA was amplified by polymerase chain reaction (PCR) using Phusion® high-fidelity polymerase (for subsequent cloning) or Taq DNA polymerase (for verification).

PCR-amplified DNA was column purified with either QIAquick® gel or QIAquick® PCR purification kits according to manufacturer's directions (Qiagen).

Standard Molecular Cloning

PCR products and destination vectors were digested with the appropriate restriction enzyme(s) and 1X restriction buffer as per the manufacturer's recommendations (New England Biolabs). Ligation reactions were performed with T4 DNA ligase and 1X T4 ligase buffer at room temperature for 10 minutes or 4°C overnight. To ensure fidelity of the PCR reaction, cloned PCR-amplified products were sequenced using an Applied Biosystems Model 3130 Genetic analyzer (DNA Sequencing Facilities, Interdisciplinary Center for Biotechnology Research, University of Florida).

DNA Electrophoresis

The size of PCR products and linearized plasmids relative to a NEB 1 kB DNA ladder or NEB 100 bp DNA ladder were analyzed by horizontal agarose gel electrophoresis using 0.8% - 2% (w/v) agarose gels containing in 1X TAE buffer (40 mM Tris acetate, 2 mM EDTA, pH 8.5). Samples (5 µl) were mixed with 1 µl loading buffer (0.25% (w/v) bromophenol blue, 40% (w/v) sucrose) and loaded onto the gel. Following sufficient separation time (typically 45 min at 100V), the gel was visualized under UV light using a Kodak Gel Logic 2200 Imaging System with Kodak molecular imaging software (Carestream, New Haven, CT).

Plasmid Isolation and Transformation

When high-purity plasmid DNA was required, plasmid DNA was isolated with a QIAprep® mini kit according to manufacturer's directions (Qiagen). For checking plasmid constructs by restriction enzyme digestion, 2 ml of an overnight culture of cells

harboring the plasmid grown in LB were centrifuged at 14,000 rpm in a benchtop micro-centrifuge (Eppendorf 5417C) for 5 min. The cell pellet was resuspended in 50 µl lysis buffer (0.2 N NaOH, 0.5% (w/v) SDS, 20% (w/v) sucrose) and vortexed. The mixture was then boiled for 5 min and allowed to cool to room temperature. 1.5 µl 4 M KCl was added and the solution was vortexed. Cell debris was removed by centrifuging 3 min at 14,000 rpm (Eppendorf 5417C centrifuge). 5 µl of the supernatant was removed used in a 10 µl restriction enzyme digestion with appropriate restriction enzymes according to the manufacturer's guidelines (NEB).

E. coli Top 10 chemically-competent cells were transformed as per manufacturer's directions (Life Technologies). *E. coli* MG1655 derivatives and BL21 (DE3) were made competent with $CaCl_2$. 2 ml of LB in a 5 ml capped test tube was inoculated with 50 µl of an overnight culuture grown in LB. The culture was allowed to incubate with shaking (200 rpm). At the end of 2 hrs, the culture was centrifuged at 4,000 rpm (Eppendorf 5417C centrifuge) and resuspended in 200 µl 0.1 M $CaCl_2$. 50 ng of plasmid DNA was added to the cells and allowed to incubate on ice. At the end of 20 min, the cells were heat-shocked at 42°C for 35 sec. The cells were inbated on ice for 2 min, 1 ml LB was added and the cells were allowed to recover at 37°C with shaking for 1 hr. The cells were then plated on solid LB medium supplemented with the appropriate antibiotic.

P. aerginosa strains were transformed by electroporation as previously described (Choi *et al.*, 2006) with the following changes. Instead of using an overnight culture, 50 ml of LB were inoculated with 1 ml of an overnight culture and grown for 2 hours at 37°C then washed in preparation for electroporation.

Site-Directed Mutagensis

Site-directed mutants were created by the overlap extension or megaprimer PCR method as described in (Sambrook and Russell, 2001). Phusion polymerase was used for each PCR reaction. QIAquick® gel extraction kit was used for purification of each PCR product. For expression in *E. coli*, mutated genes were inserted into pBAD24 (Guzman *et al.*, 1995) between the unique *Nco*I and *Xba*I restriction sites. Primers used in the site-directed mutagenesis are listed Appendix F, Table F-5 and Table F-6.

P1 Transduction

Transfer of mutant alleles between *E. coli* strains was performed by bacteriophage P1 transduction as described by Miller (Miller, 1972). When required, the resistance marker was excised from the chromosome using FLP recombinase (Cherepanov and Wackernagel, 1995). Deletions were confirmed by locus-specific PCR amplification using primers that anneal to DNA either flanking the gene or anneal internal to the gene.

Generation of *P. aeruginosa* Mutants

The Gm resistance gene from pEX18Gm was PCR amplified with primers containing FRT sites (Zhou and Sadowski, 1994) at the 5' end (Hoang *et al.*, 1998). Upstream and downstream regions flanking the gene of interest were PCR amplified from PAO1 genomic DNA and gel extracted using the QIAquick® gel extraction kit. Generation of the gene deletion construct was performed by PCR overlap extension as previously described (Choi and Schweizer, 2005) and inserted into the appropriate restriction enzyme sites of pEX18Tc (Hoang *et al.*, 1998). For gene replacement, a *sacB*-based strategy (Schweizer and Hoang, 1995) was employed. The appropriate parent strain was transformed with the pEX18Tc-derivative by electroporation. After recovery in LB, shaking at 200 rpm for 2 hr, the cells were plated on LB solid medium

plus Gm and incubated overnight. Gm resistant colonies were screened for Tet sensitivity. If only single homologous recombination events occurred as indicated by resistance to Gm and Tet, those colonies were inoculated into 5 ml LB, grown overnight, and plated onto LB (minus NaCl) and 10% (w/v) sucrose to select for excision of the plasmid from the chromosome. Deletions were confirmed by PCR using both primers external to the upstream and downstream flank regions and primers internal to the gene. When required, the Gm cassette was excised with FLP recombinase (Hoang *et al.*, 1998).

RNA Techniques

RNA Isolation

For whole-cell RNA isolation, a 5 ml aliquot of culture was taken at the appropriate time from a growing culture in the appropriate medium, then centrifuged at room temperature and the cell pellet was resuspended in TRIzol® LS reagent then frozen at -80°C. Once all the samples were collected, the samples were thawed at room temperature and RNA was extracted with chloroform. The aqueous phase was then further purified using the RNeasy® mini kit according to manufacturer's directions (Qiagen). Trace DNA was removed with TURBO™ DNase according to manufacturer's guidelines (Ambion). Total RNA concentration was determined using a NanoDrop™ ND-1000 spectrophotometer at an absorbance of 260 nm.

cDNA Synthesis and qRT-PCR

cDNA was reverse-transcribed from total RNA using an iScript™ cDNA synthesis kit. qRT-PCR reactions were performed using iQ™ SYBR® Green Supermix, according to the manufacture's guidelines. Routinely, one nanogram of total mRNA was reverse-transcribed, and then 1 μl of cDNA was added to a 20 μl SYBR® Green reaction mix

containing 2.5% DMSO. qPCR on the generated cDNA was conducted in an iCycler MyiQ™ 2 real-time system (Bio-Rad) in triplicate. Reactions containing 0.1 ng of total mRNA from each sample served as controls for DNA contamination. Primers used in the qPCR reactions are listed in Appendix F, Table F-6. Standard curves were generated with serial dilutions of plasmid DNA containing the target gene. The cycling conditions were as follows: 1 cycle at 95°C for 3 min, 40 cycles of 95°C for 10 sec and 58°C for 30 sec. Product uniformity was determine using melt curves. Data was analyzed using iQ™ 5 optical system software (Bio-Rad) and the quantity of transcript presenct was determined by comparison to the standard curve.

Protein Techniques

Protein Overexpression in *E. coli*

Overexpression of the Zur protein from *P. aeruginosa* was achieved as previously described (Gabriel *et al.*, 2008). Briefly, 6 liters of LB supplemented with a final concentration of 50 µM $ZnSO_4$ and 100 mg ml^{-1} Amp was inoculated with 10 mL per 1 L from an overnight culture grown in LB. Cells were grown to an optical density of 0.8 at 600 nm. IPTG (final concentration, 1 mM) was added and the cultures, which were then incubated at 30°C for 6 hrs.

Protein Quantification

Protein concentrations were determined using a Quick Start Bradford kit with bovine gamma-globulin as the standard as per the manufacturer's directions for 1 ml assays (BioRad). This kit is based on the Bradford method (Bradford, 1976).

Protein Separation and Chromatography

Cells were harvested by centrifugation at 5,000 rpm (SORVALL® RC-5B Superspeed® centrifuge) at 4°C for 20 min. The pellet was stored at -80°C. The cell

pellet was resuspended in ice-cold lysis buffer (20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.3 and 2 mM DTT) and cells were lysed using TEEN B Lysing matrix and a FastPrep®-24 (MP Biomedicals, Aurora, OH). The lysate was cleared by centrifuging at 5,000 rpm at 4°C for 20 min. Lysate was then transferred to 2 ml microcentrifuge tubes and centrifuged at 13,500 rpm (Eppendorf 5425R benchtop centrifuge) at 4°C for 20 min in preparation for downstream chromatography.

Cell lysates (to determine overexpression of target gene) and chromatography fractions were monitored for purity by staining with Coomassie blue after separation by 10 or 12% SDS-PAGE according to (Sambrook and Russell, 2001). Size of the protein band was estimated based on comparison to a broad range protein marker (BioRad).

Cleared lysate was loaded at 0.5 ml min^{-1} on a 25 mL Ni Sepharose™ High Performance column pre-equilibrated with 200 ml lysis buffer. The column was then washed with 200 ml wash buffer (lysis buffer with 20 mM imidazole instead of 10 mM). Protein was eluted with a linear gradient of imidazole from 20 mM to 500 mM; 65-5 ml fractions were collected.

Immunoblotting

Cell lysates were separated by gel electrophoresis using 15% SDS-PAGE gels. Proteins was transferred to an Immobilon-P polyvinylidene fluoride (PVDF) membrane using a Mini Trans-Blot® cell and a Mini-PROTEAN® 3 vertical electrophoresis system (Bio-rad) with 1X Tris-glycine buffer (3.029 g L^{-1} Tris-Base, 14.41g L^{-1} glycine) at 40 mAmps for 90 min. Transferred membranes were blocked with Tris buffered saline with 0.3% nonfat milk for 1 hr at room temperature. Primary antibody, polyclonal DksA or DksA2 generated in rabbit (Blaby-Haas, *et al*, 2010), was prepared by diluting 1:500 in blocking buffer. The membrane was incubated with antibody for 1 hr then washed 3

times with 1X wash buffer (1.21 g L^{-1} Tris-Base, 8.77 g L^{-1} NaCl, 5 ml 10% Tween, pH 7.4 with HCl). For each wash, the membrane was incubated for 10 min. The secondary antibody (anti-rabbit IgG (whole molecule) alkaline phosphatase conjugate antibody developed in goat) was prepared by diluting 1:5000 in blocking buffer. The membrane was incubated with the secondary antibody for 1 hr, then washed 3 times for 10 min each wash with wash buffer. CPD-Star® reagent was added and the membrane was exposed to a Kodak x-ray film for 30 sec to 15 min. Western blot analysis was independently repeated twice. Specificity of primary antibodies was confirmed by blotting against whole cell extracts from *P. aeruginosa* PAO1, Δ*dksA*, and Δ*zur* grown in LB.

Electrophoretic Gel Shift Assays (EMSA)

Preparation of DNA substrate

The DNA fragment used in the EMSA experiment was generated by PCR. One oligonuclotide was prelabeled at the 5'-end with biotin. The PCR fragment was purified using a QIAquick® spin column. Additionally, unlabeled DNA fragments for competition assays were generated by annealing two complementary oligonucleotides combined at equimolar concentration in 1X T4 ligase buffer. The oligonucleotide mixture was incubated at 95°C for 5 min in a heat block. The heat block was removed from the incubator and allowed to cool to room temperature.

Gel-shift assays

Protein of varying concentration and 1.5 ng of biotin-labeled DNA was added to 20 µl of binding buffer (20 mM Tris-HCl (pH 8.0), 50 mM KCl, 1 mM DTT, 5% glycerol, 0.1 mg ml^{-1} bovine serum albumin, 5 µg ml^{-1} sheared salmon sperm DNA). The mixture was incubated for 20 min at room temperature. Samples were loaded on a 6% acrylamide-

bisacrylamide nondenaturing gel in 40 mM Tris-acetate buffer (without EDTA). Electrophoresis was performed at 120 V and 4°C using pre-chilled 40 mM Tris-acetate buffer (without EDTA) as the running buffer.

DNA was then transferred to a positively charged nylon membrance using a Mini Trans-Blot® cell and a Mini-PROTEAN® 3 vertical electrophoresis system at 400 mA for 40 min in 0.5x TBE. Transferred DNA was cross-linked with a Fisher-Scientific FB-UVXL-1000 UV Crosslinker by using the optimal cross-link setting. Biotin-labeled DNA was detected using a Phototope®-Star Detection Kit following manufacturer's directions. Membranes were exposed to a Kodak x-ray film for 10-15 min. Density of DNA bands was measured using Kodak molecular imaging software (Carestream, New Haven, CT).

Pyocyanin Assay

Overnight cultures of strains grown in LB were normalized to an optical density of 1.0 at 660 nm and then diluted 100-fold into fresh LB. The cultures were incubated at 37°C. At different time intervals, a 5 ml aliquot of culture was transferred to a 15 ml centrifuge tube and centrifuged at 4,000 rpm (SORVALL® RC-5B Superspeed® centrifuge) for 10 min. The supernatant was transferred to a fresh 15 ml centrifuge tube. Pyocyanin was then extracted from the supernatant with 3 ml chloroform, which was then centrifuged at 4,000 rpm for 5 min. The chloroform phase (will be a blueish-green color) was removed and 1 ml 0.2 M HCl was added. The sample was vortexed and then centrifuged at 4,000 rpm for 5 min. The HCl phase (a pinkish color) was removed into a 1 ml plastic cuvette. Absorbance of the HCl-pyocyanin phase was measured at 520 nm on a BioSpec-mini spectrometer. A sample of uninculated medium that was also processed as per above and was accordingly used as the blank. The concentration of

pyocyanin was calculated by multiplying the absorbance at 520 nm by 17.072 (Kurachi, 1958).

Detection of Hydrogen Production

Overnight cultures of strains grown in LB were diluted 100-fold into 1 ml of LB supplemented with 0.3% glucose in a 12 x 75 mm heavy wall tube, which was then sealed with a rubber stopper. The tubes were degassed and filled with N_2 and incubated. Hydrogen production was determined after injecting a 100 µl sample with a Hamilton gas tight syringe into a gas chromatograph.

Element Analysis

Preparation of cell fractions and element analysis was performed per the prorocol provided by (Robinson, N. *et al*, 2008). This protocol is detailed below.

Sample Preparation

Overnight cultures grown in LB were diluted 1000-fold into 500 ml LB. The cultures were grown to an optical density of 0.7 at 600 nm. The cells were centrifuged and washed twice with 20 mM HEPES (pH 8.8) and resuspended in 5 ml of the same buffer. A pestle and mortar was equilibrated in liquid nitrogen. Once equilibrated, the cell suspension was added to the liquid nitrogen is a drop-wise manner and then the cells were ground thoroughly to a fine powder. The cell powder was placed in an anaerobic chamber and allowed to thaw for 10 min. The cell paste was placed in an airtight centrifuge tube and centrifuged at 160,000 X g, 30 min, 4ºC. The supernatant was transferred in the anaerobic chamber to a fresh 15 ml centrifuge tube.

Anion Exchange Chromatography

In the anaerobic chamber, the cleared lysate (corresponding to 30 mg of protein) was loaded onto a 1 ml HiTrap™ Q HP anion exchange column (GE Healthcare) and

washed with 10 ml HEPES (pH 8.8). Protein was eluted by step gradient using 1 ml aliquots of the same buffer with 100 mM, 200 mM, 300 mM, 400 mM, 500 mM and 1 M NaCl.

Size Exclusion Chromatography

Each fraction from the anaion exchange separation was then further separated by HPLC size exclusion chromotagraphy. 200 ul aliquots from each fraction of eluant was loaded onto a TSK-SW3000 column (Tosoh Biosciences) pre-equilibrated with 10 mM Tris-HCl (pH 7.5), 1 M NaCl and 10 mM EDTA at 1.0 ml min^{-1}. 30 - 0.5 ml fractions were collected.

ICP-MS

Three hundred microliter aliquots of each fraction from the size exclusion separation were diluted into 2.7 ml 2.5% (v/v) Suprapur® HNO_3 (Merck) and analyzed by ICP-MS on a Thermo Electron Corporation X-series inductively-coupled mass spectrometer (Kevin Waldron, University of Newcastle, Newcastle, England).

CHAPTER 3
BIOINFORMATIC ANALYSIS OF THE COG0523 FAMILY

Background

Although the other subfamilies of the G3E family of P-loop GTPases are relatively well characterized, only sparse reports are found on COG0523 proteins. The first member of the COG0523 family to be identified was found in *Pseudomonas denitrificans* and named CobW (Crouzet *et al.*, 1991). Disruption of *cobW*, which is found in a gene cluster with cobalamin synthesis genes, leads to a defect in the production of cobalamin (Crouzet *et al.*, 1991). Cobalamins are organometallic compounds that are composed of a Co(III) bound in an octahedral configuration to a corrin ring (Hodgkin *et al.*, 1955). The proposed role for CobW is in presentation of cobalt to the cobaltchetelase component of the pathway, which is responsible for inserting cobalt into the corrin ring (Heldt *et al.*, 2005). However, 20 years after its discovery, there is as of yet no experimental evidence for this function. Since "cobalamin biosynthesis protein" was the first role assigned to a member of COG0523, this annotation has been propagated by sequence similarity (e.g. BLAST) to nearly all COG0523 genes in current databases.

Other partially characterized members of COG0523 include the nitrile hydratase activators, which are required for iron-type nitrile hydratase (NHase) activity (Nojiri *et al.*, 1999). NHases are important enzymes in the industrial production of acrylamide and employ either an octahedral non-heme iron (III) or non-corrin cobalt (III) in the hydration of nitriles to amides (Yamada and Kobayashi, 1996). The active site residues of both NHase metal-types are highly conserved and the same coordination geometry was determined for both Co(III) and Fe(III) (Endo *et al.*, 2001). The two types of NHases,

however, explicitly incorporate the correct metal (Nojiri *et al.*, 2000). This specificity is thought to be due to an activator protein, which is required for full activity (Nojiri *et al.*, 2000, Nishiyama *et al.*, 1991, Zhou *et al.*, 2008, Nojiri *et al.*, 1999, Hashimoto *et al.*, 1994). In the case of iron-type NHase, the corresponding activator protein is a COG0523 protein and is referred to as Nha3. As with CobW, even if the involvement of Nha3 in iron-type NHase activation is documented, its exact role is not known.

It has been postulated that it has an insertase role involved in incorporation of iron into the active site of the hydratase (Lu *et al.*, 2003). When the iron-dependent NHase from *Rhodococcus* sp. N-771 was expressed in *E. coli* without Nha3 in cobalt-supplemented media, it incorporated cobalt instead of iron (Nojiri *et al.*, 2000). Therefore, Nha3 may not only be involved in incorporating iron, but also in ensuring that competing metal ions are excluded. In addition, the coexpression of Nha3 with NHase was found to be unnecessary with the coexpression of the GroESL chaperones (Stevens *et al.*, 2003), which suggests that Nha3 is involved in ensuring proper folding of NHase.

The most recently studied COG0523 protein is YciC from *B. subtilis*. Several observations led to the assumption that *yciC* coded for a component of a low-affinity zinc transporter (Gaballa and Helmann, 1998). The gene, *yciC*, is repressed by Zur, a zinc-responsive transcription factor, and deletion of *yciC* in combination with a deletion in the high-affinity zinc transporter results in an EDTA-sensitive growth defect (Gaballa and Helmann, 1998, Gaballa *et al.*, 2002). As with CobW and Nha3, the mechanism responsible for this phenotype remains elusive.

As a result of these studies, genes encoding a COG0523 protein have been arbitrarily assigned either a function in cobalamin biosynthesis, in the activation of NHase, or as a low-affinity zinc transporter without the means to automatically distinguish between these different functions. Nevertheless, each of these functions is related in a general sense to intracellular metal handling. The diversity of the metals putatively associated with COG0523, cobalt, iron, or zinc, suggests that there might be different subgroups identifiable within the COG0523 family.

This chapter presents the efforts to better understand the COG0523 family through comparative genomic techniques. COG0523 proteins occur in all kingdoms of life, and most sequenced genomes encode one or more homologs. An analysis is presented with the purpose to distinguish between the putative functional subcategories that are evident in the literature and enable predictions for the characterization of uncharacterized COG0523 genes. Particularly, the members of COG0523 found in eukaryotic and archaeal genomes are completely uncharacterized. Analysis and functional assignment of prokaryotic genes has been demonstrated to be a powerful technique to annotate eukaryotic genes (de Crécy-Lagard and Hanson, 2007), indicating the importance of investigating the family with bioinformatic tools. Furthermore, to enrich these analyses, the results are placed in context with the current understanding on the other (non-COG0523) members of the G3E family of P-loop GTPases.

Results

Sequence Attributes

The amino acid sequences of 887 COG0523 proteins from all kingdoms of life were compared. COG0523 gene sequences in the SEED database were identified by

homology to known COG0523 members and the presence of the conserved CXCC motif and P-loop GTPase domain as found in the corresponding protein sequences. This definition led to the reannotation of genes that did not meet these criteria. *cobW* gene sequences were identified based on homology to *cobW* from *P. denitrificans* (Crouzet *et al.*, 1991) and occurrence within cobalamin biosynthesis operons and/or downstream of a putative B_{12}- responsive riboswitches (Rodionov *et al.*, 2003), a regulatory RNA element modulating gene expression in response to changing B_{12} concentrations (Nahvi *et al.*, 2002).

Comparison with other G3E family members

The region of highest similarity between COG0523 and the other members of the G3E family (G3E) was the GTPase domain, defined by the canonical Walker A, Walker B, and G4 motifs (Figure 3-1) (Leipe *et al.*, 2002). Located within the GTPase domain, all members of COG0523 had a conserved, putative, metal-binding CXCC motif (Figure 3-1).

In addition to the GTPase domain, MeaB and most COG0523 proteins contained an additional C-terminal domain that was not found in UreG or HypB. On average, COG0523 was 99 and 147 amino acids larger than HypB and UreG, respectively, and only 26 residues larger than MeaB. While the N-terminal GTPase domain was well conserved among COG0523 members, this extra C-terminal domain was highly variable (Figure 3-1).

Some HypB proteins are known to contain histidine-rich regions (His-stretches) (Fu *et al.*, 1995). Therefore, the presence of these motifs in COG0523 proteins was examined. Located in the C-terminal domain, approximately 40% of the sequences analyzed contained a histidine-rich region (Figure 3-1); 365 COG0523 proteins contain

the minimal HXHXHXH motif, where X represents 0 - 4 residues. Some proteins contain a His-stretch with up to 29 histidines, such as Ava_3717 from *Anabaena variabilis:* HSHDHHDHDHDHDHSTCEHDHHDHEHDHSACSHDHHDHDHSACGHDHHDHEHHHHHSDH.

Correlation between His-stretches and metallochaperone activity

In [NiFe]-hydrogense metallocenter biosynthesis, HypB is proposed to serve as a metallochaperone, delivering nickel to the maturation complex, which is facilitated by the presence of the His-stretch (Fu *et al.*, 1995, Olson *et al.*, 1997). The HypB protein from *E. coli* does not have a His-stretch, and, accordingly, a separate protein is proposed to serve as the metallochaperone component (Zhang *et al.*, 2005). Therefore, the presence of a His-stretch could be indicative of metallochaperone activity.

Unlike HypB, most UreG proteins do not contain His-stretches and a separate protein UreE is proposed to serve as the metallochaperone (Remaut *et al.*, 2001, Benoit and Maier, 2003, Musiani *et al.*, 2004). However, an analysis of UreG proteins in the SEED database revealed that several UreG proteins do contain His-stretches. This observation provided a chance to test the hypothesis that the presence of His-stretches in G3E family proteins is evidence of metallochaperone activity.

A comparative genomic analysis involving *ureG* was performed (Appendix A, Figure A-1). As with COG0523, a distribution of UreG proteins with and without a His-stretch was found (Table 3-1). Furthermore, in genomes where UreG had a His-stretch, the gene encoding UreE was missing (Table 3-1). The absence of *ureE* did not correlate with the presence of *slyD*, the metallochaperone component of hydrogenase maturation (Zhang *et al.*, 2005), as would be expected if SlyD performed the metallochaperone role

in those organisms (Table 3-1). *B. japonicum* USDA 110 and *Frankia* sp. Ccl3 were two exceptions to this trend as they lacked both *ureE* and a His-stretch extension in UreG.

Phylogenomic Analysis

Literature reports suggest that COG0523 can be divided into subgroups based on putative function even though the exact mechanism behind those functions is unknown. These known functional subgroups include cobalamin biosynthesis, NHase activation, and zinc homeostasis. As of yet, efforts have not been made to determine the extent to which COG0523 can be divided among these subgroups and whether other roles are possible.

Gene clustering

The non-COG0523 members of G3E (*hypB*, *ureG* and *meaB*) were commonly found next to either genes encoding their target metalloenzyme or other accessory factors that are also involved in metallocenter biosynthesis (Appendix A, Figure A-2). Therefore, a comparative genomic analysis was performed to investigate the gene neighborhoods containing COG0523 genes (Appendix A, Figures A-3 and A-4). Previously, an analysis of cobalamin biosynthesis pathways in bacteria revealed that *cobW* was consistently found next to other cobalamin biosynthesis genes (Rodionov *et al.*, 2003) and the NHase activator was found downstream from structural genes for NHase (Nojiri *et al.*, 1999, Hashimoto *et al.*, 1994). As summarized in Figure 3-2 and detailed in Appendix B, COG0523 genes were found in fourteen significant gene clusters. Significance was determined by the SEED database score and corresponds to the number of genomes a gene cluster was found and the phylogenetic distance between those genomes (Overbeek *et al.*, 1999a). It should be noted that there may be functionally significant gene clusters whose scores were not above the threshold set by

this algorithm. The interaction between NHase and COG0523 has been experimentally shown but the NHase/COG0523 gene cluster was not detected in this computational analysis.

In comparison to the other G3E members, COG0523 was more widespread and more promiscuous in regards to gene neighborhoods. 32.8% of prokaryotic genomes within the SEED database contained *ureG* and 96% of those genes were in a putative operon with the other known urease accessory factors. 29.5% of prokaryotic genomes contained *hypB* and 95% of those genes were in a putative operon with characterized hydrogenase accessory factors. *meaB* was found in 26% of genomes and 63% of those genes were in a cluster with either the gene that encodes MCM or a gene that encodes the adenosylcobalamin-binding domain of MCM. In contrast, 54.6% of prokaryotic genomes contained at least one COG0523 gene and these genes were found in at least 14 significant gene neighborhoods (Figure 3-2; Appendix B).

Most genomes contain only one homolog of *hypB*, *ureG*, or *meaB* (Appendix A, Figure A-2). Conversely, 62.4% of prokaryotic genomes that contained a COG0623 gene had at least two. Up to 11 COG0523 genes could be found in a single genome, as seen in *Cyanothece* sp. ATCC 51142.

Phylogenetic reconstruction

To analyze the presence of multiple COG0523 subgroups, gene neighborhood analyses performed with the SEED database, literature reports, and gene regulation predictions provided by (Rodionov *et al.*, 2003), the RegRecise collection (Novichkov *et al.*, 2010), and the SEED database (Overbeek *et al.*, 2005) were mapped onto a phylogenetic tree reconstruction of representative COG0523 proteins. The COG0523

distance tree was built with 177 full-length COG0523 sequences. Protein names can be found in Appendix D, Table D-3. These sequences were chosen based on three criteria.

First, from each of the 14 COG0523 gene clusters found in the gene clustering analysis above, sequences were chosen for inclusion in the tree based on their PCBBH score (PCBBH is the acronym for “pair of close bidirectional hits” and the PCBBH score is a measure of the phylogenetic distance between the pair, which correlates with significance). Because of how the score is calculated, certain COG0523 genes will have higher scores than other COG0523 genes (proportional to the predicted significance of the clustering). For example, if a COG0523 gene in a Firmicute genome (A) was recently acquired from the genome of a γ-proteobacterium (B), then those two COG0523 genes (A and B) will have a higher PCBBH score than two COG0523 genes from γ-proteobacteria (B and C). This is because 1) genes A and B have higher sequence similarity than B and C, and 2) the genomes to which A and B belong are phylogentically more distant than the genomes to which genes B and C belong. For each of the 14 gene clusters observed, the COG0523 genes with the highest PCBBH score were chosen.

Second, all COG0523 proteins whose genes were identified as being downstream of predicted Zur-binding sites (Novichkov *et al.*, 2010) were included.

Third, the COG0523 proteins from six eukaryotes were also used, including 10 COG0523 homologs from the alga *Chlamydomonas reinhardtii. C. reinhardtii* is a model organism for the study of trace metal homeostasis (Merchant and Bogorad, 1986a, Allen *et al.*, 2007, Merchant *et al.*, 2006) and the response of these COG0523 genes to zinc, copper, iron and manganese deficiency was available. Therefore, these data

represented an opportunity to test whether this phylogenomic analysis was in agreement with the results of that analysis.

RV0106 from *M. tuberculosis* CDC1551 was used as an outgroup in the phylogenetic reconstruction. This protein, while having similarity to COG0523, does not contain the canonical CXCC motif (CXSC). In addition, it is missing the canonical Walker A motif of the GTPase domain, suggesting that this COG0523-like protein does not have GTPase activity.

The phylogenomic analysis resulted in the identification of fifteen subgroups (summarized in Figure 3-2, detailed in Appendix B and Appendix C). To be designated as a subgroup, a group of proteins had to be monophyletic (forms a clade; composed of an ancestor and all of its descendents (Ashlock, 1971)) and the corresponding genes should belong to similar genomic neighborhoods and/or share conserved regulatory sites: coenzyme B_{12}-responsive riboswitches or Zur-binding sites.

Two exceptions were made. Subgroups 1 and 5 appeared to be paraphyletic. The clade composed of subgroup 1 also contained subgroup 2 and the clade that contained subgroup 5 also contained subgroup 4 (Appendix C). Subgroup 1 and 2 were separated because subgroup 1 proteins were encoded by genes linked to zinc homeostasis through gene clustering or the presence of putative Zur-binding sites, and subgroup 2 proteins appeared to be NHase activators based on literature reports and gene clustering. Subgroup 5 and 4 were separated due to respective gene clusters.

Five subgroups (1, 2, 5, 12 and 13) are analyzed in more detail below. The remaining 10 subgroups are detailed in Appendix B.

Zur-regulated COG0523 proteins – subgroups 1, 5 and 13

Sixty-eight COG0523 genes were found to be downstream of a potential Zur-binding site mainly in Firmicutes and γ-, β-, and α-proteobacteria. Two COG0523 genes were found downstream of a putative Zur site in Cyanobacteria, *Prochlorococcus marinus, Nostoc* sp. PCC 7120 and several *Cyanothece* spp. While most proteins encoded by Zur-regulated COG0523 members were found in subgroup 1 (75%), several paralogs were found in other subgroups. For instance, in *Pseudomonas entomophila, Pseudomonas fluorescens*, and *Pseudomonas putida* there were two COG0523 homologs per genome predicted to be downstream of a Zur-binding site. From the phylogenetic analysis one paralog was assigned to subgroup 1, while the other was assigned to subgroup 11 (Appendix C). Zur-regulated COG0523 paralogs were also found in subgroup 5, 8, 10 and 14 (Appendix B and Appendix C).

The nitrile hydratase activator subgroup – subgroup 2

Based on this analysis, less than 0.7% of the COG0523 family was represented by the NHase activators (Subgroup 2, Figure 3-2). A complete list of identified iron-type NHase activators from both Genbank and SEED databases is given in Appendix B. In the literature, these proteins are referred to as Nha3, P44K, or P47K, depending on the organism in which the protein was identified (Appendix B). Here, this subgroup of COG0523 is referred to as Nha3. Nha3 was found clustered exclusively with the genes encoding the two subunits of the iron-type NHase (Figure 3-3), which can be distinguished from the cobalt-type NHase by the strictly conserved metal binding motifs CSLCSCT for Fe(III) and CTLCSCY for Co(III) (Banerjee *et al.*, 2002). In addition, this COG0523 subgroup appeared to be closely related with the COG0523 proteins expected to be involved in zinc homeostasis (Figure 3-2, Appendix C).

The CobW subgroup – subgroup 12

Although "cobalamin biosynthesis protein" is the most highly propagated annotation for COG0523 members, this comparative genomic and phylogenetic analysis suggested that true CobW proteins (Subgroup 12, Figure 3-2) represent only 12.5% of the COG0523 family. In three α-proteobacteria genomes from the Rhodospirillaceae family, *cobW* genes belonged to the cobalamin biosynthesis gene clusters that were not preceded by putative B_{12}-regulated riboswitches. *cobW* orthologs in Cyanobacteria were neither clustered with B_{12} biosynthesis genes nor putatively regulated by a B_{12} riboswitch. However, these orthologs were included in subgroup 12 since they were highly similar to other CobW proteins and the corresponding genes co-occurred with the cobalamin biosynthesis genes of the aerobic pathway.

cobW was often found adjacent to the cobalt chelatase component, *cobN* (Figure 3-4) and all CobW proteins analyzed contained a His-stretch, which on average was composed of 7 histidines (the least being 4 histidines and the most being 15).

COG0523 proteins in Archaea

Although COG0523 was previously assumed to be missing from Archaea (Leipe *et al.*, 2002), the availability of recently sequenced genomes revealed that out of 44 archaeal genomes in the SEED database, eight genomes contained at least one COG0523 homolog. Eight homologs were found in *Methanosarcina acetivorans* C2A containing (Appendix A, Figure A-4).

Most Archaeal members belonged to subgroup 13, members of which co-localized with genes for corrinoid-dependent methyltransferase (Figure 3-5). In *Methanosarcina barkeri, M. acetivorans, Methanosarcina mazei*, and *Methanococcus maripaludis*, COG0523 genes clustered with genes involved in methanol:CoM methylation: *mtaA,*

mtaB, (both are zinc-dependent (Sauer and Thauer, 1997)), *mtaC* (corrinoid protein (Sauer *et al.*, 1997)) and *ramM* (iron-sulfur protein (Ferguson *et al.*, 2009)) (Figure 3-5). Clustering between COG0523 and methanogenesis genes was not limited to Archaea but also found in *Clostridium botulinum* (Figure 3-5), which was the reason for the high functional coupling score between COG0523 genes and methanol:CoM methylation. Another clostridium, *Desulfitobacterium hafniense* DCB-2, encoded a COG0523 gene that clustered with a MeTr homolog (methyltetrahydrofolate:corrinoid/ iron-sulfur protein methyltransferase) (Figure 3-5). In support of the functional coupling between COG0523 genes and *ramM, MM1072* (*M. mazei* COG0523 gene) was induced to the same extent as its neighboring *ramM* homolog, *MM1071*, during growth in high salt conditions (2.38- and 2.21-fold, respectively) (Pflüger *et al.*, 2007).

COG0523 proteins in Eukarya

COG0523 genes were found to be widespread in eukaryotic genomes. Most eukaryotic organisms contained one to four homologs (Appendix A, Figure A-5). Gene clustering is not very informative in eukaryotes but most eukaryotic COG0523 homologs including *Homo sapiens* belonged to subfamily 5 (Figure 3-6). The prokaryotic members of subfamily 5 were found to cluster on the genome with genes that encode WD40-repeat proteins (Figure 3-7). WD40-repeat proteins form a β-propeller structure thought to mediate protein-protein interactions (Smith *et al.*, 1999). Subfamily 5 COG0523 genes also were found to cluster the genes encoding the components of the high-affinity zinc transporter and creatinase-encoding genes (Figure 3-7). Several prokaryotic members of subfamily 5 were predicted to be downstream of a Zur-binding site (Appendix C).

Discussion

Characterized members of the G3E family are proposed to perform two roles in metallocenter assembly: 1) facilitating incorporation of the cofactor in an energy-dependent manner into the target protein's catalytic site (the insertase role), and 2) storage and delivery of a metal cofactor to a target metalloprotein (the metallochaperone role). G3E family proteins have been found to function as either metal-insertases or as a dual function metallochaperone/insertase.

Little is known regarding the function of COG0523; however, the other G3E subfamilies are relatively well characterized. As a way to infer the function of COG0523 proteins, a comparison with those subfamilies was performed. The assumption was made that general features such as metallochaperone and/or insertase activity is maintained among COG0523. As a way to distinguish these two roles, an analysis of His-stretches was performed, since for HypB the His-stretch seems to correlate with metallochaperone function.

To strengthen the argument that a His-stretch in a G3E protein eliminates the need to an additional metallochaperone, a comparative genomic analysis of UreG was performed. Although, the presence of His-rich regions was previously thought to be absent from bacterial orthologs of UreG and only present in plant orthologs (Witte *et al.*, 2001), one of the comparative genomic analyses presented here revealed that several bacterial orthologs do contain His-rich regions. As expected, if the His stretch signifies metallochaperone activity, a His-stretch in UreG was found in genomes that lacked the gene for the characterized metallochaperone component UreE (Table 3-1). This observation suggests that the His-stretch can compensate for the lack of metallochaperone in urease maturation. *B. japonicum* USDA 110 and *Frankia* sp. Ccl3

were found to be two exceptions to this trend. In these cases, the nickel-metallochaperone involved in urease maturation could be HypB, which was present in both of these organisms (Appendix A, Figure A-2).

Indeed, in *Helicobacter pylori,* HypB is required for activity of both [NiFe]-hydrogenase and urease (Olson *et al.*, 2001), and a physical interaction between UreG and HypB has been verified (Stingl *et al.*, 2008). These two observations suggest that HypB may be responsible for delivery of nickel to UreG. Although the traditional urease metallochaperone UreE is encoded in the *H. pylori* genome, it is missing the His-stretch found in other UreE proteins. The addition of a His-stretch to the *H. pylori* UreE was found to eliminate the need for HypB in the maturation of urease (Benoit and Maier, 2003). These results further support the conclusion that the His-stretch functions in metallochpaerone activity.

Another putative metal-binding motif, CXCC, was found in all COG0523 proteins analyzed. This motif is commonly found in metallothioneins, which are involved in metal homeostasis (Gutiérrez *et al.*, 2009). The CXCC motif was found to be essential for the activity of the NHase activator, a member of COG0523 proposed to be involved in the incorporation of iron into iron-type NHase (Lu *et al.*, 2003). In the crystal structure of YjiA, a COG0523 protein from *E. coli*, this motif was found in a three-stranded β-meander, which was directly attached to the switch I region of the protein (Khil *et al.*, 2004).Characteristically, GTPases are composed of two regions, switch I and switch II, that contain residues which bind to the γ-phosphate of GTP (Vetter and Wittinghofer, 2001). Upon GTP hydrolysis, the γ-phosphate is cleaved releasing the switch region residues that then relax into the GDP-bound form of the protein (Milburn *et al.*, 1990).

Therefore, depending on whether GTP or GDP is bound, the protein can have two structural conformations (Milburn *et al.*, 1990). Since the CXCC motif is attached to the switch I region, binding of GTP may affect its relative position in the tertiary structure of the protein (Khil *et al.*, 2004).

COG0523 proteins on average appeared to have an extra C-terminal domain that was not found in either HypB or UreG proteins. The size of COG0523 proteins was on average more similar to the size of MeaB proteins. Size could be an indication of the number of other accessory proteins required for activation of the target metalloprotein. The smallest G3E protein, UreG, is the GTPase component of a complex composed of UreD and UreF, where the three proteins act together in the activation of urease (Soriano and Hausinger, 1999). In contrast, activation of MCM appears to only require delivery of the cofactor by adenosyltransferase and the activity of MeaB (Padovani *et al.*, 2008).

The amino acid sequence of the extra C-terminal domain found in COG0523 proteins was highly variable (Figure 3-1). Indeed, COG0523 proteins fall under the category of "segmentally variable genes (SVGs)," as defined by Zheng *et al.* (Zheng *et al*., 2004). SVGs are genes that code for proteins that have highly variable regions interspersed with well-conserved regions. The authors observed that SVGs encode proteins that are involved in adaptation to environmental stresses and proposed that highly variable domains are an indication of protein-protein interaction specificity or specificity of small molecule binding.

Of the subgroups defined here, several of them (5 out of 15) were associated with zinc either through putative regulation by Zur or through co-localization on the genome

with genes known to be involved in the response to zinc. A survey of the literature reveals that members of COG0523 are often implicated in the virulence of pathogens whose hosts are known to induce zinc-limitation as a defense strategy.In *M. tuberculosis*, a COG0523-like gene, *RV0106*, (shown to be repressed by Zur (Maciag *et al.*, 2007)) was up-regulated during human macrophage infection (Cappelli *et al.*, 2006) (although RV0106 shows homology to COG0523, it was found to be missing the canonical GTPase motifs, and the second cysteine of the CXCC motif was not conserved). In the closely related *Mycobacterium avium* subsp. *paratuberculosis*, this gene was found on a pathogenicity island (Stratmann *et al.*, 2004) and the corresponding protein was the second strongest antigen consistently reactive with cattle sera infected with *M. avium* or *Mycobacterium bovis* (Bannantine *et al.*, 2008). COG0523 genes were also found in a pathogenicity island from *Enterococcus faecalis* (McBride *et al*., 2009). Loss of the COG0523 gene in *Brucella suis* rendered this bacterium incapable of intramacrophagic replication (Köhler *et al.*, 2002), while loss of the COG0523 gene in *Burkholderia pseudomallei* resulted in the inability to infect *Caenorhabditis elegans* (Gan *et al.*, 2002). An ortholog from *Francisella tularensis* was expressed exclusively in bacteria separated from infected murine spleen tissue (Twine *et al.*, 2006). This gene was down-regulated in the *Francisella novicida* Δ*pmrA* mutant (Mohapatra *et al.*, 2007). PmrA is a transcription factor found to be essential for survival/growth inside human and murine macrophage cell lines (Mohapatra *et al.*, 2007).

In plants, an opposing defense strategy may be employed, as repression of zinc uptake machinery is required for full virulence of the plant pathogens, *X. campestris* and

X. *oryzae* (Tang *et al.*, 2005, Huang *et al.*, 2008, Yang *et al.*, 2007). In contrast to animal pathogens and further supporting a role for COG0523 in zinc homeostasis, two COG0523 gene homologs of *A. tumefaciens* as well as the genes encoding the high-affinity zinc transporter, ZnuABC, were down-regulated in response to plant signals (Yuan *et al.*, 2008).

The phylogenomic analysis presented here further suggests that, in addition to the bacterial orthologs, the eukaryotic orthologs of COG0523 may be involved in zinc homeostasis. In support of this assertion, two COG0523 gene homologs from the alga *C. reinhardtii* were found to be induced under zinc-deficient conditions compared to zinc-replete conditions (Haas *et al.*, 2009). The corresponding protein for one of these genes was assigned to subgroup 1 and the other was assigned to subgroup 5. As discussed, both of these subgroups are enriched with prokaryotic members whose genes either physically cluster on the genome with genes involved in zinc homeostasis or are putatively regulated by Zur.

Conclusions

These analyses provide a foundation for understanding the function of known and yet to be identified COG0523 proteins. The sequence analysis suggests that COG0523 proteins have characteristics similar to and distinct from the non-COG0523 members of the G3E family. On the one hand, sequence motifs suggest that like characterized G3E family members COG0523 proteins are GTPases and most have metal-binding motifs. The distribution of His-stretch motifs suggests that COG0523 proteins could function as metallochaperones and/or insertases. Unlike the other members of G3E, COG0523 proteins can be separated into numerous subgroups that are supported by computational approaches and literature reports. The diversity of genomic co-

localization suggests that COG0523 is more diverse than the other subfamilies of G3E. Both the metal specificity and the protein target(s) might vary from one subgroup to another.

While known roles in cobalamin biosynthesis and response to zinc limitation predominate, this analysis implies members of COG0523 are not limited to those roles. Based on genome context (co-localization and/or presence of a coenzyme B_{12} riboswitch) and protein similarity analyses, only 12.5% of sequenced COG0523 from the SEED database are true CobW proteins and assigned to the cobalamin biosynthesis pathway. Only ~30% of COG0523 members analyzed are linked to zinc homeostasis either through putative Zur sites (~8%) or co-localization with genes involved in the response to zinc starvation (~20%). In addition the third known role, NHase activator, only applies to less than 1% of sequenced COG0523 genes. Over half of COG0523 may perform a role in the activity of unknown proteins.

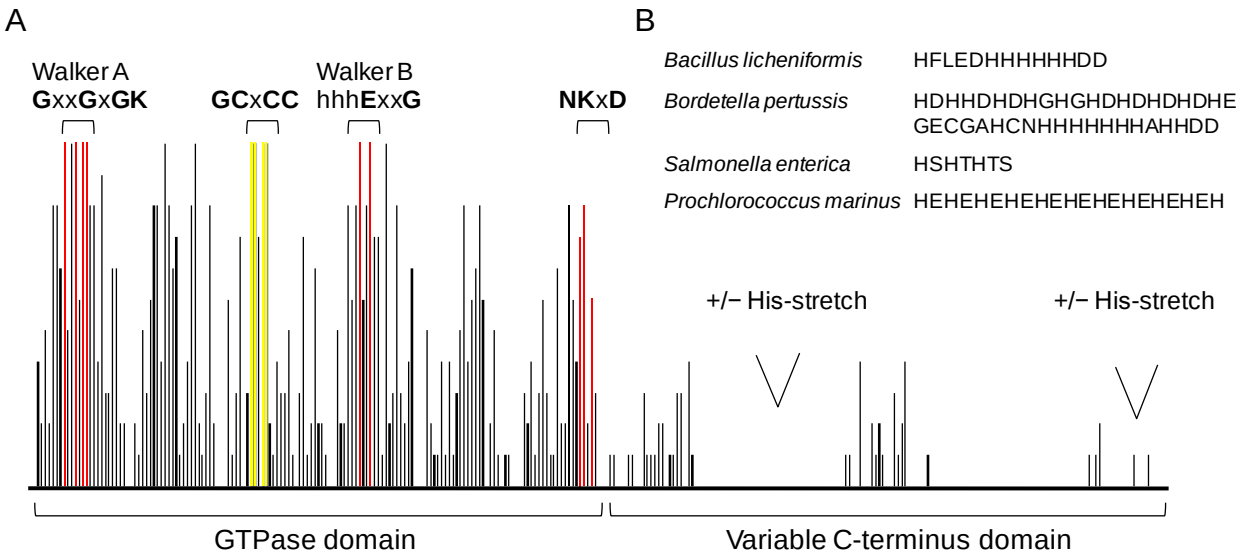


Figure 3-1. COG0523 amino acid conservation plot. A) Plot of amino acid conservation. The conserved GTPase motifs are highlighted in red. The conserved GCXCC motif is highlighted in yellow. The most common positions of His-stretches are shown. The taller the bar, the higher the conservation of amino acid residues is at that position. For instance, the Walker A motif (GXXGXGK) and the cysteine motif (CXCC) are absolutely conserved. B) Typical histidine-rich sequences found in COG0523 homologs from specified genomes. Protein identifiers used in this analysis are available in Appendix D, Table D-1.

Table 3-1. Co-occurrence profile of ureE and UreG His-stretch. Presence (+) or absence (-) of ureE, a His-stretch in UreG, and the presence of slyD in representative genomes from the SEED database.

Organism	ureE	UreG histidine stretch	*slyD*
Anaeromyxobacter sp. Fw109-5	−	HDHSLHSGHDHGLGPGSFHDRGAPH	+
Arabidopsis thaliana	−	HDHHHHHHDHEHDH	−
B. japonicum	−	−	−
Cytophaga hutchinsonii	−	HLDHFDSPGHFHHRELIH	+
Frankia sp. Ccl3	−	−	−
Gibberella zeae	−	HSHDGQSHSHDGFNAQEHGHSH	+
Herpetosiphon aurantiacus	−	HVHDDHHHHHHH (C-terminus)	−
Magnaporthe grisea	−	HSHSHDGSAPHSHSHDGSTFNAQEHGHSH	+
M. bovis	−	HSHPHSH	−
Mycobacterium marinum	−	HSHDHTHDHH	−
M. tuberculosis	−	HSHPHSH	−
Mycobacterium vanbaaleni	−	HFLDGQPHGH	−
Neurospora crassa	−	HTHSHDHGDGGHHHHPHSHSHDFNSQSGFNAQEHGHSH	+
Nocardia farcinica	−	HDHAH	−
Schizosaccharomyces pombe	−	HKGGSDDSTHHHTHDYDHHNHDHHGHDHHSHDSSSNSSSEAARLQFIQEHGHSH	−
Sorangium cellulosum	−	HDPGEHGHGRHDHDHDHDHVHDHDHDHDHVHGGGHRHAHEHEHAHEHAHGHEHGHAHAHAHAHAHEHAHGHTHEHWAH	+
Streptomyces avermitilis	−	HLDHAHTH	−
S. coelicolor	−	HLDHHH	−
Verminephrobacter eiseniae	−	HHLHH	+
Bacillus cereus	+	−	−
Corynebacterium glutamicum	+	−	−
Haloarcula marismortui	+	−	−
Rhizobium leguminosarum	+	−	−
Ureaplasma urealyticum	+	−	−
H. pylori	+	−	+

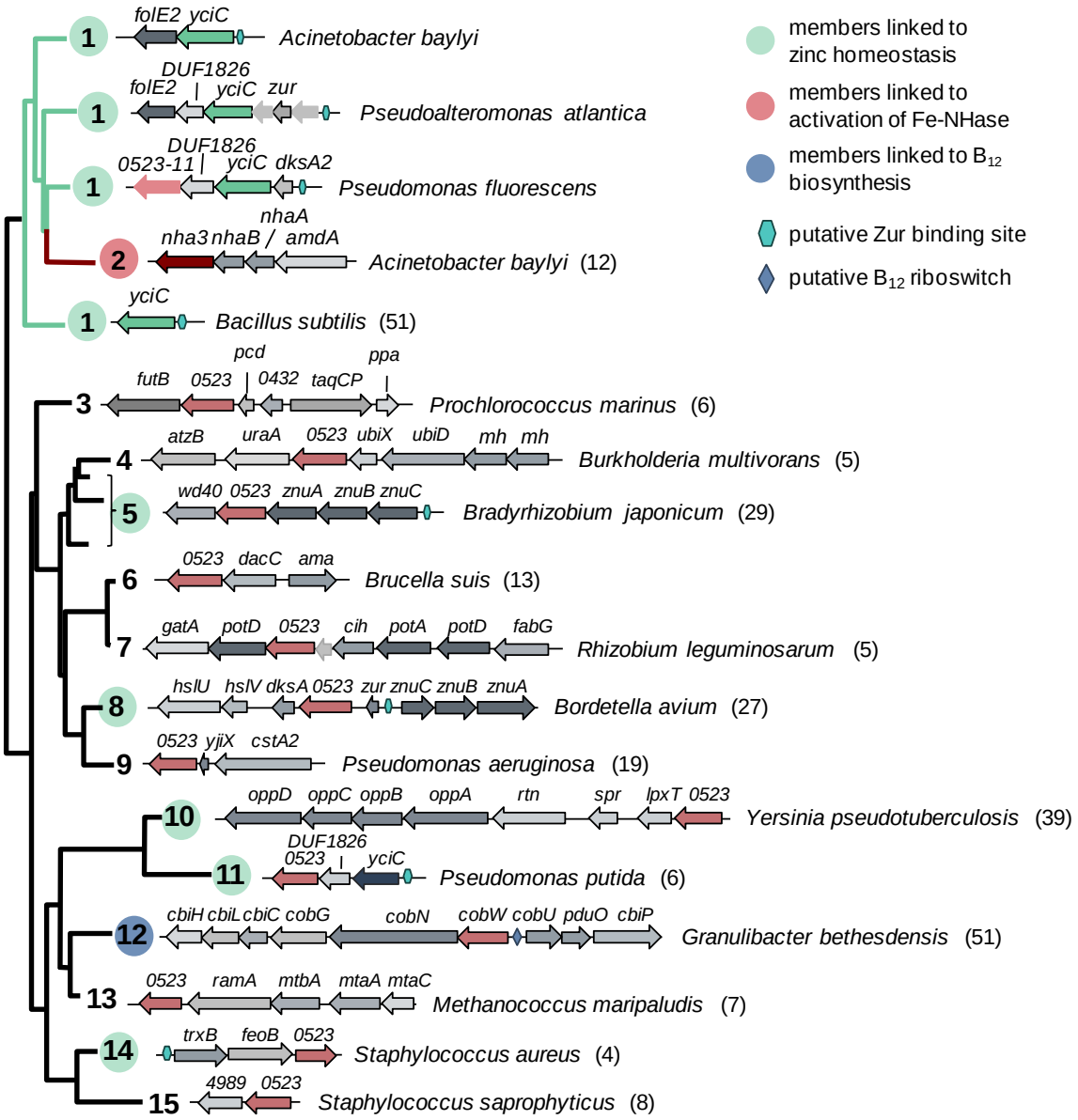


Figure 3-2. Phylogenomic analysis of COG0523. Collapsed COG0523 distance tree; each subgroup (3,4 and 6-15) was collapsed to its common node. Since subgroup 5 and 1are paraphyletic, branches were collapsed to multiple nodes. Representative gene neighborhoods for each subfamily are shown and corresponding genome. Numbers in parentheses refer to the number of species where the gene cluster occurs. For subfamily 1, this number refers to the number of species that contain a putative Zur-regulated *yciC*. Abbreviations: '*0523*', COG0523 homologs; '*yciC*', subfamily 1; '*cobW*', subfamily 12; '*nha3*', subfamily 2. For subfamily 1, '*0523-11*' refers to a subfamily 11 COG0523 homolog that is found in the same gene cluster as *yciC*. All other gene abbreviations can be found in Appendix D, Table D-2. The full COG0523 tree is available in Appendix C.

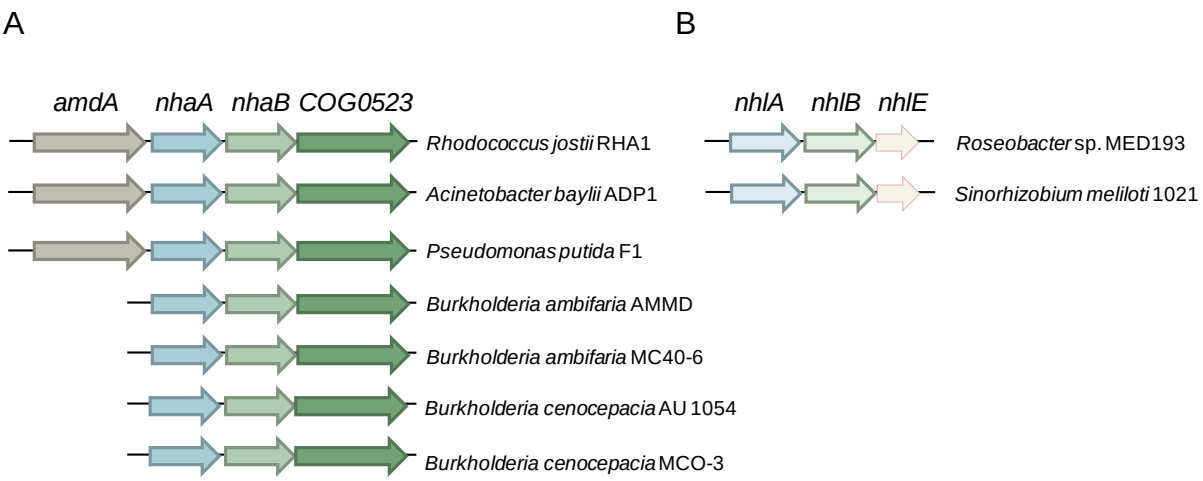


Figure 3-3. Genome context of predicted nitrile hydratase activators. A) Gene clusters of predicted Fe-type nitrile hydratase subunits and corresponding activator proteins. A COG0523 family gene, encoding the activator, is found upstream of the α- and β- subunits of the Fe-type NHase genes. NhaA contains the Fe(III)-binding motif, CSLCSCT. B) Gene cluster of representative, predicted Co-type nitrile hydratase activator subunits. *nhlE*, which shares no sequence homology with COG0523, is found upstream of the Co-type nitrile hydratase subunit genes. NhlA contains the Co(III)-binding motif, CTLCSCY.

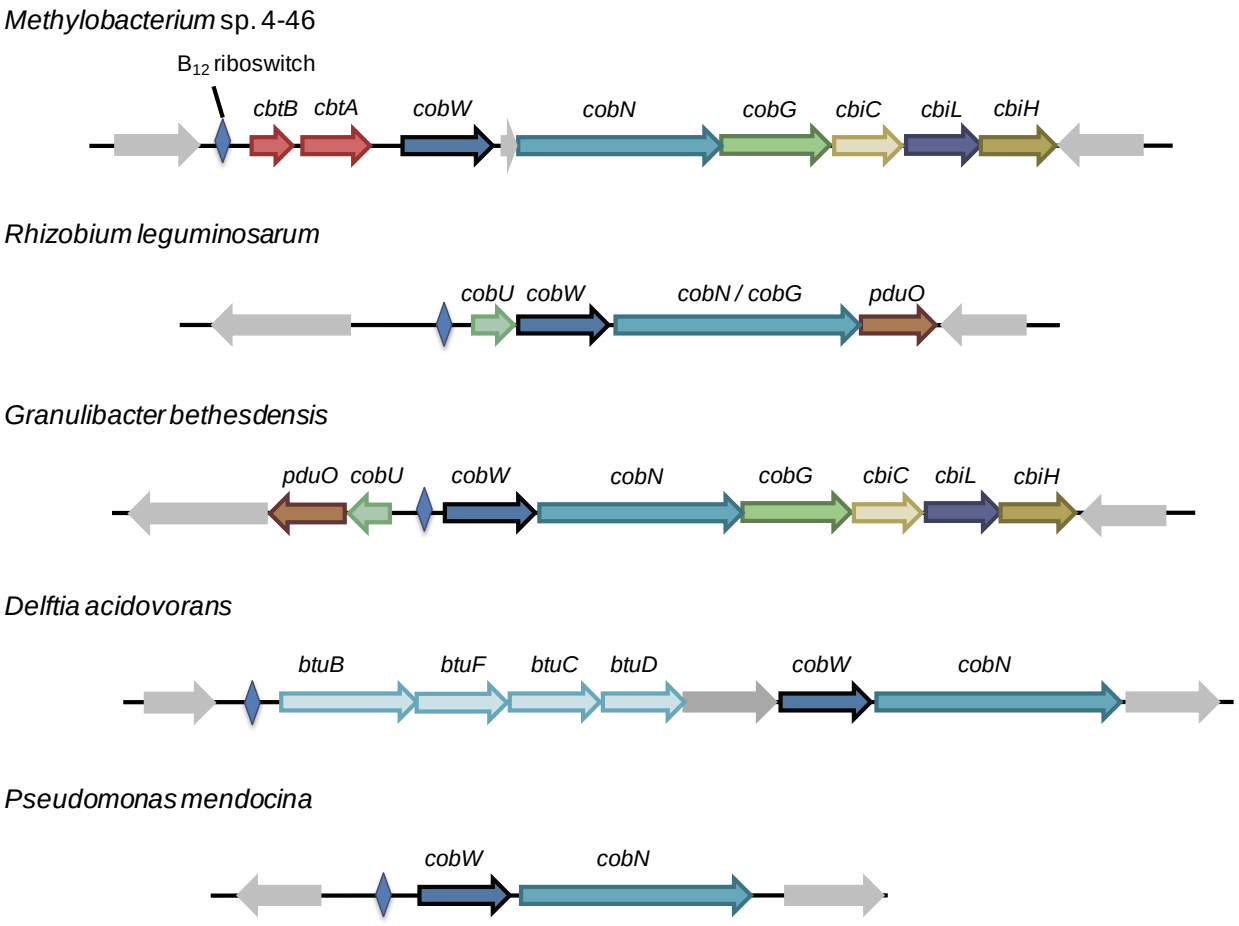


Figure 3-4. Genome context of subgroup 12 members. Abbreviations: *cobW*, subfamily 12 COG0523 paralog; *cbtB*, cobalt transporter subunit; *cbtA*, cobalt transporter subunit; *cobN*, cobaltochelatase subunit; *cobG*, precorrin-3B synthase; *cbiC*, cobalt-precorrin-8x methylmutase; *cbiL*, cobalt-precorrin-2 C20-methyltransferase; *cbiL*, cobalt-precorrin-3b C17-methyltransferase; *cbiH*, cobalt-precorrin-6x reductase; *cobU*, adenosylcobinamide-phosphate guanylyltransferase; *pduO*, cob(I)alamin adenosyltransferase; *btuB*, outer membrane vitamin B_{12} receptor; *btuF*, vitamin B_{12} ABC transporter, B_{12}-binding component; *btuC*, bitamin B_{12} ABC transporter, permease component; *btuD*, vitamin B_{12} ABC transporter, ATPase component. Grey arrows represent genes annotated as “hypothetical.”

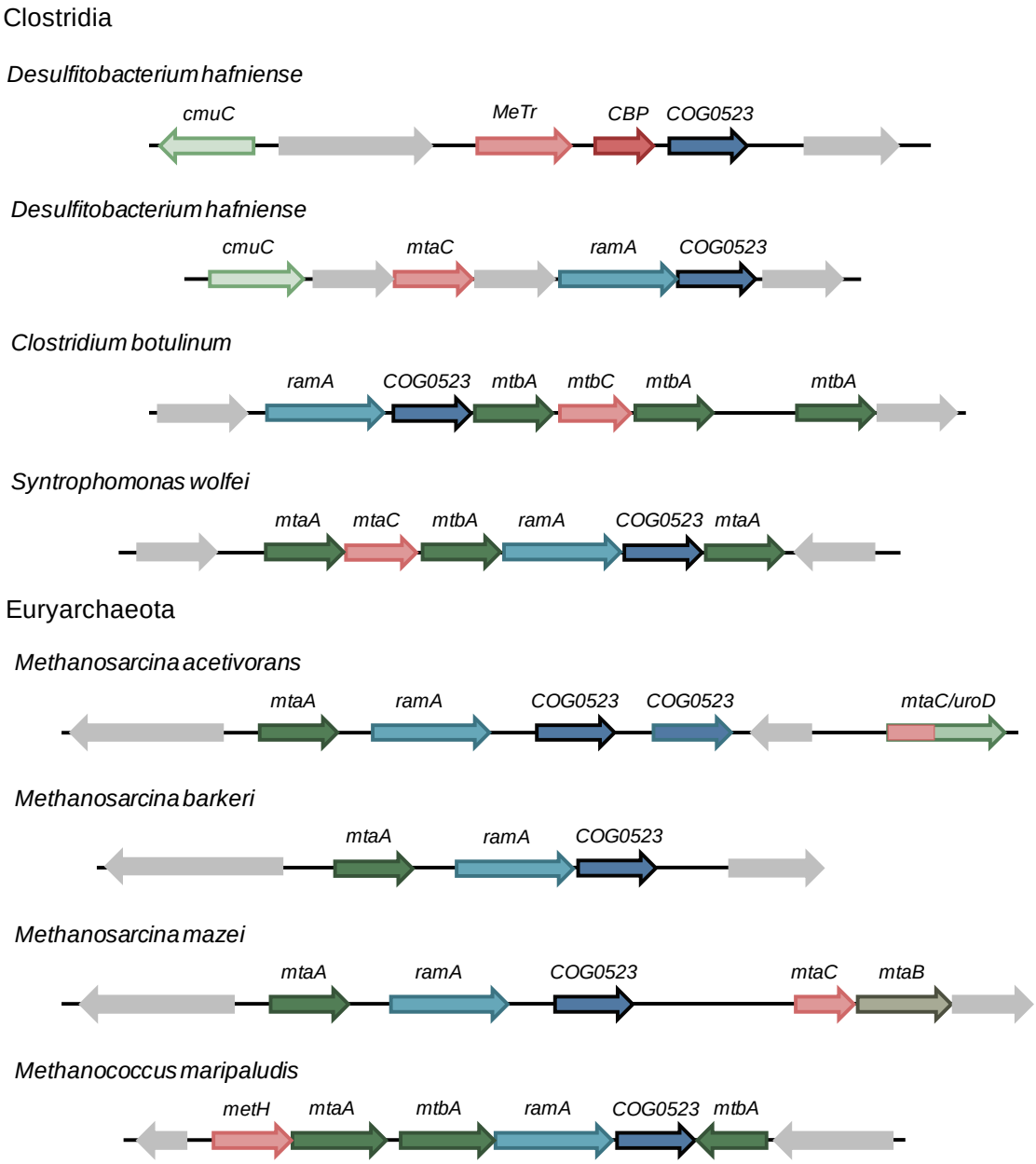


Figure 3-5. Genome context of subgroup 13 members. Subgroup 13 genes co-localize with genes encoding methyltransferases, corrinoid-binding proteins, and a protein responsible for corrinoid recycling. Abbreviations: *cmuC*, corrinoid methyltransferase-like; *metH*, methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase; CBP, B_{12} binding domain of corrinoid proteins; *mtaC/mtbC*, corrinoid-binding protein; *ramA*, iron-sulfur protein that mediates the ATP-dependent reductive activation of Co(II) corrinoid to the Co(I) state; *mtbA*, methylcobalamin:coenzyme M methyltransferase, methylaminie-specific; *mtaA*, methylcobalamin:coenzyme M methyltransferase, methanol-specific; *mtaB*, methanol:corrinoid methyltransferase.

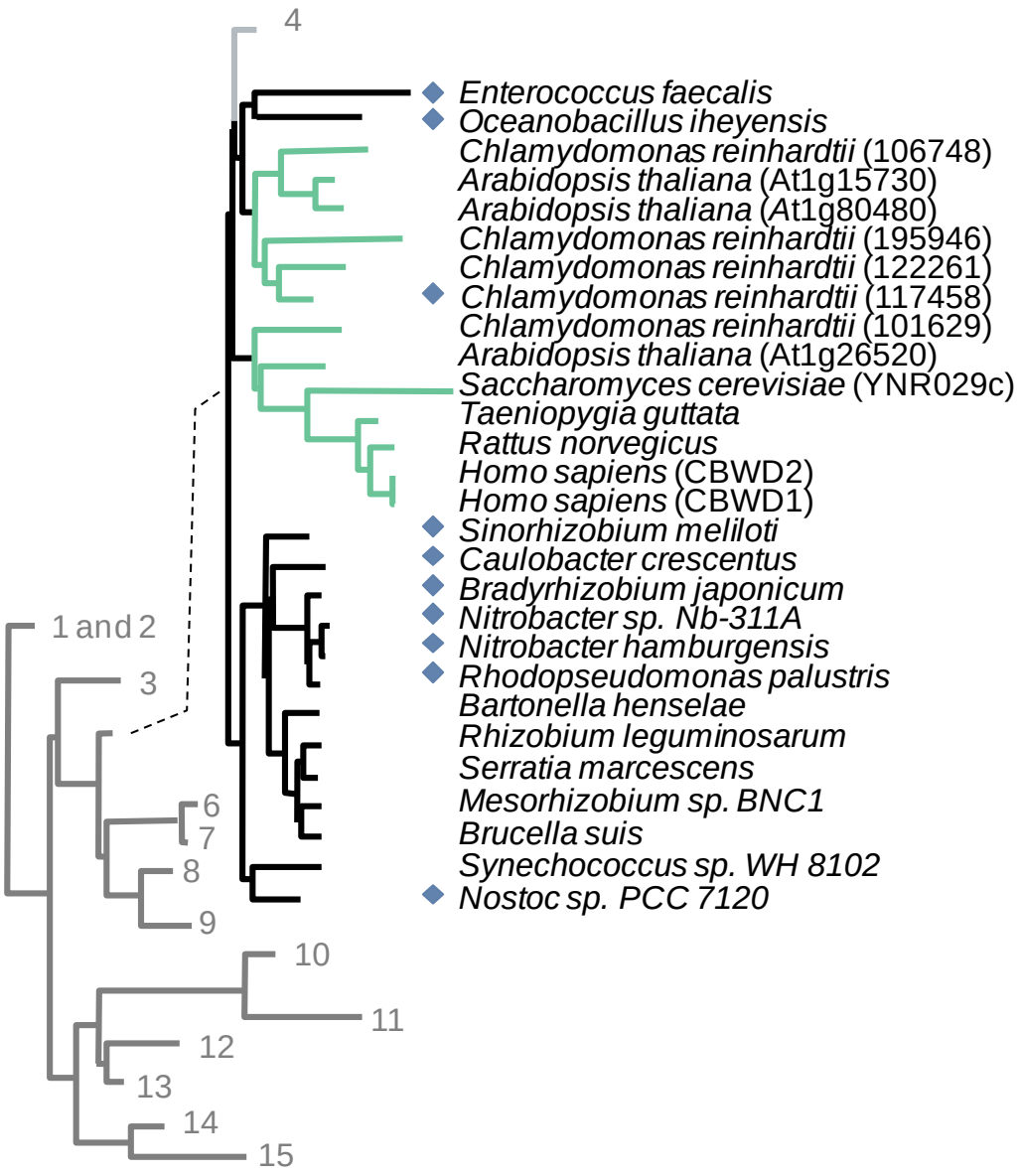


Figure 3-6. Phylogeny of eukaryotic COG0523 members. Lightly shaded tree represents the collapsed COG0523 distance tree. Branches are labeled with corresponding subgroup number. Subgroup 5 tree: branches representing eukaryotic homologs are colored green. A blue diamond next to an organisms name indicates a putative Zur-binding site is found upstream of the corresponding gene. For the *C. reinhardtii* ortholog, the blue diamond indicates confirmed induction of corresponding gene to zinc deficiency as reported in (Haas *et al.*, 2009).

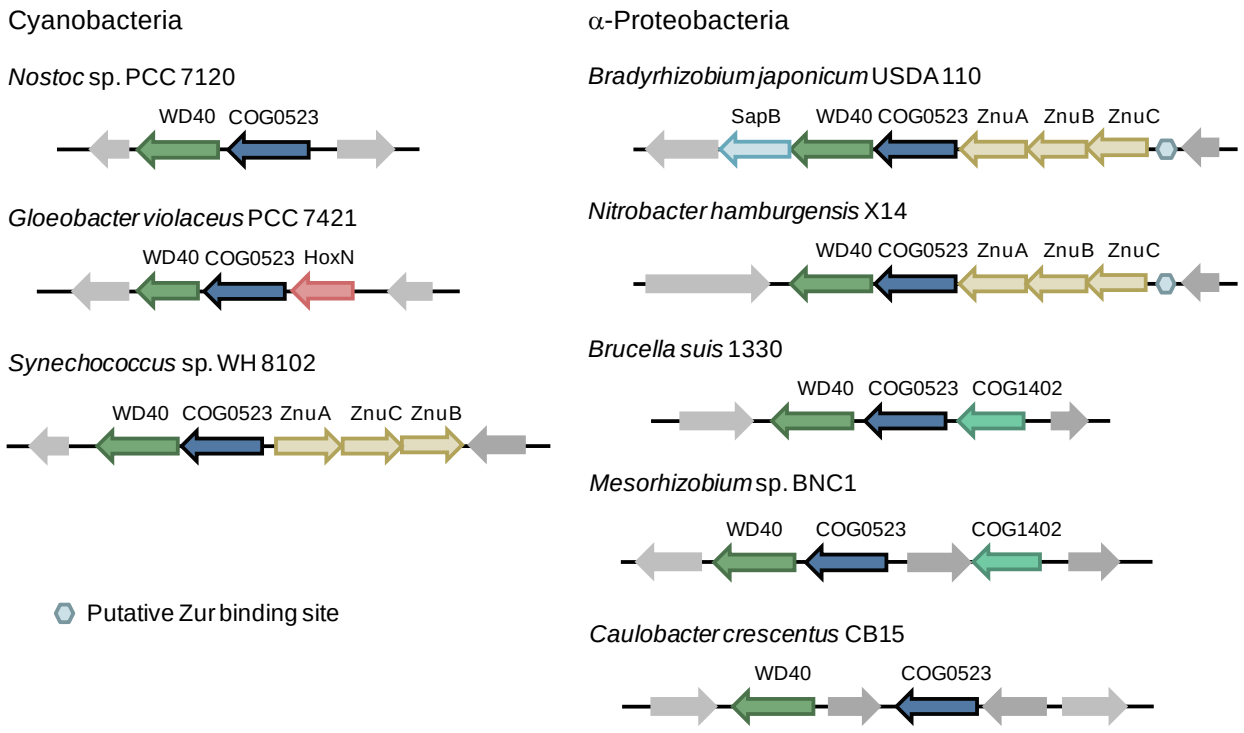


Figure 3-7. Genome context of subgroup 5 members. Subgroup 5 genes co-localize on the genome with genes encoding metal transporters, HoxN/HupN/NixA family cobalt transporter (*hoxN*) and high-affinity zinc transporter (*znuABC*), WD40 repeat-containing proteins (*WD40*), uncharacterized membrane family proteins (s*apB*), and creatine amidohydrolase-like proteins (*COG1402*).

CHAPTER 4
INVESTIGATION INTO THE FUNCTION OF THE UNCHARACTERIZED GENE *YEIR*

Background

E. coli is the prototypical model organism and is used both to gain insight into gene function and cellular processes and simply as another molecular biology tool. Due to the length of time this organism has been studied and ease of genome manipulation, the genome of *E. coli* is by far the best annotated of all sequenced genomes. Yet approximately only 59% of the gene annotations in *E. coli* are supported by experimental data and only the function for half of these are actually experimentally validated (Keseler *et al.*, 2009, 2010). That leaves 71% of the genes in *E. coli* with annotations based solely on sequence similarity to genes in other organisms, which may or may not themselves be experimentally validated. The pitfalls associated with annotating based on sequence similarity alone are well documented and have been thoroughly reviewed (Schnoes *et al.*, 2009, Devos and Valencia, 2001, Brenner, 1999). To overcome the inherent error in annotations by sequence similarity, several bioinformatic techniques have been developed for producing or refining gene annotations (Poptsova and Gogarten, 2010, Overbeek *et al.*, 2005). Sometimes referred to as guilt-by-association, these techniques have lead to a plethora of gene function discoveries (for some recent examples, see (Yang *et al.*, 2006a, El Yacoubi *et al.*, 2006, Rodionov *et al.*, 2007, Rodionov *et al.*, 2009, Phillips *et al.*, 2010).

The COG0523 family phylogenomic analysis described in Chapter 3 was used as a basis to predict functions for the two COG0523 proteins in *E. coli*: YjiA and YeiR. In the EcoCyc database (www.ecocyc.org), *yjiA* is annotated as a “P-loop guanosine triphosphatase” and *yeiR* is annotated as a “predicted enzyme” (Keseler *et al.*, 2009).

From the phylogenomic analysis presented in Chapter 3, YjiA was assigned to subgroup 9. Subgroup 9 proteins are encoded by genes that cluster physically on the genome with two genes, encoding a paralog of a carbon-starvation protein, CstA, and a small uncharacterized protein, YjiX. The function of these two proteins is currently unknown. In the case of YjiA, guilt-by-association methods do not provide an obvious functional prediction, as the genes that *yjiA* associates with are also of unknown function.

YeiR was assigned to subgroup 10 of COG0523, which has an enrichment of proteins whose genes are downstream from putative DNA-binding sites for Zur. Zur is a well-characterized transcription factor that represses gene expression in the presence of zinc (Gaballa and Helmann, 1998, Patzer and Hantke, 1998). As zinc concentrations within the cell drop, Zur repression is lifted and genes involved in adapting to zinc-deplete conditions are expressed such as high-affinity zinc transporters and back-up enzymes (Patzer and Hantke, 1998, Sankaran *et al.*, 2009, Gabriel and Helmann, 2009). Work from the Helmann lab has shown that deletion of the Zur-regulated COG0523 gene *yciC* in combination with a *ycdH* deletion in *B. subtilis* leads to a cell growth defect in the presence of the metal chelator, EDTA, and more recently zinc-deficient medium (Gaballa and Helmann, 1998, Gabriel and Helmann, 2009). The gene *ycdH* encodes a homolog to the high-affinity zinc transporter that has been characterized in several bacteria (Patzer and Hantke, 1998, Yatsunyk *et al.*, 2008, Lewis *et al.*, 1999, Davis *et al.*, 2009). Phylogenetically, YeiR is a member of a distinct COG0523 clade separate from the YciC clade (Figure 4-1). Although this observation resulted in assignment of YeiR to subgroup 10 and YciC to subgroup 1, YeiR is closely

related to other putatively zinc-regulated COG0523 proteins (Figure 4-1). Therefore, YeiR could also be involved in survival when the cell is faced with a zinc-depleted environment. Application of this same logic, suggests that YjiA is not predicted to be involved in zinc homeostasis.

Based on the phylogenomic study presented in Chapter 3 and literature reports suggesting a role for a subset of COG0523 proteins in the response to zinc-depletion, a study on the involvement of YeiR in metal-related process particularly growth during zinc-depletion and cadmium toxicity was performed. Since the other G3E family proteins, HypB, UreG and MeaB, are chaperones involved in maturation of a target metalloenzyme, YeiR may also be a metal chaperone. A mutagenesis study on conserved, putative metal-binding motifs found in YeiR and a preliminary experiment with the purpose of finding a target for YeiR are also presented.

Results

EDTA Sensitivity

The effect of a deletion of *yeiR* on optimal growth during zinc-depletion was tested. To obtain *E. coli* MG1655 Δ*yeiR,* the *yeiR*::kan allele from *E. coli* BW25113 Δ*yeiR*::kan (JW2161) was transduced by P1 phage into *E. coli* MG1655 by the method of (Miller, 1972). Kanamycin-resistant recipient colonies were selected and verified by locus-specific PCR. The Δ*yeiR*::kan strain was then transformed with pCP20, which encodes FLP recombinase (native to *S. cerevisiae*) and catalyzes the site-specific recombination reaction between the two FRT-sites flanking the resistance cassette, thus removing the resistance marker from the chromosome (Cherepanov and Wackernagel, 1995). Loss of

the resistance gene was confirmed by plating on appropriate antibiotics and by locus-specific PCR.

The parent (WT) and mutant strains were grown in a low-phosphate minimal medium commonly used for the characterization of metal-related physiology and was previously used to study low-affinity zinc transporters in *E. coli* (Grass *et al.*, 2002). In order to create zinc-depletion, EDTA was added to the medium at various concentrations. EDTA is a hexadentate ligand and one molecule of EDTA has the potential to bind one metal ion in a ring structure (Wheelwright, 1953). Once bound, the EDTA-metal complex is soluble (Plumb *et al.*, 1950) and equilibrium between bound and unbound metal in the medium is formed. Due to the difficulty in specifically limiting growth conditions for zinc, EDTA is a popular choice for investigating physiology in the response to zinc-deficiency (Gaballa and Helmann, 1998, Grass *et al.*, 2002, Patzer and Hantke, 1998, Zhao and Eide, 1996). The *E. coli* cell in particular has been shown to be well equipped for acquiring zinc from the environment. Measurements of extracellular and intracellular zinc concentrations have revealed that the cell was able to concentrate intracellular zinc to a level ~2,000-fold higher than the growth medium (Outten and O'Halloran, 2001). This trait makes depleting growth conditions of zinc difficult without the aid of a metal-chelator such as EDTA.

Growth of the WT and Δ*yeiR* strains was compared in the presence of EDTA at a concentration from 0 – 2 mM. Growth was monitored in Bioscreen C machine (Growth Curves USA) that measures the optical density of up to 200 cultures in plate format practically simultaneously (200 µl culture were well). This system was used previously to study the effect of *B. subtilis* and *Proteus mirabilis* gene deletions on growth in the

presence of zinc-depletion (Sankaran *et al.*, 2009, Gabriel and Helmann, 2009, Nielubowicz *et al.*, 2010).

In the absence of EDTA, both the WT and Δ*yeiR* strains had the same growth rate and reached the same final density (Figure 4-2A). The WT strain did not exhibit a noticeable growth defect until EDTA was present at 1.2 mM (Figure 4-3A). Growth was progressively more perturbed with higher concentrations of EDTA until growth was abolished in the presence of 2 mM EDTA. In contrast, the Δ*yeiR* mutant exhibited an observable sensitivity to EDTA at and above 1.1mM EDTA (Figure 4-3B). In the presence of concentrations at and above 1.3 mM EDTA, growth was inhibited and the cells did not recover (Figure 4-3B). The *yeiR* deletion could be complemented by uninduced expression of *yeiR* from the P_{BAD} of pBAD24 (Figure 4-2B). The EDTA-sensitive growth defect of the Δ*yeiR* mutant was not dependent on growth in the Bioscreen C machine and was reproducible with manual growth curves (Figure 4-4).

Typically, deletion of the genes encoding the high-affinity zinc transporter ZnuABC, increases growth defects caused by zinc-depletion (Grass *et al.*, 2002, Gabriel and Helmann, 2009). Therefore, to further evaluate whether the observed growth phenotype in the presence of EDTA could be due specifically to chelation of zinc, the *znuABC*::cam allele from *E. coli* GR352 (Grass *et al.*, 2002) was transferred to MG1655 and the Δ*yeiR* strain by P1 transduction. Chloramphenicol-resistant recipient colonies were selected and verified by locus-specific PCR.

The Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants were grown in LP medium with or without EDTA. As with the WT and Δ*yeiR* strains, the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants had the same growth curves in the absence of EDTA

(Figure 4-5A). The Δ*znuABC*::cam mutant did not display a noticeable sensitivity to EDTA until a concentration of 30 µM, whereas the additional deletion of *yeiR* shifted the sensitivity of the strain to 10µM EDTA (Figure 4-6). At 75 µM EDTA, the Δ*znuABC*::cam mutant was found to be as sensitive to EDTA as the Δ*yeiR* Δ*znuABC*::cam mutant (Figure 4-7A).

The difference in final density between the two strains at 20 µM EDTA was comparable to the difference in final density between the WT and Δ*yeiR* strains at 1.2 mM EDTA (Figure 4-7B). The *yeiR* deletion could be complemented by uninduced expression of *yeiR* from the P_{BAD} of pBAD24 (Figure 4-5B).

Rescue of EDTA Growth Defect by Zinc

As EDTA is a non-specific metal chelator, it was not known which metal ion's depletion was causing the observed phenotype. Therefore, various metal salts were added one at a time to the growth medium supplemented with 1.4 mM EDTA to determine which metal could suppress the growth defect observed with the Δ*yeiR* mutant. The EDTA-dependent growth defect of the Δ*yeiR* mutant was suppressed only by the addition of 25 µM Zn(II), whereas the same concentration of Ni(II), Cu(II), Co(II), Mn(II) or Fe(III) did not rescue cell growth (Figure 4-8).

The metal rescue profile was different for the Δ*znuABC*::cam Δ*yeiR* mutant. With the high-affinity zinc transporter present, a molar ratio of 56:1 (EDTA:Zn) was sufficient to see suppression of the phenotype due to deletion of *yeiR* (Figure 4-9A). However, with the genes encoding the high-affinity zinc transporter deleted, a molar ratio of 2:1 (EDTA:Zn) was required to see suppression of the phenotype (Figure 4-9B). In addition, 10 µM cobalt appeared to suppress the growth defect of the Δ*znuABC*::cam Δ*yeiR*

mutant in the presence of 20 µM EDTA to the same extent as 10 µM zinc (Figure 4-9B). Nickel at 10 µM was able to restore growth to 65% of the strain's growth in LP medium without EDTA (Figure 4-9B).

Cadmium Sensitivity

It has been proposed that intracellular zinc-depletion can be mimicked by the presence of Cd(II), as Cd(II) is thought to displace zinc in protein metal-sites since cadmium has a high-affinity for those sites (Graham *et al.*, 2009, Zhang *et al.*, 2003a). Therefore, to further probe the zinc-suppressible EDTA-dependent growth defect, growth of WT pBAD24, Δ*yeiR* pBAD24 and Δ*yeiR* pCH011 (pBAD24-*yeiR*) strains were compared in the presence of Cd(II) at a concentration from 20 – 50 µM (Figure 4-10). The Δ*yeiR* mutant was found to exhibit a clear difference in growth as compared to the WT strain in the presence of 30 µM cadmium (Figure 4-10B). The growth curve of the Δ*yeiR* strain in the presence of 30 µM cadmium was similar to the growth curve of the WT strain in 50 µM cadmium (Figure 4-10E). The *yeiR* deletion could be complemented by uninduced expression of *yeiR* from the P_{BAD} of pBAD24 (Figure 4-10).

The effect of a combination of CdII) and Zn(II), Co(II), Cu(II), Mn(II), Ni(II) or Fe(III) on the growth of the Δ*yeiR* and WT strains was also tested. The presence of zinc was able to partially suppress the cadmium-sensitive growth defect of the Δ*yeiR* mutant (Figure 4-11A) and manganese was able to suppress to a lesser extent (Figure 4-11B). The presence of zinc did not significantly suppress the cadmium-toxicity observed with the WT strain (Figure 4-12A). In contrast, growth of the Δ*yeiR* mutant in the presence of 40 µM zinc plus 40 µM cadmium was comparable to growth in 30 µM cadmium (Figure 4-13B). The presence of cobalt, nickel or iron did not lead to suppression or further

exacerbation of the cadmium-growth defect (Figure 4-13A), while the addition of copper further exacerbated growth of the WT and Δ*yeiR* strains (Figure 4-13B).

As was observed with EDTA, the cadmium–sensitive phenotype was exacerbated if the *znuABC* genes were deleted. The Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants were grown in LP medium in the presence of 2.5 – 20 µM Cd(II) (Figure 4-14). The difference in growth between the WT and Δ*yeiR* strains in the presence of 20 µM cadmium was similar to the difference in growth between the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutant strains in the presence of 2.5 µM cadmium (Figure 4-14E). As with the EDTA growth experiments, the difference in growth between these strains diminished with higher concentrations of cadmium; the Δ*znuABC*::cam mutant was as sensitive to the presence of 20 µM cadmium as the Δ*znuABC*::cam Δ*yeiR* mutant (Figure 4-14D). The *yeiR* deletion could be complemented by uninduced expression of *yeiR* from the P_{BAD} of pBAD24 (Figure 4-14).

Effect of Amino Acid Substitutions on the Activity of YeiR, *in Vivo*

The robust EDTA-sensitivity phenotype allowed probing of the importance of the conserved motifs found in COG0523 proteins. Like all COG0523 protein sequences, YeiR contains a CXCC motif where X is a non-polar amino acid in the N-terminal GTPase domain (Figure 4-15). For YeiR, this motif is at positions 63 – 65 in the protein sequence and X is a methionine. Each of the three cysteines and the methionine in this motif was mutated to alanine using the PCR-overlap strategy (Sambrook and Russell, 2001). The mutated genes were inserted between the *Nco*I and *Xba*I sites of pBAD24 and the resulting plasmids were transformed into the Δ*yeiR* mutant. The resulting strains were grown in LP medium with or without 1.4 mM EDTA (Figure 4-16). All strains

had identical growth curves when grown in LP without EDTA (Figure 4-16A). However, when grown in LP medium with EDTA, mutation of any one of the three cysteines or methionine to alanine abolished the ability of the corresponding gene *in trans* to complement the *yeiR* deletion (Figure 4-16B).

In addition to the cysteine rich motif, a short His-stretch was found in the C-terminus of YeiR (Figure 4-15). This region of the COG0523 protein sequences commonly contains His-stretches (Chapter 3). To test if these histidines play a substantial role in the activity of YeiR, each residue was mutated to alanine as described above. The mutated genes were inserted between the *Nco*I and *Xba*I sites of pBAD24 and the resulting plasmids were transformed into the Δ*yeiR* mutant. The resulting strains were grown in LP medium with or without 1.4 mM EDTA (Figure 4-17). Mutation of either H209 or H211 to A had no observable effect on the ability of the corresponding gene *in trans* to complement the *yeiR* deletion (Figure 4-17B). Mutation of H207 to A led to presumably a less active YeiR, because the Δ*yeiR* pCH140 (encodes YeiR H207A) mutant had a decreased growth rate in the presence of 1.4 mM EDTA compared to the Δ*yeiR* pCH011 (encodes WT YeiR) mutant (Figure 4-17B). Deletion of all three histidines to alanine severely perturbed the ability of the corresponding gene to complement the *yeiR* deletion (Figure 4-17B).

ICP-MS Analysis

Based on the prediction that COG0523 family proteins are chaperones involved in metallocenter biosynthesis (Chapter 3), a search for a potential target for YeiR was performed. If YeiR is involved in metal insertion, then deletion of *yeiR* could affect the abundance of the metal-bound form of the target. Analysis of whole-cell metal content

by ICP-MS did not yield a difference between the WT and mutant strains (data not shown). Therefore a protein fractionation-coupled ICP-MS technique was employed to analyze single metalloproteins (Tottey *et al.*, 2008). Detection of metal-bound proteins from the cytoplasmic fraction of the WT and Δ*yeiR* strains was performed using two-dimensional protein purification and detection of metal by ICP-MS. Purification was performed under native conditions to ensure metalloproteins did not lose their native metal cofactor. Anaerobic conditions were used (except for size-exclusion chromatography) as the presence of oxygen can also alter metal-protein speciation.

The cytoplasmic extract was loaded onto an anion exchange column and protein was eluted in steps of 100 mM, 200 mM, 300 mM, 400 mM, 500 mM and 1 M NaCl. Each fraction was then further fractionated by size-exclusion chromatography and 35 fractions were collected which were then analyzed by ICP-MS to determine the presence of copper, manganese, cobalt, nickel and zinc. The metal profiles for copper, manganese and coablt were not significantly different between the two strains (Figure 4-18). However, the nickel pool extracted from the Δ*yeiR* strain was noticeably reduced (Figure 4-18). Also a putative Ni/Zn- bound protein was present in the extract from the mutant but not from the parent (Figure 4-19). However, these peaks could represent two separate proteins (one nickel-bound and one zinc-bound) that elute in the same fractions. The nickel peaks that are not found in the Δ*yeiR* strain extract were most likely bound to small molecular weight species as they eluted off the size-exclusion column in the later fractions and the peaks were relatively sharp.

The nickel-phenotype was reproduced by the Robinson lab (University of Newcastle), therefore, this observation was investigated further. To date, only two

proteins in *E. coli* are known to require nickel for activity: [NiFe]-hydrogenase and glyoxalase I. Therefore, based on the previous results, the activity of these two proteins was determined for the WT and mutant strains.

Hydrogenase activity

Perturbation of hydrogenase activity seemed an unlikely result of the nickel defect observed in the ICP-MS analysis of the Δ*yeiR* mutant. The *E. coli* strain MC4100 commonly used to assay hydrogenase activity (Wu and Mandrand-Berthelot, 1986, Penfold *et al.*, 2003, Jacobi *et al.*, 1992) carries a 6,678-bp deletion (compared to *E. coli* MG1655) that encompasses a deletion of the *yeiR* gene (Peters *et al.*, 2003). Activity of hydrogenase from MC4100 was the same as from *E. coli* W3110, which has an intact *yeiR* gene (Pinske and Sawers, 2010). To confirm that deletion of *yeiR* does not affect hydrogenase activity as suggested indirectly from previous reports, dihydrogen evolution was measured with a gas chromatograph from 1 ml cultures grown overnight in stoppered culture tubes under a nitrogen atmosphere. Whereas deletion of the known hydrogenase maturation factor, *hypF*, noticeably affected dihydrogen production, deletion of *yeiR* had no effect compared to the parent strain (Figure 4-20A).

Glyoxalase I activity

The glyoxalase system composed of two enzymes, glyoxalase I and glyoxalase II, is responsible for the detoxification of cytotoxic α-keto aldehydes, such as methylglyoxal (Racker, 1951). While glyoxalase I from other sources has been found to use zinc as the catalytic cofactor (Aronsson *et al.*, 1978), glyoxalase I from *E. coli* has been shown to use nickel (Clugston *et al.*, 1998). It has not been determined whether additional factors are involved in ensuring that glyoxalase I from *E. coli* attains the nickel cofactor.

Glyoxalase I activity was assayed by testing the sensitivity of the WT and Δ*yeiR* strains to exogenous methylglyoxal. Deletion of the gene encoding glyoxalase I (*gloA*) results in the inability of the strain to grow in the presence of 0.5 mM methylglyoxal (Kim *et al.*, 2007), however, deletion of *yeiR* had no observable affect on the ability of cells to grow in the presence of exogenous methylglyoxal (Figure 4-20B).

Discussion and Conclusions

The growing experimental information collected in the past few years for some subfamilies of the G3E P-loop GTPase proteins (HypB, UreG and MeaB) indicate that they act as metal insertases and/or metallochaperones. However, a large number of proteins in the COG0523 subfamily are predominately uncharacterized (Chapter 3). This study on the biological role of the *E. coli* COG0523 gene, *yeiR*, helps to shed some light on this poorly defined family and strengthens the hypothesis that a fundamental characteristic of these genes is involvement in metal-related processes.

The results presented here suggest that the previously uncharacterized COG0523 gene *yeiR* is involved in a zinc-related process. Presence of *yeiR* either *in cis* (on the chromosome) or *in trans* (cloned into pBAD24) was found to be required for the optimal growth of *E. coli* MG1655 in the presence of high concentrations of EDTA or Cd(II). These phenotypes are thought to be due specifically to zinc-depletion as both growth defects were exacerbated by the deletion of the genes encoding the high-affinity zinc-transporter, *znuABC*, or rescued by the addition of zinc.

To maintain zinc homeostasis, several bacteria express a high-affinity zinc uptake system to transport zinc across the membrane in an energy-dependent manner (Patzer and Hantke, 1998). Although, several other transporters have been shown to participate

in the uptake of zinc under growth conditions of zinc sufficiency (Grass *et al.*, 2002, Beard *et al.*, 2000), the high-affinity zinc transport system is expressed specifically to transport zinc into the cell under growth conditions of zinc deficiency. This transporter is from the ABC family and is composed of three proteins: ZnuC serves as the ATPase component, while ZnuA, the periplasmic zinc metallochaperone, delivers zinc to ZnuB, the membrane permease (Patzer and Hantke, 1998). Although ZnuA is able to bind various metal ions, the structural changes induced specifically by zinc are thought to be responsible for the high specificity of this transporter (Yatsunyk *et al.*, 2008).

As suggested previously (Berducci *et al.*, 2004), the results presented here imply that *znuABC* is able to compete with EDTA for binding to zinc in the medium. As such, approximately 60-fold less EDTA is needed to see the Δ*yeiR* strain's growth defect when *znuABC* is also deleted from the chromosome. The different metal suppression profiles collected for the Δ*yeiR* and Δ*znuABC*::cam Δ*yeiR* mutants could also be a direct result of the competition between the transporter and EDTA. With *znuABC* present on the chromosome, only 25 µM zinc was able to suppress the presence of 1.4 mM EDTA (Figure 4-9A). However, with *znuABC* deleted, 10 µM zinc was needed to suppress the presence of 20 µM EDTA (Figure 4-9B).

In addition to zinc, cobalt and to a lesser extent nickel were also able to suppress the EDTA-dependent growth defect exhibited by the Δ*znuABC*::cam Δ*yeiR* mutant (Figure 4-9B). At this ratio of EDTA:metal (2:1), it is likely that cobalt and nickel are displacing zinc from EDTA making it available to the cell as has been suggested previously (Patzer and Hantke, 1998). The stability constants between EDTA and Co(II) and EDTA and Zn(II) are roughly the same (log K equals 18.1 and 18.3, respectively)

(Stumm and Morgan, 1996). The stability constant between EDTA and Ni(II) is roughly two order of magnitude higher (log K equal to 20.4) (Stumm and Morgan, 1996).

A remote possibility is that cobalt and nickel are transported into the cell and can compensate for the lack of zinc. *In vitro*, the activity of several zinc-dependent enzymes is slightly less, the same or in some cases higher with a metal cofactor other than zinc (for recent examples see (Campos-Bermudez *et al.*, 2007, Cámara *et al.*, 2008, Hall *et al.*, 2007)). Under zinc limitation and supplementation with cobalt, the zinc in carbonic anhydrase of the marine diatom, *Thalassiosira weissflogii*, is substituted with cobalt *in vivo* and remains active (Yee and Morel, 1996). However, expression of the high-affinity Co(II)-transporter from *Salmonella typhimurium* LT2 did not lead to suppression of the EDTA defect in the presence or absence of exogenous cobalt (data not shown).

The first characterization of ZnuABC from *E. coli* was performed in the strain MC4100 (Patzer and Hantke, 1998), which contains a deletion in *yeiR* (Peters *et al.*, 2003). The authors found that growth of the *znuA*::MudX and *znuB*::MudX mutants on solid medium was inhibited in the presence of 400 µM EDTA. A repeat of those experiments (with the MG1655-derived strains carrying deletions in the zinc transporter genes with or without deletion of *yeiR*), revealed that a 10-fold increase in EDTA concentration was required to inhibit the growth of the Δ*znuABC::*cam strain with *yeiR* present (data not shown). Therefore during the original phenotype elucidation of *znuA* and *znuB*, the authors may have observed the effect of the *yeiR* deletion in concert with the deletion of the transporter.

To further support a role for *yeiR* in metal-related processes, the Δ*yeiR* strain displayed increased sensitivity to the toxic metal Cd(II). One of the effects of cadmium

on the cell has been suggested to be intracellular zinc–limitation (Graham *et al.*, 2009, Zhang *et al.*, 2003a). Protein metal sites where the metal ion is bound to cysteine residues should be sensitive to cadmium competition, because cadmium binds to sulfur groups with a higher affinity (Helbig *et al.*, 2008). Accordingly, the Δ*znuABC*::cam mutant, which is unable to transport zinc in a high-affinity capacity (Patzer and Hantke, 1998), is more sensitive to cadmium than the isogenic parent and the Δ*znuABC*::cam Δ*yeiR* mutant is more sensitive still. Also in support of the hypothesis that cadmium induces zinc-limitation, attempts to suppress the cadmium-phenotype of the Δ*yeiR* strain led to partial rescue of growth by zinc. The presence of equimolar zinc increased the tolerance of the mutant strain to cadmium (Figure 4-11A). Perhaps, to a certain extent, increasing the zinc concentration in the medium ensures loading of zinc-dependent proteins with zinc rather than cadmium. Another explanation is that zinc inhibits cadmium uptake. Zinc has been previously shown to inhibit uptake of cadmium by *E. coli* (Laddaga and Silver, 1985).

Interestingly, manganese was also able to partially suppress the growth defect (to a lesser extent than zinc) (Figure 4-11B). Manganese has not been shown to inhibit cadmium-uptake by *E. coli* (Laddaga and Silver, 1985). An explanation for the ability of manganese to suppress the observed growth defect could be the stimulation of a manganese-dependent SOD by the presence of manganese. It has been proposed that one of the reasons for cadmium-toxicity is superoxide generation (Stohs and Bagchi, 1995). Deletion of both iron-dependent and manganese-dependent SOD was shown to increase the sensitivity of *E. coli* to cadmium (Geslin *et al.*, 2001). Therefore, increased activity due to cofactor ability could enhance the activity of manganese-SOD and thus

increase resistance of *E. coli* to cadmium. However, neither zinc nor manganese suppressed the cadmium-specific growth defect of the WT strain.

Ensuring proper metal allocation in the active sites of metalloenzymes is critical for the survival of any organism. Studies involving the maturation of [NiFe]-hydrogenase and of urease have provided the most thorough picture of metallocenter assembly (Kuchar and Hausinger, 2004, Leach and Zamble, 2007). The mechanism and function of each accessory factor required for maturation of these enzymes is still not fully understood, but in both cases a G3E P-loop GTPase is involved in the incorporation of the Ni(II) ion into the enzyme's active site. By analogy, members of COG0523, which have motifs that classify them as G3E P-loop GTPases, could have similar roles in the activation of target enzymes. More specifically, YeiR may be acting as a chaperone involved in the activity of a target metalloeznyme whose activity is essential under zinc-deficient growth conditions.

To further support a role in a metal-related process, the putative metal-binding site $C_{63}MCC_{66}$ was found to be essential for the ability of the corresponding gene *in trans* to complement the *yeiR* deletion (Figure 4-16B). The short histidine stretch was also found to be putatively involved in activity (Figure 4-17B). However, single mutations of the histidine motif were not as detrimental as single deletions in the cysteine motif. Indeed mutation of either H209 or H211 had no observable effect on the ability of the corresponding gene to rescue growth, and H207 only affected the growth rate and not the yield. Mutation of all three histidines appears to be required to sufficiently inactivate the gene product of *yeiR*. Deletion of the His-stretch found in the G3E protein HypB from *B. japonicum* was found to cause a partial defect in hydrogenase activity, which

could be restored by providing high concentration of nickel to the growth medium (Olson *et al.*, 1997).

The characterized and partially characterized P-loop GTPases of the G3E family are commonly found in the same operon as their target enzyme. Co-expression of the chaperone and the target is important; without the chaperone, the target metalloenzyme is inactive. However, deletion of any of the genes surrounding *yeiR* did not result in an observable growth defect in the presence of EDTA (data not shown), suggesting that these genes do not encode the missing target protein. As an alternative conclusion, YeiR may have a general chaperone role and therefore does not target a single enzyme.

If *yeiR* encodes a metal chaperone, it was hypothesized that the intracellular pool could be disrupted by deletion of *yeiR.* Indeed, deletion of *yeiR* appeared to affect several peaks from the ICP-MS analysis of the cytoplasmic fraction after native two-dimensional separation as compared to extract from the WT strain. Most noticeably, nickel-bound to putative small molecular weight species appeared to have disappeared in the mutant strain. The results of the ICP-MS analysis suggest a potential role of *yeiR* in nickel-metabolism. However, no other link with nickel was made during the course of this study.

This study provides critical insights into the potential role COG0523 genes and more specifically the previously uncharacterized gene *yeiR* in metal-related processes. The *B. subtilis* COG0523 gene *yciC* was previously shown to display an EDTA-dependent growth defect in a Δ*znuA* background (Gaballa and Helmann, 1998, Gaballa *et al.*, 2002), which was later confirmed to be due specifically to zinc-depletion (Gabriel

and Helmann, 2009). Similar to *yciC, yeiR* is linked to growth in zinc-deficient conditions. Unlike *yciC*, deletion of the high affinity zinc-transporter is not necessary to observe a growth defect in the presence of extracellular metal-depletion. Also unlike *yciC*, which is regulated by zinc through Zur (Gabriel *et al.*, 2008), *yeiR* has never been directly identified by any of the studies performed on the response of *E. coli* to zinc-depletion (Sigdel *et al.*, 2006, Graham *et al.*, 2009). Zur regulatory sites are not predicted to exist upstream of *yeiR* (Novichkov *et al.*, 2010). The results presented here on *yeiR* may suggest an uncharacterized zinc homeostatic process in *E. coli* during zinc-deplete conditions.

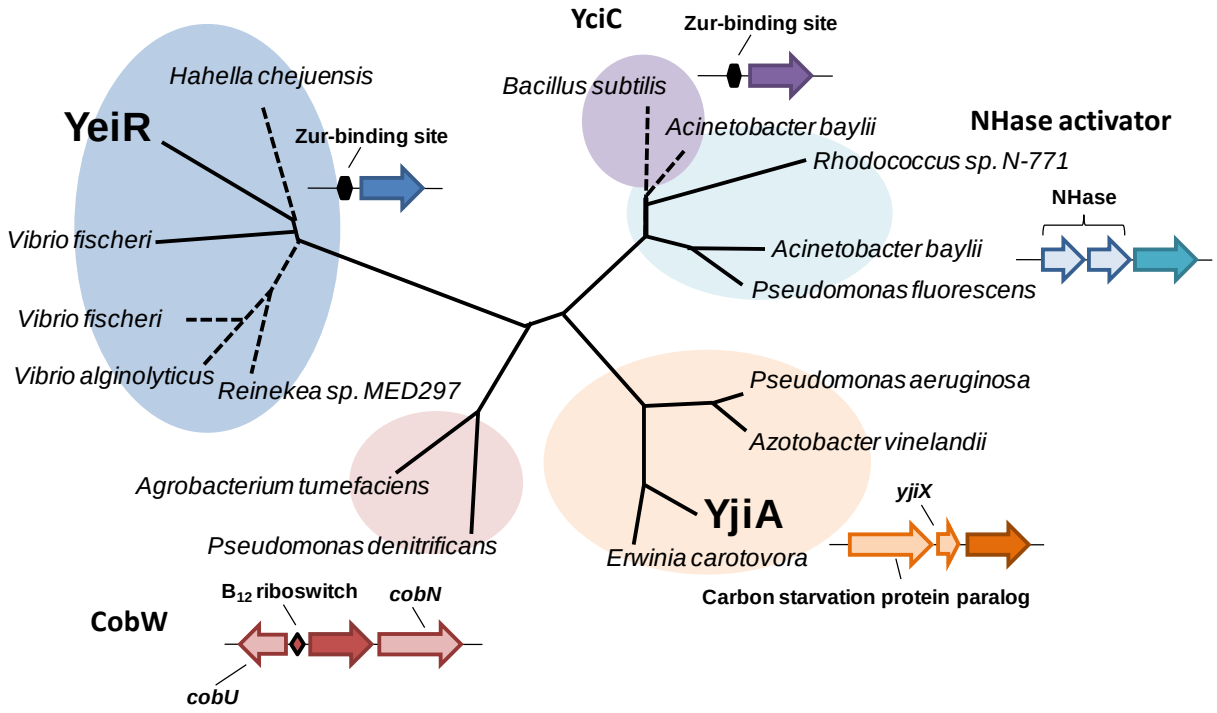


Figure 4-1. Phylogenetic tree reconstruction of chosen COG0523 subgroups. Distance tree of the three COG0523 subgroups for which experimental work is available (CobW (subgroup 12), YciC (subgroup 1), and NHase activator (subgroup 2)) and the two uncharacterized COG0523 proteins in *E. coli* YeiR (subgroup 10) and YjiA (subgroup 9). The background for each subgroup is shaded. Proteins encoded by a gene putatively regulated by Zur (Novichkov *et al.*, 2010) are indicated by a dotted line. Representative genome context for each subgroup is given.

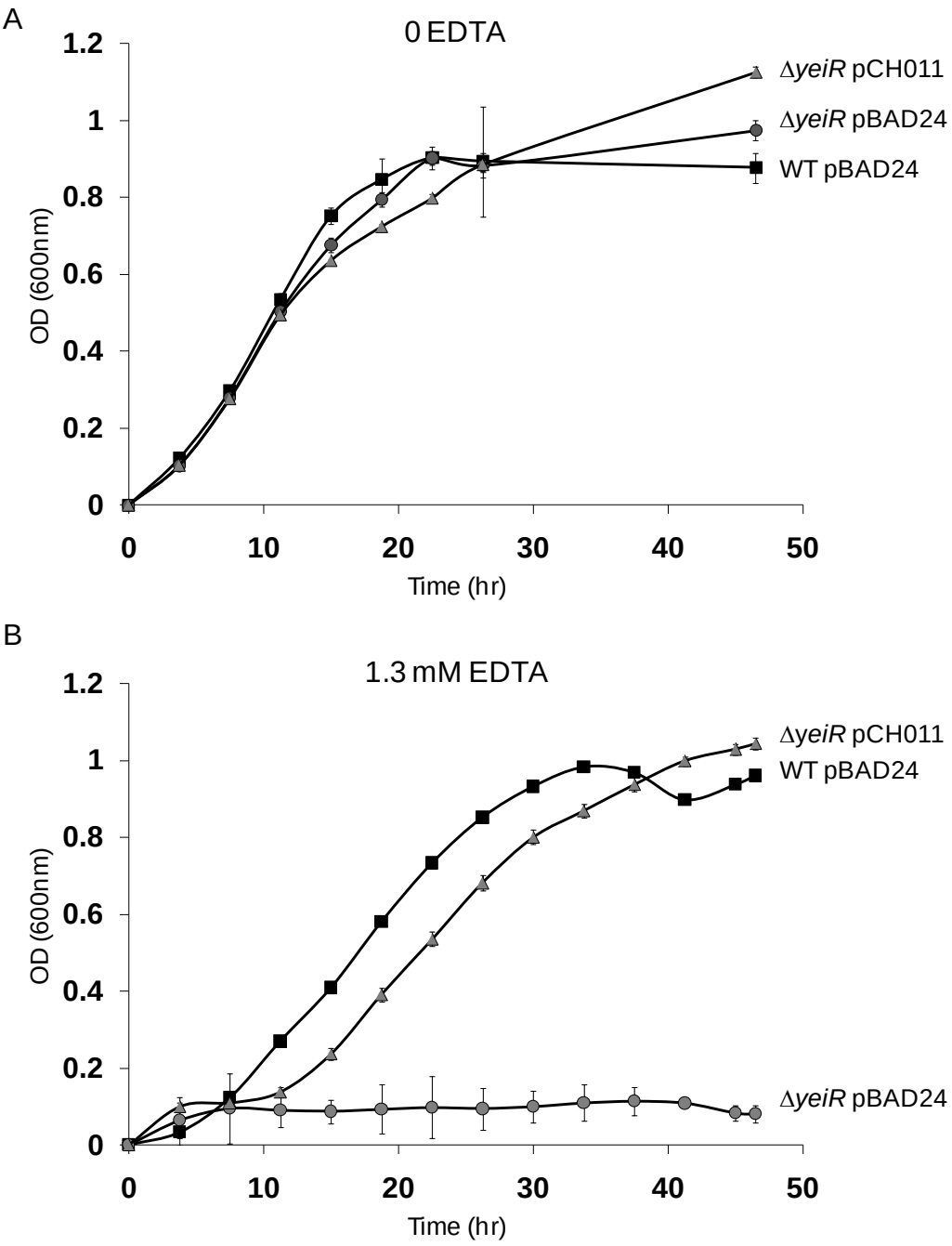


Figure 4-2. Growth curves of *E. coli* MG1655 (WT) and Δ*yeiR* strains grown in LP medium without or with 1.3 mM EDTA. A) Growth curves of the WT pBAD24, Δ*yeiR* pBAD24, and Δ*yeiR* pCH011 (P_{BAD}–*yeiR*) strains in LP medium. B) Growth curves of the WT pBAD24, Δ*yeiR* pBAD24, and Δ*yeiR* pCH011 (P_{BAD}–*yeiR*) strains in LP medium plus 1.3 mM EDTA. Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of three replicates.

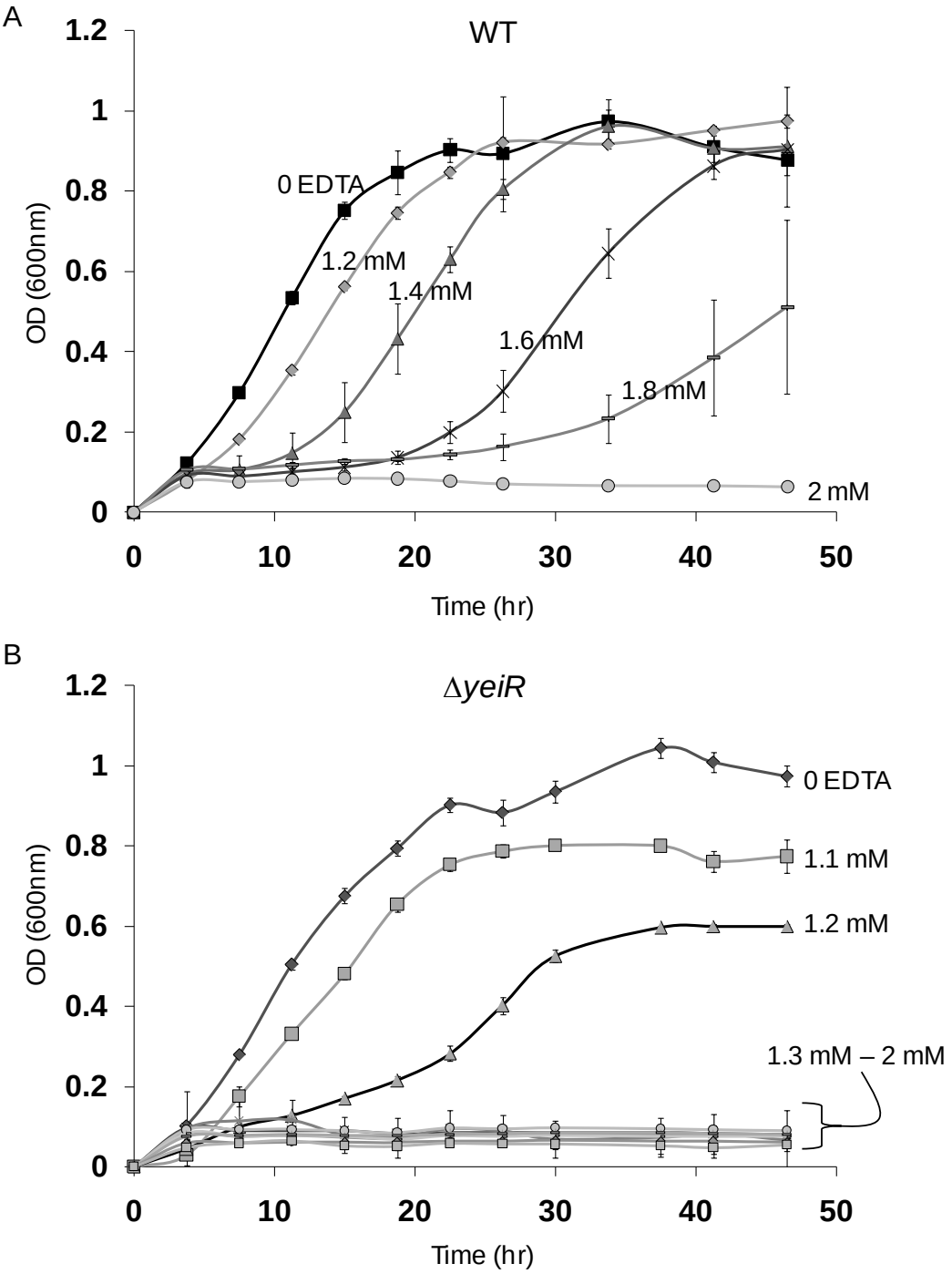


Figure 4-3. Growth curves of *E. coli* MG1655 (WT) and Δ*yeiR* strains grown in LP medium with a range of EDTA concentrations. A) Growth curves of the WT strain grown in the presence of 0, 1.2 mM, 1.4 mM, 1.6 mM, 1.8 mM or 2 mM EDTA. B) The Δ*yeiR* strain grown in the presence of 0, 1.1 mM, 1.2 mM (or 1.3 mM – 2 mM EDTA as shown (in steps of 0.1 mM). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of three replicates.

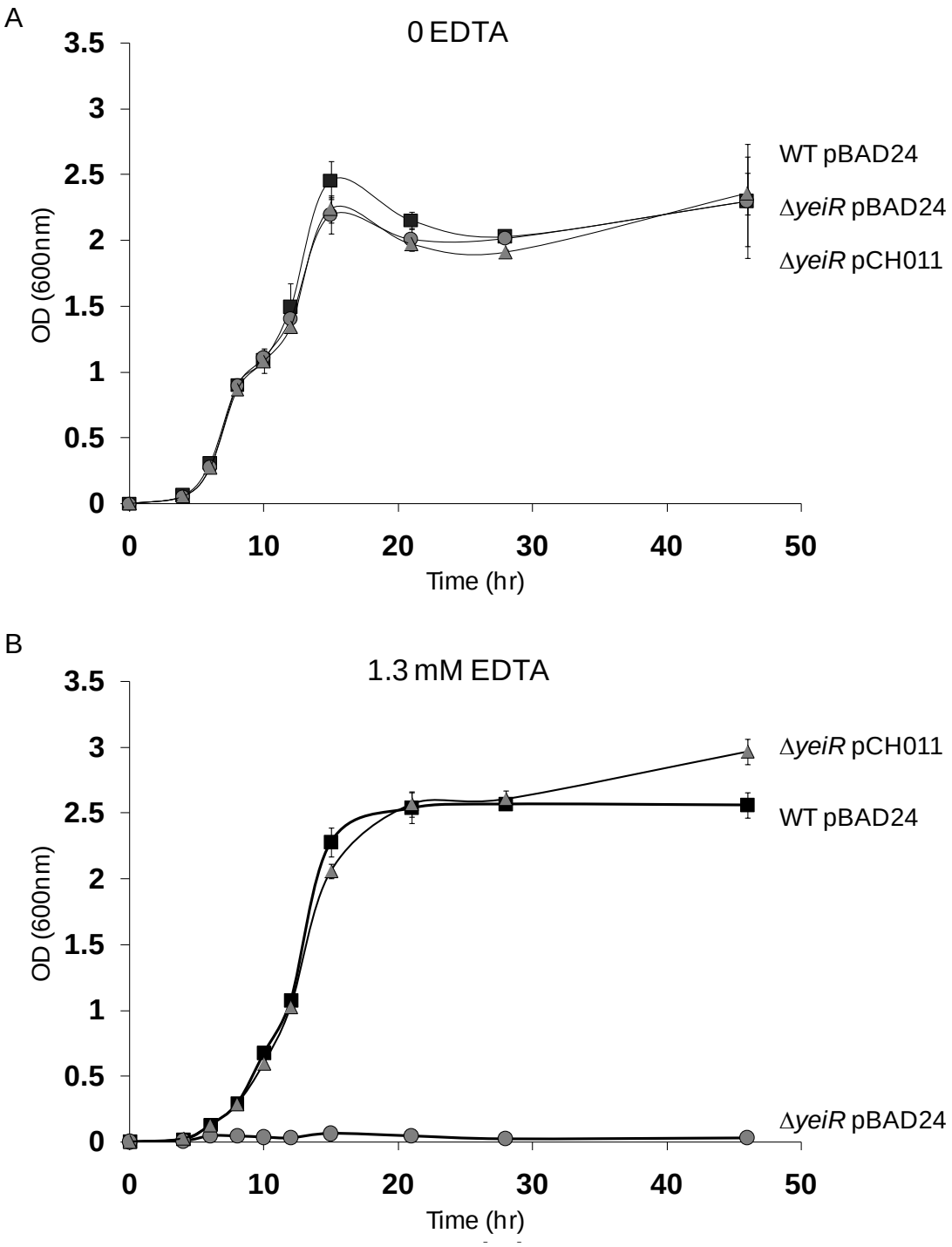


Figure 4-4. Manual growth curves. Manual growth curves of *E. coli* MG1655 (WT) pBAD24, Δ*yeiR* pBAD24, and Δ*yeiR* pCH011 (P_{BAD}–*yeiR*) strains in LP medium (A) or LP medium plus 1.3 mM EDTA (B). 200 mL cultures were grown in 500 mL Erlenmeyer flasks at 200 rpm. The optical density was measured at 600 nm with a BioSpec-mini (Shimadzu Technologies) spectrometer. Error bars represent ± the standard deviation of three replicates.

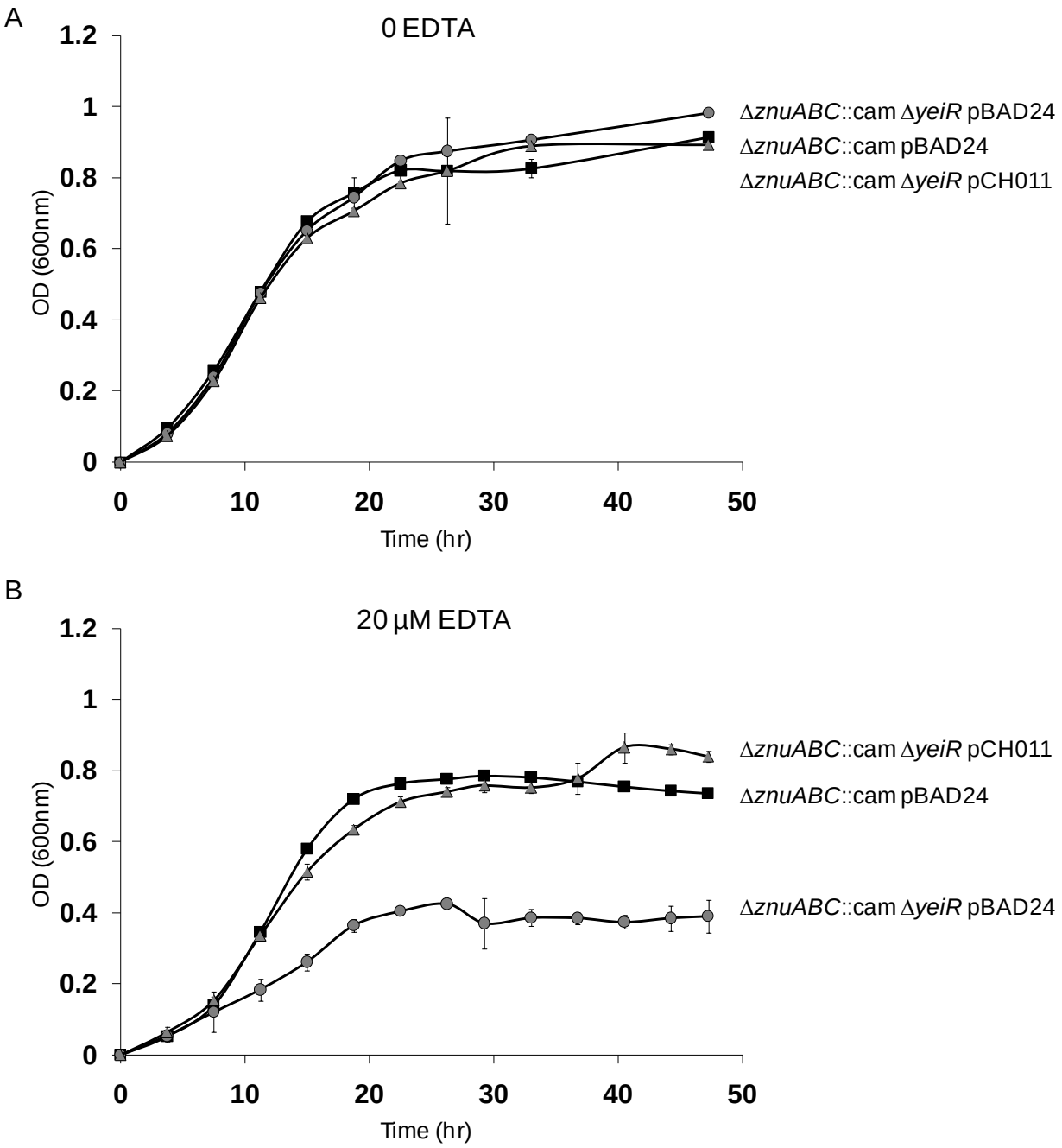


Figure 4-5. Growth curves of the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants grown in LP medium without or with 20 μM EDTA. A) Growth curves of the Δ*znuABC*::cam pBAD24, Δ*znuABC*::cam Δ*yeiR* pBAD24, and Δ*znuABC*::cam Δ*yeiR* pCH011 (P_{BAD}–yeiR) mutants in LP medium. B) Growth curves of the Δ*znuABC*::cam pBAD24, Δ*znuABC*::cam Δ*yeiR* pBAD24, and Δ*znuABC*::cam Δ*yeiR* pCH011 mutants in LP medium plus 20 μM EDTA. Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of three replicates.

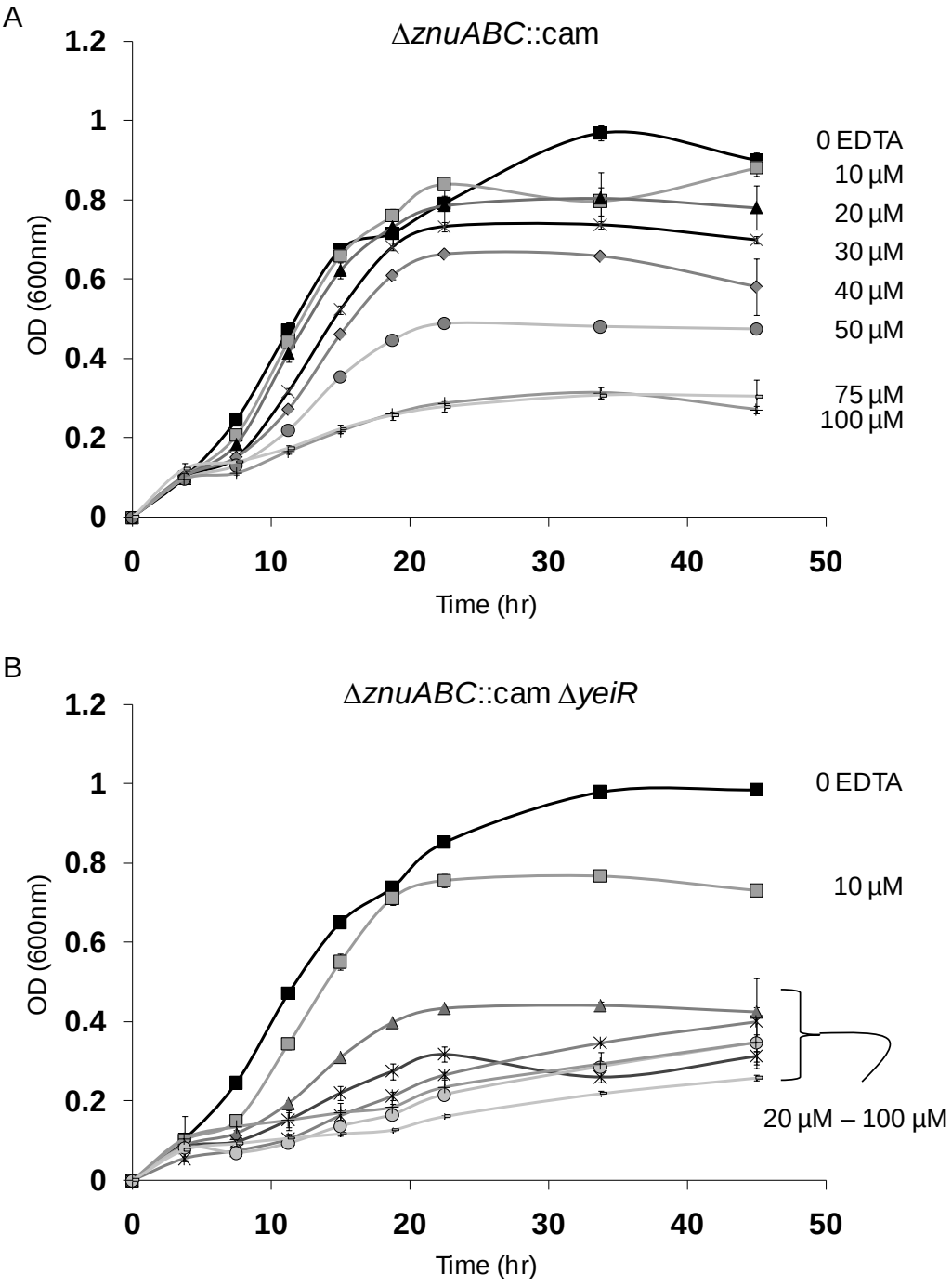


Figure 4-6. Growth curves of the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants grown in LP medium with a range of EDTA concentrations. A) Growth curves of the Δ*znuABC*::cam mutant grown in the presence of 0, 10 µM, 20 µM, 30 µM, 40 µM, 50 µM, 75 µM or 100 µM EDTA. B) The Δ*znuABC*::cam Δ*yeiR* mutant grown in the presence of 0, 10 µM, 20 µM, 30 µM, 40 µM, 50 µM, 75 µM or 100 µM EDTA. Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of three replicates.

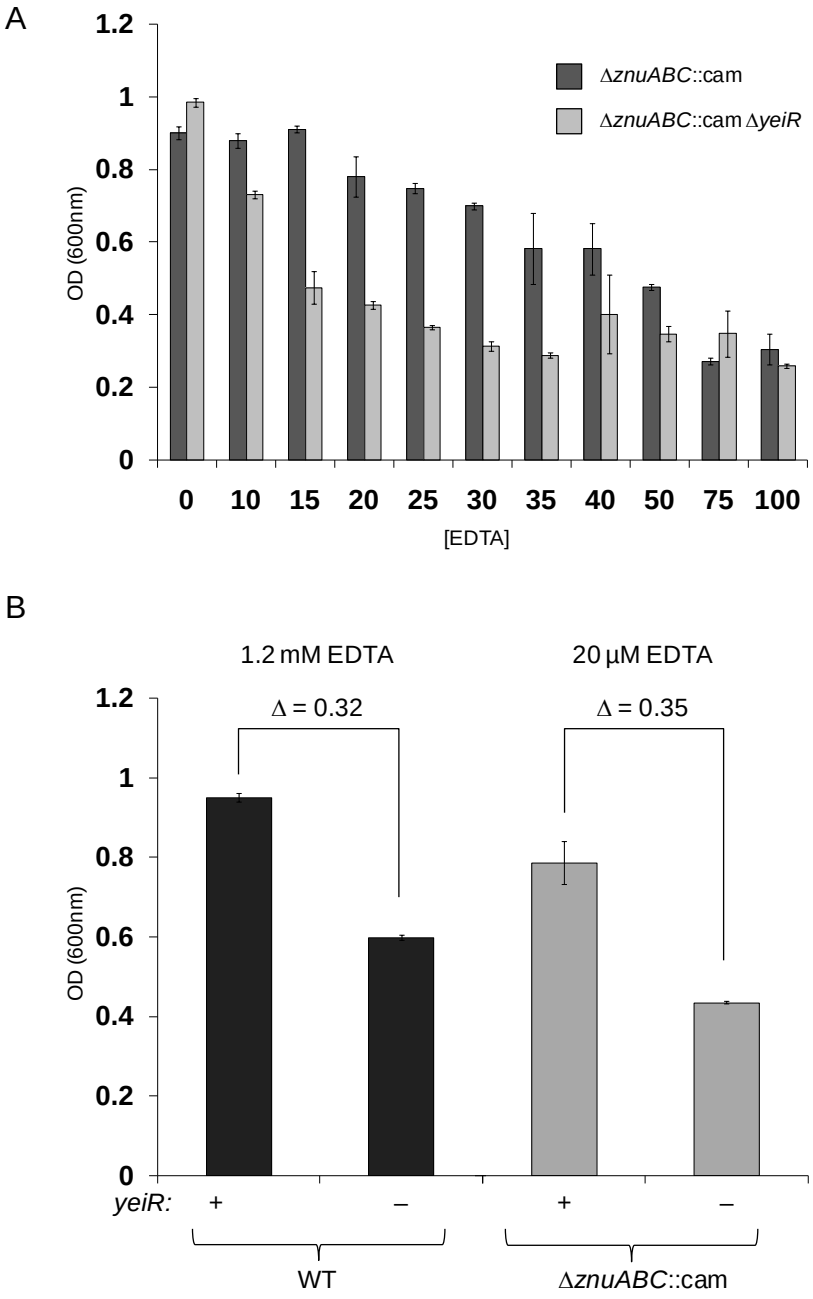


Figure 4-7. Optical density (OD; measured at 600 nm) of cultures. A) OD of the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants in stationary phase (45 hrs after inoculation) that were grown in LP medium in the absence or presence of EDTA at the indicated concentrations (µM). B) Comparison of final OD of the WT and Δ*yeiR* strains grown in LP medium plus 1.2 mM EDTA with the Δ*znuABC*::cam and Δ*znuABC*::cam Δ*yeiR* mutants grown in LP medium plus 20 µM EDTA. OD values were generated with a Bioscreen C. The difference in OD between parent and mutant is given. Error bars represent ± the standard deviation of three replicates.

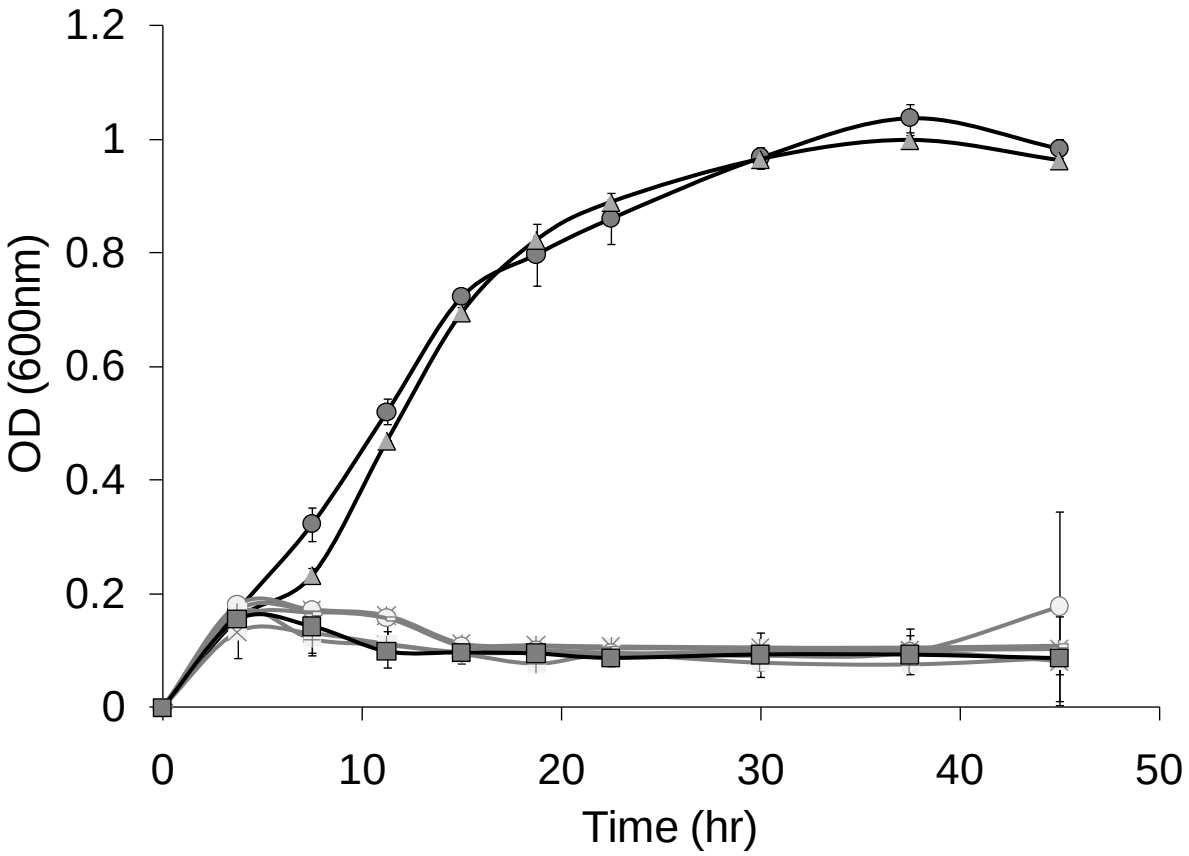


Figure 4-8. Rescue of the Δ*yeiR* strain EDTA-sensitive growth defect by zinc. Growth curves of the Δ*yeiR* strain grown in LP medium in the absence of EDTA (dark grey circle) or in the presence of 1.4 mM EDTA (grey square) plus various metals at 25 μM: zinc (triangle), cobalt (X), copper (star), iron (rectangle), manganese (open circle) or nickel (+). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

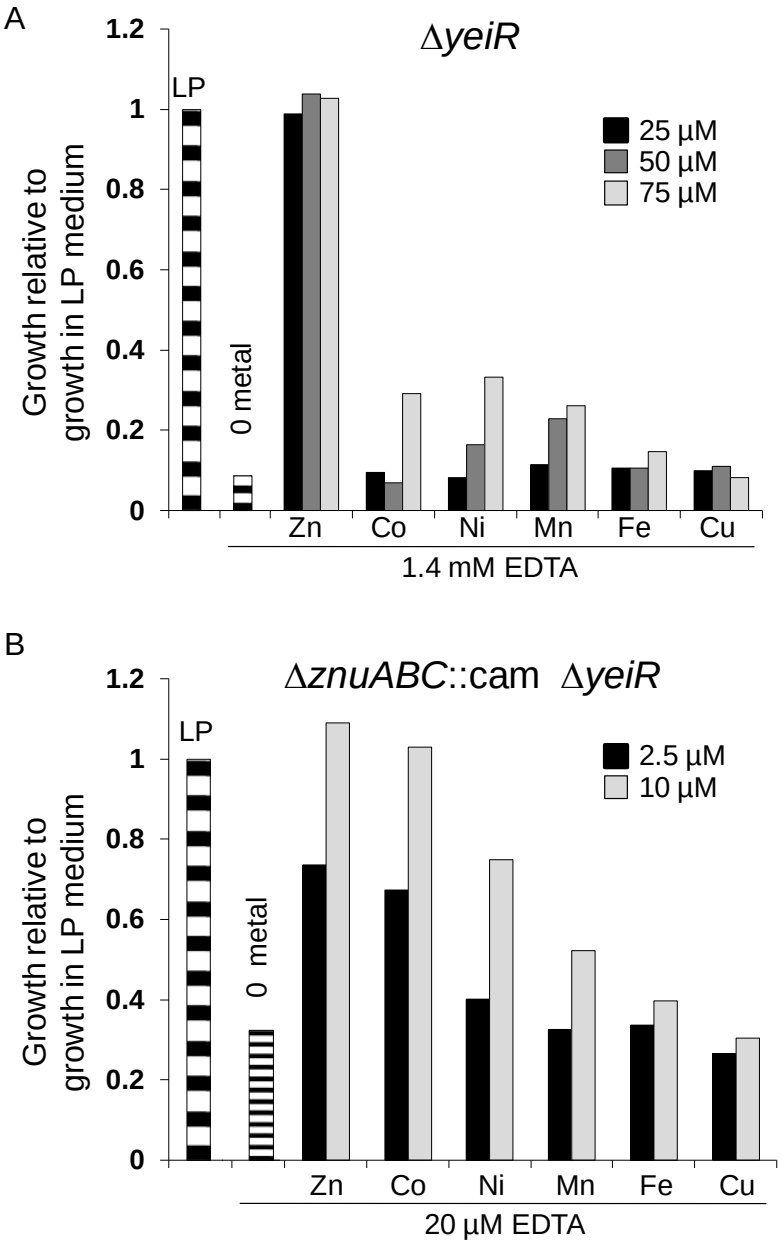


Figure 4-9. Result of adding EDTA and metal to medium. A) Growth of the Δ*yeiR* strain in LP medium in the absence of EDTA (LP) or presence of 1.4 mM EDTA +/– the indicated metals at either 2.5 µM or 10 µM metal. The values shown are a ratio of the final OD of the Δ*yeiR* mutant in the indicated medium relative to the final OD of the Δ*yeiR* mutant in LP medium without metal or EDTA. B) Growth of the Δ*znuABC*::cam Δ*yeiR* mutant in LP medium in the absence of EDTA (LP) or presence of 20 µM EDTA +/– the indicated metals at either 2.5 µM or 10 µM metal. The values shown are a ratio of the final OD of the Δ*znuABC*::cam Δ*yeiR* mutant in the indicated medium relative to the final OD of the Δ*znuABC*::cam Δ*yeiR* mutant in LP medium without metal or EDTA.

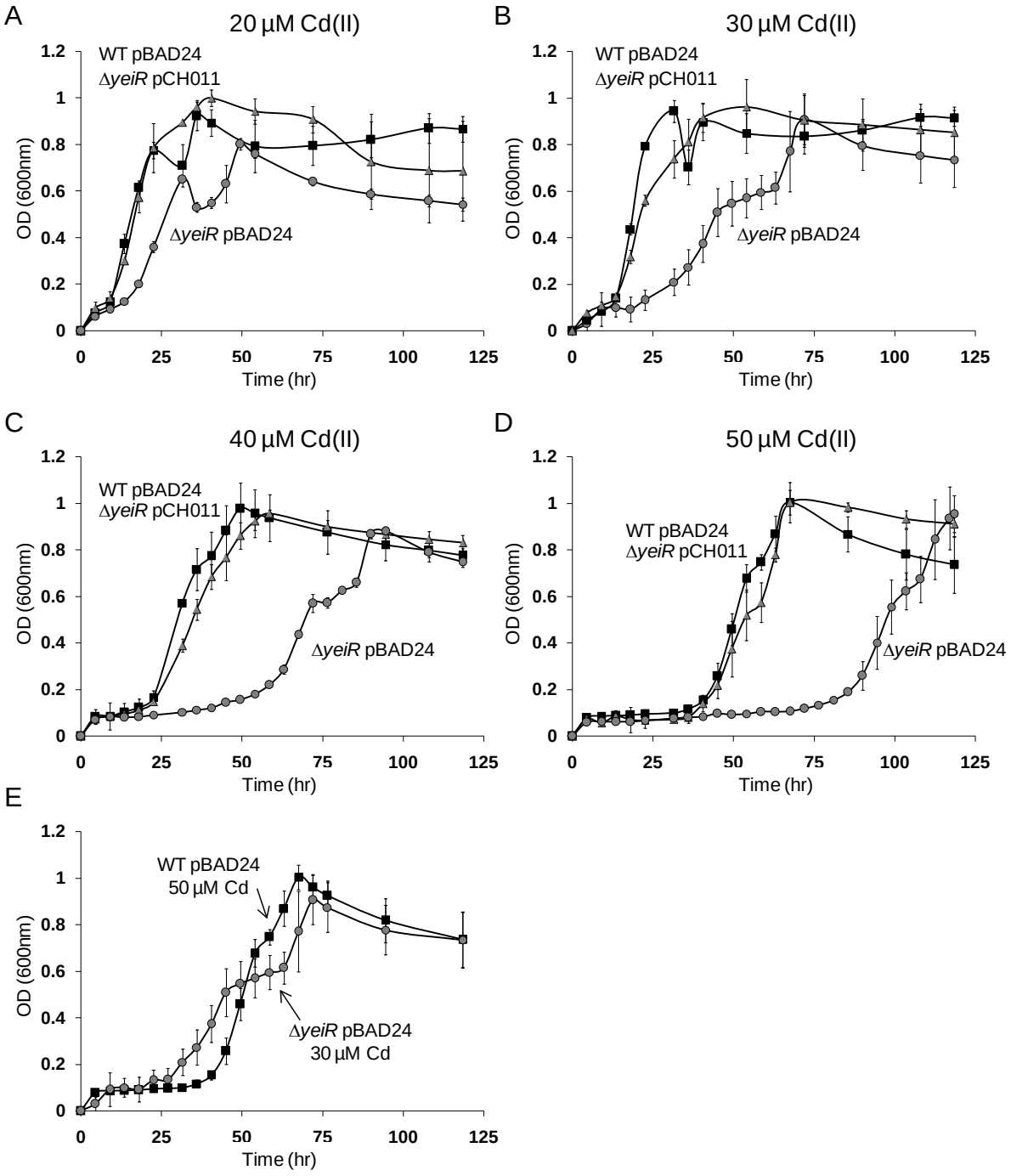


Figure 4-10. Effect of cadmium on growth. Growth curves of *E. coli* MG1655 (WT) pBAD24, Δ*yeiR* pBAD24, and Δ*yeiR* pCH011 (P_{BAD}–*yeiR*) strains in LP medium with Cd(II) at 20 µM (A), 30 µM (B), 40 µM (C) or 50 µM (D). E) Growth curves of WT pBAD24 in the presence of 50 µM Cd and Δ*yeiR* pBAD24 in the presence of 30 µM Cd. Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

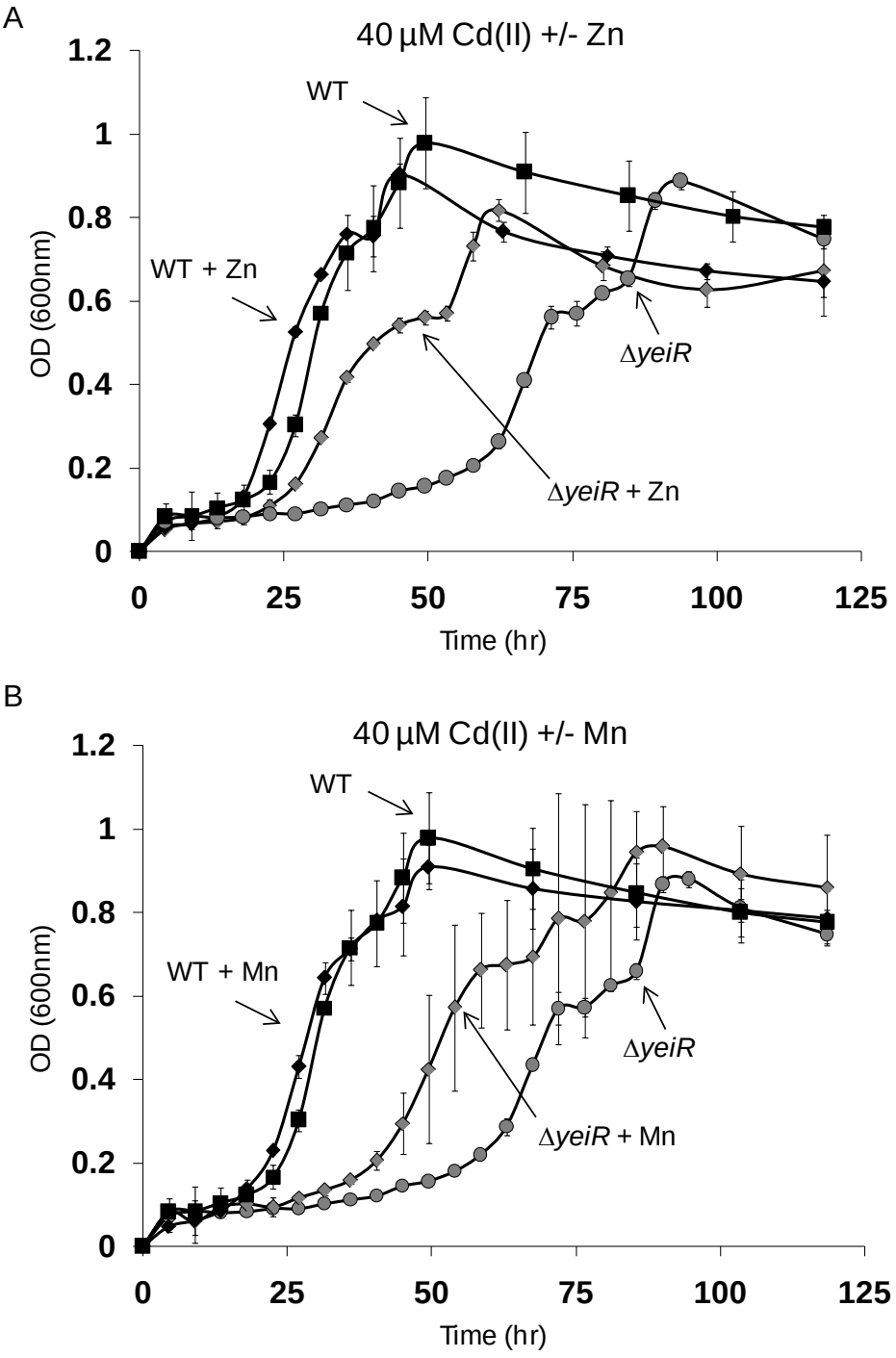


Figure 4-11. Partial rescue of the cadmium-sensitive growth defect of the Δ*yeiR* strain with zinc or manganese. A) LP medium plus 40 µM Cd(II) +/– 40 µM Zn(II); WT or Δ*yeiR* strains without Zn(II) and WT or Δ*yeiR* strains with Zn(II). B) LP medium plus 40 µM Cd +/– 40 µM Zn; WT or Δ*yeiR* strains without Mn(II) and WT or Δ*yeiR* strains with Mn(II). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

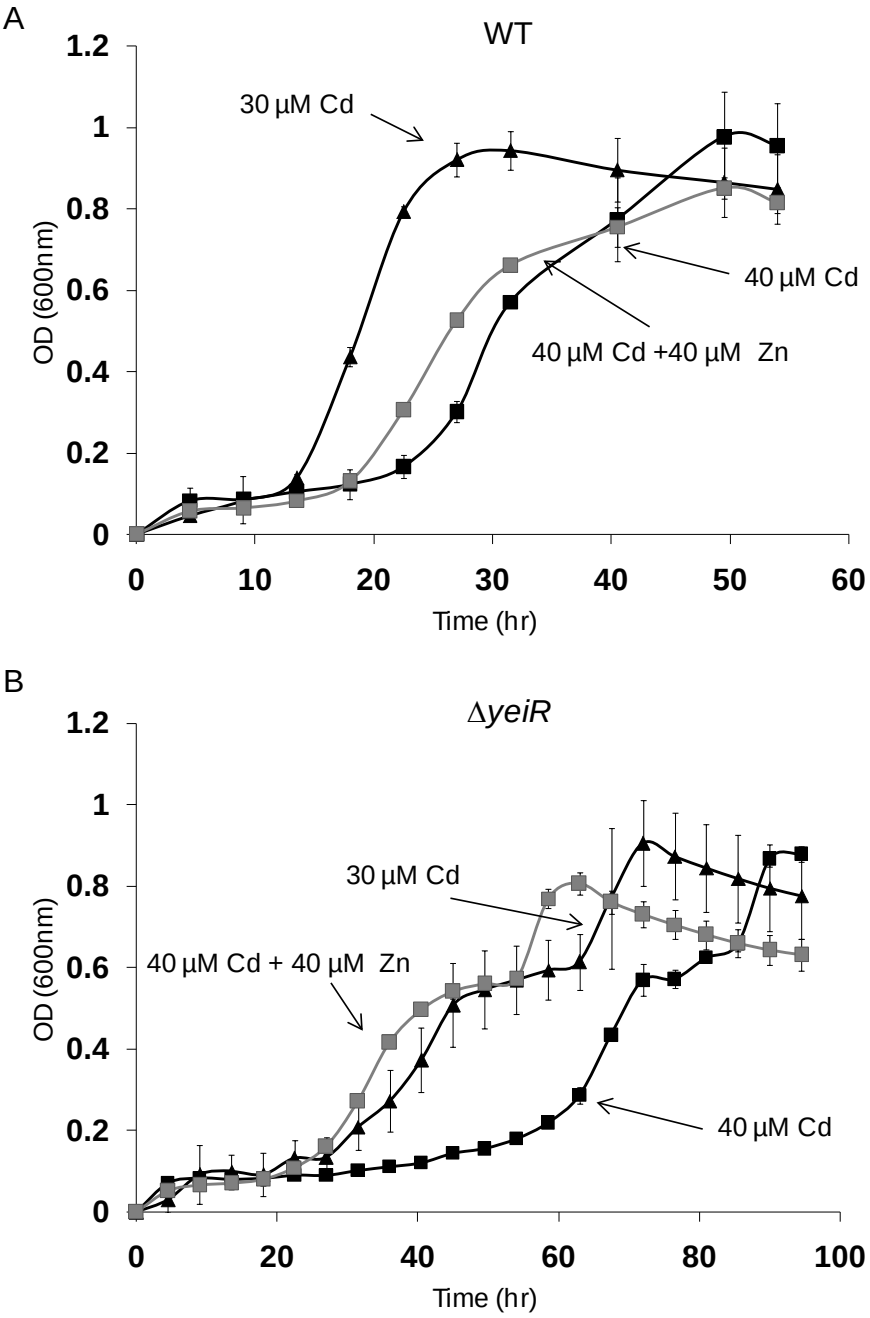


Figure 4-12. Suppression of cadmium-toxicity with zinc. Growth of the WT and Δ*yeiR* strains in: LP medium plus 30 µM Cd(II), LP medium plus 40 µM Cd(II), LP medium plus 40 µM Cd(II) and 40 µM Zn(II). A) The presence of Zn at 40 µM does not rescue growth of WT cells grown in LP medium plus 40 µM Cd(II). B) The presence of equimolar Zn(II) does partially rescue growth of the mutant strain. Growth in LP medium plus 40 µM Cd(II) and 40 µM Zn(II) is equivalent to growth in LP medium with 30 µM Cd(II). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

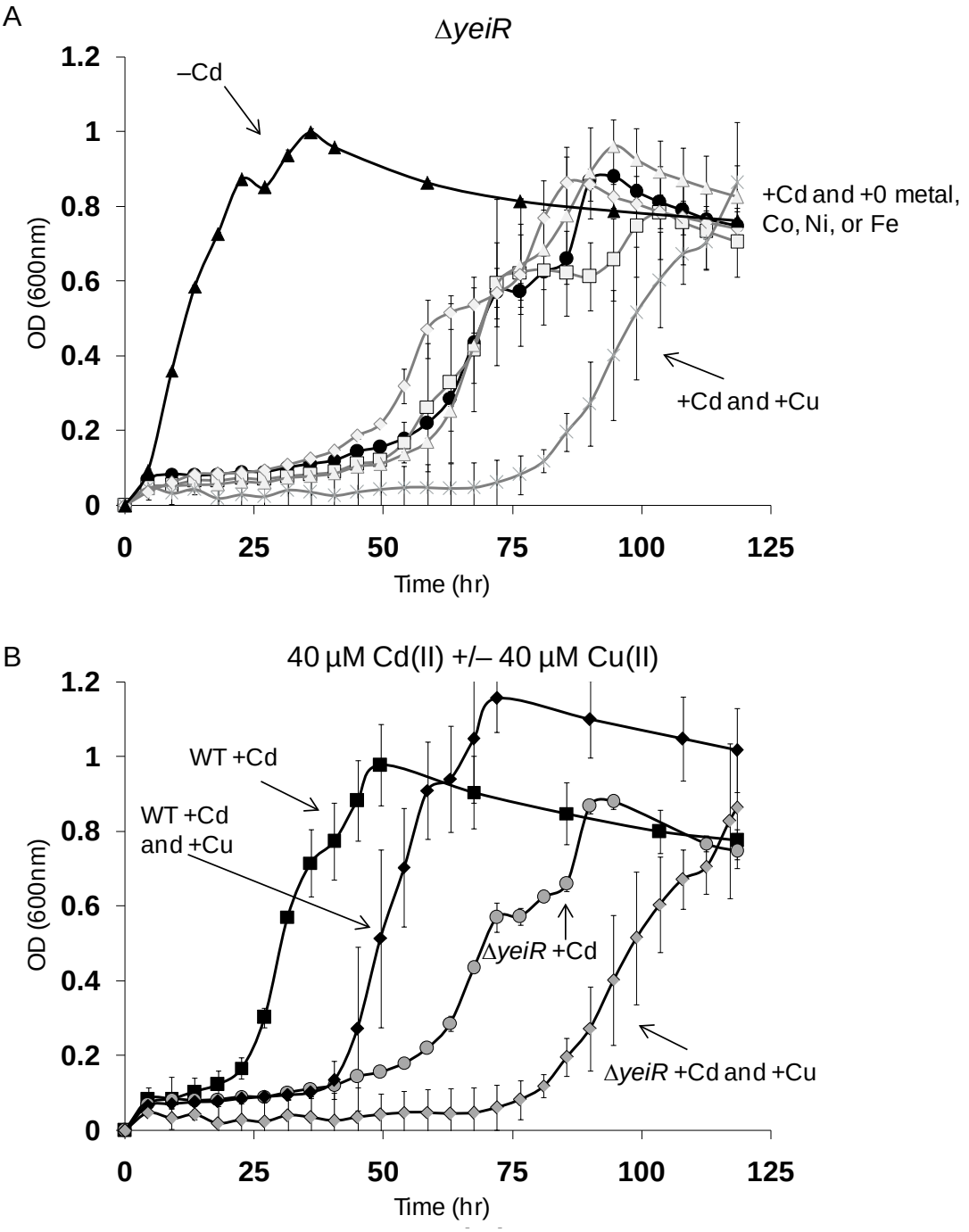


Figure 4-13. Addition of cobalt, copper, nickel or iron plus cadmium on the growth of the Δ*yeiR* strain in LP medium. A) Growth curves of the Δ*yeiR* strain without or with 40 µM Cd(II) plus 40 µM Fe(III), Co(II), Ni(II) or Cu(II). B) Growth curves of the WT and Δ*yeiR* strains in the presence of 40 µM Cd(II) or the WT and Δ*yeiR* strains in the presence of 40 µM Cd(II) plus 40 µM Cu(II). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

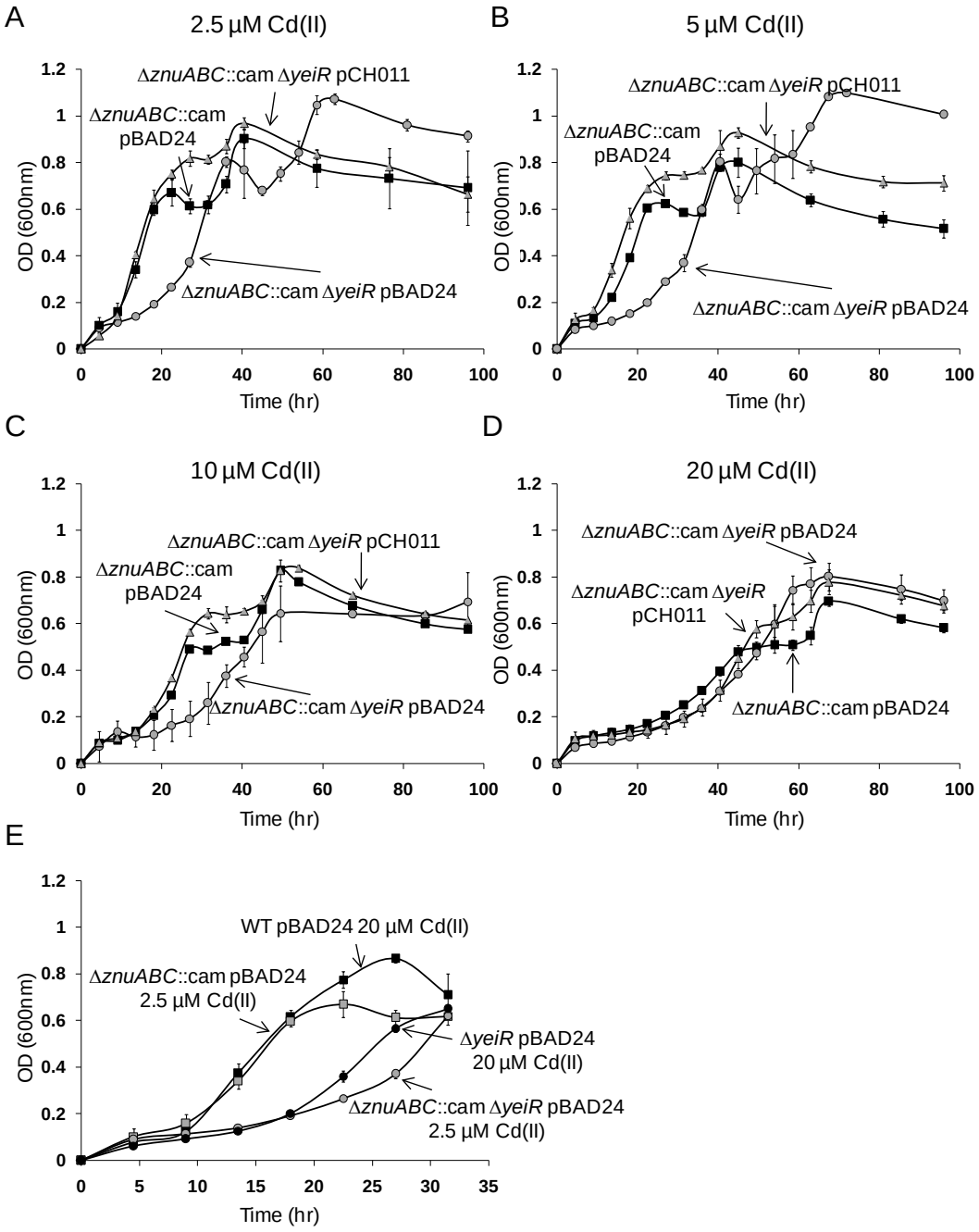


Figure 4-14. Effect of cadmium on growth of Δ*znuABC*::cam strains. Growth curves of *E. coli* MG1655 Δ*znuABC*::cam pBAD24, Δ*znuABC*::cam Δ*yeiR* pBAD24, and Δ*znuABC*::cam Δ*yeiR* pCH011 (P_{BAD}–*yeiR*) strains in LP medium with Cd(II) at 2.5 µM (A),5 µM (B), 10 µM (C) or 20 µM (D). E) Growth curves of the WT pBAD24 and Δ*yeiR* pBAD24 strains in the presence of 20 µM Cd(II) and the Δ*znuABC*::cam pBAD24 and Δ*znuABC*::cam Δ*yeiR* pBAD24 mutants in the presence of 2.5 µM Cd(II). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

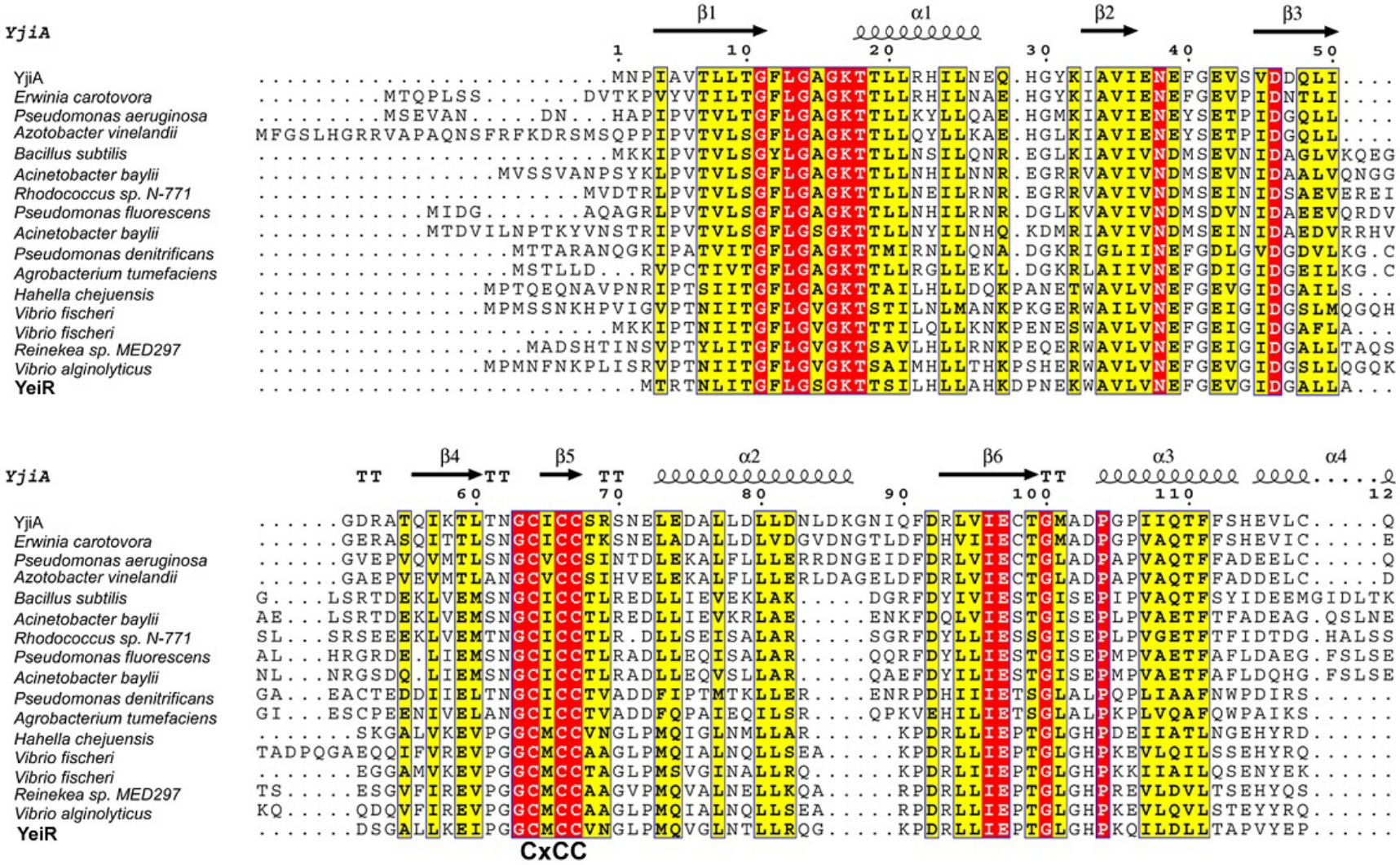


Figure 4-15. Protein sequence alignment of COG0523 proteins. These COG0523 proteins were used for the phylogenetic tree reconstruction in Figure 4-1. Secondary structure as determined from the crystal structure of the YjiA homolog is given at the top of the alignment (PDB: 1NIJ). The motifs analyzed in the mutagenesis study are shown in bold.

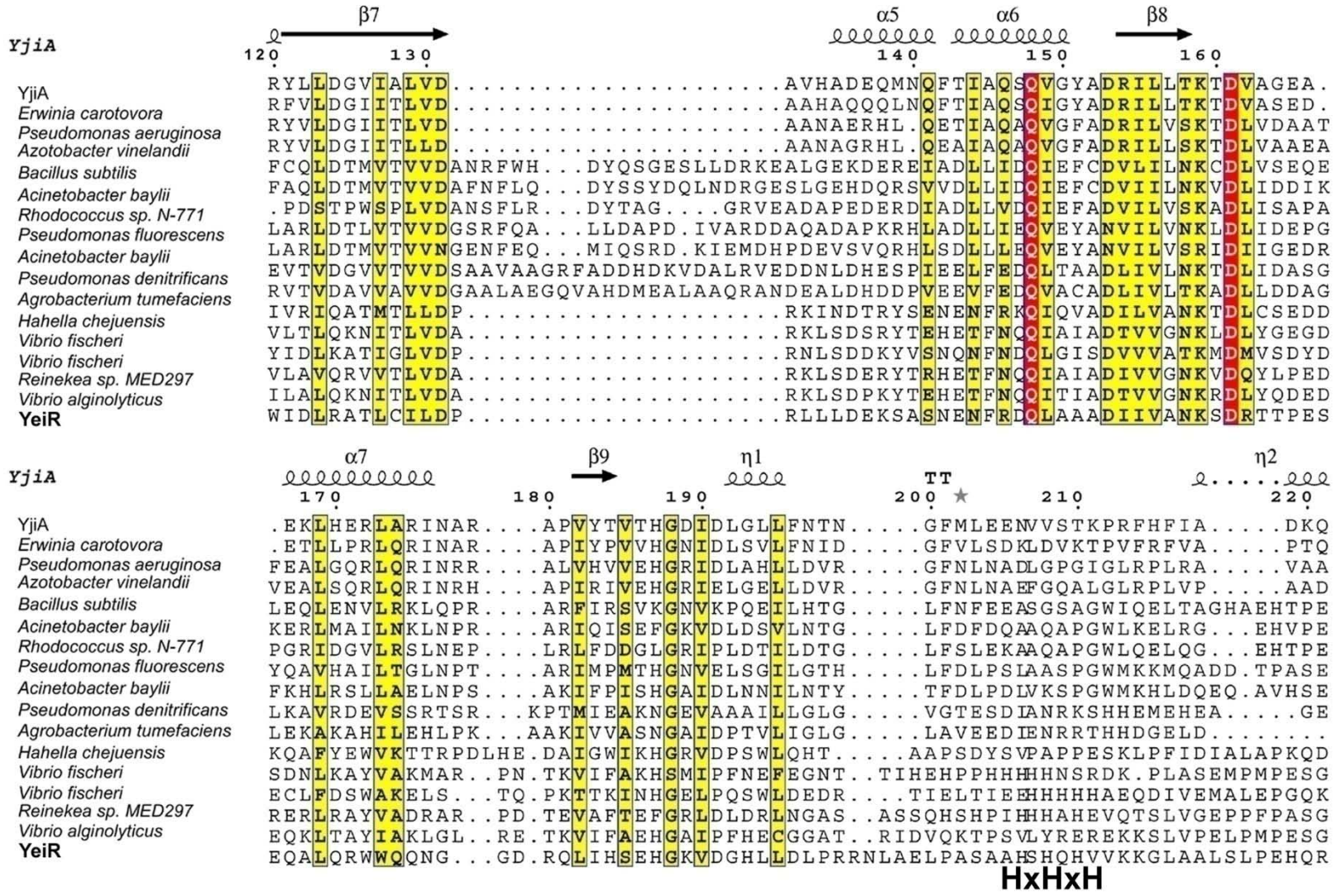


Figure 4-15. Continued.

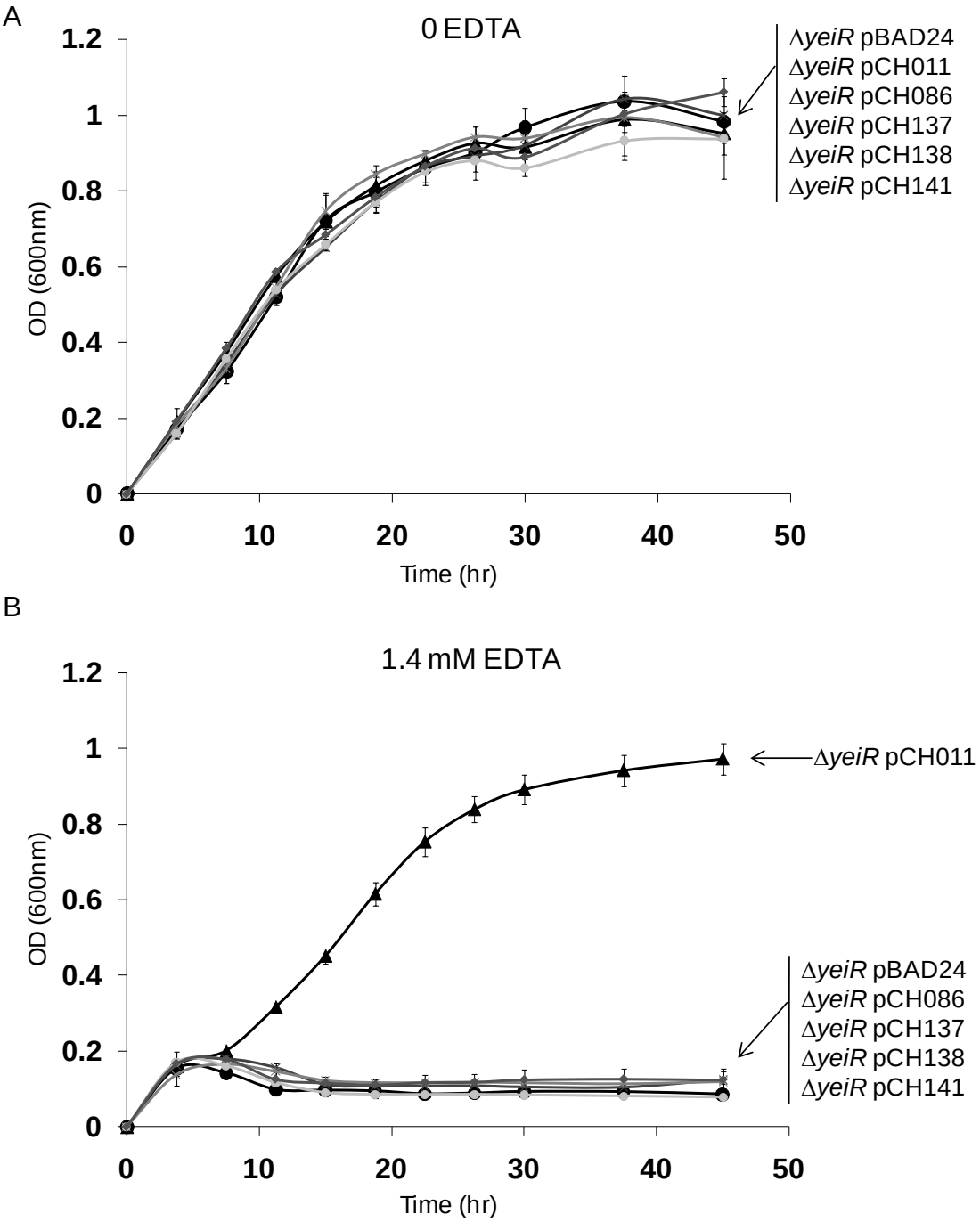


Figure 4-16. Effect of $C_{63}MCC_{66}$ mutations on the ability of the corresponding gene to complement the deletion of *yeiR*. Growth curves of the Δ*yeiR* pBAD24, Δ*yeiR* pCH011 (encodes WT YeiR), Δ*yeiR* pCH086 (encodes YeiR C63A), Δ*yeiR* pCH141 (encodes YeiR M64A), Δ*yeiR* pCH137 (encodes YeiR C65A), and Δ*yeiR* pCH138 (encodes YeiR C66A) mutants in LP medium without EDTA (A) or with 1.4 mM EDTA (B). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

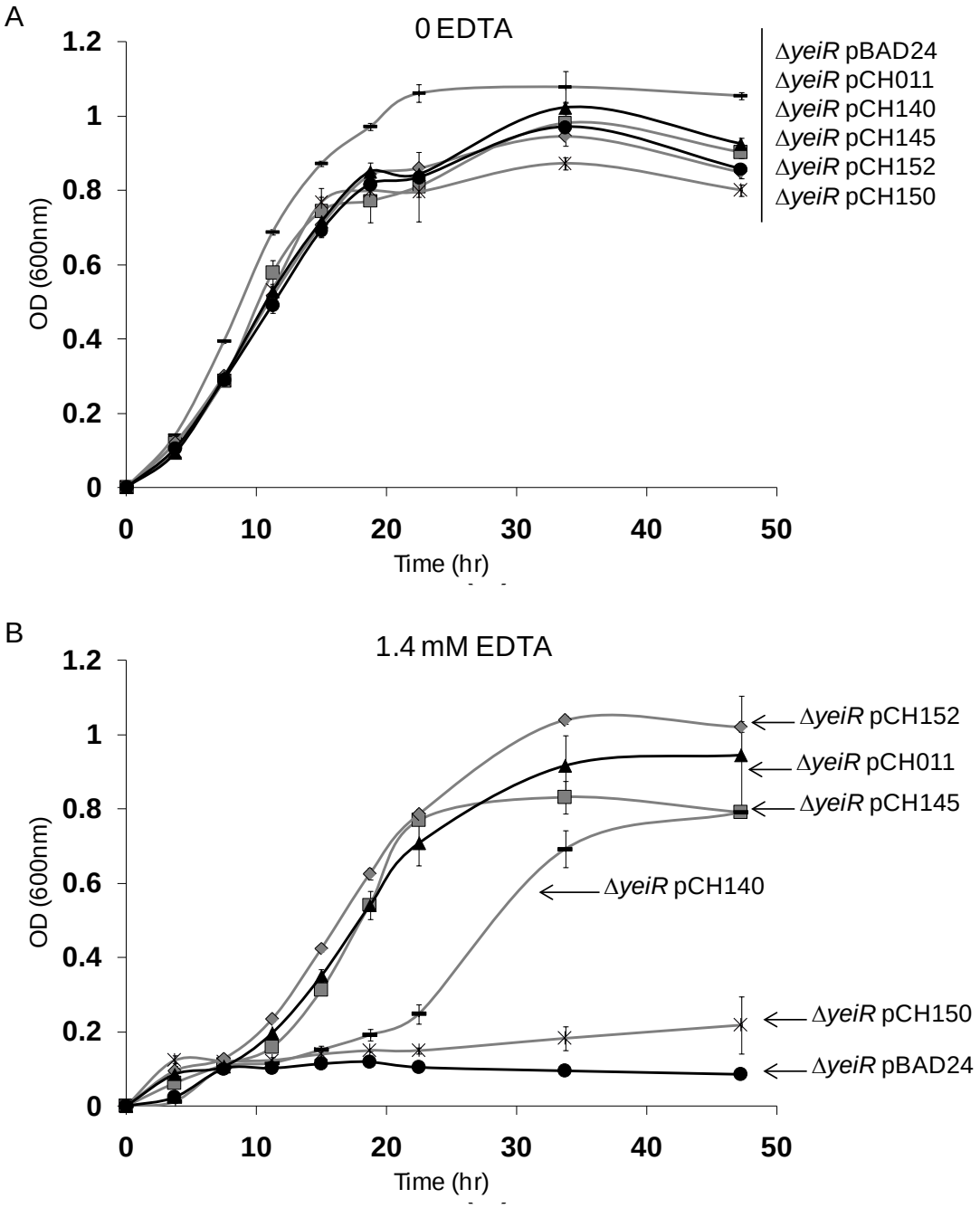


Figure 4-17. Effect of $H_{207}XHXH_{211}$ mutations on the ability of the corresponding gene *in trans* to complement the deletion of *yeiR*. Growth curves of the Δ*yeiR* pBAD24, Δ*yeiR* pCH011 (encodes WT YeiR), Δ*yeiR* pCH140 (encodes YeiR H207A) (**–**), Δ*yeiR* pCH145 (encodes YeiR H209A), Δ*yeiR* pCH152 (encodes YeiR H211A) and Δ*yeiR* pCH150 (encodes YeiR H207A H209A H211A) mutants were grown in LP medium without EDTA (A) or with 1.4 mM EDTA (B). Growth curves were generated with a Bioscreen C. Error bars represent ± the standard deviation of 3 replicates.

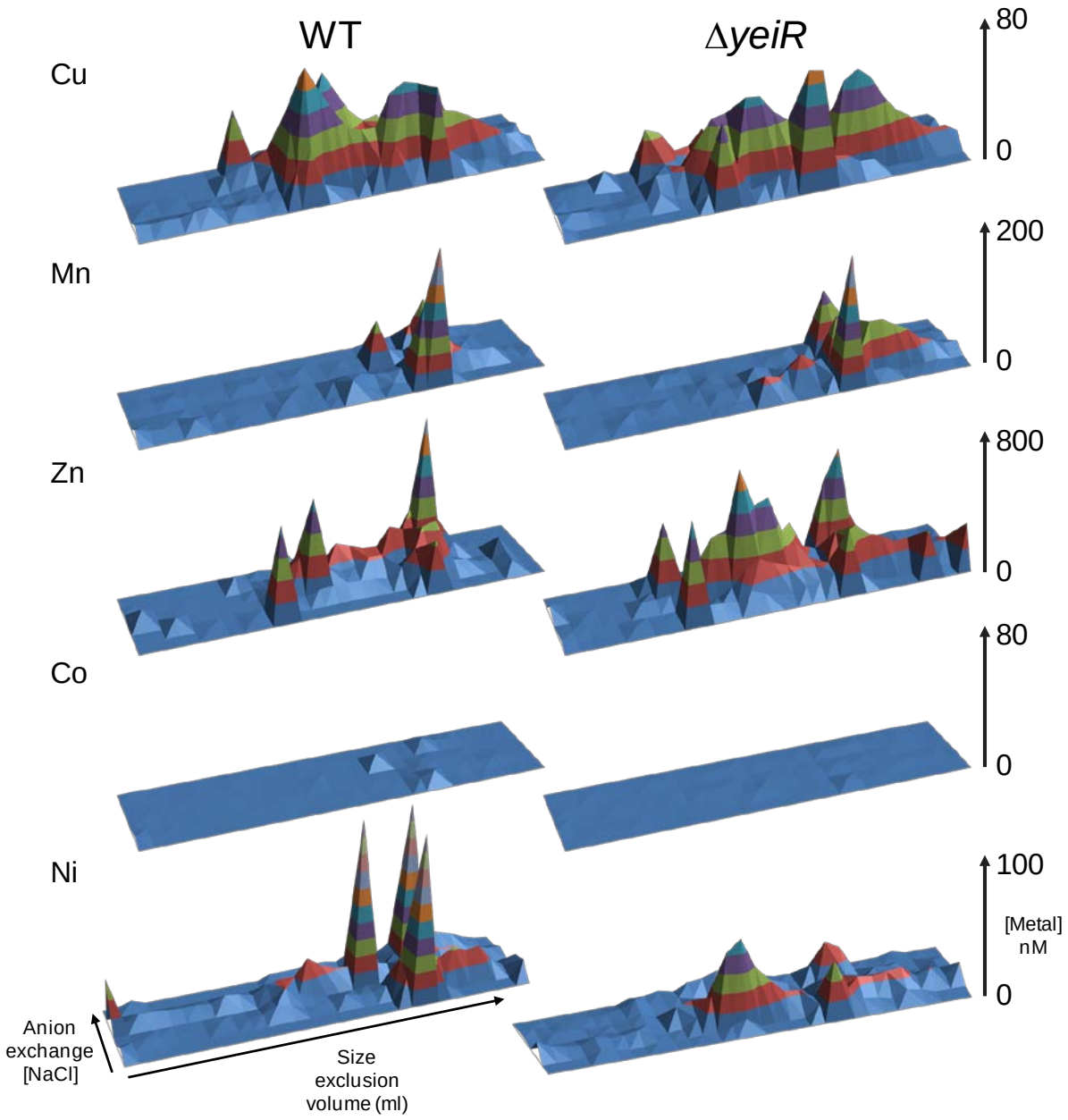


Figure 4-18. Native two-dimensional separation analyzed by ICP-MS for five elements. Native two-dimensional separation of a cytoplasmic extract from *E. coli* MG1655 (WT) and the Δ*yeiR* mutant analyzed for copper, manganese, zinc, cobalt and nickel by ICP-MS. Individual metal concentrations are represented as surfaces. The colors represent incremental increases in concentration (nM): copper, 10; manganese, 20; zinc, 100; cobalt, 10; nickel, 10. The bottom left graph shows the axis labels that are the same for each graph. The arrow on the right shows the scale to which each metal concentration is plotted.

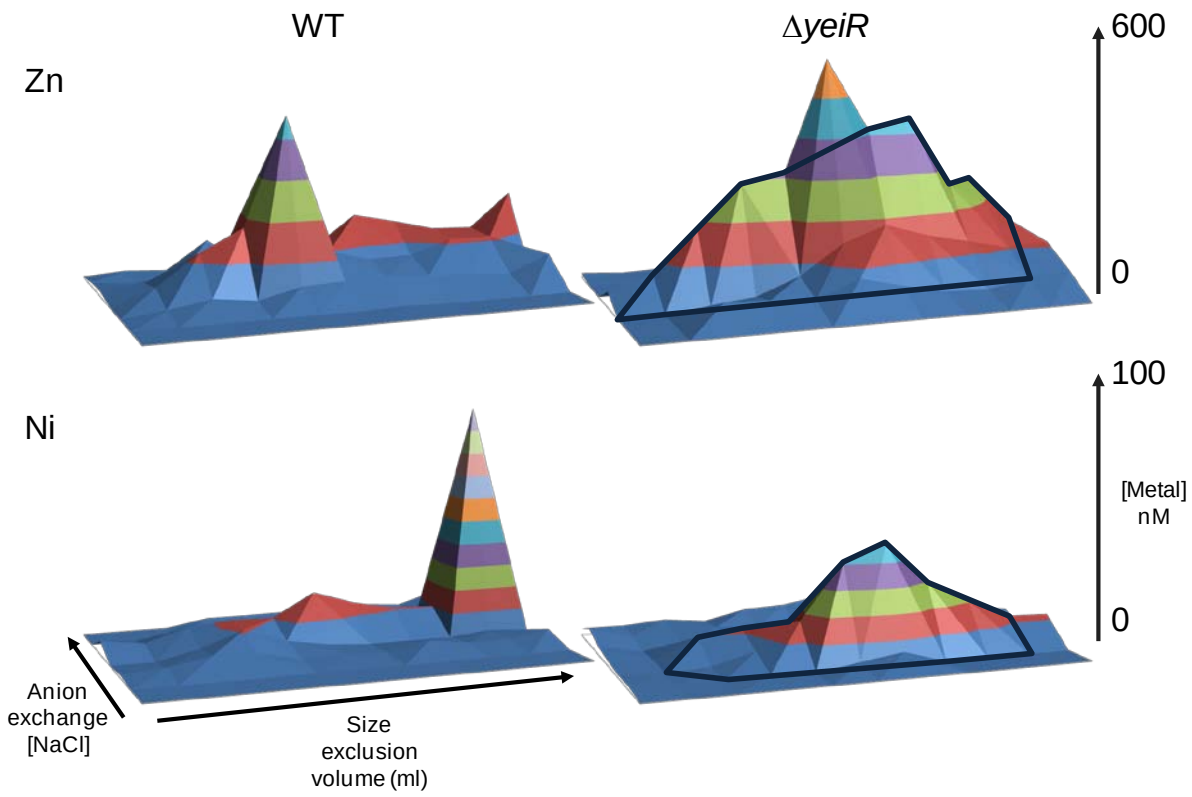
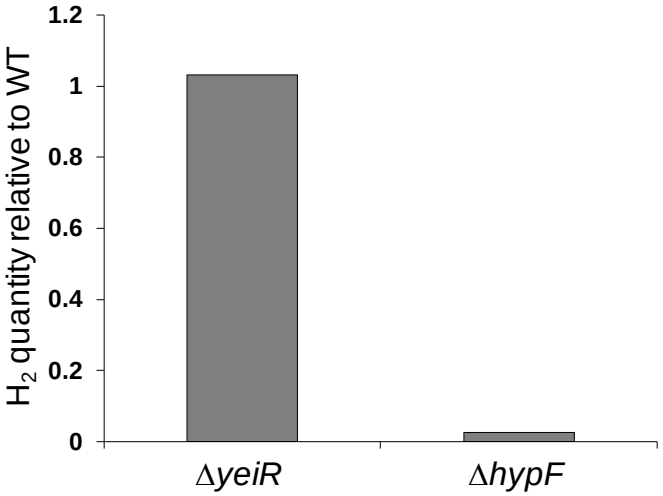


Figure 4-19. Native two-dimensional separation analyzed by ICP-MS for zinc and nickel. Fractions 16-25 for each NaCl separation are shown for zinc and nickel analysis. A protein peak (outlined in dark blue) that contains zinc and nicekl at a ratio of 8:1 is seen in the Δ*yeiR* mutant but not in the WT strain extract.

A



B

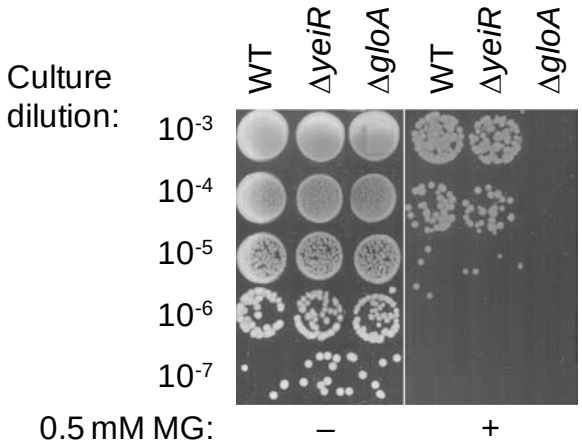


Figure 4-20. Assay of hydrogenase and glyoxalase I activity. A) The concentration of H_2 produced by overnight cultures of the WT, Δ*yeiR*, and Δ*hypF* strains was measured by gas chromatography. The concentration produced by the Δ*yeiR* and Δ*hypF* mutants relative to the WT strain is shown. B) Glyoxalase I activity was assayed by measuring sensitivity of the WT, Δ*yeiR* and Δ*gloA* strains on M9 minimal medium (1.5% agar) to the presence of 0.5 mM mehtylglyoxal (MG). Overnight cultures of each strain grown in LB were normalized to an OD of 1.0, serially diluted and 10µl of the appropriate dilution were plated as shown.

CHAPTER 5
PARALOGS OF ZINC-DEPENDENT PROTEINS

Background

Numerous enzymes have evolved to require a cofactor for activity. These non-protein constituents such as metal ions can endow a protein with catalytic activity or enable otherwise thermodynamically unfavorable structural protein folds (Fischer *et al.*, 2010, Andreini *et al.*, 2008b). Cofactors are essential for the activity of the user enzyme. If that enzyme is required for an essential process, then the host organism has an absolute requirement for the cofactor. In some cases, these cofactors must be imported from the environment, and as such may or may not be readily available. To overcome dependence on a particular cofactor, several isofunctional forms of cofactor-dependent enzymes have evolved (Galperin *et al.*, 1998a, Galperin *et al.*, 1998b, Matthews *et al.*, 2003, Graham *et al.*, 2009, Sankaran *et al.*, 2009, Macauley *et al.*, 2009).

For instance, four classes of superoxide dismutase, which catalyzes the conversion of superoxide radical to oxygen and hydrogen peroxide, have been discovered. Each class utilizes a different catalytic metal cofactor: copper, iron, manganese or nickel (Abreu and Cabelli, 2010). By encoding two or more of these SOD classes, an organism can ensure that SOD activity is present even if a particular essential metal cofactor is not available. Several *Streptomyces* spp. genomes encode both a nickel- and iron-containing SOD isozyme (Youn *et al.*, 1996). Therefore, when one of those metal ions is unavailable, free radical detoxification can still occur with use of the other metal. A further optimization of flexible cofactor requirement is differential regulation of the distinct isofunctional genes by those cofactors (Rodionov, 2007). In the case of *Streptomyces* spp., when nickel is present in the growth medium, expression of

the gene encoding iron-SOD is repressed and expression of the gene encoding nickel-SOD is induced (Kim *et al.*, 1998, Kim *et al.*, 2000).

In some cases, distinct isoforms have evolved that do not require any cofactor. Some organisms contain a methionine synthase that is dependent on the cofactor B_{12}, while in other organisms this enzyme is independent of B_{12} (Gophna *et al.*, 2005). Some organisms encode both isoforms (Gophna *et al.*, 2005). In some of those cases, the gene encoding the B_{12}-independent isozyme is regulated by a B_{12} riboswitch and is expressed when B_{12} is not available (Rodionov *et al.*, 2003).

A similar regulatory strategy has been described for zinc availability. Zinc-independent proteins are negatively regulated by the transcription factor Zur and expressed under zinc-limiting conditions to replace zinc-dependent proteins. The best characterized examples of differential regulation of zinc-dependent and –independent isofunctional proteins are the C^+ and C^- ribosomal protein paralogs (Makarova *et al.*, 2001, Panina *et al.*, 2003). In contrast to the main copies that contain Cys4 Zn-ribbon motifs (and are thus called C^+), C^- ribosomal protein duplicates lack the key Cys residues, do not bind zinc, and are repressed by Zur. When zinc is scarce, these C^- paralogs are expressed and substitute for the C^+ proteins in ribosomes (Natori *et al.*, 2007, Gabriel and Helmann, 2009). This mechanism is proposed to increase cell survival in zinc-limiting growth conditions by supplying functional copies of zinc-free proteins for the newly made ribosomes. At the same time, a pool of zinc is liberated through dissociation from the existing ribosomes and subsequent degradation of the C^+ proteins.

In addition to the ribosomal paralogs, the use of a back-up GTP cyclohydrolase I (GCYH-IA, *folE*) was also discovered as part of the Zur-mediated response to zinc-depletion (Sankaran *et al.*, 2009). In this case, the main constitutively expressed GCYH-IA, which catalyzes the first step in *de novo* folate biosynthesis, is an essential zinc-dependent protein (Auerbach *et al.*, 2000). When Zur repression is relieved, a GCYH-IB (*folE2*), that does not require zinc as a cofactor, is expressed (Sankaran *et al.*, 2009). Only remote sequence similarity is found between FolE and FolE2. This back-up strategy appears to ensure that folate biosynthesis is functional even with decreased activity of the main GCYH-IA due to zinc-depletion.

Genome context analysis revealed that a significant proportion of the Zur-regulated COG0523 genes were located within chromosomal gene clusters that contain genes encoding uncharacterized paralogs of various zinc-dependent proteins (Chapter 3 and Figure 5-1). It remains to be determined whether these paralogs, whose genes are positioned downstream from putative Zur binding-sites, functionally replace their zinc-dependent counterparts when the activity of the latter is adversely affected by the absence of zinc. Alternatively, these proteins may play other roles in the cell that are distinct from those of their paralogs. Of these paralogs, the comparative genomic analysis presented in Chapter 3 suggested functional coupling between COG0523 and the DksA paralog. This prediction was strengthened by the physically clustering of these two genes in the absence of putative co-regulation through Zur.

DksA was initially identified in *E. coli* (EC) as a suppressor of the *dnaK* phenotype (Kang and Craig, 1990). Since then, DksA was shown to act synergistically with (p)ppGpp to control the bacterial response to stress and starvation (also known as the

stringent response) (Paul *et al.*, 2004, Potrykus and Cashel, 2008). The crystal structure of EC DksA revealed that it belongs to a class of bacterial transcription factors (which includes GreA, GreB and GfhI) that bind within the RNA polymerase (RNAP) secondary channel near the active site located at the base of this channel (Perederina *et al.*, 2004).

The stringent response enables rapid and global change of gene expression following nutrient stress, which leads to a rapid increase in ppGpp levels (Magnusson *et al.*, 2005, Srivatsan and Wang, 2008). DksA/ppGpp strongly inhibits transcription of rRNA genes while activating genes for amino acid biosynthesis and transport. Both effects utilize the main activity of DksA: destabilization of open promoter complexes through interaction with RNAP. At the *rrn* promoters, open complexes are very unstable (Gaal *et al.*, 1997), and further destabilization essentially abolishes transcription of rRNA genes (Rutherford *et al.*, 2009). Conversely, RNAP forms very stable complexes at amino acid promoters such as *hisG* (Paul *et al.*, 2005). DksA and ppGpp destabilize these complexes and increase transcription *in vitro* (Paul *et al.*, 2005). *In vivo*, a part of the control could be through liberating RNAP from *rrnB* promoters that account for 70% of the total RNA synthesis in rapidly growing cells (Zhou and Jin, 1998). The end result of this dual control is the restored balance between ribosome production and available amino acid pools. Interestingly, ppGpp and DksA may play independent, or even opposing, roles at some *E. coli* promoters (Magnusson *et al.*, 2007, Aberg *et al.*, 2008, Dalebroux *et al.*, 2010) and during replication (Tehranchi *et al.*, 2010, Trautinger *et al.*, 2005).

DksA proteins characterized to date contain a canonical Cys4 Zn-finger motif. Structural analysis of the EC DksA suggests that this motif plays a key role by

maintaining the fold of the globular domain and its orientation relative to the “catalytic” coiled coil (Perederina *et al.*, 2004). The zinc ion is chelated by two cysteines from each domain and cannot be mobilized after extensive dialysis in the presence of chelators; mutation of this motif renders DksA nonfunctional (Paul *et al.*, 2004, Perron *et al.*, 2005).

In this chapter, an initial phylogenetic and sequence analysis of zinc-dependent paralogs found in the putative Zur regulons of bacterial genomes (Novichkov *et al.*, 2010) is presented with a focus on the putatively Zur-regulated paralog DksA. The DksA paralog of *P. aeruginosa* was chosen for experimental validation of the hypothesis that these DksA paralogs are zinc-independent functional paralogs of the main DksA proteins and are specifically expressed during zinc-depletion. *P. aeruginosa* was chosen to test this hypothesis due to the simple reason that the deletion of the *dksA* gene was previously shown to have a robust growth defect (Jude *et al.*, 2003) and regulon predictions (Novichkov *et al.*, 2010) suggested that the uncharacterized *dksA* paralog should be repressed by zinc through Zur. Although the DksA protein from *E. coli* is by far the best characterized, the *E. coli* genome does not encode the DksA paralog. This observation, however, was useful for an initial screen of the DksA paralog’s activity before performing experiments directly in *P. aeruginosa.*

Results

Putatively Zur-Regulated Paralogs with Conserved Zinc-Binding Residues

An analysis of putative Zur regulons in Cyanobacteria, and γ- and β-proteobacteria (Novichkov *et al.*, 2010) revealed ten families of genes annotated as zinc-dependent enzymes. These zinc-dependent enzymes include N-acetylmuramoyl-L-alanine amidase (AmiA), phosphoribosyl-AMP cyclohydrolase (HisI), dihydroorotase (PyrC), γ-class

carbonic anhydrase (Cam), porphobilinogen synthase (HemB), cysteinyl-tRNA synthetase (CysRS), threonyl-tRNA synthetase (ThrRS), 6-carboxy-5,6,7,8-tetrahydropterin synthase (QueD), GTP cychlohydrolase Type IA (FolE) and the C4-type zinc finger regulator DksA. These putatively Zur-regulated genes specifically occur in genomes that also encode a homologous gene and are most likely part of a zinc-deficiency response mechanism similar to the Zur-regulated ribosomal paralogs or FolE2.

Unlike the Zur-regulated ribosomal paralogs, however, the zinc-binding residues in the PyrC, QueD, AmiA, HisI, FolE, CysRS and ThrRS sequences appear to be conserved (Appendix D). The f*olE* paralog present in the putative Zur regulons of cyanobacterial genomes is distinct from the Zur-regulated *folE2* genes found in Fimicutes and Proteobacteria. While FolE and FolE2 share only remote sequence similarity, the characterized FolE proteins and the cyanobacterial FolE paralogs share significant sequence similarity (Appendix D). Results from the bioinformatic analysis of the PyrC and QueD paralogs is given in more detail below, as the metal-binding residues are non-canonical in the PyrC paralogs and the QueD paralogs have a variant catalytic motif.

PyrC

Dihydroorotase (DHOase) encoded by the *pyrC* gene catalyzes the third step in pyrimidine biosynthesis, a reversible cyclization of carbomyl aspartate to form dihydroorotate (Lieberman and Kornberg, 1953). Characterized DHOases are zinc-dependent proteins (Sander *et al.*, 1965, Taylor *et al.*, 1976, Ogawa and Shimizu, 1995, Williams *et al.*, 1995, Thoden *et al.*, 2001). Two classes of DHOases have been described in the literature: type I and type II. The type II class is thought to have evolved

more recently since members are relatively similar compared to the type I class whose members are widely divergent from one another (Fields *et al.*, 1999). Type I DHOases are found in multienzyme complexes (carbamyl phosphate synthetase-dihydroorotase-aspartate transcarbamylase (CAD) is a multisubunit enzyme that catalyzes the first three steps in pyrimidine biosynthesis) (Simmer *et al.*, 1990) or as uncomplexed enzymes (Porter *et al.*, 2004). Type II DHOases are monofunctional enzymes (Washabaugh and Collins, 1984).

Alcanivorax borkumensis, *Burkholderia cepacia*, *C. metallidurans*, *Hahella chejuensis*, *Vibrio alginolyticus*, *Acinetobacter baylyi*, *P. aeruginosa*, *P. fluorescens*, and *P. putida* encode one type II DHOase and two type I DHOase proteins (Figure 5-2A). The type II DHOases found in these genomes are thought to serve as the main DHOase and are represented by the characterized DHOase from *E. coli* (Lee *et al.*, 2005). One of the type I DHOases encoded in these genomes are homologous to the PyrC' from *Pseudomonas* spp., which is a non-catalytic structural subunit of the aspartate transcarbamoylase enzyme (Schurr *et al.*, 1995). The second type I DHOase encoded by these genomes is referred to here as PyrC2 and the corresponding gene is found downstream of putative Zur binding (Novichkov *et al.*, 2010). The PyrC2 proteins are currently uncharacterized, except for the PyrC2 from *P. aeruginosa,* which was shown to have DHOase activity, but the authors could not explain the redundancy between encoding the type II DHOase and PyrC2 (Brichta *et al.*, 2004).

Phylogenetic analysis of the PyrC2 proteins reveals that they share significant sequence similarity with the PyrC from *Porphyromonas gingivalis* for which a crystal structure is available (Figure 5-2A). Interestingly, the *P. gingivalis* PyrC was shown to

have a binuclear zinc site and these zinc-binding residues are conserved in the PyrC2 proteins whose corresponding genes are found in Zur regulons (Figures 5-2B and C). The zinc-binding residues found in the *P. gingivalis* PyrC crystal structure are mostly identical to the zinc-binding residues found in the *E. coli* PyrC crystal structure; including a metal-bridging lysine that is carboxylated in both cases. However, there is a slight deviation in the metal-binding residues of these two proteins: one of the histidines bound to the alpha-zinc ion in the *E. coli* ortholog is a glutamine in the *P. gingivalis* ortholog (Figure 5-2B). This difference suggests that a metal other than zinc would be bound under native conditions; the borderline Lewis acid zinc does not form as strong a bond with a hard Lewis base such as glutamine compared with a borderline Lewis base such as histidine (Gurd and Wilcox, 1956, Lesburg *et al.*, 1997). Which begs the question: why would the cell express a DHOase under zinc-deficient conditions that has a lower affinity for zinc?

QueD

Similar to the PyrC paralogs, a phylogenetic reconstruction of the QueD protein family reveals two main lineages, which are referred to here as QueD and QueD2 (Figure 5-3A). One of the QueD proteins in *Azotobacter vinelandii*, *B. cepacia*, and *C. metallidurans* is a QueD, and the other protein is a QueD2 (Figure 5-3A). The *queD2* genes in these genomes are putatively regulated by Zur. Proteins from both lineages have been shown to be involved in queuosine biosynthesis (Reader *et al.*, 2004, McCarty *et al.*, 2009). A crystal structure representing a QueD2 protein is not available, but, based on sequence analysis, the zinc-binding residues found in the *P. aeruginosa* QueD crystal structure (PDB:2OBA) are conserved in the QueD2 protein sequences (Figure 5-3B).

However, the position of the putative active site cysteine is not conserved in the QueD2 proteins. Instead of three amino acids away from the zinc-binding HGH motif as found in QueD, the cysteine is four amino acids away in the QueD2 proteins (Figure 5-3B). QueD and QueD2 are homologs of 6-pyruvoyl tetrahydropterin synthase (PTPS), an enzyme involved in tetrahydrobiopterin biosynthesis (McCarty *et al.*, 2009). For PTPS, the zinc ion is proposed to coordinate the substrate for proton abstraction by the cysteine residue's thiolate moiety (Bürgisser *et al.*, 1995). Perhaps, the novel catalytic motif of QueD2 proteins is indicative of a non-zinc metal cofactor, which correctly positions the substrate for catalysis by the displaced cysteine. Indeed, in the crystal structure of human PTPS (PDB: 3I2B), a nickel ion in found bound to the corresponding metal-ligands (Figure 5-3C).

Zur-Regulated Paralogs without the Canonical Zinc-Binding Residues

Cam

Carbonic anhydrase catalyzes the reversible hydration of CO_2 (Meldrum and Roughton, 1933). Until recently, all carbonic anhydrase classes were assumed to use zinc as the catalytic cofactor (Christianson and Fierke, 1996). However, plasticity of the metal site has become apparent (Yee and Morel, 1996, Lane and Morel, 2000, Lane *et al.*, 2005). In particular, the γ -class of carbonic anhydrases appears to use iron as the *in vivo* catalytic cofactor (Tripp *et al.*, 2004, Macauley *et al.*, 2009). The carbonic anhydrase genes present in the computationally reconstructed Zur regulons of γ- and β-proteobacteria (Novichkov *et al.*, 2010) share significant sequence similarity with the prototypical γ -class carbonic anhydrase from *M. thermophila* (Appendix D, Figure D-6).

HemB

HemB catalyzes the first common step in the biosynthesis of tetrapyrroles (Nandi and Shemin, 1968). The existence of zinc binding and non-zinc binding variants of HemB is documented in the literature (Jaffe, 2003). As expected, the HemB paralog putatively regulated by Zur in the genome of *P. putida* has highest sequence similarity to the non-zinc HemB isoforms and accordingly the zinc-binding residues are not conserved (Apendix D, Figure D-7).

DksA

DksA proteins are RNAP-binding factors that affect the interaction between RNAP and target promoters causing changes in gene expression. Proteins belonging to the DksA/TraR superfamily are present throughout the bacterial kingdom (Marchler-Bauer *et al.*, 2009) and the majority of these proteins are of unknown function. For instance, in addition to the characterized DksA protein (PA4723), which has been shown to be involved in the stringent response, the *P. aeruginosa* genome (Stover *et al.*, 2000) encodes four other proteins from this superfamily (Figure 5-4A). Three of these DksA-like proteins (PA4577, PA4870, and PA0612) contain the characteristic CXXC-(X_{17})-CXXC Zn-finger motif but otherwise have low sequence homology to DksA (24%, 16%, and 14% identity, respectively). The fifth DksA-like protein (PA5536), which is referred to here as DksA2, has significant sequence homology to the classical DksA (34% identity) but contains a CXXT-(X17)-CXXA motif instead of the typical Cys4 Zn-finger motif (Figure 5-5A). DksA2 proteins are found in several genomes belonging to proteobacterial species and can be identified based on sequence similarity to the DksA protein of *E. coli* and the presence of a variant Cys4 Zn-finger motif: CXX[S/T]-(X_{17})-[C/S/T]XXA.

Some genomes such as *P. aeruginosa* contain both a *dksA* and a *dksA2* gene. In these situations, *dksA2* genes are found downstream of putative Zur-binding sites (*Serratia marcescens, Klebsiella pneumoniae, H. chejuensis, Pseudomonas* spp., and *Methylobacillus flagellatus*) (Novichkov *et al.*, 2010) (Figure 5-4). As such, *dksA2* is often clustered physically on the chromosome with factors known to be involved in the response to zinc-depletion, such as *znuABC* (Figure 5-1).

Complementation of the *dksA* deletion of *E. coli* by the *dksA2* of *P. aeruginosa*

The *P. aeruginosa* DksA protein (PA DksA; PA4723) has been the focus of several studies examining its roles in the quorum-sensing circuitry, rRNA transcription and survival during antibiotic stress (Branny *et al.*, 2001, Jude *et al.*, 2003, Perron *et al.*, 2005, Viducic *et al.*, 2006). DksA2 (PA5536), on the other hand, is annotated as a conserved hypothetical protein (*Pseudomonas* genome database (Winsor *et al.*, 2009)). To test the prediction that DksA2 can functionally replace the canonical DksA, complementation experiments with *E. coli* were used initially. Unlike *P. aeruginosa,* the *E. coli* genome does not contain a *dksA2* gene.

E. coli Δ*dksA* is unable to grow on minimal media lacking leucine, valine, glycine, phenylalanine or threonine (Brown *et al.*, 2002). As shown in Figure 5-6A, *dksA2* expressed *in trans* from P_{BAD} was able to complement the *E. coli dksA* gene deletion, suggesting that *dksA2* may have a similar function to *dksA*. Compared to EC *dksA* or PA *dksA,* where uninduced expression from P_{BAD} was sufficient for complementation, rescue by *dksA2* required a higher concentration of the inducer arabinose (0.002 - 0.2%). As previously shown for EC *dksA* (Potrykus *et al.*, 2006), overexpression of PA

dksA was toxic in the presence of 0.2% arabinose. Toxicity was not observed when expressing *dksA2* with the arabinose concentrations used in this study.

The zinc ion found in the crystal structure of *E. coli* DksA appears to orient the CT helix relative to the coiled-coil domain (Figure 5-5B). The coiled-coil is proposed to insert into the secondary channel of RNAP and the CT helix is proposed to hug the outside of RNAP, positioning DksA (Perederina *et al.*, 2004). In the Cys4 Zn-finger motif of DksA2, two of the four cysteines are conserved (Figure 5-5A). Degeneracy of this motif could be an indication that only two of the four cysteines in the Zn-finger motif of DksA are necessary to maintain function. To test if the variant Zn-finger motif found in DksA2 is adequate for maintaining the function of DksA, the conserved Cys114 and Cys135 in PA DksA were mutated to the corresponding residues found in DksA2 (Thr and Ala, respectively). Both substitutions eliminated the ability of PA DksA to complement the *E. coli* dksA *gene* deletion (Figure 5-6B). Expression of these mutant PA *dksA* genes was not found to be more toxic than expression of the wild-type PA *dksA* gene (Figure 5-6C). Toxicity was relieved to a certain extent by both mutations. This result is consistent with a key role of a complete Zn-finger motif in the function of DksA.

Complementation of a *P. aeruginosa dksA* mutant by *dksA2*

Phenotypic studies were then performed directly in *P. aeruginosa* PAO1. Like in *E. coli*, deletion of the *dksA* gene in *P. aeruginosa* leads to a growth defect in M9 minimal media with glucose (0.2% w/v) as a sole carbon source (Jude *et al.*, 2003), an effect which was reproduced in M9 minimal medium with glycerol as a sole carbon source (Figure 5-7). Similarly to the situation observed in a heterologous *E. coli* host, *dksA2* expressed from P_{BAD} (of pBAD24) rescued the growth defect of the *P. aeruginosa*

Δ*dksA* strain (Figure 5-7), strongly suggesting that *dksA2* can functionally replace PA *dksA*.

The growth defect of the Δ*dksA* strain could be suppressed by the addition of 100 μM EDTA or by combining the *dksA* deletion with the deletion of the gene encoding the Zur homolog (PA5499, *np20*) (Figure 5-8). In both cases, suppression was dependent on the presence of *dksA2 in cis* (on the chromosome) (Figures 5-8A and B) or *in trans* (expressed from P_{BAD} of pHERD20T) (Figure 5-9). The addition of zinc, but not other transition metals tested, counteracted the suppression effect of EDTA (Figures 5-8A and D). By contrast, zinc did not affect growth of the Δ*dksA* Δ*zur* strain (Figure 5-8A).

During these experiments, the observation was made that pyocyanin was differentially produced in the various strains and that these trends mimicked the growth defects observed above (Figure 5-10A). Pyocyanin is a secreted virulence factor that is thought to play a role in the tissue damage of infected hosts (Caldwell *et al.*, 2009). The synthesis of this metabolite is regulated by quorum sensing, and PA DksA was initially characterized in a complementation screen of a quorum-sensing mutant (Branny *et al.*, 2001). During growth in LB at 37^{o}C, the Δ*dksA* strain produced less than 10% of the pyocyanin produced by the parent strain (Figure 5-10B). Pyocyanin production was restored by expressing *dksA2 in trans* (Figure 5-10B).

***dksA2* is regulated by zinc through Zur**

Suppression of the *P. aeruginosa* Δ*dksA* growth defect by EDTA or deletion of *zur* together with the promoter region organization of the *dksA2* gene (Figure 5-4B), suggest that *dksA2* is regulated by Zur in *P. aeruginosa*, and that its expression may be induced under zinc limitation. Therefore, an analysis of the effect of EDTA and

extracellular zinc on *dksA* and *dksA2* transcript levels by qRT-PCR and on DksA and DksA2 protein levels by Western blotting was performed. As shown in Figure 5-11A, the abundance of *dksA* transcript was not significantly affected by either EDTA or zinc and was found to gradually decrease throughout the growth cycle as previously described (Perron *et al.*, 2005). In contrast, as shown in Figure 5-11B, *dksA2* was not significantly expressed in the WT strain until late logarithmic/early stationary phase when grown in M9 medium, the same growth condition that gave rise to the Δ*dksA* strain growth defect described above. In the presence of EDTA, which was able to suppress the Δ*dksA* strain phenotype, the level of *dksA2* transcript was increased during early/mid logarithmic phase (Figure 5-11B). The *dksA2* transcript was undetectable throughout the growth cycle when the cells were grown in the presence of 25 μM $ZnSO_4$ (Figure 5-11B). In the absence of Zur, zinc failed to repress *dksA2* expression compared to the strain grown in M9 medium without supplementation (Figure 5-12A). The abundance of DksA and DksA2 (as discerned from the Western blot) paralleled changes in the transcript levels (Figures 11-C and 12-C).

As expected from the expression results above, His_6-Zur was found to bind specifically to the upstream region of *dksA2* in the presence of zinc (Figure 5-13). In the absence of zinc, binding was observed but only in the presence of higher concentrations of purified His_6-Zur (Figure 5-13C), while in the presence of 100 μM EDTA, binding was abolished (Figure 5-13C). Competition with the *dksA2* upstream region DNA was achieved with specific competitor (annealed oligonucleotides (42 bp) containing the putative binding site) but not with non-specific competitor (control oligonucleotides (42 bp) which lack a consensus Zur-binding site) (Figure 5-13D). Zur

(after cleavage of the tag) had the same affinity for the *dksA2* upstream DNA (data not shown).

Deletion of *dksA2* results in a growth defect in the presence of metal chelators

DksA2 may be a condition-specific functional variant of PA DksA, and this analysis suggested that *dksA2* is expressed during zinc-depletion. Under zinc-deficient growth conditions, a larger proportion of DksA proteins may be in the apo-form (metal-free) and therefore inactive. Under these conditions, growth would then be dependent on DksA2. To test this hypothesis, *dksA2* was deleted in the WT background. Growth of the mutant was assayed in the medium that results in the Δ*dksA* growth defect but with the addition of metal chelators.

The absence of *dksA2* resulted in an observable growth defect in the presence of EDTA or TPEN, two chelators frequently used to mimic zinc limitation (Figure 5-14). To confirm that this phenotype was due specifically to zinc-depletion (as opposed to that of another metal), *znuA* was deleted in the Δ*dksA2* strain background*; znuA* encodes a homolog of the periplasmic chaperone component of the high-affinity zinc transporter ZnuABC, and its deletion impairs zinc uptake (Patzer and Hantke, 1998). Growth of WT, Δ*dksA2*, Δ*znuA*::Gm^R, and Δ*dksA2* Δ*znuA*::Gm^R strains on minimal medium were compared in the presence of various concentrations of EDTA and TPEN (Figure 5-14). Deletion of *znuA* exacerbated the growth defect due to the deletion of *dksA2* in the presence of EDTA or TPEN. The growth defect was rescued by expressing *dksA2 in trans* from P_{BAD} of pHERD20T in the presence of 0.2% arabinose (Figure 5-14). As further confirmation that the EDTA-sensitive phenotype of the *dksA2* mutant was due to depletion of zinc, 25 μM of various transition metals were added to the medium

containing 1.25 mM EDTA. Only the addition of zinc was able to fully suppress the growth defect caused by the presence of EDTA (Figure 5-14C).

Discussion and Conclusions

Back-up Proteins

Zinc-depletion can have detrimental effects on bacterial cell viability and can impede an organism's ability to infect a vertebrate host. *P. aeruginosa* is a significant human opportunistic pathogen and a major cause of mortality among cystic fibrosis patients (Govan and Deretic, 1996). As part of the acute immune response, the host actively sequesters zinc, limiting its availability to the invading pathogen (Lusitani *et al.*, 2003, Liuzzi *et al.*, 2005, Corbin *et al.*, 2008). For the invading pathogen, one common solution is to induce the expression of a high-affinity zinc-transporter and transport available zinc from the environment. Indeed, high-affinity zinc-transporters have been shown to be required for virulence of several pathogens (Dahiya and Stevenson, 2010, Davis *et al.*, 2009, Ammendola *et al.*, 2007, Yang *et al.*, 2006b, Kim *et al.*, 2004, Campoy *et al.*, 2002). Other strategies may include expression of zinc-independent functional copies of key proteins, sometimes coupled to mobilization of protein-bound zinc (Panina *et al.*, 2003, Gabriel and Helmann, 2009, Akanuma *et al.*, 2006, Sankaran *et al.*, 2009), as predicted for the HemB, Cam and DksA paralogs.

An additional strategy, suggested by the presence of the PyrC, QueD, AmiA, HisI, FolE, CysRS and ThrRS paralogs, may be the induction of zinc-dependent proteins when the zinc-concentration within the cell becomes limiting. The purpose of expressing these genes could be to ensure that at least a subpopulation of the corresponding protein pool is metallated from available zinc. Assuming the distribution of zinc in the

cell is in equilibrium, increasing the concentration of one particular zinc-dependent protein may shift the equilibrium in favor of that protein acquiring zinc.

However, for these paralogs, the *in vivo* cofactor may be a metal other than zinc, even though this analysis found the zinc-binding ligands to be conserved. As an example, natively purified CucA and MncA from *Synechcystis* PCC 6803, contain copper and manganese, respectively, even though they have identical metal-binding ligands (Tottey *et al.*, 2008). For PyrC, the presence of a glutamine residue instead of histidine in the PyrC2 metal-binding site also argues that *in vivo* the metal cofactor may not be zinc. The fact that the close ortholog from *P. gingivalis* co-crystallized with a zinc ion may not be significant. When CucA and MncA were allowed to fold under *in vitro* conditions they specifically bound the “wrong” metal (Tottey *et al.*, 2008). Therefore, over-expression in *E. coli*, may have resulted in mis-metallation of the *P. gingivalis* PyrC homolog with zinc. Multiple factors affect protein metallation independent from the metal-binding ligands, such as protein location (Tottey *et al.*, 2008) and chaperones (Nojiri *et al.*, 2000, O'Halloran and Culotta, 2000). Interestingly, these Zur-regulated operons, which are rich in protein paralogs, also contain members from the predicted metal chaperone family COG0523. Therefore, if one or more of these paralogs are dependent on a chaperone for proper metal acquisition, a candidate does exist.

DksA

To further investigate the model that some of these paralogs exist to replace their zinc-dependent counterparts during conditions of zinc-deficiency, a more thorough investigation into the role of the DksA paralog was performed. The results from this study argue that *P. aeruginosa* utilizes the zinc-independent DksA2 protein as a part of

the adaptation to zinc-limited environments, specifically to compensate or replace the zinc-dependent DksA. The results supporting this model are as follows.

DksA2 lacks a conserved Zn-finger motif (Figure 5-5). Thus, DksA2 function should be independent of zinc, in contrast to DksA where it appears to be critical (Figure 5-6B). The growth assays revealed that DksA2 could (at least under the conditions tested) functionally replace DksA (Figures 5-7 and 5-11). Transcript analysis revealed that *dksA2* was expressed during metal depletion and repressed in the presence of zinc (Figure 5-12). This response was mediated by Zur, a previously uncharacterized protein in *P. aeruginosa*; expression of *dksA2* was not repressed by zinc in the absence of *zur* and Zur was found to bind specifically to the upstream region of the *dksA2* gene (Figures 5-13 and 5-14).

The observation that at least two shifts were present in the EMSA experiment (Figure 5-24A) is a common result when working with Fur family regulators (Huang *et al.*, 2008, Li *et al.*, 2009, Kallifidas *et al.*, 2010) and this result has been explained by the tendency of these proteins to form higher-order structures at the DNA-binding site (Bagg and Neilands, 1987, Owen *et al.*, 2007, Ahmad *et al.*, 2009). For instance, the Fur protein tends to polymerize at many operator sites and binds cooperatively at some sites resulting in a helical arrangement that spreads beyond the recognition site (Lecam *et al.*, 1994, Lavrrar *et al.*, 2002). The most recent model for the interaction of Fur family regulators with DNA suggests that the operator site is recognized by two protein homodimers, each contact the DNA on opposite sides of the helix. Depending on the strength of the protein-protein interaction, extended protein arrays form as the dimers

polymerize and bind with reduced stringency to adjacent DNA (Lee and Helmann, 2007).

To further support the model, deletion of *dksA2* was found to impair growth in the presence of the metal chelators, EDTA and TPEN. Under these growth conditions (minimal medium with glycerol as the carbon source and without amino acids) growth was dependent on the presence of *dksA* (Figure 5-18). Therefore, the Δ*dksA2* strain phenotype observed could be a result of an increase in the proportion of unmetallated verses metallated DksA proteins. Since zinc is likely required for the proper folding and function of DksA, the newly synthesized DksA proteins would be inactive in a zinc-depleted environment, leading to defects in gene expression control. Under these conditions, it could be advantageous to express a back-up, zinc-independent DksA variant.

Combined, these results suggest that DksA2 exerts its activity when cells encounter an environment poor in available zinc. In agreement, *dksA2* was previously found to be induced during growth in cystic fibrosis sputum (Palmer *et al.*, 2005). Sputum from cystic fibrosis patients was shown to have high levels of calprotectin (Gray *et al.*, 2008), a neutrophil protein that chelates zinc making it unavailable to pathogens (Clohessy and Golden, 1995), suggesting that the host environment is low in available zinc. DksA2 may be a previously unknown virulence factor involved in the adaptation and survival of *P. aeruginosa* during infection.

These results are consistent with a model where DksA (that requires zinc) can be functionally replaced with its paralog DksA2 (that does not require zinc) when cellular zinc levels are low. However, it is quite possible that DksA2 plays additional roles in

cellular physiology. For example, DksA2 may regulate promoters required for adaptation to various stresses, in essence acting as a specialized version of DksA. The growth defects due to the deletion of *dksA2* observed in the presence of EDTA and TPEN could result from altered regulation at yet unknown promoters.

The discovery of the DksA paralog, DksA2, adds an extra level onto the already complex global regulation of gene expression by RNAP-binding factors. Uniquely, DksA2 brings to light the potential for novel gene regulation during zinc limitation, a condition that triggers an increase in DksA2 levels (Figure 5-12). It is also possible that DksA2 (or other DksA paralogs) may be upregulated in response to other environmental stresses. The DksA family of regulators may be the key players of an elaborate gene expression program designed to integrate diverse environmental cues and balance the cell's growth under a large variety of conditions.

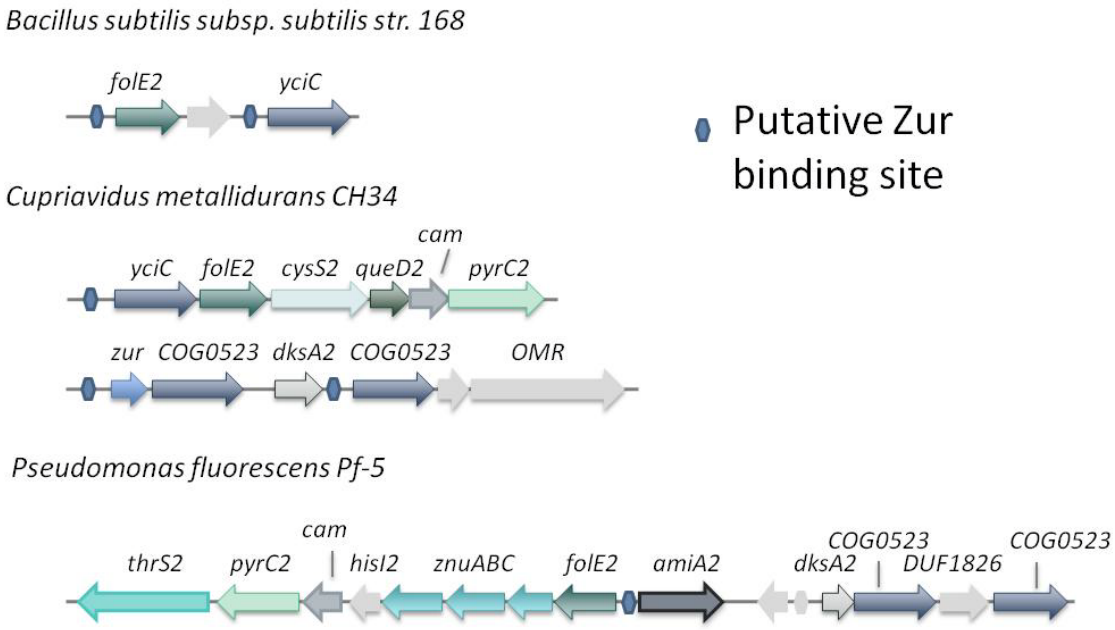


Figure 5-1. Representative gene clusters composed of Zur-regulated COG0523 members. Genes labeled *yciC* represent subfamily 1 COG0523 members. 0523-11 and 0523-8 refer to subfamilies 11 and 8, respectively. Other gene abbreviates: *folE2,* GTP cyclohydrolase IB; cysS2, paralog of cysteinyl-tRNA synthetase; *queD2*, paralog of 6-carboxy-5,6,7,8-tetrahydropterin synthase; *cam*, γ -class carbonic anhydrase; *pyrC2*, paralog of dihydroorotase; *zur*, zinc-responsive transcription factor; *dksA2*, paralog of C4-type zinc finger regulator DksA; *OMR*, putative outer membrane protein; *thrS2*, threonyl-tRNA synthetase; *hisI2*, paralog of phosphoribosyl-AMP cyclohydrolase; *znuABC*, high-affinity zinc-transporter; *amiA2*, amidase; DUF1826, Pfam protein family of unknown function; OMR, outer membrane protein.

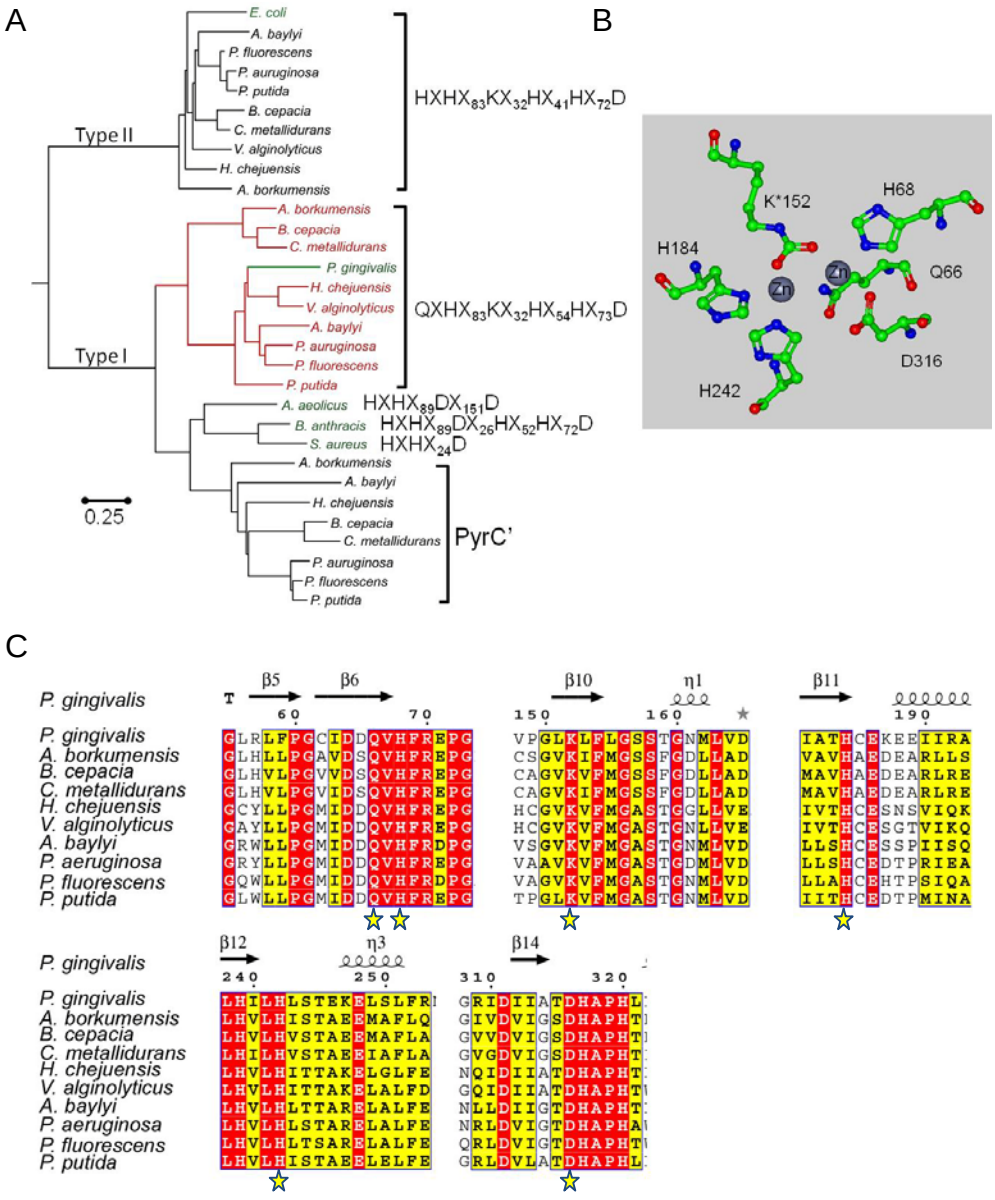


Figure 5-2. Phylogenetic and sequence analysis of the PyrC paralogs. A) Neighbor-joining tree of PyrC protein sequences from *Porphyromonas gingivalis*, *A. borkumensis*, *B. cepacia*, *C. metallidurans*, *H. chejuensis*, *V. alginolyticus*, *A. baylyi*, *P. aeruginosa*, *P. fluorescens*, and *P. putida.* Proteins for which a crystal structure is available are highlighted in green. Red indicates that the corresponding gene is putatively regulated by Zur. Deduced metal-binding motifs are given to the right. B) Binuclear zinc site from the crystal structure of *P. gingivalis* PyrC2. C) Partial protein sequence alignment of selected PyrC2 proteins. The metal-binding residues are marked with a star. Secondary structural elements were derived from the crystal structure of *P. gingivalis* PyrC2 (PDB: 2GWN).

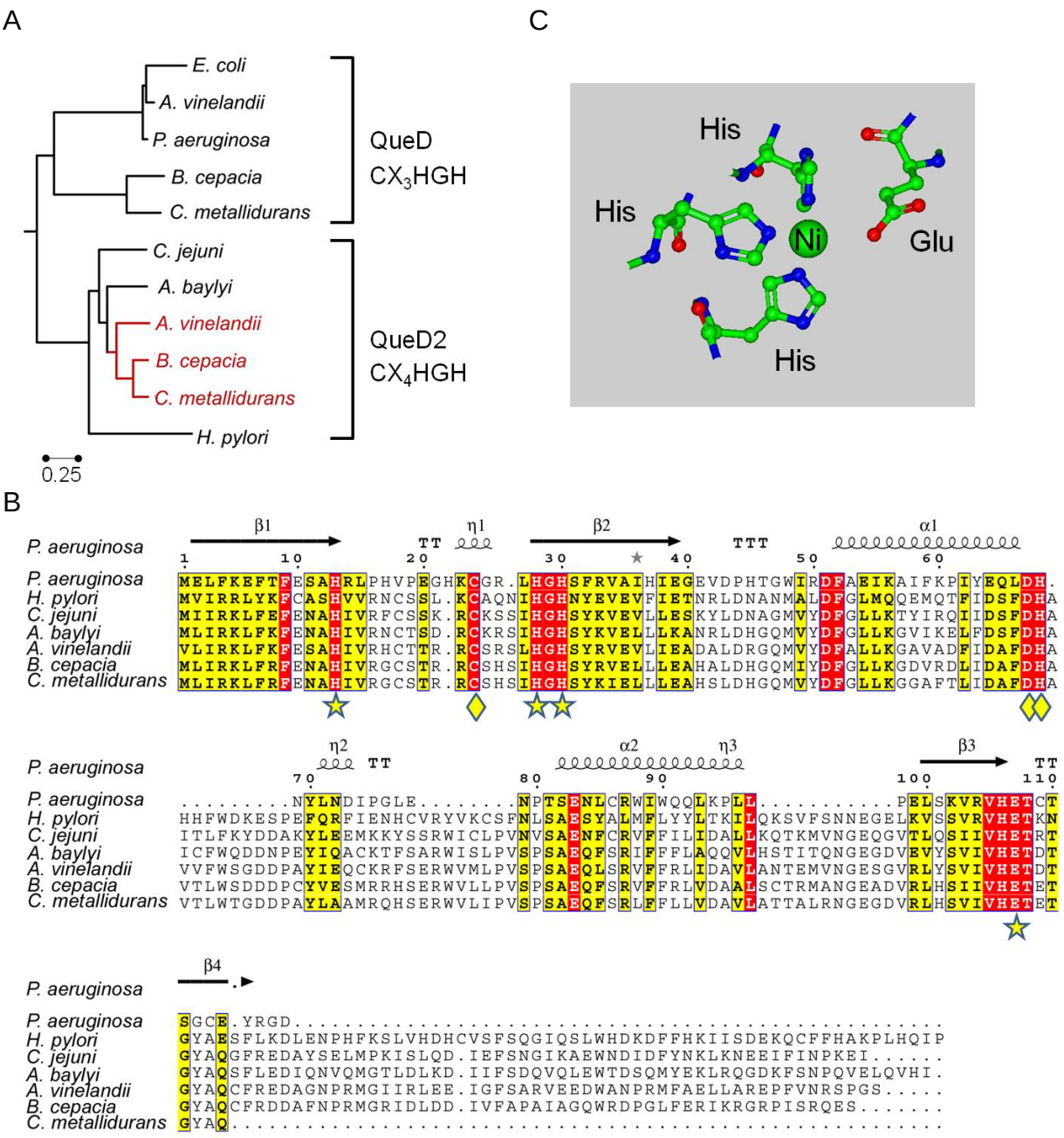


Figure 5-3. Phylogenetic and sequence analysis of the QueD paralogs. A) Neighbor-joining tree of QueD and QueD2 protein sequences. Red indicates that the corresponding gene is putatively regulated by Zur. A partial catalytic site motif is given to the right indicating position of the catalytic cysteine relative to two of the three metal-binding histidines. B) Protein sequence alignment of selected QueD2 proteins with the QueD of *P. aeruginosa*. The metal-binding residues are marked with a star. Catalytic triad residues are marked with a diamond. Secondary structural elements were derived from the crystal structure of *P. aeruginosa* QueD (PDB: 2OBA). C) Nickel site from the crystal structure of human PTPS (PDB: 3I2B).

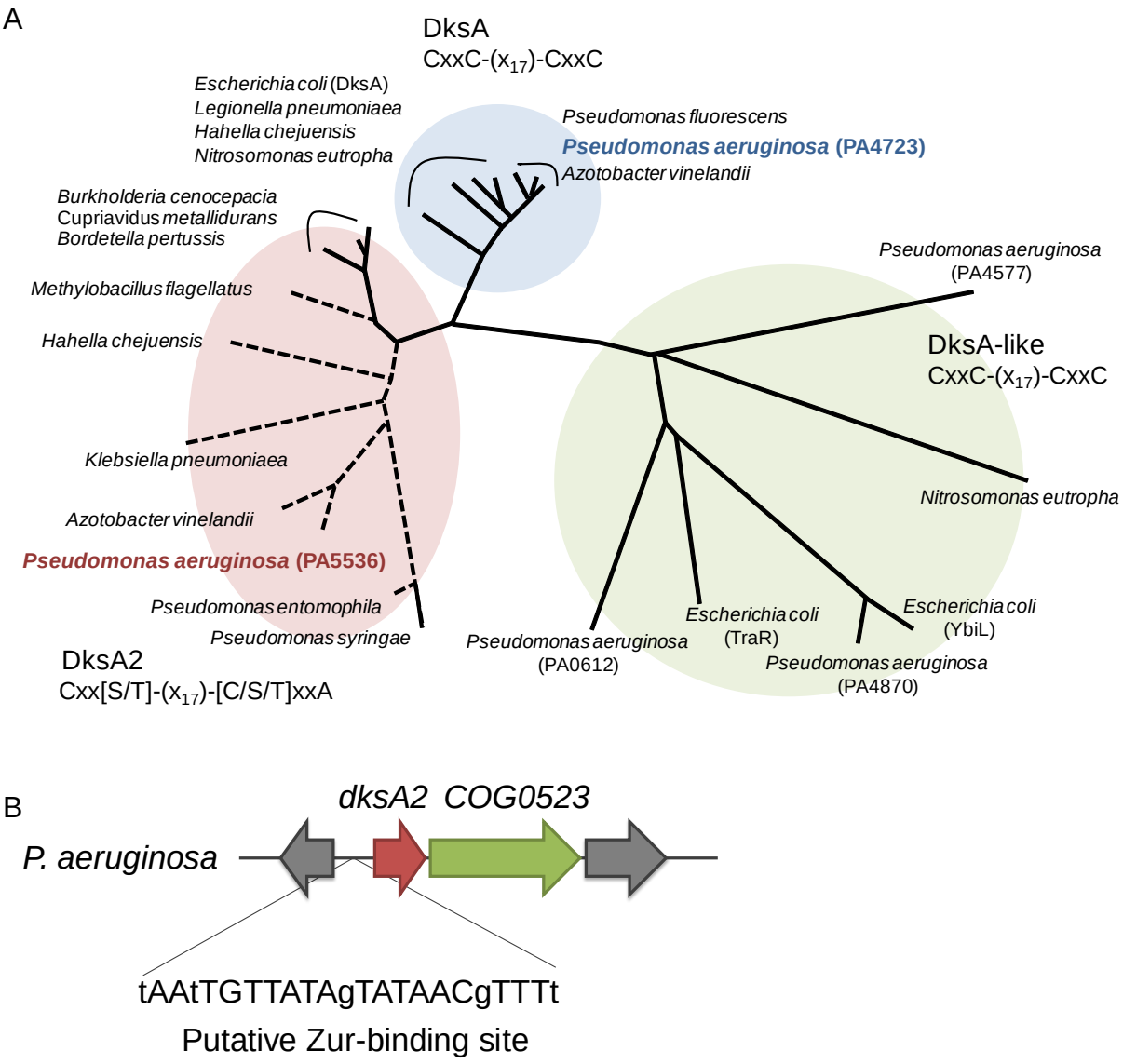
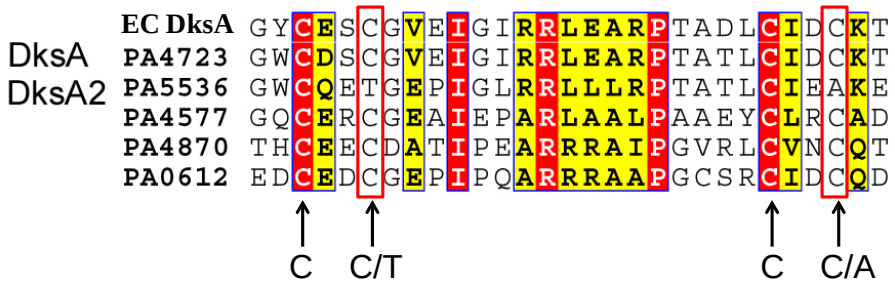


Figure 5-4. Proteins with homology to the DksA protein of *E. coli* are found with and without the Cys4 Zn-finger motif. A) Neighbor-joining tree of DksA homologs. The DksA family contains three main subgroups: DksA, DksA2 and DksA-like. The corresponding cysteine motif for each subgroup is given. Dotted branches designate that the corresponding gene is putatively regulated by Zur. B) Operon organization of the *dksA2* of *P. aeruginosa*. The putative Zur binding site (Haas *et al.*, 2009) is given.

A



B

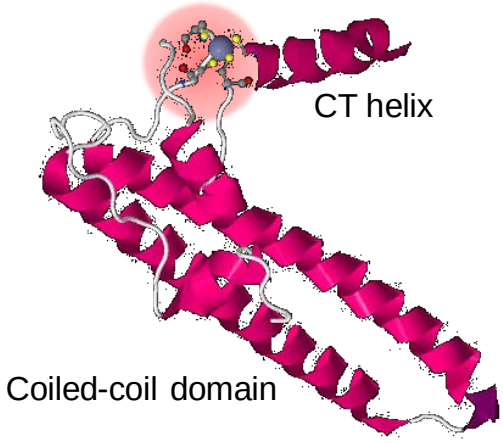


Figure 5-5. Cys4 Zn-finger domain of *E. coli* DksA and homologs from *P. aerginosa*. A) Partial sequence alignment of the DksA homologs from *E. coli* (EC) and *P. aeruginosa* (PA4723/DksA, PA5536/DksA2, PA4577, PA4870 and PA0612). The conserved and variant cysteine residues are marked. B) Cartoon of the *E. coli* DksA crystal structure (PDB: 1TJL). The Zn-finger is highlighted with a red halo and the Zn-binding cysteines are represented with balls and sticks.

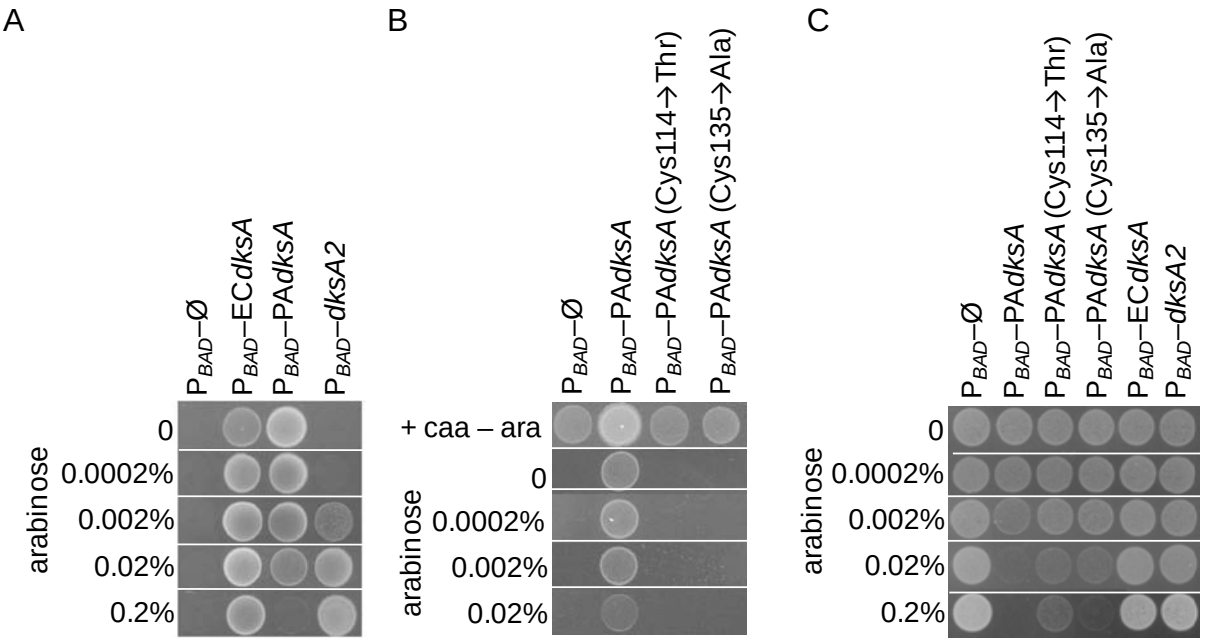


Figure 5-6. Rescue of *E. coli* Δ*dksA*::tet phenotype. Overnight cultures of *E. coli* Δ*dksA*::tet derivatives grown in LB were washed with M9 medium and normalized to an OD of 1.0. 10 µl of a 10^{-3} dilution was plated on M9 medium without or with various concentrations of arabinose. A) Complementation of the *dksA* deletion with *dksA* from *E. coli* or *P. aeruginosa* or *dksA2* from *P. aeruginosa* expressed *in trans* from the P_{BAD} of pBAD24. Strains were grown on M9 medium without casamino acids. B) Complementation of the *dksA* deletion with WT or mutated forms of *dksA* from *P. aeruginosa* expressed *in trans* from P_{BAD} of pBAD24. Strains were grown on M9 medium with (first lane) or without casamino acids (lanes 2-5). C) Expression of the indicated genes *in trans* was tested for toxicity in the WT *E. coli* background. Strains were grown on M9 medium without casamino acids.

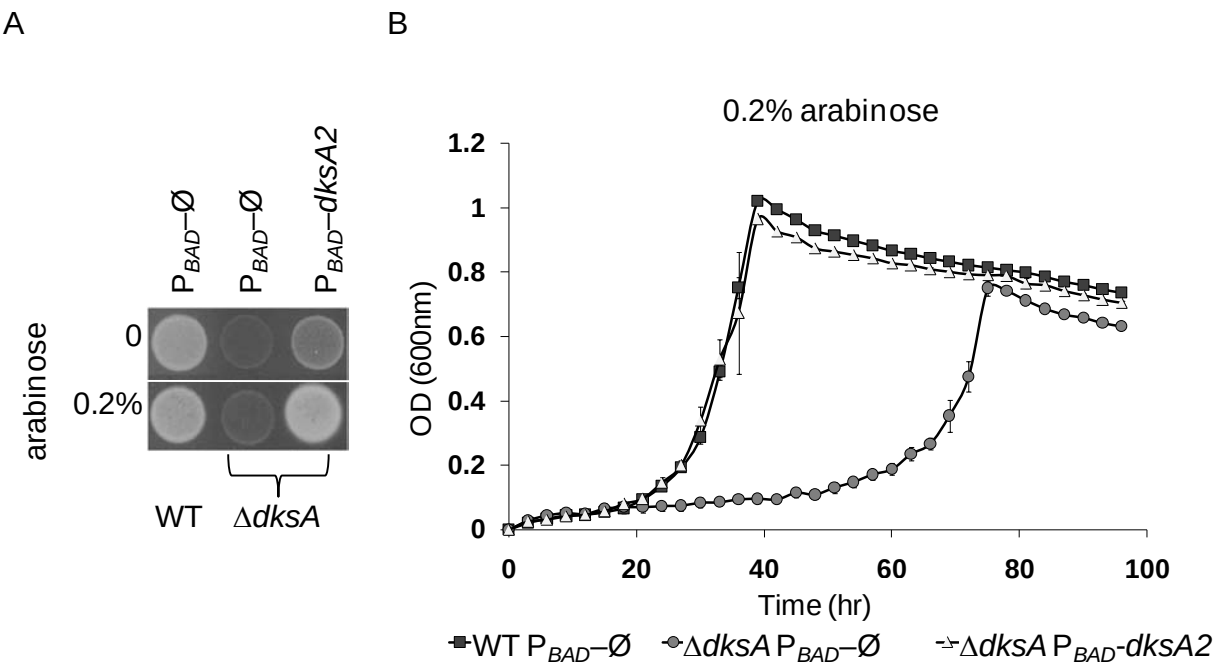


Figure 5-7. Complementation of *P. aeruginosa* Δ*dksA* with *dksA2*, *in trans*. A) Overnight cultures of *P. aeruginosa* (WT) pHERD20T (P_{BAD}–Ø), *P. aeruginosa* Δ*dksA* pHERD20T and *P. aeruginosa* Δ*dksA* pCH115 (P_{BAD}–*dksA2*) grown in LB were washed with M9 medium then normalized to an OD of 1.0. 10 µl of a 10^{-3} dilution was plated on M9 medium without or with 0.2% arabinose. B) Overnight cultures of WT pHERD20T, Δ*dksA* pHERD20T and Δ*dksA* pCH115 grown in LB were washed with M9 medium then normalized to an OD of 1.0. Each culture was then diluted 500-fold into fresh M9 medium plus 0.2% arabinose. Growth curves were monitored with a Bioscreen C. Error bars represent ± the standard deviation of three cultures.

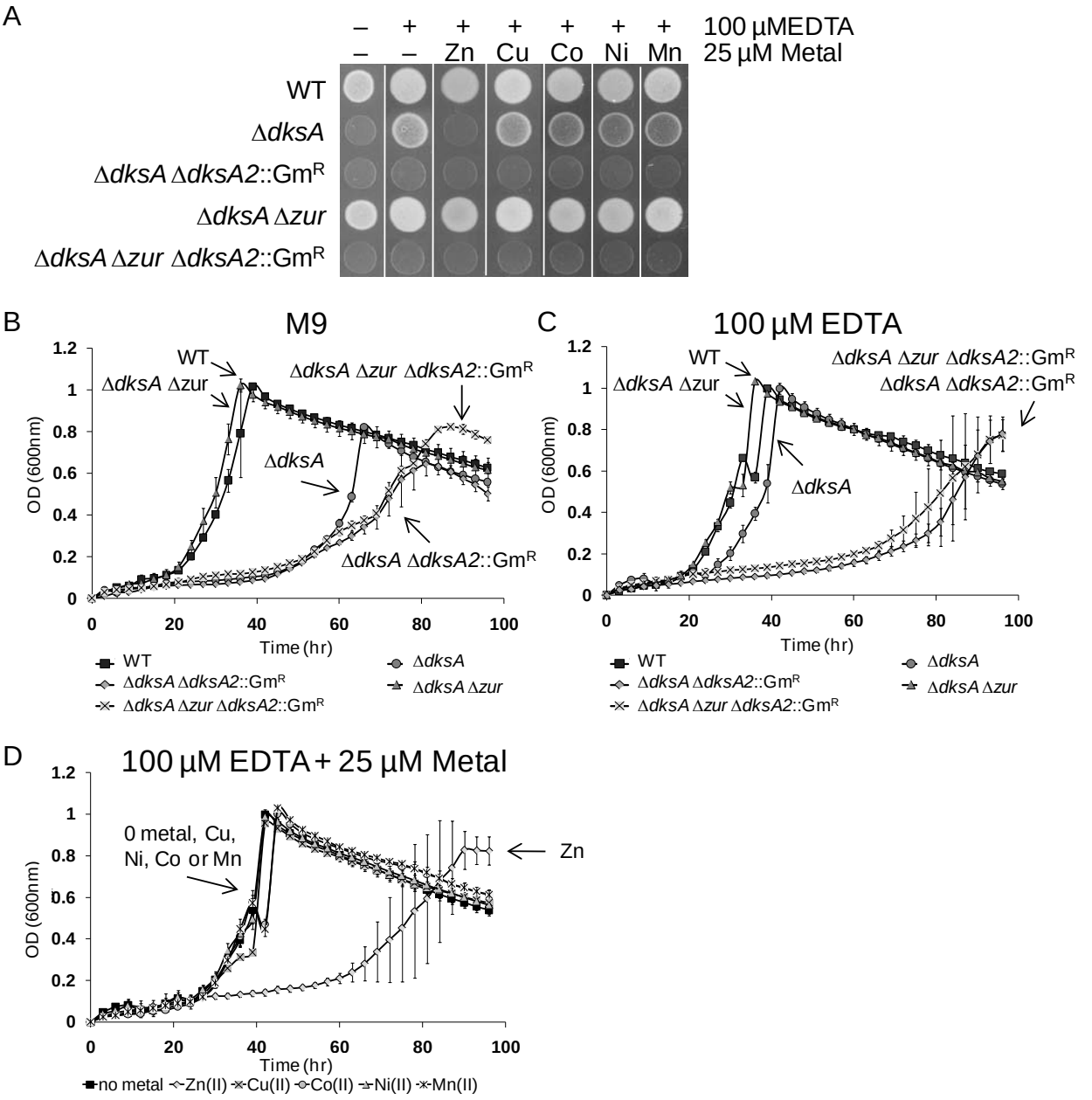


Figure 5-8. Suppression of *P. aeruginosa* Δ*dksA* growth defect. A) Overnight cultures of *P. aeruginosa* derivatives grown in LB were washed with M9 medium then normalized to an OD of 1.0. 10 µl of a 10^{-3} dilution was plated on M9 medium with the indicated supplementation. B-D) Overnight cultures of *P. aeruginosa* derivatives grown in LB were washed with M9 medium then normalized to an OD of 1.0. Each culture was then diluted 500-fold into fresh M9 medium without supplementation (B) or with 100 µM EDTA (C). D) The Δ*dksA* strain was grown in the presence of 100 µM EDTA ± 25 µM zinc, copper, cobalt, nickel or manganese. Growth curves were monitored with a Bioscreen C. Error bars represent ± the standard deviation of three cultures.

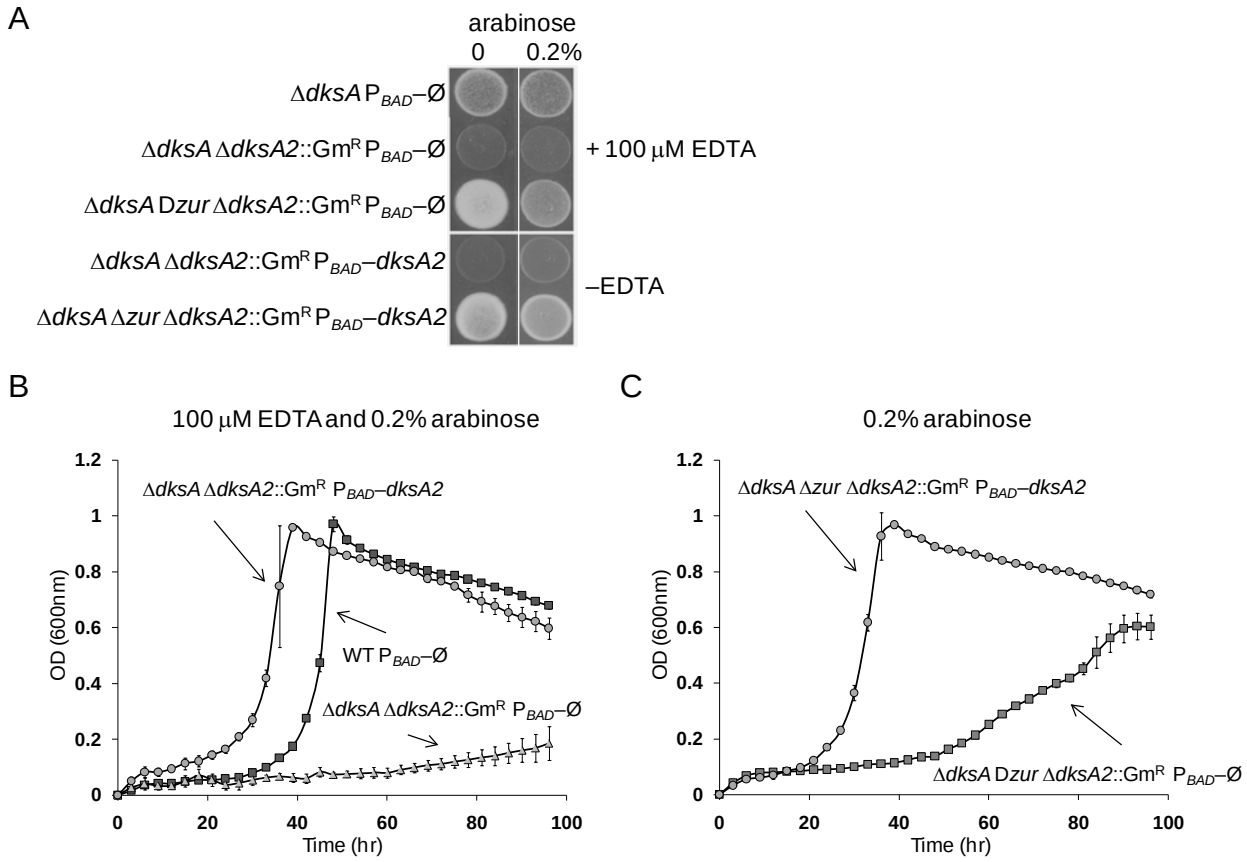


Figure 5-9. Complementation of the Δ*dksA2*::Gm^R mutant with *dksA2*, *in trans*. A) Overnight cultures of *P. aeruginosa* derivatives grown in LB were washed with M9 medium then normalized to an OD of 1.0. 10 µl of a 10^{-3} dilution was plated on M9 medium with the indicated supplementation. B) and C) Overnight cultures of *P. aeruginosa* derivatives grown in LB were washed with M9 medium then normalized to an OD of 1.0. Each culture was then diluted 500-fold into fresh M9 medium plus 0.2% arabinose and the indicated supplementation. Growth curves were monitored with a Bioscreen C. Error bars represent ± the standard deviation of three cultures.

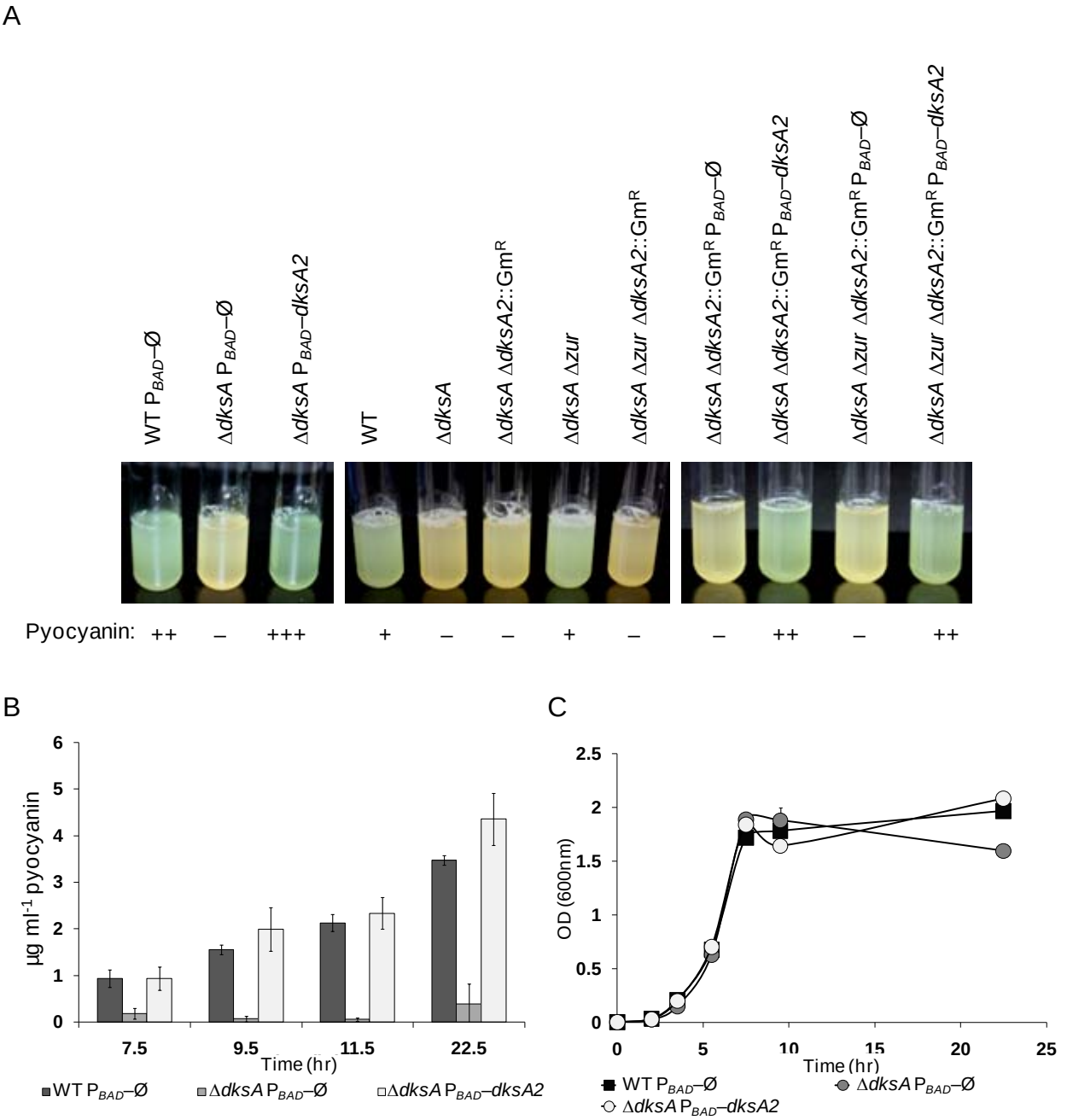


Figure 5-10. Pyocyanin defect of *P. aeruginosa* Δ*dksA* is rescued by expression of *dksA2*. A) Pyocyanin production of various *P. aeruginosa* mutant strains harboring the indicated constructs *in trans*. 5 ml cultures of LB were inoculated with a colony and grown overnight (~16 hrs) at 37°C. B) Pyocyanin production of WT pHERD20T, Δ*dksA* pHERD20T and Δ*dksA* pCH115 at various time points during in LB at 37°C. C) Corresponding growth curve of the strains assayed in (B).

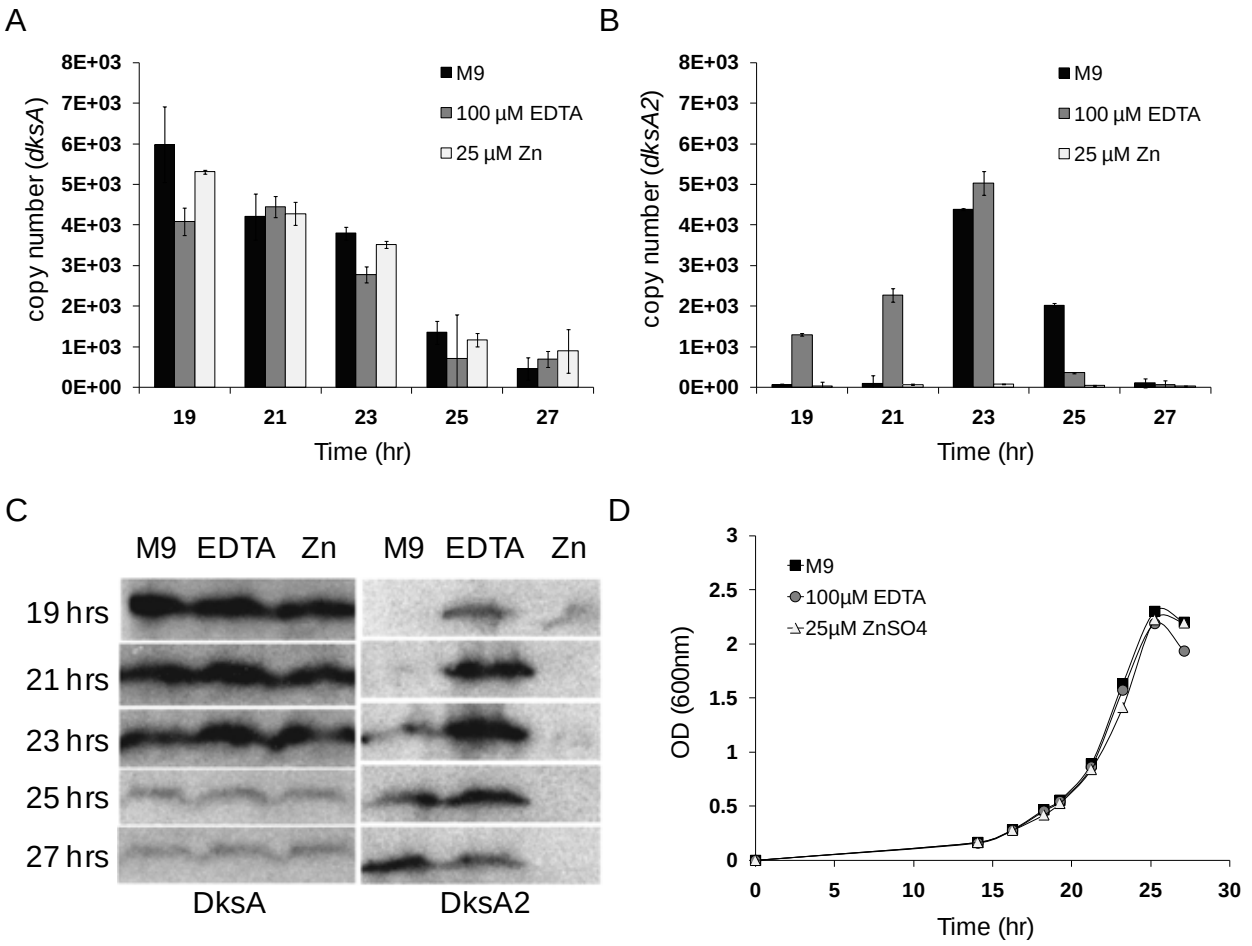


Figure 5-11. Transcript abundance of PA *dksA* and *dksA2* and protein abundance of PA DksA and DksA2. Average abundance of *dksA* (A) and *dksA2* (B) transcript in total mRNA extracted from *P. aeruginosa* at 19, 21, 23, 25 and 27 hrs of growth in M9 medium without supplementation (M9), plus 100 µM EDTA or plus 25 µM $ZnSO_4$. Copy number refers to the number of *dksA* or *dksA2* transcripts per 0.1 ng total mRNA. Error bars represent ± one standard deviation of three replicates. C) Immunoblot of DksA and DksA2 from total protein extracts from *P. aeruginosa* harvested at the same time points as for the transcript isolation in A and B. Blots were exposed to either anti-DksA or anti-DksA2 antibodies as indicated. D) Corresponding growth curves of *P. aeruginosa* cultured in M9 medium without (M9) or with 25 µM $ZnSO_4$.

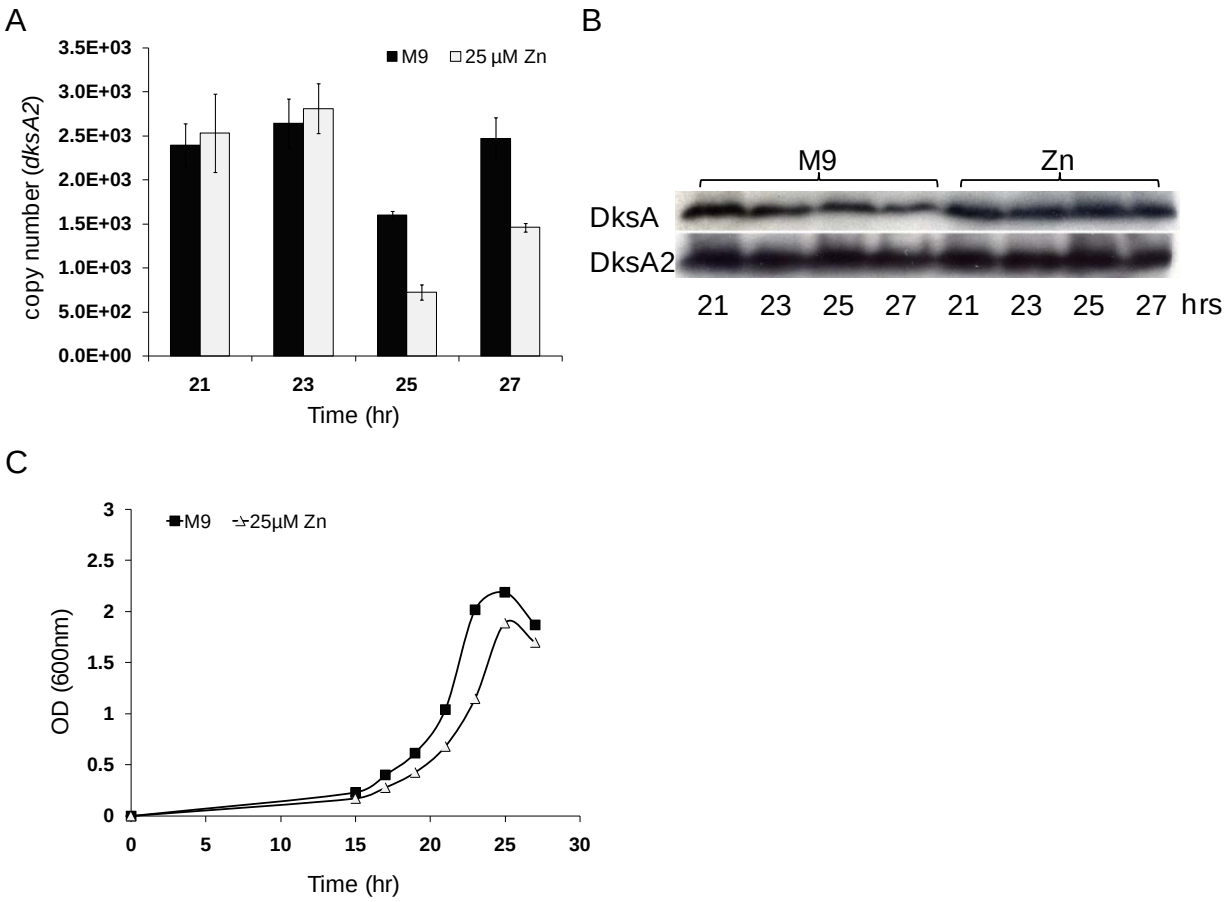


Figure 5-12. Effect of *zur* deletion of the transcript abundance of *dksA2* and protein abundance of DksA2. A) Average abundance of *dksA2* transcript in total mRNA extracted from *P. aeruginosa* Δ*zur*::Gm^R at 21, 23, 25 and 27 hours of growth in M9 medium without (M9) or with 25 µM $ZnSO_4$. Copy number refers to the number of *dksA2* transcripts per 0.1 ng total mRNA. Error bars represent ± one standard deviation of three replicates. B) Immunoblot of DksA and DksA2 from total protein extracts from *P. aeruginosa* Δ*zur*::Gm^R harvested at the same time points as for the transcript isolation in A. Blots were exposed to either anti-DksA or anti-DksA2 antibodies as indicated. C) Corresponding growth curve of *P. aeruginosa* Δ*zur*::Gm^R cultured in M9 medium without (M9) or with 25 µM $ZnSO_4$.

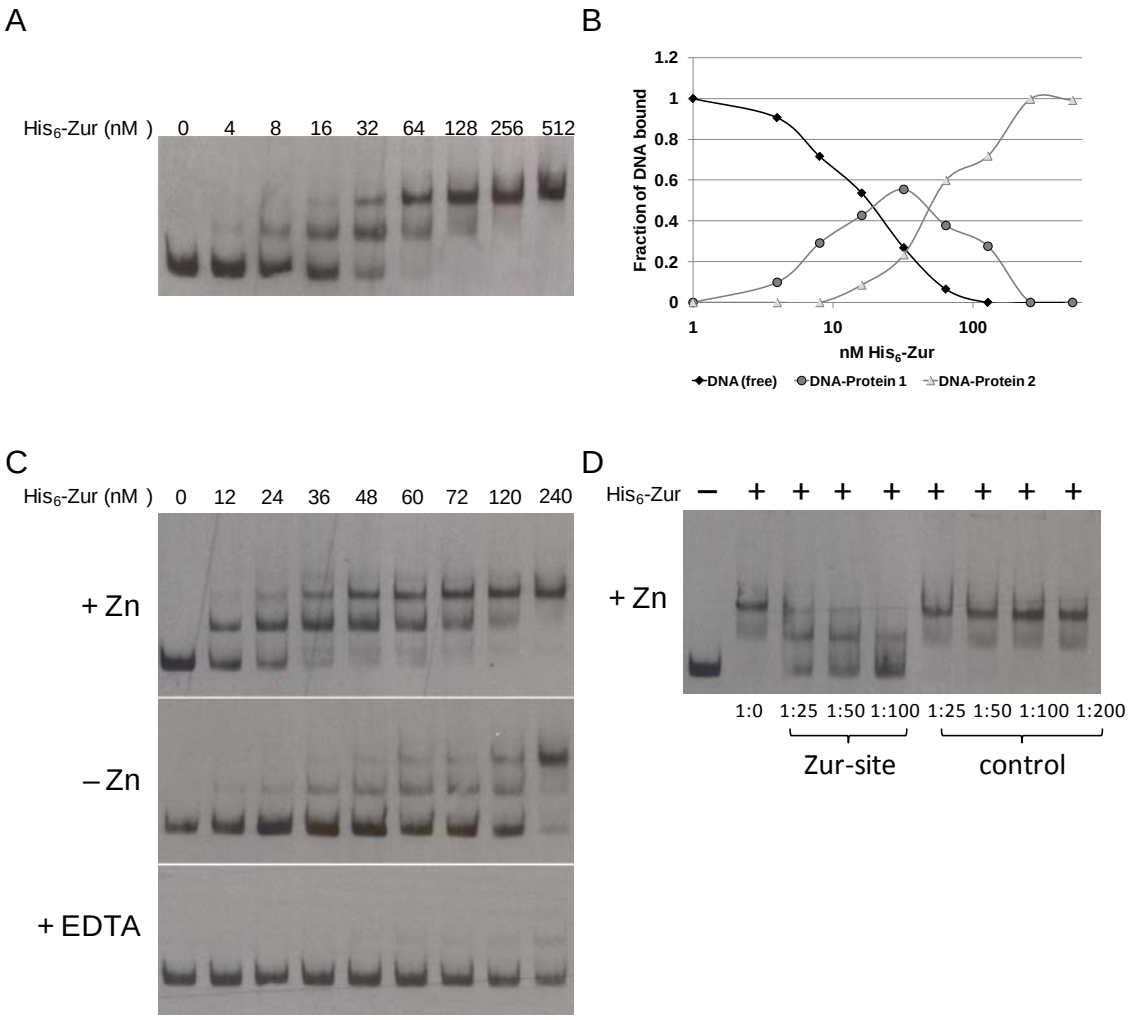


Figure 5-13. Complex formation between the upstream DNA region of *dksA2* and purified His_6-Zur. Complexes were separated by non-denaturing polyacrylamide gel electrophoresis and visualized with a Phototrope kit (NEB). A) Increasing concentrations of purified His_6-Zur were incubated in the presence of 100 µM $ZnSO_4$ and 1.5 ng of a 223 bp DNA fragment labeled with biotin at the 5' end. The fragment represents the DNA region including and surrounding the putative Zur-binding site (shown in Figure 5-4B). The concentration of His_6-Zur shown corresponds to the monomeric unit. B) Graphical representation of the concentration of free (DNA (free)) and complexed DNA (DNA-Protein 1 and DNA-Protein 2) at each concentration of His_6-Zur from A. DNA-Protein 1 refers to the smaller complex and DNA-Protein 2 refers to the larger complex. C) The binding buffer contained 100 µM $ZnSO_4$, no added Zn or 100 µM EDTA, as indicated. D) +/– 240 nM of His_6-Zur as indicated plus 1.5 ng biotin-labeled DNA with increasing amounts of either specific competitor (annealed oligonucleotides that contains the putative Zur binding site) or nonspecific competitor (oligonucleotides that do not contain a putative site).

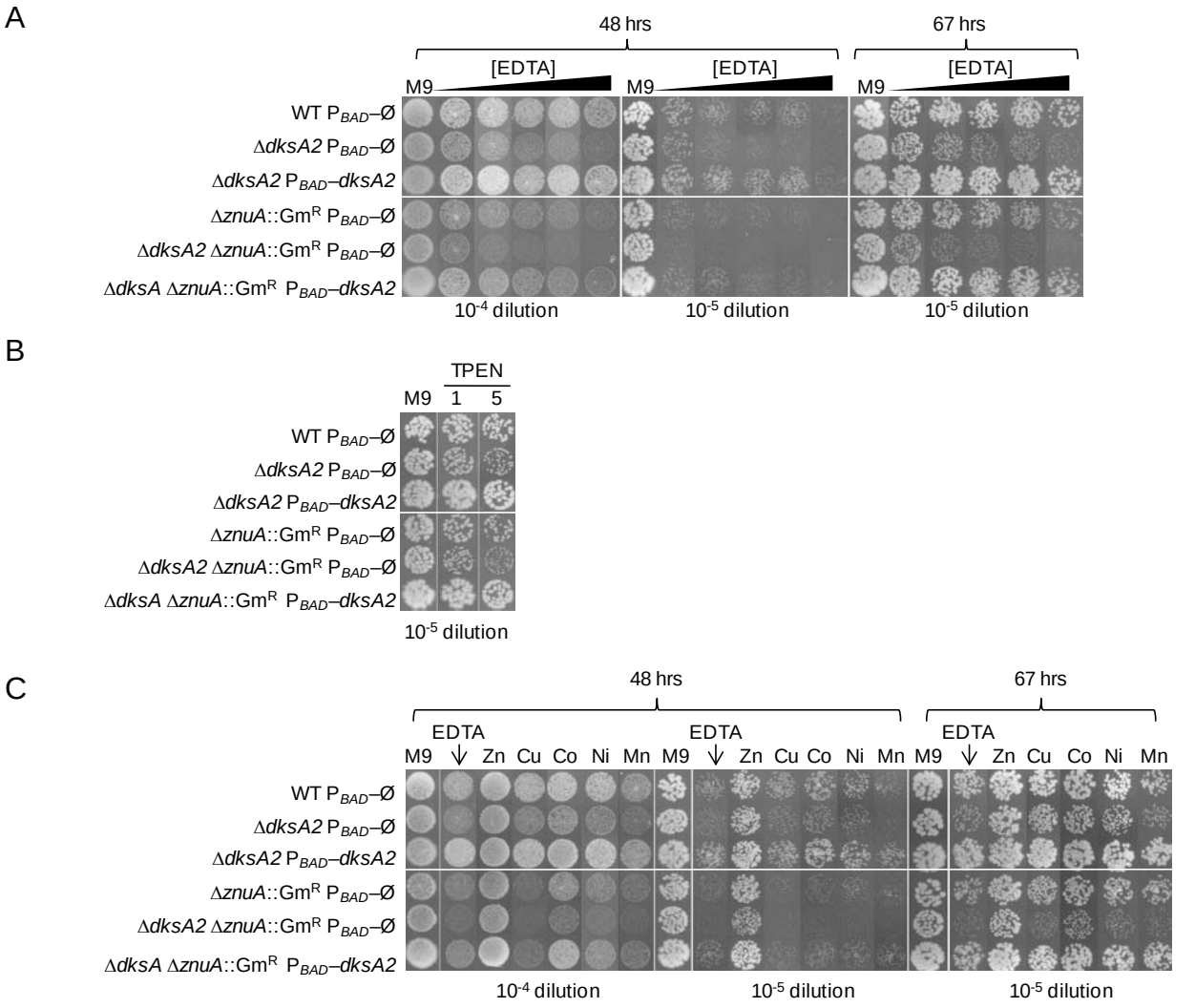


Figure 5-14. Growth defect of *P. aeruginosa* Δ*dksA2* in the presence of EDTA or TPEN. Overnight cultures of *P. aeruginosa* derivatives grown in LB were washed with M9 medium then normalized to an OD of 1.0. 10 µl of the indicated dilution was plated on M9 medium plus 0.2% arabinose. A) Strains were plated on M9 medium without EDTA (M9) or with 1, 1.25, 1.5, 1.75, or 2 mM EDTA and imaged at 48 or 67 hrs. B) Strains were plated on M9 medium without TPEN (M9) or with TPEN at either 1 µM (1) or 5 µM (5) and imaged after 48 hrs. C) Strains were plated on M9 medium without EDTA (M9) or with 1.25 mM EDTA (EDTA) +/– 25 µM of the indicated metals. Plates were imaged at the indicated time after incolulation. The images are composites of individual plates. Each plate was composed of one medium type. All six strains were plated on the same plate.

CHAPTER 6
SUMMARY AND OVERALL CONCLUSIONS

The overall goal of this study was to advance our understanding of microbial zinc homeostasis as it relates to the adaptation of cells to zinc-depletion, with specific attention on the uncharacterized *yeiR* gene from *E. coli* and the uncharacterized *dksA2* gene from *P. aeruginosa*. Additionally, a secondary goal was to advance our understanding of the COG0523 protein family through phylogenomic techniques.

A broad question, that remains mostly unanswered, is how do cells ensure proper allocation of metals? As of yet, we do not have a complete picture of how each metal is trafficked in the bacterial cell and most likely the question will have a different answer depending on the organism. Because of its potential impact on answering this question, a phylogenomic analysis of the COG0523 family was performed to gain a deeper understanding of this family. Since the discovery of the first copper metallochaperone (Pufahl *et al.*, 1997), the perception of how metal proteins acquire their cofactor was revolutionized and researchers have begun to look for these factors. Based on the results from the phylogenomic analysis presented here, the COG0523 family may represent a rich source of metal chaperones that aid in the acquisition of metal by metal-dependent enzymes. Specifically, the YeiR protein of *E. coli* may be the missing zinc chaperone that is involved in ensuring a yet unknown protein or proteins acquire zinc, a function that may be essential only when the intracellular zinc concentration is severely depleted.

Two significant studies in the past decade have led to a reevaluation and advancement of our understanding of how bacteria respond when deprived of zinc. First, was the discovery that bacteria encode a high-affinity zinc transporter and that the

corresponding genes are specifically regulated through a zinc-sensing transcription factor (Patzer and Hantke, 1998, Gaballa and Helmann, 1998). The second discovery was the realization that a pool of unliganded zinc in the cytoplasm of *E. coli* does not exist (Outten and O'Halloran, 2001). Since these two discoveries, researchers have sought to discover additional mechanisms that enable growth in the face of zinc-limitation and decipher the mechanism by which the bacterial cell ensures proper allocation of zinc among metal-dependent proteins.

During growth in zinc-replete or moderately zinc-deplete conditions, expression of the high-affinity zinc transporter (Patzer and Hantke, 1998) coupled with liberation of zinc from the dissociable ribosomal pool (Graham *et al.*, 2009) may be adequate for ensuring that the appropriate zinc-dependent enzymes are metallated and growth is sustained. *E. coli* is known to encode one zinc-independent ribosomal paralog. Expression of this gene and integration into the ribosome would theoretically lead to the potential release of 50,000 zinc ions.

It is difficult to estimate how many doublings this pool of zinc would sustain in the absence of exogenous zinc. The zinc quota for *E. coli* is 200,000 zinc ions per cell (Outten and O'Halloran, 2001), so in the absence of exogenous zinc, 50,000 zinc ions would not last very long. There would also be an increasing pressure from competing metals as the zinc concentration within the cytoplasm drops. It would be beneficial if a chaperone was present to direct this pool of zinc to essential zinc-dependent proteins to the exclusion of non-essential zinc-binding proteins. The phylogenomic analysis of COG0523 proteins combined with the YeiR study leave open the possibility that YeiR could serve as such a stress response factor. Further study with particular attention on

determining any interactions between YeiR and other proteins during conditions of zinc-replete and -deplete conditions would need to be undertaken.

A less controversial mechanism for sustaining growth when zinc becomes depleted is expression of a back-up DksA protein. As DksA is a critical factor in maintaining homeostasis in response to nutrient limitation, one can imagine its importance in zinc-deplete growth conditions, which are most likely poor in other nutrients. As DksA is dependent on the presence of zinc for maintenance of structure, it may be inactive or less active in zinc-deplete conditions leading to pleotrophic disregulation of homeostasis. The model proposed here is that a zinc-independent DksA back-up, DksA2, is produced to compensate for the lack of the main DksA protein. This study provides evidence that, at least for the conditions tested, DksA2 can functionally replace DksA. Additionally, DksA2 appears to be present in the cell specifically when zinc is depleted and regulation of the *dksA2* gene is mediated through the zinc-sensing repressor Zur. Concurrent work in the Artsimovitch lab at the Ohio State University supports the role of DksA2 serving as a back-up DksA. DksA2 was found to bind in the secondary channel of RNAP and, based on *in vitro* transcription assays, DksA2 and DksA affected transcription in a similar manner at tested promoters (Blaby-Haas *et al.*, 2011).

This study has provided evidence for two novel strategies used by bacteria to sustain growth when deprived by zinc. The previously uncharacterized YeiR protein from *E. coli* is involved in sustaining growth in the presence of EDTA, which can be suppressed specifically by the addition of exogenous zinc. Additionally, YeiR is involved in sustaining growth in the presence of cadmium. The exact function of YeiR in these

processes, however, is still unknown. The previously uncharacterized DksA2 protein from *P. aeruginosa* is also involved in sustaining growth in the presence of EDTA, which can be suppressed specifically by the additions of exogenous zinc. The function of DksA in this process is better understood but whether DksA2 functions outside the proposed model is not known. For instance, as a RNAP-binding factor that is specifically expressed when zinc is depleted, DksA2 is, in essence, a zinc-responsive transcription factor. Further exploration of the role of DksA2 in the cell specifically when zinc is limited will provide further insight.

APPENDIX A
SUBSYSTEMS

Due to the size of subsystems, screenshots are provided here that provide examples for reference by the main text. The reader is directed to view the full subsystems at http://theseed.uchicago.edu/FIG/SubsysEditor.cgi. A brief description is provided, to enable the reader to navigate the website and the provided subsystems.

The link provided will bring the reader to the main directory for subsystems stored on the SEED database. The two subsystems built by the author are located in this directory and can be accessed by following the link for “G3E family of P-loop GTPases (metallocenter biosynthesis)” or “COG0523.”

On the SEED website, the subsystem is divided into multiple tabs:

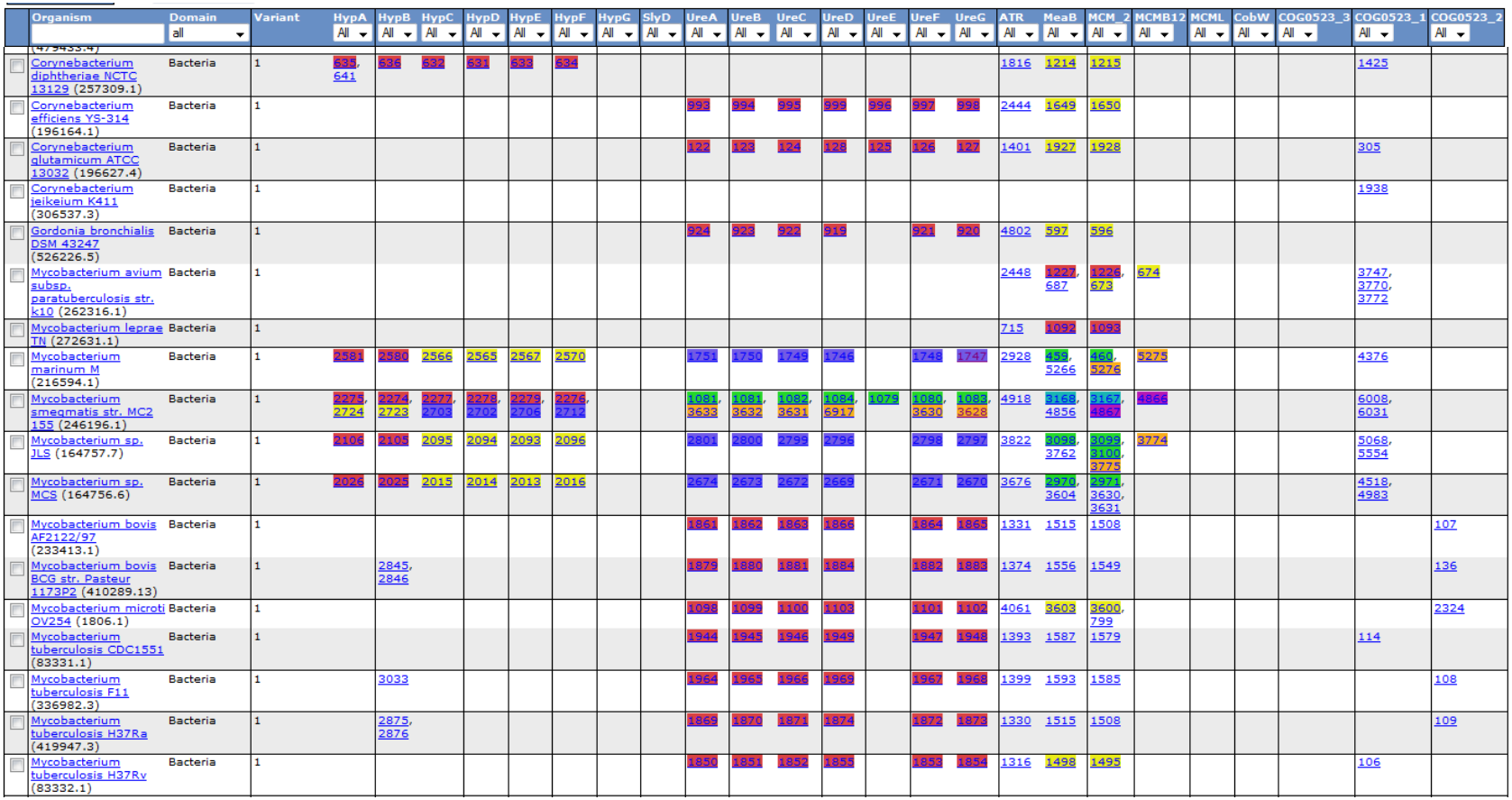
“Subsystem Info” gives a description of the subsystem including user provided background and relevant publications.

“Functional Roles” gives a list of the functional roles included in the subsystem. This page lists the user defined gene abbreviations used for each functional role in the subsystem. Links to relevant publications for each functional role are provided if available.

The “Spreadsheet” is the main element of the subsystem and provides a visual tool to discern the presence of specific functional roles among the various genome sequences provided by the SEED database. Organism names and the domain to which that organism belongs are provided to the left. The “Variant” column is used by the user to differentiate between organisms based on their interpretation of the spreadsheet. For instance, genomes that do not encode the pathway of interest can be assigned a variant code that differentiates those genomes from genomes that contain the pathway.

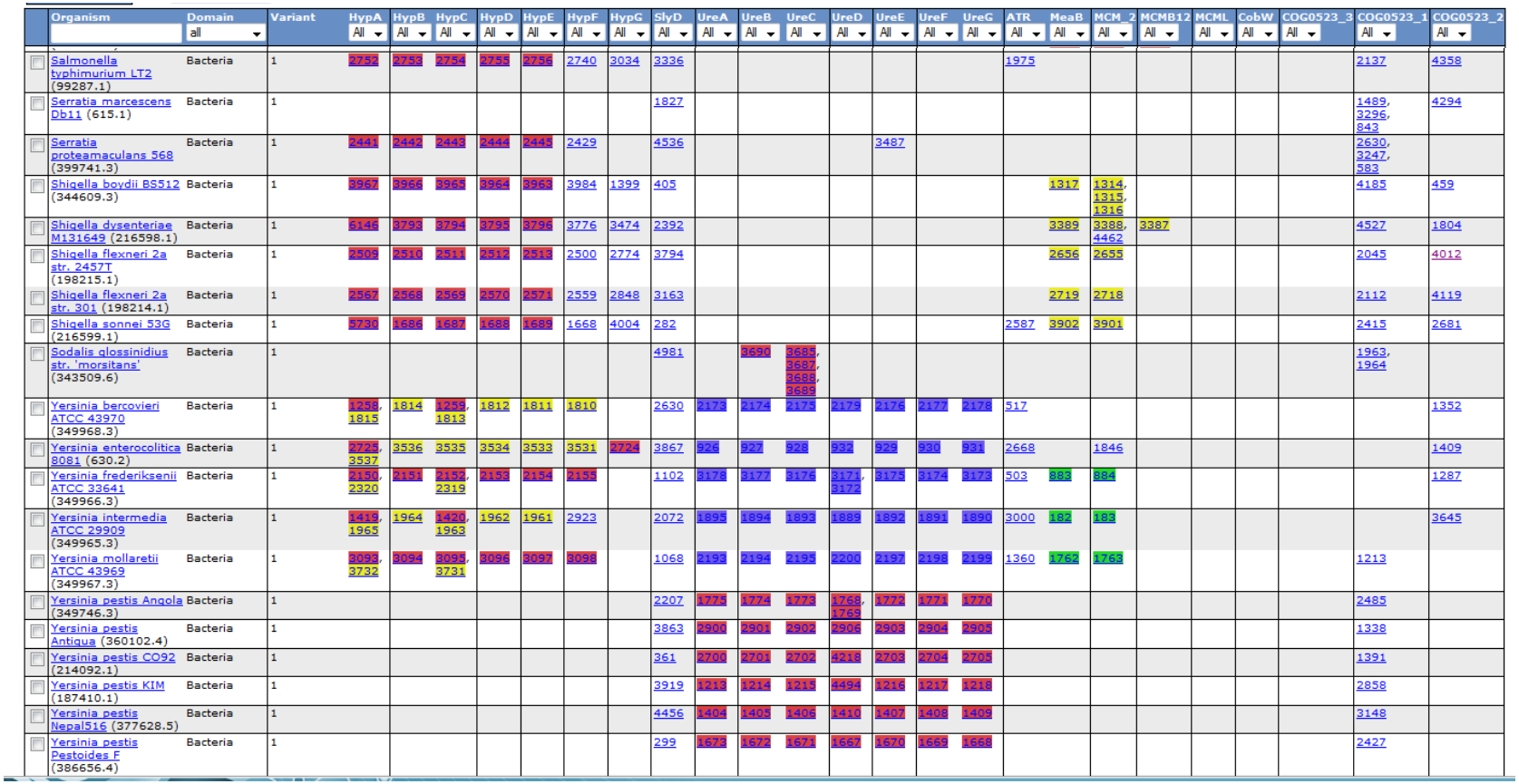
The presence of a gene with a certain functional role is represented by a number. The number provides a link to the gene page of that gene in that organism. The gene page provides more information such as gene sequence, protein sequence, links to other databases, and gene clustering information. If a cell in the spreadsheet is empty, then that gene is missing in that genome. Sometimes the gene is missing because it is mis-annotated. There are tools within SEED to find these genes and re-annotate them. Subsystems, therefore, commonly are used to find mis-annotations and repair them.

Clustering between genes in the spreadsheet is shown with coloring. In an organism, each gene cluster is represented by a different color because multiple gene clusters may exist that contain genes belonging to the functional roles added to the subsystem.



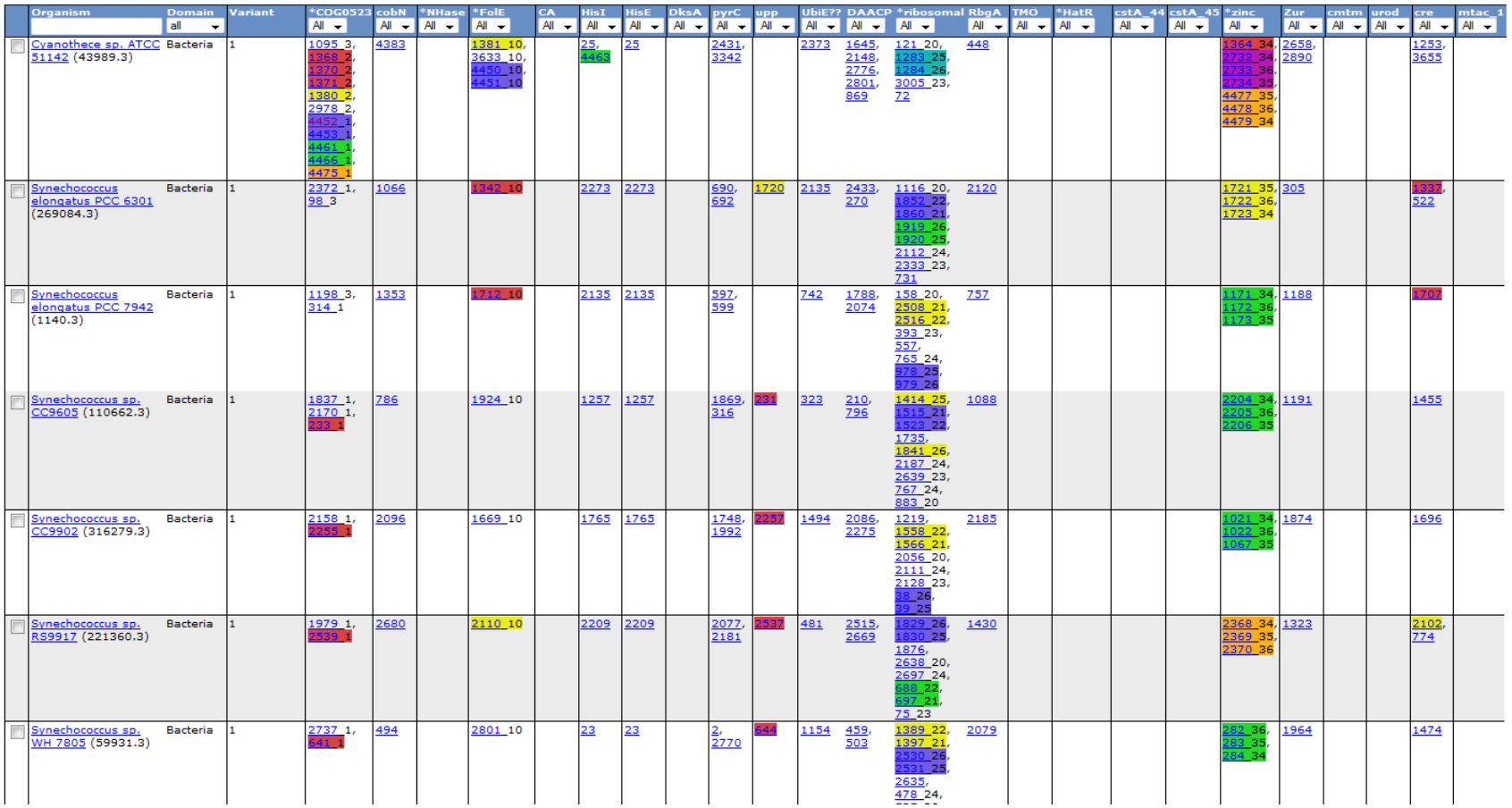
Organism	Domain	Variant	HypA	HypB	HypC	HypD	HypE	HypF	HypG	SlyD	UreA	UreB	UreC	UreD	UreE	UreF	UreG	ATR	MeaB	MCM_2	MCMB12	MCML	CobW	COG0523_3	COG0523_1	COG0523_2
Corynebacterium diphtheriae NCTC 13129 (257309.1)	Bacteria	1	635, 641	636	632	631	633	634										1816	1214	1215					1425	
Corynebacterium efficiens YS-314 (196164.1)	Bacteria	1									993	994	995	996	996	997	998	2444	1649	1650						
Corynebacterium glutamicum ATCC 13032 (196627.4)	Bacteria	1									122	123	124	125	125	126	127	1401	1927	1928					305	
Corynebacterium jeikeium K411 (306537.3)	Bacteria	1																							1938	
Gordonia bronchialis DSM 43247 (526226.5)	Bacteria	1									924	923	922	916		921	920	4802	597	596						
Mycobacterium avium subsp. paratuberculosis str. k10 (262316.1)	Bacteria	1																2448	1227, 687	1226, 673	674				3747, 3770, 3772	
Mycobacterium leprae TN (272631.1)	Bacteria	1																715	1092	1093						
Mycobacterium marinum M (216594.1)	Bacteria	1	2581	2580	2566	2565	2567	2570			1751	1750	1749	1748		1748	1747	2928	459, 5266	460, 5276	5275				4376	
Mycobacterium smegmatis str. MC2 155 (246196.1)	Bacteria	1	2275, 2724	2274, 2723	2277, 2703	2278, 2702	2275, 2706	2276, 2712			1081, 3633	1081, 3632	1082, 3631	1084, 6917	1079	1080, 3630	1083, 3628	4918	3168, 4856	3167, 4867	4866				6008, 6031	
Mycobacterium sp. JLS (164757.7)	Bacteria	1	2106	2105	2095	2094	2093	2096			2801	2800	2799	2796		2798	2797	3822	3098, 3762	3095, 3100, 3775	3774				5068, 5554	
Mycobacterium sp. MCS (164756.6)	Bacteria	1	2026	2025	2015	2014	2013	2016			2674	2673	2672	2669		2671	2670	3676	2970, 3604	2971, 3630, 3631					4518, 4983	
Mycobacterium bovis AF2122/97 (233413.1)	Bacteria	1									1861	1862	1863	1866		1864	1865	1331	1515	1508						107
Mycobacterium bovis BCG str. Pasteur 1173P2 (410289.13)	Bacteria	1		2845, 2846							1879	1880	1881	1884		1882	1883	1374	1556	1549						136
Mycobacterium microti OV254 (1806.1)	Bacteria	1									1098	1099	1100	1103		1101	1102	4061	3603	3600, 799						2324
Mycobacterium tuberculosis CDC1551 (83331.1)	Bacteria	1									1944	1945	1946	1949		1947	1948	1393	1587	1579					114	
Mycobacterium tuberculosis F11 (336982.3)	Bacteria	1		3033							1964	1965	1966	1969		1967	1968	1399	1593	1585						108
Mycobacterium tuberculosis H37Ra (419947.3)	Bacteria	1		2875, 2876							1869	1870	1871	1874		1872	1873	1330	1515	1508						109
Mycobacterium tuberculosis H37Rv (83332.1)	Bacteria	1									1850	1851	1852	1855		1853	1854	1316	1498	1495					106	

Figure A-1. Comparative genomic analysis of *ureG* containing gene clusters. A screnshot of the "G3E family of P-loop GTPases (metallocenter biosynthesis)" subsystem is shown. The urease maturation pathway is composed of UreA, UreB and UreC (structural subunits of urease), UreD and UreF (urease accessory factors), UreG (GTPase component and G3E family member) and UreE (proposed metallochaperone compoenent). Most genomes that encode urease, also encode UreD, UreE, UreF and UreG, as shown for the genomes of *Corynebacterium afficiens* and *Corynebacterium glutamicum*. Some genomes however are missing the gene for UreE, as shown for multiple *Mycobacterium* spp. genomes.



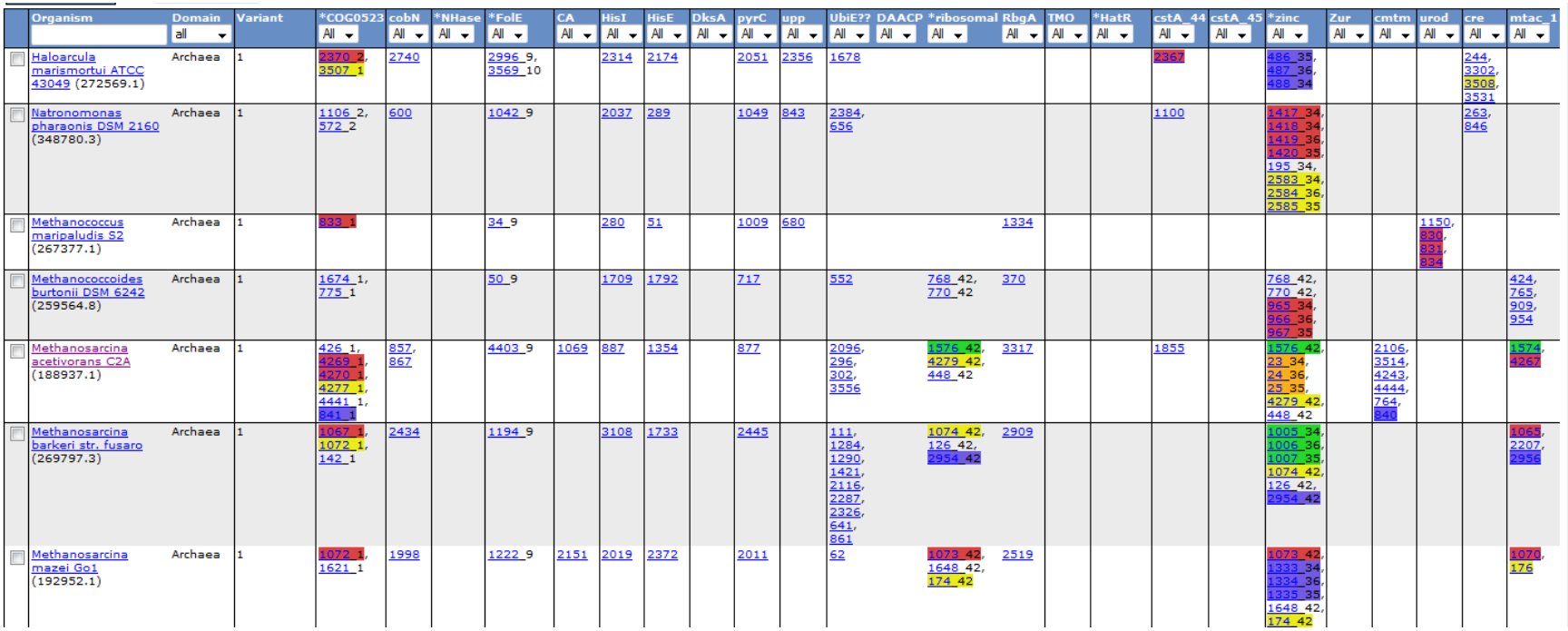
Organism	Domain	Variant	HypA	HypB	HypC	HypD	HypE	HypF	HypG	SlyD	UreA	UreB	UreC	UreD	UreE	UreF	UreG	ATR	MeaB	MCM_2	MCMB12	MCML	CobW	COG0523_3	COG0523_1	COG0523_2
Salmonella typhimurium LT2 (99287.1)	Bacteria	1	2752	2753	2754	2755	2756	2740	3034	3336								1975							2137	4358
Serratia marcescens Db11 (615.1)	Bacteria	1								1827															1489, 3296, 843	4294
Serratia proteamaculans 568 (399741.3)	Bacteria	1	2441	2442	2443	2444	2445	2429		4536					3487										2630, 3247, 583	
Shigella boydii BS512 (344609.3)	Bacteria	1	3967	3966	3965	3964	3963	3984	1399	405									1317	1314, 1315, 1316					4185	459
Shigella dysenteriae M131649 (216598.1)	Bacteria	1	3146	3795	3794	3795	3796	3776	3474	2392									3389	3388, 4462	3387				4527	1804
Shigella flexneri 2a str. 2457T (198215.1)	Bacteria	1	2508	2510	2511	2512	2513	2500	2774	3794									2656	2655					2045	4012
Shigella flexneri 2a str. 301 (198214.1)	Bacteria	1	2567	2568	2569	2570	2571	2559	2848	3163									2719	2718					2112	4119
Shigella sonnei 53G (216599.1)	Bacteria	1	3730	1686	1687	1688	1689	1668	4004	282								2587	3902	3901					2415	2681
Sodalis glossinidius str. 'morsitans' (343509.6)	Bacteria	1								4981		3690	3685, 3687, 3688, 3689												1963, 1964	
Yersinia bercovieri ATCC 43970 (349968.3)	Bacteria	1	1258, 1815	1814	1259, 1813	1812	1811	1810		2630	2173	2174	2175	2176	2176	2177	2178	517								1352
Yersinia enterocolitica 8081 (630.2)	Bacteria	1	2725, 3537	3536	3535	3534	3533	3531	2724	3867	926	927	928	932	929	930	931	2668		1846						1409
Yersinia frederiksenii ATCC 33641 (349966.3)	Bacteria	1	2150, 2320	2151	2152, 2319	2153	2154	2155		1102	3178	3177	3176	3171, 3172	3175	3174	3173	503	882	884						1287
Yersinia intermedia ATCC 29909 (349965.3)	Bacteria	1	1418, 1965	1964	1420, 1963	1962	1961	2923		2072	1895	1894	1893	1889	1892	1891	1890	3000	182	183						3645
Yersinia mollaretii ATCC 43969 (349967.3)	Bacteria	1	3093, 3732	3094	3095, 3731	3096	3097	3098		1068	2193	2194	2195	2200	2197	2198	2199	1360	1762	1763					1213	
Yersinia pestis Angola (349746.3)	Bacteria	1								2207	1775	1774	1773	1768, 1769	1772	1771	1770								2485	
Yersinia pestis Antiqua (360102.4)	Bacteria	1								3863	2900	2901	2902	2906	2903	2904	2905								1338	
Yersinia pestis CO92 (214092.1)	Bacteria	1								361	2700	2701	2702	4218	2703	2704	2705								1391	
Yersinia pestis KIM (187410.1)	Bacteria	1								3919	1213	1214	1215	4494	1216	1217	1218								2858	
Yersinia pestis Nepal516 (377628.5)	Bacteria	1								4456	1404	1405	1406	1410	1407	1408	1409								3148	
Yersinia pestis Pestoides F (386656.4)	Bacteria	1								299	1673	1672	1671	1667	1670	1669	1668								2427	

Figure A-2. Gene clustering of *hypB, ureG* and *meaB* with their target metalloenzyme or other accessory factors. A screenshot of the "G3E family of P-loop GTPases (metallocenter biosynthesis)" subsystem is shown. The gene encoding HypB is commonly found in a gene cluster encoding the [NiFe]-hydrogenase accessory factors HypA, HypC, HypD, HypE, HypF and HypG. The gene encoding UreG is commonly found in a gene cluster encoding urease (*ureABC*) and other accessory factors (*ureDEF*). The gene encoding MeaB is commonly found nect to the gene encoding methylmalonly-CoA mutase, *meaB*.



Organism	Domain	Variant	*COG0523	cobN	*NHase	*FolE	CA	HisI	HisE	DksA	pyrC	upp	UbiE??	DAACP	*ribosomal	RbgA	TMO	*HatR	cstA_44	cstA_45	*zinc	Zur	cmtm	urod	cre	mtac_1
Cyanothece sp. ATCC 51142 (43989.3)	Bacteria	1	1095_3, 1368_2, 1370_2, 1371_2, 1380_2, 2978_2, 4452_1, 4453_1, 4461_1, 4466_1, 4475_1	4383		1381_10, 3633_10, 4450_10, 4451_10		25, 4463	25		2431, 3342		2373	1645, 2148, 2776, 2801, 869	121_20, 1283_25, 1284_26, 3005_23, 72	448					1864_34, 2732_34, 2733_36, 2734_35, 4477_35, 4478_36, 4479_34	2658, 2890			1253, 3655	
Synechococcus elongatus PCC 6301 (269084.3)	Bacteria	1	2372_1, 98_3	1066		1342_10		2273	2273		690, 692	1720	2135	2433, 270	1116_20, 1852_22, 1860_21, 1919_26, 1920_25, 2112_24, 2333_23, 731	2120					1721_35, 1722_36, 1723_34	305			1337, 522	
Synechococcus elongatus PCC 7942 (1140.3)	Bacteria	1	1198_3, 314_1	1353		1712_10		2135	2135		597, 599		742	1788, 2074	158_20, 2508_21, 2516_22, 393_23, 557, 765_24, 978_25, 979_26	757					1171_34, 1172_36, 1173_35	1188			1707	
Synechococcus sp. CC9605 (110662.3)	Bacteria	1	1837_1, 2170_1, 233_1	786		1924_10		1257	1257		1869, 316	231	323	210, 796	1414_25, 1515_21, 1523_22, 1735, 1841_26, 2187_24, 2639_23, 767_24, 883_20	1088					2204_34, 2205_36, 2206_35	1191			1455	
Synechococcus sp. CC9902 (316279.3)	Bacteria	1	2158_1, 2255_1	2096		1669_10		1765	1765		1748, 1992	2257	1494	2086, 2275	1219, 1558_22, 1566_21, 2056_20, 2111_24, 2128_23, 38_26, 39_25	2185					1021_34, 1022_36, 1067_35	1874			1696	
Synechococcus sp. RS9917 (221360.3)	Bacteria	1	1979_1, 2539_1	2680		2110_10		2209	2209		2077, 2181	2537	481	2515, 2669	1829_26, 1830_25, 1876, 2638_20, 2697_24, 688_22, 697_21, 75_23	1430					2368_34, 2369_35, 2370_36	1323			2102, 774	
Synechococcus sp. WH 7805 (59931.3)	Bacteria	1	2737_1, 641_1	494		2801_10		23	23		2, 2770	644	1154	459, 503	1389_22, 1397_21, 2530_26, 2531_25, 2635, 478_24,	2079					282_36, 283_35, 284_34	1964			1474	

Figure A-3. Gene clusters involving COG0523 genes in bacterial genomes. A screenshot of the "COG0523" subsystem is shown. As an example, in the *Cyanothece* sp. ATCC 51142 genome, 11 COG0523 genes are found. Of these, 4 genes cluster physically on the genome with genes involved in zinc uptake (zinc). Three genes cluster with three of four genes for GTP cyclohydrolase I proteins (FolE). Two genes cluster with a gene for phosphoribosyl-AMP cyclohydrolase (HisI).



Organism	Domain	Variant	*COG0523	cobN	*NHase	*FolE	CA	HisI	HisE	DksA	pyrC	upp	UbiE??	DAACP	*ribosomal	RbgA	TMO	*HatR	cstA_44	cstA_45	*zinc	Zur	cmtm	urod	cre	mtac_1
Haloarcula marismortui ATCC 43049 (272569.1)	Archaea	1	2370_2, 3507_1	2740		2996_9, 3569_10		2314	2174		2051	2356	1678						2367		486_35, 487_36, 488_34				244, 3302, 3508, 3531	
Natronomonas pharaonis DSM 2160 (348780.3)	Archaea	1	1106_2, 572_2	600		1042_9		2037	289		1049	843	2384, 656						1100		1417_34, 1418_34, 1419_36, 1420_35, 195_34, 2583_34, 2584_36, 2585_35				263, 846	
Methanococcus maripaludis S2 (267377.1)	Archaea	1	833_1			34_9		280	51		1009	680				1334								1150, 830, 831, 834		
Methanococcoides burtonii DSM 6242 (259564.8)	Archaea	1	1674_1, 775_1			50_9		1709	1792		717		552		768_42, 770_42	370					768_42, 770_42, 965_34, 966_36, 967_35					424, 765, 909, 954
Methanosarcina acetivorans C2A (188937.1)	Archaea	1	426_1, 4269_1, 4270_1, 4277_1, 4441_1, 841_1	857, 867		4403_9	1069	887	1354		877		2096, 296, 302, 3556		1576_42, 4279_42, 448_42	3317			1855		1576_42, 23_34, 24_36, 25_35, 4279_42, 448_42		2106, 3514, 4243, 4444, 764, 840			1574, 4267
Methanosarcina barkeri str. fusaro (269797.3)	Archaea	1	1067_1, 1072_1, 142_1	2434		1194_9		3108	1733		2445		111, 1284, 1290, 1421, 2116, 2287, 2326, 641, 861		1074_42, 126_42, 2954_42	2909					1005_34, 1006_36, 1007_35, 1074_42, 126_42, 2954_42					1065, 2207, 2956
Methanosarcina mazei Go1 (192952.1)	Archaea	1	1072_1, 1621_1	1998		1222_9	2151	2019	2372		2011		62		1073_42, 1648_42, 174_42	2519					1073_42, 1333_34, 1334_36, 1335_35, 1648_42, 174_42					1070, 176

Figure A-4. Gene clusters involving COG0523 genes in archaeal genomes. A screenshot of the "COG0523" subsystem is shown. Among archaeal genomes that contain homologs of COG0523, gene clusters are found between a COG0523 gene and a creatine amidohydrolase gene (*Haloarcula marismortui*), a COG0523 gene and a carbon starvation gene (*Haloarcula marismortui*), a COG0523 gene and genes for uroporphyrinogen decarboxylase genes (*Natronomonas pharaonis*), COG0523 genes and methylcobalamin:coenzyme M methyltransferase genes (*M. acetivorans*, *M. barkeri*, and *M. mazei*), and a COG0523 gene and a ribosomal protein gene and zinc transport genes (*M. mazei*).

Organism	Domain	Variant	*COG0523	cobN	*NHase	*FolE	CA	HisI	HisE	DksA	pyrC	upp	UbiE??	DAACP	*ribosomal RbgA	TMO	*HatR	cstA_44	cstA_45	*zinc	Zur	cmtm	urod	cre	mtac_1
Gibberella zeae PH-1 (229533.1)	Eukaryota	1	5100_1, 5145_1, 7836_1			5854_10		865	865		9577, 9638	2476, 4155, 7475	4845												
Magnaporthe grisea 70-15 (242507.1)	Eukaryota	1	10285_1, 8908_1			1472_10, 6352_10		5123	5123		5351	10996													
Neurospora crassa (5141.1)	Eukaryota	1	6312_1			6413_10		8508	8508		6926	4049, 5818	3404												
Eremothecium gossypii (33169.1)	Eukaryota	1	230_1			4646_10		3896	3896		1310, 2603	3244													
Saccharomyces cerevisiae (baker's yeast) (4932.3)	Eukaryota	1	5506_1			2912_10		620	620		4554	3157, 5489													
Schizosaccharomyces pombe (4896.1)	Eukaryota	1	2111_1			3451_10		1696	1696		3025, 4541	3475, 3480, 4755, 571													
Fugu rubripes (Fugu) (31033.3)	Eukaryota	1	10784_2, 10785_2, 10786_2, 10787_2			15323_10, 3990_10, 6540_10, 6984_10						15619, 15620, 32374				17243									
Tetraodon nigroviridis (Green puffer) (99883.3)	Eukaryota	1	21474_1			4015_10, 566_10						16102	26632, 6343			6989									
Mus musculus (House mouse) (10090.3)	Eukaryota	1	4872_1			28902_10					11379	22003, 7714													
Rattus norvegicus (Norway rat) (10116.3)	Eukaryota	1	3919_1			11162_10						31775				23802									
Homo sapiens (Human) (9606.3)	Eukaryota	1	18530_1, 30962_1, 31379_1, 31381_1			9621_10, 9622_10					17657	33006													
Pan troglodytes (Chimpanzee) (9598.2)	Eukaryota	1	7570_1			10012_10, 10014_10					6701	22531, 22532, 37558													
Canis familiaris (dog) (9615.3)	Eukaryota	1	685_2, 686_2			26679_10, 26680_10						14666, 29740				1152									
Gallus gallus (Chicken) (9031.3)	Eukaryota	1	28080_1			20014_10						17033													
Anopheles gambiae str. PEST (180454.1)	Eukaryota	1	1718_1			10498_10					4844	2669, 590, 8132			12107			290		133_35					
Arabidopsis thaliana (3702.1)	Eukaryota	1	1879_1, 2968_1, 7636_1			12751_10		3535	3535		20157	15007, 15035, 16682, 16683, 20574, 25794, 5120			18097										

Figure A-5. Eukaryotic genomes that contain COG0523 genes. A screenshot of the “COG0523” subsystem is shown.

APPENDIX B
DETAILED DESCRIPTION OF COG0523 GENE CLUSTERS

Table B-1. Subgroup 2. Identified Fe-nitrile hydratase activator sequences.

Genome	Locus-tag	Genbank accession no.	Reference
Microbacterium sp. AJ115	nha3	CAG29801.1	(O'Mahony *et al.*, 2005)
Pseudomonas chlororaphis B23	P47K	P31521.1	(Nishiyama *et al.*, 1991)
Pseudomonas sp. K-9	nhr	BAD98534.1	(Kato *et al.*, 2005)
Rhodococcus erythropolis A4	nhr3	CAQ16890.1	(Kubác *et al.*, 2008)
Rhodococcus sp. N-771	nha3	BAA36599.1	(Lu *et al.*, 2003, Nojiri *et al.*, 1999)
Rhodococcus sp. N-774	ORF1188	BAA06274.1	(Hashimoto *et al.*, 1994)
Rhodococcus globerulus A-4	nhr3	BAC99082.1	(Xie *et al.*, 2003)
Rhodococcus jostii RHA1	RHA1_ro00362	ABG92198.1	
A. baylyi ADP1	ACIAD1614	YP_046285.1	
P. putida F1	Pput_2730	ABQ78864.1	
Burkholderia ambifaria AMMD	Bamb_6542	YP_778420.1	
B. ambifaria MC40-6	BamMC406_6256	ACB68689.1	
Burkholderia cenocepacia AU 1054	Bcen_4084	YP_623946.1	
B. cenocepacia MC0-3	Bcenmc03_3235	YP_001776881.1	

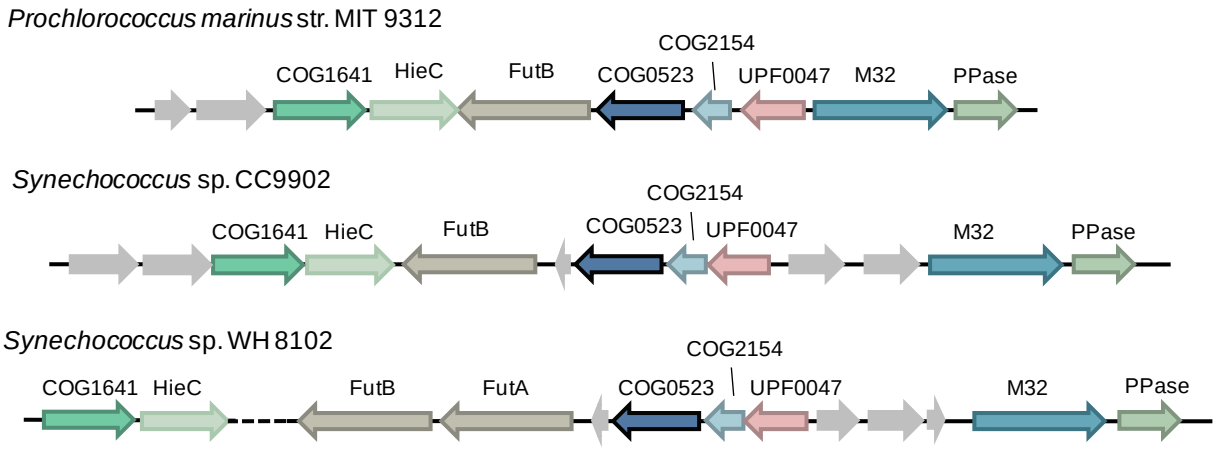
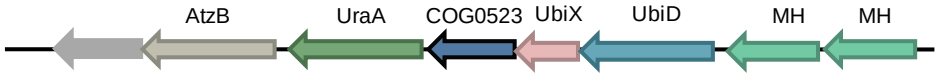


Figure B-1. Genome context of subgroup 3 members. Subgroup 3 genes co-localize on the genome with homologs of uncharacterized conserved proteins (COG1641), putative transmembrane protein (HieC), Ferric iron ABC transporter, permease protein (FutB) and periplasmic binding protein (FutA), Pterin-4-alpha-carbinolamine dehydratase (COG2154), protein of unknown function (UPF0047), carboxypeptidase Taq (M32) metallopeptidase (M32), and inorganic pyrophosphatase (PPase). Characterized members of COG2154 from animals have been shown to be responsible for recycling pterin cofactors generated by aromatic amino acid hydroxylases (AAHs) (for a review see (Thöny *et al.*, 2000)). In bacteria, the absence of AAH in many COG2154-encoding genomes suggests its role in the recycling of pterin cofactors from other pterin-dependent enzymes(Naponelli *et al.*, 2008). UPF0047 appears to be a metalloenzyme; the crystal structure of UPF0047 from *Sulfolobus tokodaii* has been solved and found to co-crystalize with a zinc ion in a potential catalytic site(Tanaka *et al.*, 2005). Over-expression of the UPF0047 homolog from *E. coli* as well as *Thermotoga, Sulfolobus,* and *Pyrococcus* was found to complement the thiamine auxotrophy of a Δ*thiE. coli* mutant (Morett *et al.*, 2008).

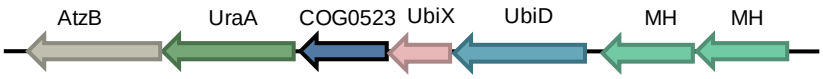
Table B-2. Genomes that contain the subgroup 3 gene cluster.

Genome	Locus_tag
Synechococcus sp. CC9311	Sync_2045
Synechococcus sp. CC9605	Syncc9605_0672
Synechococcus sp. RS9917	RS9917_09326
Synechococcus sp.WH 7805	WH7805_13878
Synechococcus sp. WH 8102	SYNW1795
P. marinus (5 genomes)	PMT9312_0491 (str. MIT 9312)

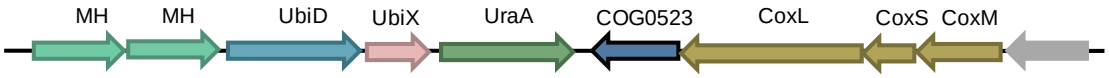
Burkholderia cepacia R18194 and *Burkholderia cenocepacia* J2315



Burkholderia multivorans ATCC 17616 and *Pseudomonas fluorescens* PfO-1



Sinorhizobium meliloti 1021



Ralstonia eutropha JMP134

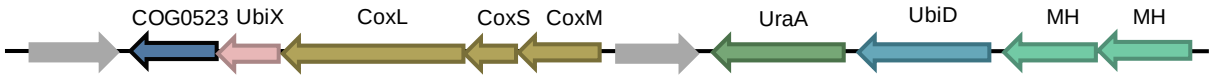


Figure B-2. Genome context of subgroup 4 members. Subgroup 4 genes co-localize on the genome with genes encoding metal-dependent hydrolases (MH), xanthine/uracil permease (UraA), 3-polyprenyl-4-hydroxybenzoate carboxy-lyase (UbiX and UbiD), hydroxyatrazine ethylaminohydrolase (AtzB), molybdenum-containing hydroxylase (CoxLSM). Several of these genes encode various proteins involved in the metabolism and degradation of aromatic compounds. In addition to molybdenum, Mo-containing hydroxylases have been found to contain [2Fe-2S] clusters (For a review see (Hille, 2005)).

Table B-3. Genomes that contain the subgroup 4 gene cluster.

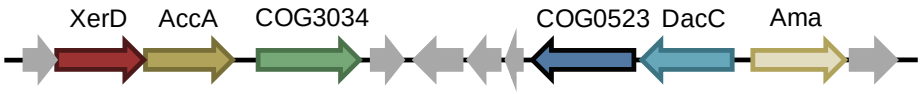
Organism	Locus_tag
Sinorhizobium meliloti 1021	SMb20133
Burkholderia cenocepacia J2315	BCAM2270
B. cepacia R18194	Bcep18194_B0634
Ralstonia eutropha JMP134	Reut_A3078
P. fluorescens PfO-1	Pfl_3432

Table B-4. Genomes that contain the subgroup 5 gene cluster.

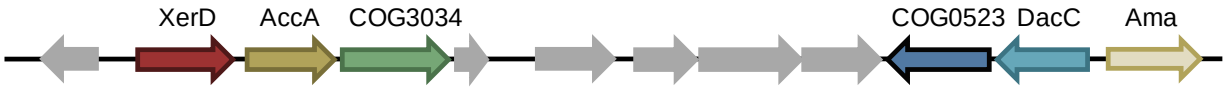
Genome	Locus_tag
Synechococcus sp. WH 8102	SYNW2482
Gloeobacter violaceus PCC 7421	glr0534
A. variabilis ATCC 29413	Ava_0285
Nostoc punctiforme PCC 73102	Npun_R3841
Nostoc sp. PCC 7120	all1751
P. marinus (5 genomes)	P9211_15201 (str. MIT 9211)
Caulobacter sp. K31	Caul_0196
Caulobacter crescentus CB15	CC0321
Aurantimonas sp. SI85-9A1	SI859A1_00649
Bartonella henselae str. Houston-1	BH12980
B. japonicum USDA 110	bll7768
Nitrobacter hamburgensis X14	Nham_3425
Nitrobacter winogradskyi Nb-255	Nwi_0917
Rhodopseudomonas palustris (4 genomes)	RPE_4794 (str. BisB53)
Brucella abortus biovar 1 str. 9-941	BruAb2_0247
Brucella canis ATCC 23365	BCAN_B1006
Brucella melitensis 16M	BMEII0308
B. suis 1330	BRA0987
Brucela suis ATCC 23445	BSUIS_B0982
Mesorhizobium loti MAFF303099	mll5156
Mesorhizobium sp. BNC1	MBNC02001409
Rhizobium leguminosarum bv. *viciae* 3841	RL4148
Sinorhizobium meliloti 1021	SMc02978
A. tumefaciens str. C58	Atu4502
Acidiphilium cryptum JF-5	Acry_1304
Gluconobacter oxydans 621H	GOX1617
Granulibacter bethesdensis CGDNIH1	GbCGDNIH1_0170
Bordetella bronchiseptica RB50	BB0682
Bordetella parapertussis 12822	BPP0675

Brucella suis 1330

Sinorhizobium meliloti 1021



Bartonella henselae str. Houston-1



Rhizobium leguminosarum bv. viciae 3841

Aurantimonas sp. SI85-9A1

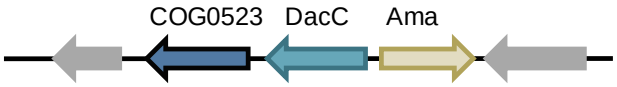


Figure B-4. Genome context of subgroup 6 members. Subgroup 6 genes co-localize with genes encoding integrase/recombinase (XerD), acetyl-coenzyme A carboxyl transferase alpha chain (AccA), uncharacterized protein (COG3034), D-alanyl-D-alanine carboxypeptidase (DacC), and N-acyl-L-amino acid amidohydrolase (Ama).

Table B-5. Genomes that contain the subgroup 6 gene cluster.

Organism	Locus_tag
Parvularcula bermudensis HTCC2503	PB2503_00115
Aurantimonas sp. SI85-9A1	SI859A1_02956
Bartonella henselae str. Houston-1	BH16310
Brucella abortus biovar 1 str. 9-941	BruAb1_2010
Brucella canis ATCC 23365	BCAN_A2081
Brucella melitensis 16M	BMEI0036
B. suis (2 genomes)	BR2035 (str. 1330)
Methylobacterium extorquens PA1	Mext_1219
Mesorhizobium loti MAFF303099	mll3580
Mesorhizobium sp. BNC1	MBNC02000233
A. tumefaciens str. C58	Atu3633
Rhizobium leguminosarum bv. viciae 3841	RL4362
Sinorhizobium meliloti 1021	SMc00684

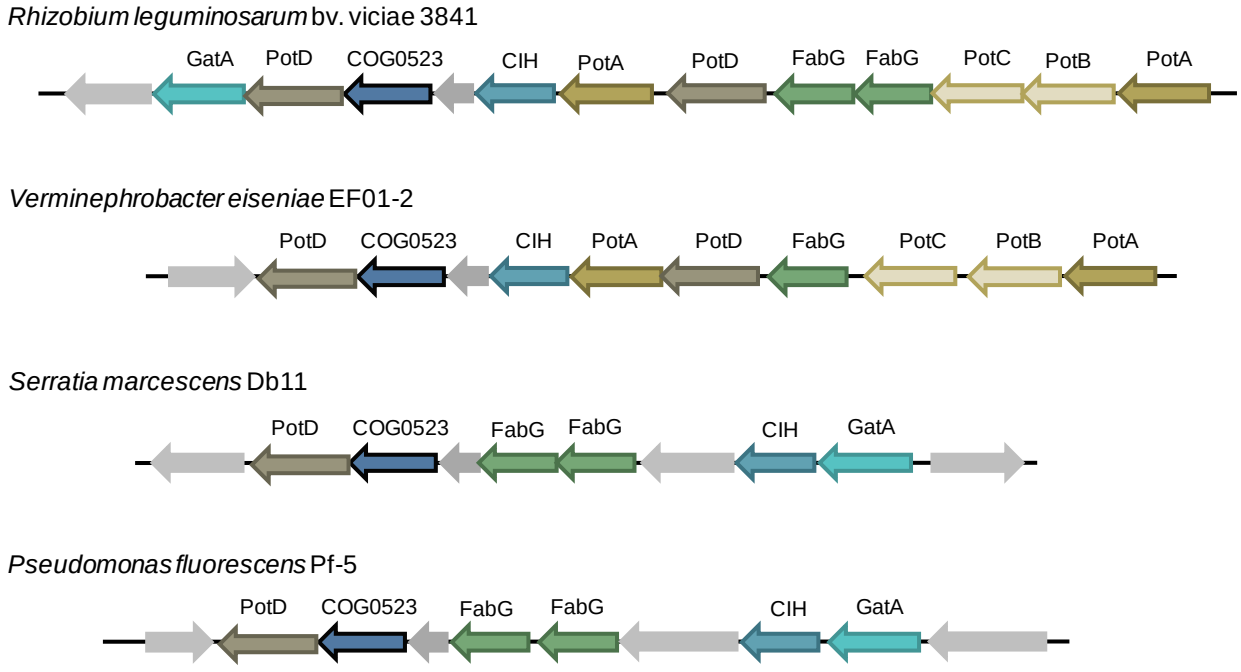
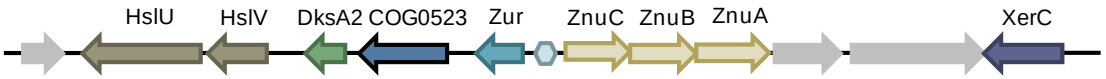


Figure B-5. Genome context of subgroup 7 members. Subgroup 7 genes co-localize with genes encoding Asp-tRNAAsn/Gu-tRNAGln amidotransferase A subunit and related amidases family protein (GatA), spermine/putrescine periplasmic binding protein (PotD), spermine/putrescine import ATP-binding protein (PotA), spermine/putrescine transport permease protein (PotB and PotC), cyclic imide hydrolase (CIH), and 3-ketoacyl-(acyl-carrier-protein) reductase (FabG).

Table B-6. Genomes containing subgroup 7 gene cluster.

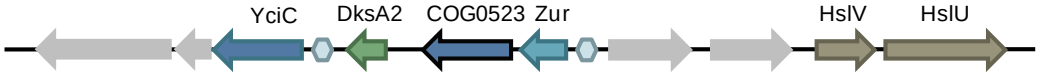
Organism	Locus_tag
Rhizobium leguminosarum bv. viciae 3841	pRL120796
Verminephrobacter eiseniae EF01-2	Veis_1001
S. marcescens DB11	fig|615.1.peg.4294
P. fluorescens Pf-5	PFL_1367
Pseudomonas syringae (3 genomes)	PSPTO4198 (pv. tomato str. DC3000)

Bordetella spp.



Ralstonia solanacearum GMI1000

Ralstonia eutropha JMP134



Burkholderia spp.

Figure B-6. Genome context of subgroup 8 members. Subgroup 8 genes cluster with ATP-dependent hsl protease ATP-binding subunit (HslU), ATP-dependent protease (HslV), C4-type zinc finger DksA/TraR family protein (DksA2), zinc uptake regulation protein (Zur), high affinity zinc transporter (ZnuABC), integrase/recombinase (XerC), and the COG0523 Subgroup 2 paralog (YciC).

Table B-7. Genomes containing subgroup 8 gene cluster.

Organism	Locus_tag
Bordetella avium	BAV0145
Bordetella bronchiseptica RB50	BB0181
Bordetella parapertussis 12822	BPP0179
B. pertussis Tohoma I	BP3084
Burkholderia ambifaria AMMD	Bamb_3135
Burkholderia cenocepacia(4 genomes)	Bcen_2475 (str. AU 1054)
B. cepacia R18194	Bcep18194_A6439
Burkholderia dolosa AUO158	BDAG_00307
Burkholderia multivorans ATCC 17616	Bmul_3084
B. cepacia R1808	Bucepa02001131
Burkholderia vietnamiensis strain G4	Bcep1808_3172
Burkholderia fungorum	Bcep2993
Burkholderia mallei(3 genomes)	BMA10229_A2171 (str. 10229)
Burkholderia xenovorans LB400	Bxe_A4377
Burkholderia pseudomallei(4 genomes)	BURPS1655_K0002 (str. 1655)
C. metallidurans CH34	Rmet_0127
Ralstonia eutropha JMP134	Reut_A0163
Polynucleobacter sp. QLW-P1DMWA-1	Pnuc_2015
Ralstonia solanacearum GMI1000	RSc0047
Acidovorax avenae	Aave_0807
Acidovorax sp. JS42	Ajs_3684
Delftia acidovorans SPH-1	Daci_1456
Polaromonas sp. JS666	Bpro_1063
Rhodoferax ferrireducens DSM 15236	Rfer_3439
Verminephrobacter eiseniae EF01-2	Veis_4556
Herminiimonas arsenicoxydans	HEAR2959
Leptothrix cholodni SP-6	Lcho_3834

Proteus mirabilis HI4320

Pseudomonas aeruginosa PAO1

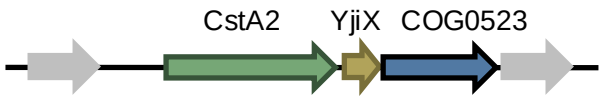


Figure B-7. Genome context of subgroup 9 members. Subgroup 9 genes co-localize with genes encoding a carbon starvation protein paralog (CstA2) and a protein of unknown function (YjiX).

Table B-8. Genomes containing subgroup 9 gene cluster.

Genome	Locus_tag
Enterobacter sp. 638	Ent638_0509
E. coli (17 genomes)	YjiA
K. pneumoniae MGH 78578	KPN_04774
Erwinia carotovora subsp. *atroseptica* SCRI1043	ECA1189
P. mirabilis HI4320	PMI0144
Salmonella bongori 12149	fig|12149.1.peg.4488
S. enterica (5 genomes)	STY4888 (subsp. Enterica serovar Typhi str. CT18)
S. typhimurium LT2	STM4530
S. marcescens Db11	fig|615.1.peg.4294
Serratia proteamaculans 568	Spro_0580
Shigella dysenteriae Sd197	SDY_4605
Shigella flexneri 2a (2 genomes)	S4640 (str. 2457T)
Shigella sonnei (2 genomes)	SSO_4485 (str. Ss046)
A. vinelandii	Avin3722
P. aeruginosa	PA4604 (str. PAO1)
P. entomophila L48	PSEEN0805
P. fluorescens (3 genomes)	PFL_5350 (str. Pf-5)
P. putida (4 genomes)	Pput_4501 (str. F1)
Pseudomonas syringae (3 genomes)	Psyr_4271 (pv. Syringae B728a)

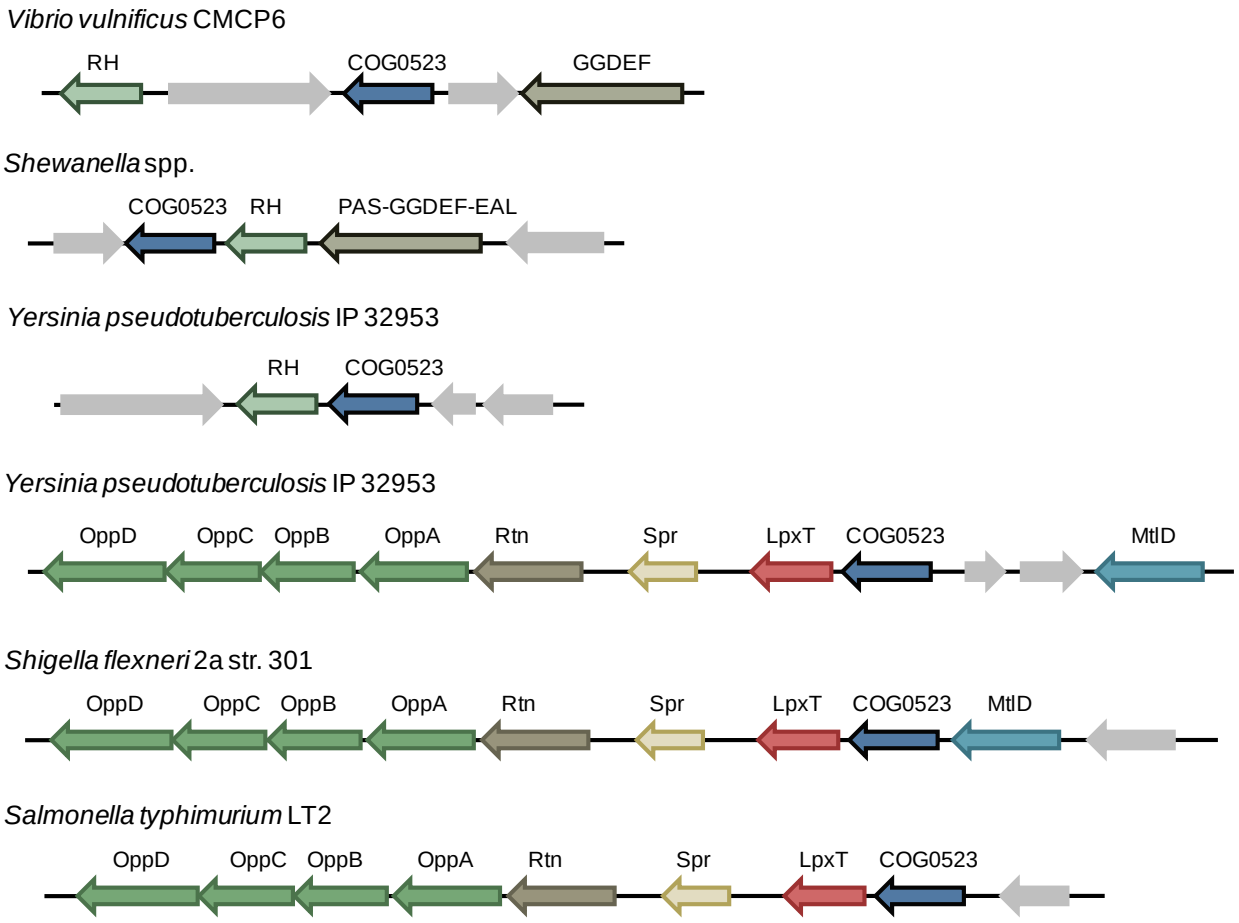
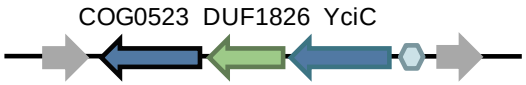


Figure B-8. Genome context of subgroup 10 members. Subgroup 10 genes co-localize with genes encoding signal transduction proteins (GGDEF, PAS-GGDEF-EAL, or Rtn) or an ATP-dependent RNA helicase (RH). Abbreviations: MtlD, mannitol-1-phosphate/altronate dehydrogenases; LpxT, undecaprenyl pyrophosphate phosphatase; Spr, predicted peptidase; OppABCD, putative ATP-dependent oligopeptide permease.

Table B-9. Genomes containing subgroup 10 gene cluster.

Genome	Locus_tag
Enterobacter sp. 638	Ent638_2769
E. coli (17 genomes)	YeiR
K. pneumoniae MGH 78578	KPN_02606
Erwinia carotovora subsp. *atroseptica* SCRI1043	ECA2733
Salmonella bongori 12149	fig|12149.1.peg.2276
S. enterica (5 genomes)	STY2448 (subsp. *enterica* serovar *Typhi* str. CT18)
S. typhimurium LT2	STM2212
S. marcescens Db11	fig|615.1.peg.3296
Serratia proteamaculans 568	Spro_3241
Shigella dysenteriae Sd197	SDY_0906
Shigella flexneri 2a (2 genomes)	SF2260 (str. 301)
Shigella sonnei(2 genomes)	SSO_2229 (str. Ss046)
Yersinia bercovieri ATCC 43970	YberA_01001247
Yersinia enterocolitica 8081	YE1435
Yersinia frederiksenii ATCC 33641	YfreA_01001244
Yersinia intermedia ATCC 29909	YintA_01003502
Yersinia mollaretii ATCC 43969	YmolA_01001161
Yersinia pestis(7 genomes)	YPA_0995 (str. Antiqua)
Yersinia pseudotuberculosis(3 genomes)	YPTB1311 (str. IP 32953)
Shewanella amozonensis SB2B	Sama_2446
Shewanella baltica(4 genomes)	Sbal_1338 (str. OS155)
Shewanella denitrificans OS217	Sden_2575
Shewanella frigidimarina NCIMB 400	Sfri_2795
Shewanella halifaxensis HAW-EB4	Shal_3090
Shewanella sp. PV-4	Shew_2776
Shewanella oneidensis MR-1	SO1502
Shewanella pealeana ATCC 700345	Spea_3001
Shewanella putrefaciens CN-32	Sputcn32_1255
Shewanella sediminis HAW-EB3	Ssed_3334
Shewanella sp. MR-4	Shewmr4_2751
Shewanella sp. MR-7	Shewmr7_2829
Shewanella sp. W3-18-1	Sputw3181_2849
Shewanella sp. ANA-3	Shewana3_2927
V. alginolyticus 12G01	V12G01_06973
Vibrio parahaemolyticus RIMD 2210633	VPA0589
Vibrio sp. Ex25	VEx2w_01001184
Vibrio sp. MED222	MED222_06985
Vibrio splendidus 12B01	V12B01_03988
Vibrio vulnificus(2 genomes)	VV20385 (str. CMCP6)

Pseudomonas putida KT2440



Pseudomonas aeruginosa PAO1

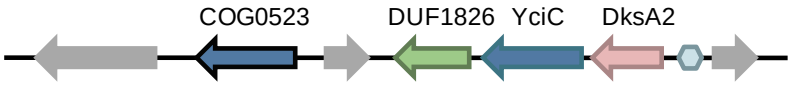


Figure B-9. Genome context of subgroup 11 members. Subgroup 11 genes co-localize with genes encoding an uncharacterized protein (DUF1826), COG0523 homolog from subfamily 2 (YciC), and C4-type zinc finger DksA/TraR family protein (DksA2).

Table B-10. Genomes containing subgroup 11 gene cluster.

Genome	Locus_tag
P. aeruginosa (3 genomes)	PA5532 (str. PAO1)
Pseudomonas mendocina ymp	Pmen_4533
P. entomophila L48	PSEEN5511
P. fluorescens (3 genomes)	PFL_6171 (str. Pf-5)
P. putida (4 genomes)	Pput_5267 (str. F1)
Pseudomonas syringae(3 genomes)	Psyr_5072 (pv. Syringae B278a)

Table B-11. Genomes containing subgroup 12.

Organism	CobW Locus Tag	first gene in operon (if operon is predicted to be regulated by riboswitch)
A. tumefaciens str. C58	Atu2805	Atu2806*
B. japonicum USDA 110	blr3261	blr3262*
Bradyrhizobium sp. BTAi1	BBta_3143	BBta_3141
Brucella abortus biovar 1 str. 9-941	BruAb1_1308	BruAb1_1311
Brucella canis ATCC 23365	BCAN_A1330	BCAN_A1333
Brucella melitensis 16M	BMEI0694	BMEI0691*
B. suis 1330	BR1307	BR1310
Dinoroseobacter shibae DFL 12	Dshi_0160	Dshi_0162
Granulibacter bethesdensis	GbCGDNIH1_0657	GbCGDNIH1_0657
Loktanella vestfoldensis SKA53	SKA53_10074	SKA53_10069
Mesorhizobium loti MAFF303099	mlr1375	mlr1374*
Methylobacterium sp. 4-46	M446_2436	M446_2433
Novosphingobium aromaticivorans	Saro_0343	Saro_0343
Oceanicola batsensis HTCC2597	OB2597_10309	OB2597_10304
Paracoccus denitrificans PD1222	Pden_2531	Pden_2530
Rhizobium leguminosarum	RL2831	RL2832
Rhodobacter sphaeroides 2.4.1	RSP_2828	RSP_2829*
Rhodobacterales bacterium	RB2654_11398	RB2654_11383
Rhodopseudomonas palustris	RPE_2235	RPE_2233
Roseobacter denitrificans OCh 114	RD1_3828	RD1_3829
Roseobacter sp. MED193	MED193_19609	MED193_19614
Roseovarius nubinhibens ISM	ISM_12605	ISM_12600
Roseovarius sp. 217	ROS217_08014	ROS217_07994
Silicibacter pomeroyi DSS-3	SPO2862	SPO2861
Silicibacter sp. TM1040	TM1040_2209	TM1040_2208
Sinorhizobium meliloti 1021	SMc04304	SMc04305*
Sulfitobacter sp. EE-36	EE36_05363	EE36_05358
Sulfitobacter sp. NAS-14.1	NAS141_10286	NAS141_10281
Burkholderia ambifaria AMMD	Bamb_1556	Bamb_1556
B. cepacia R18194	Bcep18194_A4805	Bcep18194_A4805
Burkholderia cenocepacia	Bcen_1176	Bcen_1176
Burkholderia dolosa AUO158	BDAG_01599	BDAG_01600
Burkholderia multivorans	Bmul_1583	Bmul_1584
B. cepacia R1808	Bucepa02000286	Bucepa02000285
Burkholderia vietnamiensis	Bcep1808_1601	Bcep1808_1600
Burkholderia fungorum	Bcep1785	Bcep1785
Burkholderia mallei 10399	BMA10399_E0832	BMA10399_E0831
Burkholderia xenovorans LB400	Bxe_B1647	Bxe_B1647
B. pseudomallei 1655	BURPS1655_A1847	BURPS1655_A1848*
Chromobacterium violaceum	CV1561	CV1573
Delftia acidovorans SPH-1	Daci_5747	Daci_5752
Ralstonia solanacearum GMI1000	RSp0617	RSp0628*

Table B-11. Continued.

Organism	CobW Locus Tag	first gene in operon (if operon is predicted to be regulated by riboswitch)
P. aeruginosa PAO1	PA2945	PA2945*
Pseudomonas mendocina ymp	Pmen_4587	Pmen_4587
P. entomophila L48	PSEEN2432	PSEEN2432
P. fluorescens PfO-1	Pfl_3099	Pfl_3099*
P. putida F1	Pput_2268	Pput_2268*
Pseudomonas syringae	PSPPH_2224	PSPPH_2224*
Cyanothece sp. ATCC 51142	cce_1187	–
Synechococcus elongatus	syc0712_d	–
Synechocystis sp. PCC 6803	slr0502	–
Thermosynechococcus elongatus	tll1624	–
Gloeobacter violaceus PCC 7421	gvip013	–
A. variabilis ATCC 29413	Ava_3407	–
Nostoc punctiforme PCC 73102	Npun_F0650	–
Trichodesmium erythraeum	Tery_1027	–
Magnetospirillum gryphiswaldense	MGR_0093	–
Magnetospirillum magneticum	amb4426	–
Rhodospirillum rubrum	Rru_A0671	–

Table B-12. Genomes containing subgroup 13 gene cluster.

Genome	Locus_tag
M. maripaludis S2	MMP0833
M. acetivorans C2A	MA4382 MA4381
M. barkeri str. fusaro	Mbar_A1056
M. mazei Go1	MM1072
D. hafniense DCB-2	Dhaf_0359
D. hafniense DCB-2	Dhaf_3562
Clostridium botulinum	CLB_1516
Syntrophomonas wolfei subsp. *wolfei* str. Goettingen	Swol_0421

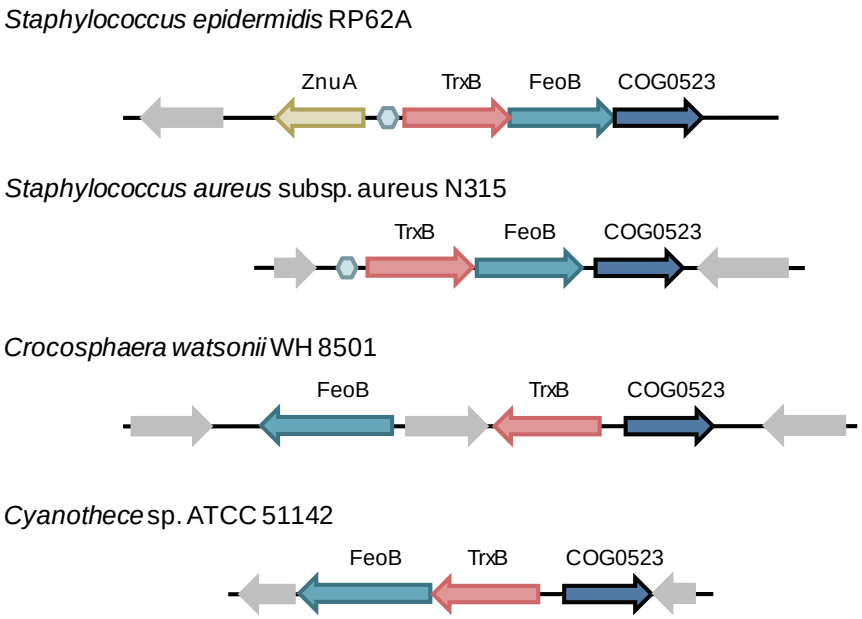
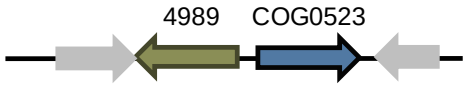


Figure B-10. Genome context of subgroup 14 members. Subgroup 14 genes co-localize with genes encoding zinc ABC transport periplasmic-binding protein (ZnuA), thioredoxin reductase (TrxB), and iron transport protein (FeoB).

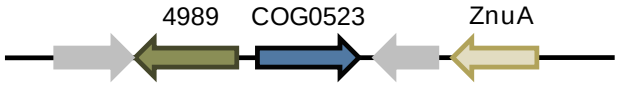
Table B-13. Genomes containing subgroup 14 gene cluster.

Genome	Locus_tag
S. aureus (13 genomes)	SAOUHSC_02901 (subsp. aureus NCTC 8325)
Staphylococcus epidermidis (2 genomes)	SE0188 (ATCC 12228)
Crocosphaera watsonii WH 8501	CwatDRAFT_5580
Cyanothece sp. ATCC 51142	Cce_4848

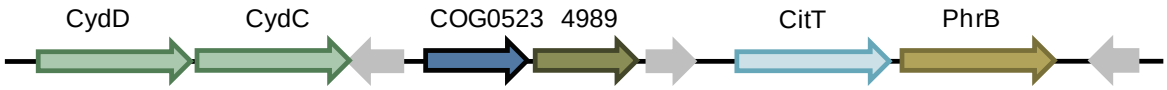
Bacillus anthracis



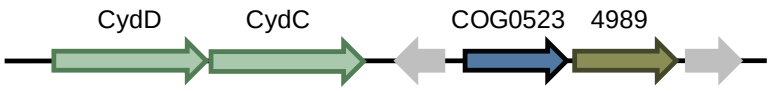
Bacillus cereus ATCC 14579 and *Bacillus thuringiensis* serovar konkukian str. 97-27



Staphylococcus aureus subsp. aureus Mu50



Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305



Staphylococcus epidermidis RP62A

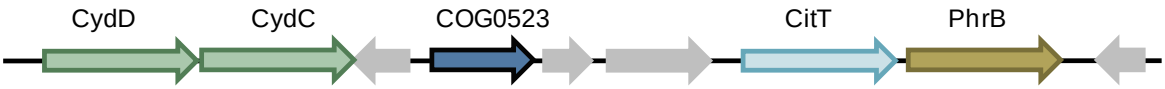


Figure B-11. Genome context of subgroup 15 members. Subgroup 15 genes co-localize with genes encoding a putative oxidoreductase COG4989 (4989), ABC-type transport system involved in cytochrome bd biosynthesis (CydD and CydC), sodium/di- and tricarboxylate cotransporter (CitT), and deoxyribodipyrimidine photolyase (PhrB).

Table B-14. Genomes containing subgroup 15 gene cluster.

Genome	Locus_tag
Bacillus anthracis (8 genomes)	BA2021 (str. Ames)
Bacillus cereus (5 genomes)	BCE2101 (str. ATCC 10987)
Bacillus thuringiensis (3 genomes)	BT9727_1849 (serovar konkukian str. 97-27)
Bacillus weihenstephanensis KBAB4	BcerKBAB4_1883
Staphylococcus haemolyticus JCSC1435	SH2206
S. aureus subsp. *aureus* Mu50	SAV0687
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	SSP2031
Staphylococcus epidermidis RP62A	SERP0344

APPENDIX C
COG0523 DISTANCE TREE

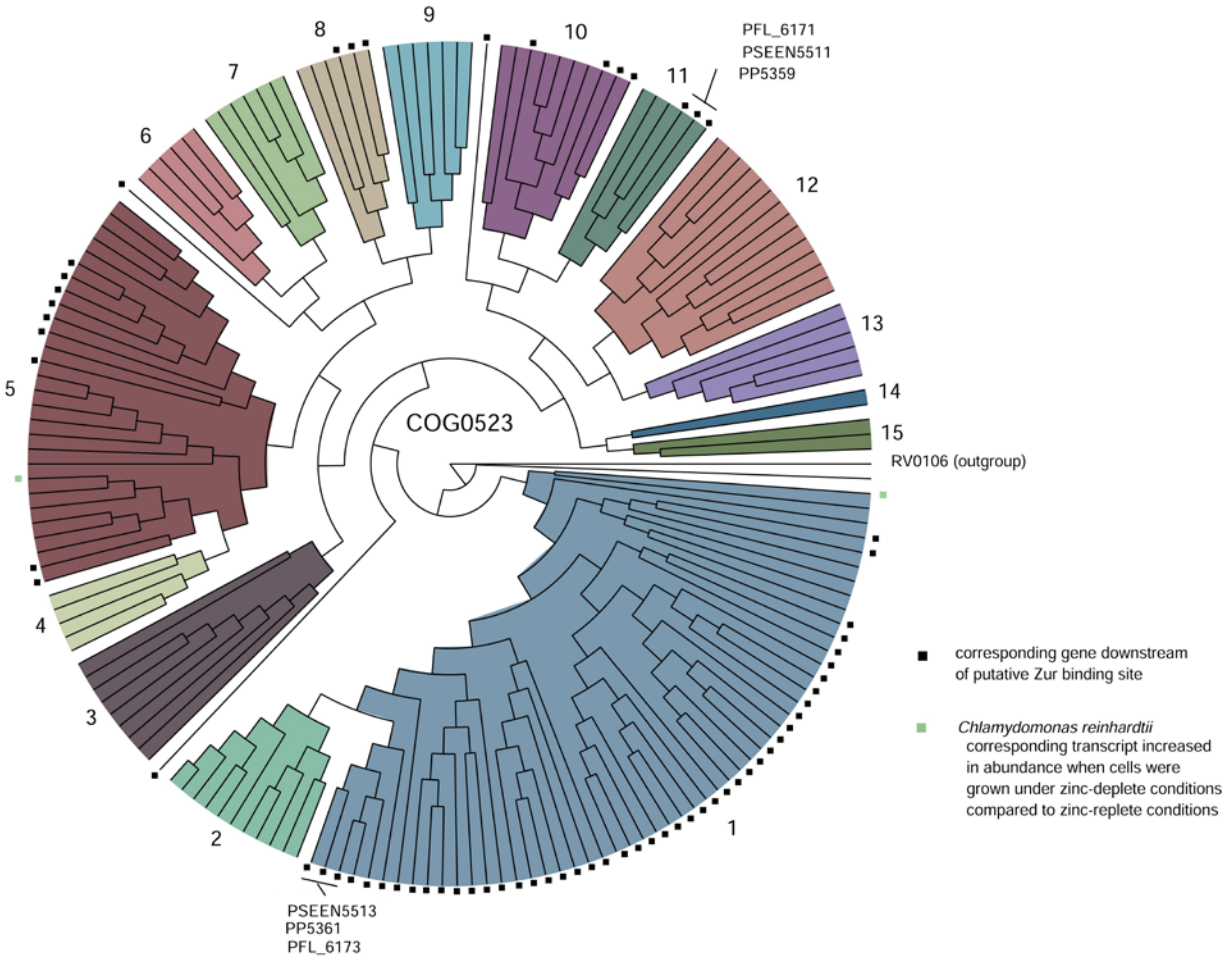


Figure C-1. Phylogenetic reconstruction of selected COG0523 proteins. Each designated subgroup is shaded and labeled. The branches representing proteins encoded by putative Zur-regulated genes are marked with a black square. The branches representing *C. reinhardtii* COG0523 homologs encoded by the genes induced by zinc deficiency are marked with a green square. Branches representing the *Pseudomonas* paralogs discussed in the text are labeled. Protein IDs for each branch can be found in Appendix D, Table D-3.

APPENDIX D
PROTEIN IDENTIFIERS AND GENE ABBREVIAITONS

Table D-1. Proteins used in COG0523 amino acid conservation analysis.

Protein	Internal identifier for SEED database (unless otherwise specified)
YciC	fig|224308.1.peg.337
ACIAD1741	fig|62977.3.peg.1665
Bcen_4084	fig|331271.3.peg.1276
nhr3	CAQ16890 (GenBank)
PMT9312_0491	fig|74546.3.peg.853
SYNW1795	fig|84588.1.peg.1788
Bcep18194_B0634	fig|269483.3.peg.168
Smb20133	fig|266834.1.peg.4768
bll7768	fig|224911.1.peg.7768
all1751	fig|103690.1.peg.2058
RL4362	fig|216596.1.peg.4497
BR2035	fig|204722.1.peg.1967
BB3253	fig|257310.1.peg.3236
PFLU1078	fig|216595.1.peg.3357
BAV0145	fig|521.1.peg.1238
Rmet_0127	fig|266264.4.peg.532
Sm4294	fig|615.1.peg.4294
PMI0144	fig|584.1.peg.507
SO1502	fig|211586.1.peg.1391
Sb2276	fig|12149.1.peg.2276
PP5359	fig|160488.1.peg.5290
PFLU6083	fig|216595.1.peg.113
PMT9312_0786	fig|74546.3.peg.895
SYNW1127	fig|84588.1.peg.1122
BR1307	fig|204722.1.peg.1267
PP3508	fig|160488.1.peg.3475
MA0856	fig|188937.1.peg.841
Desu2034	fig|49338.1.peg.1829
BT9727_1765	fig|281309.1.peg.1737
SERP2386	fig|176279.3.peg.2036
SSP2031	fig|342451.4.peg.2183
BLi00765	fig|279010.5.peg.1115
BLi01933	fig|279010.5.peg.2636
BP3084	fig|257313.1.peg.2727
SPA0639	fig|295319.3.peg.2592
Pro0851	fig|167539.1.peg.849

Table D-2. Gene abbreviations for Figure 3-2.

Subgroup	Abbreviations	Functional Roles
1	folE2	GTP cyclohydrolase I (EC 3.5.4.16) type 2
1	DUF1826	protein of unknown function
1	zur	zinc uptake regulation protein
1	dksA2	C4-type zinc finger DksA/TraR family protein
2	nhaB	Fe-type nitrile hydratase beta subunit
2	nhaA	Fe-type nitrile hydratase alpha subunit
2	amdA	amidase
3	futB	Ferric iron ABC transporter. permease protein
3	pcd	Pterin-4-alpha-carbinolamine dehydratase
3	0432	COG0432, protein of unknown function, UPF0047
3	taqCP	carboxypeptidase Taq (M32) metallopeptidase
3	ppa	inorganic pyrophosphatase
4	atzB	hydroxyatrazine ethylaminohydrolase
4	uraA	xanthine/uracil permease
4	ubiX	3-polyprenyl-4-hydroxybenzoate carboxy-lyase
4	ubiD	3-polyprenyl-4-hydroxybenzoate carboxy-lyase
4	mh	metal-dependent hydrolase
8	hslU	ATP-dependent hsl protease ATP-binding subunit
8	hslV	ATP-dependent protease
8	dksA	C4-type zinc finger DksA/TraR family protein
8	zur	zinc uptake regulation protein
8	znuC	Zinc ABC transporter, ATP-binding protein
8	znuB	Zinc ABC transporter, inner membrane permease protein
8	znuA	Zinc ABC transporter, periplasmic-binding protein
9	yjiX	protein of unknown function
9	cstA2	carbon starvation protein paralog
10	oppD	putative ATP-dependent oligopeptide permease component
10	oppC	putative ATP-dependent oligopeptide permease component
10	oppB	putative ATP-dependent oligopeptide permease component
10	oppA	putative ATP-dependent oligopeptide permease component
10	rtn	putative EAL domain-containing lipoprotein
10	spr	predicted peptidase
10	lpxT	undecaprenyl pyrophosphate phosphatase
11	DUF1826	protein of unknown function
12	prp	pentapeptide repeat protein
12	upp	uracil phosphoribosyltransferase
12	ilvD	dihydroxyacid dehydratase/phopshogluconate dehydratase
13	cbiH	Cobalt-precorrin-6x reductase
13	cbiL	Cobalt-precorrin-3b C17-methyltransferase
13	cbiC	cobalt-precorrin-8x methylmutase
13	cobG	Precorrin-3B synthase
13	cobN	cobaltochelatase subunit
13	cobU	Adenosylcobinamide-phosphate guanylyltransferase
13	pduO	Cob(I)alamin adenosyltransferase
13	cbiP	Cobyric acid synthase
14	ramA	iron-sulfur protein that mediates the ATP-dependent reductive activation of Co(II) corrinoid to the Co(I) state
14	mtbA	Methylcobalamin:coenzyme M methyltransferase, methylaminie-specific
14	mtaA	Methylcobalamin:coenzyme M methyltransferase, methanol-specific
14	mtaC	corrinoid-binding protein
15	4989	putative oxidoreductase COG4989
16	znuA	Zinc ABC transporter, periplasmic-binding protein

Table D-3. Proteins used in COG0523 phylogenetic tree reconstruction. The proteins are listed in order that they appear in the COG0523 distance tree found in Appendix C starting with RV0106 and moving clockwise around the tree.

Protein	Internal identifier for SEED database (unless otherwise specified)	Protein	Internal identifier for SEED database (unless otherwise specified)
RV0106 (outgroup)	fig|83332.1.peg.106	101629	*C. reinhardtii* draft genome v. 3.0
520982	*C. reinhardtii* draft genome v. 4.0	At1g26520	fig|3702.1.peg.2968
123019	*C. reinhardtii* draft genome v. 3.0	YNR029c	fig|4932.3.peg.5506
73360	*C. reinhardtii* draft genome v. 3.0	*Taeniopygia guttata*	XP_002196540 (Genbank)
106402	*C. reinhardtii* draft genome v. 3.0	*Rattus norvegicus*	fig|10116.3.peg.3919
Rmet_0125	fig|266264.4.peg.530	CBWD1	fig|9606.3.peg.30962
Reut_A0161	fig|264198.3.peg.673	CBWD2	fig|9606.3.peg.18530
RSc0045	fig|267608.1.peg.45	all1751	fig|103690.1.peg.2058
BURPS1710b_0218	fig|320372.3.peg.2822	SYNW2482	fig|84588.1.peg.2473
BCAL3529	fig|216591.1.peg.525	SMc03799	fig|266834.1.peg.4476
Bucepa02004096	fig|269482.1.peg.3952	CC0321	fig|190650.1.peg.320
XCC0257	fig|190485.1.peg.257	bll7768	fig|224911.1.peg.7768
XAC0276	fig|190486.1.peg.276	RPA0861	fig|258594.1.peg.858
XF1830	fig|160492.1.peg.1823	NB311A_16784	fig|314253.3.peg.2776
BH1790	fig|272558.1.peg.1790	Nham_3425	fig|323097.3.peg.3530
BT9727_1611	fig|281309.1.peg.1582	BH12980	fig|283166.1.peg.1162
BCE1836	fig|222523.1.peg.1825	BRA0987	fig|204722.1.peg.3060
BLO5309	fig|279010.5.peg.800	Meso_2842	fig|266779.1.peg.1374
ABC4020	fig|66692.3.peg.3995	SMc02978	fig|266834.1.peg.4164
OB3429	fig|221109.1.peg.3433	RHE_CH03625	ABC92380(Genbank)
BLO2348	fig|279010.5.peg.1115	Mfla_1230	fig|265072.7.peg.1253
YciC	fig|224308.1.peg.337	Rru_A1611	fig|1085.1.peg.2756
SH0144	fig|279808.3.peg.157	Meso_3058	fig|266779.1.peg.225
SA0410	fig|158879.1.peg.422	BR2035	fig|204722.1.peg.1967
SERP0080	fig|176279.3.peg.352	mll3580	fig|266835.1.peg.2760
FTT1000c	fig|177416.3.peg.1070	SMc00684	fig|266834.1.peg.3981
BMEII0179	fig|224914.1.peg.2238	RL4362	fig|216596.1.peg.4497
Atu3181	fig|176299.3.peg.3142	BB3253	fig|257310.1.peg.3236
RHE_CH02713	ABC91484 (Genbank)	BAV2401	fig|521.1.peg.71
RL3179	fig|216596.1.peg.3284	pRL120138	fig|216596.1.peg.5848
GOX2212	fig|290633.1.peg.2148	PSPPH_3929	fig|264730.3.peg.4199

Table D-3. Continued.

Protein	Internal identifier for SEED database (unless otherwise specified)	Protein	Internal identifier for SEED database (unless otherwise specified)
TM1040_0080	fig|292414.1.peg.1132	PFLU1078	fig|216595.1.peg.3357
ROS217_16055	fig|314264.3.peg.2684	PFL_1367	fig|220664.3.peg.3069
ISM_00325	fig|89187.3.peg.72	fig|615.1.peg.1489	fig|615.1.peg.1489
Mfla_1231	fig|265072.7.peg.1254	BAV0145	fig|521.1.peg.1238
PP3323	fig|160488.1.peg.3296	BURPS1710b_0389	fig|320372.3.peg.3067
LV196	fig|573.1.peg.161	Bcep18194_A6439	fig|269483.3.peg.7256
Rmet_1098	fig|266264.4.peg.1503	RSc0047	fig|267608.1.peg.47
Bcen_3034	fig|331271.3.peg.590	Reut_A0163	fig|264198.3.peg.675
Bcep18194_B0316	fig|269483.3.peg.318	Rmet_0127	fig|266264.4.peg.532
ACIAD1741	fig|62977.3.peg.1665	PA4604	fig|208964.1.peg.4604
ABO_1679	fig|393595.12.peg.1684	PFLU5331	fig|216595.1.peg.7492
fig|615.1.peg.843	fig|615.1.peg.843	PMI0144	fig|584.1.peg.507
Csal_0192	fig|290398.4.peg.672	fig|615.1.peg.4294	fig|615.1.peg.4294
V12G01_06993	fig|314288.3.peg.2404	fig|54388.1.peg.219	fig|54388.1.peg.219
Mmwyl1_1446	fig|400668.6.peg.1471	STM4530	fig|99287.1.peg.4358
PatlDRAFT_1639	fig|342610.3.peg.1415	KPN_04774	fig|573.2.peg.3898
MED297_11295	fig|314283.3.peg.895	CV3067	fig|243365.1.peg.3067
HCH_02800	fig|349521.5.peg.2494	SO1502	fig|211586.1.peg.1391
PA5535	fig|208964.1.peg.5532	Shew_2776	fig|323850.3.peg.814
Avin1667	fig|354.1.peg.3168	HCH_01308	fig|349521.5.peg.1191
Psyr_5075	fig|205918.4.peg.275	fig|198214.1.peg.2112	fig|198214.1.peg.2112
PFL_6173	fig|220664.3.peg.101	fig|12149.1.peg.2276	fig|12149.1.peg.2276
PP5361	fig|160488.1.peg.5292	YPTB1311	fig|273123.1.peg.1426
PSEEN5513	fig|384676.6.peg.4780	VV20385	fig|216895.1.peg.3313
nhr	BAD98534 (Genbank)	VF2370	fig|312309.3.peg.2370
P47K	P31521 (Genbank)	MED297_11310	fig|314283.3.peg.898
ACIAD1614	fig|62977.3.peg.1431	V12G01_06973	fig|314288.3.peg.2400
Bcen_4084	fig|331271.3.peg.1276	PA5532	fig|208964.1.peg.5529
Bamb_6542	fig|339670.3.peg.5927	PSPPH_5154	fig|264730.3.peg.5102
RHA1_ro00362	ABG92198 (Genbank)	PFLU6083	fig|216595.1.peg.113
nhr3	CAQ16890 (Genbank)	PFL_6171	fig|220664.3.peg.99
nhr3	BAC99082 (Genbank)	PSEEN5511	fig|384676.6.peg.4778

Table D-3. Continued.

Protein	Internal identifier for SEED database (unless otherwise specified)	Protein	Internal identifier for SEED database (unless otherwise specified)
nha3	CAG29801 (Genbank)	PP5359	fig|160488.1.peg.5290
nha3	BAA36599 (Genbank)	PMT9312_0786	fig|74546.3.peg.895
ORF1188	BAA06274 (Genbank)	Syncc9902_1218	fig|316279.3.peg.2255
V12G01_06988	fig|314288.3.peg.2403	Syncc9605_1264	fig|110662.3.peg.233
PMT9312_0491	fig|74546.3.peg.853	SYNW1127	fig|84588.1.peg.1122
Ava_3717	fig|240292.3.peg.4870	PMT0556	fig|74547.1.peg.556
Pro0489	fig|167539.1.peg.488	Pro0851	fig|167539.1.peg.849
PMT1284	fig|74547.1.peg.1279	PP3508	fig|160488.1.peg.3475
Syncc9902_1690	fig|316279.3.peg.2158	BMA1171	fig|243160.4.peg.121
SYNW1795	fig|84588.1.peg.1788	BURPS1655_A1847	fig|331109.3.peg.1603
Syncc9605_0672	fig|110662.3.peg.1837	Bcep1808_1601	fig|269482.1.peg.277
Tery_4617	fig|203124.1.peg.4333	BR1307	fig|204722.1.peg.1267
Ava_1948	fig|240292.3.peg.3028	RL2831	fig|216596.1.peg.2928
BCAM2270	fig|216591.1.peg.4280	SMc04304	fig|266834.1.peg.3252
Bcep18194_B0634	fig|269483.3.peg.168	blr3261	fig|224911.1.peg.3262
Bmul_3607	fig|395019.3.peg.3725	143868	*C. reinhardtii* draft genome v. 3.0
SMb20133	fig|266834.1.peg.4768	Desu2034	fig|49338.1.peg.1829
Reut_A3078	fig|264198.3.peg.2234	Desu0949	fig|49338.1.peg.2656
EF3204	fig|226185.1.peg.2991	MM1072	fig|192952.1.peg.1072
OB3433	fig|221109.1.peg.3437	Meth3844	fig|2208.1.peg.3706
106748	*C. reinhardtii* draft genome v. 3.0	MA0856	fig|188937.1.peg.841
At1G80480	fig|3702.1.peg.7636	SERP2386	fig|176279.3.peg.2036
At1G15730	fig|3702.1.peg.1879	BT9727_1765	fig|281309.1.peg.1737
195946	*C. reinhardtii* draft genome v. 3.0	SAV0687	fig|158878.1.peg.687
117458	*C. reinhardtii* draft genome v. 3.0	SERP0344	fig|176279.3.peg.657
122261	*C. reinhardtii* draft genome v. 3.0	SSP2031	fig|342451.4.peg.2183

APPENDIX E
ANALYSIS OF PUTATIVELY ZUR-REGULATED PARALOGS

AmiA

AmiA catalyzes amide bond cleavage between the lactyl group of muramic acid and the amino group of l-alanine (Lupoli *et al.*, 2009). In *E. coli*, the three paralogs AmiA, AmiB, and AmiC are responsible for splitting the murein septum during cell division (Heidrich *et al.*, 2001). The *amiA* paralogs found in Zur regulons share significant sequence similarity with *amiA*, *amiB* and *amiC* including presence of the N-terminal domain responsible for targeting the protein to the periplasm and localization at the septal ring (Figure E-1A) (de Souza *et al.*, 2008). A crystal structure of the AmiA homolog CwlV from *Paenibacillus polymyxa* is available (PDB: 1JWQ) (Ishikawa *et al.*, 1999). The catalytic zinc ion is chelated by two His and one Glu residues. These residues are conserved in the AmiA paralogs whose corresponding genes are putatively regulated by Zur (Figure E-1B).

HisI

HisI catalyzes the third step in histidine biosynthesis, hydrolysis of the phosphoribosyl-AMP adenine ring (D'Ordine *et al.*, 1999). The proteins of the *hisI* paralogs found in Zur regulons of γ-proteobacteria are closely related to the HisI homologs from α-proteobacteria (Figure E-2A). The only crystal structure available for a HisI homolog is from *Methanobacterium thermoautotrophicum*. The crystallization methods included saturation with cadmium, so the zinc-binding residues were deduced from the cadmium-binding residues. These residues are conserved in the HisI paralogs analyzed (Figure E-2B).

CysRS and ThrRS

CysRS and ThrRS are aminoacyl tRNA-synthetases for cysteine and threonine, respectively. For both proteins, the zinc ion is essential in amino acid discrimination (Zhang *et al.*, 2003b, Sankaranarayanan *et al.*, 2000). CysRS proteins whose genes belong to the putative Zur regulon in *C. metallidurans* and *B. pseudomallei* spp. have highest similarity to the CysRS orthologs from α-proteobacteria (Figure E-3A). The zinc-binding residues observed in the crystal structure of *E. coli* CysRS (PDB: 1LI5) are conserved in these paralogs (Figure E-3B).

thrS genes are found in the putative Zur regulons of cynobacterial and proteobacterial genomes (Figure E-4A). The zinc-binding residues observed in the *E. coli* ThrRS (PDB: 1QF6) are also conserved in the ThrRS paralogs, whose genes are putatively regulated bv Zur (Figure E-4).

FolE

Several cyanobacterial genomes, *Cyanothece* sp. ATCC 51142*, Cyanothece* sp. PCC 8801*, Microcystis aeruginosa, Nostoc* sp. PCC 7120*, Synechococcus* sp. PCC 7002*, Trichodesmium erythraeum,* and *A. variabilis*, contain at least two *folE* genes, which are proposed to encode GTP cyclohydrolase I Type A. One of these *folE* genes is predicted to be regulation by Zur. However, the zinc-binding ligands observed in the FolE protein from *H. sapiens* are conserved in these proteins (Figure E-5).

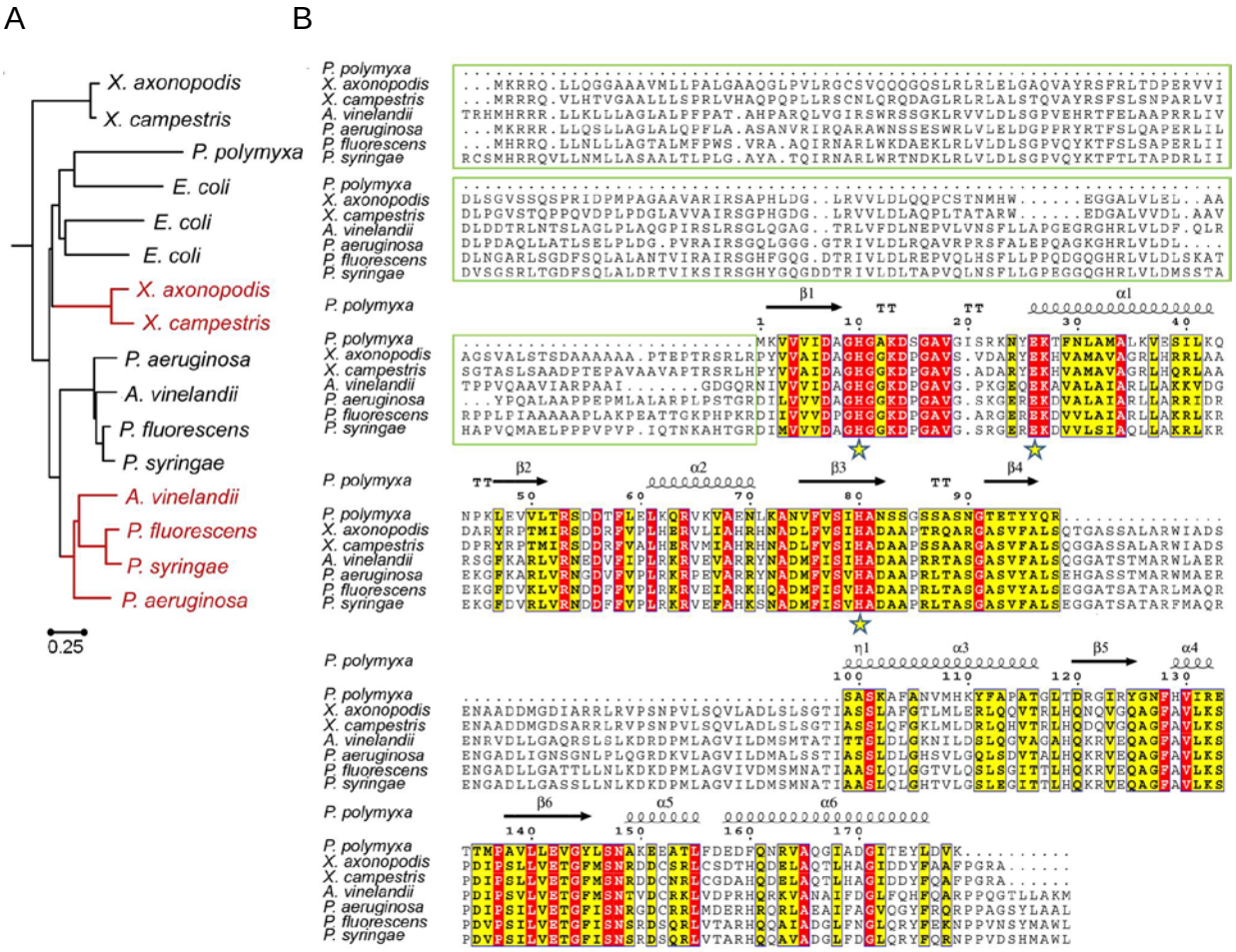


Figure E-1. Phylogenetic and sequence analysis of the AmiA paralogs. A) Neighbor-joining tree of AmiA protein sequences. Red indicates that the corresponding gene is putatively regulated by Zur. B) Protein sequence alignment of AmiA proteins whose corresponding genes are putatively regulated by Zur with CwlV from *P. polymyxa*. The metal-binding residues are marked with a star. The N-terminal periplasmic targeting domain is marked with a green box. Secondary structural elements were derived from the crystal structure of *P. polymyxa* CwlV (PDB: 1JWQ).

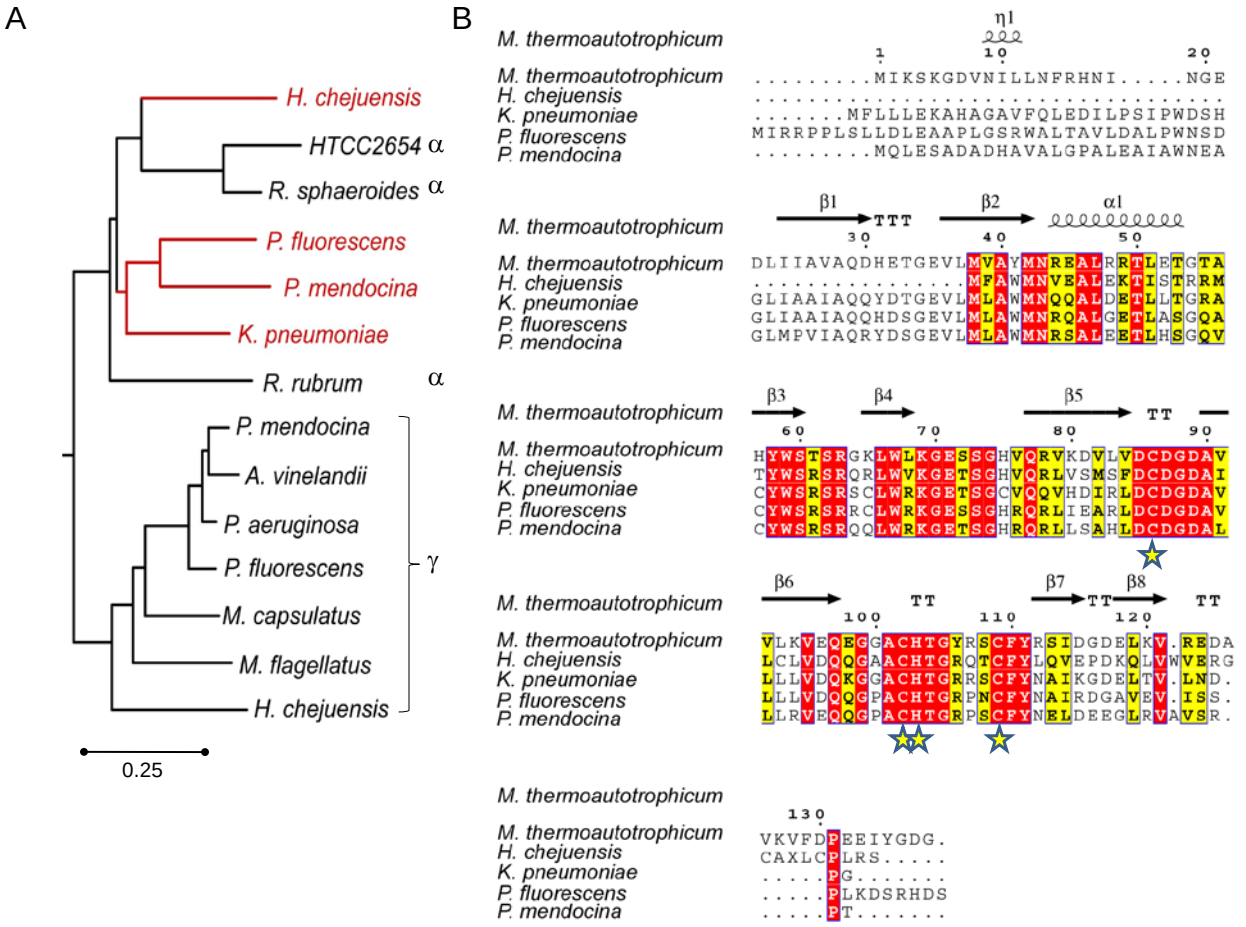
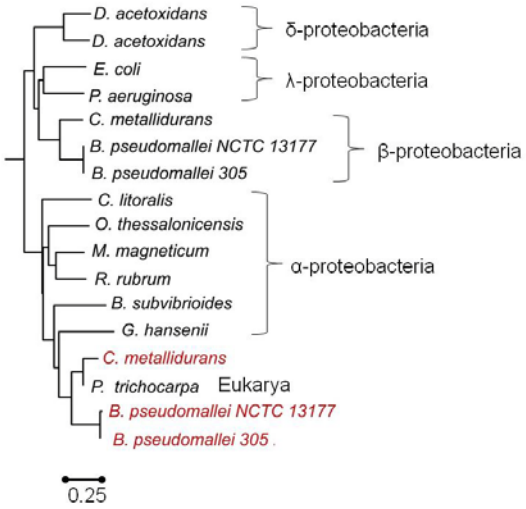


Figure E-2. Sequence and phylogenetic analysis of HisI paralogs. A) Neighbor-joining tree of HisI protein sequences. Red indicates that the corresponding gene is putatively regulated by Zur. The class of proteobacteria to which the genome belongs in shown on the right. B) Protein sequence alignment of HisI proteins whose corresponding genes are putatively regulated by Zur with HisI from *M. thermoautotrophicum*. The metal-binding residues are marked with a star. Secondary structural elements were derived from the crystal structure of *M. thermoautotrophicum* HisI (PDB: 1ZPS).

A



B

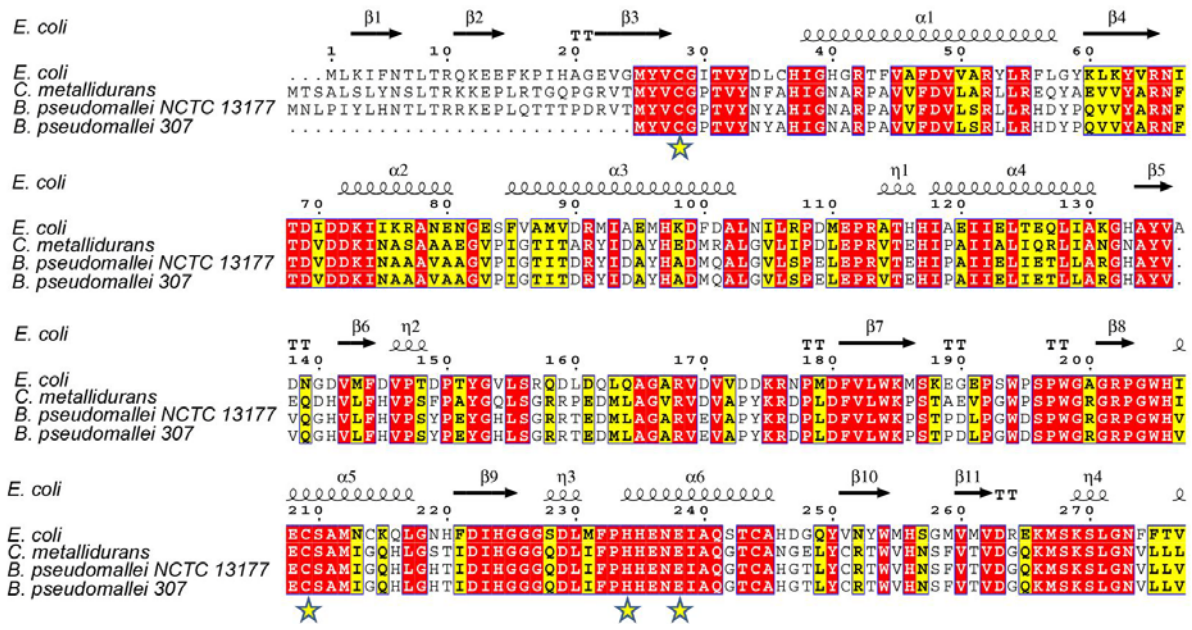
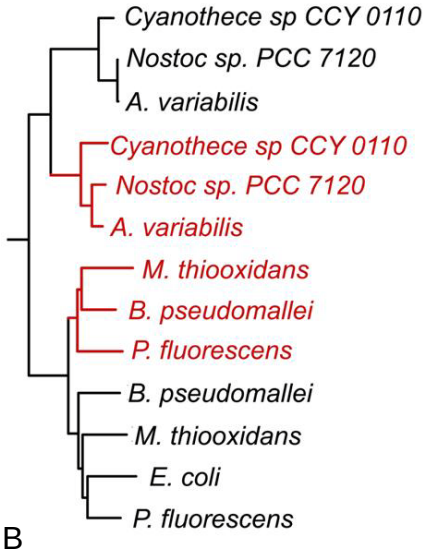


Figure E-3. Sequence and phylogenetic analysis of CysRS paralogs. A) Neighbor-joining tree of CysRS protein sequences. Red indicates that the corresponding gene is putatively regulated by Zur. The class of proteobacteria to which the genome belongs in shown on the right. B) Protein sequence alignment of CysRSproteins whose corresponding genes are putatively regulated by Zur with CysRS from *E. coli*. The metal-binding residues are marked with a star. Secondary structural elements were derived from the crystal structure of *E. coli* CysRS (PDB: 1LI5).

A



B

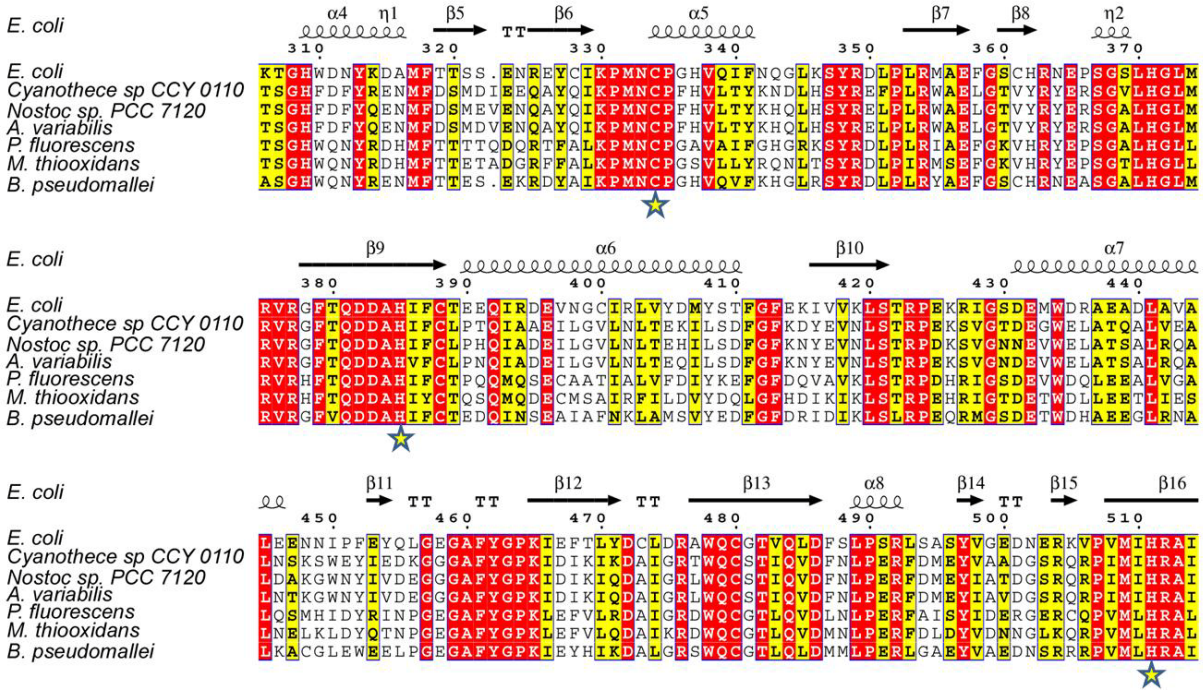


Figure E-4. Sequence and phylogenetic analysis of ThrRS paralogs. A) Neighbor-joining tree of ThrRS protein sequences. Red indicates that the corresponding gene is putatively regulated by Zur. B) Protein sequence alignment of ThrRS proteins whose corresponding genes are putatively regulated by Zur with ThrRS from *E. coli*. The metal-binding residues are marked with a star. Secondary structural elements were derived from the crystal structure of *E. coli* ThrRS (PDB:1EVK).

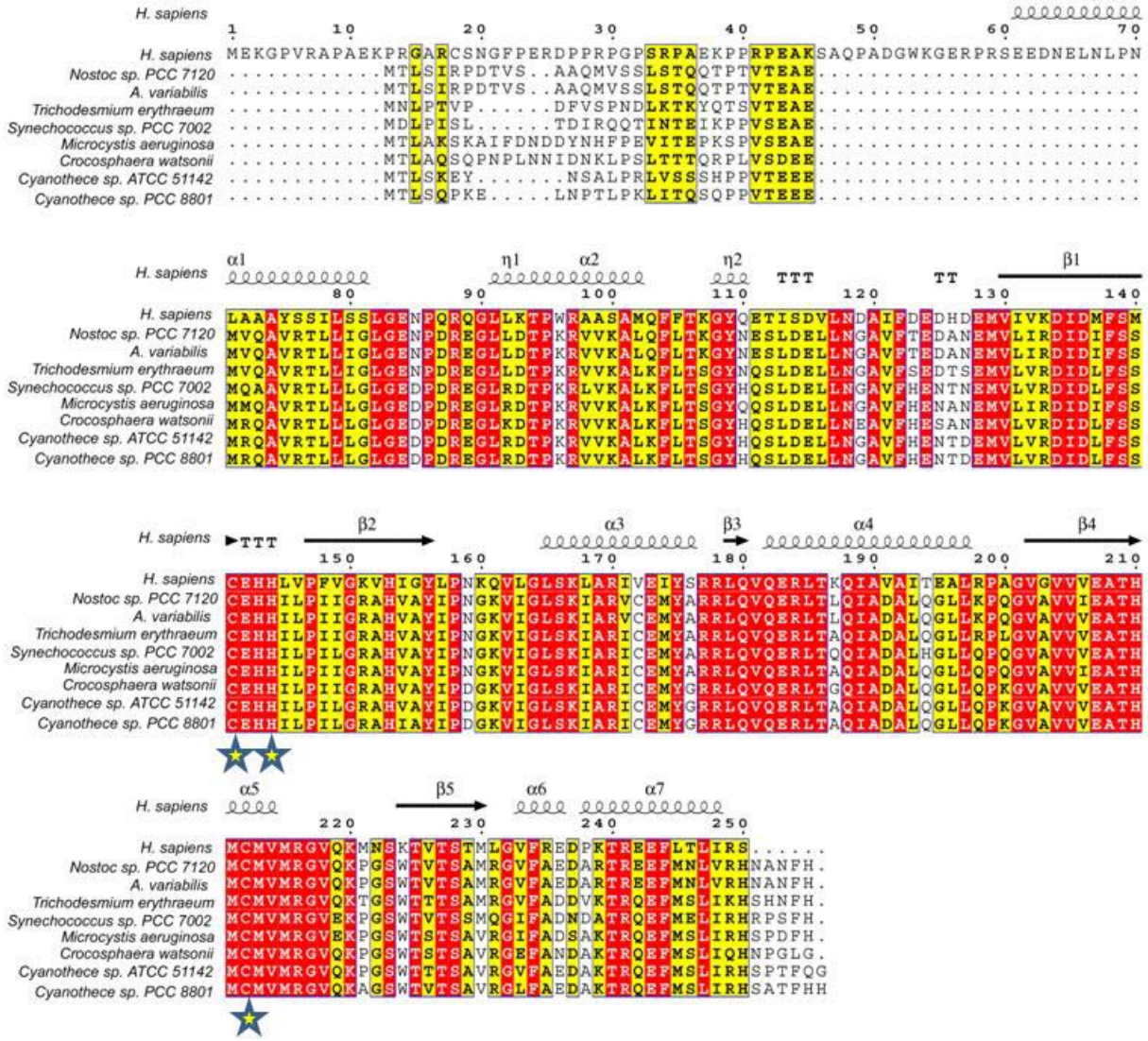


Figure E-5. Amino acid sequence comparison between the FolE from *H. sapiens* and cyanobacterial genomes. The cyanobacterial proteins are predicted to be encoded by Zur-regulated genes. The zinc-binding ligands observed in the crystal structure of human FolE (PDB: 1FB1) are pointed-out with a star.

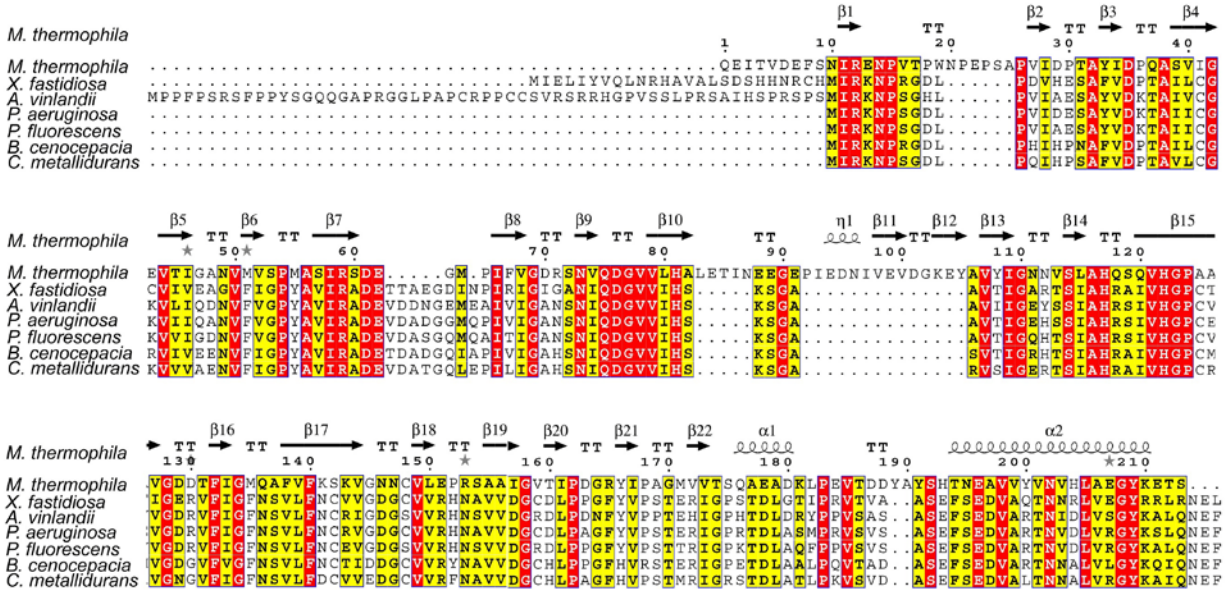


Figure E-6. Amino acid sequence alignment between the γ -class carbonic anhydrases. Alignment between Cam from *M. thermophila* and the carbonic anhydrase proteins, whose genes are found in the putative Zur regulons of *Xyella fastidiosa, A. vinelnadii, P. aeruginosa, P. fluorescens, B. cenocepacia* and *C. metallidurans*. Secondary structural elements were derived from the crystal structure of *M. thermophila* Cam (PDB: 1QRE).

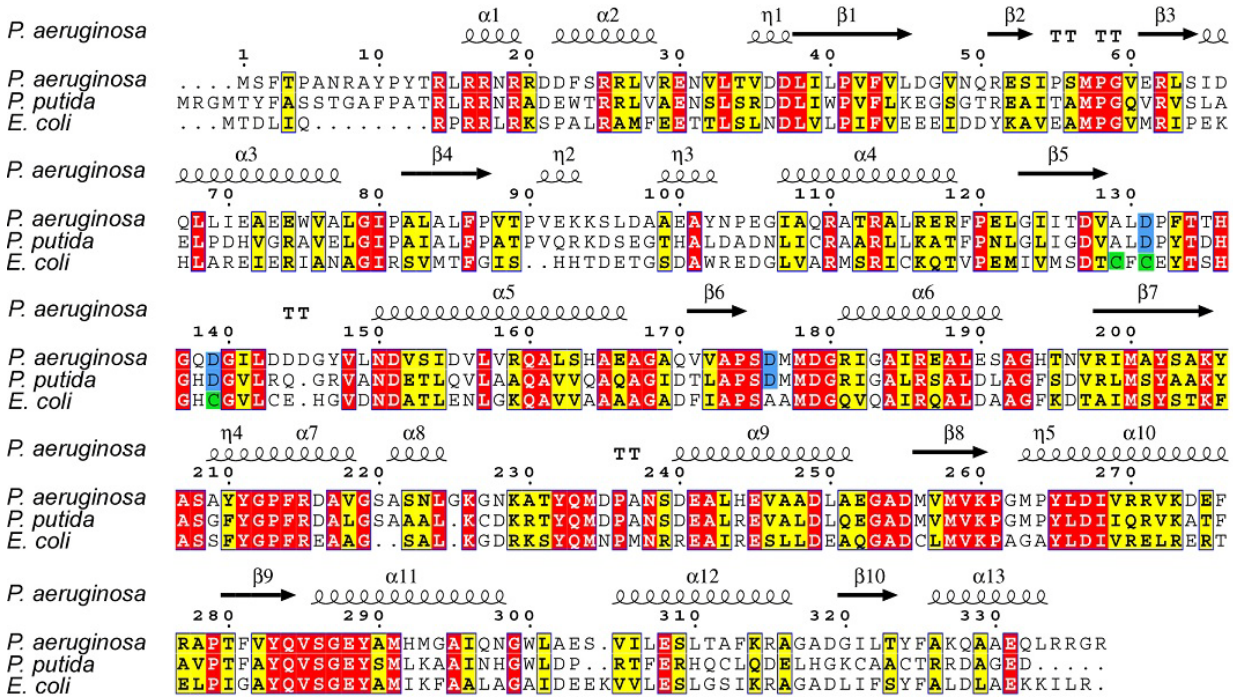


Figure E-7. Protein sequence alignment HemB proteins. Alignment of the *P. putida* HemB2 with the Mg(II)-dependent HemB from *P. aeruginosa* and the Zn(II)-dependent HemB from *E. coli*. Secondary structural elements were derived from the crystal structure of *P. aeruginosa* HemB (PDB: 1B4K). The respective metal-binding residue are highlighted: blue boxes outline the Mg-binding residues observed in the crystal structure of *P. aeruginosa* HemB (PDB: 1B4K); green boxes outline the Zn-binding residues observed in the crystal structure of *E. coli* HemB (PDB: 1I8J).

APPENDIX F
STRAINS, PLASMIDS AND OLIGONUCLEOTIDES

Table F-1. Strains used in Chapter 3.

Strain	Genotype	Source
E. coli K12 MG1655	F, λ, *rph1*	*E. coli* Genetic Stock Center
E. coli BW25113 JW2161	Δ*yeiR*::kan	(Baba *et al.*, 2006)
E. coli GR352	Δ*znuABC*::cam	(Grass *et al.*, 2002)
VDC4286	MG1655 Δ*yeiR*::kan	This study
VDC4311	MG1655 Δ*yeiR*	This study
VDC4320	VDC4311 pBAD24	This study
VDC4325	VDC4311 pCH011	This study
VDC4352	VDC4311 pCH090	This study
VDC4358	MG1655 pBAD24	This study
VDC4361	VDC4311 Δ*znuABC*::cam	This study
VDC4683	VDC4311 pCH141	This study
VDC4684	VDC4311 pCH137	This study
VDC4685	VDC4311 pCH138	This study
VDC4686	VDC4311 pCH145	This study
VDC4705	VDC4361 pBAD24	This study
VDC4706	VDC4361 pCH011	This study
VDC4713	MG1655 Δ*znuABC*::cam	This study
VDC4716	VDC4713 pBAD24	This study
VDC4721	VDC4311 pCH150	This study
VDC4732	VDC4311 pCH151	This study

Table F-2. Strains used in Chapter 4.

Strain	Genotype	Source
E. coli K12 MG1655	F, λ, rph1	E. coli K12 MG1655
P. aeruginosa PAO1	Wild-type strain	*P. aeruginosa* type strain (ATCC 33351)
VDC4098	MG1655 Δ*dksA::*Tet^R	This study
VDC4133	VDC4098 pBAD24	This study
VDC4241	VDC4098 pCH078	This study
VDC4263	VDC4098 pCH080	This study
VDC4614	VDC4098 pCH040	This study
VDC4239	VDC4098 pCH071	This study
VDC4240	VDC4098 pCH075	This study
VDC4489	PAO1 Δ*PA4723*	This study
VDC4525	VDC4489 Δ*PA5536::*Gm^R	This study
VDC4499	VDC4489 Δ*PA5499*	This study
VDC4510	VDC4499 Δ*PA5536::*Gm^R	This study
VDC4607	PAO1 pHERD20T	This study
VDC4666	VDC4489 pHERD20T	This study
VDC4667	VDC4489 pCH115	This study
VDC4610	VDC4525 pHERD20T	This study
VDC4611	VDC4525 pCH115	This study
VDC4612	VDC4510 pHERD20T	This study
VDC4613	VDC4510 pCH115	This study
VDC4640	PAO1 Δ*PA5499::*Gm^R	This study
VDC4615	PAO1 Δ*PA5536*	This study
VDC4635	VDC4615 Δ*PA5498*:: Gm^R	This study
VDC4650	Δ*PA5498*:: Gm^R pHERD20T	This study
VDC4652	VDC4635 pHERD20T	This study
VDC4653	VDC4635 pCH115	This study

Table F-3. Plasmids used in Chapter 3.

Plasmid	Description	Reference
pBAD24	*E. coli* expression vector, Amp^R	(Guzman *et al.*, 1995)
pCH011	*yeiR* cloned into the *Nco*I/*Xba*I sites of pBAD24	This study
pCH090	Product of YeiR C63A mutagenesis cloned into pBAD24	This study
pCH137	Product of YeiR C65A mutagenesis cloned into pBAD24	This study
pCH138	Product of YeiR C66A mutagenesis cloned into pBAD24	This study
pCH140	Product of YeiR H207A mutagenesis cloned into pBAD24	This study
pCH141	Product of YeiR M64A mutagenesis cloned into pBAD24	This study
pCH145	Product of YeiR H209A mutagenesis cloned into pBAD24	This study
pCH150	Product of YeiR H207A H209A H211A mutagenesis cloned into pBAD24	This study
pCH151	Product of YeiR H211A mutagenesis cloned into pBAD24	This study

Table F-4. Plasmids used in Chapter 4.

Plasmid	Description	Reference
pBAD24	*E. coli* expression vector, Amp^R	(Guzman *et al.*, 1995)
pHERD20T	*P. aeruginosa* shuttle vector, Amp^R	(Qiu *et al.*, 2008)
pCH078	*PA4723* cloned into *Nco*I/*Xba*I sites of pBAD24	This study
pCH080	EC *dksA* cloned ineto *Nco*I/*Xba*I sites of pBAD24	This study
pCH040	*PA5536* cloned into *Nco*I/*Xba*I sites of pBAD24	This study
pCH071	Product of PA4723 C114T mutagenesis cloned into *Nco*I/*Xba*I sites of pBAD24	This study
pCH075	Product of PA4723 C135A mutagenesis cloned into *Nco*I/*Xba*I sites of pBAD24	This study
pCH115	*PA5536* cloned into *Nco*I/*Xba*I sites of pHERD20T	This study
pCH103	pEX18Tc derivative carrying *PA4723* deletion construct	This study
pCH107	pEX18Tc derivative carrying *PA5536* deletion construct	This study
pCH108	pEX18Tc derivative carrying *PA5499* deletion construct	This study
pCH131	pEX18Tc derivative carrying *PA5498* deletion construct	This study

Table F-5. Oligonucleotides used in Chapter 3.

Name	Sequence (written 5' to 3')
chkout*yeiR*outFor	CATCGAACAAGCCTGCAAC
chkout*yeiR*outRev	ATCAATCGGCAACCAGAATCC
chkin*yeiR*For	CGGTAGTCGCTAAACGAAGC
chkin*yeiR*For	ACCGCACCAGTCTATGAACC
chkout*znu*outFor	GGCGATTTTGTCATCCAGTT
chkout*znu*outRev	AATTGTTGCTGGCGGTAATC
chkin*znu*For	ACATGCATCTTTGGCTTTCC
chkinznuRev	CTCGTTACCAACCTGCGTTT
*yeiR*For*Nco*I	TAACCATGGCCACCAGGACCAACCT
*yeiR*Rev*Xba*I	CGTCTAGATTACGCGGTAGTCGCTAA
YeiRC63AFor	GAGATCCCCGGCGGCGCCATGTGCTGCGTTAATGG
YeiRC63ARev	CCATTAACGCAGCACATGGCGCCGCCGGGGATCTC
YeiRC65AFor	GCGGCTGCATGGCCTGCGTTAATGG
YeiRC65ARev	CCATTAACGCAGGCCATGCAGCCGC
YeiRC66AFor	CGGCTGCATGTGCGCCGTTAATGGTTTAC
YeiRC66ARev	GTAAACCATTAACGGCGCACATGCAGCCG

Table F-6. Oligonucleotides used in Chapter 4.

Oligonucleotide	Sequence (written 5' to 3')
dksAFor*Nco*I	CCATGGAAGAAGGGCAAAACCGT
dksARev*XbaI*	TCTAGATTAGCCAGCCATCTGTTTTTC
PA4723For*Nco*I	CCATGGCCACCAAAGCAAAACAACA
PA4723Rev*Xba*I	TCTAGATCAGGAGCCGAGTTGCTTCTC
PA5536For*Nco*I	TATATACCATGGCCGAACAGGAACTGCTTGCC
PA5536Rev*Xba*I	TATTTATCTAGATCAGTTGTGCCGCACGTG
PA4723mutC114	CGACTCCACCGGCGTC
PA4723mutC135	CATCGACGCGAAGACCC
ExternalFor	CGACTCACTATAGG
ExternalRev	GAATTCGCCCTTCCATGGCCACCAAAGC
FRT/GmFor	GAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAG GAACTTCTTAGGTGGC GGTACTTGGGT
FRT/GmRev	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAG GAACTTCAGCGGTGGTAACGGCGCA
5'PA4723For*EcoR*I	TTTTTGAATTCATGACATGGAAAGGCTCGAT
5'PA4723RevFRT	CTTCGGAATAGGAACTTCTTTGCTTTGGTGGACATGAA
3'PA4723ForFRT	AGAAAGTATAGGAACTTCAAATCCGCGAGAAGCAACT
3'PA4723Rev*Hind*III	TTTTTAAGCTTAGCAAGTGCAGGCGTAGG
5'PA5536For*EcoR*I	TTTTTGAATTCCGTAGCGCGTATTGGTCTG
5'PA5536RevFRT	CTTCGGAATAGGAACTTCAGCAGTTCCTGTTCGGTCAT
3'PA5536ForFRT	AGAAAGTATAGGAACTTCCGGCACAACTGAAGAGGTTA
3'PA5536Rev*Hind*III	TTTTTAAGCTTGGTGGACTCAATCAGC AGGT
5'PA5499For*EcoR*I	TTTTTGAATTCCCTCTTCCTCGAAATTGACG
5'PA5499RevFRT	CTTCGGAATAGGAACTTCGTCTTGGGCGCAATCTTGTA
3'PA5499ForFRT	AGAAAGTATAGGAACTTCGGACGCCTGATGGACAAC
3'PA5499Rev*Hind*III	TTTTTAAGCTTATGAGGCGGTAGAGTTCAGC
5'PA5498For*EcoR*I	TTTTTGAATTCCTGATATCCGGCTGTTCCAG
5'PA5498RevFRT	CTTCGGAATAGGAACTTCGGCGGCACTCACATGAAAG
3'PA5498ForFRT	AGAAAGTATAGGAACTTCAAACCCGTGACTGATGGC
3'PA5498Rev*Hind*III	TTTTTAAGCTTACACGGTGCAGTTGATGGT
chkECdksAFor	CGATAGTGCGTGTTAAGGAG
chkECdksARev	CGTGATGGAACGGCTGTAAT
chkoutPA4723For	GAGATCCGGTGTCAGTTGGT
chkoutPA4723Rev	GCCAGTTGGTAGGCGTAGAG
chkoutPA5536For	GTATATGCCCTCTCCGAGCA
chkoutPA5536Rev	GGAGATCAGGTCGATCTTGC
chkoutPA5499For	TCGTGGAAGACGAAGAAAGG
chkoutPA5499Rev	GTGCTTGTGGACAGGAGTGA
chkinPA5499For	GGTGGTCCATGGACAAGATTGCGCCCAAGAC
chkinPA5499Rev	GGTGGTCTGCAGTCAGGCGTCCTTCTGGTCCC
chkoutPA5498For	CGATCAGGGTGACGATCTG
chkoutPA5498Rev	AGGGTGCTGAACTCATAGCC
chkinPA5498For	GAGAACTTCCTGCCCAAGGT

Table F-6. Continued.

Oligonucleotide	Sequence (written 5’ to 3’)
chkinPA5498Rev	CCTCTTCCTCGAAATTGACG
PA4723RTFor	TCCACCAAAGCAAAACAACA
PA4723RTRev	TACGGTCGACCTCTTCCATC
PA5536RTFor	AAGCCCAGCAGGACTTCTTC
PA5536RTRev	TGTCGAGCAGCTTCTTCTCC
PA5499For*Nde*I	GGTGGTCATATGTACAAGATTGCGCCCA
PA5499Rev*Hind*III	GATCAAGCTTTCAGGCGTCCTTCTGGTC
Zur*dksA2*EMSAFor	/5Biosg/CAGGCAGTAAAGGCGGAGGA
Zur*dksA2*EMSARev	CAGCAGGTCGCGGAAGAAGT
PA5536EMSAZursiteFor	TTGTGGTGTTAATTGTTATAGTATAACGTTTTAAATTTCCCA
PA5536EMSAZursiteRev	TGGGAAATTTAAAACGTTATACTATAACAATTAACACCACAA
PA4457EMSAFor	CGGTCTGAGCGAAACCGGCCTTGGTCAGCCGTTCCTCATGCA
PA4577EMSARev	TGCATGAGGAACGGCTGACCAAGGCCGGTTTCGCTCAGACCG

LIST OF REFERENCES

Aberg, A., V. Shingler, and C Balsalobre (2008) Regulation of the *fimB* promoter: a case of differential regulation by ppGpp and DksA *in vivo*. *Mol Microbiol* **67**: 1223-1241.

Abreu, I. A. and , D. E. Cabelli (2010) Superoxide dismutases-a review of the metal-associated mechanistic variations. *Biochim Biophys Acta* **1804**: 263-274.

Ahmad, R., B. Brandsdal, I. Michaud-Soret, and N. Willassen (2009) Ferric uptake regulator protein: Binding free energy calculations and per-residue free energy decomposition. *Proteins* **75**: 373-386.

Akanuma, G., H. Nanamiya, Y. Natori, N. Nomura, and F. Kawamura (2006) Liberation of zinc-containing L31 (RpmE) from ribosomes by its paralogous gene product, YtiA, in *Bacillus subtilis*. *J Bacteriol* **188**: 2715-2720.

Alber, B., C. Colangelo, J. Dong, C. Stalhandske, T. Baird, C. Tu, C. Fierke, *et al*. (1999) Kinetic and spectroscopic characterization of the gamma-carbonic anhydrase from the methanoarchaeon *Methanosarcina thermophila*. *Biochemistry* **38**: 13119-13128.

Allen, M. D., J. A. del Campo, J. Kropat, and S. S. Merchant (2007) FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient *Chlamydomonas reinhardtii*. *Eukaryot Cell* **6**: 1841-1852.

Ammendola, S., P. Pasquali, C. Pistoia, P. Petrucci, P. Petrarca, G. Rotilio, and A. Battistoni (2007) High-affinity Zn^{2+} uptake system ZnuABC is required for bacterial zinc homeostasis in intracellular environments and contributes to the virulence of *Salmonella enterica*. *Infect Immun* **75**: 5867-5876.

Andreini, C., L. Banci, I. Bertini, S. Elmi, and A. Rosato (2007) Non-heme iron through the three domains of life. *Proteins* **67**: 317-324.

Andreini, C., L. Banci, I. Bertini, and A. Rosato (2006) Zinc through the three domains of life. *J Proteome Res* **5**: 3173-3178.

Andreini, C., L. Banci, I. Bertini, and A. Rosato (2008a) Occurrence of copper proteins through the three domains of life: a bioinformatic approach. *J Proteome Res* **7**: 209-216.

Andreini, C., I. Bertini, G. Cavallaro, G. L. Holliday, and J. M. Thornton (2008b) Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem* **13**: 1205-1218.

Andrews, S., B. Berks, J. McClay, A. Ambler, M. Quail, P. Golby, and J. Guest (1997) A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology* **143**: 3633-3647.

Ansari, A. Z., J. E. Bradner and T. V. O'Halloran (1995) DNA-bend modulation in a repressor-to-activator switching mechanism. *Nature* **374**: 371-375.

Aronsson, A. C., E. Marmstål and B. Mannervik (1978) Glyoxalase I, a zinc metalloenzyme of mammals and yeast. *Biochem Biophys Res Commun* **81**: 1235-1240.

Ashlock, P. D (1971) Monophyly and associated terms. *Systematic Zoology* **20**: 63-69.

Attwood, T. K (2000) Genomics. The Babel of bioinformatics. *Science* **290**: 471-473.

Auerbach, G., A. Herrmann, A. Bracher, G. Bader, M. Gutlich, M. Fischer, M. Neukamm, *et al. (*2000) Zinc plays a key role in human and bacterial GTP cyclohydrolase I. *Proc Natl Acad Sci U S A* **97**: 13567-13572.

Ba, L., M. Doering, T. Burkholz and C. Jacob (2009) Metal trafficking: from maintaining the metal homeostasis to future drug design. *Metallomics* **1**: 292-311.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. Datsenko, *et al. (*2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006.0008.

Babitzke, P. and C. Yanofsky (1993) Reconstitution of bacillus-subtilis trp attenuation invitro with trap, the trp rna-binding attenuation protein. *Proc Natl Acad Sci U S A* **90**: 133-137.

Bagg, A. and J. Neilands (1987) Molecular mechanism of regulation of siderophore-mediated iron assimilation. *Microbiol Rev* **51**: 509-518.

Ballantine, S. P. and D. H. Boxer (1985) Nickel-containing hydrogenase isoenzymes from anaerobically grown *Escherichia coli* K-12. *J Bacteriol* **163**: 454-459.

Banerjee, A., R. Sharma and U. C. Banerjee (2002) The nitrile-degrading enzymes: current status and future prospects. *Appl Microbiol Biotechnol* **60**: 33-44.

Bannantine, J. P., M. L. Paustian, W. R. Waters, J. R. Stabel, M. V. Palmer, L. Li and V. Kapur (2008) Profiling bovine antibody responses to *Mycobacterium avium* subsp. *paratuberculosis* infection by using protein arrays. *Infect Immun* **76**: 739-749.

Beard, S., R. Hashim, G. Wu, M. Binet, M. Hughes and R. Poole (2000) Evidence for the transport of zinc(II) ions via the pit inorganic phosphate transport system in *Escherichia coli*. *FEMS Microbiol Lett* **184**: 231-235.

Beard, S. J., M. N. Hughes and R. K. Poole (1995) Inhibition of the cytochrome bd-terminated NADH oxidase system in *Escherichia coli* K-12 by divalent metal cations. *FEMS Microbiol Lett* **131**: 205-210.

Beck, C. and R. Warren (1988) Divergent promoters, a common form of gene organization. *Microbiol Rev* **52**: 318-326.

Bekker, A., H. D. Holland, P. L. Wang, D. Rumble, H. J. Stein, J. L. Hannah, L. L. Coetzee, *et al.* (2004) Dating the rise of atmospheric oxygen. *Nature* **427**: 117-120.

Benoit, S. and R. J. Maier (2003) Dependence of *Helicobacter pylori* urease activity on the nickel-sequestering ability of the UreE accessory protein. *J Bacteriol* **185**: 4787-4795.

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and E. W. Sayers (2009) GenBank. *Nucleic Acids Res* **37**: D26-31.

Berducci, G., A. P. Mazzetti, G. Rotilio and A. Battistoni (2004) Periplasmic competition for zinc uptake between the metallochaperone ZnuA and Cu,Zn superoxide dismutase. *FEBS Lett* **569**: 289-292.

Berg, J. and Y. Shi (1996) The galvanization of biology: a growing appreciation for the roles of zinc. *Science* **271**: 1081-1085.

Bilokapic, S., T. Maier, D. Ahel, I. Gruic-Sovulj, D. Söll, I. Weygand-Durasevic and N. Ban (2006) Structure of the unusual seryl-tRNA synthetase reveals a distinct zinc-dependent mode of substrate recognition. *EMBO J* **25**: 2498-2509.

Binet, M. R. and R. K. Poole (2000) Cd(II), Pb(II) and Zn(II) ions regulate expression of the metal-transporting P-type ATPase ZntA in *Escherichia coli*. *FEBS Lett* **473**: 67-70.

Blaby-Haas, C. E., R. Furman, D. A. Rodionov, I. Artsimovitch and V. de Crécy-Lagard (2011) Role of a Zn-independent DksA in Zn homeostasis and stringent response. *Mol Microbiol* **79**: 700-715.

Blanchette, M. and M. Tompa (2002) Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res* **12**: 739-748.

Blasco, F., J. Dos Santos, A. Magalon, C. Frixon, B. Guigliarelli, C. Santini and G. Giordano (1998) NarJ is a specific chaperone required for molybdenum cofactor assembly in nitrate reductase A of *Escherichia coli*. *Mol Microbiol* **28**: 435-447.

Blencowe, D. and A. Morby (2003) Zn(II) metabolism in prokaryotes. *FEMS Microbiol Rev* **27**: 291-311.

Blindauer, C., M. Harrison, A. Robinson, J. Parkinson, P. Bowness, P. Sadler and N. Robinson (2002) Multiple bacteria encode metallothioneins and SmtA-like zinc fingers. *Mol Microbiol* **45**: 1421-1432.

Blumenthal, T (1998) Gene clusters and polycistronic transcription in eukaryotes. *Bioessays* **20**: 480-487.

Bork, P (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* **10**: 398-400.

Bou-Abdallah, F., S. Adinolfi, A. Pastore, T. Laue and N. Dennis Chasteen (2004) Iron binding and oxidation kinetics in frataxin CyaY of *Escherichia coli*. *J Mol Biol* **341**: 605-615.

Boyd, J. M., A. J. Pierik, D. J. Netz, R. Lill and D. M. Downs (2008) Bacterial ApbC can bind and effectively transfer iron-sulfur clusters. *Biochemistry* **47**: 8195-8202.

Bradford, M. M (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248-254.

Branny, P., J. Pearson, E. Pesci, T. Köhler, B. Iglewski and C. Van Delden (2001) Inhibition of quorum sensing by a *Pseudomonas aeruginosa dksA* homologue. *J Bacteriol* **183**: 1531-1539.

Bremer, H. and P. Dennis (2008) Feedback control of ribosome function in *Escherichia coli*. *Biochimie* **90**: 493-499.

Brennan, B., G. Alms, M. Nelson, L. Durney and R. Scarrow (1996) Nitrile hydratase from *Rhodococcus rhodochrous* J1 contains a non-corrin cobalt ion with two sulfur ligands. *J Amer Chem Soc* **118**: 9194-9195.

Brennan, S. and J. Matthews (1997) Hb Auckland [alpha 87(F8)HIS->ASN]: A new mutation of the proximal histidine identified by electrospray mass spectrometry. *Hemoglobin* **21**: 393-403.

Brenner, S. E (1999) Errors in genome annotation. *Trends Genet* **15**: 132-133.

Brichta, D. M., K. N. Azad, P. Ralli and G. A. O'Donovan (2004) *Pseudomonas aeruginosa* dihydroorotases: a tale of three *pyrC*s. *Arch Microbiol* **182**: 7-17.

Brocklehurst, K., J. Hobman, B. Lawley, L. Blank, S. Marshall, N. Brown and A. Morby (1999) ZntR is a Zn(II)-responsive MerR-like transcriptional regulator of *zntA* in *Escherichia coli*. *Mol Microbiol* **31**: 893-902.

Brown, L., D. Gentry, T. Elliott and M. Cashel (2002) DksA affects ppGpp induction of RpoS at a translational level. *J Bacteriol* **184**: 4455-4465.

Bruland, K. (1980) Oceanographic distributions of cadmium, zinc, nickel, and copper in the north pacific. *Earth Planet. Sci. Lett.* **47**: 176-198.

Busenlehner, L. S., M. A. Pennella and D. P. Giedroc (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev* **27**: 131-143.

Bürgisser, D. M., B. Thöny, U. Redweik, D. Hess, C. W. Heizmann, R. Huber and H. Nar (1995) 6-Pyruvoyl tetrahydropterin synthase, an enzyme with a novel type of active site involving both zinc binding and an intersubunit catalytic triad motif; site-directed mutagenesis of the proposed active center, characterization of the metal binding site and modelling of substrate binding. *J Mol Biol* **253**: 358-369.

Caldwell, C., Y. Chen, H. Goetzmann, Y. Hao, M. Borchers, D. Hassett, L. Young, *et la.* (2009) *Pseudomonas aeruginosa* exotoxin pyocyanin causes cystic fibrosis airway pathogenesis. *Am J Pathol* **175**: 2473-2488.

Campbell, D. R., K. E. Chapman, K. J. Waldron, S. Tottey, S. Kendall, G. Cavallaro, C. Andreini, *et al.* (2007) Mycobacterial cells have dual nickel-cobalt sensors: sequence relationships and metal sites of metal-responsive repressors are not congruent. *J Biol Chem* **282**: 32298-32310.

Campos-Bermudez, V. A., N. R. Leite, R. Krog, A. J. Costa-Filho, F. C. Soncini, G. Oliva and A. J. Vila (2007) Biochemical and structural characterization of *Salmonella typhimurium* glyoxalase II: new insights into metal ion selectivity. *Biochemistry* **46**: 11069-11079.

Campoy, S., M. Jara, N. Busquets, A. Pérez De Rozas, I. Badiola and J. Barbé (2002) Role of the high-affinity zinc uptake ZnuABC system in *Salmonella enterica* serovar typhimurium virulence. *Infect Immun* **70**: 4721-4725.

Cappelli, G., E. Volpe, M. Grassi, B. Liseo, V. Colizzi and F. Mariani (2006) Profiling of *Mycobacterium tuberculosis* gene expression during human macrophage infection: upregulation of the alternative sigma factor G, a group of transcriptional regulators, and proteins with unknown function. *Res Microbiol* **157**: 445-455.

Carter, E., N. Flugga, J. Boer, S. Mulrooney and R. Hausinger (2009) Interplay of metal ions and urease. *Metallomics* **1**: 207-221.

Cherepanov, P. P. and W. Wackernagel (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**: 9-14.

Chevenet, F., C. Brun, A. Bañuls, B. Jacq and R. Christen (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**: 439.

Choi, K., A. Kumar and H. Schweizer (2006) A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. *J Microbiol Methods* **64**: 391-397.

Choi, K. and H. Schweizer (2005) An improved method for rapid generation of unmarked *Pseudomonas aeruginosa* deletion mutants. *BMC Microbiol* **5**: 30.

Christianson, D. W. and C. A. Fierke (1996) Carbonic anhydrase: Evolution of the zinc binding site by nature and design. *Acc. Chem. Res.* **29**: 331–339.

Claverys, J. P. (2001) A new family of high-affinity ABC manganese and zinc permeases. *Res Microbiol* **152**: 231-243.

Clohessy, P. and B. Golden (1995) Calprotectin-mediated zinc chelation as a biostatic mechanism in host defence. *Scand J Immunol* **42**: 551-556.

Clugston, S., J. Barnard, R. Kinach, D. Miedema, R. Ruman, E. Daub and J. Honek (1998) Overproduction and characterization of a dimeric non-zinc glyoxalase I from *Escherichia coli*: evidence for optimal activation by nickel ions. *Biochemistry* **37**: 8754-8763.

Coleman, J. E. (1967) Metal ion dependent binding of sulphonamide to carbonic anhydrase. *Nature* **214**: 193-194.

Conant, J. (1923) An electrochemical study of hemoglobin. *J Biol Chem* **57**: 401-414.

Cooper, B. F., V. Sideraki, D. K. Wilson, D. Y. Dominguez, S. W. Clark, F. A. Quiocho and F. B. Rudolph (1997) The role of divalent cations in structure and function of murine adenosine deaminase. *Protein Sci* **6**: 1031-1037.

Corbin, B., E. Seeley, A. Raab, J. Feldmann, M. Miller, V. Torres, K. Anderson, *et al.* (2008) Metal chelation and inhibition of bacterial growth in tissue abscesses. *Science* **319**: 962-965.

Cornishbowden, A. (1985) Nomenclature for incompletely specified bases in nucleic-acid sequences - recommendations 1984. *Nucleic Acids Res* **13**: 3021-3030.

Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard and B. O. Palsson (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**: 92-96.

Crichton, R. and J. Pierre (2001) Old iron, young copper: From Mars to Venus. *Biometals* **14**: 99-112.

Crouzet, J., S. Levy-Schil, B. Cameron, L. Cauchois, S. Rigault, M. C. Rouyez, F. Blanche, *et al.* (1991) Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair *Pseudomonas denitrificans* DNA fragment containing five *cob*

genes and identification of structural genes encoding Cob(I)alamin adenosyltransferase, cobyric acid synthase, and bifunctional cobinamide kinase-cobinamide phosphate guanylyltransferase. *J Bacteriol* **173**: 6074-6087.

Culotta, V., L. Klomp, J. Strain, R. Casareno, B. Krems and J. Gitlin (1997) The copper chaperone for superoxide dismutase. *J BiolChem **272***: 23469-23472.

Cvetkovic, A., A. Menon, M. Thorgersen, J. Scott, F. N. Poole, F. J. Jenney, W. Lancaster, *et al.* (2010) Microbial metalloproteomes are largely uncharacterized. *Nature* **466**: 779-782.

Cámara, B., M. Marín, M. Schlömann, H. J. Hecht, H. Junca and D. H. Pieper (2008) *trans*-Dienelactone hydrolase from *Pseudomonas reinekei* MT1, a novel zinc-dependent hydrolase. *Biochem Biophys Res Commun* **376**: 423-428.

D'Ordine, R., T. Klem and V. Davisson (1999) N1-(5‘-Phosphoribosyl)adenosine−5‘-monophosphate cyclohydrolase: Purification and characterization of a unique metalloenzyme. *Biochemistry* **38**: 1537–1546.

D'souza, V. and R. Holz (1999) The methionyl aminopeptidase from *Escherichia coli* can function as an iron(II) enzyme. *Biochemistry* **38**: 11079-11085.

da Silva, J. J. R. F. and R. J. P. Williams (2001) *The biological chemistry of the elements: The inorganic chemistry of life*. New York: Oxford University Press.

Dahiya, I. and R. Stevenson (2010) The ZnuABC operon is important for *Yersinia ruckeri* infections of rainbow trout, *Oncorhynchus mykiss* (Walbaum). *J Fish Dis* **33**: 331-340.

Dalebroux, Z., B. Yagi, T. Sahr, C. Buchrieser and M. Swanson (2010) Distinct roles of ppGpp and DksA in *Legionella pneumophila* differentiation. *Mol Microbiol* **76**: 200-219.

Dalet, K., E. Gouin, Y. Cenatiempo, P. Cossart and Y. Héchard (1999) Characterisation of a new operon encoding a Zur-like protein and an associated ABC zinc permease in Listeria monocytogenes. *FEMS Microbiol Lett* **174**: 111-116.

Dandekar, T., B. Snel, M. Huynen and P. Bork (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**: 324-328.

Davis, L., T. Kakuda and V. DiRita (2009) A *Campylobacter jejuni znuA* orthologue is essential for growth in low-zinc environments and chick colonization. *J Bacteriol* **191**: 1631-1640.

Dayhoff, M. O. (1968) *Atlas of Protein Sequence and Structure*. Washington DC: National Biomedical Research Foundation.

de Crécy-Lagard, V. and A. Hanson (2007) Finding novel metabolic genes through plant-prokaryote phylogenomics. *Trends Microbiol* **15**: 563-570.

de Souza, R. F., V. Anantharaman, S. J. de Souza, L. Aravind and F. J. Gueiros-Filho (2008) AMIN domains have a predicted role in localization of diverse periplasmic protein complexes. *Bioinformatics* **24**: 2423-2426.

Demerec, K. and P. Hartman (1959) Complex loci in microorganisms. *Annual Rev Microbiol* **13**: 377-406.

Devirgiliis, C., C. Murgia, G. Danscher and G. Perozzi (2004) Exchangeable zinc ions transiently accumulate in a vesicular compartment in the yeast *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* **323**: 58-64.

Devos, D. and A. Valencia (2001) Intrinsic errors in genome annotation. *Trends Genet* **17**: 429-431.

Dixon, N. E., T. C. Gazzola, R. L. blakeley and B. Zermer (1975) Letter: Jack bean urease (EC 3.5.1.5). A metalloenzyme. A simple biological role for nickel? *J Am Chem Soc* **97**: 4131-4133.

Drake, H. L., S. I. Hu and H. G. Wood (1980) Purification of carbon monoxide dehydrogenase, a nickel enzyme from Clostridium thermocaceticum. *J Biol Chem* **255**: 7174-7180.

Dudev, M., J. Wang, T. Dudev and C. Lim (2006) Factors governing the metal coordination number in metal complexes from Cambridge structural database analyses. *J Phys ChemB* **110**: 1889-1895.

Dudev, T. and C. Lim (2008) Metal binding affinity and selectivity in metalloproteins: Insights from computational studies. *Annu Rev Biophys* **37**: 97-116.

Eaves, D. J., T. Palmer and D. H. Boxer (1997) The product of the molybdenum cofactor gene *mobB* of *Escherichia coli* is a GTP-binding protein. *Eur J Biochem* **246**: 690-697.

Edgar, R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Edwards, J. S., R. U. Ibarra and B. O. Palsson (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19**: 125-130.

Edwards, J. S. and B. O. Palsson (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* **97**: 5528-5533.

Egami, F. (1975) Origin and early evolution of transition element enzymes. *J Biochem* **77**: 1165-1169.

Eicken, C., M. A. Pennella, X. Chen, K. M. Koshlap, M. L. VanZile, J. C. Sacchettini and D. P. Giedroc (2003) A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J Mol Biol* **333**: 683-695.

Eisen, J. A. (1998a) A phylogenomic study of the MutS family of proteins. *Nucleic Acids Res* **26**: 4291-4300.

Eisen, J. A. (1998b) Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res* **8**: 163-167.

Eisen, J. A., D. Kaiser and R. M. Myers (1997) Gastrogenomic delights: a movable feast. *Nat Med* **3**: 1076-1078.

El Yacoubi, B., S. Bonnett, J. N. Anderson, M. A. Swairjo, D. Iwata-Reuyl and V. de Crécy-Lagard (2006) Discovery of a new prokaryotic type I GTP cyclohydrolase family. *J Biol Chem* **281**: 37586-37593.

Ellefson, W. L., W. B. Whitman and R. S. Wolfe (1982) Nickel-containing factor F430: chromophore of the methylreductase of Methanobacterium. *Proc Natl Acad Sci U S A* **79**: 3707-3710.

Endo, G. and S. Silver (1995) CadC, the transcriptional regulatory protein of the cadmium resistance system of *Staphylococcus aureus* plasmid pI258. *J Bacteriol* **177**: 4437-4441.

Endo, I., M. Nojiri, M. Tsujimura, M. Nakasako, S. Nagashima, M. Yohda and M. Odaka (2001) Fe-type nitrile hydratase. *J Inorg Biochem* **83**: 247-253.

Epstein, W. and J. Beckwith (1968) Regulation of gene expression. *Annu Rev Biochem* **37**: 411-436.

Feist, A. M. and B. Palsson (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol* **26**: 659-667.

Felsenstein, J (1997) An alternating least squares approach to inferring phylogenies from pairwise distances. *Syst Biol* **46**: 101–111.

Feng, Y., M. Li, H. Zhang, B. Zheng, H. Han, C. Wang, J. Yan, *et al.* (2008) Functional definition and global regulation of Zur, a zinc uptake regulator in a *Streptococcus suis* serotype 2 strain causing streptococcal toxic shock syndrome. *J Bacteriol* **190**: 7567-7578.

Ferguson, T., J. A. Soares, T. Lienard, G. Gottschalk and J. A. Krzycki (2009) RamA, a protein required for reductive activation of corrinoid-dependent methylamine

methyltransferase reactions in methanogenic archaea. *J Biol Chem* **284**: 2285-2295.

Fields, C., D. Brichta, M. Shephardson, M. Farinha and G. O'Donovan (1999) Phylogenetic analysis and classification of dihydroorotases: a complex history for a complex enzyme. *Paths Pyrimidines* **7**: 49-63.

Finney, L. and T. O'Halloran (2003) Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. *Science* **300**: 931-936.

Fischer, J. D., G. L. Holliday, S. A. Rahman and J. M. Thornton (2010) The structures and physicochemical properties of organic cofactors in biocatalysis. *J Mol Biol* **403**: 803-824.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.

Fork, D. C. and W. Urbach (1965) Evidence for the localization of plastocyanin in the electron-transport chain of photosynthesis. *Proc Natl Acad Sci U S A* **53**: 1307-1315.

Fritz, G., D. Griesshaber, O. Seth and P. Kroneck (2001) Nonaheme cytochrome c, a new physiological electron acceptor for [Ni,Fe] hydrogenase in the sulfate-reducing bacterium *Desulfovibrio desulfuricans* Essex: Primary sequence, molecular parameters, and redox properties. *Biochemistry* **40**: 1317-1324.

Fu, C., J. W. Olson and R. J. Maier (1995) HypB protein of *Bradyrhizobium japonicum* is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer. *Proc Natl Acad Sci U S A* **92**: 2333-2337.

Gaal, T., M. S. Bartlett, W. Ross, C. L. Turnbough and R. L. Gourse (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**: 2092-2097.

Gaballa, A. and J. D. Helmann (1998) Identification of a zinc-specific metalloregulatory protein, Zur, controlling zinc transport operons in *Bacillus subtilis*. *J Bacteriol* **180**: 5815-5821.

Gaballa, A., T. Wang, R. W. Ye and J. D. Helmann (2002) Functional analysis of the *Bacillus subtilis* Zur regulon. *J Bacteriol* **184**: 6508-6514.

Gabriel, S. and J. Helmann (2009) Contributions of Zur-controlled ribosomal proteins to growth under zinc starvation conditions. *J Bacteriol* **191**: 6116-6122.

Gabriel, S. E., F. Miyagi, A. Gaballa and J. D. Helmann (2008) Regulation of the *Bacillus subtilis yciC* gene and insights into the DNA-binding specificity of the zinc-sensing metalloregulator Zur. *J Bacteriol* **190**: 3482-3488.

Galperin, M. and E. Koonin (2010) From complete genome sequence to 'complete' understanding? *Trends Biotechnol* **28**: 398-406.

Galperin, M. Y., A. Bairoch and E. V. Koonin (1998a) A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Sci* **7**: 1829-1835.

Galperin, M. Y. and E. V. Koonin (1998) Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. *In Silico Biol* **1**: 55-67.

Galperin, M. Y. and E. V. Koonin (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol* **18**: 609-613.

Galperin, M. Y., D. R. Walker and E. V. Koonin (1998b) Analogous enzymes: independent inventions in enzyme evolution. *Genome Res* **8**: 779-790.

Gan, Y. H., K. L. Chua, H. H. Chua, B. Liu, C. S. Hii, H. L. Chong and P. Tan (2002) Characterization of *Burkholderia pseudomallei* infection and identification of novel virulence factors using a *Caenorhabditis elegans* host system. *Mol Microbiol* **44**: 1185-1197.

Gasper, R., A. Scrima and A. Wittinghofer (2006) Structural insights into HypB, a GTP-binding protein that regulates metal binding. *J BiolChem* **281**: 27492-27502.

Gelfand, M., P. Novichkov, E. Novichkova and A. Mironov (2000) Comparative analysis of regulatory patterns in bacterial genomes. *Brief Bioinform* **1**: 357-371.

GERALD, P. S. and M. L. EFRON (1961) Chemical studies of several varieties of Hb M. *Proc Natl Acad Sci U S A* **47**: 1758-1767.

Gerdes, S., O. Kurnasov, K. Shatalin, B. Polanuyer, R. Sloutsky, V. Vonstein, R. Overbeek, *et al.* (2006) Comparative genomics of NAD biosynthesis in cyanobacteria. *J Bacteriol* **188**: 3012-3023.

Gerendas, J., J. Polacco, S. Freyermuth and B. Sattelmacher (1999) Significance of nickel for plant growth and metabolism. *J Plant Nutr Soil Sci* **162**: 241-256.

Geslin, C., J. Llanos, D. Prieur and C. Jeanthon (2001) The manganese and iron superoxide dismutases protect *Escherichia coli* from heavy metal toxicity. *Res Microbiol* **152**: 901-905.

Giedroc, D. P. and J. E. Coleman (1986) Structural and functional differences between the two intrinsic zinc ions of *Escherichia coli* RNA polymerase. *Biochemistry* **25**: 4969-4978.

Glerum, D., A. Shtanko and A. Tzagoloff (1996) Characterization of COX17, a yeast gene involved in copper metabolism and assembly of cytochrome oxidase. *J Biol Chem* **271**: 14504-14509.

Gophna, U., E. Bapteste, W. Doolittle, D. Biran and E. Ron (2005) Evolutionary plasticity of methionine biosynthesis. *Gene* **355**: 48-57.

Gouet, P., E. Courcelle, D. Stuart and F. Métoz (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305-308.

Govan, J. and V. Deretic (1996) Microbial pathogenesis in cystic fibrosis: Mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microbiol Rev* **60**: 539-574.

Graham, A., S. Hunt, S. Stokes, N. Bramall, J. Bunch, A. Cox, C. McLeod, *et al.* (2009) Severe zinc depletion of *Escherichia coli*: roles for high affinity zinc binding by ZinT, zinc transport and zinc-independent proteins. *J Biol Chem* **284**: 18377-18389.

Grass, G., B. Fan, B. P. Rosen, S. Franke, D. H. Nies and C. Rensing (2001) ZitB (YbgR), a member of the cation diffusion facilitator family, is an additional zinc transporter in *Escherichia coli*. *J Bacteriol* **183**: 4664-4667.

Grass, G., S. Franke, N. Taudte, D. H. Nies, L. M. Kucharski, M. E. Maguire and C. Rensing (2005a) The metal permease ZupT from *Escherichia coli* is a transporter with a broad substrate spectrum. *J Bacteriol* **187**: 1604-1611.

Grass, G., M. Otto, B. Fricke, C. J. Haney, C. Rensing, D. H. Nies and D. Munkelt (2005b) FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. *Arch Microbiol* **183**: 9-18.

Grass, G., M. Wong, B. Rosen, R. Smith and C. Rensing (2002) ZupT is a Zn(II) uptake system in *Escherichia coli*. *J Bacteriol* **184**: 864-866.

Gray, R., G. MacGregor, D. Noble, M. Imrie, M. Dewar, A. Boyd, J. Innes, *et al.* (2008) Sputum proteomics in inflammatory and suppurative respiratory diseases. *Am J Respir Crit Care Med* **178**: 444-452.

Gribble, T. and H. Schwartz (1965) Effect of protoporphyrin on hemoglobin synthesis. *Biochim Biophys Acta* **103**: 333-338.

Grotz, N., T. Fox, E. Connolly, W. Park, M. L. Guerinot and D. Eide (1998) Identification of a family of zinc transporter genes from *Arabidopsis* that respond to zinc deficiency. *Proc Natl Acad Sci U S A* **95**: 7220-7224.

Guffanti, A. A., Y. Wei, S. V. Rood and T. A. Krulwich (2002) An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K^+ and H^+. *Mol Microbiol* **45**: 145-153.

Gurd, F. R. and P. E. WilcoX (1956) Complex formation between metallic cations and proteins, peptides and amino acids. *Adv Protein Chem* **11**: 311-427.

Guss, J. M. and H. C. Freeman (1983) Structure of oxidized poplar plastocyanin at 1.6 A resolution. *J Mol Biol* **169**: 521-563.

Gutiérrez, J., F. Amaro and A. Martín-González (2009) From heavy metal-binders to biosensors: ciliate metallothioneins discussed. *Bioessays* **31**: 805-816.

Gutteridge, J. M. and S. Wilkins (1983) Copper salt-dependent hydroxyl radical formation. Damage to proteins acting as antioxidants. *Biochim Biophys Acta* **759**: 38-41.

Guzman, L., D. Belin, M. Carson and J. Beckwith (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J Bacteriol* **177**: 4121-4130.

Haas, C. E., D. A. Rodionov, J. Kropat, D. Malasarn, S. S. Merchant and V. de Crécy-Lagard (2009) A subset of the diverse COG0523 family of putative metal chaperones is linked to zinc homeostasis in all kingdoms of life. *BMC Genomics* **10**: 470.

Hall, R. S., D. F. Xiang, C. Xu and F. M. Raushel (2007) N-Acetyl-D-glucosamine-6-phosphate deacetylase: substrate activation via a single divalent metal ion. *Biochemistry* **46**: 7942-7952.

Hantke, K. (1981) Regulation of ferric iron transport in *Escherichia coli* K12: isolation of a constitutive mutant. *Mol Gen Genet* **182**: 288-292.

Harding, M. (2004) The architecture of metal coordination groups in proteins. *Acta Crystallogr D Biol Crystallogr* **60**: 849-859.

Harding, M. M. (1999) The geometry of metal-ligand interactions relevant to proteins. *Acta Crystallogr D Biol Crystallogr* **55**: 1432-1443.

Hashimoto, Y., M. Nishiyama, S. Horinouchi and T. Beppu (1994) Nitrile hydratase gene from *Rhodococcus* sp. N-774 requirement for its downstream region for efficient expression. *Biosci Biotechnol Biochem* **58**: 1859-1865.

Hassan, M. T., D. van der Lelie, D. Springael, U. Römling, N. Ahmed and M. Mergeay (1999) Identification of a gene cluster, czr, involved in cadmium and zinc resistance in *Pseudomonas aeruginosa*. *Gene* **238**: 417-425.

Hauryliuk, V. (2006) GTPases of the prokaryotic translation apparatus. *Mol Biol* **40**: 688-701.

Hayashi, A., T. Fujita, M. Fujimura and K. Titani (1980) A new abnormal fetal hemoglobin, Hb FM-Osaka (alpha 2 gamma 2 63His replaced with Tyr). *Hemoglobin* **4**: 447-448.

Heidrich, C., M. F. Templin, A. Ursinus, M. Merdanovic, J. Berger, H. Schwarz, M. A. de Pedro, *et al.* (2001) Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**: 167-178.

Heinz, U., M. Kiefer, A. Tholey and H. Adolph (2005) On the competition for available zinc. *J Biol Chem* **280**: 3197-3207.

Helbig, K., C. Grosse and D. H. Nies (2008) Cadmium toxicity in glutathione mutants of *Escherichia coli*. *J Bacteriol* **190**: 5439-5454.

Heldt, D., A. Lawrence, M. Lindenmeyer, E. Deery, P. Heathcote, S. Rigby and M. Warren (2005) Aerobic synthesis of vitamin B_{12}: ring contraction and cobalt chelation. *Biochem Soc Trans* **33**: 815-819.

Hightower, K. E. and C. A. Fierke (1999) Zinc-catalyzed sulfur alkyation:insights from protein farnesyltransferase. *Curr Opin Chem Biol* **3**: 176-181.

Hille, R. (2005) Molybdenum-containing hydroxylases. *Arch Biochem Biophys* **433**: 107-116.

Hoang, T., R. Karkhoff-Schweizer, A. Kutchma and H. Schweizer (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* **212**: 77-86.

Hodgkin, D., J. Pickworth, J. Robertson, K. Trueblood, R. Prosen and J. White (1955) The crystal structure of the hexacarboxylic acid derived from B_{12} and the molecular structure of the vitamin. *Nature* **176**: 325-328.

Hojas-Bernal, R., P. McNab-Martin, V. Fairbanks, M. Holmes, J. Hoyer, D. McCormick and K. Kubik (1999) Hb Chile [beta 28(B10)Leu -> Met]: an unstable hemoglobin associated with chronic methemoglobinemia and sulfonamide or methylene blue-induced hemolytic anemia. *Hemoglobin* **123**: 125-134.

Hollingsworth, N. M., L. Ponte and C. Halsey (1995) MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes Dev* **9**: 1728-1739.

Hong, J., B. Mullin and C. An (1996) Physical and genetic map of a major *nif* gene cluster from *Frankia* strain FaC1. *Mol Cells* **6**: 279-284.

Huang, D. L., D. J. Tang, Q. Liao, H. C. Li, Q. Chen, Y. Q. He, J. X. Feng, *et al.* (2008) The Zur of *Xanthomonas campestris* functions as a repressor and an activator of

putative zinc homeostasis genes via recognizing two distinct sequences within its target promoters. *Nucleic Acids Res* **36**: 4295-4309.

Hwang, Y. and D. Miller (1987) A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein. *J Biol Chem* **262**: 13081-13085.

Ilbert, M., V. Mejean, M. Giudici-Orticoni, J. Samama and C. Iobbi-Nivol (2003) Involvement of a mate chaperone (TorD) in the maturation pathway of molybdoenzyme TorA. *J BiolChem* **278**: 28787-28792.

Irving, H. and R. Williams (1948) Order of stability of metal complexes. *Nature* **162**: 746-747.

Ishikawa, S., S. Kawahara and J. Sekiguchi (1999) Cloning and expression of two autolysin genes, *cwlU* and *cwlV*, which are tandemly arranged on the chromosome of *Bacillus polymyxa* var. colistinus. *Mol Gen Genet* **262**: 738-748.

Iwig, J. S., S. Leitch, R. W. Herbst, M. J. Maroney and P. T. Chivers (2008) Ni(II) and Co(II) sensing by *Escherichia coli* RcnR. *J Am Chem Soc* **130**: 7592-7606.

Jacob, C., W. Maret and B. Vallee (1998) Control of zinc transfer between thionein, metallothionein, and zinc proteins. *Proc Natl Acad Sci U S A* **95**: 3489-3494.

Jacobi, A., R. Rossmann and A. Böck (1992) The hyp operon gene products are required for the maturation of catalytically active hydrogenase isoenzymes in *Escherichia coli*. *Arch Microbiol* **158**: 444-451.

Jacobitz, S. and P. E. Bishop (1992) Regulation of nitrogenase-2 in *Azotobacter vinelandii* by ammonium, molybdenum, and vanadium. *J Bacteriol* **174**: 3884-3888.

Jacobson, M., K. Brigle, L. Bennett, R. Setterquist, M. Wilson, V. Cash, J. Beynon, *et al.* (1989) Physical and genetic-map of the major *nif* gene-cluster from *Azotobacter vinelandii*. *J Bacteriol* **171**: 1017-1027.

Jaffe, E. (2003) An unusual phylogenetic variation in the metal ion binding sites of porphobilinogen synthase. *Chem Biol* **10**: 25-34.

Jensen, W. (2003) The place of zinc, cadmium, and mercury in the periodic table. *J Chem Edu* **80**: 952-961.

Jeon, W. B., J. Cheng and P. W. Ludden (2001) Purification and characterization of membrane-associated CooC protein and its functional role in the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum*. *J Biol Chem* **276**: 38602-38609.

Jude, F., T. Köhler, P. Branny, K. Perron, M. Mayer, R. Comte and C. van Delden (2003) Posttranscriptional control of quorum-sensing-dependent virulence genes by DksA in *Pseudomonas aeruginosa*. *J Bacteriol* **185**: 3558-3566.

Kallifidas, D., B. Pascoe, G. Owen, C. Strain-Damerell, H. Hong and M. Paget (2010) The Zinc-responsive regulator Zur controls expression of the coelibactin gene cluster in *Streptomyces coelicolor*. *J Bacteriol* **192**: 608-611.

Kang, P. and E. Craig (1990) Identification and characterization of a new *Escherichia coli* gene that is a dosage-dependent suppressor of a *dnaK* deletion mutation. *J Bacteriol* **172**: 2055-2064.

Kasahara, M. and Y. Anraku (1972) Inhibition of the respiratory chain of *Escherichia coli* by zinc ions. *J Biochem* **72**: 777-781.

Kasahara, M. and Y. Anraku (1974) Succinate- and NADH oxidase systems of *Escherichia coli* membrane vesicles. Mechanism of selective inhibition of the systems by zinc ions. *J Biochem* **76**: 967-976.

Katayama, A., A. Tsujii, A. Wada, T. Nishino and A. Ishihama (2002) Systematic search for zinc-binding proteins in *Escherichia coli*. *Eur J Biochem* **269**: 2403-2413.

Kathuria, S. and A. C. Martiny (2011) Prevalence of a calcium-based alkaline phosphatase associated with the marine cyanobacterium *Prochlorococcus* and other ocean bacteria. *Environ Microbiol* **13**: 74-83.

Kato, Y., S. Yoshida and Y. Asano (2005) Polymerase chain reaction for identification of aldoxime dehydratase in aldoxime- or nitrile-degrading microorganisms. *FEMS Microbiol Lett* **246**: 243-249.

Katoh, S. (1960) A new copper protein from *Chlorella ellisoidea*. *Nature* **186**: 533-534.

Katoh, S., I. Suga, I. Shiratori and A. Takamiya (1961) Distribution of plastocyanin in plants, with special reference to its localization in chloroplasts. *Arch Biochem Biophys* **94**: 136-141.

Katoh, S. and A. Takamiya (1961) A new leaf copper protein 'plastocyanin', a natural Hill oxidant. *Nature* **189**: 665-666.

Kehl-Fie, T. and E. Skaar (2010) Nutritional immunity beyond iron: a role for manganese and zinc. *Curr Opin Chem Biol* **14**: 218-224.

Kerby, R., S. Hong, S. Ensign, L. Coppoc, P. Ludden and G. Roberts (1992) Genetic and physiological characterization of the *Rhodospirillum rubrum* carbon monoxide dehydrogenase system. *J Bacteriol* **174**: 5284-5294.

Keseler, I., C. Bonavides-Martínez, J. Collado-Vides, S. Gama-Castro, R. Gunsalus, D. Johnson, M. Krummenacker, *et al.* (2009) EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res* **37**: D464-470.

Khil, P. P., G. Obmolova, A. Teplyakov, A. J. Howard, G. L. Gilliland and R. D. Camerini-Otero (2004) Crystal structure of the *Escherichia coli* YjiA protein suggests a GTP-dependent regulatory function. *Proteins* **54**: 371-374.

Kidd, S. P. and N. L. Brown (2003) ZccR: a MerR-like regulator from *Bordetella pertussis* which responds to zinc, cadmium, and cobalt. *Biochem Biophys Res Commun* **302**: 697-702.

Kim, E. J., H. J. Chung, B. Suh, Y. C. Hah and J. H. Roe (1998) Transcriptional and post-transcriptional regulation by nickel of sodN gene encoding nickel-containing superoxide dismutase from *Streptomyces coelicolor* Müller. *Mol Microbiol* **27**: 187-195.

Kim, I., J. Kim, B. Min, C. Lee and C. Park (2007) Screening of genes related to methylglyoxal susceptibility. *J Microbiol* **45**: 339-343.

Kim, J. K., S. B. Mulrooney and R. P. Hausinger (2005) Biosynthesis of active *Bacillus subtilis* urease in the absence of known urease accessory proteins. *J Bacteriol* **187**: 7150-7154.

Kim, J. S., J. H. Jang, J. W. Lee, S. O. Kang, K. S. Kim and J. K. Lee (2000) Identification of cis site involved in nickel-responsive transcriptional repression of *sodF* gene coding for Fe- and Zn-containing superoxide dismutase of *Streptomyces griseus*. *Biochim Biophys Acta* **1493**: 200-207.

Kim, S., K. Watanabe, T. Shirahata and M. Watarai (2004) Zinc uptake system (*znuA* locus) of *Brucella abortus* is essential for intracellular survival and virulence in mice. *J Vet Med Sci* **66**: 1059-1063.

Kisker, C., H. Schindelin, B. Alber, J. Ferry and D. Rees (1996) A left-handed beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *Embo J* **15**: 2323-2330.

Kleiner, D. (1978) Inhibition of the respiratory system in *Azotobacter vinelandii* by divalent transition-metal ions. *FEBS Lett* **96**: 364-366.

Kochan, I. (1973) The role of iron in bacterial infections, with special consideration of host-tubercle bacillus interaction. *Curr Top Microbiol Immunol* **60**: 1-30.

Konhauser, K. O., E. Pecoits, S. V. Lalonde, D. Papineau, E. G. Nisbet, M. E. Barley, N. T. Arndt, *et al.* (2009) Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* **458**: 750-753.

Koonin, E., Y. Wolf and L. Aravind (2000) Protein fold recognition using sequence profiles and its application in structural genomics. *Adv Prot Chem* **54**: 245-275.

Korotkova, N. and M. E. Lidstrom (2004) MeaB is a component of the methylmalonyl-CoA mutase complex required for protection of the enzyme from inactivation. *J Biol Chem* **279**: 13652-13658.

Koutmos, M., R. Pejchal, T. M. Bomer, R. G. Matthews, J. L. Smith and M. L. Ludwig (2008) Metal active site elasticity linked to activation of homocysteine in methionine synthases. *Proc Natl Acad Sci U S A* **105**: 3286-3291.

Krom, B., J. Warner, W. Konings and J. Lolkema (2000) Complementary metal ion specificity of the metal-citrate transporters CitM and CitH of *Bacillus subtilis*. *J Bacteriol* **182**: 6374-6381.

Kubác, D., O. Kaplan, V. Elišáková, M. Pátek, V. Vejvoda, K. Slámová, A. Tóthová, *et al.* (2008) Biotransformation of nitriles to amides using soluble and immobilized nitrile hydratase from *Rhodococcus erythropolis* A4. *J Mol Catal B: Enzymatic* **50**: 107-113.

Kuchar, J. and R. P. Hausinger (2004) Biosynthesis of metal sites. *Chem Rev* **104**: 509-525.

Kurachi, M. (1958) Studies on the biosynthesis of pyocyanine. *Bull Inst Chem Res Kyoto Univ* **36**: 188–196.

Köhler, S., V. Foulongne, S. Ouahrani-Bettache, G. Bourg, J. Teyssier, M. Ramuz and J. P. Liautard (2002) The analysis of the intramacrophagic virulome of *Brucella suis* deciphers the environment encountered by the pathogen inside the macrophage host cell. *Proc Natl Acad Sci U S A* **99**: 15711-15716.

la Cour, T. F., J. Nyborg, S. Thirup and B. F. Clark (1985) Structural details of the binding of guanosine diphosphate to elongation factor Tu from *E. coli* as studied by X-ray crystallography. *EMBO J* **4**: 2385-2388.

Laddaga, R. A. and S. Silver (1985) Cadmium uptake in *Escherichia coli* K-12. *J Bacteriol* **162**: 1100-1105.

Lam, R., V. Romanov, K. Johns, K. Battaile, J. Wu-Brown, J. Guthrie, R. Hausinger, *et al.* (2010) Crystal structure of a truncated urease accessory protein UreF from *Helicobacter pylori*. *Proteins* **78**: 2839-2848.

Lane, T. W. and F. M. Morel (2000) A biological function for cadmium in marine diatoms. *Proc Natl Acad Sci U S A* **97**: 4627-4631.

Lane, T. W., M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince and F. M. Morel (2005) Biochemistry: a cadmium enzyme from a marine diatom. *Nature* **435**: 42.

Lavrrar, J., C. Christoffersen and M. McIntosh (2002) Fur-DNA interactions at the bidirectional *fepDGC-entS* promoter region in *Escherichia coli* **322**. *J Mol Bio* : 983-995.

Lawrence, J. G. and J. R. Roth (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843-1860.

Leach, M. and D. Zamble (2007) Metallocenter assembly of the hydrogenase enzymes. *Curr Opin Chem Biol* **11**: 159-165.

Leach, M. R., J. W. Zhang and D. B. Zamble (2007) The role of complex formation between the *Escherichia coli* hydrogenase accessory factors HypB and SlyD. *J Biol Chem* **282**: 16177-16186.

Lecam, E., D. Frechon, M. Barray, A. Fourcade and E. Delain (1994) Observation of binding and polymerization of Fur repressor onto operator-containing DNA with electron and atomic-force microscopes. *Proc Natl Acad Sci U S A* **91**: 11816-11820.

Lee, J. and J. Helmann (2007) Functional specialization within the Fur family of metalloregulators. *Biometals* **20**: 485-499.

Lee, J. and E. Sonnhammer (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* **13**: 875-882.

Lee, M., C. W. Chan, J. Mitchell Guss, R. I. Christopherson and M. J. Maher (2005) Dihydroorotase from *Escherichia coli*: loop movement and cooperativity between subunits. *J Mol Biol* **348**: 523-533.

Lee, M., S. Mulrooney, M. Renner, Y. Markowicz and R. Hausinger (1992) *Klebsiella aerogenes* urease gene cluster: sequence of *ureD* and demonstration that four accessory genes (u*reD, ureE, ureF*, and *ureG*) are involved in nickel metallocenter biosynthesis. *J Bacteriol* **174**: 4324-4330.

Leipe, D. D., Y. I. Wolf, E. V. Koonin and L. Aravind (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* **317**: 41-72.

Leitch, S., M. J. Bradley, J. L. Rowe, P. T. Chivers and M. J. Maroney (2007) Nickel-specific response in the transcriptional regulator, *Escherichia coli* NikR. *J Am Chem Soc* **129**: 5085-5095.

Lesburg, C. A., C. Huang, D. W. Christianson and C. A. Fierke (1997) Histidine --> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase II alter metal coordination geometry but retain catalytic activity. *Biochemistry* **36**: 15780-15791.

Levinson, G. and G. A. Gutman (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* **15**: 5323-5338.

Lewis, D., J. Klesney-Tait, S. Lumbley, C. Ward, J. Latimer, C. Ison and E. Hansen (1999) Identification of the ZnuA-encoded periplasmic zinc transport protein of *Haemophilus ducreyi*. *Infect Immun* **67**: 5060-5068.

Lewis, P. J., S. D. Thaker and J. Errington (2000) Compartmentalization of transcription and translation in *Bacillus subtilis*. *EMBO J* **19**: 710-718.

Li, Y., Y. Qiu, H. Gao, Z. Guo, Y. Han, Y. Song, Z. Du, *et al.* (2009) Characterization of Zur-dependent genes and direct Zur targets in *Yersinia pestis*. *BMC Microbiol* **9**: 128.

Lieberman, I. and A. Kornberg (1953) Enzymic synthesis and breakdown of a pyrimidine, orotic acid. I. Dihydro-orotic dehydrogenase. *Biochim Biophys Acta* **12**: 223-234.

Lin, Z., M. Nei and H. Ma (2007) The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution. *Nucleic Acids Res* **35**: 7591-7603.

Lindsay, J. A. and S. J. Foster (2001) zur: a Zn(2+)-responsive regulatory element of *Staphylococcus aureus*. *Microbiology* **147**: 1259-1266.

Lindskog, S. and J. E. Coleman (1973) The catalytic mechanism of carbonic anhydrase. *Proc Natl Acad Sci U S A* **70**: 2505-2508.

Liolios, K., I. M. Chen, K. Mavromatis, N. Tavernarakis, P. Hugenholtz, V. M. Markowitz and N. C. Kyrpides (2010) The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* **38**: D346-354.

Liu, J., S. J. Dutta, A. J. Stemmler and B. Mitra (2006) Metal-binding affinity of the transmembrane site in ZntA: implications for metal selectivity. *Biochemistry* **45**: 763-772.

Liu, M., G. Gui, B. Wei, J. Preston, L. Oakford, U. Yuksel, D. Giedroc, *et al.* (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem* **272**: 17502-17510.

Liu, M. and T. Romeo (1997) The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J Bacteriol* **179**: 4639-4642.

Liu, T., S. Nakashima, K. Hirose, M. Shibasaka, M. Katsuhara, B. Ezaki, D. P. Giedroc, *et al.* (2004) A novel cyanobacterial SmtB/ArsR family repressor regulates the

expression of a CPx-ATPase and a metallothionein in response to both Cu(I)/Ag(I) and Zn(II)/Cd(II). *J Biol Chem* **279**: 17810-17818.

Liuzzi, J. P., L. A. Lichten, S. Rivera, R. K. Blanchard, T. B. Aydemir, M. D. Knutson, T. Ganz, *et al.* (2005) Interleukin-6 regulates the zinc transporter Zip14 in liver and contributes to the hypozincemia of the acute-phase response. *Proc Natl Acad Sci U S A* **102**: 6843-6848.

Llull, D. and I. Poquet (2004) New expression system tightly controlled by zinc availability in *Lactococcus lactis*. *Appl Environ Microbiol* **70**: 5398-5406.

Loke, H. K. and P. A. Lindahl (2003) Identification and preliminary characterization of AcsF, a putative Ni-insertase used in the biosynthesis of acetyl-CoA synthase from *Clostridium thermoaceticum*. *J Inorg Biochem* **93**: 33-40.

Lopez-Serrano, D., F. Solano and A. Sanchez-Amat (2004) Identification of an operon involved in tyrosinase activity and melanin synthesis in *Marinomonas mediterranea*. *Gene* **342**: 179-187.

Lu, J., Y. Zheng, H. Yamagishi, M. Odaka, M. Tsujimura, M. Maeda and I. Endo (2003) Motif CXCC in nitrile hydratase activator is critical for NHase biogenesis *in vivo*. *FEBS Lett* **553**: 391-396.

Lu, M., J. Chai and D. Fu (2009) Structural basis for autoregulation of the zinc transporter YiiP. *Nat Struct Mol Biol* **16**: 1063-1067.

Lu, M. and D. Fu (2007) Structure of the zinc transporter YiiP. *Science* **317**: 1746-1748.

Lucarelli, D., S. Russo, E. Garman, A. Milano, W. Meyer-Klaucke and E. Pohl (2007) Crystal structure and function of the zinc uptake regulator FurB from *Mycobacterium tuberculosis*. *J Biol Chem* **282**: 9914-9922.

Luk, E., M. Carroll, M. Baker and V. C. Culotta (2003) Manganese activation of superoxide dismutase 2 in *Saccharomyces cerevisiae* requires MTM1, a member of the mitochondrial carrier family. *Proc Natl Acad Sci U S A* **100**: 10353-10357.

Lupoli, T. J., T. Taniguchi, T. S. Wang, D. L. Perlstein, S. Walker and D. E. Kahne (2009) Studying a cell division amidase using defined peptidoglycan substrates. *J Am Chem Soc* **131**: 18230-18231.

Luscombe, N. M., D. Greenbaum and M. Gerstein (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* **40**: 346-358.

Lusitani, D., S. Malawista and R. Montgomery (2003) Calprotectin, an abundant cytosolic protein from human polymorphonuclear leukocytes, inhibits the growth of *Borrelia burgdorferi*. *Infect Immun* **71**: 4711-4716.

Lutz, S., A. Jacobi, V. Schlensog, R. Bohm, G. Sawers and A. Bock (1991) Molecular characterization of an operon (*hyp*) necessary for the activity of the 3 hydrogenase isoenzymes in *Escherichia coli*. *Mol Microbiol* **5**: 123-135.

Macauley, S. R., S. A. Zimmerman, E. E. Apolinario, C. Evilia, Y. M. Hou, J. G. Ferry and K. R. Sowers (2009) The archetype gamma-class carbonic anhydrase (Cam) contains iron when synthesized *in vivo*. *Biochemistry* **48**: 817-819.

MacDiarmid, C. W., L. A. Gaither and D. Eide (2000) Zinc transporters that regulate vacuolar zinc storage in *Saccharomyces cerevisiae*. *EMBO J* **19**: 2845-2855.

Maciag, A., E. Dainese, G. M. Rodriguez, A. Milano, R. Provvedi, M. R. Pasca, I. Smith, *et al.* (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. *J Bacteriol* **189**: 730-740.

Magnusson, L., A. Farewell and T. Nyström (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol* **13**: 236-242.

Magnusson, L., B. Gummesson, P. Joksimovi• , A. Farewell and T. Nyström (2007) Identical, independent, and opposing roles of ppGpp and DksA in *Escherichia coli*. *J Bacteriol* **189**: 5193-5202.

Maier, T., F. Lottspeich and A. Böck (1995) GTP hydrolysis by HypB is essential for nickel insertion into hydrogenases of *Escherichia coli*. *Eur J Biochem* **230**: 133-138.

Makarova, K. and E. Koonin (2003) Filling a gap in the central metabolism of archaea: prediction of a novel aconitase by comparative-genomic analysis. *FEMS Microbiol Lett* **227**: 17-23.

Makarova, K., V. Ponomarev and E. Koonin (2001) Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins. *Genome Biol* **2**: 1-14.

Marchler-Bauer, A., J. Anderson, F. Chitsaz, M. Derbyshire, C. DeWeese-Scott, J. Fong, L. Geer, R. Geer, *et al. (*2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**: D205-210.

Maret, W. and Y. Li (2009) Coordination dynamics of zinc in proteins. *Chem Rev* **109**: 4682-4707.

Marques, H. and L. Knapton (1997) Factors affecting the rate of ligand substitution reactions of aquacobalamin (vitamin B-12a). *J Chem Soc, Dalton Trans* **20**: 3827-3833.

Martin, M., B. Byers, M. Olson, M. Salin, J. Arceneaux and C. Tolbert (1986) A *Streptococcus mutans* superoxide dismutase that is active with either manganese or iron as a cofactor *J Biol Chem* **261**: 9361-9367.

Matthews, R., A. Smith, Z. Zhou, R. Taurog, V. Bandarian, J. Evans and M. Ludwig (2003) Cobalamin-dependent and cobalamin-independent methionine synthases: Are there two solutions to the same chemical problem? *Helvetica Chimica Acta* **86**: 3939-3954.

Mayaux, J. F., T. Kalogerakos, K. K. Brito and S. Blanquet (1982) Removal of the tightly bound zinc from *Escherichia coli* trypsin-modified methionyl-tRNA synthetase. *Eur J Biochem* **128**: 41-46.

McBride, S. M., P. S. Coburn, A. S. Baghdayan, R. J. Willems, M. J. Grande, N. Shankar and M. S. Gilmore (2009) Genetic variation and evolution of the pathogenicity island of *Enterococcus faecalis*. *J Bacteriol* **191**: 3392-3402.

McCarty, R. M., A. r. d. Somogyi and V. Bandarian (2009) *Escherichia coli* QueD is a 6-Carboxy-5,6,7,8-tetrahydropterin synthase. *Biochemistry* **48**: 2301.

McCue, L., W. Thompson, C. Carmack, M. Ryan, J. Liu, V. Derbyshire and C. Lawrence (2001) Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. *Nucleic Acids Res* **29**: 774-782.

McGuire, A. M., J. D. Hughes and G. M. Church (2000) Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res* **10**: 744-757.

McMahon, R. J. and R. J. Cousins (1998) Regulation of the zinc transporter ZnT-1 by dietary zinc. *Proc Natl Acad Sci U S A* **95**: 4841-4846.

McNaught, A. D. and A. Wilkinson (1997) *IUPAC. Compendium of Chemical Terminology.* Oxford: Blackwell Scientific Publications.

Meier, T. I., R. B. Peery, K. A. McAllister and G. Zhao (2000) Era GTPase of *Escherichia coli*: binding to 16S rRNA and modulation of GTPase activity by RNA and carbohydrates. *Microbiology* **146 (Pt 5)**: 1071-1083.

Meldrum, N. U. and F. J. W. Roughton (1933) Carbonic anhydrase and the state of carbon dioxide in the blood. *Nature* **131**: 874-875.

Merchant, S. and L. Bogorad (1986a) Rapid degradation of apoplastocyanin in Cu(II)-deficient cells of *Chlamydomonas reinhardtii*. *J Biol Chem* **261**: 15850-15853.

Merchant, S. and L. Bogorad (1986b) Regulation by copper of the expression of plastocyanin and cytochrome c552 in *Chlamydomonas reinhardi*. *Mol Cell Biol* **6**: 462-469.

Merchant, S. S., M. D. Allen, J. Kropat, J. L. Moseley, J. C. Long, S. Tottey and A. M. Terauchi (2006) Between a rock and a hard place: trace element nutrition in *Chlamydomonas*. *Biochim Biophys Acta* **1763**: 578-594.

Mergeay, M., D. Nies, H. G. Schlegel, J. Gerits, P. Charles and F. Van Gijsegem (1985) *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J Bacteriol* **162**: 328-334.

Milburn, M., L. Tong, A. deVos, A. Brünger, Z. Yamaizumi, S. Nishimura and S. Kim (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**: 939-945.

Miller, J. H (1972) *Experiments in Molecular Genetics.* New York: Cold Spring Harbor Laboratory Press.

Mironov, A. A., E. V. Koonin, M. A. Roytberg and M. S. Gelfand (1999) Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res* **27**: 2981-2989.

Misra, H. P. and I. Fridovich (1972) The generation of superoxide radical during the autoxidation of hemoglobin. *J Biol Chem* **247**: 6960-6962.

Mohapatra, N. P., S. Soni, B. L. Bell, R. Warren, R. K. Ernst, A. Muszynski, R. W. Carlson, *et al. (*2007) Identification of an orphan response regulator required for the virulence of *Francisella* spp. and transcription of pathogenicity island genes. *Infect Immun* **75**: 3305-3314.

Molendijk, A. J., B. Ruperti and K. Palme (2004) Small GTPases in vesicle trafficking. *Curr Opin Plant Biol* **7**: 694-700.

Moncrief, M. and R. Hausinger (1997) Characterization of UreG, identification of a UreD-UreF-UreG complex, and evidence suggesting that a nucleotide-binding site in UreG is required for *in vivo* metallocenter assembly of *Klebsiella aerogenes* urease. *J Bacteriol* **179**: 4081-4086.

Morett, E., J. Korbel, E. Rajan, G. Saab-Rincon, L. Olvera, M. Olvera, S. Steffen, *et al. (*2003) Systematic discovery of analogous enzymes in thiamin biosynthesis. *Nature Biotech* **21**: 790-795.

Morett, E., G. Saab-Rincón, L. Olvera, M. Olvera, H. Flores and R. Grande (2008) Sensitive genome-wide screen for low secondary enzymatic activities: the YjbQ family shows thiamin phosphate synthase activity. *J Mol Biol* **376**: 839-853.

Motley, S. T., B. J. Morrow, X. Liu, I. L. Dodge, A. Vitiello, C. K. Ward and K. J. Shaw (2004) Simultaneous analysis of host and pathogen interactions during an in vivo infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment. *Cell Microbiol* **6**: 849-865.

Moura, J. J. G., M. Teixeira, I. Moura, A. V. Xavier and J. LeGall (1984) Nickel-A redox active catalytic site in hydrogenase. *J. Mol. Catal.* **23**: 303-314.

Musiani, F., B. Zambelli, M. Stola and S. Ciurli (2004) Nickel trafficking: insights into the fold and function of UreE, a urease metallochaperone. *J Inorg Biochem* **98**: 803-813.

Myers, L. C., M. P. Terranova, A. E. Ferentz, G. Wagner and G. L. Verdine (1993) Repair of DNA methylphosphotriesters through a metalloactivated cysteine nucleophile. *Science* **261**: 1164-1167.

Nahvi, A., N. Sudarsan, M. Ebert, X. Zou, K. Brown and R. Breaker (2002) Genetic control by a metabolite binding mRNA. *Chem Biol* **9**: 1043-1049.

Nanamiya, H., G. Akanuma, Y. Natori, R. Murayama, S. Kosono, T. Kudo, K. Kobayashi, *et al. (*2004) Zinc is a key factor in controlling alternation of two types of L31 protein in the *Bacillus subtilis* ribosome. *Mol Microbiol* **52**: 273-283.

Nandi, D. L. and D. Shemin (1968) Delta-aminolevulinic acid dehydratase of Rhodopseudomonas spheroides. 3. Mechanism of porphobilinogen synthesis. *J Biol Chem* **243**: 1236-1242.

Naponelli, V., A. Noiriel, M. J. Ziemak, S. M. Beverley, L. F. Lye, A. M. Plume, J. R. Botella, *et al. (*2008) Phylogenomic and functional analysis of pterin-4a-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiol* **146**: 1515-1527.

Narumiya, S. (1996) The small GTPase Rho: Cellular functions and signal transduction. *J Biochem* **120**: 215-228.

Natori, Y., H. Nanamiya, G. Akanuma, S. Kosono, T. Kudo, K. Ochi and F. Kawamura (2007) A fail-safe system for the ribosome under zinc-limiting conditions in *Bacillus subtilis. Mol Microbiol* **63**: 294-307.

Needleman, S. B. and C. D. Wunsch (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443-453.

Neilands, J. B. (1995) Siderophores: structure and function of microbial iron transport compounds. *J Biol Chem* **270**: 26723-26726.

Neu, H. (1967) 5'-Nucleotidase of *Escherichia coli* .I. Purification and properties. *J Biol Chem* **242**: 3896-3904.

Niederhoffer, E. C., C. M. Naranjo, K. L. Bradley and J. A. Fee (1990) Control of *Escherichia coli* superoxide dismutase (*sodA* and *sodB*) genes by the ferric uptake regulation (*fur*) locus. *J Bacteriol* **172**: 1930-1938.

Nielubowicz, G. R., S. N. Smith and H. L. Mobley (2010) Zinc uptake contributes to motility and provides a competitive advantage to *Proteus mirabilis* during experimental urinary tract infection. *Infect Immun* **78**: 2823-2833.

Nies, D. H. (1992) CzcR and CzcD, gene products affecting regulation of resistance to cobalt, zinc, and cadmium (czc system) in *Alcaligenes eutrophus*. *J Bacteriol* **174**: 8102-8110.

Nies, D. H. (1995) The cobalt, zinc, and cadmium efflux system CzcABC from *Alcaligenes eutrophus* functions as a cation-proton antiporter in *Escherichia coli*. *J Bacteriol* **177**: 2707-2712.

Nishiyama, M., S. Horinouchi, M. Kobayashi, T. Nagasawa, H. Yamada and T. Beppu (1991) Cloning and characterization of genes responsible for metabolism of nitrile compounds from *Pseudomonas chlororaphis* B23. *J Bacteriol* **173**: 2465-2472.

Nojiri, M., H. Nakayama, M. Odaka, M. Yohda, K. Takio and I. Endo (2000) Cobalt-substituted Fe-type nitrile hydratase of *Rhodococcus* sp. N-771. *FEBS Lett* **465**: 173-177.

Nojiri, M., M. Yohda, M. Odaka, Y. Matsushita, M. Tsujimura, T. Yoshida, N. Dohmae, *et al.* (1999) Functional expression of nitrile hydratase in *Escherichia coli*: requirement of a nitrile hydratase activator and post-translational modification of a ligand cysteine. *J Biochem* **125**: 696-704.

Novichkov, P., O. Laikova, E. Novichkova, M. Gelfand, A. Arkin, I. Dubchak and D. Rodionov (2010) RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res* **38**: D111-118.

O'Halloran, T. and V. Culotta (2000) Metallochaperones, an intracellular shuttle service for metal ions. *J Biol Chem* **275**: 25057-25060.

O'Mahony, R., J. Doran, L. Coffey, O. J. Cahill, G. W. Black and C. O'Reilly (2005) Characterization of the nitrile hydratase gene clusters of *Rhodococcus erythropolis* strains AJ270 and AJ300 and *Microbacterium* sp. AJ115 indicates horizontal gene transfer and reveals an insertion of IS1166. *Anton Leeuw Int J G* **87**: 221-232.

Ogawa, J. and S. Shimizu (1995) Purification and characterization of dihydroorotase from *Pseudomonas putida*. *Arch of Microbiol* **164**: 353-357.

Olson, J., C. Fu and R. Maier (1997) The HypB protein from *Bradyrhizobium japonicum* can store nickel and is required for the nickel-dependent transcriptional regulation of hydrogenase. *Mol Microbiol* **24**: 119-128.

Olson, J. W., N. S. Mehta and R. J. Maier (2001) Requirement of nickel metabolism proteins HypA and HypB for full activity of both hydrogenase and urease in *Helicobacter pylori*. *Mol Microbiol* **39**: 176-182.

Oparin, A. I. (1938) *Origin of life*. New York: Macmillan Co.

Ose, D. and I. Fridovich (1979) Manganese-containing superoxide dismutase from *Escherichia coli*: reversible resolution and metal replacements. *Arch Biochem Biophys* **194**: 360-364.

Osterman, A. and R. Overbeek (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* **7**: 238-251.

Outten, C. and T. O'Halloran (2001) Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* **292**: 2488-2492.

Outten, C. E., F. W. Outten and T. V. O'Halloran (1999) DNA distortion mechanism for transcriptional activation by ZntR, a Zn(II)-responsive MerR homologue in *Escherichia coli*. *J Biol Chem* **274**: 37517-37524.

Outten, C. E., D. A. Tobin, J. E. Penner-Hahn and T. V. O'Halloran (2001a) Characterization of the metal receptor sites in *Escherichia coli* Zur, an ultrasensitive zinc(II) metalloregulatory protein. *Biochemistry* **40**: 10417-10423.

Outten, F. W., D. L. Huffman, J. A. Hale and T. V. O'Halloran (2001b) The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *J Biol Chem* **276**: 30670-30677.

Outten, F. W., C. E. Outten, J. Hale and T. V. O'Halloran (2000) Transcriptional activation of an *Escherichia coli* copper efflux regulon by the chromosomal MerR homologue, CueR. *J Biol Chem* **275**: 31024-31029.

Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, V. de Crécy-Lagard, *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: 5691-5702.

Overbeek, R., M. Fonstein, M. D'Souza, G. Pusch and N. Maltsev (1999a) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A* **96**: 2896-2901.

Overbeek, R., M. Fonstein, M. D'Souza, G. Pusch and N. Maltsev (1999b) Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol* **1**: 93-108.

Owen, G., B. Pascoe, D. Kallifidas and M. Paget (2007) Zinc-responsive regulation of alternative ribosomal protein genes in *Streptomyces coelicolor* involves *zur* and *sigmaR*. *J Bacteriol* **189**: 4078-4086.

Padovani, D. and R. Banerjee (2006) Assembly and protection of the radical enzyme, methylmalonyl-CoA mutase, by its chaperone. *Biochemistry* **45**: 9300-9306.

Padovani, D. and R. Banerjee (2009) A G-protein editor gates coenzyme B12 loading and is corrupted in methylmalonic aciduria. *Proc Natl Acad Sci U S A* **106**: 21567-21572.

Padovani, D., T. Labunska and R. Banerjee (2006) Energetics of interaction between the G-protein chaperone, MeaB, and B_{12}-dependent methylmalonyl-CoA mutase. *J Biol Chem* **281**: 17838-17844.

Padovani, D., T. Labunska, B. A. Palfey, D. P. Ballou and R. Banerjee (2008) Adenosyltransferase tailors and delivers coenzyme B_{12}. *Nat Chem Biol* **4**: 194-196.

Pai, E., U. Krengel, G. Petsko, R. Goody, W. Kabsch and A. Wittinghofer (1990) Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *EMBO J* **9**: 2351-2359.

Pal, C. and L. Hurst (2004) Evidence against the selfish operon theory. *Trends Genet* **20**: 232-234.

Palmer, K., L. Mashburn, P. Singh and M. Whiteley (2005) Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. *J Bacteriol* **187**: 5267-5277.

Palmiter, R. D. and S. D. Findley (1995) Cloning and functional characterization of a mammalian zinc transporter that confers resistance to zinc. *EMBO J* **14**: 639-649.

Palsson, B (2006) *Systems biology: properties of reconstructed networks*. New York: Cambridge University Press.

Panina, E., A. Mironov and M. Gelfand (2001) Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res* **29**: 5195-5206.

Panina, E. M., A. A. Mironov and M. S. Gelfand (2003) Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc Natl Acad Sci U S A* **100**: 9912-9917.

Parr, R. and R. Pearson (1983) Absolute hardness: companion parameter to absolute electronegativity. *J Am ChemSoc* **105**: 7512-7516.

Pasquali, P., S. Ammendola, C. Pistoia, P. Petrucci, M. Tarantino, C. Valente, M. L. Marenzoni, *et al.* (2008) Attenuated *Salmonella enterica* serovar *Typhimurium l*acking the ZnuABC transporter confers immune-based protection against challenge infections in mice. *Vaccine* **26**: 3421-3426.

Patzer, S. and K. Hantke (1998) The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*. *Mol Microbiol* **28**: 1199-1210.

Patzer, S. and K. Hantke (2000) The zinc-responsive regulator Zur and its control of the *znu* gene cluster encoding the ZnuABC zinc uptake system in *Escherichia coli*. *J Biol Chem* **275**: 24321-24332.

Paul, B., M. Barker, W. Ross, D. Schneider, C. Webb, J. Foster and R. Gourse (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118**: 311-322.

Paul, B., M. Berkmen and R. Gourse (2005) DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci U S A* **102**: 7823-7828.

Peariso, K., C. Goulding, S. Huang, R. Matthews and J. Penner-Hahn (1998) Characterization of the zinc binding site in methionine synthase enzymes of *Escherichia coli*: The role of zinc in the methylation of homocysteine. *J Am Chem Soc* **120**: 8410-8416.

Peers, G. and N. Price (2006) Copper-containing plastocyanin used for electron transport by an oceanic diatom. *Nature* **441**: 341-344.

Pellegrini, M., E. Marcotte, M. Thompson, D. Eisenberg and T. Yeates (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* **96**: 4285-4288.

Penfold, D. W., C. F. Forsterand and M. L. E. (2003) Increased hydrogen production by *Escherichia coli* strain HD701 in comparison with the wild-type parent strain MC4100. *Enzyme Microb Technol* **33**: 185-189.

Percy, M. J., N. V. McFerran and T. R. Lappin (2005) Disorders of oxidised haemoglobin. *Blood Rev* **19**: 61-68.

Perederina, A., V. Svetlov, M. Vassylyeva, T. Tahirov, S. Yokoyama, I. Artsimovitch and D. Vassylyev (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* **118**: 297-309.

Permina, E. A., A. E. Kazakov, O. V. Kalinina and M. S. Gelfand (2006) Comparative genomics of regulation of heavy metal resistance in Eubacteria. *BMC Microbiol* **6**: 49.

Perron, K., R. Comte and C. van Delden (2005) DksA represses ribosomal gene transcription in *Pseudomonas aeruginosa* by interacting with RNA polymerase on ribosomal promoters. *Mol Microbiol* **56**: 1087-1102.

Peters, J., T. Thate and N. Craig (2003) Definition of the *Escherichia coli* MC4100 genome by use of a DNA array. *J Bacteriol* **185**: 2017-2021.

Petrarca, P., S. Ammendola, P. Pasquali and A. Battistoni (2010) The Zur-regulated ZinT protein is an auxiliary component of the high-affinity ZnuABC zinc transporter that facilitates metal recruitment during severe zinc shortage. *J Bacteriol* **192**: 1553-1564.

Pflüger, K., A. Ehrenreich, K. Salmon, R. P. Gunsalus, U. Deppenmeier, G. Gottschalk and V. Müller (2007) Identification of genes involved in salt adaptation in the archaeon *Methanosarcina mazei* Gö1 using genome-wide gene expression profiling. *FEMS Microbiol Lett* **277**: 79-89.

Phillips, G., V. Chikwana, A. Maxwell, B. El-Yacoubi, M. Swairjo, D. Iwata-Reuyl and V. de Crécy-Lagard (2010) Discovery and characterization of an amidinotransferase involved in the modification of archaeal tRNA. *J Biol Chem* **285**: 12706-12713.

Pinchuk, G., D. Rodionov, C. Yang, X. Li, A. Osterman, E. Dervyn, O. Geydebrekht, *et al.* (2009) Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. *Proc Natl Acad Sci U S A* **106**: 2874-2879.

Pinske, C. and R. Sawers (2010) The role of the ferric-uptake regulator Fur and iron homeostasis in controlling levels of the [NiFe]-hydrogenases in *Escherichia coli*. *Int J Hydrogen Energ* **35**: 8938-8944.

Plumb, R., A. Martell and F. Bersworth (1950) Spectrophotometric determination of displacement series of metal complexes of the sodium salts of ethylenediaminetetraacetic acid. *J Phys Colloid Chem* **54**: 1208-1215.

Poptsova, M. and J. Gogarten (2010) Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology* **156**: 1909-1917.

Porter, T., Y. Li and F. Raushel (2004) Mechanism of the dihydroorotase reaction. *Biochemistry* **43**: 16285-16292.

Poschenrieder, C., R. Tolrà and J. Barceló (2006) Can metals defend plants against biotic stress? *Trends Plant Sci* **11**: 288-295.

Potrykus, K. and M. Cashel (2008) (p)ppGpp: still magical? *Annu Rev Microbiol* **62**: 35-51.

Potrykus, K., D. Vinella, H. Murphy, A. Szalewska-Palasz, R. D'Ari and M. Cashel (2006) Antagonistic regulation of *Escherichia coli* ribosomal RNA rrnB P1 promoter activity by GreA and DksA. *J Biol Chem* **281**: 15238-15248.

Predki, P. F. and B. Sarkar (1994) Metal replacement in "zinc finger" and its effect on DNA binding. *Environ Health Perspect* **102**: 195-198.

Price, M., K. Huang, A. Arkin and E. Alm (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* **15**: 809-819.

Priest, J., J. Watterson, R. Jones, A. Faassen and B. Hedlund (1989) Mutant fetal hemoglobin causing cyanosis in a newborn. *Pediatrics* **83**: 734-736.

Proudfoot, M., E. Kuznetsova, G. Brown, N. Rao, M. Kitagawa, H. Mori, A. Savchenko, *et al.* (2004) General enzymatic screens identify three new nucleotidases in *Escherichia coli*: Biochemical characterization of SurE, YfbR, and YjjG. *J BiolChem* **279**: 54687-54694.

Pruitt, K. D., T. Tatusova and D. R. Maglott (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**: D61-65.

Pufahl, R. A., C. P. Singer, K. L. Peariso, S. J. Lin, P. J. Schmidt, C. J. Fahrni, V. C. Culotta, *et al.* (1997) Metal ion chaperone function of the soluble Cu(I) receptor Atx1. *Science* **278**: 853-856.

Qin, Z., L. McCue, W. Thompson, L. Mayerhofer, C. Lawrence and J. Liu (2003) Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites. *Nature Biotech* **21**: 435-439.

Qiu, D., F. Damron, T. Mima, H. Schweizer and H. Yu (2008) P_{BAD}-based shuttle vectors for functional analysis of toxic and highly regulated genes in *Pseudomonas* and *Burkholderia* spp. and other bacteria. *Appl Environ Microbiol* **74**: 7422-7426.

Racker, E. (1951) The mechanism of action of glyoxalase. *J Biol Chem* **190**: 685-696.

Rae, T., P. Schmidt, R. Pufahl, V. Culotta and T. O'Halloran (1999) Undetectable intracellular free copper: the requirement of a copper chaperone for superoxide dismutase. *Science* **284**: 805-808.

Rajagopalan, P., X. Yu and D. Pei (1997) Peptide deformylase: A new type of mononuclear iron protein. *J Am Chem Soc* **119**: 12418-12419.

Ravanel, S., B. Gakiere, D. Job and R. Douce (1998) The specific features of methionine biosynthesis and metabolism in plants. *Proc Natl Acad Sci U S A* **95**: 7805-7812.

Reader, J. S., D. Metzgar, P. Schimmel and V. de Crécy-Lagard (2004) Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J Biol Chem* **279**: 6280-6285.

Reddi, A. R. and B. R. Gibney (2007) Role of protons in the thermodynamic contribution of a Zn(II)-Cys4 site toward metalloprotein stability. *Biochemistry* **46**: 3745-3758.

Reed, J., I. Famili, I. Thiele and B. Palsson (2006a) Towards multidimensional genome annotation. *Nat Rev Genet* **7**: 130-141.

Reed, J., T. Patel, K. Chen, A. Joyce, M. Applebee, C. Herring, O. Bui, *et al.* (2006b) Systems approach to refining genome annotation. *Proc Natl Acad Sci U S A* **103**: 17480-17484.

Remaut, H., N. Safarov, S. Ciurli and J. Van Beeumen (2001) Structural basis for Ni(2+) transport and assembly of the urease active site by the metallochaperone UreE from Bacillus pasteurii. *J Biol Chem* **276**: 49365-49370.

Rensing, C. and G. Grass (2003) *Escherichia coli* mechanisms of copper homeostasis in a changing environment. *FEMS Microbiol Rev* **27**: 197-213.

Rensing, C., B. Mitra and B. P. Rosen (1997) The zntA gene of *Escherichia coli* encodes a Zn(II)-translocating P-type ATPase. *Proc Natl Acad Sci U S A* **94**: 14326-14331.

Reyes-Caballero, H., A. J. Guerra, F. E. Jacobsen, K. M. Kazmierczak, D. Cowart, U. M. Koppolu, R. A. Scott, *et al.* (2010) The metalloregulatory zinc site in *Streptococcus pneumoniae* AdcR, a zinc-activated MarR family repressor. *J Mol Biol* **403**: 197-216.

Rice, P., I. Longden and A. Bleasby (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276-277.

Richards, M. P. and R. J. Cousins (1975) Mammalian zinc homeostasis: requirement for RNA and metallothionein synthesis. *Biochem Biophys Res Commun* **64**: 1215-1223.

Robinson, N., S. Whitehall and J. Cavet (2001) Microbial metallothioneins. *Adv Microb Physiol* **44**: 183-213.

Robinson, N., K. Waldron, S. Tottey and C. Bessant (2008) Metalloprotein metal pools: identification and quantification by coupling native and non-native separations through principal component analysis. *Protocol Exchange* doi:10.1038/nprot.2008.236.

Robinson, N. and D. Winge (2010) Copper Metallochaperones. *Annu Rev Biochem* **79**: 537-562.

Rodionov, D., J. De Ingeniis, C. Mancini, F. Cimadamore, H. Zhang, A. Osterman and N. Raffaelli (2008) Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators. *Nucleic Acids Res* **36**: 2047-2059.

Rodionov, D., M. Gelfand, J. Todd, A. Curson and A. Johnston (2006a) Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria. *PLoS Comput Biol* **2**: 1568-1585.

Rodionov, D., P. Hebbeln, A. Eudes, J. ter Beek, I. Rodionova, G. Erkens, D. Slotboom, *et al.* (2009) A novel class of modular transporters for vitamins in prokaryotes. *J Bacteriol* **191**: 42-51.

Rodionov, D., P. Hebbeln, M. Gelfand and T. Eitinger (2006b) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. *J Bacteriol* **188**: 317-327.

Rodionov, D., O. Kurnasov, B. Stec, Y. Wang, M. Roberts and A. Osterman (2007) Genomic identification and *in vitro* reconstitution of a complete biosynthetic pathway for the osmolyte di-myo-inositol-phosphate. *Proc Natl Acad Sci U S A* **104**: 4279-4284.

Rodionov, D. A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* **107**: 3467-3497.

Rodionov, D. A., A. G. Vitreschak, A. A. Mironov and M. S. Gelfand (2003) Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. *J Biol Chem* **278**: 41148-41159.

Rodnina, M., H. Stark, A. Savelsbergh, H. Wieden, D. Mohr, N. Matassova, F. Peske, *et al.* (2000) GTPase mechanisms and functions of translation factors on the ribosome. *Biol Chem* **381**: 377-387.

Ross-Macdonald, P. and G. S. Roeder (1994) Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. *Cell* **79**: 1069-1080.

Rowe, J. L., G. L. Starnes and P. T. Chivers (2005) Complex transcriptional control links NikABCDE-dependent nickel transport with hydrogenase expression in *Escherichia coli*. *J Bacteriol* **187**: 6317-6323.

Rulísek, L. and J. Vondrásek (1998) Coordination geometries of selected transition metal ions (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) in metalloproteins. *J Inorg Biochem* **71**: 115-127.

Rutherford, S., C. Villers, J. Lee, W. Ross and R. Gourse (2009) Allosteric control of *Escherichia coli* rRNA promoter complexes by DksA. *Genes Dev* **23**: 236-248.

Sachadyn, P. (2010) Conservation and diversity of MutS proteins. *Mutat Res* **694**: 20-30.

Salomone-Stagni, M., B. Zambelli, F. Musiani and S. Ciurli (2007) A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis. *Proteins* **68**: 749-761.

Salzberg, S. L. (2007) Genome re-annotation: a wiki solution? *Genome Biol* **8**: 102.

Sambrook, J. and D. Russell (2001) *Molecular Cloning: a laboratory manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sander, E., L. Wright and D. McCormick (1965) Evidence for function of a metal ion in the activity of dihydroorotase from *Zymobacterium oroticum*. *J Biol Chem* **240**: 3628-3630.

Sankaran, B., S. A. Bonnett, K. Shah, S. Gabriel, R. Reddy, P. Schimmel, D. A. Rodionov, *et al.* (2009) Zinc-independent folate biosynthesis: genetic, biochemical, and structural investigations reveal new metal dependence for GTP cyclohydrolase IB. *J Bacteriol* **191**: 6936-6949.

Sankaranarayanan, R., A. C. Dock-Bregeon, B. Rees, M. Bovee, J. Caillet, P. Romby, C. S. Francklyn, *et al.* (2000) Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat Struct Biol* **7**: 461-465.

Saraste, M., P. R. Sibbald and A. Wittinghofer (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* **15**: 430-434.

Sauer, K., U. Harms and R. K. Thauer (1997) Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*. Purification, properties and encoding genes of the corrinoid protein MT1. *Eur J Biochem* **243**: 670-677.

Sauer, K. and R. Thauer (1997) Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*. Zinc dependence and thermodynamics of the methanol:cob(I)alamin methyltransferase reaction. *Eur J Biochem* **249**: 280-285.

Schmidt, P., C. Kunst and V. Culotta (2000) Copper activation of superoxide dismutase 1 (SOD1) *in vivo*: Role for protein-protein interactions with the copper chaperone for SOD1. *J Biol Chem* **275**: 33771-33776.

Schneider, T. and R. Stephens (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **18**: 6097-6100.

Schnoes, A., S. Brown, I. Dodevski and P. Babbitt (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* **5**: e1000605.

Schurr, M., J. Vickrey, A. Kumar, A. Campbell, R. Cunin, R. Benjamin, M. Shanley, *et al.* (1995) Aspartate transcarbamoylase genes of *Pseudomonas putida*: requirement for an inactive dihydroorotase for assembly into the dodecameric holoenzyme. *J Bacteriol* **177**: 1751-1759.

Schweizer, H. and T. Hoang (1995) An improved system for gene replacement and *xylE* fusion analysis in *Pseudomonas aeruginosa*. *Gene* **158**: 15-22.

ScotT, A. (1993) How nature synthesizes vitamin B_{12} - a survey of the last 4 billion years. *Angew Chemie* **32**: 1223-1243.

Scrutton, M. C., C. W. Wu and D. A. Goldthwait (1971) The presence and possible role of zinc in RNA polymerase obtained from *Escherichia coli*. *Proc Natl Acad Sci U S A* **68**: 2497-2501.

Segev, N. (2010) GTPases in intracellular trafficking: An overview. *Semin Cell Dev Biol* **22**: 1-2.

Serres, M. and M. Riley (2006) Genomic analysis of carbon source metabolism of *Shewanella oneidensis* MR-1: Predictions versus experiments. *J Bacteriol* **188**: 4601-4609.

Shapiro, L. and R. Losick (1997) Protein localization and cell fate in bacteria. *Science* **276**: 712-718.

Shearer, J., I. Kung, S. Lovell, W. Kaminsky and J. Kovacs (2001) Why is there an "Inert" metal center in the active site of nitrile hydratase? Reactivity and ligand dissociation from a five-coordinate Co(III) nitrile hydratase model. *J Am Chem Soc* **123**: 463-468.

Shepard, E. M., S. E. McGlynn, A. L. Bueling, C. S. Grady-Smith, S. J. George, M. A. Winslow, S. P. Cramer, *et al.* (2010) Synthesis of the 2Fe subcluster of the [FeFe]-hydrogenase H cluster on the HydF scaffold. *Proc Natl Acad Sci U S A* **107**: 10448-10453.

Shi, H., K. Bencze, T. Stemmler and C. Philpott (2008) A cytosolic iron chaperone that delivers iron to ferritin. *Science* **320**: 1207-1210.

Shin, J. H., S. Y. Oh, S. J. Kim and J. H. Roe (2007) The zinc-responsive regulator Zur controls a zinc uptake system and some ribosomal proteins in *Streptomyces coelicolor* A3(2). *J Bacteriol* **189**: 4070-4077.

Sigdel, T., J. Easton and M. Crowder (2006) Transcriptional response of *Escherichia coli* to TPEN. *J Bacteriol* **188**: 6709-6713.

Siiman, O., N. M. Young and P. R. Carey (1974) Resonance Raman studies of "blue" copper proteins. *J Am Chem Soc* **96**: 5583-5585.

Simmer, J., R. Kelly, A. Rinker, B. Zimmermann, J. Scully, H. Kim and D. Evans (1990) Mammalian dihydroorotase: nucleotide-sequence, peptide sequences, and evolution of the dihydroorotase domain of the multifunctional protein CAD. *Proc Nat Acad Sci U S A* **87**: 174-178.

Singh, A. and P. Bragg (1974) Inhibition of energization of *Salmonella typhimurium* membrane by zinc ions. *FEBS Lett* **40**: 200-202.

Smith, T. F., C. Gaitatzes, K. Saxena and E. J. Neer (1999) The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* **24**: 181-185.

Sokabe, M., T. Ose, A. Nakamura, K. Tokunaga, O. Nureki, M. Yao and I. Tanaka (2009) The structure of alanyl-tRNA synthetase with editing domain. *Proc Natl Acad Sci U S A* **106**: 11028-11033.

Song, H. K., S. B. Mulrooney, R. Huber and R. P. Hausinger (2001) Crystal structure of *Klebsiella aerogenes* UreE, a nickel-binding metallochaperone for urease activation. *J Biol Chem* **276**: 49359-49364.

Sood, S. M., M. X. Wu, K. A. Hill and C. W. Slattery (1999) Characterization of zinc-depleted alanyl-tRNA synthetase from *Escherichia coli*: role of zinc. *Arch Biochem Biophys* **368**: 380-384.

Soriano, A., G. J. Colpas and R. P. Hausinger (2000) UreE stimulation of GTP-dependent urease activation in the UreD-UreF-UreG-urease apoprotein complex. *Biochemistry* **39**: 12435-12440.

Soriano, A. and R. P. Hausinger (1999) GTP-dependent activation of urease apoprotein in complex with the UreD, UreF, and UreG accessory proteins. *Proc Natl Acad Sci U S A* **96**: 11140-11144.

Srivatsan, A. and J. Wang (2008) Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* **11**: 100-105.

Stepanova, E., A. Shevelev, S. Borukhov and K. Severinov (2009) Mechanisms of action of RNA polymerase-binding transcription factors that do not bind to DNA. *Biofizika* **54**: 773-790.

Stevens, J. M., N. Rao Saroja, M. Jaouen, M. Belghazi, J. M. Schmitter, D. Mansuy, I. Artaud, *et al.* (2003) Chaperone-assisted expression, purification, and characterization of recombinant nitrile hydratase NI1 from *Comamonas testosteroni*. *Protein Expr Purif* **29**: 70-76.

Stingl, K., K. Schauer, C. Ecobichon, A. Labigne, P. Lenormand, J. C. Rousselle, A. Namane, *et al.* (2008) *In vivo* interactome of *Helicobacter pylori* urease revealed by tandem affinity purification. *Mol Cell Proteomics* **7**: 2429-2441.

Stohs, S. and D. Bagchi (1995) Oxidative mechanisms in the toxicity of metal ions. *Free Radic Biol Med* **18**: 321-336.

Stormo, G. (2000) DNA binding sites: representation and discovery. *Bioinformatics* **16**: 16-23.

Stormo, G., T. Schneider, L. Gold and A. Ehrenfeucht (1982) Use of the perceptron algorithm to distinguish translational initiation sites in *Escherichia coli*. *Nucleic Acids Res* **10**: 2997-3011.

Stover, C., X. Pham, A. Erwin, S. Mizoguchi, P. Warrener, M. Hickey, F. Brinkman, *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**: 959-964.

Stratmann, J., B. Strommenger, R. Goethe, K. Dohmann, G. F. Gerlach, K. Stevenson, L. L. Li, *et al.* (2004) A 38-kilobase pathogenicity island specific for *Mycobacterium avium* subsp. *paratuberculosis* encodes cell surface proteins expressed in the host. *Infect Immun* **72**: 1265-1274.

Strzepek, R. and P. Harrison (2004) Photosynthetic architecture differs in coastal and oceanic diatoms. *Nature* **431**: 689-692.

Stumm, W. and J. J. Morgan (1996) *Aquatic chemistry*. New York: John Wiley and Sons, Inc.

Su, S. S. and P. Modrich (1986) *Escherichia coli* mutS-encoded protein binds to mismatched DNA base pairs. *Proc Natl Acad Sci U S A* **83**: 5057-5061.

Suthers, P. F., A. Zomorrodi and C. D. Maranas (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol Syst Biol* **5**: 301.

Szpunar, J. (2004) Metallomics: a new frontier in analytical chemistry. *Anal Bioanal Chem* **378**: 54-56.

Takahashi-Íñiguez, T., H. García-Arellano, M. A. Trujillo-Roldán and M. E. Flores (2010) Protection and reactivation of human methylmalonyl-CoA mutase by MMAA protein. *Biochem Biophys Res Commun* **404**: 443-447.

Tan, K., G. Moreno-Hagelsieb, J. Collado-Vides and G. D. Stormo (2001) A comparative genomics approach to prediction of new members of regulons. *Genome Res* **11**: 566-584.

Tanaka, Y., K. Tsumoto, E. Tanabe, Y. Yasutake, N. Sakai, M. Yao, I. Tanaka, *et al.* (2005) Crystal structure of the hypothetical protein ST2072 from *Sulfolobus tokodaii*. *Proteins* **61**: 1127-1131.

Tang, D. J., X. J. Li, Y. Q. He, J. X. Feng, B. Chen and J. L. Tang (2005) The zinc uptake regulator Zur is essential for the full virulence of *Xanthomonas campestris* pv. campestris. *Mol Plant Microbe Interact* **18**: 652-658.

Taylor, W., M. Taylor, W. Balch and P. Gilchrist (1976) Purification of properties of dihydroorotase, a zinc-containing metalloenzyme in *Clostridium oroticum*. *J Bacteriol* **127**: 863-873.

Tehranchi, A., M. Blankschien, Y. Zhang, J. Halliday, A. Srivatsan, J. Peng, C. Herman, *et al.* (2010) The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* **141**: 595-605.

Thelwell, C., N. Robinson and J. Turner-Cavet (1998) An SmtB-like repressor from *Synechocystis* PCC 6803 regulates a zinc exporter. *Proc Nat Acad Sci U S A* **95**: 10728-10733.

Thoden, J., G. J. Phillips, T. Neal, F. Raushel and H. Holden (2001) Molecular structure of dihydroorotase: a paradigm for catalysis through the use of a binuclear metal center. *Biochemistry* **40**: 6989-6997.

Thompson, J. D., T. J. Gibson and D. G. Higgins (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* Chapter 2: Unit 2.3.

Thöny, B., G. Auerbach and N. Blau (2000) Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem J* **347**: 1-16.

Torreilles, J. and M. C. Guérin (1990) Nickel (II) as a temporary catalyst for hydroxyl radical generation. *FEBS Lett* **272**: 58-60.

Torrents, E., I. Grinberg, B. Gorovitz-Harris, H. Lundstrom, I. Borovok, Y. Aharonowitz, B. Sjoberg, e*t al.* (2007) NrdR controls differential expression of the *Escherichia coli* ribonucleotide reductase genes. *J Bacteriol* **189**: 5012-5021.

Tottey, S., D. Harvie and N. Robinson (2005) Understanding how cells allocate metals using metal sensors and metallochaperones. *Acc Chem Res* **38**: 775-783.

Tottey, S., D. Harvie and N. Robinson (2007) Understanding How Cells Allocate Metals. In: Molecular Microbiology of Heavy Metals*.* D. H. N. a. S. Silver (ed). Berlin: Springer, pp. 3-35.

Tottey, S., K. Waldron, S. Firbank, B. Reale, C. Bessant, K. Sato, T. Cheek, *et al.* (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **455**: 1138-1142.

Trautinger, B., R. Jaktaji, E. Rusakova and R. Lloyd (2005) RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol Cell* **19**: 247-258.

Tripp, B. C., C. B. Bell, F. Cruz, C. Krebs and J. G. Ferry (2004) A role for iron in an ancient carbonic anhydrase. *J Biol Chem* **279**: 6683-6687.

Twine, S. M., N. C. Mykytczuk, M. D. Petit, H. Shen, A. Sjöstedt, J. Wayne Conlan and J. F. Kelly (2006) *In vivo* proteomic analysis of the intracellular bacterial pathogen, *Francisella tularensis*, isolated from mouse spleen. *Biochem Biophys Res Commun* **345**: 1621-1633.

Vallee, B. L. and D. S. Auld (1990) Active-site zinc ligands and activated H_2O of zinc enzymes. *Proc Natl Acad Sci U S A* **87**: 220-224.

Vallee, B. L. and R. J. Williams (1968) Metalloenzymes: the entatic nature of their active sites. *Proc Natl Acad Sci U S A* **59**: 498-505.

Venkateswaran, K., D. Moser, M. Dollhopf, D. Lies, D. Saffarini, B. MacGregor, D. Ringelberg, *et al.* (1999) Polyphasic taxonomy of the genus Shewanella and description of *Shewanella oneidensis* sp. nov. *Int J Syst Bacteriol* **49**: 705-724.

Vetter, I. and A. Wittinghofer (2001) The guanine nucleotide-binding switch in three dimensions. *Science* **294**: 1299-1304.

Viducic, D., T. Ono, K. Murakami, H. Susilowati, S. Kayama, K. Hirota and Y. Miyake (2006) Functional analysis of *spoT, relA* and *dksA* genes on quinolone tolerance in *Pseudomonas aeruginosa* under nongrowing condition. *Microbiol Immunol* **50**: 349-357.

Vignais, P., B. Billoud and J. Meyer (2001) Classification and phylogeny of hydrogenases. *FEMS Microbiol Rev* **25**: 455-501.

Volbeda, A., M. H. Charon, C. Piras, E. C. Hatchikian, M. Frey and J. C. Fontecilla-Camps (1995) Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**: 580-587.

Wagner, E., S. Altuvia and P. Romby (2002) Antisense RNAs in bacteria and their genetic elements. *Adv Genet* **46**: 361-398.

Waldron, K. J. and N. J. Robinson (2009) How do bacterial cells ensure that metalloproteins get the correct metal? *Nat Rev Microbiol* **7**: 25-35.

Washabaugh, M. and K. Collins (1984) Dihydroorotase from *Escherichia coli*. Purification and characterization. *J Biol Chem* **259**: 3293-3298.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191.

Waugh, R. and D. Boxer (1986) Pleiotropic hydrogenase mutants of *Escherichia coli* K12: growth in the presence of nickel can restore hydrogenase activity. *Biochimie* **68**: 157-166.

Weinberg, E. (1975) Metal starvation of pathogens by hosts. *Bioscience* **25**: 314-318.

Werner, A. (1913) Information on the assymetric cobalt atom. IX. *Berichte Der Deutschen Chemischen Gesellschaft*: 3674-3683.

Westin, G. and W. Schaffner (1988) A zinc-responsive factor interacts with a metal-regulated enhancer element (MRE) of the mouse metallothionein-I gene. *EMBO J* **7**: 3763-3770.

Wheelwright, E. J., Spedding, F. H. , Schwarzenbach, G. (1953) The stability of the rare earth complexes with ethylenediaminetetraacetic acid. *J Am Chem Soc* **75**: 4196-4201.

Williams, N., M. Manthey, T. Hambley, S. Odonoghue, M. Keegan, B. Chapman and R. Christopherson (1995) Catalysis by hamster dihydroorotase: zinc-binding, site-directed mutagenesis, and interaction with inhibitors. *Biochemistry* **34**: 11344-11352.

Williams, R. (1997) The natural selection of the chemical elements. *Cell Mol Life Sci* **53**: 816-829.

Winsor, G., T. Van Rossum, R. Lo, B. Khaira, M. Whiteside, R. Hancock and F. Brinkman (2009) Pseudomonas Genome Database: facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Res* **37**: D483-488.

Witte, C., E. Isidore, S. Tiller, H. Davies and M. Taylor (2001) Functional characterisation of urease accessory protein G (ureG) from potato. *Plant Mol Biol* **45**: 169-179.

Wittmann, T. and C. Waterman-Storer (2001) Cell motility: can Rho GTPases and microtubules point the way? *J Cell Sci* **114**: 3795-3803.

Worlock, A. J. and R. L. Smith (2002) ZntB is a novel Zn^{2+} transporter in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **184**: 4369-4373.

Wu, L. F. and M. A. Mandrand-Berthelot (1986) Genetic and physiological characterization of new *Escherichia coli* mutants impaired in hydrogenase activity. *Biochimie* **68**: 167-179.

Wu, S., R. Fallon and M. Payne (1997) Over-production of stereoselective nitrile hydratase from *Pseudomonas putida* 5B in *Escherichia coli*: activity requires a novel downstream protein. *Appl Microbiol Biotech* **48**: 704-708.

Xie, S. X., Y. Kato, H. Komeda, S. Yoshida and Y. Asano (2003) A gene cluster responsible for alkylaldoxime metabolism coexisting with nitrile hydratase and amidase in *Rhodococcus globerulus* A-4. *Biochemistry* **42**: 12056-12066.

Xiong, A. and R. K. Jayaswal (1998) Molecular characterization of a chromosomal determinant conferring resistance to zinc and cobalt ions in *Staphylococcus aureus*. *J Bacteriol* **180**: 4024-4029.

Xu, B., B. Trawick, G. A. Krudy, R. M. Phillips, L. Zhou and P. R. Rosevear (1994) Probing the metal binding sites of *Escherichia coli* isoleucyl-tRNA synthetase. *Biochemistry* **33**: 398-402.

Xu, Y., L. Feng, P. D. Jeffrey, Y. Shi and F. M. Morel (2008) Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* **452**: 56-61.

Yagi, T. (1970) Solubilization, purification and properties of particulate hydrogenase from *Desulfovibrio vulgaris*. *J Biochem* **68**: 649-57.

Yamada, H. and M. Kobayashi (1996) Nitrile hydratase and its application to industrial production of acrylamide. *Biosci Biotechnol Biochem* **60**: 1391-1400.

Yamakura, F. (1978) A study on the reconstitution of iron-superoxide dismutase from *Pseudomonas ovalis*. *J Biochem* **83**: 849-857.

Yamakura, F., K. Kobayashi, H. Ue and M. Konno (1995) The pH-dependent changes of the enzymatic-activity and spectroscopic properties of iron-substituted manganese superoxide-dismutase: a study on the metal-specific activity of Mn-containing superoxide-dismutase. *Eur J Biochem* **227**: 700-706.

Yamakura, F. and K. Suzuki (1980) Cadmium, chromium, and manganese replacement for iron in iron-superoxide dismutase from *Pseudomonas ovalis*. *J Biochem* **88**: 191-196.

Yanai, I., J. Mellor and C. DeLisi (2002) Identifying functional links between genes using conserved chromosomal proximity. *Trends Genet* **18**: 176-179.

Yang, C., D. Rodionov, X. Li, O. Laikova, M. Gelfand, O. Zagnitko, M. Romine, A. Obraztsova, *et al.* (2006a) Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J Biol Chem* **281**: 29872-29885.

Yang, W., Y. Liu, L. Chen, T. Gao, B. Hu, D. Zhang and F. Liu (2007) Zinc uptake regulator (*zur*) gene involved in zinc homeostasis and virulence of *Xanthomonas oryzae* pv. oryzae in rice. *Curr Microbiol* **54**: 307-314.

Yang, X., T. Becker, N. Walters and D. Pascual (2006b) Deletion of *znuA* virulence factor attenuates *Brucella abortus* and confers protection against wild-type challenge. *Infect Immun* **74**: 3874-3879.

Yatsunyk, L., J. Easton, L. Kim, S. Sugarbaker, B. Bennett, R. Breece, I. Vorontsov, *et al.* (2008) Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from *Escherichia coli*. *J Biol Inorg Chem* **13**: 271-288.

Yee, D. and F. M. M. Morel (1996) *In vivo* substitution of zinc by cobalt in carbonic anhydrase of a marine diatom. *Limnol Oceanogr* **41**: 573-577.

Youn, H. D., E. J. Kim, J. H. Roe, Y. C. Hah and S. O. Kang (1996) A novel nickel-containing superoxide dismutase from *Streptomyces* spp. *Biochem J* **318**: 889-896.

Yuan, Z. C., E. Haudecoeur, D. Faure, K. F. Kerr and E. W. Nester (2008) Comparative transcriptome analysis of *Agrobacterium tumefaciens* in response to plant signal salicylic acid, indole-3-acetic acid and gamma-amino butyric acid reveals signalling cross-talk and Agrobacterium--plant co-evolution. *Cell Microbiol* **10**: 2339-2354.

Zambelli, B., M. Stola, F. Musiani, K. De Vriendt, B. Samyn, B. Devreese, J. Van Beeumen, *et al.* (2005) UreG, a chaperone in the urease assembly process, is an intrinsically unstructured GTPase that specifically binds Zn^{2+}. *J Biol Chem* **280**: 4684-4695.

Zhang, B., O. Georgiev, M. Hagmann, C. Günes, M. Cramer, P. Faller, M. Vasák, *et al.* (2003a) Activity of metal-responsive transcription factor 1 by toxic heavy metals and H_2O_2 in vitro is modulated by metallothionein. *Mol Cell Biol* **23**: 8471-8485.

Zhang, C. M., T. Christian, K. J. Newberry, J. J. Perona and Y. M. Hou (2003b) Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase. *J Mol Biol* **327**: 911-917.

Zhang, J. W., G. Butland, J. F. Greenblatt, A. Emili and D. B. Zamble (2005) A role for SlyD in the *Escherichia coli* hydrogenase biosynthetic pathway. *J Biol Chem* **280**: 4360-4366.

Zhang, Y. and V. N. Gladyshev (2009) Comparative genomics of trace elements: emerging dynamic view of trace element utilization and function. *Chem Rev* **109**: 4828-4861.

Zhang, Y. and V. N. Gladyshev (2010) General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. *J Biol Chem* **285**: 3393-3405.

Zhang, Y., H. Marrakchi and C. Rock (2002) The FabR (YijC) transcription factor regulates unsaturated fatty acid biosynthesis in *Escherichia coli*. *J BiolChem* **277**: 15558-15565.

Zhang, Y., D. Rodionov, M. Gelfand and V. Gladyshev (2009) Comparative genomic analyses of nickel, cobalt and vitamin B12 utilization. *Bmc Genomics* **10**:78

Zhao, H. and D. Eide (1996) The yeast ZRT1 gene encodes the zinc transporter protein of a high-affinity uptake system induced by zinc limitation. *Proc Natl Acad Sci U S A* **93**: 2454-2458.

Zhao, H. and D. J. Eide (1997) Zap1p, a metalloregulatory protein involved in zinc-responsive transcriptional regulation in *Saccharomyces cerevisiae*. *Mol Cell Biol* **17**: 5044-5052.

Zheng, H., M. Chruszcz, P. Lasota, L. Lebioda and W. Minor (2008) Data mining of metal ion environments present in protein structures. *J Inorg Biochem* **102**: 1765-1776.

Zheng, Y., R. Roberts, S. Kasif and C. Guan (2005) Characterization of two new aminopeptidases in *Escherichia coli*. *J Bacteriol* **187**: 3671-3677.

Zheng, Y., R. J. Roberts and S. Kasif (2004) Segmentally variable genes: a new perspective on adaptation. *PLoS Biol* **2**: E81.

Zhou, Y. and D. Jin (1998) The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proc Natl Acad Sci U S A* **95**: 2908-2913.

Zhou, Z., Y. Hashimoto, K. Shiraki and M. Kobayashi (2008) Discovery of posttranslational maturation by self-subunit swapping. *Proc Natl Acad Sci U S A* **105**: 14849-14854.

Zhou, X. D. and P. D. Sadowski (1994) Cleavage-dependent ligation by the FLP recombinase. *Biochemistry* **33**: 12746-12751.

Zuckerkandl, E. and L. Pauling (1965) Molecules as documents of evolutionary history. *J Theor Biol* **8**: 357-366.

BIOGRAPHICAL SKETCH

Crysten Elizabeth Blaby-Haas was born October of 1983 in Gainesville, Florida. She received a Bachelor of Science in microbiology and cell science from the University of Florida in May of 2006. She worked as an undergraduate research assistant in the laboratory of Prof. James F. Preston on developing *Paenibacillus* JDR-2 and *Bacillus subtilis* for the conversion of plant biomass to fermentable substrates that can be used in the production of alternative fuels. In January of 2007, she joined the laboratory of Dr. Valérie de Crécy-Lagard, whose expertise she drew upon to employ comparative genomics and construct hypothesis for testing gene function using genetics. She has presented work at several national and international conferences and received a fellowship to attend the 59th Meeting of Nobel Laureates in Germany. In 2010, she married Dr. Ian Blaby, with whom she will continue research in the laboratory of Prof. Sabeeha Merchant at the University of California, Los Angeles.