UNIVERSITY OF CALIFORNIA, SAN DIEGO

Bioinformatic Characterization of the Copper and Heavy Metal Families of P-type ATPases

A Thesis submitted in partial satisfaction of the requirement for the degree

Master of Science

in

Biology

by

Danielle Elaine Harake

Committee in charge:

Professor Milton Saier, Jr., Chair
Professor Stephen Baird
Professor Christopher Wills

2007

The Thesis of Danielle Elaine Harake is approved:

_____

_____

_____

Chair

University of California, San Diego

2007

DEDICATION

In recognition of their constant support and inspiration, this thesis is dedicated to my family and to all the members of the Saier lab, in particular Professor Milton Saier, Jr.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures ix

List of Tables xi

Acknowledgements xii

Abstract xiii

Introduction 1

Chapter 1: The 3.A.3. Superfamily: P-type ATPases 12

1.1 Phylogenetic Analysis 14

1.2 Families of Functionally Characterized P-type ATPases 17

1.3 Families of Functionally Uncharacterized P-type ATPases 17

1.4 Hydropathy Analysis 18

Chapter 2: The Copper P-type ATPases 20

2.1 Phylogenetic Analysis 22

2.2 Analyses of Sequences of Unusual Length 25

2.3 Hydropathy Analysis 32

2.4 HMMTop Analysis 33

2.5 SOPMA Analysis 35

2.6 EMBOSS Pepwheel Analysis 37

2.7 16S rRNA Analysis 38

Chapter 3: The Heavy Metal P-type ATPases 41

3.1 Phylogenetic Analysis 42
3.2 Analyses of Sequences of Unusual Length 46
3.3 Hydropathy Analysis 51
3.4 HMMTop Analysis 52
3.5 SOPMA Analysis 54
3.6 EMBOSS Pepwheel Analysis 56
3.7 16S rRNA Analysis 57
Chapter 4: Conserved Motifs in Copper P-type ATPases 59
4.1 Motif 1 61
4.2 Motif 2 62
4.3 Motif 3 62
4.4 Motif 4 64
4.5 Motif 5 65
4.6 Motif 6 65
4.7 Motif 7 66
4.8 Motif 8 67
4.9 Motif 9 67
Chapter 5: Conserved Motifs in Heavy Metal P-type ATPases 68
5.1 Motif 1 69
5.2 Motif 2 70
5.3 Motif 3 71
5.4 Motif 4 71

5.5 Motif 5 72

5.6 Motif 6 72

5.7 Motif 7 73

5.8 Motif 8 73

5.9 Motif 9 74

Chapter 6: Homology Analyses of Copper and Heavy Metal P-type ATPases 75

6.1 Homology Analyses of Copper P-type ATPases 76

6.2 Homology Analyses of Heavy-Metal P-type ATPases 97

Chapter 7: Analyses of P-type ATPases by Organismal Domain 125

7.1 Archaea 127

7.2 Bacteria 129

7.3 Eukaryota 129

Discussion 131

Appendix 134

Figure 1A. 135

Figure 1.B 136

Figure 2.A 137

Figure 3.A.1 138

Figure 3.B.1 139

Figure 3.C.1 140

Figure 3.D 141

Figure 4. 142

Figure 5.A 143

Figure 6.A.1 144

Figure 6.B.1 145

Figure 6.C.1 146

Figure 6.D 147

Figure 7. 148

Figure 8. 149

Figure 9. 150

Table 1. 151

Table 2. 155

Table 3. 170

Table 4. 171

Table 5. 178

Table 6. 195

Table 7. 198

Table 8. 210

Table 9. 212

References 214

LIST OF FIGURES*

Figure 1.A: The phylogenetic tree of the 3.A.3 Superfamily. The clustering patterns of the families and Sub-Superfamilies (SSFs) are indicated within the figure. …….. 135

Figure 1.B.1: The AveHas plots representing the sequences within the Sub-Superfamilies created from the proteins of the 3.A.3 Superfamily. …….. 136

Figure 2.A: The phylogenetic tree of the 385 protein sequences representing the Copper P-type ATPase Family (dendogram format). The locations of the twenty clusters representing these sequences are indicated in the figure. …….. 137

Figure 2.B: The Phylogenetic Tree of the proteins of the Copper P-type ATPase Family (cladogram format). The locations of the twenty clusters found within these sequences are indicated in the figure. …….. *

Figure 3.A.1 AveHas analysis for cluster 1 of the Copper P-type ATPase Family. …….. 138

Figure 3.B.1 HMMTop analysis for cluster 1 of the Copper P-type ATPase Family. …….. 139

Figure 3C.1 SOMPA analysis for cluster 1 of the Copper P-type ATPase Family. …….. 140

Figure 3.D: The EMBOSS Pepwheel analyses of a segment from the protein sequence Ssu1, from cluster 15. This cluster was determined to have the best amphipathic peak in the region between TMS2 and TMS3 of all of the clusters representing the Copper P-type ATPases. …….. 141

Figure 4: The 16S rRNA phylogenetic tree for the prokaryotic genera present among the sequences representing the Copper P-type ATPase Family. …….. 142

Figure 5.A: The phylogenetic tree of the sequences representing the proteins of the Heavy Metal P-type ATPase Family (dendogram format). The locations of the 17 clusters found within these sequences are indicated in the figure. …….. 143

Figure 5.B: The phylogenetic tree of the sequences representing the proteins of the Heavy Metal P-type ATPase Family (cladogram format). The locations of the seventeen clusters found within these sequences are indicated in the figure. …….. 144

* For the purpose of clarity when examining the figures corresponding with this thesis, all figures are available in their entirety at: http://www.biology.ucsd.edu/~msaier/supmat/P-type_ATPase. Condensed views of select figures are also provided in the appendix.

Figure 6A.1: AveHas analysis for cluster 1 of the Heavy Metal P-type ATPase Family. …….. 145

Figure 6.B.1: HMMTop analysis for cluster 1 of the Heavy Metal P-type ATPase Family. …….. 146

Figure 6C.1: SOPMA analysis for cluster 1 of the Heavy Metal P-type ATPase Family. …….. 147

Figure 6D: The EMBOSS Pepwheel analyses of a segment from the protein sequence Ssp3, from cluster 10. This cluster was determined to have the best amphipathic peak in the region between TMS2 and TMS3 of all of the clusters representing the Heavy Metal P-type ATPases. …….. 148

Figure 7: The 16S rRNA phylogenetic tree containing one sequence for every genus present in the sequences representing the Heavy Metal P-type ATPase Family. …….. 149

Figure 8: The AveHas plot constructed from one sequence from each of the twenty clusters representing the sequences of the Copper P-type ATPase Family. This plot was used to identify the region amongst these sequences that contained the strongest amphipathic peak, which was then investigated using the EMBOSS Pepwheel program. …….. 150

Figure 9: The AveHas plot constructed from one sequence from each of the seventeen clusters representing the sequences of the Heavy Metal P-type ATPase Family. This plot was used to identify the region amongst these sequences that contained the strongest amphipathic peak, which was then investigated using the EMBOSS Pepwheel program. …….. 151

Figure S1: The multiple alignment of the sequences representing the 3.A.3 Superfamily in TCDB …….. *

Figure S2: The multiple alignment of the sequences representing the Copper P-type ATPase Family. …….. *

Figure S3: The multiple alignment for the sequences representing the Heavy Metal P-type ATPase Family. …….. *

LIST OF TABLES[†]

Table 1: The proteins sequences representing the families within the 3.A.3. Superfamily in TCDB. …….. 151

Table 2: The protein sequences representing the Copper P-type ATPase Family (385 sequences). …….. 155

Table 3: Motif analyses for the twenty clusters representing the proteins of the Copper P-type ATPase Family. …….. 170

Table 4: Information regarding the protein sequences used to generate the 16S rRNA phylogenetic tree representing the genura present in Copper P-type ATPase Family. …….. 171

Table 5: The protein sequences representing the Heavy Metal P-type ATPase Family (311 sequences). …….. 178

Table 6: Motif analyses for the seventeen clusters representing the Heavy Metal P-type ATPase Family. …….. 195

Table 7: The organisms whose 16S rRNAs were used to construct a phylogenetic tree representing each genus found among the 311 protein sequences representing the Heavy Metal P-type ATPase Family …….. 198

Table 8: An overview of the sequence characteristics of the clusters representing Copper P-type ATPase Family …….. 210

Table 9: An overview of the sequence characteristics of the clusters representing the Heavy Metal P-type ATPase Family …….. 212

[†] Condensed views of these tables are provided in the appendix, and are also available at: http://www.biology.ucsd.edu/~msaier/supmat/P-type_ATPase.

ACKNOWLEDGEMENTS

I would like to take this opportunity to acknowledge and sincerely thank Professor Milton Saier, Jr. for his guidance and support in the completion of my research project. I am greatly indebted to him for his mentorship and for allowing me join him in his research. I am particularly grateful to him for renewing in me a passion for learning and an eagerness to follow the path of discovery. Additionally, I would like to thank all of the members of the Saier Lab, whose help and encouragement were invaluable and whose friendship I shall always hold dear.

ABSTRACT OF THE THESIS

Bioinformatic Characterization of the Copper and Heavy Metal Families of P-type ATPases

by

Danielle Elaine Harake

Master of Science in Biology

University of California, San Diego, 2007

Professor Milton Saier, Jr., Chair

P-type ATPases are classified as the 3.A.3 Superfamily in the Transport Classification Database (TCDB) and are an important group of protein pumps involved in the translocation of specific ions or phospholipids across biological membranes. The 3.A.3. Superfamily presently consists of thirty-four families of P-type ATPases, only ten of which are functionally characterized. Methodical analyses of sequences from each of these families are required to elucidate the mechanism by which they transport their substrates and to identify their distinguishing functional characteristics. A phylogenetic tree was constructed to confirm that sequences representing the 3.A.3. Superfamily in TCDB reflected their family assignments, and that the sequence characteristics of each

"sub-superfamily" cluster within this tree were examined. The Copper and Heavy Metal P-type ATPase families were subsequently selected for more in-depth analyses. Protein sequences representing these two families were collected using NCBI psi-BLAST, multiply aligned, and analyzed using protein and 16S rRNA phylogenetic trees. The sequence similarities within each family, with regard to sequence length and conservation of specific amino acid motifs, were examined and compared. Additionally, the secondary structure patterns, regions of amphipathicity and hydrophobicity, and predicted numbers and locations of transmembrane segments (TMSs) for these sequences were analyzed. Most sequences from both families showed conservation of nine well-described motifs and exhibited a consistent pattern of eight TMSs. Sequence homology analyses revealed that while most sequences clustered with others from their genus or phylogenetic group, some sequences were so divergent from their neighbors that they could indicate instances of horizontal gene transfer.

Introduction

P-type ATPases are a diverse group of protein transporters involved in either the uptake or efflux of ions across a membrane in concert with ATP hydrolysis. These proteins have been identified in numerous species belonging to Archaea, Bacteria and Eukaryota, and are collectively classified as belonging to the 3.A.3 Superfamily in the Transport Classification Database (TCDB). Despite the substrate binding differences observed amongst P-type ATPases and the wide range of organisms in which they are found, they do consistently exhibit some key points of similarity. P-type ATPases show strong conservation of nine well-described sequence motifs (Møller J.V., et al. 1995). Furthermore, all P-type ATPases depend on ATPase activity to engage in their specific ion binding activities, and they all rely on the formation of an intermediate conformation which requires the phosphorylation of a specific aspartate residue (Møller J.V., et al., 1995).

At present, ten families within the 3.A.3 Superfamily have been distinguished based on substrate ion type and, in some instances, by some discernible functional property. An additional twenty-four families have been identified which consist of P-type ATPases that, at this time, are functionally unclassified. Particular attention was paid in this thesis to Families 5 and 6, the Copper and Heavy Metal Families, respectively. In addition, ninety-three full-length sequences listed as representatives of the 3.A.3 Superfamily in the Transport Classification Database (TCDB) were collected and analyzed. The protocol employed for subsequent analyses of these proteins was essentially identical to the protocol previously used in the analyses of

archaeal homologues (Saier, unpublished data). In brief summation, redundant and partial sequences were removed from the initial individual compilation of sequences for both Family 5 and Family 6, respectively, leaving only full-length sequences for analysis. Sequence alignments were conducted, and phylogenetic trees and hydropathy plots were constructed so as to provide a means for identifying and evaluating the conservation of specific motifs and to facilitate the discovery of unique sequence characteristics. The multiple sequence alignments can be found online at http://www.biology.ucsd.edu/~msaier/supmat/P-type_ATPase.

Table 1 provides a summary of the ninety-three protein sequences listed as representative members of the P-type ATPase Superfamily (3.A.3). Sixty-nine of these sequences can be found distributed throughout ten functionally-characterized families within the 3.A.3 Superfamily. The remaining twenty-four sequences each represents a family of P-type ATPases which, as of yet, are functionally uncharacterized. While some of these proteins appear to cluster with functionally characterized proteins, as indicated by their clustering patterns on the phylogenetic tree, preliminary analysis indicates that they are phylogenetically distinct from one another. Collectively, these subfamilies are visually displayed in Figure 1.A where, for the most part, they appear to cluster well based on their functional properties, if not precisely by their family assignments.

Data for the Copper P-type ATPases were collected by conducting a non-redundant protein psi-BLAST (NCBI) search using as query the protein sequences listed on TCDB as representatives of Family 5. Redundant sequences with greater than

ninety percent sequence similarity were removed from each BLAST result using the program CD-Hit (M.R. Yen and M.H. Saier, unpublished program). Redundant sequences were eliminated again using the same program after all BLAST results were combined, yielding 385 distinct protein sequences (Table 2). After retrieving the corresponding sequence data from NCBI a multiple alignment was generated and a phylogenetic tree was constructed and examined (Figures 2.A and 2.B). Additionally, one nucleotide sequence of 16S ribosomal RNA was collected for every genus present among the 385 protein sequences, and a 16S rRNA phylogenetic tree was constructed and analyzed (Figure 4 and Table 4). As the protein sequence phylogenetic tree was too dense to properly assess the clustering properties of the sequences, it was divided into twenty smaller clusters (Figures 2.A and 2.B; Table 2). These clusters were created based on the most distinct branches visible in the phylogenetic tree. The clusters containing more than one sequence were individually multiply aligned, and the resulting data were then used to perform motif analyses on each cluster. Additionally, these data were used to perform a series of analyses using the programs AveHAS or WHAT, HMMTOP, SOPMA and EMBOSS Pepwheel (Figures 3A.1-.20, 3.B.1-.20, 3C.1-.20, 3D).

Nearly identical procedures were performed to collect and analyze the data for Family 6, the Heavy metal P-type ATPases, as were used to collect the data for Family 5. After collecting the initial set of sequences and eliminating redundancies, a total of 311 proteins sequences remained. Accordingly, these sequences were multiply aligned and a phylogenetic tree was generated and evaluated. Likewise, one 16S ribosomal

RNA sequence was collected to represent each genus present amongst the 311 protein sequences. As in Family 5, the phylogenetic tree containing all of the sequences representing Family 6 was too dense to use for detailed data analysis. Consequently, the tree was divided into seventeen clusters based on the most distinct grouping patterns observed in the phylogenetic tree (Figures 5.A-5.B; Table 5). Multiple alignments were generated for the clusters containing more than one protein sequence, and motif analyses were performed. These data were consequently used to carry out motif, AveHAS, WHAT, HMMTOP, SOPMA and EMBOSS Pepwheel analyses (Figures 6A.1-.17, 6B.1-.17, 6C.1-.17, 6D).

Computational Methods

Sequence Collection

The data representing the protein sequences of the Copper P-type ATPase Family (Family 5) were gathered by first individually examining each of the sample proteins listed as representative members of Family 5, which, at the time, was only five proteins. This was accomplished by running each of the full-length sequences listed on TCDB as a representative of Family 5 separately through NCBI's non-redundant psi-BLAST program (http://www.ncbi.nlm.nih.gov/BLAST/). All sequences with e-values above threshold were collected and compiled into one large group, which was then greatly reduced in size by eliminating sequences that were

redundant, fragmented or too short. Elimination of redundancies was accomplished using the program CD-Hit, which removed sequences with greater than ninety-percent similarity. After the redundancies were eliminated within each of these groups, all of the sequences were combined, and, again, redundant sequences were eliminated using the same criteria with CD-Hit, thereby reducing the total to 385 sequences. The corresponding TinySeq XML format (NCBI) of these proteins was obtained and modified using the script MakeTable4 to generate a format of the sequences compatible for running the sequences through a multiple alignment. The MakeTable4 program also generated a file with 16S rRNA nucleotide sequences representing one example of nearly every prokaryotic genus present among the protein sequences analyzed (M.R. Yen and M.H. Saier, unpublished program). 16S rRNA nucleotide sequences for the remaining genera were manually retrieved from NCBI's Core Nucleotide database and combined with those collected by the Make Table4 program. Using the multiple alignment program ClustalX and the phylogenetic tree-generating program TreeView PPC, these nucleotide sequences were aligned and a 16S rRNA phylogenetic tree was created.

Virtually identical methods were employed to gather the proteins representing Family 6, the Heavy Metal P-type ATPases, resulting in the collection of 311 sequences. Likewise, the Make Table4 program was used to collect a sample 16S rRNA nucleotide sequence for nearly every genus of prokaryotic species found in the protein sequences representing Family 6, as well as to modify the TinySeq XML format for the sequences from Family 6 so as to prepare them for running a multiple alignment. Representative 16S rRNA nucleotide sequences for the genera that were

not found in the Make Table4 program were manually retrieved from NCBI's Core Nucleotide database and added to the previously obtained sequences from Make Table4. These sequences were then aligned using ClustalX and a 16S rRNA phylogenetic tree was generated using TreeView PPC.

By contrast to the procedures used to obtain data for Family 5 and Family 6, the collection of the sequences for the analysis of the 3.A.3 Superfamily was less complicated. The data pertaining to each of the 93 protein sequences listed on the Transport Classification Database (TCDB) as representatives of the 3.A.3 Superfamily were compiled and converted to a TinySeq XML format compatible for running a multiple alignment, as detailed above. While most of the relevant sequence data were obtained through NCBI, two of the proteins were not yet listed on NCBI at the time of this report. The sequence for one of these proteins, obtained from *Cyanidioschyzon merolae* 10D and designated as Cme1, was obtained from TransportDB (http://membranetransport.org). The other sequence, obtained form *Thalissiosira thermophila* SB210 and designated as Tpe1 (Thever M.D.; 2007). Both Cme1 and Tpe1 are, at the present time, functionally uncharacterized P-type ATPases and belong to the subfamilies 3.A.3.12.1 (Family 12) and 3.A.3.21.1 (Family 21), respectively.

16S rRNA Analysis

The 16S ribosomal RNA sequences used for analysis of the prokaryotic genera present in Families 5 and 6 were obtained from the NCBI database. The 18S ribosomal RNAs were not collected for the eukaryotic genera, as eukaryotes represented only a

small fraction of the total sequences from either Family 5 or 6, and, for the most part, clustered by themselves. Prokaryotic organisms with unclassified genera were omitted from the 16S rRNA analyses for these two families. A total of 138 different genera were identified in Family 5, and 136 different genera were identified in Family 6. When possible, 16S rRNA corresponding to the complete genomes of each genus were collected for analysis. However, the complete 16S rRNA could not be located for all of the prokaryotic genera. In these instances, the 16S rRNA of the partial genome of an organism belonging to the desired genus was used instead, given that it was similar in length to the 16S rRNA obtained from the fully sequenced genomes of other organisms collected. The sequences were aligned using the multiple alignment program ClustalX (see below), and phylogenetic trees for both Family 5 and Family 6 were constructed using TreeView PPC (see below). These phylogenetic trees were subsequently analyzed to evaluate the clustering patterns of each organism.

Multiple Alignments and Phylogenetic Trees

The ClustalX program was used to perform multiple alignments of the 16S rRNA sequences, as well as the protein sequences representing Families 5 and 6. Additionally, it was used to perform a multiple alignment of the protein sequences representing the 3.A.3 Superfamily, as well as multiple alignments on the six sub-superfamilies (SSFs) created based on the clustering patterns of sequences observed in the 3.A.3 Superfamily. After completing these alignments, corresponding

phylogenetic trees were generated. The vast number of sequences present in Families 5 and 6 obscured detailed analysis of their respective phylogenetic trees. Thus, to facilitate further analysis, the sequences in each family were divided into smaller clusters based on the most distinct grouping patterns found in their respective phylogenetic trees (Figures 2.A-.B and 5.A-5.B). This resulted in the production of twenty groups within Family 5 and seventeen groups within Family 6. Accordingly, multiple alignments were run for each group consisting of more than one protein sequence. The completed alignment data were subsequently used to analyze the conservation of motifs and the similarities, amphipathicities, hydrophobicities, and topological patterns of the sequences in each group using the programs AveHAS, WHAT, HMMTOP, SOPMA and EMBOSS Pepwheel (see below).

Hydropathy Analysis

The AveHAS program was run on almost each of the twenty groups representing the proteins of the Copper P-type ATPase Family and the seventeen groups representing the proteins of the Heavy Metal P-type ATPase Family. However, group # 7 in Family 5 and groups # 11 and 17 in Family 6 only contained one sequence each. Thus, the WHAT program was run in lieu of AveHAS for these three groups, based on program specifications (http://tcdb.ucsd.edu/avehas.html and (http://tcdb.ucsd.edu/what.html). These two programs are designed to display regions of hydrophobicity and amphipathicity, as well as predict the numbers and locations of transmembrane sequences (TMS) present in each sequence or group of sequences it

analyzes. The AveHAS program also provides an indication of the average similarity of the proteins it analyzes.

In addition to analyzing the protein sequences in many of the clusters in both Families 5 and 6, the AveHAS program was also run on twenty protein sequences from the Copper P-type ATPases and seventeen protein sequences from the Heavy Metal P-type ATPases, one per cluster. This was done to provide additional means of comparing these two families, and, more specifically, to identify the most amphipathic region in each of these two families to examine using the EMBOSS Pepwheel program.

Motif Analysis

One of the most distinguishing factors of the P-type ATPases is their display of nine well conserved motifs, which are examined in depth for Families 5 and 6 in Chapters 4 and 5, respectively. Thorough examination of these motifs was conducted via manual identification in the multiple alignments of each of the twenty Copper and seventeen Heavy metal groups. The alignment locations, levels of conservations, and appearance or absence of residue substitutions in each of these motifs were documented in Microsoft Excel spreadsheets, as seen in Tables 3 and 6. The motifs from the groups belonging to Families 5 and 6 were further scrutinized using the SOMPA program to analyze trends in motif topology and the HMMTop program to identify the location of each motif in relation to the protein sequence's TMSs.

Homology Analysis

Each of the twenty clusters of protein sequences representing the Copper P-type ATPase and the seventeen clusters of protein sequences representing the Heavy Metal P-type ATPases were examined to determine the phylogenetic relationships of the individual sequences within each cluster. Several sub-clusters were discovered within each cluster. The branching distances and genera of the protein sequences within these sub-clusters were compared using the 16S rRNA and protein phylogenetic trees corresponding with either the Copper or Heavy Metal families (Figures 2.A-2.B, 4, 5.A-5.B, and 7). Sequences that clustered closely together in their protein phylogenetic tree were either all from the same genus, or were from genera that were found within the same cluster in their 16S rRNA tree, were predicted to be orthologous to one another. Although the sequences clustering closely together most frequently were from the same or a closely related genus, others were found in adjacent or very distant clusters in their 16S rRNA tree. These sequences were not likely to be orthologous to one another and, depending on their branching differences in both the 16S rRNA and protein phylogenetic trees, some appeared to be possible examples of horizontal gene transfer.

Phylogenetic Domain Analysis

It is not uncommon for P-type ATPases to exhibit greater sequence similarity with other P-type ATPases with the same substrate than they might with a P-type ATPase from the same organism, but with a different substrate. However, most P-type ATPases have certain conserved functions, like phosphate recycling, and conserved regions like a phosphorylation site and an ATP-binding site. Additionally, the functions and characteristics of P-type ATPases are influenced by the phylogenetic kingdom to which they belong (Møller J.V., et al., 1995). Consequently, the sequences of the Copper and Heavy Metal P-type ATPases (Families 5 and 6, respectively) were further analyzed by examining the characteristics of sequences based on their phylogenetic kingdoms. The Bacterial domain was, by far, the most represented domain amongst the sequences in both families. The number of protein sequences belonging to Archaea and Eukaryota in these two families was comparatively small, with twenty archaeal sequences and fifteen eukaryotic sequences found in Family 5 and forty-seven archaeal sequences and eight eukaryotic sequences found in Family 6.

Chapter 1: 3.A.3: The P-Type ATPase Superfamily

3.A.3: The P-Type ATPase Superfamily

The P-type ATPase Superfamily, which is classified as the 3.A.3 Superfamily (TCDB), is a collection of proteins associated with the transport of ions across a membrane in concert with the hydrolysis of ATP. It is comprised of both prokaryotic and eukaryotic proteins, and is classified into several distinct families. In some cases, these families are further broken down into subfamilies based on distinguishable genomic differentiations within a given family. At present, ten families within the 3.A.3 Superfamily have been classified based on substrate (ion) type. An additional twenty-four families have been created from, as of yet, functionally unclassified P-type ATPases.

Table 1 provides a summary of the data corresponding to the 93 full-length protein sequences listed on TCDB as representatives of the 3.A.3 Superfamily. Sixty-nine of these sequences are members of Families 1-10 within the 3.A.3 Superfamily, which each contains distinct, functionally characterized P-type ATPases. The remaining twenty-four proteins represent Families 11-34 which, as of yet, are functionally uncharacterized P-type ATPases. Preliminary examinations suggest that, while some of these proteins appear to cluster near proteins that belong to various functionally classified families, these twenty-four families are phylogenetically different from one another. Collectively, these proteins are visually displayed in Figure 1.A where, for the most part, they appear to cluster well based on their functional properties, if not precisely by their family assignments.

1.1: Phylogenetic Analysis

The phylogenetic tree for the 3.A.3 Superfamily, the P-type ATPases, is depicted in Figure 1.A. As expected, the protein sequences with known substrates (ions) primarily aligned in accordance with their family groupings. The clustering patterns seen in this phylogenetic tree generally support previous research, which indicated that greater sequence similarity generally exists between P-type ATPases with the same substrate, even from different organisms, than between those from the same organism, but different substrate type (Axelsen and Palmgren; 1998). However, there were some exceptions to these clustering patterns observed. While the protein sequences that represented Families 1 and 9 (3.A.3.1 and 3.A.3.9, respectively) clustered closely without interruption from proteins belonging to other families, they themselves broke up the continuity of the sequences representing Family 2 (3.A.3.2). Despite the difference in ion substrate types, the division of Family 2 by Families 1 and 9 was somewhat expected, as Na^+/K^+, H^+/K^+, fungal Na^+, and Ca^{2+} P-type ATPases are collectively classified as type II ATPases (Møller J.V., et al., 1996; Axelsen and Palmgren, 1998). Unlike type I ATPases, which include the Copper and Heavy Metal Families and have, on average, eight TMSs, the type II ATPases generally consist of ten TMSs. Type II ATPases have been further subdivided into type IIA ATPases, type IIB ATPases, type IIC ATPases and type IID ATPases. The first of these two subdivisions of type II ATPases are involved in Ca^{2+} transport. The third subdivision is comprised of both Na^+/K^+ and H^+/K^+ transporters, whereas the last subdivision is made up of both Ca^{2+} and Na^+ transporters. Although differences in ion

substrate types are found even within the different categories of type II ATPases, previous research supports the existence of such relationships and suggests that their ancient origins may have necessitated the co-evolution of these transporters (Axelsen and Palmgren, 1998; Benito, et al., 2000).

Family 1, which was represented by four protein sequences in Figure 1.A, was comprised of P-type ATPases transporting Na^+, K^+, H^+, or NH_4^+ ATPases. Similarly, Family 9, which was also represented by four protein sequences in Figure 1.A, consisted of P-type ATPases primarily associated with Na^+ and/or K^+ efflux. By contrast, Family 2 was one of two families within TCDB associated with Ca^{2+} ion pumps, the other being Family 10. Family 2 was represented by eighteen sequences in Figure 1.A, which were further divided into Family 2A, 2B and 2C. Collectively, these three components of Family 2 appeared to be composed of Ca P-type ATPases that were located in one of the following regions: the trans-Golgi network, the plasma membrane, or endomembranes, which include the Endoplasmic Reticulum. Interestingly, it did not appear that the three divisions in Family 2 seen in Figure 1.A exclusively contained protein sequences from one region versus another.

In Figure 1.A, the only sequence representing Family 4 (3.A.3.4), which is involved in $Mg^{2+/}Ni^{2+}$ uptake, was depicted as clustering within Family 3 (3.A.3.3). These two families are classified as type IIIB and type IIIA ATPases, respectively (Axelsen and Palmgren, 1998). While Family 3 is primarily associated with the translocation of protons, other studies have indicated that these two families are similar in function and location, so it was not surprising to find them clustered

together (Axelsen and Palmgren, 1998; Mukherjee, et al., 2000). Additionally, Figure 1.A depicted the proteins of Family 3 as belonging two distinct subgroups, with cluster 3 (3.A.3.3.3) belonging to its own subgroup, and the remaining clusters (1, 2 and 4-7) falling into a second subgroup. Family 4 was shown in Figure 1.A as forming its own subgroup between the two groups in Family 3.

The functionally uncharacterized proteins belonging to Families 12-22 clustered together in Figure 1.A, with Families 11-18 forming one subgroup, and Families 10, 12, 20 and 22 forming distinct subgroups of their own. These groups associated closely with Family 8, whose proteins in Figure 1.A fall into one distinct subgroup and are associated with phospholipid translocation, and contain Family 10, whose proteins are associated with the endoplasmic reticulum Ca^{2+} pumps. The remaining families of uncharacterized P-type ATPases can be seen in Figure 1 as follows: Family 30 appeared to cluster with Families 1, 2 and 9, but formed its own distinct subgroup. Families 23 and 24 appeared to form their own subgroup and cluster between Families 3 and one of the subfamilies of Family 2. The sequences representing Family 8 are classified as type IV ATPases, whereas the neighboring eukaryotic sequences, which are functionally uncharacterized, have been categorized as type V ATPases (Axelsen and Palmgren, 1998).

Families 26, 31, and 32 clustered alongside Family 7, but each formed their own subgroup. All of the proteins representing Family 6 in Figure 1.A appeared to form two subgroups, one containing cluster 7 (3.A.3.6.7) and the other containing all of their other clusters (1-6 and 8-10). Families 25, 33 and 34 formed a distinct

subgroup between Families 5 and 6, and Families 27 and 29 formed another subgroup with clusters 9 and 13 from Family 5 (3.A.3.5.9 and 3.A.3.5.13), which was located in Figure 1.A between the other sequences belonging to Families 5 and 6. The remaining sequences from Family 5 were all found together in a separate subgroup, and Family 28 was found in its own distinct subgroup clustering closely to Family 5, but on the side farthest from Family 6.

1.2: Families of Functionally Characterized P-type ATPase Proteins

To date, the substrate ion and, in some cases, some of the functional characteristics of ten families within the 3.A.3 Superfamily have been identified (3.A.31- 3.A.310). These families were collectively represented by a total of sixty-nine protein sequences in Figure 1.A, eighteen of which represented the Copper Family (Family 5) and ten of which represented the Heavy Metal Families (Family 6).

1.3: Families of Functionally Uncharacterized P-type ATPase Proteins

Twenty-four families within the 3.A.3 Superfamily are, at the writing of this report, functionally unclassified. As previously described, some of these families clustered next to or within families of P-type ATPases whose functionality has been characterized. While this suggests that they may share some sequence and functional similarity with those proteins, it is premature to conclude that they are, in fact,

functionally or phylogenetically similar enough to be categorized as part of the same family within the 3.A.3 Superfamily.

1.4: Hydropathy Analysis

The ninety-three protein sequences representing the 3.A.3 Superfamily were subdivided into seven groupings classified as sub-superfamilies, or SSFs, based on the major clusters observed in the corresponding phylogenetic tree (Figure 1.A). The proteins in each of these sub-superfamilies were then analyzed using the AveHAS program (http://tcdb.ucsd.edu/avehas.html). This program provided predictions of the number of transmembrane segments (TMSs) in each sub-superfamily, as well as visually indicated the levels of similarity and the points of hydrophobicity, and amphipathicity shared by the protein sequences in each sub-superfamily.

SSF 1 and SSF 7 were examined both together and separately using the AveHAS program. These two sub-superfamilies are thought to be Type I P-type ATPases, and both display very similar sets of eight TMSs (Figure 1.B.1-1.B.7). SSF 1 contains the characterized families of the Copper and Heavy metal P-type ATPases (Families 5 and 6, respectively). It also contains the uncharacterized Families 27 and 28, which cluster within Family 5, but each form their own distinct subgroup, and Families 25, 29, 33 and 34, which cluster together between Families 5 and 6. SSF 7 contains the characterized Family 7, the Kdp P-type ATPases, and the uncharacterized Families 26, 31 and 32.

SSF 2 is thought to contain Type II P-type ATPases, and consist of ten distinct TMSs. It is comprised of the characterized Families 1, 2 and 9, which are known to cluster together, and the uncharacterized Family 30. Both Families 1 and 9 are involved in Na^+ and/or K^+ translocation. Family 2 is one of the two families within the 3.A.3 Superfamily associated with Ca^{2+} translocation, but it is also believed to be closely linked with Na^+ translocation as well due to similarities in evolutionary origin (Benito, et. al, 2000). Like SSF 2, the protein sequences falling within SSF 3,4 have ten TMSs. SSF 3,4 contains only Families 3 and 4, the Proton and Mg^{2+}/Ni^{2+} P-type ATPases, which are known to cluster together (Mukherjee, et al., 2000). Additionally, SSF 2 clusters closely to SSF 6 (Figure 1.A), which also contains ten TMSs. SSF 6 does not contain any of the currently functionally characterized P-type ATPases. Instead, it is comprised entirely of Families 23 and 24, which each form their own distinct subgroup within SSF 6. Lastly, SSF 5, which also contains 10 TMSs, consists of the functionally characterized Families 8 and 10, which are involved in phospholipid and calcium translocation, respectively, and the uncharacterized Families 11-22. All of the members of Family 8 cluster together, forming a single subgroup within SSF 5. Likewise, Families 13-16 appear to collectively form another subgroup within SSF 5. The remaining Families within SSF 5, Families 11, 12 and 17-20, each form their own distinct subgroup.

Chapter 2: The Copper P-Type ATPases

The Copper P-Type ATPases

Copper P-type ATPases, which are classified as Family 5 in TCDB, are found in both prokaryotic and eukaryotic organisms. Many Family 5 proteins exhibit functional and phylogenetic similarities to proteins belonging to the Heavy metal P-type ATPase Family, including their tendency to have one or more heavy metal-associated domains (HMAs). Despite these similarities, though, phylogenetic analyses indicate that Copper and Heavy metal P-type ATPases are distinct from one another. Amongst the most studied proteins belonging to Family 5 include human proteins ATP7A and ATP7B. Although in-depth analyses of these proteins was not within the scope of this study, it is interesting to note that defects in these proteins are associated with Menkes and Wilson's diseases, respectively.

385 Copper P-type ATPase sequences were obtained as a result of performing a series of non-redundant psi-BLAST searches (NCBI) and subsequent eliminations of redundant sequences. The corresponding sequence data from NCBI was used to perform a multiple alignment and to generate a phylogenetic tree. Additionally, these data were used to collect and analyze the 16S rRNAs of the prokaryotic genera present (see below). As the initial phylogenetic tree was too dense to properly assess the clustering properties of the sequences present, it was subsequently divided into twenty smaller clusters. These groups were based on the most distinct branches visible in the phylogenetic tree containing all 385 sequences, and underwent motif, AveHAS, HMMTOP, SOPMA and EMBOSS Pepwheel analyses.

2.1: Phylogenetic Analysis

The original 385 protein sequences representing the Copper P-type ATPase Family were divided into twenty clusters based on the distinct branching patterns observed in the protein phylogenetic tree (Figures 2.A-2.B; Table 2). Cluster 1 was comprised of four protein sequences with an average amino acid length of 771 ± 52 residues. The two bacterial sequences present belonged to the phylogenetic groups Thermotogae and ε-proteobacteria, while the two archaeal sequences present both belonged to Crenarchaeota. Cluster 2 contained twenty-eight protein sequences, all from Actinobacteria. The average length of these sequences was 769 ± 53 residues.

Cluster 3 contained twenty-six protein sequences from several different bacterial phylogenetic groups: Acidobacteria, Δ-proteobacteria, β-proteobacteria, γ-proteobacteria, Chlorobi, and one unclassified proteobacterium. The average length of these sequences was 806 ± 16 residues. Cluster 4 consisted of twenty protein sequences, averaging 706 ± 40 amino acid residues in length. Six of these proteins belong to the archaeal phylogenetic group, Euryarchaeota. The remaining sequences belonged to the bacterial phylogenetic groups Firmicutes, Aquificae, β-proteobacteria, **Δ-**proteobacteria, Deiococci, and Chloroflexi.

Cluster 5 contained five bacterial protein sequences with an average length of 827 ± 18 amino acid residues. These sequences belonged to the phylogenetic groups Deinococci, Actinobacteria, and α-proteobacteria. Cluster 6 was comprised of forty-seven bacterial sequences from the phylogenetic groups Planctomycetes, α-

proteobacteria, β-proteobacteria, and γ-proteobacteria. These sequences had an average length of 788 ± 52 amino acid residues.

Cluster 7 was the only group amongst the twenty groups that contained only one sequence. This sequence was 807 amino acid residues long and was from an unclassified proteobacteria. Cluster 8 consisted of twenty-seven bacterial protein sequences from Cyanobacteria, Deinococci, α-proteobacteria, β-proteobacteria, and γ-proteobacteria. These sequences averaged 816 ± 52 amino acids residues in length.

Cluster 9 contained twenty-four protein sequences with an average length of 849 ± 83 amino acid residues. Unlike most of the groups of its size, all of the sequences from group 9 are from a single bacterial phylogenetic group, γ-proteobacteria. Group 10 was one of the larger groups, containing thirty-nine protein sequences which averaged 831 ± 84 amino acid residues in length. Its nine archaeal sequences were all from the phylogenetic group Euryarchaeota, whereas its thirty bacterial sequences were from a range of phylogenetic groups, which included Chloroflexi, Cyanobacteria, Chlorobi, Firmicutes, Bacteriodetes, Actinobacteria, Acidobacteria, β-proteobacteria, and **Δ-**proteobacteria.

Cluster 11 consisted of nine protein sequences which had an average length of 831 ± 71 amino acid residues. Its three archaeal sequences were all from the phylogenetic group Euryarchaeota, and its remaining sequences were from the bacterial phylogenetic groups Actinobacteria, Chlorobi, Firmicutes, and **Δ-**proteobacteria. All seven of the protein sequences belonging to cluster 12 belonged to the bacterial phylogenetic group Firmicutes, and exhibited a rather small average

amino acid length of 658 ± 41 residues. Cluster 13 was comprised of forty-five protein sequences, all eukaryotic sequences from the phylogenetic group Viridiplantae, and was the largest of all twenty groups. In addition to its large group size, cluster 13 exhibited the largest average amino acid length per sequence at 1206 ± 214 residues. These unique features may contribute to its unique AveHAS plot, as described below.

Cluster 14 was made up of nine proteins, which averaged 826 ± 76 amino acid residues. Two of its sequences belonged to the eukaryotic phylogenetic group Viridiplantae. These eukaryotic sequences were at least one-hundred residues longer than the remaining seven sequences, which all belonged to the bacterial phylogenetic group Cyanobacteria. Cluster 15 was comprised of nineteen bacterial proteins belonging to the phylogenetic groups Firmicutes, Fusobacteria, Spirochetes, **Δ-**proteobacteria, and ε-proteobacteria. These sequences had an average sequence length of 786 ± 63 amino acid residues.

Both clusters 16 and 18 were comprised of three bacterial protein sequences. The sequences found in cluster 16 averaged 806 ± 26 amino acid residues in length and were of the phylogenetic groups β-proteobacteria and γ-proteobacteria. The sequences found in cluster 18, however, had a shorter average sequence length of 725 ± 4 amino acid residues and were entirely from Firmicutes. Cluster 17 contained seventeen bacterial protein sequences from the phylogenetic groups α-proteobacteria, β-proteobacteria, γ-proteobacteria, and Chlamydiae. These sequences had an average length of 757 ± 34 amino acid residues.

Cluster 19 consisted of twenty-two bacterial protein sequences, which had an average length of 796 ± 49 amino acid residues. These sequences were almost entirely from the phylogenetic group Firmicutes, with only one sequence belonging to the phylogenetic group Spirochaetes. Lastly, cluster 20 was comprised of thirty protein sequences from the bacterial phylogenetic groups α-proteobacteria, β-proteobacteria, γ-proteobacteria. These sequences had an average length of 820 ± 92 amino acid residues.

2.2: Analyses of Segments of Unusual Length

The sequences within the Copper P-type Family that appeared notably shorter or longer than other sequences, particularly than those they clustered with in the phylogenetic tree, were subjected to a more thorough examination (Figure S2; Table 2). After identifying sequences of unusually great length in Family 5, the multiple alignment for the proteins of Family 5 was used to determine if these sequences had extra segments, as compared to their neighboring sequences. This was accomplished using the programs TCDB BLAST and NCBI BLAST. TCDB BLAST compiled and ranked a list of proteins within its database that shared similarities with the given segment of amino acid residues. NCBI BLAST performed a similar task on these proteins in its own database, but was also able to take its search one step farther through its ability to utilize NCBI's Conserved Domain Database (NCBI CDD), which provided a ranked listing of possible conserved domains present within the segment of

interest. Many of the unusual segments or sequences examined did not contain any unique conserved domains that exhibited a strong E-value. Nonetheless, several proteins did have segments or entire sequences of interest, as described below (Table 8).

Forty sequences or segments of sequences were analyzed from the proteins representing the Copper P-type ATPases (Family 5). Thirty-three segments from a collection of twenty-nine sequences that were longer than the other sequences they clustered with were analyzed. Additionally, seven full-length sequences that were unusually short in length were analyzed to assure that they did, in fact, contain the necessary domains required to be fully-functional P-type ATPases. As expected, most of the TCDB BLAST and NCBI BLAST results indicated that the segments and sequences analyzed exhibited the strongest relationship to proteins from Family 5, Family 6, or both. However, in several of the segments a protein from 9.A.2.1.1, identified as a Periplasmic mercury ion binding protein (TCDB) was also present amongst the retrieved proteins in TCDB BLAST. This protein was retrieved with values above threshold (1e-04) for Fsp4, Dra1, Det1, Dha1, Cac1, Tde1, Bma1, Bce1, Bce2, Psp1, and Reu1. It was also amongst the proteins retrieved in the TCDB BLAST for the extra segment from the sequence Asu1, but it had an E-value below threshold (0.006). In addition to the retrieval of the Periplasmic mercury ion binding protein, a few other unexpected proteins were retrieved while performing TCDB BLASTs on the segments of interest, but they all exhibited incredibly E-values below threshold, and therefore were not included in this analysis.

Many conserved domains or fragments of conserved domains were identified by NCBI CDD during analysis of the thirty-three segments that were notably larger than most other sequences within their respective clusters. Two different heavy metal associated domains (HMAs) were identified, cd00371 and pfam00403, with at least one present in all but Psp2, Rpa1, Sal1 and Bps1 from cluster 6, Nsp3 and Reu2 from cluster 10, Spn1 from cluster 12, and Sth2 from cluster 19. Other common conserved domains included COG2608 (CopZ), PRK10671 (copper transporter), COG2217 (ZntA), and PRK11033 (zntA). Amongst the segments examined in cluster 10, the conserved domains pfam04945 (YHS domain), cd01057 (AAMH_A, Aromatic and Alkene Monooxygenease Hydroxylases, subunit A), COG3350 (uncharacterized conserved protein) and smart00746 (TRASH, metallochaperone-like domain) were common, and were all found in Rpa1, Sal1, Nsp6, Bps1 in cluster 6 contained all of these domains but cd0157, and Reu2 in cluster 6 contained COG3350 and smart00746. Reu2 also contained an unusual conserved domain, PRK00807 (50S ribosomal protein L24e), but its E-value, 0.001, was below threshold. This domain was also found in the segment analyzed from Nsp6, also in cluster 10, which had an E-value above threshold, 3e-04.

In addition to exhibiting several of the more common conserved domains, Dha1 from cluster 11 also contained the conserved domain cd5062_PTKc_IGF-1R, Protein-Tyrosine-like Kinase Family; Insulin-like Growth Factor-1 Receptor; catalytic domain. Although its E-value was below threshold, 0.010, it seems possible that the sequence could, in fact contain such a domain. Spn1 from cluster 12 only had a

fragment of a single conserved domain located in its extra segment, COG4633, an uncharacterized protein conserved in bacteria. Although only part of this conserved domain is actually found in a portion of the sequence analyzed, it exhibited a strong E-value of 1e-12, and it was the only segment of any of the sequences analyzed in Family 5 in which NCBI CDD detected this domain. In cluster 19, Sth2, from *Symbiobacterium thermophilium*, an Actinobacterium, contained fragments of conserved domains not detected by NCBI CDD in any of the other segments from the sequences analyzed in Family 5. The fragments of these domains, pfam00115 (COX2, Cytochrome C oxidase subunit II, periplasmic domain) and COG2131 (SufI, Putative multicopper oxidases), both exhibited E-values above threshold, 6e-05 and 6e-04, respectively.

Although segments from sequences in clusters 2, 4, 6, 8-13, 15, 19 and 20 were analyzed, clusters 6, 10, and 20 contained the most unusually long sequences. The proteins analyzed from cluster 6 were Rfe2, Psp2, Rpa1, Sal1, and Bps1, from *Rhodoferax ferrireducens, Polaromonas* sp. JS666, *Rhodopsudomonas palustris, Xanthomonas axonopodis pv. vesicatoria,* and *Burkholderia pseudomallei*, respectively (Table 2). These sequences range between 787 to 973 amino acids long, and are from the phylogenetic groups α-proteobacteria, γ-proteobacteria, and β-proteobacteria, respectively (Table 8). The long proteins analyzed from cluster 10 were Nsp6, Nsp3, Det1, Reu2 and Mbu1. Nsp6 and Nsp3 were both from *Nocardioides* sp. JS614, and consequently from the phylogenetic group Actinobacteria. The remaining three segments were from *Dehalococcoides*

ethenogenes, *Ralstonia eutropha*, and *Methanococcoides burtonii*, respectively. These sequences were between 828 and 1071 amino acids long and are from Chloroflexi, β-proteobacteria, and Euryarchaeota, respectively. Lastly, the long proteins analyzed in cluster 20 were Bma1, Bce1, Bce2, Psp1 and Reu1, which were between 816 and 1061 amino acids in length. These proteins were from *Burkholderia mallei, Burkholderia cepacia*, *Burkholderia cenocepacia*, *Polaromonas* sp. JS666, and *Ralstonia eutropha*, respectively, and were all β-proteobacteria. Despite their length, however, they did not exhibit any unusual conserved domains, but instead exhibited a collection of HMAs cd00371 and pfam00403, COG2608 (CopZ), PRK10671 (copper transporter), COG2217 (ZntA), and PRK11033 (zntA).

Amongst the proteins with the longest sequences found in Family 5 were eukaryotic proteins labeled Hsa1, Cfa2, Dme1 and Ddi2 from cluster 13, and bacterial proteins labeled Nsp6, Bma1, Bce1, and Bce2 from clusters 10 and 20 (Table 2). These sequences were all over one-thousand amino acid residues long, and were chosen for examination either because their length was an anomaly within their cluster, or as a representative sequence of part of a cluster or an entire cluster that is unusually long. The eukaryotic sequences were from the organisms*, Homo sapiens, Canis familiaris, Drosophila melanogaster,* and *Dictyostelium discoideum,* and were all Metazoans and slime molds, although cluster 13 also consisted of a few sequences from Fungi and one sequence from Viridiplantae. Nsp6, the only bacterial sequence examined from cluster 10, was from an Actinobacterium from *Nocardiodes* sp. JS614, whereas the bacteria examined from cluster 20 were from *Burkholderia ambifaria*

AMMD, *Burkholderia cepacia* R18194, and *Burkholderia cenocepacia* HI2424, respectively, were all β-proteobacteria. Despite their differences, most of the TCDB BLAST and NCBI BLAST results indicated that the segments of these proteins that extended beyond the length of some of the shorter sequences in their clusters still were most closely related to Copper or Heavy metal P-type ATPases. No results were found on TCDB BLAST for Ddi2, nor were any conserved domains found on NCBI CDD. Also, the only result retrieved for this segment in NCBI BLAST with an E-value above threshold (9e-104) was a hypothetical protein DDBDRAFT_0217251. Similarly, in Dme1, no conserved domains were detected by NCBI CDD, nor were any results retrieved by TCDB BLAST, but one copper and one heavy metal P-type ATPase were detected at levels above threshold by NCBI BLAST.

Nearly every segment from the above long sequences analyzed indicated the presence of at least one HMA (heavy metal-associated) domain, with the exception of Nsp6, Ddi2, Cfa2, and Dme1, which did not have any conserved domains detected by NCBI CDD. The segments from three of these sequences, Ddi2, Cfa2, and Dme1, did not pull up any results on TCDB or NCBI BLAST. Nsp6 did not produce any results on NCBI BLAST and only had one protein retrieved on TCDB BLAST, a L-lysine transport protein from 2.A.3.2.4, which had a E-value below threshold (0.85). While it is uncertain at this time why these extra segments, which distinguish these proteins from others which they cluster with, do not have any conserved domains, it is possible that their extra amino acid residues still play a role in their biological function by

binding heavy metals. Further examination of this hypothesis, however, is not within the scope of this thesis.

In addition to closely examining the largest sequences found amongst the Copper P-type ATPases, the shortest of the sequences found with Family 5 were also carefully evaluated to insure that they were, in fact, full-length protein sequences. Prior to even examining these sequences for the presence of the nine known well conserved sequence motifs (see above), a couple of these sequences were selected for sequence examination via the TCDB BLAST, NCBI BLAST and NCBI CDD programs. The seven sequences identified as having shorter lengths than all or most of their neighboring sequences were Rge3, Bfu2, Sty1, Nsp4, Mth2, Ptr1, and Cte1, which were from *Rubrivivax gelatinosus, Burkholderia fungorum, Salmonella typhimurium, Nostoc sp. PCC 7120, Moorella thermoacetica, Pan troglodytes* and *Clostridium tetani*, respectively. These sequences were between 670 and 1197 amino acids in length, with Rge3 and Bfu2 from cluster 6, Neu3 from cluster 8, Nsp4 and Mth2 from cluster 10, Ptr1 from cluster 13, and Cte1 from cluster 15 (Figures 2.A and 2.B; Table 2). Rge3 and Sty1 were β-proteobacteria, Sty1 was a γ-proteobacterium, Nsp4 was a Cyanobacterium, Mth2 and Cte1 were Firmicutes, and Ptr1 was a Metazoan. Although the lengths of these proteins are shorter than most or all of their neighboring sequences, all of these sequences brought up only P-type ATPases on TCDB, most or all of which were Copper or Heavy metal P-type ATPases. Interestingly, the protein labeled Ptr1 was described as being similar to ATP7B. Additionally, the results obtained from NCBI CDD indicated that both sequences

contained an E1-E2 domain and a hydrolase domain, both with E-values above threshold, and thereby provided evidence that both sequences were, in fact, full-length P-type ATPases.

2.3: Hydropathy Analysis

The WHAT program or a newly updated version of AveHAS (TCDB) was used to analyze each of the twenty clusters from Family 5 (Figure 3.A.1- 3.A.20). The recently modified AveHAS program not only displays information regarding the similarities, hydrophobicities and amphipathicities of the sequences present in the multiple alignment, but also generates a prediction regarding the numbers and locations of TMSs present. The AveHAS program predicted 8 TMSs for clusters 1, 2, and 4-20, which were subsequently labeled as TMSs A, B, 1, 2, 3, 4, 5, and 6. It predicted the existence of 8 TMSs, which were labeled as A, B, 1, 2, 3, 4, 5, and 6. Cluster 3, by contrast, appeared to have 9 TMSs, which were labeled as Y, A, B, 1, 2, 3, 4, 5, and 6. In lieu of AveHAS, the WHAT program was used to analyze the single sequence belonging to cluster 7. The additional transmembrane segment predicted in cluster 3, TMS Y, appeared to be in front of TMSA and did not seem to impact the grouping of the other eight TMSs. All of the clusters but clusters 6 and 10 had TMSs A, B, and 1-6 arranged in a fairly consistent pattern, where TMSA and TMSB grouped closely together and were separated only by a short gap from TMS1 and TMS2, which also were close to one another. TMS1 and TMS2 were then separated by another short

gap from TMSs 3 and 4. After a much longer gap than that separating TMS2 from TMS3, and TMS5 and TMS6 were found clustering closely together. The eight TMSs detected in clusters 6 and 10, however appeared to deviate from the typical pattern of TMS groupings. Although TMSs 1-6 appeared to group together as expected, TMSA was separated by a sizeable gap from TMSB, which grouped more closely to TMSs 1 and 2.

In addition to using the AveHAS program to analyze the sequences of the individual clusters within Family 5, it was also used to analyze one protein sequence from each of the twenty clusters representing the Copper P-type ATPases. These data were used to provide a means for identifying the regions in Family 5 which exhibited the most strongly conserved amphipathic peaks in the sequence, the best of which was examined using the EMBOSS PEPWEEL program (see below).

2.4: HMMTop Analysis

The program HMMTop was employed to predict the numbers and locations of transmembrane sequences in each of the twenty clusters, which collectively represent the Copper P-type ATPase Family (http://www.tcdb.org/progs/hydropathy.php). This program is an optimization method for predicting transmembrane segments and topology. It is designed to take into account the effects of how the placement of a given set of residues in a given segment of a protein can impact other segments within that protein (Dosztányi Z., et al., 2003).

Unlike the AveHAS program, HMMTop can only be run on one protein sequence at a time. Thus, one protein sequence was selected from each of the twenty clusters that were created based on clustering patterns (see above). While using only one sequence per cluster limits the strength of the results obtained, these data provide useful information when compared to the results for each cluster obtained with the WHAT, AveHAS, and SOPMA programs (Figures 3.A.1-.20, 3.B.1-.20, and 3.C.1-.20). In agreement with the information yielded by the AveHAS plots, HMMTop analysis revealed that Motif 4 (M4) was located inside of TMS4 of all twenty clusters. Additionally, HMMTOP analysis of each sequence provided the numbers and locations of its TMSs, thereby more clearly identifying the location of each of the nine recognized conserved motifs in relation to each TMS (Møller J.V., et al., 1995).

All but two of the twenty clusters had eight TMSs, which were labeled A, B, 1, 2, 3, 4, 5, and 6 (Figures 3.B.1-.20). The proteins analyzed from cluster 3 had 9 TMSs, as was also indicated by the AveHAS analysis (Figures 3.A.1-.20). By contrast, the protein examined from cluster 6 only yielded a prediction of 7 TMS, which differed from the value predicted by the AveHAS program. The protein from cluster 6 chosen for HMMTop analysis was Cvi2, from *Chromobacterium* violaceum ATCC 12472, and was one of several sequences in group 6 belonging to β-proteobacteria. It did not differ significantly in length from the other protein sequences in cluster 6. Based on the locations of the nine known well-conserved motifs, it appeared that the transmembrane sequences designated TMSs 3-6 were present in this protein. In the AveHAS plot corresponding to cluster 6, TMSs A and B were located unusually far

apart from each other. These data suggest that the missing TMS noted by HMMTop in Cvi2 could be either TMSA or TMSB, as the lack of one of those two TMSs in a sequence belonging to cluster 6 could impact the calculations made by the AveHAS program when predicting the average distances of these two TMSs amongst its sequences.

2.5: SOPMA Analysis

Like HMMTop, the SOPMA program was conducted on one sample protein from each of the twenty clusters in Family 5 (Figure 3.B.1- 3.B.20). SOPMA is a program designed to predict secondary structures of any protein sequence, providing predictions on a residue-to-residue basis (Geourjon and Deléage, 1995). The predictions of TMSs for each of the twenty clusters generally appeared to correspond to the predictions made by HMMTop and the AveHAS or WHAT programs, and most, if not all, of the differences in predictions made by these programs can be attributed to the differences in their design.

In addition to examining the numbers and locations of TMSs in comparison to the locations of the nine known well-conserved motifs, the SOPMA program was also used to evaluate the secondary structure patterns of each of the nine motifs for all twenty clusters. The program is designed to detect a number of different secondary structures and assign them a distinct label in the form of one letter of the alphabet. However, the only letters used to describe the secondary structures within any of the

motifs for all twenty clusters were the letters C, E, H, and T, which stood for random coil, extended strand, alpha helix, and beta turn, respectively (Figures 3.C.1-.20).

The secondary structure patterns for Motifs 1, 2, 3, and 5 were conserved throughout all twenty clusters and were ttc, ccc, ttcc, and ctccee, respectively. Motif 4 exhibited either the secondary structure patterns ccch or cccc for all twenty clusters. Most of the twenty clusters exhibited a secondary structure pattern for Motif 6 that was either ttcce, or some variation close to it. The clusters that exhibited some slight variation of this pattern were as follows: Clusters 8, 16, and 20 displayed etcce, clusters 9 and 17 displayed etccc, cluster 10 displayed ttccc, and cluster 18 displayed ttcce. The secondary structure pattern observed in all twenty clusters for Motif 7 was either hhcc or cccc. Motif 8 exhibited the secondary structure pattern eeecc in all but cluster 5, which was eeetc. Lastly, Motif 9, which is comprised of 23 residues, was consistently composed of a string of several "e's" followed by several "c's", then several "h's", and finally several more "e's." These results indicated that most of the conserved motifs are in exposed surface regions of the proteins, possibly at the ends of the α–helices or β–strands. They are present in regions that are likely to include β–turns or random coils.

While the SOPMA program does not provide an actual prediction for the number of TMSs, its in-depth residue secondary structure predictions can be used to assess the predictions regarding the locations and numbers of predictions made by the WHAT or Ave Has programs and the HMMTop program. As expected, analysis of the data generated by the SOPMA program indicates that regions which were predicted by

the WHAT or Ave Has programs and the HMMTop program to have transmembrane segments do, in fact, have multiple residues with helical topologies. Additionally, the SOPMA program generates a visual plot of the patterns created by helices, sheets, turns and coils. A comparison between the distances, in amino acid residues, between peaks in the SOPMA-generated plots and those generated by the AveHAS program generally appear to match. While some discrepancies between AveHAS, HMMTop and SOPMA data were observed, these were generally minor differences and are likely attributable to the fact that data obtained from AveHAS represent the average pattern for a group of proteins, whereas HMMTop and SOPMA only analyze one protein at a time.

2.6: EMBOSS Pepwheel Analyses

EMBOSS Pepwheel is a program designed to detect and facilitate analysis of amphipathic regions in a sequence through the generation of alpha helices (http://www.tcdb.org/progs/pepwheel.php). An AveHAS plot containing one sequence from each cluster representing Family 5 indicated that the region between TMS2 and TMS3 had two well-conserved amphipathic peaks (Figures 8). Analyses of the twenty AveHAS plots corresponding to each of the clusters representing the Copper P-type ATPases indicated that these peaks were expressed most strongly in cluster 15. Therefore, the EMBOSS Pepwheel program was used to analyze the segment of amino acids between TMS2 and TMS3 in a single sequence from cluster 15, SSu1 (from

Streptococcus suis). Initial analysis of the sequence motifs found between TMS2 and TMS3, M1- M3, did not reveal any strongly amphipathic regions. However, analysis of the regions between these motifs and TMS2 and TMS3 indicated the fifteen amino acids between M2 and M3 aligned such that amphipathic residues and hydrophobic residues were almost completely separated on opposite sides of the predicted alpha helix. Two strongly hydrophilic residues, aspartic acid (D) and glutamine (Q), were surrounded by hydrophobic residues. While the disruption of the hydrophobic region by these two hydrophilic residues indicates that the amino acids in this region form a structure near the protein surface that in some way contributes to an important, conserved function, as indicated by the AveHAS plots for Families 5 and 6. One possible explanation for the presence of such a helix between M2 and M3 is to provide flexibility for the movement of the TGES loop, which has been proposed to occur as part of the conformation change that occurs between the phosphorylated and dephosphorylated states (Anthonisen, et al.; 2006).

2.7: 16S rRNA Analysis

The 16S ribosomal RNA sequences used to analyze the prokaryotic genera present in Families 5 were collected from the NCBI database, as previously described. A total of 138 different genera were identified in Family 5 after removing unclassified organisms from the search, and the corresponding 16S rRNA for each genus was used to perform a multiple alignment and to generate a phylogenetic tree (Table 4).

Although some exceptions were noted, for the most part, the clustering patterns of the genera in the 16S rRNA tree were similar to those observed among the corresponding protein sequences in the phylogenetic tree representing the Copper P-type ATPases (Figure 4).

As expected, all eleven archaeal genera, which consisted of the phylogenetic groups Thermoprotei, Crenarchaeota and Euryarchaeota, clustered closely together. The bacterial genera *Deinococus, Thermus, Chloroflexus, Aquifex* and *Thermotoga* were located clustering near these archaeal genera. These bacterial genera belong to diverse range of phylogenetic groups: Gloeobacteria, α-proteobacteria, Chloroflexi, Aquificae and Thermotogae, respectively, and appear to cluster distinctly from other bacterial organisms, even those from the same bacterial phylogenetic group (Figure 4).

The remaining bacterial genera primarily clustered together by phylogenetic groups, with several distinct clusters predominately composed of Firmicutes, Cyanobacteria, Actinobacteria, α-proteobacteria, β-proteobacteria, δ-proteobacteria and γ-proteobacteria. One notable exception to these fairly well-organized clusters, however, was a cluster composed of the genera *Rhodospirellula, Bacteroides, Cytophaga, Chlorobium, Pelodictyon, Campylobacter, Thiomicrospira, Helicobacter* and *Wolinella*. These genera were from the phylogenetic groups Planctomycetacia, Bacteroidetes, Chlorobi, ε-proteobacteria, and β-proteobacteria. Another unusual cluster observed amongst the bacterial genera was one containing *Solibacter, Leptospira, Treponema, Candidatus,* and *Dehalococcoides,* which belong to Acidobacter, Spirochaetes, Spirochaetes, Chlamydiae and Chloroflexi, respectively.

Additionally, there were a few genera that clustered amongst genera that belonged to a different phylogenetic group than their own. *Arthrobacter*, from β-proteobacteria, was found in a cluster of genera from *Actinobacteria. Fusobacter,* from Fusobacteria, was found in a cluster of genera from Firmicutes, as was *Listeria*, from Spirochaetes. Lastly, the genera *Thiobacillus* and *Xanthomonas*, from β–proteobacteria and γ-proteobacteria, respectively, were found in a cluster with γ–proteobacteria.

Chapter 3: The Heavy Metal P-Type ATPases

The Heavy metal P-Type ATPases

Heavy metal P-type ATPases are found in both prokaryotes and eukaryotes, and are classified as Family 6 in TCDB. Phylogenetic analyses indicate that while a number of Family 6 proteins exhibit functional and phylogenetic similarity with proteins found in Family 5, the proteins belonging to these two phylogenetic families are distinguishable from one another.

311 Heavy Metal P-type ATPase sequences were collected from several non-redundant psi-BLASTs (NCBI), which were combined after elimination of redundant sequences using the program CD-Hit (Table 5). A multiple alignment was performed on the combined sequences, a phylogenetic tree was generated, and the 16S rRNA of the prokaryotic genera present within these sequences were analyzed (see below). The sequences belonging to Family 6 were divided into seventeen smaller clusters based on their branching patterns in the corresponding phylogenetic tree. Multiple alignments were conducted on the clusters containing more than one protein sequence, and motif analyses were conducted on the sequences within each cluster. All clusters then were run through either the AveHAS program or the WHAT program and underwent motif, EMBOSS Pepwheel, HMMTOP, and SOPMA analyses.

3.1: Phylogenetic Analysis

The 311 sequences representing the Heavy metal P-type ATPase Family were divided into seventeen clusters based on clustering patterns in the corresponding

phylogenetic tree, as previously described (Figures 5.A-5.B). Cluster 1 was the largest of all seventeen groups and consisted of fifty-eight bacterial protein sequences, averaging 760 ± 59 amino acid residues in length. These proteins were from the phylogenetic groups -proteobacteria, -proteobacteria, -proteobacteria, and Deinococci. Cluster 2 consisted of forty-seven protein sequences, averaging 776 ± 63 amino acid residues in length. These proteins belonged to -proteobacteria, -proteobacteria, -proteobacteria, and Cyanobacteria.

Nine bacterial protein sequences were found in cluster 3. These sequences belonged to the phylogenetic groups α-proteobacteria, β-proteobacteria, γ-proteobacteria, and Planctomycetes, and had an average length of 775 ± 98 amino acid residues. Cluster 4 was comprised of only two bacterial protein sequences, which exhibited a rather small average length of only 664 ± 32 amino acid residues. These two protein sequences were from the phylogenetic groups Firmicutes and Chlamydiae.

Cluster 5 was the second largest of all seventeen groups with fifty-seven protein sequences. These sequences had an average length of 695 ± 60 amino acid residues. Three of these sequences belonged to the archaeal phylogenetic group Euryarchaeota. The remaining fifty-four sequences were from δ-proteobacteria, γ-proteobacteria, ε-proteobacteria, Firmicutes, Cyanobacteria, and Spirochaetes.

Cluster 6 contained only six protein sequences, all from the eukaryotic phylogenetic group Viridiplantae. These proteins exhibited the largest average amino acid length of all seventeen groups at 989 ± 175 proteins. Cluster 7 consisted of fourteen residues, which had a smaller average sequence length of only 640 ± 31 amino acid residues. Two of the sequences were from Euryarchaeota, while the

remaining five protein sequences were from the bacterial phylogenetic groups of Firmicutes, Chloroflexi, Actinobacteria, and α-proteobacteria.

Cluster 8 was composed of nineteen protein sequences, which were further subdivided into parts A, B, and C. These subdivisions had average sequence lengths of 639 ± 14 residues, 821 ± 2 residues, and 664 ± 13 residues, respectively. Part A was composed of the bacterial phylogenetic group Firmicutes, Part B was composed of the eukaryotic phylogenetic group Viridiplantae, and Part C was composed of the bacterial phylogenetic group Chlamydiae.

Cluster 9 consisted of twenty-five bacterial protein sequences from the phylogenetic groups Actinobacteria, Bacteroidetes, α-proteobacteria, and β-proteobacteria. These sequences had an average length of 794 ± 82 amino acid residues. Cluster 10 contained ten protein sequences, averaging 696 ± 83 amino acid residues in length. These sequences were from the bacterial phylogenetic groups Cyanobacteria, Actinobacteria, and Chloroflexi.

Cluster 11 was made up of a single protein sequence, Lpn3, from *Legionella pneumophila* subsp. pnuemophila str. Philadelphia 1, which was described as a cadmium efflux ATPase (NCBI). This γ-proteobacterium had a sequence length of 635 amino acid residues. Like cluster 11, cluster 17 consisted of only one protein sequence, Cau1, which was 684 amino acid residues long and was from *Chloroflexus aurantiacus* (group Chloroflexi).

Cluster 12 contained twenty-four proteins, all belonging to the bacterial phylogenetic group Actinobacteria. These sequences were further subdivided into

12A, 12B, and 12C, based on their amino acid lengths. The average amino acid lengths for these sequences were 653 ± 13, 716 ± 9, and 638 ± 25, for Parts A, B, and C, respectively. Both 12A and 12B primarily consisted of organisms belonging to the genus *Mycobacterium,* although 12A also contained the genera *Janibacter, Nocardia, Rhodococcus, Arthrobacter* and 12B contained the genera *Rhodococcu and, Gordonia.* 12C, by contrast, was mostly composed of organisms of the genus *Corynebacterium,* although it also contained an organism from *Brevibacterium.* Interestingly, 12A and 12C, the sub-clusters most similar in amino acid length, each contained a single organism simply described as a marine actinobacterium.

Cluster 13 consisted of seven proteins, which averaged 813 ± 60 amino acid residues in length. These sequences all belonged to the archaeal phylogenetic group Euryarchaeota. Cluster 14 was comprised of only three proteins, and like cluster 13, they were all from Euryarchaeota. However, unlike cluster 13 it exhibited a much smaller average sequence length of only 678 ± 65 amino acid residues.

Cluster 15 contained twenty-one protein sequences, all belonging to Firmicutes. These proteins had an average sequence length of 731 ± 55 amino acid residues. Cluster 16 consisted of seven bacterial protein sequences, averaging a sequence length of 759 ± 56 amino acid residues. These proteins belonged to the phylogenetic groups Cyanobacteria, Actinobacteria, and Chloroflexi.

3.2: Analyses of Segments of Unusual Length

A more thorough analysis was conducted on the sequences within the Heavy Metal P-type Family that appeared notably shorter or longer than other sequences, particularly than those they clustered within the phylogenetic tree. After identifying sequences of unusually great length in Family 6, the multiple alignment for the proteins of Family 6 was used to identify where these sequences had extra segments of amino acids, as compared to their neighboring sequences (Figure S3; Table 5). This was accomplished using the programs TCDB BLAST and NCBI BLAST. TCDB BLAST compiled and ranked a list of proteins within its database that shared similarities with the given segment of amino acid residues. NCBI BLAST performed a similar task on these proteins in its own database, but was also able to take its search one step farther through its ability to utilize NCBI's Conserved Domain Database (NCBI CDD), which provided a ranked listing of possible conserved domains present within the segment of interest. Many of the unusual segments or sequences examined did not contain any unique conserved domains that exhibited a strong E-value. Nonetheless, several proteins did have segments or entire sequences of interest, as described below (Table 9).

Thirty-five sequences or segments of sequences were analyzed from the proteins representing the Heavy Metal P-type ATPases. Twenty-nine segments from a collection of twenty-three sequences that were longer than the other sequences they clustered with were analyzed. Six full-length sequences that were unusually short in length were also analyzed to assure that they contained the basic domains required to function as P-type ATPases. The results for most of the TCDB BLAST and NCBI

BLAST queries indicated that the segments and sequences analyzed were most similar to proteins from Family 5, Family 6, or both. However, a few unique proteins were brought up in these BLASTs. The most common proteins pulled up by these searches were Periplasmic mercury binding proteins from 9.A.2.1.1, which were found at values above threshold (1e-04) in Hma3, Hma1, and Hwa1 from cluster 13 and in Atu1 from cluster 1. This protein was also found at values below threshold in Ssp1 (0.002) and Msp2 (0.002) from cluster 1. Additionally, although no results were found using TCDB BLAST for the segment from Cau2 from cluster 10, NCBI BLAST found one Heavy metal P-type ATPase sequence and numerous transcriptional regulators with E-values above threshold. The transcriptional regulators found in the NCBI BLAST search were primarily TrmB, a sugar-specific transcriptional regulator. These data correspond with the results obtained from NCBI CDD, which indicate the presence of three transcriptional regulators as conserved domains in this segment. The conserved domains detected all had E-values above threshold and were pfam01978 (TrmB), COG3355, and COG1378. Additionally, a few other unexpected proteins were detected amongst the TCDB BLAST results for several of the segments of interest. However, these proteins all exhibited E-values below threshold, and were subsequently excluded from analysis.

The twenty-nine segments from long sequences analyzed from Family 6 primarily either lacked any conserved domains or exhibited one or a collection of a small repertoire of conserved domains or fragments of conserved domains. Using NCBI CDD, two heavy metal associated domains (HMAs) were identified, cd00371

and pfam00403. These two domains were detected along with the domain COG2608 (CopZ) in Hma3, Hma1, Hwa1 and Hma2 of cluster 13, Ssp1, Atu1 and Msp2 of cluster 1, and Rme1 and both segments from Cte1 from cluster 2. Other common conserved domains included PRK10671 (copper transporter), COG2217 (ZntA), and PRK11033 (zntA), a zinc/cadmium/mercury/lead-transporting ATPase, which are all found in Hma3, Hma1, Hwa1 and Hma2 of cluster 13, and Ssp1, Atu1 and Msp2 of cluster 1.

While sequence from clusters 1, 2, 3, 5, 6, 8, 9, 10, and 13 were analyzed, clusters 1, 2, and 13 contained the most proteins with unusually long sequences, and were the only clusters with detected conserved domains. Despite their length, no conserved domains were detected in clusters 3, 5, 6, 8, 9, or 10. Ath2 and Osa2, proteins from Viridiplantae found in cluster 8, were about eight-hundred amino acids long and were the only eukaryotic proteins in a cluster predominantly consisting of bacteria from Firmicutes and Chlamydiae, which were all only about six-hundred amino acids long. While eukaryotes typically have longer sequence lengths than prokaryotes, this does not explain why eukaryotes within the same cluster would be so much longer than their prokaryote counterparts without gaining some additional function. Thus, while no conserved domains were detected within the segments analyzed for Ath2 and Osa2, it is possible that these extra amino acids provide some unknown function. Similarly, the sequences from cluster 9, Lbl1, Sru1, Asp5, Csp1, Asp4, Eli1, and Bmu2, from *Leewenhoekiella blandensis, Salinibacter rubber, Athrobacter* sp. FB24, *Caulobacter* sp. K31, *Acidovorax* sp. JS42, *Erythrobacter*

litoralis, and *Burkolderia multivorans,* are also all approximately eight-hundred amino acids in length by contrast to several sequences in cluster 9 that were approximately six-hundred amino acids long. These sequences are from the phylogenetic groups Bacteroidetes, Actinobacteria, α-proteobacteria and β-proteobacteria, and their difference in size from the shorter sequences within their cluster cannot be attributed to differences in phylogenetic groups, as several of the shorter sequences belong to some of the same phylogenetic groups as the longer sequences. It is possible, however, that these additional amino acids may confer some unknown benefits in certain environmental niches, or that bacteria in certain environmental niches were highly exposed to another organism bearing these additional residues, and incorporated them into their genetic material at a high frequency. The same reasoning may hold true for Msp1, from *Mesorhizobium* sp. BNC1, an α-proteobacterium from cluster 3 that was approximately three-hundred amino acids longer than any of the other sequences it clustered with, even other α-proteobacteria. Lastly, none of the segments of Ota1, Tca2 or Aha1 from cluster 6, from *Ostreococcus tauri, Thlaspi caerulescens,* and *Arabidopsis halleri,* respectively, indicated the presence of any conserved domains, despite the fact they were at least one-hundred amino acids longer than any of the three remaining sequences in cluster 6, which are also from Viridiplantae. While the actual usefulness of these excess residues is unknown, it is possible they provide each organism with increased fitness in their respective environmental niches.

In addition to closely examining the largest sequences found amongst the Heavy metal P-type ATPases, the shortest of the sequences found with Family 6 were

also carefully evaluated to confirm that they were, in fact, full-length protein sequences containing the E1-E2 domain and hydrolase domain characteristic of P-type ATPases. The six proteins identified as having shorter sequence lengths than most of their neighboring sequences were Ssp5, Pla2, and Pmi1 from cluster 1, Aav1 and Asp6 from cluster 2, and Csp4 from cluster 5. Ssp5, from *Sulfitobacter* sp. EE-36 and Pla2, from *Parvibaculum lavamentivorans* are both α-proteobacterial proteins from the same sub-cluster within cluster 1. Pmi1, from *Proteus mirabilis*, is a γ-proteobacterium from another sub-cluster in cluster 1. These three sequences are about six-hundred amino acids long, whereas most other sequences in cluster 1 are one- to two-hundred amino acids longer. Two unusually short sequences were examined from cluster 2: Aav1 and Asp6, from *Acidovorax avenae* and *Acidovorax* sp. JS42, respectively. These two β-proteobacteria, which were 629 and 673 amino acids long, respectively, were at least fifty to one-hundred amino acids shorter than the other sequences they clustered with. Similarly, Csp4, from *Clostridium* sp. OhILAs, was much shorter, by at least one-hundred amino acids, than its neighboring sequences in a sub-cluster of cluster 5, even though they were all from Firmicutes. Despite their short lengths, however, all six proteins have E1-E2 and hydrolase domains with E-values above threshold.

3.3: Hydropathy Analysis

As performed on the groups of sequences representing Family 5 (see above), the WHAT program or the AveHAS program (TCDB) was used to analyze the seventeen clusters belonging to Family 6 (Figure 6.A.1- 6.A.17). As previously noted, the modified AveHAS program not only provides information regarding the similarities, hydrophobicities and amphipathicities of the sequences present in a given multiple alignment, but it also generates a prediction regarding the numbers and locations of TMS present.

Fifteen of the clusters representing Family 6 contained more than one protein sequence, and thus were analyzed using the AveHAS program. All of these sequences appeared to contain eight TMSs, as expected for type I ATPases (Møller J.V, et al., 1995). Two of the seventeen clusters representing the proteins of Family 6 consisted of only one protein sequence, cluster 11 and cluster 17. As dictated by program requirements, the WHAT program was used to analyze these two clusters instead of the AveHAS program. While eight TMSs were detected in cluster 11, their predicted locations appeared to deviate from what was seen in the other clusters with eight TMSs in Family 6. By contrast, only six TMSs were predicted for cluster 17. Examination of the plot constructed by the WHAT programs indicates the presence of the last four TMSs, 3, 4, 5 and 6, but also indicates that two of the first four TMSs, A, B, 1 or 2, were missing. As cluster 17 only contains one sequence, it is not clear whether or not it was incompletely sequenced and is missing a segment containing

additional TMSs or, perhaps, it was unique from the other clusters representing Family 6.

Of the clusters of proteins representing the Heavy Metal P-type ATPases that consisted of eight TMSs, nearly every one had their TMSs spaced in a fairly consistent pattern. The locations of these TMSs closely mirrored the most common pattern seen in Family 5: TMSA and TMSB grouped closely together with TMS1 and TMS2, which were followed after a short space by TMS3 and TMS4, which were then followed by a slightly longer space by TMS5 and TMS6. The nine known well-conserved motifs were generally located between TMSs, with Motif 4 usually found partially or completely within TMS4. Cluster 6, however, appeared to deviate from this expected pattern. Although it appeared to contain eight TMSs in its AveHAS plot, there was a sizeable gap between TMSA and TMSB, instead of between TMSB and TMS1, and TMSB was grouped together with TMS1 and TMS2. All of the sequences in cluster 6 are from Viridiplantae, and it is possible that the differences observed in cluster 6's arrangement of its TMSs locations is important for ion translocation in P-type ATPases from Viridiplantae, and perhaps for other eukaryotes as well.

3.4: HMMTop Analysis

The HMMTop program (Dosztányi Z., et al., 2003) was used on one sample protein from each of the seventeen clusters in an effort to determine the numbers and locations of the transmembrane segments (TMSs) present in each group

(http://www.tcdb.org/progs/hydropathy.php). This program also illustrated the locations of each of the nine well-conserved motifs in relation to these TMSs (Figure 6.B.1- 5.6.17). As such, these data provided another useful source of information that could be compared to the results obtained for each cluster through the WHAT, AveHAS, and SOPMA programs (Figures 6.A.1-.17, 6.B.1-.17, and 6.C.1-.17).

Many of the predictions made by HMMTop for the Heavy Metal proteins chosen from each of the seventeen clusters did not match the numbers of TMSs predicted by AveHAS. Only three of the clusters were predicted to have eight TMSs, clusters 3, 11, and 15. Clusters 9, 13, and 14 were predicted to only have five TMSs, whereas clusters 7 and 12 were predicted to have seven TMSs. Cluster 10 was predicted to have ten TMSs and clusters 1, 2, 4, 6, 8, 16, and 17 were predicted to have six TMSs. Despite these notably different predictions in the numbers of TMSs, almost all of the clusters had four TMSs following Motif 3, as expected. However, clusters 8, 9, 13 and 14 had only three TMSs after Motif 3, cluster 11 had five predicted TMSs after Motif 3, and cluster 10 had six predicted TMSs after Motif 3. While several differences were noted between predictions made by AveHAS and HMMTop, these are most likely attributable to the sampling size of each of these programs. HMMTop, like the SOMPA program, only examines one protein sequence at a time. Therefore, it does not predict the characteristics of the proteins in a given cluster as well AveHAS, which bases its predictions on averages of the values obtained from all of the protein sequences in a group.

3.5: SOPMA Analysis

The SOPMA program for secondary structure prediction was conducted on one sample protein from each of the seventeen groups representing branches from Family 6 (Figure 6.C.1- 6.C.17). Primarily, the SOPMA program was used to evaluate the secondary structure patterns of each of the nine motifs for all seventeen clusters. Additionally, the SOPMA program was used to compare the predictions of the numbers and locations of TMSs for each of the seventeen clusters made by the HMMTop and AveHAS or WHAT programs.

As mentioned previously, one of the functions of the SOPMA program is to detect a number of different secondary structures and assign them a distinct label in the form of one letter of the alphabet. However, as observed in Family 5, the only letters used to describe any of the motifs for the seventeen clusters representing Family 6 were the letters C, E, H, and T, which represent random coil, extended strand, alpha helix, and beta turn, respectively. The most common secondary structure patterns for each motif are described below.

As observed in Family 5, the secondary structure patterns for Motifs 1, 2, 5, and 8 were completely conserved throughout all seventeen clusters, and were ttc, ccc, and ctccee, and eeecc, respectively. Motif 3 most commonly exhibited the secondary structure pattern ttcc, which was conserved in all groups but clusters 9 and 14, where was cccc, and cluster 8 where it was ttce. As seen in Family 5, Motif 4 exhibited either the secondary structure patterns ccch or cccc for all seventeen clusters. Again, most of

the seventeen clusters exhibited a secondary structure pattern for Motif 6 that was either ttcce, or some variation close to it. Clusters 3, 4, 5, 7, and 10 displayed the pattern etcce, and cluster 12 displayed the pattern etccc.

The secondary structure pattern hhcc or cccc was observed at Motif 7 in all but group 17, which exhibited the pattern hhhh. Additionally, this different secondary pattern was not found in any of the clusters of Family 5 at Motif 7. Due to the uniqueness of this pattern, other residues between Motif 6 and Motif 8 were examined for the possible presence of a secondary structure pattern that matched one of the expected patterns for Motif 7, but no better candidates were identified. Lastly, Motif 9, which is comprised of 23 residues, consistently displayed of a string of several "e's" followed by several "c's", then several "h's", and finally several more "e's" in all seventeen groups.

Although the SOPMA program is not ideal for predicting the number of TMSs in a protein sequence, its detailed predictions regarding the secondary structure of each residue within a sequence helps identify potential locations for TMSs, which can then be compared to data obtained using the WHAT or AveHAS programs and the HMMTop program. As expected, analysis of the data generated by the SOPMA program indicates that regions that were predicted by the WHAT or Ave Has programs and the HMMTop program to have transmembrane segments do, in fact, have multiple residues with helical topologies. Additionally, the SOPMA program generates a visual plot of the patterns created by helices, sheets, turns and coils. A comparison between the distances, in amino acid residues, between peaks in the

SOPMA-generated plots and those generated by the AveHAS program appear to closely correlate. Examination of the results showed that SOPMA did not accurately predict transmembrane helices. It did, however, indicate that most observed motifs are exposed to the surfaces of the proteins, and are primarily located in coils and β-turns.

3.6: EMBOSS Pepwheel Analysis

Analysis of an AveHAS plot containing one sequence from each of the seventeen clusters representing the Heavy Metal P-type ATPases revealed that the region between TMS2 and TMS3 had two well-conserved amphipathic peaks (Figures 9). These two peaks were also noted in an AveHAS plot containing a sequence from each of the twenty clusters representing the Copper P-type ATPases. In Family 6, these two amphipathic peaks were exhibited the most strongly in cluster 10. As such, EMBOSS Pepwheel was used to analyze the segment of amino acids found between TMS2 and TMS3 in a single protein sequence from cluster 10, SSp3 (from *Synechocystis* sp. PCC6803). As observed in Family 5, no strongly amphipathic regions were detected at any of the sequence motifs between TMS2 and TMS3, whereas analyses of the regions between these motifs and TMS2 and TMS3 indicated that the fifteen amino acids between M2 and M3 aligned such that most of the hydrophilic residues and hydrophobic residues were located on different sides of the predicted alpha helix. However, one semipolar residue, serine (S), and one strongly hydrophilic residue, asparagine (S), were located amongst hydrophobic residues.

While the clustering patterns of amphipathic and hydrophobic amino acids in the predicted alpha helix do not show precise symmetry in their arrangement, it is possible that these proteins form a helical structure that in some way contributes to an important, conserved function. As previously mentioned for this region in Family 5, it is possible that this region contributes in some way to the movement of the TGES loop, which may undergo notable conformational changes as it transitions between phosphorylated and dephosphorylated states (Anthonisen, et al.; 2006).

3.7: 16S rRNA Analysis

As previously described the 16S ribosomal RNA sequences used for analysis of the prokaryotic genera present in Family 6 was obtained from the NCBI database. After removing unclassified organisms, a total of different 136 genera were identified in Family 6. The 16S rRNA data corresponding to a single organism belonging to each genus were collected for analysis (see above). The nucleotide sequences were aligned using the program ClustalX and a phylogenetic tree was created (Figure 7). Analysis of this tree indicated that organisms belonging to different genera generally clustered by phylogenetic group and by nucleotide sequence length. All ten archaeal genera belonged to the phylogenetic group Euryarchaeota and clustered together, as expected. *Blastopirellula* and *Halothermothrix*, two bacterial genera from the phylogenetic groups Planctomycetes and Firmicutes, respectively, appeared to cluster more closely to these ten archaeal genera than to any other bacterial genus present amongst the

Heavy metal P-type ATPase Family protein sequences. These two genera are among the more ancient bacterial genera and, as such, it is not unusual that they were discovered to branch at a point near the center of the 16S rRNA tree from the archaeal sequences.

Another point of interest is that the genera *Treponema, Halothermothrix, Mariprofundus, Fusobacterium* and *Symbiobacterium* are not found amongst other genera belonging to the same phylogenetic group as themselves. *Treponema*, which belongs to the phylogenetic group Spirochaetes, is found in the cluster containing Chlamydiae and Cyanobacteria, but appears to exhibit a more distant phylogenetic relationship to its neighboring genera. *Halothermothrix* belongs to the phylogenetic group Firmicutes, but does not cluster with the other genera from Firmicutes. Instead, it clusters closely with Blastospirellula, from Planctomycetes, and together they cluster very closely to the archaeal genera. *Mariprofundu*s, presently described as only an unclassified proteobacterium, appears to be distinct from all of the other surrounding genera, which are from α-proteobacteria and **Δ-**proteobacteria. Similarly, *Fusobacterium*, from the phylogenetic group Fusobacteria, notably varies phylogenetically from all of the other genera it clusters with, which almost entirely consist of genera belonging to the phylogenetic group Firmicutes. The only other genus in this cluster that is not from Firmicutes is *Symbiobacterium*, which belongs instead to the phylogenetic group Actinobacteria. Unlike *Fusobacterium*, however, *Symbiobacterium* clusters closely with its neighboring genera (Figure 7).

Chapter 4: Conserved Motifs in Copper P-Type ATPases

Conserved Motifs in Copper P-Type ATPases

Motif Introduction

A notable characteristic of P-type ATPases is their strong conservation of nine distinct sequence motifs (Møller J.V., et al., 1995). Their order of appearance, in progression from the N-terminus to the C-terminus of each sequence, is as follows: PGD, PAD, TGES, PEGL, DKTGTLT, KGAPE, DPPR, MVTGD, and VAVTGDGVNDSPALKKADIGVAM. The first three of these motifs are located in the small cytosolic loop between TMS2 and TMS3, known as 'Region B.' The residues within this loop, particularly those in Motif 3, may enhance the stability of this region, thereby providing more favorable reaction kinetics for the enzyme's transition between its E1 and E2 states.

Motif 4, PEGL, was consistently located within TMS4 of the sequences from Families 5 and 6, and is believed to contribute to energy transduction. Motifs 5-8 are found within a large, catalytically active cytosolic loop, known as 'Domain C,' which immediately followed TMS 4. Motif 5, DKTGTLT, contains a phosphorylatable aspartate (D) residue and exhibited considerable conservation of its residues in both Families 5 and 6. This aspartate residue is phosphorylated during enzyme cycling. The remaining residues of motif 5 may play roles in catalysis and help maintain the structure of Domain C. Unlike motif 5, motif 6, KGAPE, was poorly conserved in Families 5 and 6, with the exception of its glycine residue. Despite its poor conservation, motif 6 is thought to play a role in ATP binding. Motif 7, DPPR,

exhibited the best conservation of its first and last residues in Families 5 and 6. Interestingly, these two residues are thought to be important for phosphorylation of the ATPase. The final three residues in motif 8, MVTGD, are also believed to play a critical role in phosphorylation of the enzyme, and were notably well-conserved in both Families 5 and 6. Lastly, Motif 9, VAVTGDGVNDSPALKKADIGVAM, forms part of a flexible hinge region that joins Domain C with the C-terminal domain (Møller J.V., et al, 1995). This 'hinge' region, or 'Region J,' helps provide the flexibility needed for conformational changes that occur during ion translocation (Møller J.V., et al., 1995).

Each motif was individually identified and examined through methodical, manual analyses of all twenty clusters in their respective multiple alignments. As expected, the multiple alignments indicated that all of the sequences in each cluster aligned at these nine motifs, with at least some degree of conservation at one or more residues within the motif. All sequence motifs can bee seen, along with their location in their corresponding multiple alignments and their degree of conservation, in Table 3.

4.1: Motif 1 (PGD)

This motif most commonly presented itself not as PGD, but as PGE, with the second residue, glycine (G), fully conserved in all twenty groups. Proline (P) was fully or partially conserved in seventeen of the twenty clusters, as designated in Table 3 by

an asterix or by one or two dots above the residue, respectively. However, neither it nor any conservatively substituted residue was present in clusters 1, 3, and 17. Aspartic acid (D) was only partially conserved in clusters 8, 14, and 16. Its most frequent substitute ion, glutamate (E), was fully or partially conserved in eleven clusters. Lack of a conserved residue in the third position of this motif was observed in clusters 3, 4, 5, 10, 12, and 13.

4.2: Motif 2 (PAD)

Motif 2, like Motif 1, is located between TMS2 and TMS3, and was consistently located within three residues of the end of Motif 1. The most strongly conserved residue in this motif was aspartic acid (D), which was completely conserved throughout all twenty clusters. Proline (P) was partially or completely conserved in fifteen of the twenty clusters, but was not conserved in clusters 2, 3, 5, 13, or 15. Alanine (A) was not partially or fully conserved in any of the twenty clusters. Instead, the most common substituted residue at the second position in this motif was Valine, which was fully or partially conserved in eleven clusters. Threonine was also partially conserved in two of the twenty clusters in lieu of alanine.

4.3: Motif 3 (TGES)

Motif 3, like Motifs 1 and 2, was consistently located between TMS2 and TMS3. Both Threonine (T) and glycine (G) were fully or partially conserved in all twenty clusters. Glutamate (E) was fully or partially conserved in all but cluster #2, where it was still expressed amongst several of the sequences in this cluster. Serine was fully or partially conserved in all but clusters 1, 3, 7, 8, 9, 17, and 18. In its place, proline was conserved in clusters 7, 8, and 9, but no replacement residues were conserved in clusters 1, 3, 17, or 18. Previous mutagenesis studies have investigated the conservation of this motif in the Ca^{2+}-ATPase (Anthonisen, et al.; 2006). Crystal structures were used to assess the effects of residue size, polarity, and charge on the reaction kinetics involved in the transition from the E1 to E2 states. It was noted that glutamate (Glu) was incredibly well conserved in this motif in all P-type ATPases. This was also found to be generally true amongst the P-type ATPases examined in this thesis, but it was found that the level of this residue's conservation falls somewhat short of Anthonisen et al.'s description of it as a "universally conserved" residue.

Concerns about actual levels of the conservation of Glu aside, analyses of the functional data produced from the mutagenesis studies indicated that Glu played an imperative role in the dephosphorylation of E2P. It was demonstrated that the length, the hydrogen bonding potential, and the negatively charged carboxylate group of Glu are all integral components of its hypothesized ability to bind to and to activate the water molecule that attacks the phosphoryl group during this process. Various mutated forms of this motif were created, where different residues were substituted for Glu. While examination of their respective crystal structures indicated that some residue

substitutions still permitted the continuation of the reaction cycle, and therefore produced seemingly viable alternatives to glutamate, all residue substitutions appeared to negatively affect the reaction kinetics at one transition point or another. Consequently, in light of the all-around desirability of Glu in motif 3 from a kinetic standpoint and evidence that Glu is essential to the catalysis in E2/E2P, it has been proposed that during the $Ca_2E1P \rightarrow E2P$ transition Glu is actively involved in the insertion of the TGES loop into the catalytic site. It also appears that the movement of this loop into the catalytic site may permit lumenal Ca^{2+} sites to open, thus permitting ion translocation.

4.4: Motif 4 (PEGL)

Motif 4 was consistently located partially within TMS4 in all twenty clusters, and most commonly took the form PCAL as opposed to the predicted motif sequence. Proline (P) was completely conserved in all twenty clusters, whereas glutamate (E) was not conserved in any cluster. In place of glutamate, cysteine was completely conserved in all but clusters 4, 5, and 15. Although no residues were fully or partially conserved in cluster 15 at the second position in the motif, histidine was completely conserved in cluster 4 and aspartic acid was completely conserved in cluster 5. Glycine was not conserved at all in the third position of motif 4. However, it was predominantly replaced by alanine, which was fully or completely conserved in all but clusters 15 and 17, where no residues exhibited any distinguishable degree of

conservation. Leucine was completely conserved in all twenty clusters except cluster 15, where it still exhibited partial conservation.

4.5: Motif 5 (DKTGTLT)

In Motif 5, which was located between TMS4 and TMS5, the initial four residues, DKTG, were partially or completely conserved in all groups except in cluster 13. It appears that in cluster 13, the aspartic acid residue was separated from the proceeding three residues by the insertion of other residues in another sequence. This insertion seems to have caused the sequences in cluster 13 to misalign at the point of this motif. The threonine residue located in the fifth position was completely conserved in all but group 3, where no residue was conserved at this location, and the threonine residue located at the end of the motif was fully conserved in all twenty clusters. Leucine was fully or partially conserved in clusters 1, 3, 4, 5, 6, 7, 10, and 11. In its place, lysine (K) was partially or fully conserved only in clusters 8 and 9, valine was fully or partially conserved in clusters 2, and 19, and isoleucine was fully or partially conserved in clusters 11, 12, and 18.

4.6: Motif 6 (KGAPE)

Motif 6 appeared to be located immediately following the end of Motif 5. With the exception of its glycine residue, which was fully or partially conserved in all but

cluster #11, Motif 6 was generally poorly conserved in its entirety. While this poor conservation of residues was somewhat disconcerting at first, comparison of the corresponding data obtained from the SOPMA analyses, (Figure 3.C.1-.20) suggested that the secondary structure present within Motif 6 was still fairly well conserved (see above).

4.7: Motif 7 (DPPR)

Motif 7, located between TMS4 and TMS5, displayed the strongest conservation at its first and last residues. Aspartic acid was fully conserved in all twenty clusters. By contrast, proline in the second residue position was only partially conserved in cluster 7, and proline in the third position was only partially conserved in clusters 12 and 16. Substituting for proline in the second position, alanine exhibited partial conservation in cluster 5 and glutamine displayed partial conservation in cluster 18. In lieu of proline in the third residue position, isoleucine was partially conserved in clusters 4, 9, 10, and 18, valine was partially conserved in clusters 6 and 11, threonine was completely conserved in cluster 7, and leucine was partially conserved in cluster 17. Arginine (R) was fully or partially conserved in clusters 1, 3, 4, 5, 14, 16, and 17. In place of arginine, Lysine was partially or fully conserved in clusters 2, 6, 7, 8, 10, 11, 12, 13, 15, 18, 19, 20. No residues were conserved in this position in cluster 9.

4.8: Motif 8 (MVTGD)

Motif 8 was generally located approximately twenty residues away from the end of Motif 7. The final two residues in this motif, glycine and aspartic acid, were completely conserved in all twenty clusters. The first residue, methionine, was partially or completely conserved in all but clusters 2, 3, 14, and 20 where, instead, it was replaced by a partially conserved leucine, clusters 2, 14, and 20, or a partially conserved isoleucine, cluster 3. Valine was partially or fully conserved in groups 7, 8, 13, and 16. Most commonly, valine was replaced by leucine, which was fully or partially conserved in clusters 4, 7, 9, 10, 12, 14, 15, 17, 18, and 20, or by isoleucine, which was fully or partially conserved in clusters 1 and 11. No residues were conserved in this final position in clusters 2, 3, 6, or 19.

4.9: Motif 9 (VAVTGDGVNDSPALKKADIGVAM)

Motif 9 exhibited a lot of sequence variation amongst the sequences in each of the twenty clusters of Family 5, but it was completely conserved at the residues "GDG"s in all but groups 17, 19, and 20, where it was partially conserved. Additionally, strong conservation at the residues "PALA," either partially or fully, was observed. Similarly, the last four residues in this motif, "GVAM," were generally well conserved, either partially or fully, although usually one or more of these residues deviated from the expected amino acid.

Chapter 5: Conserved Motifs in Heavy Metal P-Type ATPases

Conserved Motifs in Heavy Metal P-Type ATPases

Motif Introduction

The proteins of the Heavy metal P-type ATPase Family, like all known P-type ATPases, have nine known well-conserved sequence motifs (Møller J.V., et al., 1995). As described above for the proteins of Family 5, these motifs can be found in sequential progression from the N-terminus to the C-terminus of a given sequence as follows: PGD, PAD, TGES, PEGL, DKTGTLT, KGAPE, DPPR, MVTGD, and VAVTGDGVNDSPALKKADIGVAM.

Each motif was individually identified and examined in the multiple alignments of all seventeen of the groups representing the Heavy metal P-type ATPases using methods identical to those described above for the Copper P-type ATPase Family. As expected, the multiple alignments indicated that all of the sequences in each group aligned at these nine motifs, with at least some degree of conservation at one or more residues within each motif. The sequence motifs for all seventeen clusters, along with their corresponding data, can be found in Table 6.

5.1: Motif 1 (PGD)

This motif most commonly presented itself not as the residues PGD, but as PGE, with the second residue, glycine (G), being the most strongly conserved of the

three and fully conserved in all but cluster 10, where no residue was significantly conserved. Proline (P) was partially (as designated by one or two dots) or fully conserved in nine clusters (as designated by an asterix on the multiple alignment). In place of proline, alanine was partially conserved in cluster 6, and no residues were significantly conserved at this location in clusters 3, 5, 7, 8, 9, and 16. Aspartic acid was fully or partially conserved in clusters 3, 9, and 14. In substitution for aspartic acid, glutamine (E) was partially or fully conserved in clusters 4, 6, 11, 12, 13, 16, and 17, whereas no residues displayed significant levels of conservation at this location in clusters 1, 2, 5, 7, 8, 10, and 15.

5.2: Motif 2 (PAD)

As expected, Motif 2 was identified between TMS2 and TMS3, and was almost always found within three residues of the end of Motif 1. As observed in Family 5, the most strongly conserved residue in this motif was aspartic acid, which was fully or completely conserved throughout all seventeen clusters. Proline was partially or completely conserved in nine of the seventeen clusters. However, in place of proline, alanine was seen as either fully or partially conserved in clusters 11, 12, and 15. Alanine, as the residue in the second position in this motif, was only partially conserved in cluster 17. Several different residues were substituted in its place throughout the remainder of the groups in Family 6. Threonine was fully conserved in clusters 11 and 12. Valine was fully or partially conserved in clusters 7 and 16.

Leucine was fully or partially conserved in clusters 3 and 14. Methionine was partially conserved in cluster 15 and isoleucine was partially conserved in cluster 6.

5.3: Motif 3 (TGES)

As observed in Family 5, Motif 3 of Family 6, which was located between TMS2 and TMS3, exhibited high levels of conservation. Threonine, glycine, and glutamate are partially or fully conserved in all seventeen groups. Serine was partially or completely conserved in all but cluster 8, in which no substitute residue exhibited any conservation, and cluster 17, in which a proline residue was substituted and displayed partial conservation.

5.4: Motif 4 (PEGL)

As was seen in Family 5, Motif 4 was located at the edge and partially inside of TMS4 in Family 6. Likewise, a notable conservation of the residues "PCAL" was observed, as opposed to the expected residues "PEGL." Proline was fully or partially conserved in all seventeen clusters. Although no glutamate was seen in the second position, cysteine was predominantly observed in its place. Cysteine, as a substitute residue, was fully or partially conserved in all but cluster 7, which does not display significant residue conservation at this location. Alanine was partially or fully conserved in all but cluster 6, which displayed a partially conserved glycine residue.

Leucine was fully or partially conserved in all but cluster 13, which exhibited a partially conserved phenylalanine residue instead.

5.5: Motif 5 (DKTGTLT)

Motif 5 was fully or partially conserved in all seventeen clusters at residues DKT and at the threonine residue immediately following glycine. Glycine was fully or partially conserved in all but cluster 4, where no residue exhibited significant conservation at this location. Leucine was partially or fully conserved in clusters 3-5, and clusters 7-17. In place of leucine, aspartic acid was partially or fully conserved at this location in clusters 1 and 12 and isoleucine was partially conserved at this location in cluster 6.

5.6: Motif 6 (KGAPE)

Motif 6 exhibited poor conservation except at its glycine residue. This residue was partially or fully conserved in all but clusters 1 and 12. In its place a histidine residue was partially conserved in cluster1 and an asparagine residue was partially conserved in cluster 12. Despite the presence of these unique residue substitutions and poor conservation of residues in general, examination of corresponding SOPMA analysis data confirmed that, for the most part, the residues associated with Motif 6 exhibited fairly consistent secondary structure patterns (Figure 6.C.1-.17).

5.7: Motif 7 (DPPR)

The last residue of Motif 7 was generally located approximately twenty residues from the first residue of Motif 8. Although both proline residues exhibited fairly poor levels of conservation, both the first and last residue exhibited better conservation. Aspartic acid was partially or fully conserved in all seventeen clusters. Although it exhibited somewhat weaker conservation than aspartic acid, arginine was partially or fully conserved in all but clusters 5 and 8, in which no residues exhibited appreciable conservation.

5.8: Motif 8 (MVTGD)

In Motif 8 the last two residues, glycine and aspartic acid, were partially or fully conserved in all seventeen clusters. Methionine was partially or fully conserved in all but cluster 17, in which it was replaced by a leucine residue. Threonine was partially or fully conserved in all but cluster 9, where a partially conserved isoleucine residue has been substituted. Leucine was partially or completely conserved only in clusters 5 and 8. A partially conserved isoleucine residue was substituted for leucine at this position in cluster 9, and either a partially or fully conserved leucine residue was substituted at this location in clusters 1-4, 6, and clusters 10-17.

5.9: Motif 9 (VAVTGDGVNDSPALKKADIGVAM)

Considerable sequence variation was observed for Motif 9. However, the residues “GDG” were completely conserved in all seventeen clusters. Again, strong residue conservation, either partial or fully, was observed at “PALA” and “GVAM.” However, usually one or more of these residues varied from those expected for this motif. For example, instead of seeing the expected residues “GVAM”, the most commonly observed residues for that portion of Motif 9 were “GIAM”, which were fully or partially conserved in all but clusters 2 and 17. In cluster 2 this portion of Motif 9 showed partial conservation of residues “GFAM,” and cluster 17 showed partial conservation of residues “GLAV.”

Chapter 6: Homology Analyses of Copper and Heavy Metal P-type ATPases

6.1: Homology Analyses of Copper and Heavy Metal P-type ATPases

The homology of the sequences representing the Copper P-type ATPase and Heavy Metal P-type ATPases was examined, one cluster at a time. Analysis of the phylogenetic relationships of the individual sequences within each phylogenetic group revealed the formation of several sub-clusters. The branching distances and relative similarity of the genera of the protein sequences within these sub-clusters were compared using their corresponding 16S rRNA and protein phylogenetic trees (Figures 2.A-2.B and 4, and Figures 5.A-5.B and 7, respectively). Sequences that clustered closely in both the 16S rRNA and protein phylogenetic trees were generally predicted to be orthologous to one another. However, some sequences that were located near each other in the protein phylogenetic tree actually belonged to genera that were found either in adjacent or more distant clusters in the 16S rRNA tree. These sequences were not predicted to be orthologous to one another, and, in some instances, may represent instances of horizontal gene transfer.

Homology Analyses for Copper P-type ATPases

Cluster 1 was made up of two archaeal and two bacterial protein sequences. The two archaeal sequences, Ape1 and Pae2 (from *Aeropyrum pernix* and *Pyrobaculum aerophilium*), clustered closely in both the 16S rRNA tree and the protein phylogenetic tree, and therefore could be orthologs. While the bacterial sequences Tma1 (from *Thermotoga maritima)* and Wsu1 (from *Wolinella*

succinogenes) clustered closely to each other, as well as close to the archaeal sequences, in the protein phylogenetic tree, their genera were located in very distant clusters from one another in the 16S rRNA tree. As such, the clustering of Wsu1 with these other three sequences may represent an instance of horizontal gene transfer. Additionally, *Thermotoga* was located in a cluster adjacent to the archaeal cluster containing *Aeropyrum* and *Pyrobaculum* in the 16S rRNA tree. Although these two clusters branch close to the center of the tree, the proximity of *Thermotoga* to these archaeal genera suggests that a horizontal gene transfer event may have taken place.

A mixture of organisms from *Tropheryma, Brevibacterium, Propionibacterium, Arthrobacter, Corynebacterium, Mycobacterium, Frankia, Streptomyces, Nocardia, Thermobifda, Kinecoccus, and Nocardioides* are found in cluster 2. These bacteria were all Actinobacteria, and clustered closely in both the 16S rRNA and protein phylogenetic trees, thereby making them possible orthologs. While some organisms from the same genus were found adjacent to each other in the protein phylogenetic tree, in many cases they were joined or separated by organisms from other genera. For example, Pac2 (from *Propionibacterium acnes*) clustered closer to Mtu3 (from *Mycobacterium tuberculosis*), Mle1 (from *Mycobacterium leprae*), and Mtu2 (from *Mycobacterium tuberculosis*), as opposed to with Pac1 (from *Propionibacterim acnes*).

Cluster 3 contained bacterial sequences from Acidobacteria, Δ-proteobacteria, Chlorobi, γ-proteobacteria, β-proteobacteria, ε-proteobacteria, and one unclassified proteobacterium. Sus2 (from *Solibacter usitatus*), which was distant from all other

sequences in both the 16S rRNA and protein phylogenetic trees, is not likely to be orthologous to any of the other sequences in cluster 3. Cph3 (from *Chlorobium phaeobacteroides*), Cli2 (from *Chlorobium limicola*), and Pph2 (from *Pelodictyon phaeoclathratiforme*), all Chlorobi, clustered together with Msp4 (an unclassified proteobacterium) in both the 16S rRNA and protein phylogenetic trees and are likely orthologous sequences. Pae3, Mfl2, and Tde5 (from *Pseudomonas aeruginosa, Methylobacillus flagellatus,* and *Thiobacillus dentrificans*, respectively), clustered adjacent to one another in the protein phylogentic tree, but were not clustered tightly together. Similarly, in the 16S rRNA tree, the genera that these sequences belong to appeared to cluster together, but exhibited early branching, and so they may or may not be orthologs. Asp6 and Dar4 (from *Azoarcus* sp. EbN1 and *Dechloromonas aromatica*, respectively) were found next to these three proteins, but were distant enough in both the 16S rRNA and protein phylogenetic trees that they are not likely to be orthologs. However, they did cluster closely enough to one another in these two trees that they themselves could be orthologs. Sde2, Sam2, Sfr2, Sba2, Ppr4, and Son2 (from *Shewanella denitrificans, Shewanella amazonensis, Shewanella frigidimarina, Shewanella baltica,* and *Photobacterium profundum, Shewanella oneidensis*, respectively) all clustered closely together in both the 16S rRNA and protein phylogenetic trees, and are likely to be orthologous sequences. Likewise, Vch4, Vvu2, Vpa2, Vsp2 (from *Vibrio cholerae, Vibrio vulnificus, Vibrio parahaemolyticus,* and *Vibrio* sp. Ex25, respectively) were found clustering closely in both the 16S rRNA and protein phylogenetic trees, as were Ngo2 and Nme3 (from *Neisseria gonorrhoeae* and

Neisseria menigitidis, respectively). The proximity of these four *Vibrio* sequences suggests that they are orthologous, as does the proximity of these two *Neisseria* sequences. Lastly, the proteins Tde6, Cla1, and Cup1 (from *Thiomicrospira denitrificans, Campylobacter lari,* and *Campylobacter upsaliensis*, respectively) were found clustered closely together in both the 16S rRNA and protein phylogenetic trees, and so are likely to be orthologs.

Cluster 4 contained several archaeal sequences and a mixture of bacterial sequences from Firmicutes, Aquificae, β-proteobacteria, Δ-proteobacteria, Deinococci, and Chloroflexi. Several of the sequences found clustering together belonged to the same genus and are very likely to be orthologs. These sequences included Efa5, Efa4, Ehi1 (from *Enterococcus faecalis, Enterococcus faecium,* and *Enterococcus hirae,* respectively), as well as Sau2 and Sep2 (from *Staphylococcus aureus* and *Staphylococcus epidemidis*) and Lpl2 and Lca1 (from *Lactobacillus plantarum* and *Lactobacillus casei*). These sequences, along with Ppe1 (from *Pedicoccus pentosacaeus*), clustered next to one another in both the 16S rRNA and protein phylogenetic trees. While it is apparent that the genera these sequences belong to are all found within the same cluster in the 16S rRNA tree, the branching distances observed in both trees suggests that, collectively, these sequences may or may not be orthologs. Aae1, Oih2, Mth3, Tde4, Bba2, Afu3, Mba2, Mac2, Mma4 (from *Aquifex aeolicus, Oceanobacillus iheyensis, Methanothermobacter thermautotrophicus, Thiobacillus denitrificans, Bdellovibrio bacteriovorus, Archaeoglobus fulgidus, Methanosarcina barkeri,* and *Methanosarcina mazei*, respectively), also all grouped

together in cluster 4, albeit at some distance from one another. Interestingly, these sequences were a mix of bacteria and archaea, and did not all cluster with other sequences from the same phylogenetic domain. For instance, the sequence Mth3 does not cluster closely in the phylogenetic tree with the other archaeal sequences in its sub-cluster, Afu3, Mba2, and Mac2, and so it may not orthologous to any of those sequences. Similarly, the bacterial sequences Bba2, Tde4, Oih2, and Aae1 do not all cluster adjacently in the protein phylogenetic tree and are very distant from one another in the 16S rRNA phylogenetic tree, making them unlikely to be orthologs and possibly even examples of horizontal gene transfer. Lastly, the sequences Tth2, Cau2, and Hma3 (from *Thermus thermophilus, Chloroflexus aurantiacus,* and *Haloarcula marismortui*, respectively) clustered together in the protein phylogenetic tree, and *Thermus* and *Chloroflexus* were found together in a cluster adjacent to *Haloarcula* in the 16S rRNA tree. However, instead of clustering more closely to the other bacterial sequence, Cau2 was closer to the archaeal sequence Hma3. Tth2 and Cau2 belong to genera that are amongst the more ancient bacterial sequences and branch very close to the center of the 16S rRNA tree from the cluster that contains the archaeal sequences. As these two sequences belong to two separate domains they may represent an occasion of horizontal gene transfer

The sequences found in cluster 6 were primarily from α–proteobacteria, β-proteobacteria, and γ–proteobacteria, but it also contained one sequence from Planctomycetes. Cvi2, Xax1, and Asp4 (from *Chromobacterium violaceum*, *Xanthomonas axonopdis* and *Azoarcus* sp. EbN1, respectively) branched distantly

from each other and from the remaining sequences in cluster 6 in both the protein phylogenetic and the 16S rRNA trees. By contrast, the sequences Bvi2, Bvi3, Neu6, Neu4, Neu5, Neu7, Rfe2, Rge3, Psp2, and Rfe3 (from *Dechloromonas aromatica, Ralstonia metallidurans, Burkholderia vietnamiensis, Burkholderia vietnamiensis, Nitrosomonas eutropha, Nitrosomonas eutropha, Nitrosomonas eutropha, Nitrosomonas europaea, Rhodoferax ferridreducens, Rubrivivax gelatinosus, Polaromonas* sp. JS666, and *Rhodoferax*, respectively) all clustered closely together in both the 16S rRNA and protein phylogenetic trees and could all be orthologous to one another. While not all of the sequences clustered most closely to others from the same genera, the proximity of all of their genera in the 16S rRNA tree suggests that they all cluster close enough to one another to exchange genetic material. Mde2, Spo3, Rle1, Mlo1, Msp3, Tde3, Sme4, and Rpa1 (from *Microbulbifer degradans, Silicibacter pomeroyi, Rhizobium leguminosarum, Mesorhizobium loti, Mesorhizobium* sp. BNC1, *Thiomicrospira denitrificans, Sinorhizobium meliloti*, and *Rhodopseudomonas palustris*, respectively) also clustered closely in both the 16S rRNA and protein phylogenetic trees, with the exceptions of Mde2 and Tde3. While Tde3 was located more distantly than any of the other proteins in this cluster in both the 16S rRNA and protein phylogenetic trees, Mde2 was close to Spo3 in the protein phylogenetic tree, but very distant from any of the other genera from these sequences in the 16S rRNA tree. These data suggest that neither Tde3 nor Mde2 are orthologous to the other sequences they cluster with, or to each other, and that they each represent instances of horizontal gene transfer.

Also found in cluster 6 were the sequences Bja1, Nha1, Nha2, Nha3, Nwi1, Nar1, Sal1, Nar2, Sal2, Sal3 (from *Bradyrhizobium japonicum, Nitrobacter hamburgensis, Nitrobacter hamburgensis, Nitrobacter hamburgensis, Nitrobacter winofradskyi, Novosphingobium aromaticivorans, Novosphingobium aromaticivorans, Sphingopyxis alaskensis, Sphingopyxis alaskensis,* and, *Sphingopyxis alaskensis,* respectively). These protein sequences exhibited similar distance patterns in both the 16S rRNA and protein phylogenetic trees and clustered relatively closely to Rle1, Mlo1, Msp3, Tde3, Sme4, and Rpa1, possibly indicating that all of these sequences are orthologous. Sty2, Kpn1, Sma2, (from *Salmonella typhimurium, Klebsiella pneumoniae*, and *Serratia marcescens*, respectively*)* and Lpn4, Lpn3, Lpn1, Lpn2 (all from *Legionella pneumophila*) were also found in cluster 6. These seven sequences were separated into two sub-clusters with the sequences Sty2, Kpn1, and Sma2 clustered together in one sub-cluster, and Lpn4, Lpn3, Lpn1, and Lpn2 clustered together in the other. The sequences within each of these sub-clusters were found close together in both the 16S rRNA and protein phylogenetic trees, and are likely to be orthologous to one another. However, these two sub-clusters of cluster 6 were distant from each other in both trees, and, collectively, are not likely to be orthologous. Lastly, the sequences Rba1, Eli1, Par2, Pcr2, Bps1, Bfu2, Bvi4 (from *Rhodopirellula baltica, Erythrobacter litoralis, Psychrobacter arcticus, Psychrobacter cryohalolentis, Burkholderia pseudomallei, Burkholderia fungorum*, and *Burkholderia vietnamiensis*, respectively) were also found within cluster 6. The two sequences from *Psychrobacter* clustered close to one another and distantly from

the other sequences surrounding it in the protein phylogenetic tree, as did the three sequences from *Burkholderia*. While is likely that the sequences from the same genera are orthologous to one another, collectively, the four genera represented by these seven sequences were not very close to each other in the 16S rRNA tree, as they were either in adjacent clusters or branched at very distant points within the same cluster, and so they may or may not all be orthologous to each other.

Clusters 5, 7 and 18 contained far fewer sequences than most of the other clusters. Cluster 5 contained five bacterial sequences from Deinococci, Actinobacteria, and α–proteobacteria. Dge4, from *Deinococcus geothermalis*, and Nha4, from *Nitrobacter hamburgensis*, branched distantly in the protein phylogenetic tree and were found in different clusters from each other and from the remaining protein sequences in the 16S rRNA tree. As such, it is unlikely that these two sequences are orthologous to each other or to any of the remaining three sequences in cluster 5. By contrast, Nsp8, from *Nocardioides* sp. JS614, Nfa2 and Nfa3 (both from *Nocardia farcinica*) clustered closely together in the protein phylogenetic tree and exhibited similar clustering patterns in the 16S rRNA tree as well. Consequently, it is likely that all three sequences are orthologous to one another. Only one sequence was found in cluster 7, Msp2. This sequence was an unclassified proteobacterium from *Magnetococcus* sp. MC-1. It was the only sequence from its genera out of all of the sequences representing the Copper P-type ATPases, and was very distant from the other genera it clustered with in the 16S rRNA tree. Lastly, cluster 18 contained three sequences, all from Firmicutes, Lla1, Efa3, and Ehi1 (from *Lactococcus lactis,*

Enterococcus faecium, and *Enterococcus hirae*, respectively). Efa3 and Ehi1 were from the same genus, and Lla1 clustered closely to these sequences in the protein phylogenetic tree and 16S rRNA tree, thus making it likely that all three sequences are orthologs.

With the exception of one uncultured bacterium, the sequences found in cluster 8 were from α–proteobacteria, β-proteobacteria, γ–proteobacteria, Cyanobacteria and Deinococci. Ilo1, from *Idiomarina loihiensis*, was very distant from other sequences within cluster 8, but was located in the same cluster in the 16S rRNA tree as Sty1 and Pmi1. However, Ilo1, Sty1, and Pmi1 (from *Salmonella typhimurium* and *Proteus mirabilis*, respectively) were not clustered close to each other in the protein phylogenetic tree, and are not likely to be orthologs. Neu2, Msp1, New3, Sty1, Bbr1, and Avi1 (from *Nitrosomonas eutropha, Mesorhizobium* sp. BNC1, *Nitrosomonas eutropha, Salmonella typhimurium, Bordetella bronchiseptica,* and *Azotobacter vinelandii*, respectively) were also found in cluster 8. Although these sequences were all found relatively near one another in the protein phylogenetic tree, only *Nitrosomonas* and *Bordetella* were found within the same cluster in the 16S rRNA tree, and therefore could possibly be orthologs. The genera of the remaining sequences are very distant from each other in the 16S rRNA tree, and these sequences could be examples of horizontal gene transfer. Additionally, Atu1, Bme1, Sme3, Sme1, Sme2, Atu2, and bac2 (from *Agrobacterium tumefaciens, Brucella melitensis, Sinorhizobium meliloti, Sinorhizobium medicae, Sinorhizobium meliloti, Agrobacterium tumefaciens,* and an uncultured bacterium, respectively) were found in cluster 8. While the two

sequences from *Agrobacterium* did not cluster adjacently to one another in the protein phylogenetic tree, it is still likely that they are orthologs and that their distance from each other can be attributed to genetic exchange between other nearby sequences. Collectively, *Agrobacterium, Salmonella,* and *Brucella* cluster very closely within the same cluster in the 16S rRNA tree, and so it is likely that Atu1, Bme1, Sme3, Sme1, Sme2, Atu2 are orthologous to each other. Since the sequence bac2 was from an uncultured bacterium, no 16S rRNA sequence could be entered into the 16S rRNA tree to compare it to the other genera representing the sequences of the Copper P-type ATPases. Subsequently, it cannot be determined at this time whether or not bac2 is orthologous to the sequences surrounding it, or if it is from a very distant genus and may be an example of horizontal gene transfer.

Also found in cluster 8 were the protein sequences Pde1, Jsp1, Ssp2, Rsp1, and Spo2 (from *Paracoccus denitrificans, Jannaschia* sp. CCS1, *Silicibacter* sp. TM1040, *Rhodobacter sphaeroides,* and *Silicibacter pomeryoi*, respectively). The clustering patterns of these sequences were similar in both the 16S rRNA and protein phylogenetic trees, and as they all clustered closely it is quite possible that these sequences are all orthologs. Lastly, the sequences Tth1, Dra1, and Gvi1 (from *Thermus thermophilus, Deinococcus radiodurans,* and *Gloeobacter violaceus,* respectively), as well as Dge3, Dge1, and Dge2 (all from *Deinococcus geothermalis*) were found in cluster 8. Although Dra1 does not cluster amongst the other sequences in cluster 8 from *Deinococcus,* it is likely that it is still orthologous to these sequences, and that its distance from them can be attributed to an exchange of genetic material

with other nearby sequences. Also, *Deinococcus* and *Thermus* were next to each other in the 16S rRNA tree, and therefore the sequences in cluster 8 belonging to these genera could be orthologs. By contrast, *Gloeobacter* was located in a cluster of the 16S rRNA tree that was adjacent to the one containing *Deinococcus* and *Thermus*, and so Gvi1 is most likely not orthologous to the other sequences in cluster 8.

The sequences found in cluster 9 were exclusively from γ–proteobacteria. The sequences Apl1, Aau1, and Msu1 (from *Actinobacillus pleuropneumoniae, Actinobacillus succinogenes,* and *Mannheimia succiniciproducens,* respectively) clustered together in both the 16S rRNA and protein phylogenetic trees and could be orthologous. While the sequences Aau1 and Msu1 clustered more closely in the protein phylogenetic tree than Aau1 clustered with Apl1, it is possible that Msu1 and Aau1 were able to exchange some of their genetic material, thereby causing them to cluster more closely together than expected. Eca1, Plu1, Sma1, Ype1, Eco1, and Sen1 (from *Erwinia carotovora, Photorhabdus luminescens, Serratia marcescens, Yersinia pestis, Escherichia coli,* and *Salmonella enterica*, respectively) clustered closely to one another in both the 16S rRNA and protein phylogenetic trees, and therefore could all be orthologous. These sequences were distant enough from Apl2, Aau1, and Msu1, the first three sequences examined in cluster 9, that they were found in a separate cluster in the 16S rRNA tree, and, as such, are not likely to be orthologous to them. Ppr3, from *Photobacterium profundum*, was located in a sub-cluster of cluster 9 along with Vfi1, Vch3, Vch1, Vch2, Vvu1, Vpa1,Vsp1 (all from *Vibrio*), as well as with Sba1 and Son1 (from *Shewanella baltica* and *Shewanella oneidensis*). While the

sequences from the same genera clustered the closest together, all of the sequences clustered closely in both the 16S rRNA and protein phylogenetic trees, and could all be orthologs. Lastly, the sequences Cbu1, Mca1, Csa1, Ilo2, and Sfr1 (from *Coxiella burnetti, Methylococcus capsulatus, Chromohalobacter salexigens, Idiomarina loihiensis,* and *Shewanella frigidimarina*, respectively) were also found together in cluster 9. Cbu1, Mca1, Csa1 were more distant from Ilo2, and Sfr1, and were located within the same 16S rRNA cluster, making them possible orthologs. Ilo2 and Sfr2 clustered closely together, and were located near each other in a cluster of the 16S rRNA tree adjacent to the one containing Cbu1, Mca1, and Csa1. Thus, it is possible that Ilo2 and Sfr2 are orthologous, but it is unlikely that these five sequences are all orthologous to one another.

Cluster 10 contained bacterial sequences from Chloroflexi, Actinobacteria, Δ-proteobacteria, Acidobacteria, Cyanobacteria, Bacteroidetes, and Firmicutes, as well archaeal sequences from Euryarchaeota. Cau1, Nsp6, Rxy1, Ade1, Sus1, Nsp3, Ppr2 (from *Chloroflexus aurantiacus, Nocardioides* sp. JS614, *Rubrobacter xylanophilus*, *Anaeromyxobacter dehalogenans, Solibacter usitatus, Nocardioides* sp. JS614, and *Pelobacter propionicus*, respectively) were found together in cluster nine, with the first four sequences forming one sub-cluster and the remaining three sequences forming another. Despite their separation, it is likely that the two sequences from *Nocardioides* are still orthologous to one another, and that their separation can be attributed to the trading of genetic material between two similar sequences. Examination of the 16S rRNA tree indicated that *Nocardiodes* and *Rubrobacter* were

located together in one cluster, as were *Pelobacter* and *Anaeromyxobacter* in a different cluster. The remaining genera, *Chloroflexus* and *Solibacter*, were not located in either of these two clusters, but rather were each located in their own separate clusters in the 16S rRNA tree. Subsequently, it appears that Nsp6, Nsp3, and Rxy1 could all be orthologs, as could Ppr2 and Ade1. Cau1 and Sus1, on the other hand, are fairly distant in the 16S rRNA tree from the proteins they cluster with in the protein phylogenetic tree and therefore could possibly be examples of horizontal gene transfer. The sequences Ssp1, Sel1, Cwa1, Ter1, Npu1, Nsp2, Ava1, and Nsp1 (from *Synechocystis* sp. PCC 6803, *Synechoccus elongates, Crocosphaera watsonii, Trichodesmium erythraeum, Nostoc punctiforme, Nostoc* sp. PCC 7120, *Anabaena variabilis*, and *Nostoc* sp. PCC 7120, respectively) were also seen in cluster 10. All of these sequences clustered closely in both the 16S rRNA and protein phylogenetic trees. Nsp1 and Ava1 clustered more closely together than Nsp1 did with the other sequences from *Nostoc,* which suggests that these two sequences may have exchanged genetic material with one another or both taken up similar pieces of genetic material from a foreign source. As such, it is still likely that these protein sequences could all be orthologous to each other.

Also found in cluster 10 were the sequences Bfr1, Chu1, Bfr2, and Bth1 (from *Bacteroides fragilis, Cytophaga hutchinsonii, Bacteroides fragilis*, and *Bacteroides thetaiotamicron*, respectively).These sequences all clustered closely in the protein phylogenetic tree, and their genera, which were located alongside each other in the 16S rRNA tree, exhibited very short branches. As such, it appears highly likely that

these sequences are all orthologous to each other. The sequences Dac1, Det1, and Dsp1 (from *Desulfuromonas acetoxidans, Dehalococcoides ethenogenes*, and *Dehalococcoides* sp. CBDB1, respectively) also were located together in cluster 10. While these three sequences all clustered together in the protein phylogenetic tree, with Det1 and Dsp1 the closest together, the genera *Desulfuromonas* and *Dehaloccoides* were located very distantly from each other in the 16S rRNA tree. As such, while both sequences from *Dehalococcoides* are likely to be orthologous, Dsp1 is not likely to be orthologous, and may even be an example of horizontal gene transfer. Sfu1, Reu2, and Nsp4 (from *Syntrophobacter fumaroxidans, Ralstonia eutropha*, and *Nostoc* sp. 7120, respectively) were all located around the edge of a sub-cluster of cluster 10, and an examination of the protein phylogenetic tree indicated that they were quite distant from each other and the remaining sequences in their sub-cluster. They were also quite distant from each other in the 16S rRNA tree, making them unlikely to be orthologs and possibly even examples of horizontal gene transfer.

Other sequences that grouped together in cluster 10 include Mth2, Mbu1, Mba1, Mac1, and Mma3 (from *Moorella thermoacetica, Methanococcoides burtonii, Methanosarcina barkeri, Methanosarcina acetivorans*, and *Methanosarcina mazei*, respectively). Mth2 was a bacterial sequence and was quite distant from the other four sequences it its sub-cluster, with regard to location on the 16S rRNA tree. Its location near to archaeal sequences in the protein phylogenetic tree indicates that it might have taken up some genetic material from an archaeal sequence, and thus may be an example of horizontal gene transfer. The remaining sequences from this sub-cluster

were archaea and clustered closely in both the 16S rRNA and protein phylogenetic trees. Consequently, it is likely that they are orthologous to each other, and not orthologous to Mth2. Afu1, Pfu1, Tko1, Mma2, and Mth1 (from *Archaeoglobus fulgidus, Pyrococcus furiosus, Thermococcus Kodakarensis, methanococcus maripaludis*, and *Methanothermobacter thermautotrophicus*, respectively) were also found in cluster 10. These archaeal sequences were further divided into two sub-clusters. The first sub-cluster was composed of Afu1, Pfu1, and Tko1, and the other sub-cluster was composed of Mma2 and Mth1. Although these two sub-clusters are separated . However, it is possible that they are all orthologous to one another, as they all clustered closely together in both the 16S rRNA and protein phylogenetic trees. Lastly, the sequences Cli1, Cte2, Cph2, and Pph1 (from *Chlorobium limicola, Chlorobium tepidum*, *Chlorobium phaeobacteroides*, and *Pelodictyon phaeoclathratiforme,* respectively) were also found in cluster 10. Three of these sequences belonged to *Chlorobium* and were very likely to be orthologous. The last sequence was from *Pelodictyon*, which was located right alongside *Chlorobium* in the 16S rRNA tree. The proximity of these sequences in the protein phylogenetic tree and their late branching in the 16S rRNA tree suggest that all four sequences are orthologous.

Cluster 11 contained three archaeal sequences from Euryarchaeota and six bacterial sequences from Actinobacteria, Chlorobi, Firmicutes, and Δ–proteobacteria. Mtu1 (from *Mycobacterium tuberculosis*) was distant from other proteins in cluster 11 in both the 16S rRNA and protein phylogenetic trees, and so is most likely not

orthologous to any of the other sequences. By contrast, the archaeons Hsp1, Hma1, and Hma2 (from *Halobacterium* sp. NRC-1, *Haloarcula marismortui,* and *Haloarcula marismortui*, respectively), were found close to one another in both the 16S rRNA and protein phylogenetic trees, and are most likely orthologs. Of the remaining sequences in cluster 11, Cph1, Dha1, Ppr1, Gme1, and Gsu1 (from *Chlorobium phaeobacteroides, Desulfitobacterium hafniense, Pelobacter propionicus, Geobacter metallireducens,* and *Geobacter sulfurreducens*, respectively), only Gme1 and Gsu1 clustered close to each other in both the 16S rRNA and protein phylogenetic trees, and therefore could be orthologs. The other three sequences exhibited distant branching in both the 16S rRNA and protein phylogenetic trees, and therefore are not likely to be orthologous to any other protein in cluster 11.

All seven sequences found in cluster 12 were Firmicutes and are likely to all be orthologs. Lmos2, Lde1, Spn1, Lpl1, Lac1, Lga1, and Ljo1 (from *Listeria moncytogenes, Lactobacillus delbrueckii, Streptococcus pneumoniae, Lactobacillus plantarum, Lactobacillus gasseri,* and *Lactobacillus johnsonii*) clustered closely to each other in both the 16S rRNA and protein phylogenetic trees, and exhibited similar branching patterns. While these proteins are likely to all be orthologous, it appears that some exchange of genetic material between closely related organisms, as Lpl1 clusters more closely to Lga1 and Ljo1, from *Lactobacillus*, than Lde1 does.

All sequences in cluster 13 were eukaryotes and were from Viridiplantae, Fungi, Mycetozoa, and Metazoa. No comparison was made between the protein phylogenetic tree and the 18S rRNA of the eukaryotic genera. No 18S rRNA tree was

constructed because all but two eukaryotes, which were Viridiplantae and clustered with Cyanobacteria, clustered together in the protein phylogenetic tree. While not all organisms from the same genera were adjacent to each other in the protein phylogenetic tree, all sequences were located close to one another. Therefore, it is likely that the sequences within each phylogenetic group, if not all sequences in cluster 6, are orthologs. Cluster 14 contained the remaining two eukaryotic sequences, and also contained seven bacterial sequences from Cyanobacteria. As Viridiplantae are thought to have evolved from Cyanobacteria, it is not unexpected to find that they cluster more closely to Cyanobacteria than to other eukaryotes. The Cyanobacteria, Sel2, Ter2, Npu2, Ava2, Snp7, Ssp3, and Tel1 (from *Synechococcus elongates, Trichodesmium erythraeum, Nostoc punctiforme, Anabaena variabilis, Nostoc* sp. PCC 7120, *Synechocysistis* sp. PCC 6803, and *Thermosynechoccus elongatus*, respectively) clustered closely to one another in both the 16S rRNA and protein phylogenetic trees, albeit some formation of sub-clusters in the protein phylogenetic tree and some possible differences in branching distances in the 16S rRNA tree. Despite these differences, these sequences may all be close enough to be orthologous to one another.

Cluster 15 was composed of several bacteria from Firmicutes, Spirochaetes, Δ-proteobacteria, Fusobacteria, and ε–proteobacteria. Although there were other sequences from *Streptococcus* in cluster 15, Ssu1 (from *Streptococcus suis*) clustered more closely to the two sub-clusters containing the sequences Cpe1, Cte1, Cac1, Efa1, and Efa2 (from *Clostridium perfringens, Clostridium tetani, Clostridium*

acetobutylicum, Enteroccus faecalis, and *Enterococcus faecium,* respectively), which were separated by genus. While it is highly likely that the sequences from the same genus are orthologous to one another, collectively the sequences from these three genera may or may not be orthologs as their branches were somewhat distant from each other in both the 16S rRNA and protein phylogenetic trees.

Cluster 16 consisted of only three proteins, two γ–proteobacteria and one β–proteobacteria. The sequences Mca2 and Mca3 were both from *Methylococcus capsulatus,* and since they are from the same organism and cluster adjacent to one another in the protein phylogenetic tree, they are most likely orthologs. However, Dar3 (from *Dechloromonas aromatica*) clustered at a greater distance from the other two proteins in cluster 16 in the protein phylogenetic tree and was located in an adjacent cluster in the 16S rRNA tree, and so is most likely not orthologous to Mca2 and Mca3.

Cluster 17 primarily contained γ–proteobacteria and β–proteobacteria, although it also contained one α–proteobacterium and one uncultured bacterium. The sequences Pam1, Zmo1, Bba1, Asp1, and Gox1 (from *Candidatus Protochlamydia amoebophila, Zymomonas mobilis, Bdellovibrio bacteriovorus, Acinetobacter* sp. ADP1, and *Gluconobacter oxydans,* respectively) clustered together, albeit somewhat distantly in the protein phylogenetic tree. However, they were all very distant from one another in the 16S rRNA tree, and may be an example of horizontal gene transfer. By contrast, the sequences Hin1, Pmu1, Ngo1, Nme1, Nme2, Par1, and Pcr1 (from *Haemophilus influenzae, Paseutella multocida, Neisseria gonorrhoeae, Neisseria*

meningitidis, Neisseria meningitides, Psychrobacter arcticus, and *Psychrobacter cryohalolentis*, respectively), clustered more closely to each other than Pam1, Zmo1, Bba1, Asp1, and Gox1 did in both the 16S rRNA and protein phylogenetic trees. Thus, while it is highly likely that the sequences from the same genus are orthologs, it is also possible that these seven sequences, collectively, are orthologous to each other.

Cluster 19 was composed entirely of gram positive bacteria. With the exception of one Actinobacterium, all of the sequences from cluster 19 were Firmicutes. In both the 16S rRNA and protein phylogenetic trees, Lin2 (from *Leptospira interrogans*), was too distant to make it a likely ortholog of any of the other sequences in cluster 19. Bsu1, Ban1, Bli1, Bcl1, Bha1 (from *Bacillus subtilis, Bacillus anthracis, Bacillus licheniformis, Bacillus clausii, Bacillus* and *Bacillus halodurans*, respectively) were all from the same genus and clustered closely together, making them probable orthologs. While a slight gap appeared to separate these five sequences into two sub-clusters, with Bsu1, Ban1, and Bli1 in one sub-cluster and Bcl1 and Bha1 in the other, this could be attributed to an exchange of genetic material between sequences within each of these sub-clusters of bacteria. The sequences Bce3, Gka1 and Oih1 (from *Bacillus cereus, Geobacillus kaustophilus,* and *Oceanobacillus iheyensis*, respectively) were found in a sub-cluster of cluster 19 near the sequences Sha1, Ssa1, Sau1, and Sep1 (from *Staphylococcus haemolyticus, Staphylococcus saphropyticus, Staphylococcus aureus,* and *Staphylococcus epidermidis*, respectively), which formed another sub-cluster. While it is highly likely that all of the sequences from *Staphylococcus* are orthologous to one another, collectively these proteins

appeared to have similar branching patterns in both the 16S rRNA and protein phylogenetic trees, and so could all be orthologs. Although it is interesting that Bce3 would cluster more closely in the protein phylogenetic tree to Gka1 than to other sequences from *Bacillus* in cluster 19, it is possible that an exchange of genetic material occurred between these two sequences, as they are from closely related genera. A similar explanation could explain why the sequence Bcl2, from *Bacillus clausii*, clustered more closely in the protein phylogenetic tree to Esi1, Lin1, and Lmo1 (from *Exiguobacterium sibricum, Listeria innocua, and Listeria monocytogenes*) than to any other sequence from *Bacillus*. Again, these bacteria exhibited similar branching patterns in both the 16S rRNA and protein phylogenetic trees, and their proximity in both of these trees suggests that they could all be orthologous to each other. Lastly, the sequences Hmo1, Sth2, Swo1, Cth1, and Tte1 (from *Heliobacillus mobilis, Symbiobacterium thermophilum, Syntrophomonas wolfei, Clostridium thermocellum,* and *Thermoanaerobacter tengcongensis,* respectively) were all from different genera that were found near to each other in both the 16S rRNA and protein phylogenetic trees. However, the branching distances of these organisms and genera in both trees, and the differences in location between some of the sequences in the phylogenetic tree and the locations of their genera in the 16S rRNA tree indicates that these sequences may or may not be orthologs.

The sequences from cluster 20 were mostly γ–proteobacteria, β–proteobacteria, although there were a few α–proteobacteria, as well. The sequences Mde1, from *Microbulbifer degradans*, and Mfl1, from *Methylobacillus,* were very distant both

from each other and from the other sequences in cluster 20 in both the 16S rRNA and protein phylogenetic trees, and so are not likely to be orthologous to each other or any of these sequences. Rfe1, from *Rhodoferax ferrireducens*, clustered next to Bfu1, Bma1, Bvi1, Bam1, Bce1, and Bce2 (from *Burkholderia fungorum, Burkholderia mallei, Burkholderia vietnamiensis, Burkholderia ambifaria, Burkholderia cepacia,* and *Burkholderia cenocepacia,* respectively). Although Rfe1 clustered somewhat more distantly from the six sequences from *Burkholderia* in the protein phylogenetic tree, the genera *Burkholderia* and *Rhodoferax* were next to each other in the 16S rRNA tree and branched at approximately the same distance within their cluster, thereby making them possible orthologs. Sam1, Cps1, and Sde1 (from *Shewanella amozonensis, Colwellia psychrerythraea,* and *Shewanella denitrificans*, respectively), were clustered together in both the 16S rRNA and protein phylogenetic trees. Although *Colwellia* and *Shewanella* were found within the same cluster in the 16S rRNA tree, they branched very early on in the cluster. However, as Cps1 and Sde1 clustered more closely together than Sam1 and Sde1, it appears that genetic material could have been exchanged between these two sequences and suggests that, collectively, these three sequences could be orthologous to one another. The sequences Mma2, from *Magentospirillum magnetotacticum*, and Rru1, from *Rhodospirllum rubrum*, were very distant from other sequences in cluster 20. Although they were also somewhat distant from each other, and exhibited distant branching in the 16S rRNA tree, these two sequences were found within the same cluster in both the 16S rRNA and protein phylogenetic tree and could be orthologs.

6.2: Homology Analyses of Heavy Metal ATPases

Cluster 1 contained sequences from α–proteobacteria, β–proteobacteria, γ–proteobacteria, and Deinococci. The sequences Neu1, Nwi1, Nha2, Rpa3, and Pla2 (from *Nitrosomonas eutropha, Nitrobacter winogradskyi, Nitrobacter hamburgensis, Rhodopseudomonas palustris*, and *Parvibaculum lavamentivorans*, respectively) were grouped together in the protein phylogenetic tree. As expected, the sequences from *Nitrobacter* exhibited closer branching to one another than to Rpa3 and Pla2, which also clustered closely together. Neu1 from *Nitrosomonas* also clustered closely in the protein phylogenetic tree with the sequences from *Nitrobacter,* but these two genera were found in separate clusters in the 16S rRNA tree, making it unlikely that they are orthologs. With the exception of the sequence Neu1, branching patterns similar to those seen in the protein phylogenetic tree were also observed in the 16S rRNA tree. While there was some distance between these sequences in both trees, they were all close enough together to possibly be orthologous to each other. The sequences Ssp1, Pcr1, Dge2, Dge1, and Dra1 (from *Shewanella* sp. W3-18-1, *Psychrobacter cyohalolentis, Deinococcus geothermalis*, *Deinococcus geothermalis*, and *Deinococcus radiodurans*, respectively) were also found next to one another in cluster 1. All of the sequences from *Deinococcus* grouped closer to one another than to any other sequence in this sub-cluster, and are most likely orthologous. The sequences Pcr1 and Ssp1 clustered closely together in the protein phylogenetic tree. While they

were also found within the same cluster in the 16S rRNA tree, they branched away from each other at a point close to the center of the tree. As such, they could still be orthologous to one another, but it appears that they are much more distantly related that the other sequences found in this sub-cluster. Cluster 1 also contained the sequences Atu1, Rsp2, Pde1, Pde2, Ogr1, Ssp5, Oha1, and Rsp4 (from *Agrobacterium tumefaciens, Rhodobacter sphaeroides, Paracoccus denitrificans, Paracoccus denitrificans, Oceanicola granulosus, Sulfitobacter* sp. EE-36, *Oceanicola batsensis*, and *Rhodobacter sphaeroides*, respectively). Not all of these sequences clustered the closest to other sequences from the same genus. However, this phenomenon could be attributed to an exchange of genetic material between closely related sequences. Incidentally, all of the genera of these sequences clustered closely to each other in both the 16S rRNA and protein phylogenetic trees, and therefore these sequences may all be orthologs. Also located in cluster 1 were the sequences: Rru1, Msp3, Xau1, Nha1, Rpa1, Bsp4, Rpa2, Bsp5, Ret1, and Rle1 (from *Rhodospirillum rubrum*, *Mesorhizobium* sp. BNC1, *Xanthobacter autrophicus, Nitrobacter hamburgensis, Rhodopsuedomonas palustris, Bradyrhizobium* sp. BTAi1, *Rhodopseudomonas palustris, Bradyrhizobium* sp. BTAi1, *Rhizobium etli,* and *Rhizobium leguminosarum*, respectively). Again, these sequences exhibited some variations in their clustering patterns, with regard to type of genus. However, these sequences all clustered within the same 16S rRNA cluster and exhibited similar branching patterns as those observed in the protein phylogenetic tree, and so could all be orthologous to one another.

Also found in cluster 1 were the sequences: Bsu1, Sme1, Sme2, Msp2, Asp2, and Mlo1 (from *Brucella suis, Sinorhizobium medicae, Sinorhizobium meliloti, Mesorhizobium* sp. BNC1, *Aurantimonas* sp. SI85-9A1, and *Mesorhizobium loti*, respectively). These sequences were further divided into two sub-cluster, with the first sub-cluster containing Bsu1, Sme1, and Sme2, and the second sub-cluster containing Msp2, Asp2, and Mlo1. Collectively, all six of these sequences were not only close to each other in the protein phylogenetic tree, but examination of their 16S rRNA tree indicated that their respective genera were grouped together in the same cluster and exhibited similar branching patterns as those seen amongst these sequences in the protein phylogenetic tree. Consequently, it is very likely that these proteins are all orthologous to each other. In addition to these sequences, Ppr5, Vfi1, Vch1, Vsp2, Vvu1, Vsp1, Vpa1, Van1, Ppr3, and Ppr4 (from *Photobacterium profundum, Vibrio fischeri, Vibrio cholerae, Vibrio* sp. MED222, *Vibrio vulnificus, Vibrio* sp. Ex25, *Vibrio parahaemolyticus, Photobacterium profundum*, and *Photobacterium profundum*, respectively) were found in cluster 1. Nearly all of the sequences from *Vibrio* clustered much closer to one another than to sequences from every other genus, with the exceptions of Vfi1 and Van1. Vfi1 was located by itself on a relatively distant branch next to most of the other sequences from *Vibrio*. By contrast, Van1 clustered more closely to Ppr3 and Ppr4, from *Photobacterium*, than to any other sequences from *Vibrio*. It is possible that these sequences underwent some event in which they exchanged or acquired genetic material from another similar organism, thereby causing them to branch apart from other sequences from other sequences from *Vibrio.*

Despite these differences, the genera of these sequences all clustered closely together in the 16S rRNA tree and so could all be orthologs. Also found in cluster 1 were the sequences Ahy1, Msp4, Eca1, Plu1, Pmi1, Spr1, Yps1, Yen1, Yfr1, Ymo1, and Yin1 (from *Aeromonas hydrophila, Marinomonas* sp. MWL1, *Erwinia carotovora, Photorhabdus luminescens, Proteus mirablis, Serratia proteamaculans, Yersinia pseudotuberculosis, Yersinia enterocolitica*, *Yersinia frederiksenii, Yersinia mollaretii*, and *Yersinia intermedia*, respectively). Of these sequences, Ahy1, Msp4, and Eca1 were separated by the greatest branch lengths. The remaining sequences formed several smaller sub-clusters, with Plu1 and Pmi1 clustering together and Spr1 clustering at a distance all of the sequences from *Yersinia*, which all clustered amongst themselves. With the exception of the genus *Marinomonas*, which was located in a nearby, but separate cluster, all of these sequences were found in the same cluster in the 16S rRNA tree, and exhibited clustering patterns similar to those observed in the protein phylogenetic tree. As such, these sequences may be orthologous to each other. Lastly, the sequences Esp1, Sen1, and Eco1 (from *Enterobacter* sp. 638, *Salmonella enterica*, and *Escherichia coli*, respectively) were found in cluster1. These sequences were close to one another in the protein phylogenetic tree, but were very distant from the other sequences in cluster 1. Despite their distance from their neighboring sequences, they themselves were very close to one another, both with regard to the protein phylogenetic tree and the 16S rRNA tree, where they exhibited short branch distances from each other. Consequently, these sequences could be orthologous to one another.

Cluster 2 was very large and contained sequences from β–proteobacteria, γ-proteobacteria, Δ-proteobacteria, and Cyanobacteria. Cvi1, Rfe1, Mfl1, and Ppr1 (from *Chromobacterium violaceum, Rhodoferax ferrireducens, Methylobacillus flagellatus*, and *Pelobacter propionicus*, respectively) formed one of the several sub-clusters within cluster 2, and appeared to branch distantly in the protein phylogenetic tree from each other and all of the other sequences in this cluster. Despite this distance, the genera of Cvi2, Rfe1, and Mfl1 were all found in the same cluster in the 16S rRNA tree, and so could possibly be orthologs. The genus for Ppr1, however, was found in a distant cluster and thus is not only unlikely to be orthologous to Cvi2, Rfe1, and Mfl1, and may be an example of a horizontal gene transfer event. Rme1, Psp1, Avi1, Rpi1, Asp1, Rme3, Cte2, and Kpn1 (from *Ralstonia metallidurans, Polunucleobacter* sp. QLW-P1DMWA-1, *Azotobacter vinelandii, Ralstonia piketti, Acidovorax* sp. JS42, *Ralstonia metallidurans, Comamonas testosteroni*, and *Klebsiella pneumoniae*, respectively) were also found together in cluster 2. Several further sub-clusters were distinguished amongst these sequences. Rme1 and Psp1 formed one sub-cluster, and Avi1 was located by itself at a distance from Rpi1, Asp1 and Rme3, which formed another sub-cluster, as well as from Cte2 and Kpn1, which also clustered together. In addition to grouping together in the protein phylogenetic tree, the genera of the first two sets of sequences, Rme1 and Psp1 and Rpi1, Asp1 and Rme3, all clustered close to one another in the 16S rRNA tree, which suggests that they could be orthologous. While Cte2 and Kpn1, the two sequences from the third sub-cluster within this group of proteins, clustered together in the protein phylogenetic

tree, their genera were very distant from each other in the 16S rRNA tree. While the genus *Comamonas* was located close within the same cluster in the 16S rRNA tree as the genera from the sequences in the first two sub-clusters, *Klebsiella* was found in an adjacent cluster and is most likely too distant to be orthologous to the other sequences it clustered with. Avi1, which branched distantly from the other sequences around, was also not located in the same cluster as the genera from Rme1, Psp1, Rpi1, Asp1 and Rme3, and so is unlikely to be orthologous to them. Although its genus was located in the same cluster containing *Klebsiella* in the 16S rRNA tree, these two genera branched close to the center of the tree and were also somewhat distant from each other in the protein phylogenetic tree as well. Thus, Avi1 and Kpn1 may or may not be orthologous to each other.

Additional sequences that grouped together in cluster 2 were Pme1, Pae1, Psy1, Pfl1, Pfl2, Ppu2, Ppu1, and Pen1 (from *Pseudomonas mendocina, Pseudomonas aeruginosa, Pseudomonas syringae, Pseudomonass fluorescens, Pseudomonas putida, Pseudomonas putida*, and *Pseduomonas entomophila*, respectively). Although these sequences were subdivided into three smaller sub-clusters, they all grouped together closely in the protein phylogenetic tree and were from the same genus, and are almost certainly orthologs of each other. The sequences Rme2, Reu1, and Reu2 (from *Ralstonia metallidurans, Ralstonia eutropha*, and *Ralstonia eutropha*, respectively) were found together in a sub-cluster of cluster 2 that was adjacent to the sequences from *Pseudomonas*. While these three sequences were close to one another in the protein phylogenetic tree and of the same genus, making them likely to be orthologs of

one another, their distance from the sequences in *Pseudomonas* in both the protein phylogenetic and 16S rRNA trees suggests that, collectively, these sequences are not orthologous to each other. Also found in a neighboring sub-cluster of cluster 2 were the sequences Dar1, Bpn1, Bps2, Bmu1, Bvi1, and Bsp2 (from *Dechloromonas aromatica, Burkholderia phytofirmans, Burkholderia pseudomallei, Burkholderia multivorans, Burkholderia vietnamiensis*, and *Burkholderia* sp. 383, respectively). The sequence from *Dechloromonas* clustered at a slightly greater distance from the rest of the sequences in this sub-cluster, which all cluster closely together and are from the same genus, *Burkholderia*. Nonetheless, these two genera clustered near to one another in the 16S rRNA tree and so all of these sequences could be orthologs. The sequences Sgl1, Bar1, and Bbr1 (from *Sodalis glossinidius, Bordetella avium*, and *Bordetella bronchieseptica*, respectively) and Ppr2, Rso1, and Rpi2 (from *Pelobacter propionicus, Ralstonia aolanacearum*, and *Ralstonia pickettii*, respectively) were also found branching slightly more distantly from their surrounding sub-clusters of sequences in cluster 2. While the sequences within these two small sub-clusters grouped closely amongst themselves, examination of the locations of their genera in the 16S rRNA tree indicated that in the first of these sub-clusters *Sodalis* and *Bordetella* were in separate, but nearby, sub-clusters, and that *Pelobacter* and *Ralstonia* were in separate and distant sub-clusters. Thus, while the sequences from the same genus in these sub-clusters are most likely orthologs, it is unlikely that the sequences from different genera are orthologous to one another. Additionally, the distance between the genera of Ppr2, Rso1, and Rpi2 indicated that the proximity of

Ppr2 to the two sequences from *Ralstonia* may have occurred as a result of a horizontal gene transfer event. Also found in cluster 2 were the sequences Ssp2, Aav1, Pna1, Dar2, Cte3, Asp3, Asp6, Cte4, and Dac1 (from *Synechococcus* sp. WH 5701, *Acidovorax avenae, Polaromonas napthalenivorans, Dechloromonas aromatica, Comamonas testosterone, Acidovorax* sp. JS42, *Acidovorax* sp. JS42, *Comamonas testosteroni*, and *Delftia acidovorans*, respectively). Altogether, these sequences formed two sub-clusters that branched somewhat distantly from each other, with Ssp2, Aav1, Pna1, Dar2, Cte3, Asp3 located in the first sub-cluster and Asp6, Cte4, and Dac1 located in the second sub-cluster. Of the sequences in the first of these sub-clusters, only Cte3 and Asp3 appeared to group closely to each other. By contrast, the three sequences in the second of these sub-clusters all appeared to group much closer to one another, even though two of them belonged to genera that were also present in the first cluster. This separation of sequences from the same genera indicates that some of these sequences may have undergone an exchange of genetic material with each other, or acquired or lost some quantity of genetic material, therefore causing them to cluster more distantly from other sequences from the same distance. Despite the branch distances between the sequences of these two sub-clusters in the protein phylogenetic tree, the genera of all the sequences but Ssp2 were located in the same cluster of the 16S rRNA tree and may be orthologous to each other. The genus *Synechococcus* was very distant in the 16S rRNA tree from the cluster containing the other genera from these two sub-clusters. Ssp2 also branched quite distantly in the protein phylogenetic tree from the other sequences in its sub-cluster. As such, Ssp2 is

not likely to be orthologous to these sequences, and may be an instance of horizontal gene transfer. Lastly, the sequences Cte1, Lpn1 and Lpn2 (from *Comamonas testosterone, Legionella pneumophila*, and *Legionella pneumophila*, respectively) were also found in cluster 2. While these sequences branched distantly in the protein phylogenetic tree from the other sub-clusters of cluster 2, they grouped closely together in both in this tree. However, while the two sequences from *Legionella* are likely to be orthologous to one another, the genera *Legionella* and *Comamonas* were located in separate, albeit adjacent, sub-clusters in the 16S rRNA tree and so Cte1 is most likely not orthologous to Lpn1 and Lpn2.

Cluster 3 and cluster 4 of Family 6 both only contained a few sequences. Cluster 3 was composed entirely of sequences from α–proteobacteria. The most distantly clustering sequences found in cluster 3 were Bma1, Ftu1, and Sfu1 (from *Blastopirellula marina, Francisella tularensis*, and *Syntrophobacter fumaroxidans*, respectively). In addition to being distant from one another in the protein phylogenetic tree, the genera corresponding with these sequences were located in different clusters from one another, making them unlikely to be orthologs and possibly even instances of horizontal gene transfer. Msp1, Mma7, and Mma3 (from *Mesorhizobium* sp. BNC1, *Magnetospirillum magentotacticum*, and *Magnetospirillum magneticum*, respectively) were also found in cluster 3, but were much closer to one another in the protein phylogenetic tree, and their genera were located in the same cluster in the 16S rRNA tree. While they did branch somewhat distantly from each other with the same cluster of the 16S rRNA tree, it is still possible that these sequences, especially the two

sequences from *Magnetospirillum*, were orthologous. Also located together in cluster 3 were the sequences Rsp3, Rsp1, and Dsh1 (from *Roseobacter* sp. MED193, *Roseobacter* sp. MED193, and *Dinoroseobacter shibae*, respectively). Like Msp1, Mma7, and Mma3, these three sequences clustered close to each other in the protein phylogenetic tree. However, unlike these sequences, the genera of Rsp3, Rsl1, and Dsh1 branched more closely together in the same cluster in the 16S rRNA tree, and are therefore more likely to be orthologous to one another. By contrast to cluster 3, only two sequences were found in cluster 4: Chy1 and Pam1 (from *Carboxydothermus hydrogenoformans*, and *Candidatus Protochlamydia amoebophila*, respectively). These two sequences branched very distantly from one another in both the protein phylogenetic tree and the 16S rRNA tree. Such branching distances in both trees suggest that these sequences are not orthologous to one another, and that perhaps their proximity to one another in the phylogenetic tree may signify the occurrence of a horizontal gene transfer event.

Cluster 5 contained bacterial sequences from Firmicutes, Fusobacteria, Δ-proteobacteria, ε-proteobacteria, Actinobacteria, Cyanobacteria, Bacteroidetes, γ-proteobacteria, and Spirochaetes, and archaeal sequences from Euryarchaeota. These sequences primarily grouped together with other sequences from the same genus or phylogenetic group, although this was not always the case. Among the sequences in cluster 5 with the greatest branch lengths were Lin1, Lsa2, and Wsu1 (from *Lawsonia intracellularis, Lactobacillus salivarius*, and *Wolinella succinogenes*, respectively). While these sequences appeared to be located more closely to one another in the

protein phylogenetic tree than to any other sequences in cluster 5, they branched very distantly from one another in both this tree and the 16S rRNA tree. *Lawsonia* and *Wolinella* were located in the same cluster in the 16S rRNA tree, but branched close to the center of the tree, making it uncertain whether or not they are orthologs. Lin1, by contrast, not only branched the most distantly from the other sequences, but its genus, *Lactobacillus,* was located in a cluster on the opposite side of the 16S rRNA tree from *Lawsonia* and *Wolinella*. Given these sequences' distant branching in both trees, it is unlikely that these sequences are orthologs, especially Lin1. Additionally, the clustering patterns of these sequences in the 16S rRNA tree were not thought to indicate the occurrence of horizontal gene transfer, as significant branch lengths were also observed in the protein phylogenetic tree as well. The sequences Efa1, Efa2, Fnu1, and Cce1 (from *Enterococcus faecalis, Enterococcus faecium, Fusobacterium nucleatum*, and *Clostrium cellulolyticum*, respectively) were also found within cluster 5. Efa1 and Efa2 clustered together in the protein phylogenetic tree and belonged to the same genus, therefore making them likely orthologs. Fnu1 and Cce1, by contrast, were more spread out from each other in the protein phylogenetic tree and in the 16S rRNA tree. Although these two sequences were found within the same cluster of the 16S rRNA tree, they branched away from a point almost at the center of the tree. As such, these sequences may or may not be orthologous to one another. Also, the cluster in the 16S rRNA tree in which *Fusobacterium* and *Clostridium* were found in one cluster and *Enterococcus* was found in an adjacent cluster, thereby making it less likely that, collectively, these four sequences are orthologs. The sequences Bce1,

Bce2, Bsu2, and Bli1 (from *Bacillus cereus, Bacillus cerus, Bacillus subtilis*, and *Bacillus licheniformis*, respectively) were also found together in cluster5. Although the protein phylogenetic tree indicated that these sequences all clustered closely, they exhibited some separation from each other and created two sets of sub-clusters, which each contained two sequences. This partitioning of sequences into two separate sub-clusters could indicate that the sequences from *Bacillus* that clustered most closely to one another underwent similar events that affected their genetic composition. Despite the slight distance between some of these sequences, they were from the same genus are most likely to be orthologous. The sequences Cth1, Csp3, Dha1, Csu1, Ame2, Csp4, Cbe1, and Cno1 (from *Clostridium thermocellum, Clostridium* sp. OhILAs, *Desulitobacterium hafniense, Alkaliphilus metalliredigenes, Clostridium* sp. OhILAs, *Clostridium beijerincki*, and *Clostridium novyi*, respectively) were also found in cluster 5. Not all of the sequences from *Clostridium* clustered adjacent to one another, which suggested that they could have exchanged genetic material with other closely clustering sequences from different genera. Despite their separation, Cth1, Csp3, Csu1, Csp4, and Cbe1 were still relatively close to one another in the protein phylogenetic tree and they are all from the same genus. As such, these sequences could be orthologous to one another. While the remaining three sequences from this sub-cluster, Dha1, Csu1, and Ame1, were located within close proximity to one another in both the protein phylogenetic and 16S rRNA trees, making them possible orthologs, the genera from these three sequences were distant from *Clostridium* in the

16S rRNA tree. Thus, it does not appear that these eight sequences are all orthologous to one another.

Also located in cluster 5 were the sequences Tet1, Tet2, Cdi1, Cpe1, Cte5, and Cac1 (from *Thermoanaerobacter ethanolicus, Thermoanaerobacter ethanolicus, Clostridium difficile, Clostridium perfringens, Clostridium tetani*, and *Clostridium acetobutylicum*, respectively). These sequences were divided into three sub-clusters, each containing two sequences. Two of the sub-clusters contained the sequences from *Clostridium*, and while these sub-clusters exhibited relatively distant branching from each other in the protein phylogenetic tree, they could still be orthologous to one another. The remaining sub-cluster contained the two sequences from *Thermoanaerobacter*, which grouped tightly together and were most likely orthologs. Although the two genera represented by the sequences in these three clusters were found within the same cluster of the 16S rRNA tree, they branched away from each other at a point that was very close to the center of the tree. Subsequently, these sequences may or may not all be orthologous to each other. Mla1, Esi1, and Lsa1 (from *Methanocorpusculum labreanum, Exiguobacterium sibiricum*, and *Lactobacillus sakei*, respectively) were also found together in a sub-cluster of cluster 5. Esi1 and Lsa1 grouped the most closely together of the three sequences in the protein phylogenetic tree and their genera were located with the same cluster of the 16S rRNA tree, therefore making them likely to be orthologous to one another. Whereas Esi1 and Lsa1 were bacterial sequences, Mla1 was an archaeal sequence and was in a cluster that represented an entirely different phylogenetic domain in the 16S

rRNA tree than Esi1 and Lsa1. The proximity of Mla1 to two bacteria suggests that horizontal gene transfer may have occurred. Other sequences from cluster 5 included Hor1, Pab1, and Mma4 (from *Halothermothrix orenii, Pyrococcus abyssi*, and *Methanococcus maripaludis*, respectively). While Pab1 and Mma4 clustered more closely to one another than to Hor1, they were both archaeal sequences and displayed similar branching points in the 16S rRNA tree. Such similarities make it likely that Pab1 and Mma4 are orthologs. Although Hor1 is a bacterial sequence, its genus was found in a cluster near the archaeal genera and it separated from the branch that contained all of the archaeal genera at a point near the center of the 16S rRNA tree. Despite its proximity to Pab1 and Mma4, Hor1 is from a different phylogenetic domain. Thus, it is likely that its proximity to these archaeal sequences in the protein phylogenetic tree can be attributed to a horizontal gene transfer event.

Also found in cluster 5 were the sequences Sth3, Mth1, Dha2, and Dha3 (from *Symbiobacterium thermophilum, Moorella thermoaceitca, Desulfitobacterium hafniense*, and *Desulfitobacterium hafniense*, respectively). Although these sequences were not all from different genera, they all clustered closely to one another in the protein phylogenetic tree and the 16S rRNA tree, and were all likely to be orthologs. The sequences Sth3, Mth1, Dha2, and Dha3 also clustered next to Npu2, Ssp7, Obp1, and Lsp2 (from *Nostoc punctiforme, Synechocystis* sp. PCC 6803, *Oscillatoria brevis*, and *Lyngbya* sp. PCC 8106, respectively) in cluster 5. Like Sth3, Mth1, Dha2, and Dha3, Npu2, Ssp7, Obp1, and Lsp2 all clustered closely in both the 16S rRNA tree and the protein phylogenetic tree, which suggested that they were all orthologous to

one another. However, the genera of Sth3, Mth1, Dha2, and Dha3 were located in a separate cluster in the 16S rRNA tree from the one containing the genera representing the sequences Npu2, Ssp7, Obp1, and Lsp2, thereby making it unlikely that all of these sequences were orthologous.

Other sequences found in cluster 5 were Chu1, Chu2, Fjo1, Orh1, and Spu1 (from *Cytophaga hutchinsonii, Cytophaga hutchinsonii, Flavobacterium johnsoniae, Ornithobacterium rhinotracheale*, and *Shewanella putrfaciens*, respectively). These sequences grouped closely to one another in the protein phylogenetic tree, and, with the exception of Spu1, were all found clustering together in the 16S rRNA tree at distances that were similar to the patterns observed in the protein phylogenetic tree. Spu1, by contrast, was located in a very distant cluster from the rest of these sequences in the 16S rRNA tree, and may even be an instance of horizontal gene transfer. Also found in cluster 5 were the sequences Lbl2, Tsp1, Rbi1, Gfo2, Csp5, Lpl3, Brf1, and Bth2 (from *Leeuwenhoekiella blandensis, Tenacibaculum* sp. MED152, *Robignitalea biformata, Gramella forsetti, Cellulophaga* sp. MED134, *Leewenhoekiella blandensis*, *Bacteroides fragilis*, and *Bacteroides thetaiotaomicron,* respectively). These sequences all grouped together in cluster 5, forming several small, close sub-clusters. Additionally, the genera corresponding to these sequences were all located in the same cluster in the 16S rRNA tree, and appeared to exhibit similar clustering patterns as those seen in the phylogenetic tree. While these sequences all grouped together in both trees and could all be orthologs, the two sequences from *Bacteroides* branched at a notably greater distance from the other sequences in both trees, and, consequently,

may or may not be orthologous to the other six proteins they clustered with. Lastly, the sequences Tde1, Csa2, Hhe1, Hac1, and Hfe1 (from *Treponema denticola, Caldicellulosiruptor saccharolyticus, Helicobacter hepaticus, Helicobacter acinonychis*, and *Helicobacter felis*, respectively) were also found in cluster 5. The three sequences from *Helicobacter*, Hhe1, Hac1, and Hfe1, are most likely orthologous to each other as they grouped closely together and were notably distant from even their neighboring proteins Tde1 and Csa2. While Tde1 and Csa2 also appeared to be in close proximity to one another, their genera were neither located in the same cluster nor in the same cluster as Hhe1, Hac1, and Hfe1. Thus, Tde1 and Csa2 were unlikely to be orthologous to each other or to Hhe1, Hac1, and Hfe1, and may even be examples of horizontal gene transfer.

All sequences in cluster 6 were eukaryotes from Viridiplantae, and included Ota1, Mtr1, Aha2, Ath1, Tca2, and Aha1 (from *Ostreococcus tauri, Medicago truncatula, Arapidopsis halleri, Arabidopsis thaliana, Thlaspi caerulescens*, and *Arabidopsis halleri*, respectively). As in Family 5, no 18S rRNA tree was constructed because all but two eukaryotes, which were Viridiplantae and clustered with Firmicutes and Chlamydiae, were located together in the protein phylogenetic tree. Although not all organisms from the same genera were located adjacent to each other, their close proximity in the protein phylogenetic tree suggests that sequences could still be similar enough to be orthologous to one another.

The sequences found in cluster 7 were from the bacterial phylogenetic groups Firmicutes, Actinobacteria, Chloroflexi, and α–proteobacteria, and from the eukaryotic

phylogenetic group Euryarchaeota. Csp2, Lme1, Lsa3 (from *Clostridium* sp. OhILAs, *Leuconstoc mesenteroides,* and *Lactobacillus salivarius*, respectively) were located in a sub-cluster that was quite distant from the other sequences in cluster 7, as were Ame1 and Hau1 (from *Alkaliphilus metalliredigenes* and *Herpetosiphon*, respectively). Csp2, Lme1, and Lsa3 exhibited similar clustering patterns in both the 16S rRNA and protein phylogenetic trees, and were close enough to each other in both that they could all be orthologs. By contrast, Ame1 and Hau1 are very distant from each other in the 16S rRNA tree, as Ame1 is a bacterial sequence and Hau1 is an archaeal sequence. This considerable difference makes it unlikely that they are orthologous, and suggests that this could be an instance of horizontal gene transfer. Met2, Hma5, Cgl2, Cgl1, Cef2, Nsp4, Bli3, Rxy2, and Pde3 (from an uncultured methanogenic archaeon*, Haloarcula marismortui, Corynebacterium glutamicum, Corynebacterium glutamicum Corynebacterium efficiens, Nocardioides* sp. JS614, *Brevibacterium linens, Rubrobacter xylanophilus*, and *Paracoccus denitrificans*, respectively) were also found in cluster 7. These sequences were further divided into four sub-clusters: met2 and Hma5 in the first, Cgl2, Cgl2 and Cef2 in the second, Nsp4 and Bli3 in the third, and Rxy2 and Pde3 in the fourth. As Hma5 and met2 were archaeal sequences and the remaining sequences in these sub-clusters were bacterial sequences, these two sequences may be instances of horizontal gene transfer. With the exception of Rxy2 and Pde3, the bacterial sequences were all located close to one another in the same cluster of the 16S rRNA tree, and could be orthologs. Rxy2 was located in the same cluster as Cgl2, Cgl1, Nsp4, and Bli3, but it branched very close to

the center of the tree, and so may or may not be orthologous to these sequences. By contrast, Pde3 was from the genus *Paraoccus*, which was very distant in the 16S rRNA tree from the genera of the other sequences. Consequently, Pde3 could be an instance of horizontal gene transfer.

Cluster 8 was primarily composed of Firmicutes and Chlamydiae, although it also contained two eukaryotic sequences from Viridiplantae, Ath2 and Osa1 (from *Arabidopsis thaliana* and *Oryza sativa*, respectively). As previously discussed, Viridiplantae are thought to have evolved from Cyanobacteria. Although no sequences from Cyanobacteria were found in cluster 8, the genera from Chlamydiae were located in a cluster adjacent to one containing genera from Cyanobacteria in the 16S rRNA tree. While these two clusters exhibited very early branching from one another, it is possible that Heavy Metal P-type ATPases from Viridiplantae are more closely related to the phylogenetic group Chlamydia than Cyanobacteria. If this is true, then some time after the division between genera from Chlamydiae and genera from Cyanobacteria occurred, the Family 6 genera from Viridiplantae separated from the genera from Chlamydiae. The branch distances observed, however, between members of these phylogenetic groups are significant, and it is unlikely that the sequences from Viridiplantae are orthologous to any of the bacterial sequences in cluster 8. Examination of the phylogenetic tree indicates that cluster 8 is comprised of several sub-clusters of bacterial sequences. The sequences Efa4, Esi2, Bcl2, Bha2, Bsp6, Bce3, Bth1, Bsu3, Bli1 (from *Enteroccous faecium, Exiguobacterium sibiricum, Bacillus clausii, Bacillus halodurans, Bacillus* sp. NRRL B-14911, *Bacillus cereus,*

Bacillus thuringiensis, Bacillus subtills, and *Bacillus licheniformis,* respectively) were located close together in both the protein phylogenetic and 16S rRNA trees, and are therefore likely to all be orthologous to one another. Lwe1 and Lmo2 clustered adjacent to each other in the protein phylogenetic tree (from *Listeria weishimeri* and *Listeria monocytogenes,* respectively), as did Efa3 and Spy1 (from *Enterococcus faecium* and *Streptococcus pyogenes*, respectively). While Lwe1 and Lmo2 were from the same genus and are very likely to be orthologs, Efa3 and Spy1 were distant from the other sequences in cluster 8 and were from genera that are in adjacent clusters in the 16S rRNA tree. As such, they are less likely to be orthologous to each other or to other sequences in cluster 8. The sequences Ctr1, Cpn1, Cab1, and Cfe1 (from *Chlamydia trachomatis, Chlamydophila pneumoniae, Chlamydophila abortus*, and *Chlamydophila felis*, respectively) formed another sub-cluster found in cluster 8. Although Ctr1 was from *Chlamydia* and was somewhat distant in the protein phylogenetic tree from the other proteins in this sub-cluster, which were all from *Chlamydophila*, these two genera were only a short distance from each other in the 16S rRNA tree. Thus, it is likely that these four sequences were orthologous to one another.

The sequences found in cluster 9 were from Actinobacteria, α-proteobacteria, β-proteobacteria, γ-proteobacteria, and Bacteroidetes. The most distant sub-clusters with cluster 9 were Cef1, Lbl1 and Gfo1, Sru1 and Asp5, Dac2 and Nar1, and Mfl2, Ssp6, and Hne1. Cef1, from *Corynebacterium efficiens*, was distant in both the 16S rRNA and protein phylogenetic trees from the other sequences in cluster 9. Similarly,

Lbl1, and Gfo1 (from *Leeuwenhoekiella blandensis,* and *Gramella forsetii*, respectively), which cluster closely enough in both trees to likely be orthologous to each other, were distant from the other sequences in cluster 9 and, like Cef1, are not likely to be orthologous to any of the other sequences in cluster 9. While Sru1 and Asp5 (from *Salinibacter rubber* and *Arthrobacter* sp. FB24, respectively) have a relatively short branch distance from each other in the protein phylogenetic tree, their genera were very distant from each other in the 16S rRNA tree. As such, their proximity to one another in the protein phylogenetic tree could indicate the occurrence of horizontal gene transfer. Similarly, Dac2 and Nar1 (from *Delftia acidovorans* and *Novosphingobium aromaticivorans*, respectively) were also located next to each other in the protein phylogenetic tree, but their genera were far apart in the 16S rRNA tree and could represent an instance of horizontal gene transfer. Mfl2, Ssp6, and Hne1 (from *Mycobacterium flavescens*, *Sulfitobacter* sp. NAS-14.1, and *Hyphomonas neptunium*, respectively) also formed one of the more distant sub-clusters within cluster 9. Although Mfl2 was not very distant from Ssp6 and Hne1 in the protein phylogenetic tree, its genus, *Mycobacterium,* was very distant from the genera *Sulfitobacter* and *Hyphomonas* in the 16S rRNA tree, which were found in the same cluster. Consequently, while Ssp6 and Hne1 could be orthologous to each other, *Mycobacterium* was so distant from the other two genera in the 16S rRNA tree that Mfl2 may be an example of horizontal gene transfer.

Also found in cluster 9 were the sequences Csp1, Ccr1, Asp4, Bmu2 (from *Caulobacter* sp. K31, *Caulobacter crescentus, Acidovorax* sp. JS42, and *Burkholderia*

multivorans, respectively). Csp1 and Ccr1 were from the same genus and are very likely to be orthologous to each other. Bmu2 and Asp4, while not from the same genus, were found in the same cluster of the 16S rRNA and clustered next to each other in the protein phylogenetic tree, and could also be orthologous to one another. Although these four sequences all are found relatively close to each other in the protein phylogenetic tree, they are not found in the same cluster in the 16S rRNA tree and are most likely not all orthologous to each other. Eli1, Ppu3, and Sal1(from *Erythrobacter litoralis, Pseudomonas putida*, and *Sphingopyxis alaskensis*, respectively) were also found in cluster 9. Although these sequences were next to each other in the protein phylogenetic tree, they exhibited early branching points. While Eli1 and Sal1 did not cluster as closely together in the protein phylogenetic tree as Sal1 and Ppu3, the genera *Sphinopyxis* and *Erythrobacter* were found in the same cluster of the 16S rRNA tree, whereas *Pseudomonas* is located in a nearby cluster. As such, Eli1 and Sal1 may be orthologous to each other, but is unlikely that they are also orthologous to Ppu3. Also seen in cluster 9 were the sequences Rme4, Rpa4, and Pla1 (from *Ralstonia metallidurans, Rhodopseudomonas palustris,* and *Parvibaculum lavamentivorans*, respectively). Although the branch distances between these sequences in the protein phylogenetic tree suggested that these sequences were not incredibly close to one another, the genera *Parvibaculum* and *Rhodospeudomonas* were adjacent to each other in the 16S rRNA tree and exhibited similar branching patterns as the protein phylogenetic tree. Thus, Rpa4 and Pla1 could be orthologous to each other. The genus *Ralstonia*, however was located in a separate, but nearby,

cluster, and suggests that Rme4 is most likely not orthologous to Rpa4 and Pla1. Lastly, Bsp7, Xau2, Xau3, Ssp4, and Msp5 (from *Bradyrhizobium* sp. BTAi1, *Xanthobacter autotrophicus, Xanthobacter autotrophicus, Sphigomonas* sp. SKA58, and *Mesorhizobium* sp. BNC1, respectively) were also found in cluster 9. Bsp7, Xau2, and Xau3 were located close together in the protein phylogenetic tree, as were their genera, *Bradyrhizobium* and *Xanthobacter*, in the 16S rRNA tree. Therefore it is likely that these three sequences are orthologous to each other. Msp5 and Ssp4 were also close to each other in the protein phylogenetic tree and their genera are found within the same cluster of the 16S rRNA tree, so they too could be orthologs.

The sequences found in cluster 10 were all either from Chloroflexi or from Cyanobacteria. Ssp3, Lsp1, Nsp2, Npu1, and Nsp3 (from *Synechocystis* sp. PCC6803, *Lyngbya* sp. PCC 8106, *Nodularia spumigena, Nostoc punctiforme,* and *Nostoc* sp. PCC 7120, respectively) were all located close together in a sub-cluster of cluster 10. The genera of these sequences were also located closely in the same cluster of the 16S rRNA tree, which indicated that all five of these sequences could be orthologs. Sth2, Cau2, Cag2, Rca2, and Rsp5 (from *Symbiobacterium thermophilum, Chloroflexus aurantiacus, Chloroflexus aggregans, Roseiflexus castenholsii*, and *Roseiflexus* sp. RS-1, respectively) were also found in a sub-cluster of cluster 10. Sth2 was the most distant of these five sequences in the protein phylogenetic. The remaining four sequences clustered closer together, and in accordance to their genus. While the genera *Chloroflexus* and *Roseiflexus* were located in the same cluster of the 16S rRNA tree adjacent to one another, *Symbiobacterium* was located in a very distant cluster. As

such, it seems likely that while Cau2, Cag2, Rca2, and Rsp5 are most likely orthologs, Sth2 may be the result of a horizontal gene transfer event.

Two of the clusters representing the sequences of the Heavy Metal ATPases only contained a single sequence: cluster 11 and cluster 17. Cluster 11 contained the sequence Lpn3, from *Legionella pneumophila,* and cluster 17 contained the sequence Tth1, from *Thermus thermophilus*. Tth1 was the only sequence from the genus *Thermus,* which could explain why it did not cluster with the other sequences from the phylogenetic group Deinococci in cluster1. However, Lpn3 was not the only sequence from *Legionella* in Family6, and so it is possible that it may have undergone horizontal gene transfer, making it notably distinct from the other sequences from Legionella in Family 6. Also, as Lpn3 is only 635 amino acids long, and the other two sequences from *Legionella*, Lpn1 and Lpn2, are 729 and 713 amino acids long, respectively. As such, it is possible that Lpn1 and Lpn2 underwent a gene duplication event contained an extra domain, thereby making them longer and more distant from Lpn3.

Cluster 12 was entirely composed of sequences from Actinobacteria. Nfa1, Msp6, Mva1, Mfl4, Rer1, and Mfl6 (from *Nocardia farcina, Mycobacterium* sp. JLS, *Mycobacterium vanbaalenii, Mycobacterium flavescens, Rhodococcus erythropolis*, and *Mycobacterium flavescens*, respectively) were located next to one another in cluster 12. While most of the sequences from *Mycobacterium* clustered alongside each other, Mfl6 clustered more closely to Rer1, from *Rhodococcus*. This variation could be attributed to the exchange of genetic material between sequences from genera that

cluster near one another in their corresponding 16S rRNA tree, as is seen with these two genera. Collectively, *Mycobacterium, Rhodococcus, and Nocardia* are all located very close to one another in the same cluster of the 16S rRNA tree, and so Nfa1, Msp6, Mva1, Mfl4, Rer1, and Mfl6 are likely to be orthologous sequences. Jsp1, Nsp5, Asp8, mar1, Aau2, Aau1, and Asp7 (from *Janibacter* sp. HTCC2649, *Nocardiodes* sp. JS614, *Arthrobacter aurescens,* a marine actinobacterium*, Arthrobacter aurescens*, and *Arthrobacter* sp. FB24, respectively) were also found in cluster 12. Although the genus of mar1 is unknown, and it is only described as a marine actinobacterium, clustered very closely with Asp8 in the protein phylogenetic tree and these two sequences are most likely orthologous to one another, as well as to the other three sequences from *Arthrobacter*, Aau2, Aau1, and Asp7. The sequences Jsp1 and Nsp5, which clustered closely to each other in both the protein phylogenetic and 16S rRNA trees, were also found in the same cluster in the 16S rRNA tree as *Nocardiodes* and *Arthrobacter.* Thus, the sequences Jsp1, Nsp5, Asp8, mar1, Aau2, Aau1, and Asp7 could all be orthologs. Gwe1, Rer2, Mfl7, Msp7, Mfl3, and Mfl5 (from *Gordonia westfalica, Rhodococcus erythropolis, Mycobacterium flacescens, Mycobacterium* sp. MCS, *Mycobacterium flavescens*, and *Mycobacterium flavescens*, respectively) were also found together in cluster 12, with the sequences from *Mycobacterium* grouped more closely amongst themselves than with either Gwe1 or Rer2. As all three of the genera represented by these sequences clustered closely within the 16S rRNA tree, there is a strong likelihood that these sequences, especially the ones from *Mycobacterium*, are orthologous to one another. Lastly, the sequences

mar2, Cje1, Bli4, Cef3, and Cef4 (from a marine actinobacterium, *Corynebacterium jeikeium, Brevibacterium linens, Corynebacterium efficiens*, and *Corynebacterium efficiens*, respectively) were also found in cluster 12. Although the separation of Cje1 from the other sequences from *Corynebacterium* indicates that it may have exchanged some genetic material with Bli4, it is still likely Cje1, Bli4, Cef3, and Cef4 were all orthologous to one another, as both *Corynebacterium* and *Brevibacterium* were found within the same cluster of the 16S rRNA tree and branched at distances that were similar to what was observed in the protein phylogenetic tree. Although it is possible that mar2 is also orthologous to these sequences, it exhibited a greater branch length in the protein phylogenetic tree than the other sequences surrounding it. Also, as mar2 is only described as a marine actinobacterium, it cannot be determined whether or not it is similar enough to be orthologous to these sequences, or if its cluster location in the protein phylogenetic tree was caused by a horizontal gene transfer event or the exchange of genetic material from sequences from nearby, but separate clusters in the 16S rRNA tree.

Clusters 13 and 14 were amongst the smallest clusters representing Family 6, and were both exclusively composed of archaeal sequences. Cluster 13 contained the archaeal sequences Nph1, Hsp2, Hma3, Hma1, Hwa1, Hma2, Hma4 (from *Natronomonas pharaonis, Halobacterium* sp. NRC-1, *Haloarcula marismortui, Haloarcula marismortui, Haloquadratum walsbyi, Haloarcula marismortui*, and *Haloarcula marismortui*, respectively). These sequences all clustered closely together in both the protein phylogenetic and 16S rRNA trees, making it very likely that they

are orthologous. Although Hwa1, from *Halquadratum,* grouped more closely to Hma3 and Hma1 than the two other sequences from *Haloarcula* did, this variation is most likely attributable to the exchange of genetic material between sequences from two closely clustering genera. As such, it is likely that all of these sequences are orthologous to each other. Cluster 14 only contained three sequences: Mst1, Mma2, Mth2 (from *Methanosphaera stadtmanae, Methanosphaera marisigri,* and *Methanothermobacter thermautrophicus,* respectively). These sequences were found very close to one another in the protein phylogenetic tree and their genera, *Methanosphaera* and *methanothermobacter* were located adjacent to one another in the16S rRNA tree. Consequently, it is very likely that all three of these sequences are orthologs.

Cluster 15 was almost entirely composed of sequences from Firmicutes, with the exception of met1, an archaeal sequence from Euryarchaeota. The sequences Gst1, Gka2, Bha1, Bcl1, Sau1, Bps1, Sep1, Lmo2, Sth1, and Lla1 (from *Geobacillus stearothermophilus, Geobacillus kaustophilus, Bacillus halodurans, Bacillus clausii, Staphylococcus aureus, Bacillus pseudofirmus, Staphylococcus epidermidis, Listeria monocytogenes, Streptococcus thermophilus*, and *Lactococcus lactis*, respectively) all clustered together in the phylogenetic tree. Gst1 and Gka2 were the most distantly branching sequences of this sub-cluster, but as they were located next to one another in the protein phylogenetic tree and they are from the same genus, they are most likely orthologous to one another. Similarly, while it appears that an exchange of genetic material may have occurred between some of the sequences, as Bps1 and Sth1 were

not located amongst the other sequences in this sub-cluster from their respective genera, these sequences all belong to genera that cluster closely in the 16S rRNA tree. Thus, all of the sequences within this sub-cluster could be orthologous to one another. Oih1, Bsp1, Gka1, Sag1, Gsu1, Tca1 (from *Oceanobacillus iheyensis, Bacillus* sp. NRRL B-14911, *Geobacillus kaustophillus, Streptococcus agalactiae, Geobacter sulfurreducens*, and *Thermosinus carboxydivorans*, respectively) were also next to one another in cluster15. Although these sequences formed two distinct sub-clusters in the protein phylogenetic tree, their genera were all found within the same cluster in the 16S rRNA tree and exhibited similar branching patterns as was observed in the protein phylogenetic tree. As such, it is likely that, despite some distance separating these sub-clusters in the protein phylogenetic tree, these sequences are all orthologs. Lastly, the sequences Swo1, Dre1, met1, Bwe1, Ssa1 (from *Syntrophomonas wolfei, Desulfotomaculum reducens,* an uncultured methanogenic archaeon, *Bacillus weihenstephanensis,* and *Staphylococcus saprophyticus*, respectively) were found grouped into two sub-clusters within cluster 15. Swo1 and Dre1, two bacteria, formed one sub-cluster, along with met1, an archaeal sequence. Although Swo1 and Dre1 were close to one another both the protein phylogenetic tree and the 16S rRNA tree, and are most likely orthologs, met1 was from a separate phylogenetic kingdom and is not likely to be orthologous to Swo1 or Dwe1. Bwe1 and Ssa1 formed the second sub-cluster amongst these five sequences, and were located close to one another in both the protein phylogenetic and 16S rRNA trees, making them likely orthologs. Additionally the genera of these two sequences are located in the same cluster as those of Swo1 and

Dre1, albeit rather distantly, and so collectively these four sequences could be orthologous to one another.

The sequences found in cluster 16 were from Cyanobacteria, Actinobacteria, Chloroflexi, and one unclassified Proteobacteria. Cluster 16 was composed of two distant sub-clusters, with Ava1, Nsp1, Mfe1, Rxy1 (from *Anabaena variabilis, Nostoc* sp. PCC 7120, *Mariprofundus ferroxydans,* and *Rubrobacter xylanophilus,* respectively) in one, and Rca1, Cag1, and Cau1 (from *Roseiflexus castenholzii, Chloroflexus aggreganss*, and *Chloroflexus aurantiacus*, respectively) in the other. In the first sub-cluster, Ava1 and Nsp1 clustered together and Mfe1 and Rxy1 clustered together. While genera of Ava1 and Nsp1 were also located adjacent to one another in the 16S rRNA tree, indicating that they are most likely orthologs, the genera of Mfe1 and Rxy1 were very distant from each other and the proximity of their clustering in the protein phylogenetic tree could represent an instance of horizontal gene transfer. The sequences found in the second sub-cluster of cluster 16, Rca1, Cag1, and Cau1, were located directly beside one another in both the 16S rRNA and protein phylogenetic trees. As such, it is very likely that these three sequences are orthologous to one another.

Chapter 7: Analyses of P-type ATPases by Organismal Domain

Analyses of P-type ATPase by Organismal Domain

Introduction

P-type ATPases, regardless of substrate type or host organism, exhibit certain characteristic features. These include a phosphorylation site and ATP-binding site, and the ability to undergo phosphate recycling. Additionally, P-type ATPases generally show conservation of nine well-known sequence motifs. While all P-type ATPases consist of several joined cytosolic and membrane-associated regions, the actual number of membrane traverses often differs between proteins with different substrate types and phylogenetic classifications. For instance, previous research indicated that many P-type ATPases tend to have six membrane traverses in the N-terminal region (type I P-type ATPases), whereas type II enzymes only have four in this region. By contrast, these type II systems frequently have more membrane traverses in their C-terminal region, generally four to six traverses, as opposed to the two to four traverses often seen in the same region in type I prokaryotes (Møller J.V., et al., 1995). These differences in TMS locations affect the organization of the intramembranous channel through which substrates ions or phospholipids pass, and may provide some insight into the genomic evolution of these different types of P-type ATPases.

Type I P-type ATPases appear to have originally evolved to allow for heavy metal transport, and most notably include the proteins of Families 5 and 6, the Copper and Heavy Metal P-type ATPases (Møller J.V., et al., 1995). Most type I P-type

ATPases are characterized by well-conserved cysteine-containing consensus motifs in the N-terminal region, which are involved in heavy metal ion binding. Family 7, the Kdp transporters, have also been classified as type I P-type ATPases. It has been proposed that channel membrane proteins for K^+ may have fused to the C-terminal end of type I P-type ATPases, thereby making them capable of K^+ translocation (Møller J.V., et al., 1995). Type II P-type ATPases generally have more membrane traverses than type I P-type ATPases, especially in the C-terminal region. It has been suggested that these extra membrane traverses are a feature necessitated by the general tendency of type II P-type ATPases to transport ions like Na^+, K^+, Ca^{2+}, and H^+, which have lower atomic masses than those transported by type I P-type ATPases (Møller J.V., et al., 1995).

7.1: Archaea

The number of archaeal sequences containing Copper and Heavy Metal P-type ATPases were twenty and fifteen, respectively, and corresponded to only a small fraction of the total number of prokaryotes possessing Families 5 and 6 ATPases (Tables 2 and 5). In Family 5, archaeal homologues were found in clusters 2, 4, 10 and 11. The two sequences in cluster 1 were from Crenarchaeota and were from the genera *Aeropyrum* and *Pyrobaculum*. These archaea were in the same cluster as two bacterial proteins from Thermotogae and ε-proteobacteria. Six Euryarchaeotal enzymes were found in cluster 4, one from *Archaeoglobus,* two from *Methanosarcina,*

one from *Methanothermobacter* and one from *Haloarcula.* Nine Euryarchaeotal enzymes were found in cluster 10. Four of these archaeal sequences were from a sub-cluster also containing bacterial sequences, and were from the genera *Methanococcoides* and *Methanosarcina*. Another four sequences were from a sub-cluster without bacteria, and were from the genera *Archaeoglobus*, *Methanococcus, Pyrococcus, Thermococcus*. The final sequence was found in another sub-cluster, along with several bacteria, and was from the genus *Methanothermobacter.* Three more archaeal sequences from Euryarchaeota were found in cluster 11, two from the genera *Haloarcula* and one from the genus *Halobacterium*.

In Family 6 seven Euryarchaeotal homologues were found in cluster 13, four from the genus *Haloarcula,* and one from each of the genera *Natromonas, Haloquadratum,* and *Halobacterium*. Three proteins from Euryarchaeota were found in cluster 14. Cluster 14 did not contain any bacterial homologues, and its three sequences were each from a separate genera, *Methanosphaera, Methanoculleus,* and *Methanothermobacter*. One uncultured methanogenic archaeal protein from Euryarchaeota was found in cluster 15, along with two bacterial ATPases. Similarly, three sequences from Euryarchaeota were found amongst bacterial sequences in cluster 5, one from each of the genera *Methanocorpusculum, Pyrococcus* and *Methanococcus*. Lastly, two archaeal sequences from Euryarchaeota were found in cluster 7, one from the genus *Pyrococcus* and the other from the genus *Methanococcus.*

7.2: Bacteria

Bacteria by far was the largest of the three domains, with 318 bacteria sequences representing Family 5 and 288 bacterial sequences in Family 6 (Tables 2 and 5). In both of Families 5 and 6, these bacterial sequences were further subdivided into many clusters and sub-clusters, which were generally based on similarities in sequence lengths and phylogenetic groups. P-type ATPases often exhibit greater sequence similarity based on substrate type, as opposed to organism type (Møller J.V., et al., 1995). Additionally, it is not uncommon for bacterial sequences to take up foreign DNA and to incorporate it into their own genetic makeup, either accidentally or intentionally, to confer enhanced levels of fitness. As such, many of the bacterial sequences in both Families 5 and 6 cluster alongside archaeal or eukaryotic sequences, or amongst protein sequences belonging to very distant phylogenetic groups.

7.3: Eukaryota

Families 5 and 6 consisted of forty-seven and eight eukaryotic sequences, respectively (Tables 2 and 5). These proteins were amongst the longest all of the sequences representing the Copper and Heavy Metal Families of the P-type ATPases. Eukaryotic sequences representing the Copper P-type ATPases were found in two different clusters, cluster 13 and cluster 14. Cluster 13 exclusively contained eukaryotic proteins, and cluster 14, which primarily consisted of bacterial sequences, contained two eukaryotic sequences from Viridiplantae. Cluster 13 was broken down into several sub-clusters. The sub-cluster representing nine sequences from

Viridiplantae included two sequences from each of the genera *Arabidopsis* and *Sorghum*, one from the genus *Zea,* and three from the genus *Oryza.* The sub-clusters containing Fungal homologues were from a total of fourteen sequences, two from each of the genera *Aspergillus* and *Gibberella*, and one from each of the genera *Schizosaccharomyces, Yarrowia, Cryptococcus, Glomerella, Kluyceromyces, Magnaporthe, Neurospora, Saccharomyces, Trametes,* and *Ustilago*. The two proteins from the sub-cluster containing Mycetozoa were both from the genus *Dictyostelium.* Lastly, a collection of twenty sequences from animal homologues in several adjacent sub-clusters in the phylogenetic tree, were from the following genera; two from each of *Anopheles, Caenorhabitus, Canis, Rattus,* and *Mus,* one each from *Drosophila, Cricetulus*, *Danio, Ovis,* and *Tetradon,* and three from *Homo.* The two eukaryotic enzymes found in cluster 14 were both from Viridiplantae. These sequences were from the genera *Arabidopsis* and *Oryza*, and were found clustered together with Cyanobacteria. The clustering of sequences from Viridiplantae with Cyanobacteria was not unexpected, as the evolution of Chloroplasts from Cyanobacteria has previously been supported by phylogenetic evidence.

As in Family 5, the eukaryotic sequences representing the Heavy Metal P-type ATPases were divided into two clusters, cluster 6 and cluster 8. The sequences found in cluster 6 were entirely from Viridiplantae, with one from each of the genera *Ostreococcus, Medicago* and *Thlaspi,* and three from the genus *Arabidopsis*. The remaining two eukaryotic sequences were found in a sub-cluster of cluster 8. Although there were homologues from Firmicutes and Chlamydiae within cluster 8, there were no bacterial sequences within this sub-cluster. The two eukaryotic proteins in cluster 8

were both from Viridiplantae, with one from the genus *Arabidopsis* and the other from the genus *Oryza*.

Discussion

P-type ATPases, which are classified the 3.A.3 Superfamily (TCDB), are a diverse phylogenetic group of protein pumps that are involved in the translocation of many different substrates (ions) in archaeal, bacterial, and eukaryotic organisms, as well as phospholipids in eukaryotes. Despite their differences, most possess nine well-conserved sequence motifs, and all are dependent on the phosphorylation of a specific aspartate residue (Møller J.V., et al., 1995).

Phylogenetic analysis of the proteins representing the different families within the 3.A.3 Superfamily revealed that most of the clustering observed in Figure 1.A correlated with similarities in substrate type. Such clustering patterns were expected, as previous research has indicated that substrate type, as opposed to type of organism, plays a more substantial role in determining sequence similarity amongst P-type ATPases (Axelsen and Palmgren; 1998). Twenty-four functionally unclassified families of P-type ATPases were recently added to the Transport Classification Database. Some of these families appeared to cluster closely to sequences with known functionalities, thus indicating that they may share functional aspects with these characterized P-type ATPases. However, most of the families of functionally unclassified P-type ATPases clustered independently from both other functionally

uncharacterized families, as well as from functionally characterized families of P-type ATPases.

A total of 385 protein sequences representing the Copper P-type ATPase family, Family 5, and 311 protein sequences representing the Heavy Metal P-type ATPase family, Family 6, were collected and separately analyzed. The methodologies used to analyze these two families were essentially the same, with the proteins of Family 5 being subdivided into twenty smaller clusters and the proteins of Family 6 being subdivided into seventeen smaller clusters. These clusters were created based on observed branching patterns in each phylogenetic tree, so as to facilitate in-depth analyses of each family. Protein sequences from each cluster of Families 5 and 6 that contained more than one sequence were multiply aligned and manually examined to determine whether or not they exhibited nine well-conserved motifs, some of which are already known to play important roles (Møller J.V., et al., 1995; Anthonisen A.N., et al., 2006). Overall, either the actual amino acids predicted for each of these motifs or a consistent pattern of secondary structure predictions were made by the SOMPA program at the location of these motifs for every cluster in both Families 5 and 6. The conservation of these motifs' primary and secondary structures suggests that changes in the primary structure at these regions may be carefully protected against, from an evolutionary standpoint, as they could be important for structure. Such changes may destroy the protein's ability to selectively transport a given substrate across a biological membrane, and thus could be incredibly disadvantageous to an organism's fitness.

In addition to motif and SOPMA analyses, the AveHAS, WHAT, HMMTop, and EMBOSS Pepwheel programs were used to further analyze the sequences. These programs indicated that most clusters within these two families exhibited a consistent pattern of eight TMSs, as expected for type I P-type ATPases. These programs also showed that sequences in both Families 5 and 6, on average, exhibited a strong amphipathic region between motifs 2 and 3. Although the actual implication for this occurrence is not yet known, it is possible that it somehow enhances the flexibility of the TGES loop, which has been proposed to radically change its conformation during the transition between phosphorylated and dephosphorylated states (Anthonisen A.N., et al., 2006).

While P-type ATPases encompass a diverse and complex group of proteins, this research indicates that each family contains certain sets of identifiable characteristics that may help elucidate their functional characteristics. Perhaps the differences that distinguish different families of P-type ATPases from one another will be identified by experimentally testing regions of sequences that are only conserved amongst proteins from a given family, rather than throughout all of the 3.A.3 Superfamily.

Appendix

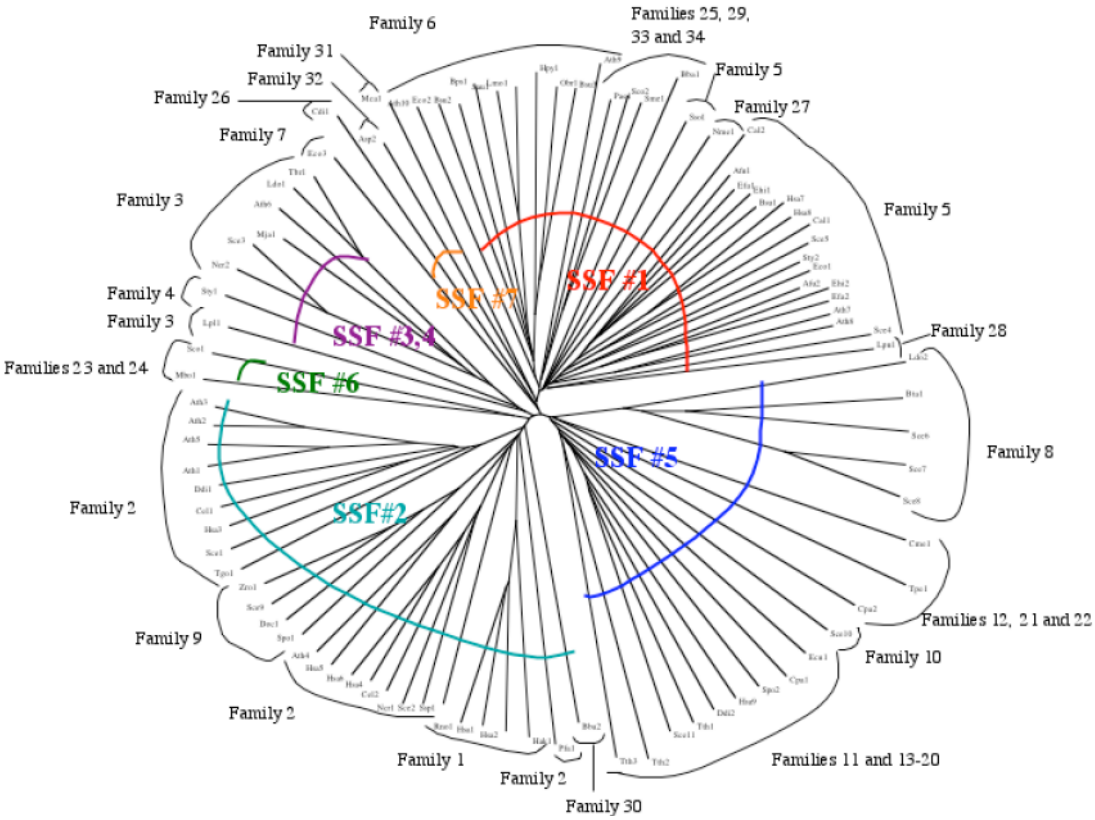

Figure 1.A: The phylogenetic tree of the 3.A.3 Superfamily. The clustering patterns of the families and sub-superfamilies within the 3.A.3 Superfamily are indicated within the figure.

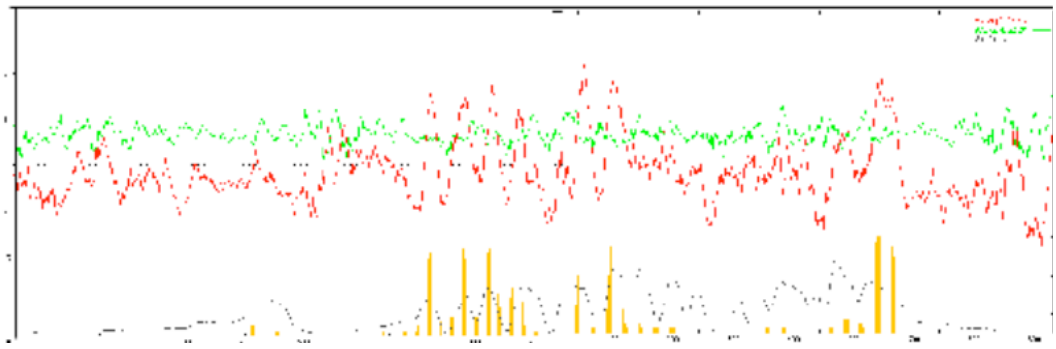

Figure 1.B.1: The AveHAS plot representing the sequences within Sub-Superfamily 1 (SSFI).

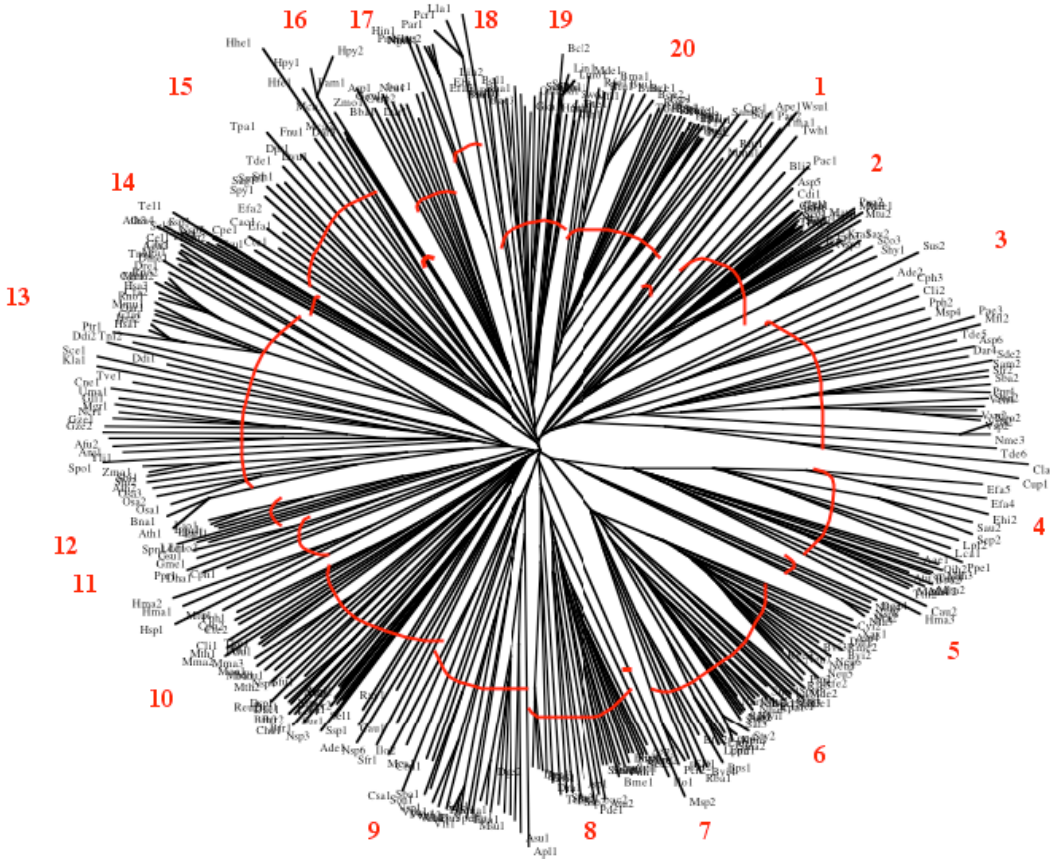

Figure 2.A: The phylogenetic tree of the 385 protein sequences representing the Copper P-type ATPase Family. The locations of the twenty clusters representing these sequences are indicated in the figure.

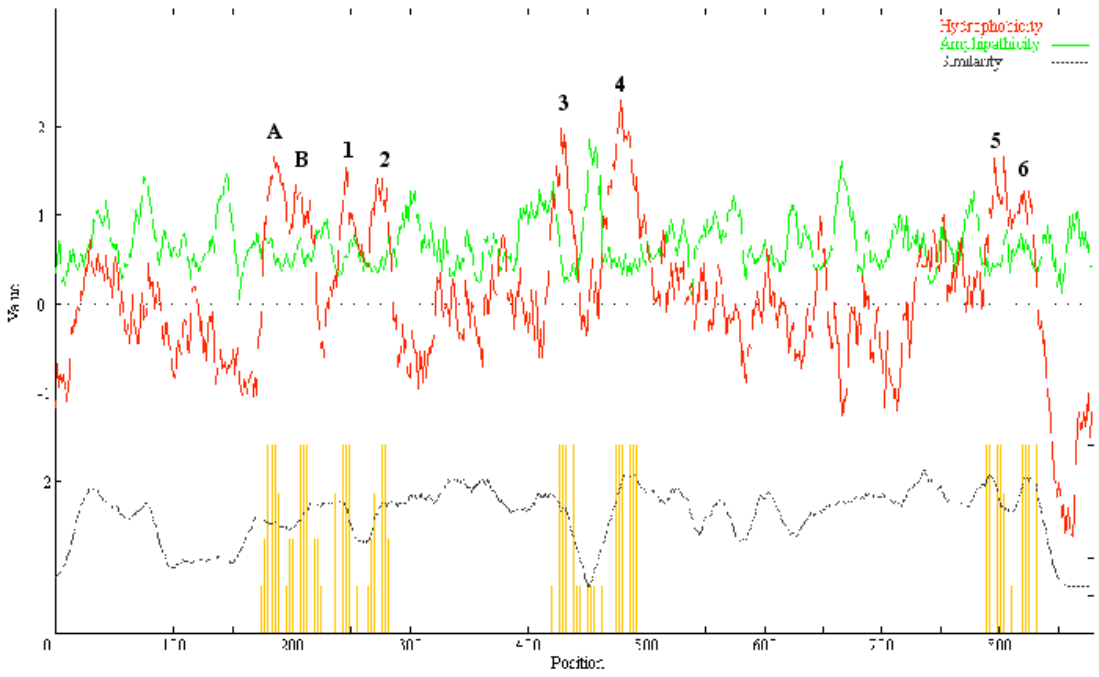

Figure 3.A.1: The AveHAS plot for sequences representing the Copper P-type ATPase Family from cluster1.

Figure 3.B.1: The HMMTop plot for the protein sequence Tma1. This sequence was taken from cluster 1 of the sequences representing the Copper P-type ATPase Family.

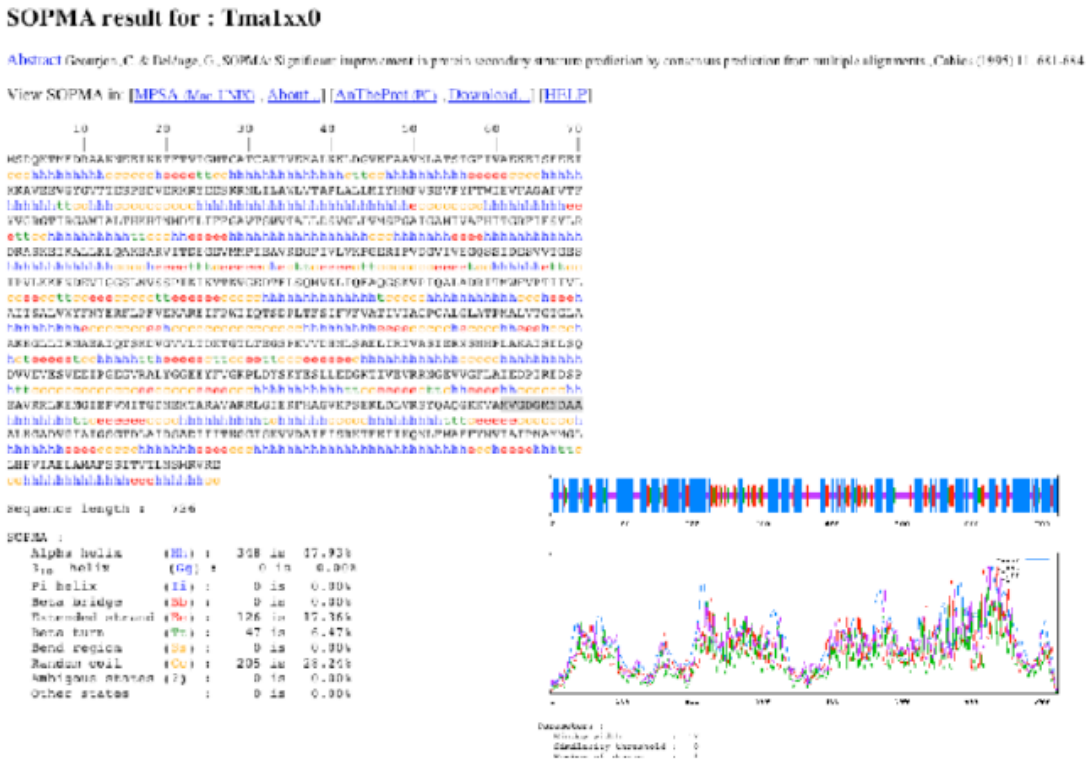

Figure 3.C.1: The SOPMA plot for the protein sequence Tma1. This sequence was taken from cluster 1 of the sequences representing the Copper P-type ATPase Family.

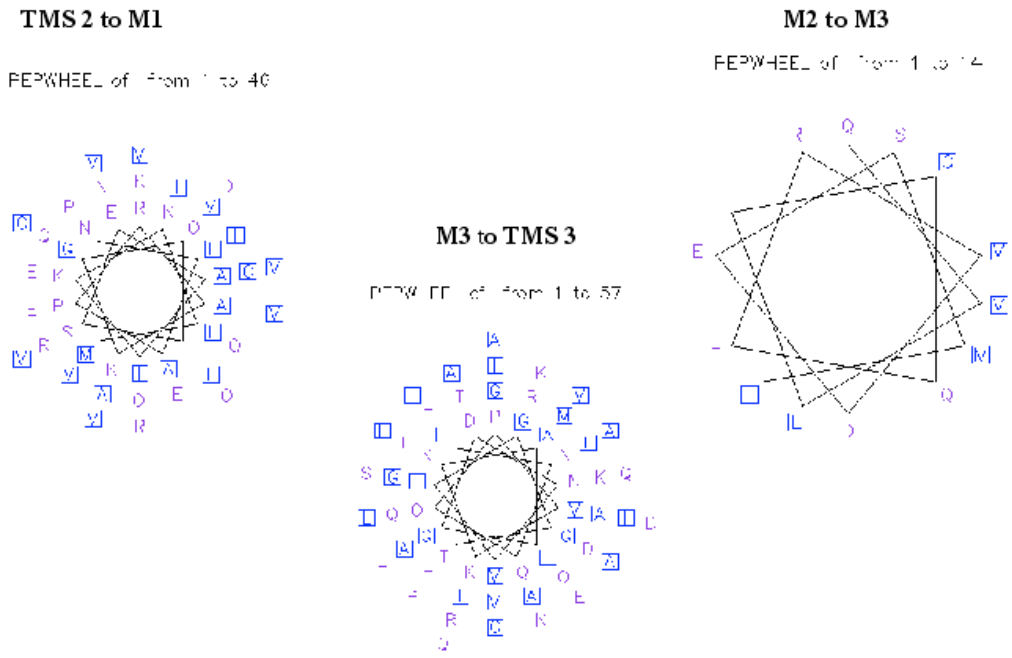

Figure 3.D: The EMBOSS Pepwheel analyses of a segment from the protein sequence Ssu1, from cluster 15. This cluster was determined to have the best amphipathic peak in the region between TMS2 and TMS3 of all of the clusters representing the Copper P-type ATPases.

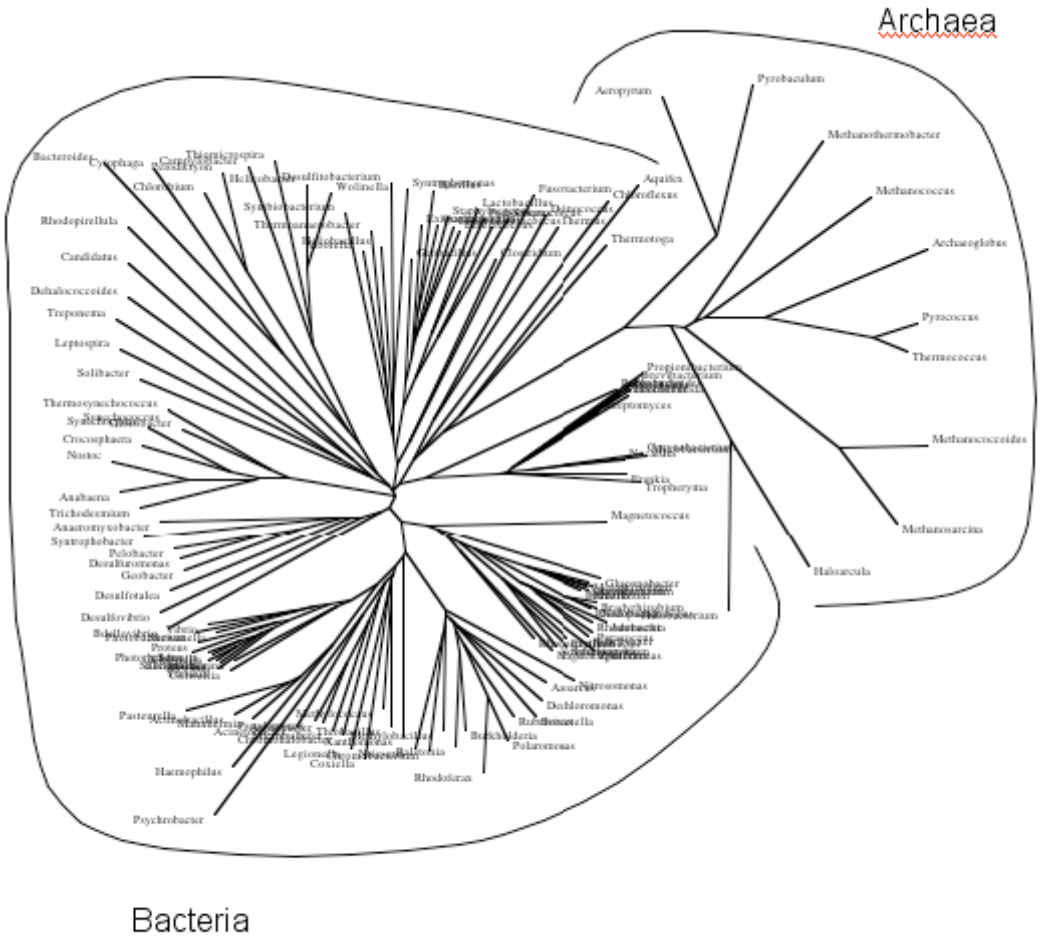

Figure 4. The 16S rRNA phylogenetic tree containing each of the prokaryotic genera found among the sequences representing the Copper P-type ATPases.

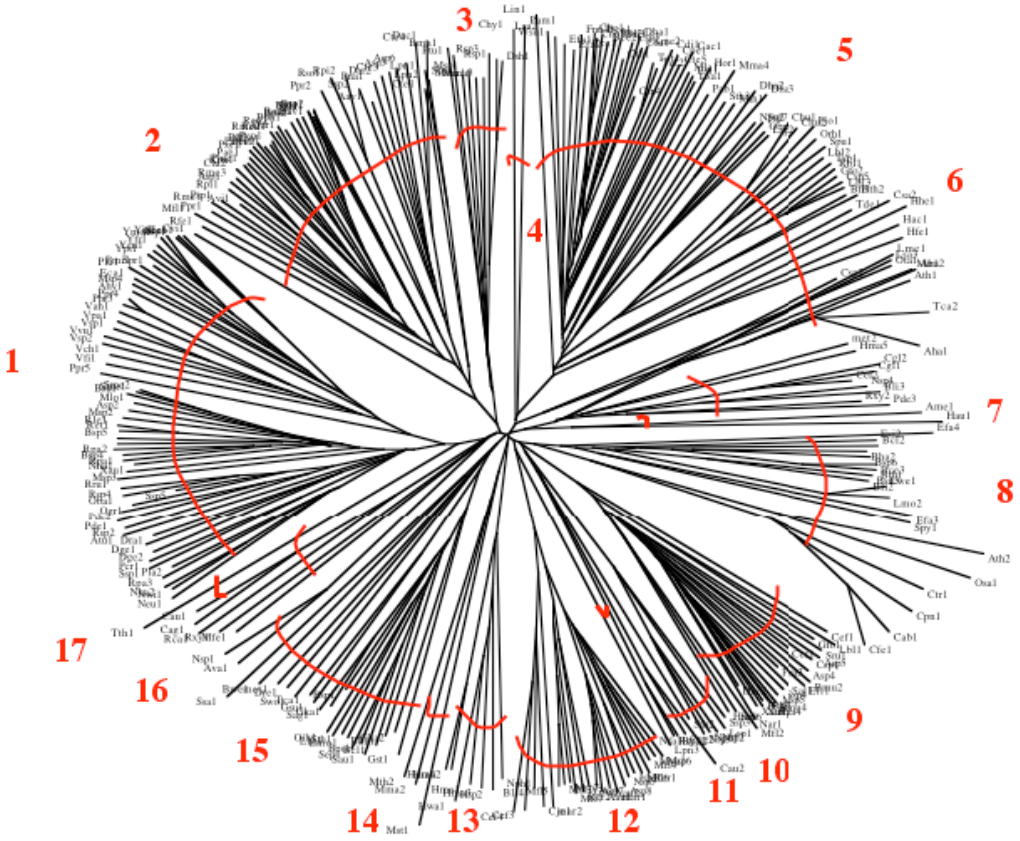

Figure 5.A: The phylogenetic tree of the 311 protein sequences representing the Heavy Metal P-type ATPase Family. The locations of the twenty clusters found within these sequences are indicated in the figure.

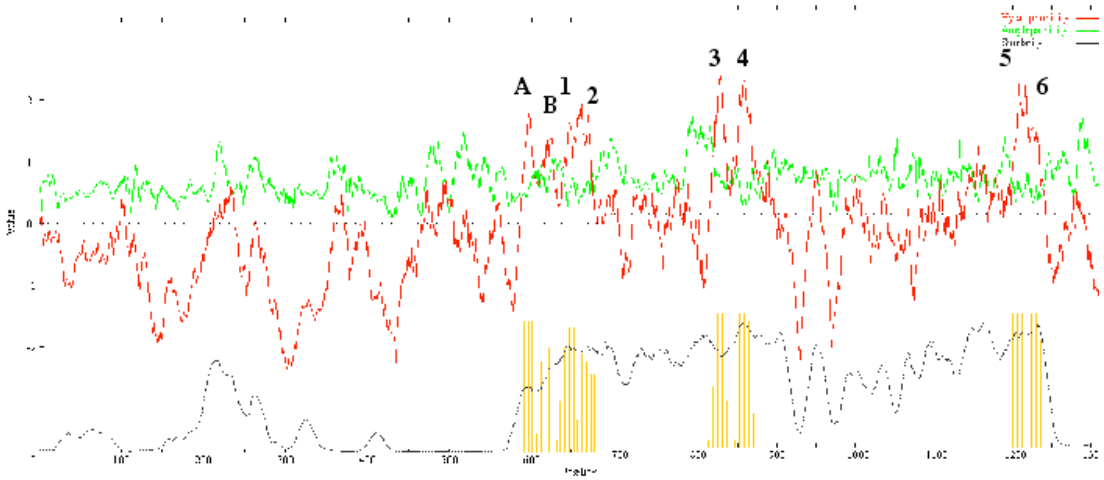

Figure 6.A.1: The AveHAS plot for sequences representing the Heavy Metal P-type ATPase Family from cluster1.

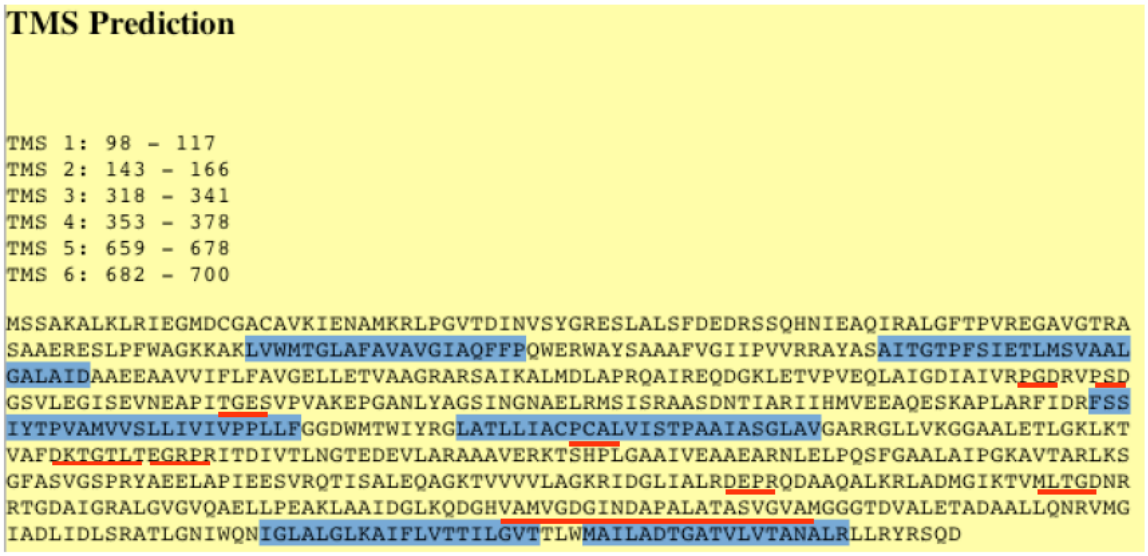

Figure 6.B.1: The HMMTop plot for the protein sequence Neu1. This sequence was taken from cluster 1 of the sequences representing the Copper P-type ATPase Family.

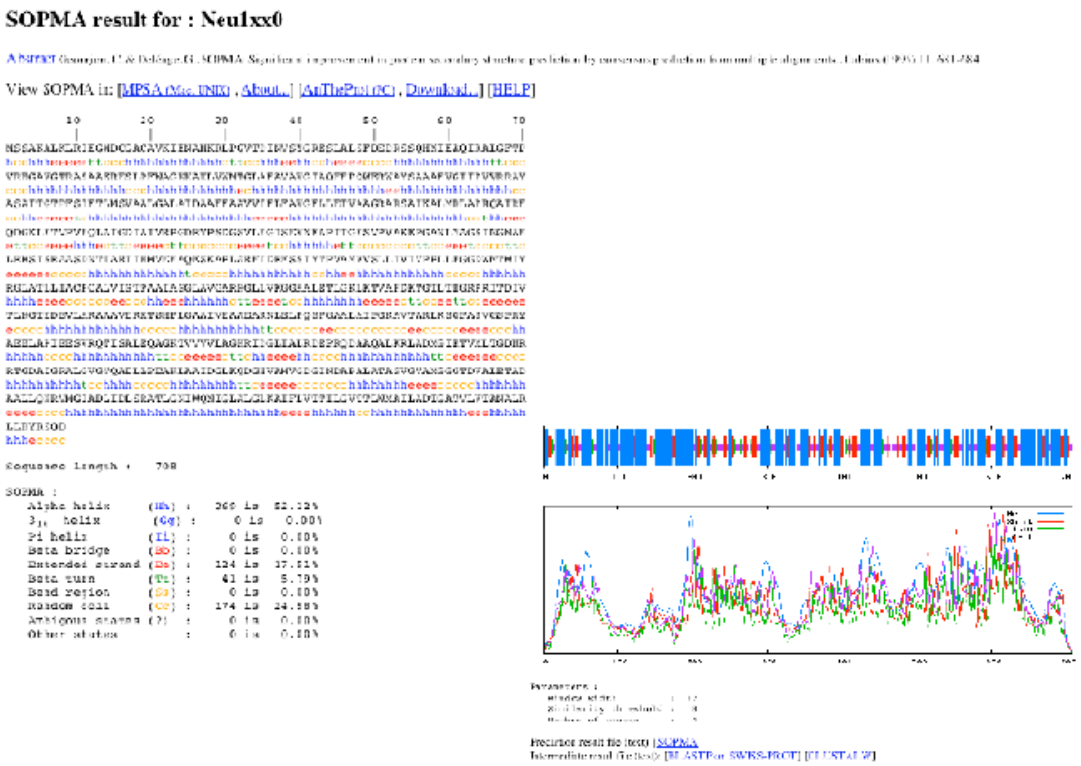

Figure 6.C.1: The SOPMA plot for the protein sequence Neu1. This sequence was taken from cluster 1 of the sequences representing the Heavy Metal P-type ATPase Family.

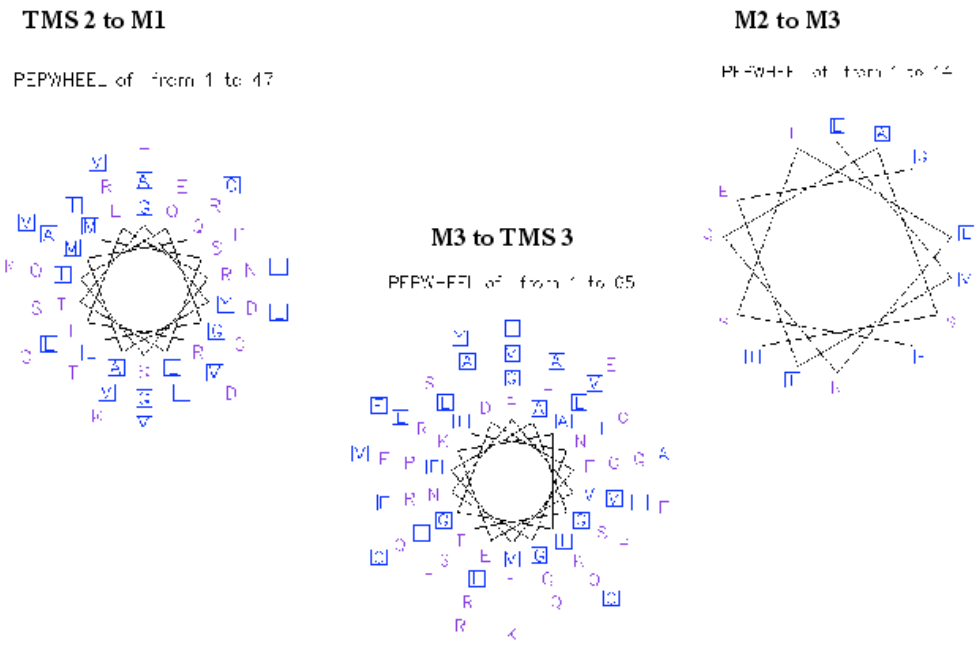

Figure 6.D: The EMBOSS Pepwheel analyses of a segment from the protein sequence Ssp3, from cluster 10. This cluster was determined to have the best amphipathic peak in the region between TMS2 and TMS3 of all of the clusters representing the Heavy Metal P-type ATPases.

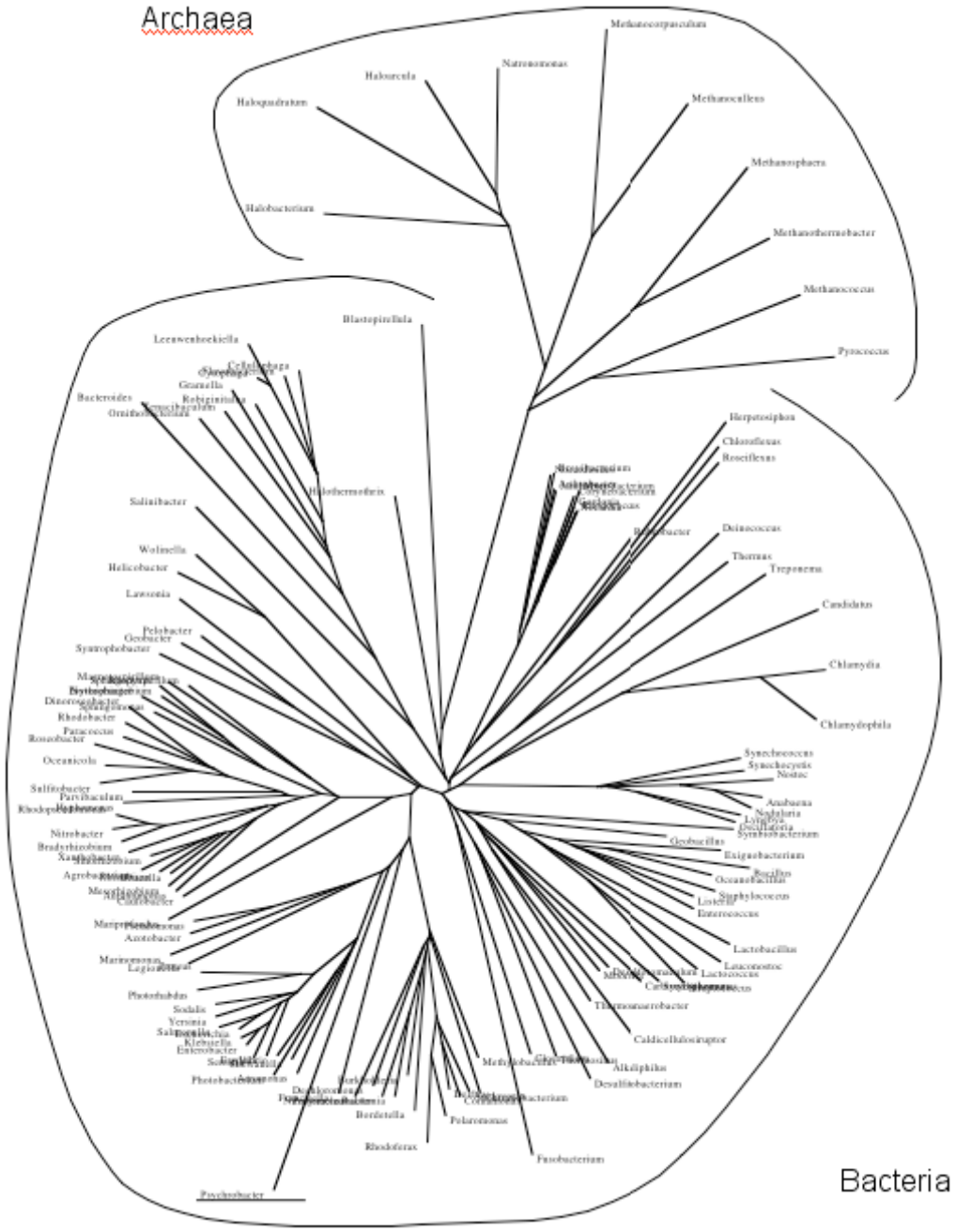

Figure 7. The 16S rRNA phylogenetic tree containing each of the prokaryotic genera found among the sequences representing the Copper P-type ATPases.

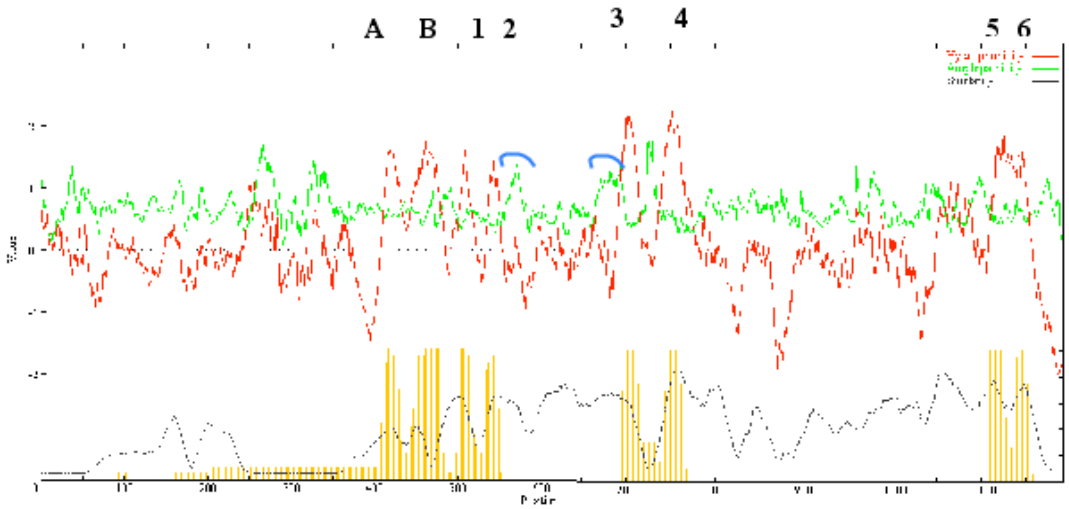

Figure 8. An AveHAS plot containing one sequence from each of the twenty clusters representing the Copper P-type ATPases. The two broadest amphipathic peaks between TMS2 and TMS3 are marked.

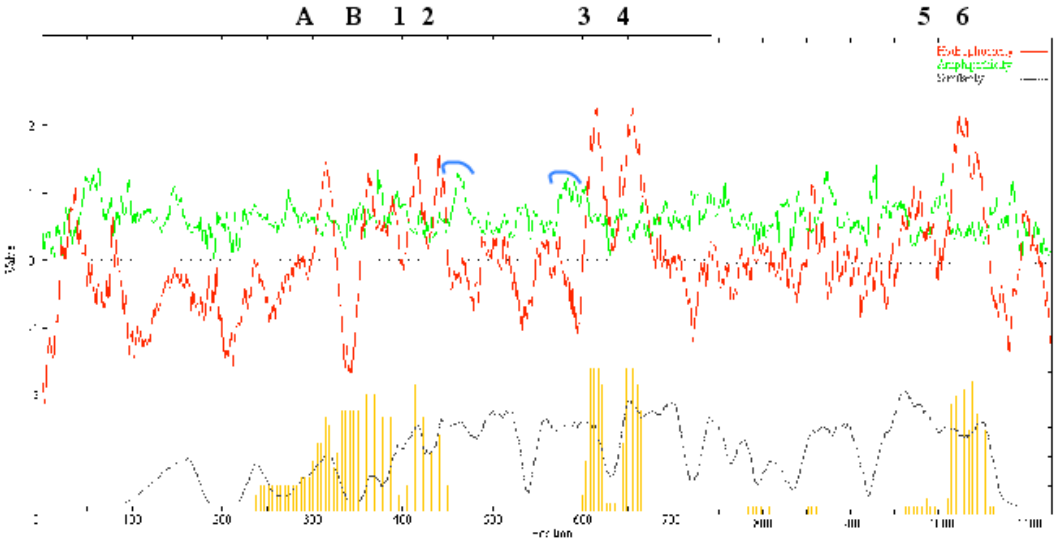

Figure 9. An AveHAS plot containing one sequence from each of the seventeen clusters representing the Heavy Metal P-type ATPases. The two broadest amphipathic peaks between TMS2 and TMS3 are marked.

Table 1. The Protein sequences of the Families within the 3.A.3 Superfamily in TCDB.

	Abbr.	SequenceDescription	Organism	Group	#AA	GI#	Group	Kingdom
1	Hsa1	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (Sodium pump subunit alpha-1) (Na(+)/K(+) ATPase alpha-1 subunit)	Homo sapiens	3.A.3.1.1	1023	114374	Metazoa	Eukaryota
2	Hsa2	Potassium-transporting ATPase alpha chain 1 (Proton pump) (Gastric H(+)/K(+) ATPase subunit alpha)	Homo sapiens	3.A.3.1.2	1035	148877240	Metazoa	Eukaryota
3	Hak1	Na-ATPase	Heterosigma akashiwo	3.A.3.1.3	1330	75213257	Raphidophyceae	Eukaryota
4	Rno1	Potassium-transporting ATPase alpha chain 2 (Proton pump) (Non-gastric H(+)/K(+) ATPase subunit alpha)	Rattus norvegicus	3.A.3.1.4	1036	1703464	Metazoa	Eukaryota
5	Hsa3	Plasma membrane calcium-transporting ATPase 4 (PMCA4) (Plasma membrane calcium pump isoform 4) (Plasma membrane calcium ATPase isoform 4) (Matrix-	Homo sapiens	3.A.3.2.1	1241	14286105	Metazoa	Eukaryota
6	Sce1	Calcium-transporting ATPase 2 (Vacuolar Ca(2+)-ATPase)	Saccharomyces cerevisiae	3.A.3.2.2	1173	728904	Fungi	Eukaryota
7	Sce2	Calcium-transporting ATPase 1 (Golgi Ca(2+)-ATPase)	Saccharomyces cerevisiae	3.A.3.2.3	950	114301	Fungi	Eukaryota
8	Ssp1	Cation-transporting ATPase pma1	Synechocystis sp. PCC 6803	3.A.3.2.4	905	2506205	Cyanobacteria	Bacteria
9	Hsa4	Calcium-transporting ATPase type 2C member 1 (ATPase 2C1) (ATP-dependent Ca(2+)	Homo sapiens	3.A.3.2.5	919	68068024	Metazoa	Eukaryota
10	Ncr1	Putative calcium P-type ATPase (Hypothetical protein	Neurospora crassa	3.A.3.2.6	1025	74698463	Fungi	Eukaryota
11	Hsa5	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (Calcium pump 2) (SERCA2) (SR Ca(2+)-ATPase 2) (Calcium-transporting ATPase sarcoplasmic reticulum type, slow twitch skeletal muscle isoform) (Endoplasmic reticulum class 1/2 Ca(2+) ATPase)	Homo sapiens	3.A.3.2.7	1042	114312	Metazoa	Eukaryota
12	Pfa1	P-type ATPase	Plasmodium falciparum	3.A.3.2.8	1228	74967369	Apicomplexa	Eukaryota
13	Hsa6	Calcium-transporting ATPase type 2C member 2 (ATPase	Homo sapiens	3.A.3.2.9	963	19924283	Metazoa	Eukaryota
14	Ath1	Calcium-transporting ATPase 8, plasma membrane-type (Ca(2+)-ATPase isoform 8)	Arabidopsis thaliana	3.A.3.2.10	1074	12643246	Viridiplantae	Eukaryota
15	Ath2	Calcium-transporting ATPase 1, chloroplast precursor (Ca(2+)-ATPase isoform 1) (Plastid envelope ATPase 1)	Arabidopsis thaliana	3.A.3.2.11	1020	30316378	Viridiplantae	Eukaryota
16	Ath3	Calcium-transporting ATPase 2, plasma membrane-type (Ca(2+)-ATPase isoform 2)	Arabidopsis thaliana	3.A.3.2.12	1014	12229639	Viridiplantae	Eukaryota
17	Ath4	Calcium-transporting ATPase 1, endoplasmic reticulum-type	Arabidopsis thaliana	3.A.3.2.13	1061	12643704	Viridiplantae	Eukaryota
18	Ath5	Calcium-transporting ATPase 9, plasma membrane-type (Ca(2+)-ATPase isoform 9)	Arabidopsis thaliana	3.A.3.2.14	1086	150421517	Viridiplantae	Eukaryota
19	Cel1	Calcium ATPase (Hypothetical protein mca-1)	Caenorhabditis elegans	3.A.3.2.15	1228	74763859	Metazoa	Eukaryota
20	Cel2	PMR1 protein (Hypothetical protein pmr-1)	Caenorhabditis elegans	3.A.3.2.16	901	75028081	Metazoa	Eukaryota
21	Ddi1	Probable calcium-transporting ATPase PAT1	Dictyostelium discoideum	3.A.3.2.17	1115	1703456	none	Eukaryota
22	Tgo1	Ca2+-ATPase	Toxoplasma gondii	3.A.3.2.18	1405	75023636	Apicomplexa	Eukaryota
23	Ncr2	Plasma membrane ATPase (Proton pump)	Neurospora crassa	3.A.3.3.1	920	114347	Fungi	Eukaryota

24	Ldo1	Probable proton ATPase 1A (LdH1A)	Leishmania donovani	3.A.3.3.2	974	20981683	Trypanosomatidae	Eukaryota
25	Lpl1	Cadmium/manganese transport ATPase	Lactobacillus plantarum	3.A.3.3.3	758	81325414	Firmicutes	Bacteria
26	Mja1	Putative cation-transporting ATPase MJ1226	Methanocaldococcus jannaschii	3.A.3.3.4	805	47606650	Euryarchaeota	Archaea
27	Tbr1	P-type H+-ATPase	Trypanosoma brucei	3.A.3.3.5	912	75013788	Trypanosomatidae	Eukaryota
28	Sce3	Plasma membrane ATPase 1 (Proton pump 1)	Saccharomyces cerevisiae	3.A.3.3.6	918	1168544	Fungi	Eukaryota
29	Ath6	ATPase 1, plasma membrane-type (Proton pump 1)	Arabidopsis thaliana	3.A.3.3.7	949	12644156	Viridiplantae	Eukaryota
30	Sty1	Magnesium-transporting ATPase, P-type 1 (Mg(2+) transport ATPase, P-type 1)	Salmonella typhimurium	3.A.3.4.1	902	543864	Gammaproteobacteria	Bacteria
31	Ehi1	Probable copper-importing ATPase A	Enterococcus hirae	3.A.3.5.1	727	416665	Firmicutes	Bacteria
32	Ehi2	Probable copper exporting ATPase B	Enterococcus hirae	3.A.3.5.2	745	416666	Firmicutes	Bacteria
33	Hsa7	Copper-transporting ATPase 2 (Copper pump 2) (Wilson disease-associated protein)	Homo sapiens	3.A.3.5.3	1465	84028176	Metazoa	Eukaryota
34	Sty2	Putative cation-transporting P-type ATPase	Salmonella typhimurium	3.A.3.5.4	824	13633955	Gammaproteobacteria	Bacteria
35	Eco1	Copper-transporting P-type ATPase	Escherichia coli	3.A.3.5.5	834	2493016	Gammaproteobacteria	Bacteria
36	Hsa8	Copper-transporting ATPase 1 (Copper pump 1) (Menkes disease-associated protein)	Homo sapiens	3.A.3.5.6	1500	1351993	Metazoa	Eukaryota
37	Afu1	Cation-transporting ATPase, P-type (PacS)	Archaeoglobus fulgidus	3.A.3.5.7	804	74549566	Euryarchaeota	Archaea
38	Cal1	Copper-transporting P-type ATPase	Candida albicans	3.A.3.5.8	1204	74623612	Fungi	Eukaryota
39	Cal2	Copper resistance-associated P-type ATPase	Candida albicans	3.A.3.5.9	1197	74698483	Fungi	Eukaryota
40	Afu2	Copper-transporting ATPase, P-type (CopB)	Archaeoglobus fulgidus	3.A.3.5.10	690	74514552	Euryarchaeota	Archaea
41	Ath7	Putative copper-transporting ATPase PAA1	Arabidopsis thaliana	3.A.3.5.11	949	12643855	Viridiplantae	Eukaryota
42	Ath8	Paa2 P-type ATPase	Arabidopsis thaliana	3.A.3.5.12	883	75145757	Viridiplantae	Eukaryota
43	Sso1	Copper-transporting ATPase	Sulfolobus solfataricus	3.A.3.5.13	695	74538531	Crenarchaeota	Archaea
44	Sce4	Probable copper-transporting ATPase (Cu(2+)-ATPase)	Saccharomyces cerevisiae	3.A.3.5.14	1216	584790	Fungi	Eukaryota
45	Efa1	TcrA	Enterococcus faecium	3.A.3.5.15	811	122613057	Firmicutes	Bacteria
46	Efa2	TcrB	Enterococcus faecium	3.A.3.5.16	710	75404579	Firmicutes	Bacteria
47	Sce5	Copper-transporting ATPase (Cu(2+)-ATPase)	Saccharomyces cerevisiae	3.A.3.5.17	1004	728935	Fungi	Eukaryota
48	Bsu1	Copper-transporting P-type ATPase copA (Protein copA)	Bacillus subtilis	3.A.3.5.18	803	7531047	Firmicutes	Bacteria
49	Sau1	Probable cadmium-transporting ATPase (Cadmium efflux	Staphylococcus aureus	3.A.3.6.1	727	115414	Firmicutes	Bacteria
50	Eco2	Lead, cadmium, zinc and mercury-transporting ATPase	Escherichia coli	3.A.3.6.2	732	586655	Gammaproteobacteria	Bacteria
51	Hpy1	Cadmium, zinc and cobalt-transporting ATPase	Helicobacter pylori	3.A.3.6.3	686	2493007	Epsilonproteobacteria	Bacteria
52	Bps1	putative heavy metal resistance membrane ATPase	Burkholderia pseudomallei K96243	3.A.3.6.4	836	53720986	Betaproteobacteria	Bacteria
53	Obr1	P type ATPase BXA1	Oscillatoria brevis	3.A.3.6.5	660	81323132	Cyanobacteria	Bacteria
54	Ath9	Putative cadmium/zinc-transporting ATPase HMA1, chloroplast precursor	Arabidopsis thaliana	3.A.3.6.6	819	12643808	Viridiplantae	Eukaryota
55	Ath10	Putative cadmium/zinc-transporting ATPase 3	Arabidopsis thaliana	3.A.3.6.7	951	12229675	Viridiplantae	Eukaryota
56	Lmo1	Probable cadmium-transporting ATPase (Cadmium efflux	Listeria monocytogenes	3.A.3.6.8	711	3121832	Firmicutes	Bacteria
57	Bsu2	YkvW protein	Bacillus subtilis	3.A.3.6.9	637	81815533	Firmicutes	Bacteria
58	Bsu3	YvgW protein	Bacillus subtilis	3.A.3.6.10	702	81815569	Firmicutes	Bacteria

59	Eco3	Potassium-transporting ATPase A chain (Potassium-translocating ATPase A chain) (ATP phosphohydrolase A chain) (Potassium-binding and translocating subunit A)	Escherichia coli	3.A.3.7.1	682	2506206	Gammaproteo-bacteria	Bacteria
60	Bta1	Probable phospholipid-transporting ATPase IA (Chromaffin granule ATPase II) (ATPase class I type 8A member 1)	Bos taurus	3.A.3.8.1	1149	8134328	Metazoa	Eukaryota
61	Sce6	Probable phospholipid-transporting ATPase DRS2	Saccharomyces cerevisiae	3.A.3.8.2	1355	728905	Fungi	Eukaryota
62	Ldo2	Putative miltefosine transporter	Leishmania donovani	3.A.3.8.3	1097	75008488	Trypanosomatid ae	Eukaryota
63	Sce7	Probable phospholipid-transporting ATPase DNF1	Saccharomyces cerevisiae	3.A.3.8.4	1571	728906	Fungi	Eukaryota
64	Sce8	Probable phospholipid-transporting ATPase DNF2	Saccharomyces cerevisiae	3.A.3.8.5	1612	2493010	Fungi	Eukaryota
65	Sce9	Sodium transport ATPase 1	Saccharomyces cerevisiae	3.A.3.9.1	1091	114302	Fungi	Eukaryota
66	Spo1	Calcium-transporting ATPase 3	Schizosaccharomyces pombe	3.A.3.9.2	1037	114303	Fungi	Eukaryota
67	Doc1	P-type ATPase 2	Debaryomyces occidentalis	3.A.3.9.3	1082	74675873	Fungi	Eukaryota
68	Zro1	Na+-ATPase	Zygosaccharomyces rouxii	3.A.3.9.4	1048	74676231	Fungi	Eukaryota
69	Sce10	Probable cation-transporting ATPase 1	Saccharomyces cerevisiae	3.A.3.10.1	1215	731415	Fungi	Eukaryota
70	Ecu1	CATION-TRANSPORTING ATPase	Encephalitozoon cuniculi	3.A.3.11.1	1146	74697535	Fungi	Eukaryota
71	Cme1	Functionally uncharacterized P-type ATPase family 12 (FUPA12) (one member; 1998	Cyanidioschyzon merolae 10D	3.A.3.12.1	1998	*not in NCBI: CMR432C	* some type of Protozoan	Eukaryota
72	Hsa9	Probable cation-transporting ATPase 13A2	Homo sapiens	3.A.3.13.1	1180	14285364	Metazoa	Eukaryota
73	Sce11	Probable cation-transporting ATPase 2	Saccharomyces cerevisiae	3.A.3.14.1	1472	2493012	Fungi	Eukaryota
74	Ddi2	Hypothetical protein	Dictyostelium discoideum AX4	3.A.3.15.1	1158	74997157	none	Eukaryota
75	Tth1	E1-E2 ATPase family protein	Tetrahymena thermophila SB210	3.A.3.16.1	1328	121979716	Oligohymenoph orea	Eukaryota
76	Spo2	Probable cation-transporting ATPase C29A4.19c	Schizosaccharomyces pombe	3.A.3.17.1	1096	6707665	Fungi	Eukaryota
77	Cpa1	cation-transporting ATPase 2 with 8 transmembrane domains	Cryptosporidium parvum Iowa II	3.A.3.18.1	1491	66359870	Apicomplexa	Eukaryota
78	Tth2	E1-E2 ATPase family protein	Tetrahymena thermophila SB210	3.A.3.19.1	1807	118355868	Oligohymenoph orea	Eukaryota
79	Tth3	E1-E2 ATPase family protein	Tetrahymena thermophila SB210	3.A.3.20.1	1072	121971402	Oligohymenoph orea	Eukaryota
80	Tpe1	Functionally uncharacterized P-type ATPase family 21	Thalassiosira pseudonana	3.A.3.21.1	1372	*not in NCBI: ORF00905	* some type of Protozoan	Eukaryota
81	Cpa2	Possible MgtA, cation transport ATpase, signal peptide, 12 transmembrane domains	Cryptosporidium parvum	3.A.3.22.1	1434	74777644	Apicomplexa	Eukaryota
82	Sco1	Cation-transporting ATPase	Streptomyces coelicolor	3.A.3.23.1	802	81858200	Actinobacteria	Bacteria
83	Mbo1	Cation-transporting ATPase	Mycobacterium bovis	3.A.3.24.1	1625	81835012	Actinobacteria	Bacteria
84	Sco2	Putative cation-transporting ATPase	Streptomyces coelicolor	3.A.3.25.1	776	81858708	Actinobacteria	Bacteria
85	Cdi1	Putative cation transport protein	Corynebacterium diphtheriae	3.A.3.26.1	841	81401907	Actinobacteria	Bacteria
86	Nme1	Cation transport ATPase, E1-E2 family	Neisseria meningitidis serogroup B	3.A.3.27.1	823	81784642	Betaproteobacte ria	Bacteria
87	Lpn1	Heavy metal transporting P-type ATPase, cation transporting	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	3.A.3.28.1	852	81378261	Gammaproteo-bacteria	Bacteria
88	Bba1	Cation transport ATPase, E1-E2 family precursor	Bdellovibrio bacteriovorus	3.A.3.29.1	798	81616796	Deltaproteo-bacteria	Bacteria

89	Bba2	Cation-transporting ATPase	Bdellovibrio bacteriovorus	3.A.3.30.1	825	81829271	Deltaproteo-bacteria	Bacteria
90	Mca1	heavy metal translocating P-type ATPase	Methylococcus capsulatus str. Bath	3.A.3.31.1	839	53804062	Gammaproteo-bacteria	Bacteria
91	Asp2	Putative cation transport ATPase	Azoarcus sp. EbN1	3.A.3.32.1	694	81358540	Betaproteo-bacteria	Bacteria
92	Sme1	Putative cation transport P-type ATPase	Sinorhizobium meliloti	3.A.3.33.1	746	81854505	Alphaproteo-bacteria	Bacteria
93	Pae1	Probable cation-transporting P-type ATPase	Pseudomonas aeruginosa	3.A.3.33.4	661	81857196	Gammaprote-obacteria	Bacteria

Table 2. The protein sequences representing the Copper P-type ATPases

	Abbr.	Sequence Description	Organism	#AA	GI#	Group	King	Cluster#	Ave. AA length for cluster (2 decimal places)
1	Tma1	cation-transporting ATPase, P-type	Thermotoga maritima MSB8	726	4980818	Thermotogae	B	1	770.5 ± 51.53
2	Wsu1	CATION-TRANSPORTING ATPASE (P-TYPE)	Wolinella succinogenes	732	34482565	Epsilonproteo-bacteria	B	1	
3	Ape1	cation-transporting ATPase	Aeropyrum pernix K1	835	14601418	Crenarchaeota	A	1	
4	Pae2	cation-transporting ATPase (P-type)	Pyrobaculum aerophilum str. IM2	789	18313568	Crenarchaeota	A	1	
5	Twh1	metal cation-transporting ATPase	Tropheryma whipplei TW08/27	716	28572954	Actinobacteria (HighGC G+)	B	2	769.39 ± 53.34
6	Bli2	COG2217: Cation transport ATPase	Brevibacterium linens BL2	735	62425502	Actinobacteria (HighGC G+)	B	2	
7	Pac1	cation-transporting ATPase	Propionibacterium acnes KPA171202	752	50843675	Actinobacteria (HighGC G+)	B	2	
8	Asp5	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Arthrobacter sp. FB24	779	66963956	Actinobacteria (HighGC G+)	B	2	
9	Cdi1	Putative cation-transporting P-type ATPase	Corynebacterium diphtheriae NCTC 13129	743	38234821	Actinobacteria (HighGC G+)	B	2	
10	Cef1	putative cation-transporting ATPase	Corynebacterium efficiens YS-314	757	23492243	Actinobacteria (HighGC G+)	B	2	
11	Cgl1	Cation transport ATPases	Corynebacterium glutamicum ATCC 13032	755	21323151	Actinobacteria (HighGC G+)	B	2	
12	Cje1	putative cation-transporting P-type ATPase	Corynebacterium jeikeium K411	706	68263309	Actinobacteria (HighGC G+)	B	2	
13	Mav1	CtpA	Mycobacterium avium subsp. paratuberculosis K-10	742	41410382	Actinobacteria (HighGC G+)	B	2	
14	Fsp4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Frankia sp. Ccl3	996	68172203	Actinobacteria (HighGC G+)	B	2	
15	Aau1	metal transporter ATPase	Arthrobacter aurescens	773	42558790	Actinobacteria (HighGC G+)	B	2	
16	Fsp1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Frankia sp. EAN1pec	792	68232964	Actinobacteria (HighGC G+)	B	2	
17	Sco1	putative metal transporter ATPase	Streptomyces coelicolor A3(2)	753	8894835	Actinobacteria (HighGC G+)	B	2	
18	Nfa1	putative cation transporter	Nocardia farcinica IFM 10152	750	54022124	Actinobacteria (HighGC G+)	B	2	
19	Tfu1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Thermobifida fusca YX	752	72163247	Actinobacteria (HighGC G+)	B	2	
20	Sav1	putative cation-transporting P-type ATPase	Streptomyces avermitilis MA-4680	750	29608991	Actinobacteria (HighGC G+)	B	2	
21	Sco2	probable cation-transporting P-type ATPase.	Streptomyces coelicolor A3(2)	760	6714779	Actinobacteria (HighGC G+)	B	2	
22	Fsp2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Frankia sp. EAN1pec	785	68229303	Actinobacteria (HighGC G+)	B	2	
23	Fsp3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Frankia sp. EAN1pec	810	68233340	Actinobacteria (HighGC G+)	B	2	
24	Kra1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Kineococcus radiotolerans SRS30216	867	67988497	Actinobacteria (HighGC G+)	B	2	

25	Nsp5	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nocardioides sp. JS614	785	71366152	Actinobacteria (HighGC G+)	B	2	
26	Sav2	putative cation-transporting P-type ATPase	Streptomyces avermitilis MA-4680	751	29604256	Actinobacteria (HighGC G+)	B	2	
27	Sco3	probable cation-transporting ATPase	Streptomyces coelicolor A3(2)	762	6689151	Actinobacteria (HighGC G+)	B	2	
28	Shy1	unknown	Streptomyces hygroscopicus	762	63033835	Actinobacteria (HighGC G+)	B	2	
29	Pac2	cation-transporting P-type ATPase A	Propionibacterium acnes KPA171202	747	50842178	Actinobacteria (HighGC G+)	B	2	
30	Mtu3	PROBABLE CATION TRANSPORTER P-TYPE ATPASE A CTPA	Mycobacterium tuberculosis H37Rv	761	15607234	Actinobacteria (HighGC G+)	B	2	
31	Mle1	cation-transporting ATPase	Mycobacterium leprae TN	750	15828087	Actinobacteria (HighGC G+)	B	2	
32	Mtu2	cation-transporting ATPase, E1-E2 family	Mycobacterium tuberculosis CDC1551	752	13879590	Actinobacteria (HighGC G+)	B	2	
33	Sus2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Solibacter usitatus Ellin6076	824	67929240	Acidobacteria	B	3	806.04 ± 15.91
34	Ade2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Anaeromyxobacter dehalogenans 2CP-C	807	66858031	Deltaproteo-bacteria	B	3	
35	Cph3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chlorobium phaeobacteroides BS1	807	67939261	Chlorobi	B	3	
36	Cli2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chlorobium limicola DSM 245	809	67918515	Chlorobi	B	3	
37	Pph2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pelodictyon phaeoclathratiforme BU-1	818	68551122	Chlorobi	B	3	
38	Msp4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Magnetococcus sp. MC-1	818	68246176	unclassified Proteobacteria	B	3	
39	Pae3	COG2217: Cation transport ATPase	Pseudomonas aeruginosa UCBPP-PA14	811	53727752	Gammaproteo-bacteria	B	3	
40	Mfl2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methylobacillus flagellatus KT	806	68213218	Betaproteo-bacteria	B	3	
41	Tde5	heavy metal translocating P-type ATPase	Thiobacillus denitrificans ATCC 25259	834	74318062	Betaproteo-bacteria	B	3	
42	Asp6	probable cation transport P-type ATPase	Azoarcus sp. EbN1	817	56478348	Betaproteo-bacteria	B	3	
43	Dar4	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	812	71906350	Betaproteo-bacteria	B	3	
44	Sde2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella denitrificans OS-217	813	69945884	Gammaproteo-bacteria	B	3	
45	Sam2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella amazonensis SB2B	795	68548180	Gammaproteo-bacteria	B	3	
46	Sfr2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella frigidimarina NCIMB 400	797	69952331	Gammaproteo-bacteria	B	3	
47	Sba2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella baltica OS155	799	68543438	Gammaproteo-bacteria	B	3	
48	Son2	cation transport ATPase, E1-E2 family	Shewanella oneidensis MR-1	799	24373906	Gammaproteo-bacteria	B	3	

49	Ppr4	putative cation transport ATPase, E1-E2 family	Photobacterium profundum SS9	769	54309022	Gammaproteo-bacteria	B	3	
50	Vch4	COG2217: Cation transport ATPase	Vibrio cholerae V51	790	75819988	Gammaproteo-bacteria	B	3	
51	Vvu2	cation transport ATPase	Vibrio vulnificus YJ016	789	37679860	Gammaproteo-bacteria	B	3	
52	Vpa2	cation transport ATPase, E1-E2 family	Vibrio parahaemolyticus RIMD 2210633	787	28898313	Gammaproteo-bacteria	B	3	
53	Vsp2	COG2217: Cation transport ATPase	Vibrio sp. Ex25	787	75854754	Gammaproteo-bacteria	B	3	
54	Ngo2	putative P-type cation-transporting ATPase	Neisseria gonorrhoeae FA 1090	818	59801112	Betaproteo-bacteria	B	3	
55	Nme3	putative P-type cation-transporting ATPase	Neisseria meningitidis Z2491	823	7380097	Betaproteo-bacteria	B	3	
56	Tde6	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Thiomicrospira denitrificans ATCC 33889	817	71149739	Epsilonproteo-bacteria	B	3	
57	Cla1	cation-transporting ATPase, P-type (copA)	Campylobacter lari RM2100	781	57241791	Epsilonproteo-bacteria	B	3	
58	Cup1	cation-transporting ATPase, P-type (copA)	Campylobacter upsaliensis RM3195	830	57241905	Epsilonproteo-bacteria	B	3	
59	Ath3	metal-transporting P-type ATPase	Arabidopsis thaliana	949	7270300	Viridiplantae	E	14	825.78 ± 76.00
60	Osa4	putative potential copper-transporting ATPase	Oryza sativa (japonica cultivar-group)	959	50947041	Viridiplantae	E	14	
61	Sel2	copper transporting P-ATPase	Synechococcus elongatus PCC 7942	790	436954	Cyanobacteria	B	14	
62	Ter2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Trichodesmium erythraeum IMS101	773	71677457	Cyanobacteria	B	14	
63	Npu2	COG2217: Cation transport ATPase	Nostoc punctiforme PCC 73102	808	23130092	Cyanobacteria	B	14	
64	Ava2	Copper-translocating P-type ATPase	Anabaena variabilis ATCC 29413	813	75907770	Cyanobacteria	B	14	
65	Nsp7	cation-transporting P-type ATPase	Nostoc sp. PCC 7120	815	17132916	Cyanobacteria	B	14	
66	Ssp3	cation-transporting ATPase; E1-E2 ATPase	Synechocystis sp. PCC 6803	780	16331210	Cyanobacteria	B	14	
67	Tel1	cation-transporting P-type ATPase	Thermosynechococcus elongatus BP-1	745	22295646	Cyanobacteria	B	14	
68	Ssu1	COG2217: Cation transport ATPase	Streptococcus suis 89/1591	816	50591442	Firmicutes (LowGC G+)	B	15	786.37 ± 62.54
69	Cpe1	probable copper-transporting ATPase	Clostridium perfringens str. 13	889	18144214	Firmicutes (LowGC G+)	B	15	
70	Cte1	copper efflux ATPase	Clostridium tetani E88	670	28210589	Firmicutes (LowGC G+)	B	15	
71	Cac1	Heavy-metal transporting P-type ATPase	Clostridium acetobutylicum ATCC 824	818	15026756	Firmicutes (LowGC G+)	B	15	
72	Efa1	copper-translocating P-type ATPase	Enterococcus faecalis V583	828	29374937	Firmicutes (LowGC G+)	B	15	
73	Efa2	TcrA	Enterococcus faecium	811	76151977	Firmicutes (LowGC G+)	B	15	
74	Spy1	copper-exporting ATPase	Streptococcus pyogenes MGAS5005	743	71911218	Firmicutes (LowGC G+)	B	15	
75	Sag1	hypothetical protein gbs0421	Streptococcus agalactiae NEM316	744	25010494	Firmicutes (LowGC G+)	B	15	
76	Smu1	CopA	Streptococcus mutans	742	9965435	Firmicutes (LowGC G+)	B	15	
77	Sth1	cation transporting ATPase, copper transport	Streptococcus thermophilus CNRZ1066	742	55739474	Firmicutes (LowGC G+)	B	15	
78	Tde1	copper-translocating P-type ATPase	Treponema denticola ATCC 35405	891	42525527	Spirochaetes	B	15	
79	Tpa1	cation-transporting ATPase, P-type	Treponema pallidum subsp. pallidum str. Nichols	792	15640020	Spirochaetes	B	15	
80	Dps1	probable heavy-metal transporting ATPase	Desulfotalea psychrophila LSv54	816	50876360	Deltaproteo-bacteria	B	15	
81	Dvu1	copper-translocating P-type ATPase	Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough	905	46580729	Deltaproteobacteria	B	15	

82	Fnu1	Copper-exporting ATPase	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	769	19713692	Fusobacteria	B	15	
83	Hhe1	Helicobacter hepaticus ATCC 51449, complete genome	Helicobacter hepaticus ATCC 51449	747	32262231	Epsilonproteo-bacteria	B	15	
84	Hfe1	copA	Helicobacter felis	732	2660542	Epsilonproteobacteria	B	15	
85	Hpy1	CopA	Helicobacter pylori	741	1477772	Epsilonproteobacteria	B	15	
86	Hpy2	adenosine triphosphatase	Helicobacter pylori	745	1516876	Epsilonproteobacteria	B	15	
87	Efa5	copper-translocating P-type ATPase	Enterococcus faecalis V583	701	29375460	Firmicutes (LowGC G+)	B	4	706.1 ± 40.41
88	Efa4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Enterococcus faecium DO	718	68194673	Firmicutes (LowGC G+)	B	4	
89	Ehi2	ATPase	Enterococcus hirae	745	290543	Firmicutes (LowGC G+)	B	4	
90	Sau2	putative cation exporting ATPase protein	Staphylococcus aureus subsp. aureus MRSA252	681	49482924	Firmicutes (LowGC G+)	B	4	
91	Sep2	cation-transporting ATPase, E1-E2 family	Staphylococcus epidermidis RP62A	687	57865849	Firmicutes (LowGC G+)	B	4	
92	Lpl2	copper transporting ATPase	Lactobacillus plantarum WCFS1	679	28379707	Firmicutes (LowGC G+)	B	4	
93	Lca1	COG2217: Cation transport ATPase	Lactobacillus casei ATCC 334	674	62514142	Firmicutes (LowGC G+)	B	4	
94	Ppe1	COG2217: Cation transport ATPase	Pediococcus pentosaceus ATCC 25745	696	48870965	Firmicutes (LowGC G+)	B	4	
95	Aae1	cation transporting ATPase (E1-E2 family)	Aquifex aeolicus VF5	664	15606387	Aquificae	B	4	
96	Olh2	copper-transporting ATPase	Oceanobacillus iheyensis HTE831	671	23099176	Firmicutes (LowGC G+)	B	4	
97	Mth3	heavy-metal transporting CPx-type ATPase	Methanothermobacter thermautotrophicus str. Delta H	675	2621844	Euryarchaeota	A	4	
98	Tde4	heavy metal translocating P-type ATPase	Thiobacillus denitrificans ATCC 25259	689	74318758	Betaproteobacteria	B	4	
99	Bba2	Heavy-metal transporting CPX-type ATPase	Bdellovibrio bacteriovorus HD100	692	42523748	Deltaproteobacteria	B	4	
100	Afu3	copper-transporting ATPase, P-type (copB)	Archaeoglobus fulgidus DSM 4304	690	11497769	Euryarchaeota	A	4	
101	Mba2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methanosarcina barkeri str. fusaro	829	68133215	Euryarchaeota	A	4	
102	Mac2	copper P-type ATPase	Methanosarcina acetivorans C2A	764	20089064	Euryarchaeota	A	4	
103	Mma4	Copper-exporting ATPase	Methanosarcina mazei Go1	711	21227565	Euryarchaeota	A	4	
104	Tth2	copper-exporting ATPase	Thermus thermophilus HB27	687	46199673	Deinococci	B	4	
105	Cau2	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chloroflexus aurantiacus J-10-fl	709	76259864	Chloroflexi	B	4	
106	Hma3	copper-transporting ATPase CopA	Haloarcula marismortui ATCC 43049	760	55229187	Euryarchaeota	A	4	
107	Dge4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	857	66798845	Deinococci	B	5	828.6 ± 17.56
108	Nha4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	811	69931139	Alphaproteobacteria	B	5	
109	Nsp6	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nocardioides sp. JS614	818	71369384	Actinobacteria (HighGC G+)	B	5	
110	Nfa2	putative cation-transporting ATPase	Nocardia farcinica IFM 10152	827	54027733	Actinobacteria (HighGC G+)	B	5	

111	Nfa3	putative cation-transporting ATPase	Nocardia farcinica IFM 10152	830	54027676	Actinobacteria (HighGC G+)	B	5	
112	Cvi2	Cu-ATPase	Chromobacterium violaceum ATCC 12472	724	34498672	Betaproteobacteria	B	6	787.83 ± 51.63
113	Xax1	CopF	Xanthomonas axonopodis pv. vesicatoria	808	46981559	Gammaproteobacteria	B	6	
114	Asp4	haloacid dehalogenase/epoxide hydrolase family protein	Azoarcus sp. EbN1	785	56476640	Betaproteobacteria	B	6	
115	Dar2	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	772	71907887	Betaproteo-bacteria	B	6	
116	Rme2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Ralstonia metallidurans CH34	805	68558944	Betaproteo-bacteria	B	6	
117	Bvi2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia vietnamiensis G4	814	67534305	Betaproteo-bacteria	B	6	
118	Bvi3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia vietnamiensis G4	716	67543438	Betaproteo-bacteria	B	6	
119	Neu6	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrosomonas eutropha C71	794	71549362	Betaproteo-bacteria	B	6	
120	Neu4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrosomonas eutropha C71	778	71549384	Betaproteo-bacteria	B	6	
121	Neu5	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrosomonas eutropha C71	794	71550862	Betaproteo-bacteria	B	6	
122	Neu7	Haloacid dehalogenase/epoxide hydrolase family:E1-E2 ATPase	Nitrosomonas europaea ATCC 19718	722	30249201	Betaproteo-bacteria	B	6	
123	Rfe2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rhodoferax ferrireducens DSM 15236	851	74024178	Betaproteo-bacteria	B	6	
124	Rge3	COG2217: Cation transport ATPase	Rubrivivax gelatinosus PM1	675	47571761	Betaproteo-bacteria	B	6	
125	Psp2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Polaromonas sp. JS666	796	67848215	Betaproteo-bacteria	B	6	
126	Rfe3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rhodoferax ferrireducens DSM 15236	816	74021198	Betaproteo-bacteria	B	6	
127	Mde2	COG2217: Cation transport ATPase	Microbulbifer degradans 2-40	822	48863058	Gammaproteo-bacteria	B	6	
128	Spo3	copper-translocating P-type ATPase	Silicibacter pomeroyi DSS-3	785	56709147	Alphaproteo-bacteria	B	6	
129	Rle1	copper transporter ActP	Rhizobium leguminosarum bv. viciae	841	4633808	Alphaproteo-bacteria	B	6	
130	Mlo1	cation transporting P-type ATPase	Mesorhizobium loti MAFF303099	839	13474443	Alphaproteo-bacteria	B	6	
131	Msp3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	846	68192368	Alphaproteo-bacteria	B	6	
132	Tde3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Thiomicrospira denitrificans ATCC 33889	721	71151189	Epsilonproteo-bacteria	B	6	
133	Sme4	Putative cation transport P-type ATPase	Sinorhizobium meliloti 1021	733	16263042	Alphaproteo-bacteria	B	6	
134	Rpa1	putative cation transporting P-type ATPase	Rhodopseudomonas palustris CGA009	973	39934730	Alphaproteo-bacteria	B	6	

135	Bja1	heavy-metal transporting P-type ATPase	Bradyrhizobium japonicum USDA 110	823	27375811	Alphaproteo-bacteria	B	6	
136	Nha1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	801	69931404	Alphaproteo-bacteria	B	6	
137	Nha2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	818	69931153	Alphaproteo-bacteria	B	6	
138	Nha3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	833	69930453	Alphaproteo-bacteria	B	6	
139	Nwi1	Heavy metal translocating P-type ATPase	Nitrobacter winogradskyi Nb-255	831	75675104	Alphaproteo-bacteria	B	6	
140	Nar1	COG2217: Cation transport ATPase	Novosphingobium aromaticivorans DSM 12444	719	48848119	Alphaproteo-bacteria	B	6	
141	Sal1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Sphingopyxis alaskensis RB2256	787	68537607	Alphaproteo-bacteria	B	6	
142	Nar2	COG2217: Cation transport ATPase	Novosphingobium aromaticivorans DSM 12444	753	48848641	Alphaproteo-bacteria	B	6	
143	Sal2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Sphingopyxis alaskensis RB2256	773	68537513	Alphaproteo-bacteria	B	6	
144	Sal3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Sphingopyxis alaskensis RB2256	773	68537533	Alphaproteo-bacteria	B	6	
145	Sty2	putative cation transporting P-type ATPase SilP	Salmonella typhimurium	824	4206631	Gammaproteo-bacteria	B	6	
146	Kpn1	SilP	Klebsiella pneumoniae	815	38016809	Gammaproteo-bacteria	B	6	
147	Sma2	putative cation transporting P-type ATPase (silver resistance)	Serratia marcescens	815	38259447	Gammaproteo-bacteria	B	6	
148	Lpn4	copper efflux ATPase	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	736	52841258	Gammaproteo-bacteria	B	6	
149	Lpn3	hypothetical protein lpl1397	Legionella pneumophila str. Lens	735	54294329	Gammaproteo-bacteria	B	6	
150	Lpn1	hypothetical protein lpp1596	Legionella pneumophila str. Paris	735	54297546	Gammaproteo-bacteria	B	6	
151	Lpn2	copper efflux ATPase	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	735	52841854	Gammaproteo-bacteria	B	6	
152	Rba1	copper-transporting ATPase	Rhodopirellula baltica SH 1	807	32442997	Planctomycetes	B	6	
153	Eli1	putative cation transporting P-type ATPase	Erythrobacter litoralis HTCC2594	776	61101425	Alphaproteo-bacteria	B	6	
154	Par2	copper/silver efflux P-type ATPase	Psychrobacter arcticus 273-4	814	71039127	Gammaproteo-bacteria	B	6	
155	Pcr2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Psychrobacter cryohalolentis K5	814	71363512	Gammaproteo-bacteria	B	6	
156	Bps1	COG2217: Cation transport ATPase	Burkholderia pseudomallei 668	807	67736006	Betaproteobacteria	B	6	
157	Bfu2	COG2217: Cation transport ATPase	Burkholderia fungorum LB400	693	48782831	Betaproteo-bacteria	B	6	
158	Bvi4	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia vietnamiensis G4	801	67543504	Betaproteo-bacteria	B	6	
159	Msp2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Magnetococcus sp. MC-1	807	68246736	unclassified Proteobacteria	B	7	807

160	Ilo1	Cation transport ATPase	Idiomarina loihiensis L2TR	749	56459705	Gammaproteo-bacteria	B	8	815.96 ± 51.95
161	Asp3	putative cation-transporting P-type ATPase	Acinetobacter sp. BW3	663	30409104	Gammaproteobacteria	B	8	
162	Neu2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrosomonas eutropha C71	837	71549401	Betaproteo-bacteria	B	8	
163	Msp1	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	855	68190740	Alphaproteo-bacteria	B	8	
164	Neu3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nitrosomonas eutropha C71	829	71548817	Betaproteo-bacteria	B	8	
165	Sty1	putative cation transport ATPase	Salmonella typhimurium LT2	762	16418854	Gammaproteo-bacteria	B	8	
166	Bbr1	probable cation-transporting ATPase	Bordetella bronchiseptica RB50	808	33600166	Betaproteobacteria	B	8	
167	Avi1	Copper ion binding domain:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Azotobacter vinelandii AvOP	829	67158934	Gammaproteobacteria	B	8	
168	Pmi1	heavy-metal transporting P-type ATPase	Proteus mirabilis	829	1353678	Gammaproteo-bacteria	B	8	
169	Atu1	heavy-metal transporting P-type ATPase	Agrobacterium tumefaciens str. C58	836	17935100	Alphaproteo-bacteria	B	8	
170	Bme1	COPPER-TRANSPORTING ATPASE	Brucella melitensis 16M	826	17983757	Alphaproteo-bacteria	B	8	
171	Sme3	ActP copper transport ATPase	Sinorhizobium meliloti 1021	826	16263001	Alphaproteo-bacteria	B	8	
172	Sme1	P-type ATPase ActP	Sinorhizobium medicae WSM419	827	4680350	Alphaproteo-bacteria	B	8	
173	Sme2	putative copper-transporting P-type ATPase protein	Sinorhizobium meliloti 1021	827	15140904	Alphaproteo-bacteria	B	8	
174	Atu2	copper transporting ATPase	Agrobacterium tumefaciens str. C58	861	17934845	Alphaproteo-bacteria	B	8	
175	bac2	ActP	uncultured bacterium	962	37222119	environmental samples	B	8	
176	Pde1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Paracoccus denitrificans PD1222	807	69935212	Alphaproteo-bacteria	B	8	
177	Jsp1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Jannaschia sp. CCS1	744	68182783	Alphaproteo-bacteria	B	8	
178	Ssp2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Silicibacter sp. TM1040	814	69301665	Alphaproteo-bacteria	B	8	
179	Rsp1	COG2217: Cation transport ATPase	Rhodobacter sphaeroides 2.4.1	813	46192167	Alphaproteo-bacteria	B	8	
180	Spo2	copper-translocating P-type ATPase	Silicibacter pomeroyi DSS-3	828	56677433	Alphaproteo-bacteria	B	8	
181	Tth1	cation-transporting ATPase pacS	Thermus thermophilus HB27	798	46199660	Deinococci	B	8	
182	Dra1	cation-transporting ATPase	Deinococcus radiodurans R1	847	6460273	Deinococci	B	8	
183	Gvi1	cation-transporting ATPase	Gloeobacter violaceus PCC 7421	747	37523616	Cyanobacteria	B	8	
184	Dge3	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	836	66799238	Deinococci	B	8	

185	Dge1	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	838	66799324	Deinococci	B	8	
186	Dge2	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	833	66798023	Deinococci	B	8	
187	Apl1	COG2217: Cation transport ATPase	Actinobacillus pleuropneumoniae serovar 1 str. 4074	716	32034997	Gammaproteo-bacteria	B	9	848.88 ± 83.82
188	Asu1	copper-transporting P-type ATPase	Actinobacillus succinogenes 130Z	922	75431632	Gammaproteo-bacteria	B	9	
189	Msu1	ZntA protein	Mannheimia succiniciproducens MBEL55E	750	52306995	Gammaproteo-bacteria	B	9	
190	Eca1	copper-transporting P-type ATPase	Erwinia carotovora subsp. atroseptica SCRI1043	907	49610658	Gammaproteo-bacteria	B	9	
191	Plu1	Copper-transporting P-type ATPase	Photorhabdus luminescens subsp. laumondii TTO1	911	37527684	Gammaproteo-bacteria	B	9	
192	Sma1	putative copper transporting P-type ATPase efflux pump	Serratia marcescens	903	55581748	Gammaproteo-bacteria	B	9	
193	Ype1	cation-translocating ATPase	Yersinia pestis KIM	961	22124996	Gammaproteo-bacteria	B	9	
194	Eco1	COG2217: Cation transport ATPase	Escherichia coli HS	834	75194704	Gammaproteo-bacteria	B	9	
195	Sen1	copper-transporting ATPase	Salmonella enterica subsp. enterica serovar Typhi	833	16501768	Gammaproteo-bacteria	B	9	
196	Ppr3	hypothetical cation-transporting ATPase	Photobacterium profundum SS9	965	54309978	Gammaproteo-bacteria	B	9	
197	Vfi1	copper-exporting ATPase	Vibrio fischeri ES114	893	59711388	Gammaproteo-bacteria	B	9	
198	Vch3	COG2217: Cation transport ATPase	Vibrio cholerae V51	915	75820888	Gammaproteo-bacteria	B	9	
199	Vch1	COG2217: Cation transport ATPase	Vibrio cholerae RC385	915	75824332	Gammaproteo-bacteria	B	9	
200	Vch2	COG2217: Cation transport ATPase	Vibrio cholerae O395	906	75827509	Gammaproteo-bacteria	B	9	
201	Vvu1	cation transport ATPase	Vibrio vulnificus YJ016	922	37679128	Gammaproteo-bacteria	B	9	
202	Vpa1	cation transport ATPase, E1-E2 family	Vibrio parahaemolyticus RIMD 2210633	911	28897532	Gammaproteo-bacteria	B	9	
203	Vsp1	COG2217: Cation transport ATPase	Vibrio sp. Ex25	896	75855297	Gammaproteo-bacteria	B	9	
204	Sba1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella baltica OS155	744	68542465	Gammaproteo-bacteria	B	9	
205	Son1	cation transport ATPase, E1-E2 family	Shewanella oneidensis MR-1	753	24373257	Gammaproteo-bacteria	B	9	
206	Cbu1	copper-translocating P-type ATPase	Coxiella burnetii RSA 493	742	29654798	Gammaproteo-bacteria	B	9	
207	Mca1	copper-translocating P-type ATPase	Methylococcus capsulatus str. Bath	725	53803908	Gammaproteo-bacteria	B	9	
208	Csa1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chromohalobacter salexigens DSM 3043	850	67677553	Gammaproteo-bacteria	B	9	
209	Ilo2	Cation transport ATPase	Idiomarina loihiensis L2TR	753	56460326	Gammaproteo-bacteria	B	9	
210	Sfr1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella frigidimarina NCIMB 400	746	69949999	Gammaproteo-bacteria	B	9	

211	Cau1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chloroflexus aurantiacus J-10-fl	850	76258102	Chloroflexi	B	10	813.14 ± 83.90
212	Nsp6	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nocardioides sp. JS614	1071	71366115	Actinobacteria (HighGC G+)	B	10	
213	Rxy1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rubrobacter xylanophilus DSM 9941	751	68560830	Actinobacteria (HighGC G+)	B	10	
214	Ade1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Anaeromyxobacter dehalogenans 2CP-C	954	66855764	Deltaproteo-bacteria	B	10	
215	Sus1	Copper-translocating P-type ATPase:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Solibacter usitatus Ellin6076	798	67933813	Acidobacteria	B	10	
216	Nsp3	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Nocardioides sp. JS614	928	71367384	Actinobacteria (HighGC G+)	B	10	
217	Ppr2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pelobacter propionicus DSM 2379	786	71839521	Deltaproteo-bacteria	B	10	
218	Ssp1	cation-transporting ATPase; E1-E2 ATPase	Synechocystis sp. PCC 6803	745	16329860	Cyanobacteria	B	10	
219	Sel1	PacS	Synechococcus elongatus PCC 7942	747	22002521	Cyanobacteria	B	10	
220	Cwa1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Crocosphaera watsonii WH 8501	766	67921918	Cyanobacteria	B	10	
221	Ter1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Trichodesmium erythraeum IMS101	758	71677291	Cyanobacteria	B	10	
222	Npu1	COG2217: Cation transport ATPase	Nostoc punctiforme PCC 73102	760	53688476	Cyanobacteria	B	10	
223	Nsp2	cation-transporting ATPase	Nostoc sp. PCC 7120	753	17158771	Cyanobacteria	B	10	
224	Ava1	Copper-translocating P-type ATPase	Anabaena variabilis ATCC 29413	753	75910433	Cyanobacteria	B	10	
225	Nsp1	cation-transporting ATPase	Nostoc sp. PCC 7120	753	17135447	Cyanobacteria	B	10	
226	Bfr1	putative copper transport-related membrane protein	Bacteroides fragilis NCTC 9343	836	60494739	Bacteroidetes	B	10	
227	Chu1	COG2217: Cation transport ATPase	Cytophaga hutchinsonii	804	48854913	Bacteroidetes	B	10	
228	Bfr2	cation-transporting ATPase	Bacteroides fragilis YCH46	736	53713321	Bacteroidetes	B	10	
229	Bth1	cation-transporting ATPase pacS	Bacteroides thetaiotaomicron VPI-5482	738	29338397	Bacteroidetes	B	10	
230	Dac1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Desulfuromonas acetoxidans DSM 684	734	68179510	Deltaproteo-bacteria	B	10	
231	Det1	copper-translocating P-type ATPase	Dehalococcoides ethenogenes 195	828	57234243	Chloroflexi	B	10	
232	Dsp1	copper-translocating P-type ATPase	Dehalococcoides sp. CBDB1	828	73660437	Chloroflexi	B	10	
233	Sfu1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Syntrophobacter fumaroxidans MPOB	814	71544562	Deltaproteo-bacteria	B	10	

234	Reu2	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Ralstonia eutropha JMP134	814	72384011	Betaproteo-bacteria	B	10	
235	Nsp4	cation transporting ATPase	Nostoc sp. PCC 7120	724	17158728	Cyanobacteria	B	10	
236	Mth2	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Moorella thermoacetica ATCC 39073	857	68268515	Firmicutes (LowGC G+)	B	10	
237	Mbu1	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methanococcoides burtonii DSM 6242	942	68212009	Euryarchaeota	A	10	
238	Mba1	Copper ion binding domain:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methanosarcina barkeri str. fusaro	954	68133914	Euryarchaeota	A	10	
239	Mac1	P-type copper-transporting ATPase	Methanosarcina acetivorans C2A	982	20090203	Euryarchaeota	A	10	
240	Mma3	Copper-exporting ATPase	Methanosarcina mazei Go1	962	21228430	Euryarchaeota	A	10	
241	Afu1	cation-transporting ATPase, P-type (pacS)	Archaeoglobus fulgidus DSM 4304	804	11498084	Euryarchaeota	A	10	
242	Pfu1	heavy-metal transporting cpx-type atpase	Pyrococcus furiosus DSM 3638	799	18977112	Euryarchaeota	A	10	
243	Tko1	heavy-metal transporting P-type ATPase	Thermococcus kodakarensis KOD1	799	57159096	Euryarchaeota	A	10	
244	Mma2	Haloacid dehalogenase/epoxide hydrolase:ATPase, E1-E2 type:Heavy metal transport/detoxification protein	Methanococcus maripaludis S2	723	45358728	Euryarchaeota	A	10	
245	Mth1	heavy-metal transporting CPx-type ATPase	Methanothermobacter thermautotrophicus str. Delta H	790	2622654	Euryarchaeota	A	10	
246	Cli1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chlorobium limicola DSM 245	786	67918133	Chlorobi	B	10	
247	Cte2	copper-transporting ATPase, E1-E2 family	Chlorobium tepidum TLS	758	21673644	Chlorobi	B	10	
248	Cph2	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chlorobium phaeobacteroides DSM 266	773	67937188	Chlorobi	B	10	
249	Pph1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pelodictyon phaeoclathratiforme BU-1	755	68549269	Chlorobi	B	10	
250	Mtu1	cation-transporting ATPase, E1-E2 family	Mycobacterium tuberculosis CDC1551	792	13880563	Actinobacteria (HighGC G+)	B	11	831.44 ± 71.34
251	Hsp1	YvgX	Halobacterium sp. NRC-1	857	15789882	Euryarchaeota	A	11	
252	Hma1	copper-transporting ATPase	Haloarcula marismortui ATCC 43049	868	55229179	Euryarchaeota	A	11	
253	Hma2	copper-transporting ATPase CopA	Haloarcula marismortui ATCC 43049	873	55230107	Euryarchaeota	A	11	
254	Cph1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chlorobium phaeobacteroides BS1	727	67938332	Chlorobi	B	11	

255	Dha1	Copper ion binding domain:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Desulfitobacterium hafniense DCB-2	976	68207415	Firmicutes (LowGC G+)	B	11	
256	Ppr1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pelobacter propionicus DSM 2379	795	71838988	Deltaproteo-bacteria	B	11	
257	Gme1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Geobacter metallireducens GS-15	798	68004283	Deltaproteo-bacteria	B	11	
258	Gsu1	copper-translocating P-type ATPase	Geobacter sulfurreducens PCA	797	39984438	Deltaproteo-bacteria	B	11	
259	Lmo2	CtpA	Listeria monocytogenes	653	1513069	Firmicutes (LowGC G+)	B	12	658.29 ± 40.73
260	Lde1	COG2217: Cation transport ATPase	Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365	638	62516436	Firmicutes (LowGC G+)	B	12	
261	Spn1	COG2217: Cation transport ATPase	Streptococcus pneumoniae TIGR4	750	66878130	Firmicutes (LowGC G+)	B	12	
262	Lpl1	copper transporting ATPase	Lactobacillus plantarum WCFS1	641	28379476	Firmicutes (LowGC G+)	B	12	
263	Lac1	copper-transporting ATPase	Lactobacillus acidophilus NCFM	641	58338205	Firmicutes (LowGC G+)	B	12	
264	Lga1	hypothetical protein Lgas_03001578	Lactobacillus gasseri ATCC 33323	644	23003522	Firmicutes (LowGC G+)	B	12	
265	Ljo1	cation-transporting ATPase PacS	Lactobacillus johnsonii NCC 533	641	42519887	Firmicutes (LowGC G+)	B	12	
266	Ath1	ATP dependent copper transporter (RAN1)	Arabidopsis thaliana	1001	6850337	Viridiplantae	E	13	1206.27 ± 213.52
267	Bna1	copper-transporting P-type ATPase	Brassica napus	999	15636781	Viridiplantae	E	13	
268	Osa1	putative ATP dependent copper transporter	Oryza sativa (japonica cultivar-group)	926	52076515	Viridiplantae	E	13	
269	Osa2	putative copper-exporting ATPase	Oryza sativa (japonica cultivar-group)	1012	50905629	Viridiplantae	E	13	
270	Osa3	putative copper-transporting P-type ATPase	Oryza sativa (japonica cultivar-group)	978	50905963	Viridiplantae	E	13	
271	Ath2	F2K11.18	Arabidopsis thaliana	1191	6633848	Viridiplantae	E	13	
272	Sbi2	putative copper-exporting ATPase	Sorghum bicolor	1002	48374970	Viridiplantae	E	13	
273	Sbi1	putative copper-exporting ATPase	Sorghum bicolor	908	48374969	Viridiplantae	E	13	
274	Zma1	putative ATP dependent copper transporter	Zea mays	1001	48374954	Viridiplantae	E	13	
275	Spo1	SPBC29A3.01	Schizosaccharomyces pombe	904	3006137	Fungi	E	13	
276	Yli1	hypothetical protein	Yarrowia lipolytica CLIB122	933	50551739	Fungi	E	13	
277	Ani1	hypothetical protein AN3624.2	Aspergillus nidulans FGSC A4	1182	67526333	Fungi	E	13	
278	Afu2	copper-transporting ATPase, putative	Aspergillus fumigatus Af293	1187	66849203	Fungi	E	13	
279	Gze2	hypothetical protein FG08188.1	Gibberella zeae PH-1	1174	42550200	Fungi	E	13	
280	Gze1	hypothetical protein FG01501.1	Gibberella zeae PH-1	1106	42546043	Fungi	E	13	
281	Ncr1	hypothetical protein	Neurospora crassa OR74A	1181	32417818	Fungi	E	13	
282	Mgr1	hypothetical protein MGG_03724	Magnaporthe grisea 70-15	1186	39943288	Fungi	E	13	
283	Gli1	CLAP1	Glomerella lindemuthiana	1167	24528450	Fungi	E	13	
284	Uma1	hypothetical protein UM00227.1	Ustilago maydis 521	1056	71003387	Fungi	E	13	
285	Cne1	copper-exporting ATPase, putative	Cryptococcus neoformans var. neoformans JEC21	1055	57230777	Fungi	E	13	
286	Tve1	copper P-type ATPase CtaA	Trametes versicolor	983	28625435	Fungi	E	13	
287	Kla1	unnamed protein product	Kluyveromyces lactis NRRL Y-1140	975	50310791	Fungi	E	13	
288	Sce1	Ccc2p	Saccharomyces cerevisiae	1004	6320475	Fungi	E	13	

289	Ddi1	hypothetical protein DDBDRAFT_0218568	Dictyostelium discoideum AX4	985	66809993	Mycetozoa	E	13
290	Ddi2	hypothetical protein DDBDRAFT_0168129	Dictyostelium discoideum AX4	1280	66822883	Mycetozoa	E	13
291	Tni2	unnamed protein product	Tetraodon nigroviridis	1727	47214278	Metazoa	E	13
292	Ptr1	PREDICTED: similar to ATP7B	Pan troglodytes	1197	55640249	Metazoa	E	13
293	Hsa1	ATP7B	Homo sapiens	1465	1947035	Metazoa	E	13
294	Hsa2	Cu transporting ATPase P	Homo sapiens	1411	738766	Metazoa	E	13
295	Cfa1	ATPase, Cu++ transporting, beta polypeptide	Canis familiaris	1432	70608105	Metazoa	E	13
296	Oar1	ATPase, Cu++ transporting, beta polypeptide	Ovis aries	1505	57619187	Metazoa	E	13
297	Mmu1	ATPase, Cu++ transporting, beta polypeptide	Mus musculus	1462	6680758	Metazoa	E	13
298	Rno1	ATPase, Cu++ transporting, beta polypeptide	Rattus norvegicus	1451	6978561	Metazoa	E	13
299	Cfa2	PREDICTED: similar to ATPase, Cu++ transporting, alpha polypeptide isoform 2	Canis familiaris	1504	74007803	Metazoa	E	13
300	Hsa3	Menkes Disease (ATP7A)	Homo sapiens	1376	5262841	Metazoa	E	13
301	Cgr1	Copper-transporting ATPase 1 (Copper pump 1)	Cricetulus griseus	1476	1351992	Metazoa	E	13
302	Mmu2	putative copper efflux ATPase	Mus musculus	1465	458224	Metazoa	E	13
303	Rno2	ATPase, Cu++ transporting, alpha polypeptide	Rattus norvegicus	1492	16258817	Metazoa	E	13
304	Dre1	Menkes disease ATPase	Danio rerio	1482	70724999	Metazoa	E	13
305	Tni1	unnamed protein product	Tetraodon nigroviridis	1492	47222902	Metazoa	E	13
306	Dme1	CG1886-PA	Drosophila melanogaster	1219	45446920	Metazoa	E	13
307	Aga1	ENSANGP00000026574	Anopheles gambiae str. PEST	1126	55234928	Metazoa	E	13
308	Aga2	ENSANGP00000008866	Anopheles gambiae str. PEST	1145	55234929	Metazoa	E	13
309	Cbr1	Hypothetical protein CBG21197	Caenorhabditis briggsae	1241	39586559	Metazoa	E	13
310	Cel1	Hypothetical protein Y76A2A.2	Caenorhabditis elegans	1238	9367167	Metazoa	E	13
311	Dar3	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	808	71907465	Betaproteo-bacteria	B	16
312	Mca2	copper-translocating P-type ATPase	Methylococcus capsulatus str. Bath	831	53804835	Gammaproteo-bacteria	B	16
313	Mca3	copper-translocating P-type ATPase	Methylococcus capsulatus str. Bath	779	53805105	Gammaproteo-bacteria	B	16
314	Pam1	putative copper-transporting ATPase	Candidatus Protochlamydia amoebophila UWE25	729	46399353	Chlamydiae	B	17
315	Zmo1	copper-transporting ATPase	Zymomonas mobilis subsp. mobilis ZM4	740	56543385	Alphaproteo-bacteria	B	17
316	Bba1	copper-transporting ATPase copA	Bdellovibrio bacteriovorus HD100	724	42523682	Deltaproteo-bacteria	B	17
317	Asp1	P-type ATPase, copper transporting ATPase, a phophatase-like domain	Acinetobacter sp. ADP1	802	50085487	Gammaproteo-bacteria	B	17
318	Gox1	Cation-transporting ATPase	Gluconobacter oxydans 621H	791	58039131	Alphaproteo-bacteria	B	17
319	Cvi1	copper-transporting ATPase copA	Chromobacterium violaceum ATCC 12472	781	34496720	Betaproteo-bacteria	B	17
320	Asp2	copper-transporting ATPase	Azoarcus sp. EbN1	803	56478367	Betaproteo-bacteria	B	17

321	Dar1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	735	71906690	Betaproteo-bacteria	B	17	
322	Neu1	copA; copper-transporting ATPase	Nitrosomonas europaea ATCC 19718	782	30249022	Betaproteo-bacteria	B	17	
323	bac1	copper-translocating P-type ATPase	uncultured bacterium 577	797	40063228	environmental samples	B	17	
324	Hin1	probable cation-transporting ATPase	Haemophilus influenzae 86-028NP	722	68057093	Gammaproteo-bacteria	B	17	
325	Pmu1	unknown	Pasteurella multocida subsp. multocida str. Pm70	724	12722323	Gammaproteo-bacteria	B	17	
326	Ngo1	putative transport ATPase	Neisseria gonorrhoeae FA 1090	725	59801010	Betaproteo-bacteria	B	17	
327	Nme1	putative cation-transporting ATPase	Neisseria meningitidis Z2491	725	7380180	Betaproteo-bacteria	B	17	
328	Nme2	cation transport ATPase, E1-E2 family	Neisseria meningitidis MC58	720	7226567	Betaproteo-bacteria	B	17	
329	Par1	probable copper(heavy metal)-transporting P-type ATPase	Psychrobacter arcticus 273-4	786	71039177	Gammaproteo-bacteria	B	17	
330	Pcr1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Psychrobacter cryohalolentis K5	786	71364942	Gammaproteobacteria	B	17	
331	Lla1	copper/potassium-transporting ATPase	Lactococcus lactis subsp. lactis Il1403	720	15672816	Firmicutes (LowGC G+)	B	18	725 ± 4.36
332	Efa3	Copper-translocating P-type ATPase:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Enterococcus faecium DO	728	68194672	Firmicutes (LowGC G+)	B	18	
333	Ehi1	ATPase	Enterococcus hirae	727	290642	Firmicutes (LowGC G+)	B	18	
334	Lin2	heavy-metal transporting p-type ATPase	Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	739	45658809	Spirochaetes	B	19	796.36 ± 49.20
335	Bsu1	hypothetical protein BSU33500	Bacillus subtilis subsp. subtilis str. 168	803	16080403	Firmicutes (LowGC G+)	B	19	
336	Ban1	COG2217: Cation transport ATPase	Bacillus anthracis str. A2012	805	65321054	Firmicutes (LowGC G+)	B	19	
337	Bli1	Cu2+-exporting ATPase	Bacillus licheniformis ATCC 14580	814	52005067	Firmicutes (LowGC G+)	B	19	
338	Bcl1	copper-transporting ATPase	Bacillus clausii KSM-K16	809	56964995	Firmicutes (LowGC G+)	B	19	
339	Bha1	copper-transporting ATPase	Bacillus halodurans C-125	806	10173170	Firmicutes (LowGC G+)	B	19	
340	Bce3	copper-translocating P-type ATPase	Bacillus cereus ATCC 10987	798	44004521	Firmicutes (LowGC G+)	B	19	
341	Gka1	heavy metal-transporting ATPase	Geobacillus kaustophilus HTA426	798	56419437	Firmicutes (LowGC G+)	B	19	
342	Oih1	copper-transporting ATPase	Oceanobacillus iheyensis HTE831	791	23098597	Firmicutes (LowGC G+)	B	19	
343	Sha1	copper-transporting ATPase copA	Staphylococcus haemolyticus JCSC1435	795	70725497	Firmicutes (LowGC G+)	B	19	
344	Ssa1	copper-transporting ATPase	Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	794	73661606	Firmicutes (LowGC G+)	B	19	
345	Sau1	putative copper importing ATPase A	Staphylococcus aureus subsp. aureus MSSA476	802	49245779	Firmicutes (LowGC G+)	B	19	
346	Sep1'	copper-transporting ATPase copA	Staphylococcus epidermidis ATCC 12228	794	27469037	Firmicutes (LowGC G+)	B	19	
347	Hmo1	Copper-importing ATPase	Heliobacillus mobilis	839	27262376	Firmicutes (LowGC G+)	B	19	
348	Sth2	putative copper-transporting ATPase	Symbiobacterium thermophilum IAM 14863	949	51893754	Actinobacteria (HighGC G+)	B	19	

349	Swo1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Syntrophomonas wolfei str. Goettingen	799	71540755	Firmicutes (LowGC G+)	B	19	
350	Cth1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium thermocellum ATCC 27405	743	67917082	Firmicutes (LowGC G+)	B	19	
351	Tte1	Cation transport ATPase	Thermoanaerobacter tengcongensis MB4	796	20808818	Firmicutes (LowGC G+)	B	19	
352	Bcl2	copper-transporting ATPase	Bacillus clausii KSM-K16	862	56962006	Firmicutes (LowGC G+)	B	19	
353	Esi1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Exiguobacterium sibiricum 255-15	710	68055969	Firmicutes (LowGC G+)	B	19	
354	Lin1	hypothetical protein lin1967	Listeria innocua Clip11262	737	16801033	Firmicutes (LowGC G+)	B	19	
355	Lmo1	copper-translocating P-type ATPase	Listeria monocytogenes str. 4b F2365	737	46908085	Firmicutes (LowGC G+)	B	19	
356	Mde1	COG2217: Cation transport ATPase	Microbulbifer degradans 2-40	724	48863075	Gammaproteo-bacteria	B	20	811.97 ± 91.92
357	Mfl1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methylobacillus flagellatus KT	730	68212445	Betaproteo-bacteria	B	20	
358	Rfe1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rhodoferax ferrireducens DSM 15236	757	74023290	Betaproteo-bacteria	B	20	
359	Bfu1	COG2217: Cation transport ATPase	Burkholderia fungorum LB400	787	48782582	Betaproteo-bacteria	B	20	
360	Bma1	hypothetical protein Bmal10_02002739	Burkholderia mallei 10399	1061	67636459	Betaproteo-bacteria	B	20	
361	Bvi1	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia vietnamiensis G4	924	67542206	Betaproteo-bacteria	B	20	
362	Bam1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia ambifaria AMMD	946	74018699	Betaproteo-bacteria	B	20	
363	Bce1	COG2217: Cation transport ATPase	Burkholderia cepacia R18194	1031	46317831	Betaproteo-bacteria	B	20	
364	Bce2	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia cenocepacia HI2424	1021	67665187	Betaproteo-bacteria	B	20	
365	Tde2	COG2217: Cation transport ATPase	Thiobacillus denitrificans ATCC 25259	802	52006150	Betaproteo-bacteria	B	20	
366	Rge2	COG2217: Cation transport ATPase	Rubrivivax gelatinosus PM1	817	47571766	Betaproteo-bacteria	B	20	
367	Psp1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Polaromonas sp. JS666	816	67847332	Betaproteobacteri a	B	20	
368	Rge1	COG2217: Cation transport ATPase	Rubrivivax gelatinosus PM1	739	47571847	Betaproteo-bacteria	B	20	
369	Rso1	probable cation-transporting atpase transmembrane protein	Ralstonia solanacearum	748	17430371	Betaproteo-bacteria	B	20	
370	Reu1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Ralstonia eutropha JMP134	819	73543058	Betaproteo-bacteria	B	20	

371	Rme1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Ralstonia metallidurans CH34	813	68554610	Betaproteo-bacteria	B	20
372	Psy4	ORFG	Pseudomonas syringae	794	8388793	Gammaproteo-bacteria	B	20
373	Pae1	COG2217: Cation transport ATPase	Pseudomonas aeruginosa UCBPP-PA14	792	32039086	Gammaproteo-bacteria	B	20
374	Ppu2	heavy metal translocating P-type ATPase	Pseudomonas putida KT2440	799	26987324	Gammaproteo-bacteria	B	20
375	Pfl2	copper-translocating P-type ATPase	Pseudomonas fluorescens Pf-5	798	70734212	Gammaproteo-bacteria	B	20
376	Pfl1	COG2217: Cation transport ATPase	Pseudomonas fluorescens PfO-1	797	48732302	Gammaproteo-bacteria	B	20
377	Ppu1	copper transporter	Pseudomonas putida	797	22003413	Gammaproteo-bacteria	B	20
378	Psy3	copper-translocating P-type ATPase	Pseudomonas syringae pv. tomato str. DC3000	732	28867978	Gammaproteo-bacteria	B	20
379	Psy1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pseudomonas syringae pv. syringae B728a	732	66043921	Gammaproteo-bacteria	B	20
380	Psy2	copper-translocating P-type ATPase	Pseudomonas syringae pv. phaseolicola 1448A	732	71736278	Gammaproteo-bacteria	B	20
381	Sam1	Copper ion-binding:Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella amazonensis SB2B	789	68548173	Gammaproteo-bacteria	B	20
382	Cps1	copper-translocating P-type ATPase	Colwellia psychrerythraea 34H	791	71280081	Gammaproteo-bacteria	B	20
383	Sde1	Copper-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella denitrificans OS-217	793	69945673	Gammaproteo-bacteria	B	20
384	Mma1	COG2217: Cation transport ATPase	Magnetospirillum magnetotacticum MS-1	724	23015317	Alphaproteo-bacteria	B	20
385	Rru1	COG2217: Cation transport ATPase	Rhodospirillum rubrum	754	48764318	Alphaproteo-bacteria	B	20

Table 3. Motif analyses of the 20 clusters representing the Copper P-type ATPase Family

	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6	Motif 7	Motif 8	Motif 9
	PGD	PAD	TGES	PEGL	DKTGTKT	KGAPE	DPPR	MVTGD	VAVTGDGVNDSPALKKADIGVAM
1	(P)GE	PVD	TGE(P)	PCAL	DKTGTLT	(A)G(APS)	D(E I)R	MITGD	V(A)MVGDGNDA(P)AL(AG)AD(VG)AV
	335-337	340-342	358-361	483-486	526-532	533-537	658-662	681-685	724-746
2	PGE	ATD	TG(E)S	PCAL	DKTGTVT	TGTMT	D(TV)K	L(L)TGD	VAM(V)GDG(V)NDAAAL(AQ)ADLGIAM
	519-521	524-526	542-545	652-655	695-702	703-707	850-853	870-874	919-941
3	(P)G(E)	(PA)D	TGE(S)	PCAL	DKTG(T)LT	(R)G(IE)G	D(E I)R	I(L)TGD	V(M)M(V)GDGIND(AP)AL(AA)A(DT)G(VAM)
	386-388	391-393	409-412	514-517	557-563	628-632	707-710	729-733	774-796
4	PG(E)	P(A)D	TGES	PHAL	DKTGTLT	(E)G(RFG)	D(I)IR	MLTGD	(VA)M(V)GDGVNDAPAL(A)QA(D)IGIAI
	368-370	373-375	391-394	496-499	539-545	546-550	707-710	729-733	774-796
5	PG(A)	AVD	TGES	PDAL	DKTGTLT	(K)GEP(E)	DA(V)R	MLTGD	VA(M)VGDGVNDAPALA(Q)ADVGIAI
	373-375	378-380	396-399	500-503	544-550	551-555	672-675	694-698	738-760
6	PG(E)	PVD	TGES	PCAL	DKTGTLT	(E)G(K)P(K)	D(P)VK	M(L)TGD	VAMAGDGVNDAPALA(A)ADGIAM
	554-556	559-561	577-580	683-686	726-732	733-737	855-858	878-882	922-944
1 seq									
7	PGA	FSD	TGEF	PCAL	DKTGTLT	CGTPV	DPTL	MVTGD	VAMVGDGINDAPALARADVGIAL
	328-330	333-335	351-354	456-459	499-505	506-510	589-592	645-649	691-713
8	PGD	P(V)D	TGEF	PCAM	DKTGTKT	(E)G(RPE)	D(P)K	MVTGD	(VA)FVGDGINDAPALA(E)AD(V)GIAI
	390-392	395-397	413-416	519-522	562-568	569-573	762-765	784-788	828-850
9	PGE	PVD	TGEF	PCAL	DKTGTKT	(E)G(KPS)	D(P)I(K)	MLTGD	(VA)M(V)GDGINDAPAL(AQ)D(G)IAM
	548-550	553-555	571-574	677-680	720-726	727-731	867-870	889-893	933-955
10	PG(E)	PVD	TGES	PCAL	DKTGTLT	(R)G(K)P(T)	D(T)IK	MLTGD	V(A)M(V)GDGINDA(P)AL(AQ)ADIG(I)AI
	578-580	583-585	601-604	726-729	769-775	776-780	914-917	936-940	996-1018
11	PGE	PVD	TGES	PCAL	DKTGTLT	(R)G(EPE)	D(T)VK	MITGD	V(A)MVGDGINDAPALA(Q)A(D)G(I)AI
	528-530	532-535	551-554	682-685	725-731	732-736	889-892	911-915	957-979
12	PG(E)	PVD	TGES	PCAL	DKTGT/T	(V)GKP(Q)	D(V)PK	MLTGD	VAFVGDGINDAPALS(T)A(D)VGIAM
	270-272	275-277	293-296	405-408	448-454	455-459	575-578	597-601	641-663
13	PG(D)	(PA)D	TGES	PCAL	LKTGTLT	(H)G(TPV)	D(P)K	MVTGD	V(A)MVGDGINDSPAL(AA)A(D)VGIA(M)
	1270-1272	1275-1277	1293-1296	1453-1456	1534-1539	1540-1544	1817-1823	1839-1843	1908-1920
14	PGD	PVD	TGES	PCAL	DKTGTLT	(T)G(HPV)	D(TL)R	LLSGD	VAMVGDGINDAPALA(Q)A(D)VGI(A)L
	437-439	452-454	460-463	596-599	639-645	646-650	791-794	813-817	874-896
15	PGE	(P)VD	TGES	PC(A)L	DKTGTLT	(E)G(KPV)	D(TV)K	MLTGD	(VA)MVGDGINDAL(AQ)SD(V)GI(A)M
	485-487	490-492	508-511	621-624	664-671	672-676	847-850	869-873	913-935
16	PGD	PTD	TGES	PCAL	DKTGT/T	E(G)KP(E)	D(P)PR	MVTGD	VGMIGDGINDAPALAAADVGFAI
	361-363	366-368	384-387	488-491	531-537	538-542	659-662	682-686	726-748
17	(P)GE	PVD	TGE(S)	PC(A)L	DKTGTLT	(E)G(KPQ)	D(A)LR	MLTGD	VAM(V)GDGINDAPALA(A)AD(V)SFAI
	351-353	356-358	374-377	478-481	521-527	528-532	661-664	684-688	727-749
18	PGE	P(A)D	TGE(S)	PCAL	DKTGT/T	(E)GK(P/E	DQIK	MLTGD	(VG)MAGDGINDA(P)AL(AL)A(S)VGIAM
	257-259	262-264	280-283	384-387	427-433	434-438	563-566	585-589	631-653
19	PGE	P(V)D	TGES	PCAL	DKTGTVT	(N)G(KPE)	D(TV)K	M(L)TGD	(V)AMVGDGINDAPALA(T)AD(I)GIAI
	436-438	441-443	459-462	565-568	608-614	615-619	741-744	763-767	810-832
20	PGE	PVD	TGES	PCAL	DKTGTLT	(E)G(K)P(R)	D(TV)K	LLTGD	VAMVGDGIND(A)PALA(A)A(D)VG(I)AM
	583-585	588-590	606-609	710-713	753-759	760-764	928-931	950-954	1002-1024

Table 4. The organisms whose16S rRNAs were used to construct a phylogenetic tree representing each genus found among the 385 protein sequences representing the Copper P-type ATPases.

	Abb.	Description	Organism	Length	GI #
1	Gka1	B.kaustophilus 16S ribosomal RNA	Geobacillus kaustophilus	1432	39549
2	Mma1	Mycobacterium marinum 16S rRNA gene	Mycobacterium marinum	1489	44459
3	Vch1	V.cholerae gene for 16S ribosomal RNA	Vibrio cholerae	1452	49417
4	Atu1	Agrobacterium tumefaciens 16S rRNA gene	Agrobacterium tumefaciens	1489	142272
5	Fsp1	Frankia sp. 16S ribosomal RNA	Frankia sp.	1443	174593
6	Lmo1	Listeria monocytogenes 16S ribosomal RNA	Listeria monocytogenes	1553	175140
7	Lpn1	Legionella pneumophila 16S ribosomal RNA	Legionella pneumophila	1544	175168
8	Mth1	Methanococcus thermolithotrophicus 16S ribosomal RNA	Methanothermococcus thermolithotrophicus	1452	175445
9	Pae1	P.aeruginosa 16S ribosomal RNA	Pseudomonas aeruginosa	1517	175722
10	Tpa1	Treponema pallidum 16S ribosomal RNA	Treponema pallidum	1573	176249
11	Tbr1	Thermoanaerobacter finii (DSM 3389) 16S ribosomal RNA (16S rRNA) gene	Thermoanaerobacter brockii subsp. finnii	1523	349568
12	Mbu1	M.burtonii gene for 16S rRNA	Methanococcoides burtonii	1476	434307
13	Aca1	Acinetobacter anitratus ATCC 15308 16S rRNA gene	Acinetobacter calcoaceticus subsp. anitratus	1528	506689
14	Cau1	Chloroflexus aurantiacus gene for 16S ribosomal RNA, partial sequence	Chloroflexus aurantiacus J-10-fl	1401	550527
15	Lwe1	Leptospira weilii Sarmin 16S rRNA gene, partial sequence	Leptospira weilii	1486	558932
16	Bsp1	Bradyrhizobium spec. (LMG 9980) gene for 16S rRNA	Bradyrhizobium sp.	1441	563846
17	Sau1	Staphylococcus aureus 16S ribosomal RNA (16S rRNA) gene	Staphylococcus aureus	1500	576603
18	Ype1	Yersinia pestis 16S ribosomal RNA (16S rRNA) gene	Yersinia pestis	1467	576926
19	Kau1	K.aurantiacus (IFO 15268) 16S rRNA gene	Kineococcus aurantiacus	1417	609073
20	Pfu1	Pyrococcus furiosus 16S small subunit ribosomal RNA	Pyrococcus furiosus	1495	643670
21	Pca1	Pelobacter carbinolicus 16S ribosomal RNA gene, partial sequence	Pelobacter carbinolicus	1414	727426
22	Sde1	Thiomicrospira denitrificans 16S ribosomal RNA (16S rRNA) gene, partial sequence	Sulfurimonas denitrificans DSM 1251	1446	790922
23	Eac1	Exiguobacterium acetylicum 16S rRNA gene, partial sequence	Exiguobacterium acetylicum	1482	893364

24	Sth1	S.thermovulgaris 16S rRNA gene	Streptomyces thermovulgaris	1486	1071779
25	Sel1	Synechococcus elongatus gene for 16S rRNA	Synechococcus elongatus	1453	1213586
26	Eco1	E.coli (ATCC 11775T) gene for 16S rRNA	Escherichia coli	1450	1240022
27	Mde1	M.defluvii 16S rRNA gene	Methanothermobacter defluvii	1445	1430856
28	Apo1	Actinobacillus porcinus 16S ribosomal RNA gene, partial sequence	Actinobacillus porcinus	1466	1519226
29	Sla1	Silicibacter lacuscaerulensis 16S ribosomal RNA gene, partial sequence	Silicibacter lacuscaerulensis	1339	1737206
30	Sty1	Salmonella typhi 16S ribosomal RNA gene, complete sequence	Salmonella typhi	1541	1857865
31	Dsp1	Desulfitobacterium sp. 16S rRNA gene, clone 2	Desulfitobacterium sp.	1616	1915888
32	Mgl1	Moorella glycerini 16S small subunit ribosomal RNA gene, complete sequence	Moorella glycerini	1513	1916225
33	Cpe1	C.perfringens 16S rRNA gene	Clostridium perfringens	1504	2058294
34	Pmu1	16S rRNA gene of Pasteurella multocida	Pasteurella multocida	1545	2173518
35	Bbr1	Bordetella bronchiseptica 16S rRNA gene	Bordetella bronchiseptica	1532	2174260
36	Kox1	Klebsiella oxytoca gene for 16S ribosomal RNA, partial sequence	Klebsiella oxytoca	1441	2209046
37	Ssp1	Syntrophobacter sp. 16S ribosomal RNA	Syntrophobacter sp.	1484	2463455
38	Gox1	Gluconobacter oxydans gene for 16S rRNA	Gluconobacter oxydans	1476	2597906
39	Ljo1	Lactobacillus johnsonii 16S rRNA gene	Lactobacillus johnsonii	1487	2597958
40	Ppr1	Photobacterium profundum gene for 16S ribosomal RNA, strain:SS9	Photobacterium profundum SS9	1518	2924634
41	Nsp1	Nostoc ATCC53789 16S ribosomal RNA gene, partial sequence	Nostoc sp. ATCC 53789	1481	3132717
42	Rge1	Rubrivivax gelatinosus gene for 16S rRNA, partial sequence	Rubrivivax gelatinosus	1471	3327378
43	Neu1	Nitrosomonas europaea 16S ribosomal RNA gene, complete sequence	Nitrosomonas europaea	1520	3414677
44	Tco1	Trichodesmium contortum 16S ribosomal RNA gene, partial sequence	Trichodesmium contortum	1438	3559793
45	Hsp1	Haloarcula sp. gene for 16S rRNA	Haloarcula sp.	1470	4115513
46	Fnu1	Fusobacterium nucleatum subsp. nucleatum 16S ribosomal RNA gene	Fusobacterium nucleatum subsp. nucleatum	1479	4490387

47	Hsp2	Halobacterium sp. AUS-1 DNA for 16S ribosomal RNA	Halobacterium sp. AUS-1	1465	4580005
48	Sma1	Serratia marcescens 16S ribosomal RNA gene, partial sequence	Serratia marcescens	1494	4883843
49	Smu1	Streptococcus mutans 16S rRNA gene, strain NCTC 10449	Streptococcus mutans	1512	5578899
50	Cfu1	Cytophaga fucicola 16S rRNA gene, type strain NN015860, partial	Cellulophaga fucicola	1472	5701823
51	Efa1	Enterococcus faecalis 16S ribosomal RNA gene, partial sequence	Enterococcus faecalis	1510	5732229
52	Rxy1	Rubrobacter xylanophilus partial 16S rRNA gene	Rubrobacter xylanophilus	1509	6006622
53	Det1	Dehalococcoides ethenogenes 16S ribosomal RNA gene, partial sequence	Dehalococcoides ethenogenes	1434	7524025
54	Bba1	Bdellovibrio bacteriovorus strain TRA2 16S ribosomal RNA gene, partial sequence	Bdellovibrio bacteriovorus	1380	8131956
55	Abe1	Anabaena bergii 16S ribosomal RNA gene, partial sequence	Anabaena bergii	1399	8571953
56	Ato1	Azoarcus tolulyticus strain 4FB10 16S ribosomal RNA gene, partial sequence	Azoarcus tolulyticus	1454	9965637
57	1-Apr	Archaeoglobus profundus 16S ribosomal RNA gene, partial sequence	Archaeoglobus profundus	1480	10444409
58	Sth2	Symbiobacterium thermophilum DNA for 16S rRNA	Symbiobacterium thermophilum	1479	11079170
59	Twh1	Tropheryma whippelii 16S ribosomal RNA gene, partial sequence	Tropheryma whipplei str. Twist	1484	13507079
60	Msu1	Mannheimia sp. 55E 16S ribosomal RNA gene, partial sequence	Mannheimia succiniciproducens	1392	13752230
61	Ade1	Anaeromyxobacter dehalogenans strain 2CP-3 16S ribosomal RNA gene, partial sequence	Anaeromyxobacter dehalogenans	1548	14485223
62	Pda1	Pediococcus damnosus 16S rRNA gene, strain DSM 20331	Pediococcus damnosus	1561	14572624
63	Tth1	Thermotoga thermarum gene for 16S rRNA	Thermotoga thermarum	1471	14587794
64	Ssp2	Synechocystis PCC6805 gene for 16S rRNA, partial sequence	Synechocystis sp. PCC 6805	1437	16215699
65	Asp1	Aquifex sp. Ob6 partial 16S rRNA gene, strain Ob6	Aquifex sp. Ob6	1500	16944830
66	Pph1	Pelodictyon phaeum partial 16S rRNA gene, strain CIB 2401	Pelodictyon phaeum	1453	18076421

67	Por1	Pyrobaculum organotrophum gene for 16S rRNA, partial sequence, strain:JCM 9190	Pyrobaculum organotrophum	1426	18143436
68	Mac1	Methanosarcina acetivorans str. C2A, complete genome	Methanosarcina acetivorans C2A	5751492	19918815
69	Cte1	Chlorobium tepidum TLS, complete genome	Chlorobium tepidum TLS	2154946	21672841
70	Csp1	Chromobacterium sp. 70 16S ribosomal RNA ribosomal RNA gene, complete sequence	Chromobacterium sp. 70	1455	21898816
71	Tco2	Thermococcus coalescens gene for 16S rRNA	Thermococcus coalescens	1489	31711492
72	CFr1	Candidatus Fritschea eriococci strain Elm 16S ribosomal RNA and 23S ribosomal RNA genes, complete sequence	Candidatus Fritschea eriococci	4531	31747860
73	Avi1	Azotobacter vinelandii DSM576 16S ribosomal RNA gene, partial sequence	Azotobacter vinelandii	1398	33242483
74	Mca1	Methylococcus capsulatus partial 16S rRNA gene, strain Texas	Methylococcus capsulatus	1481	47457725
75	Xca1	Xanthomonas campestris pv. coriandri 16S ribosomal RNA gene, partial sequence	Xanthomonas campestris pv. coriandri	1502	47524536
76	Wsu1	Wolinella succinogenes strain ATCC 29543 16S ribosomal RNA, partial sequence	Wolinella succinogenes	1481	47558942
77	Pac1	Propionibacterium acnes isolate WD1 16S ribosomal RNA gene, partial sequence	Propionibacterium acnes	1484	50082577
78	Oih1	Oceanobacillus iheyensis strain MSU3110 16S ribosomal RNA gene, partial sequence	Oceanobacillus iheyensis	1455	50980367
79	Cgl1	Corynebacterium glutamicum strain CICC10226 16S ribosomal RNA gene, partial sequence	Corynebacterium glutamicum	1472	55735425
80	Dsp2	Desulfotalea sp. SFA4 partial 16S rRNA gene	Desulfotalea sp. SFA4	1360	56409896
81	Hmo1	Heliobacillus mobilis gene for 16S rRNA, partial sequence	Heliobacillus mobilis	1482	56692171
82	Plu1	Photorhabdus luminescens subsp. luminescens strain ATCC 29999 16S ribosomal RNA gene, partial sequence	Photorhabdus luminescens subsp. luminescens	1475	58042756
83	Jse1	Jannaschia seosinensis strain CL-SP26 16S ribosomal RNA gene, partial sequence	Jannaschia seosinensis	1384	59804128
84	Bsu1	Bacillus subtilis partial 16S rRNA gene, isolate SMF7	Bacillus subtilis	1522	60098072
85	Zmo1	Zymomonas mobilis strain XW101 16S ribosomal RNA gene, partial sequence	Zymomonas mobilis	1343	61676855

86	Cen1	Coxiella endosymbiont of Amblyomma americanum 16S ribosomal RNA gene, partial sequence	Coxiella endosymbiont of Amblyomma americanum	1469	62999434
87	Bpy1	Burkholderia pyrrocinia isolate RG6-5 16S ribosomal RNA gene, partial sequence	Burkholderia pyrrocinia	1387	63020470
88	Pfe1	Paracoccus ferrooxidans strain BDN-1 16S ribosomal RNA gene, partial sequence	Paracoccus ferrooxidans	1418	63020950
89	Dfi1	Deinococcus sp. CC-FR2-10 16S ribosomal RNA gene, partial sequence	Deinococcus ficus	1453	66394748
90	Cae1	Colwellia aestuarii strain SMK-10 16S ribosomal RNA gene, partial sequence	Colwellia aestuarii	1496	66734249
91	Paq1	Polaromonas aquatica 16S rRNA gene, strain CCUG 39797	Polaromonas aquatica	1407	68051131
92	Msp1	Magnetospirillum sp. PM2411 gene for 16S ribosomal RNA	Magnetospirillum sp. PM2411	1455	68533196
93	Gsp1	Geobacter sp. CLFeRB 16S ribosomal RNA gene, complete sequence	Geobacter sp. CLFeRB	1485	70906126
94	Dar1	Dechloromonas aromatica RCB, complete genome	Dechloromonas aromatica RCB	4501104	71845263
95	Tsp1	Thiobacillus sp. EBD bloom 16S ribosomal RNA gene, partial sequence	Thiobacillus sp. EBD bloom	1408	77455743
96	Rsp1	Rhodoferax sp. PIC-C33 16S ribosomal RNA gene, partial sequence	Rhodoferax sp. PIC-C33	1473	77994459
97	Nvu1	Nitrobacter vulgaris partial 16S rRNA gene, type strain DSM 10236T	Nitrobacter vulgaris	1441	78271519
98	Isp1	Idiomarina sp. AN-BI1D 16S rRNA gene, strain AN-BI1D	Idiomarina sp. AN-BI1D	1371	82173853
99	Hsp3	Haemophilus sp. oral clone ASCB01 16S ribosomal RNA gene, complete sequence	Haemophilus sp. oral clone ASCB01	1555	82582885
100	Mfl1	Methylobacillus flagellatus strain KT 16S ribosomal RNA gene, partial sequence	Methylobacillus flagellatus	1400	83272653
101	Dal1	Desulfuromonas alkaliphilus strain Z-0531 16S ribosomal RNA gene, partial sequence	Desulfuromonas alkaliphilus	1504	83595213
102	Tfu1	Thermobifida fusca 16S ribosomal RNA gene, partial sequence	Thermobifida fusca	1448	83722780
103	Epe1	Erwinia persicina strain GS04 16S ribosomal RNA gene, partial sequence	Erwinia persicina	1470	86161535
104	Rsp2	Ralstonia sp. PHD-12 16S ribosomal RNA gene, partial sequence	Ralstonia sp. PHD-12	1493	87312602

105	Rsp3	Rhodopseudomonas sp. TUT3627 gene for 16S rRNA, partial sequence	Rhodopseudomonas sp. TUT3627	1482	88606680
106	Csp2	Campylobacter sp. 150B 16S ribosomal RNA gene, partial sequence	Campylobacter sp. 150B	1390	89114015
107	Bdo1	Bacteroides dorei gene for 16S rRNA, partial sequence, strain:JCM 13472	Bacteroides dorei	1490	89242107
108	Msp2	Microbulbifer sp. KBB-1 16S ribosomal RNA gene, partial sequence	Microbulbifer sp. KBB-1	1500	89258479
109	Nsp2	Neisseria sp. J01 16S ribosomal RNA gene, partial sequence	Neisseria sp. J01	1502	89277202
110	Ssp3	Sphingopyxis sp. EMB 355 16S ribosomal RNA gene, partial sequence	Sphingopyxis sp. EMB 355	1407	89348125
111	Raz1	Rhodobacter azotoformans strain S3 16S ribosomal RNA gene, partial sequence	Rhodobacter azotoformans	1459	89357194
112	Rsp4	Rhizobium sp. As-2 16S ribosomal RNA gene, partial sequence	Rhizobium sp. As-2	1391	89953754
113	Rsp5	Rhizobium sp. Lv6.1Se 16S ribosomal RNA gene, partial sequence	Rhizobium sp. Lv6.1Se	1446	89954503
114	Psp1	Psychrobacter sp. AM11 partial 16S rRNA gene, isolate AM11	Psychrobacter sp. AM11	1501	90071222
115	Mti1	Mesorhizobium tianshanense strain RCAN08 16S ribosomal RNA gene, partial sequence	Mesorhizobium tianshanense	1479	90186421
116	Hca1	Helicobacter canadensis strain L231 16S ribosomal RNA gene, partial sequence	Helicobacter canadensis	1407	90194336
117	Bce1	Brucella cetaceae partial 16S rRNA gene, strain NCTC 12891	Brucella cetaceae	1430	90818672
118	Pmi1	Proteus mirabilis strain O 16S ribosomal RNA gene, partial sequence	Proteus mirabilis	1451	90856205
119	Swo1	Syntrophomonas wolfei subsp. methybutyica strain 5J-1 16S ribosomal RNA gene, partial sequence	Syntrophomonas wolfei subsp. methylbutyratica	1602	90903520
120	Csp3	Chromohalobacter sp. HS208 16S ribosomal RNA gene, partial sequence	Chromohalobacter sp. HS208	1381	90994975
121	Asp2	Arthrobacter sp. m3 16S ribosomal RNA gene, partial sequence	Arthrobacter sp. m3	1484	91771955
122	Dde1	Desulfovibrio desulfuricans isolate SRB16 16S ribosomal RNA gene, complete sequence	Desulfovibrio desulfuricans	1521	91974481

123	Esp1	Erythrobacter sp. CNU001 16S ribosomal RNA gene, partial sequence	Erythrobacter sp. CNU001	1480	91982998
124	Nre1	Novosphingobium subarcticum gene for 16S rRNA, partial sequence, strain:T7b	Novosphingobium resinovorum	1400	92019078
125	Bsp2	Brevibacterium sp. CNJ737 PL04 16S ribosomal RNA gene, partial sequence	Brevibacterium sp. CNJ737 PL04	1493	92091011
126	Nsp3	Nocardia sp. CNS044 PL04 16S ribosomal RNA gene, partial sequence	Nocardia sp. CNS044 PL04	1484	92091036
127	Nsp4	Nocardioides sp. CNJ892 PL04 16S ribosomal RNA gene, partial sequence	Nocardioides sp. CNJ892 PL04	1482	92091040
128	Ssp4	Shewanella sp. BSi20587 16S ribosomal RNA gene, partial sequence	Shewanella sp. BSi20587	1504	93009053
129	Ape1	Aeropyrum pernix genes for 16S rRNA, 23S rRNA and ITS region, complete sequence, strain:OH3	Aeropyrum pernix	6220	18250957
130	Lla1	L.lactis ribosomal RNA operon encoding 16S, 5S, and 23S ribosomal RNA, transfer RNA-Ala and transfer RNA-Asn	Lactococcus lactis	5953	44070
131	Rba1	Rhodopirellula baltica SH 1, complete genome	Rhodopirellula baltica SH 1	7145576	32470666
132	Gvi1	Gloeobacter violaceus PCC 7421, complete genome	Gloeobacter violaceus PCC 7421	4659019	37519569
133	Tth2	Thermus thermophilus HB27, complete genome	Thermus thermophilus HB27	1894877	46198308
134	Tel1	Thermosynechococcus elongatus BP-1 DNA, complete genome	Thermosynechococcus elongatus BP-1	2593857	47118315
135	Sus1	Solibacter usitatus Ellin6076	Solibacter usitatus Ellin6076	113720	67861827
136	Cwa1	Crocosphaera watsonii WH 8501	Crocosphaera watsonii WH 8501	347519	67921358
137	Msp3	Magnetococcus sp. MC-1	Magnetococcus sp. MC-1	51219	69259430
138	Rru1	Rhodospirillum rubrum ATCC 11170, complete genome	Rhodospirillum rubrum ATCC 11170	4352825	83591340

Table 5. The protein sequences representing the Heavy Metal P-type ATPase Family

	Abbr.	Sequence Description	Organism	#AA	GI#	Group	King.	Cluster #	Ave AA length for cluster (2 decimal place)
1	Neu1	heavy metal translocating P-type ATPase	Nitrosomonas eutropha C91	708	114330115	Beta-proteobacteria	B	1A	all 1: 760.16 ± 58.83
2	Nwi1	Heavy metal translocating P-type ATPase	Nitrobacter winogradskyi Nb-255	712	75677312	Alpha-proteobacteria	B	1A	
3	Nha2	Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	711	92119182	Alpha-proteobacteria	B	1A	
4	Rpa3	putative cation-transporting P-type ATPase	Rhodopseudomonas palustris CGA009	709	39936323	Alpha-proteobacteria	B	1A	
5	Pla2	putative cation-transporting P-type ATPase	Parvibaculum lavamentivorans DS-1	619	121523945	Alpha-proteobacteria	B	1A	
6	Ssp1	heavy metal translocating P-type ATPase	Shewanella sp. W3-18-1	884	120597953	Gamma-proteobacteria	B	1A	
7	Pcr1	Heavy metal translocating P-type ATPase	Psychrobacter cryohalolentis K5	738	93006175	Gamma-proteobacteria	B	1A	
8	Dge2	Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	722	94972219	Deinococci	B	1A	
9	Dge1	Heavy metal translocating P-type ATPase	Deinococcus geothermalis DSM 11300	793	94972049	Deinococci	B	1A	
10	Dra1	cation-transporting P-type ATPase	Deinococcus radiodurans R1	728	15807741	Deinococci	B	1A	
11	Atu1	hypothetical protein AGR_C_1540	Agrobacterium tumefaciens str. C58	916	15888184	Alpha-proteobacteria	B	1A	
12	Rsp2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rhodobacter sphaeroides ATCC 17029	740	83371845	Alpha-proteobacteria	B	1A	
13	Pde1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Paracoccus denitrificans PD1222	758	69934127	Alpha-proteobacteria	B	1A	
14	Pde2	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Paracoccus denitrificans PD1222	732	69936175	Alpha-proteobacteria	B	1A	
15	Ogr1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Oceanicola granulosus HTCC2516	726	89069243	Alpha-proteobacteria	B	1A	
16	Ssp5	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Sulfitobacter sp. EE-36	576	83943424	Alpha-proteobacteria	B	1A	
17	Oba1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Oceanicola batsensis HTCC2597	762	84499300	Alpha-proteobacteria	B	1A	

18	Rsp4	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Rhodobacter sphaeroides ATCC 17025	803	83368473	Alpha-proteobacteria	B	1A
19	Rru1	(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Rhodospirillum rubrum ATCC 11170	777	83592363	Alpha-proteobacteria	B	1B
20	Msp3	heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	734	110635182	Alpha-proteobacteria	B	1B
21	Xau1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Xanthobacter autotrophicus Py2	735	89358240	Alpha-proteobacteria	B	1B
22	Nha1	Heavy metal translocating P-type ATPase	Nitrobacter hamburgensis X14	734	92109658	Alpha-proteobacteria	B	1B
23	Rpa1	Heavy metal translocating P-type ATPase	Rhodopseudomonas palustris BisB5	726	91978108	Alpha-proteobacteria	B	1B
24	Bsp4	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bradyrhizobium sp. BTAi1	746	78692510	Alpha-proteobacteria	B	1B
25	Rpa2	Heavy metal translocating P-type ATPase	Rhodopseudomonas palustris BisB5	716	91975984	Alpha-proteobacteria	B	1B
26	Bsp5	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bradyrhizobium sp. BTAi1	762	78697507	Alpha-proteobacteria	B	1B
27	Ret1	probable heavy metal-transporting ATPase protein	Rhizobium etli CFN 42	748	86359309	Alpha-proteobacteria	B	1B
28	Rle1	putative transmembrane cation transport ATPase	Rhizobium leguminosarum bv. viciae 3841	756	116253999	Alpha-proteobacteria	B	1B
29	Bsu1	cadmium-translocating P-type ATPase	Brucella suis 1330	814	23502866	Alpha-proteobacteria	B	1B
30	Sme1	putative heavy metal transporting ATPase protein	Sinorhizobium medicae WSM419	744	113871168	Alpha-proteobacteria	B	1B
31	Sme2	PUTATIVE HEAVY METAL TRANSPORTING ATPASE PROTEIN	Sinorhizobium meliloti 1021	743	15963877	Alpha-proteobacteria	B	1B
32	Msp2	heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	955	110346943	Alpha-proteobacteria	B	1B
33	Asp2	metal-transporting P-type ATPase	Aurantimonas sp. SI85-9A1	724	90419255	Alpha-proteobacteria	B	1B
34	Mlo1	probable metal-transporting P-type ATPase	Mesorhizobium loti MAFF303099	749	13472245	Alpha-proteobacteria	B	1B
35	Ppr5	putative cation transport ATPase	Photobacterium profundum 3TCK	718	90413947	Gamma-proteobacteria	B	1C

36	Vfi1	lead, cadmium, zinc and mercury transporting ATPase	Vibrio fischeri ES114	771	59712129	Gamma-proteobacteria	B	1C
37	Vch1	cation transport ATPase, E1-E2 family	Vibrio cholerae O1 biovar eltor str. N16961	768	15641046	Gamma-proteobacteria	B	1C
38	Vsp2	cation transport ATPase, E1-E2 family protein	Vibrio sp. MED222	785	86146958	Gamma-proteobacteria	B	1C
39	Vvu1	cation transport ATPase	Vibrio vulnificus YJ016	797	37680562	Gamma-proteobacteria	B	1C
40	Vsp1	hypothetical protein VEx2w_02001362	Vibrio sp. Ex25	768	116186099	Gamma-proteobacteria	B	1C
41	Vpa1	cation transport ATPase, E1-E2 family	Vibrio parahaemolyticus RIMD 2210633	768	28897733	Gamma-proteobacteria	B	1C
42	Van1	putative cation transport ATPase	Vibrio angustum S14	859	90579060	Gamma-proteobacteria	B	1C
43	Ppr3	putative cation transport ATPase	Photobacterium profundum 3TCK	828	90413284	Gamma-proteobacteria	B	1C
44	Ppr4	putative cation transport ATPase	Photobacterium profundum SS9	801	54308413	Gamma-proteobacteria	B	1C
45	Ahy1	lead, cadmium, zinc and mercury transporting ATPase	Aeromonas hydrophila subsp. hydrophila ATCC 7966	832	117621449	Gamma-proteobacteria	B	1C
46	Msp4	heavy metal translocating P-type ATPase	Marinomonas sp. MWYL1	803	118750511	Gamma-proteobacteria	B	1C
47	Eca1	heavy metal-transporting ATPase	Erwinia carotovora subsp. atroseptica SCRI1043	787	50123272	Gamma-proteobacteria	B	1C
48	Plu1	lead, cadmium, zinc and mercury transporting ATPase	Photorhabdus luminescens subsp. laumondii TTO1	761	37527955	Gamma-proteobacteria	B	1C
49	Pmi1	putative P-type cation-translocating membrane ATPase	Proteus mirabilis	692	2624376	Gamma-proteobacteria	B	1C
50	Spr1	heavy metal translocating P-type ATPase	Serratia proteamaculans 568	771	118067336	Gamma-proteobacteria	B	1C
51	Yps1	P-type heavy metal efflux ATPase, ATZN	Yersinia pseudotuberculosis IP 32953	788	51594570	Gamma-proteobacteria	B	1C
52	Yen1	putative P-type cation-translocating membrane ATPase	Yersinia enterocolitica subsp. enterocolitica 8081	776	123440611	Gamma-proteobacteria	B	1C
53	Yfr1	COG2217: Cation transport ATPase	Yersinia frederiksenii ATCC 33641	774	77973654	Gamma-proteobacteria	B	1C
54	Ymo1	COG2217: Cation transport ATPase	Yersinia mollaretii ATCC 43969	775	77961783	Gamma-proteobacteria	B	1C
55	Yin1	COG2217: Cation transport ATPase	Yersinia intermedia ATCC 29909	775	77978516	Gamma-proteobacteria	B	1C
56	Esp1	heavy metal translocating P-type ATPase	Enterobacter sp. 638	728	118741095	Gamma-proteobacteria	B	1D

57	Sen1	P-type ATPase family, Pb/Cd/Zn/Hg transporting ATPase	Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	732	62182075	Gamma-proteobacteria	B	1D	
58	Eco1	Lead, cadmium, zinc and mercury transporting ATPase	Escherichia coli CFT073	732	26250084	Gamma-proteobacteria	B	1D	
59	Cvi1	lead,cadmium,zinc and mercury transporting ATPase	Chromobacterium violaceum ATCC 12472	764	34496609	Beta-proteobacteria	B	2A	all 2: 775.96 ± 63.34
60	Rfe1	Heavy metal translocating P-type ATPase	Rhodoferax ferrireducens T118	812	89899662	Beta-proteobacteria	B	2B	
61	Mfl1	Heavy metal translocating P-type ATPase	Methylobacillus flagellatus KT	748	91776900	Beta-proteobacteria	B	2C	
62	Ppr1	heavy metal translocating P-type ATPase	Pelobacter propionicus DSM 2379	791	118577287	Delta-proteobacteria	B	2C	
63	Rme1	Heavy metal translocating P-type ATPase	Ralstonia metallidurans CH34	984	94311241	Beta-proteobacteria	B	2C	
64	Psp1	Heavy metal translocating P-type ATPase	Polynucleobacter sp. QLW-P1DMWA-1	726	116269477	Beta-proteobacteria	B	2C	
65	Avi1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Azotobacter vinelandii AvOP	728	67158452	Gamma-proteobacteria	B	2C	
66	Rpi1	heavy metal translocating P-type ATPase	Ralstonia pickettii 12J	814	121530669	Beta-proteobacteria	B	2C	
67	Asp1	heavy metal translocating P-type ATPase	Acidovorax sp. JS42	799	121593647	Beta-proteobacteria	B	2C	
68	Rme3	Pb-efflux ATPase	Ralstonia metallidurans CH34	799	56130719	Beta-proteobacteria	B	2C	
69	Cte2	heavy metal translocating P-type ATPase	Comamonas testosteroni KF-1	801	118049056	Beta-proteobacteria	B	2C	
70	Kpn1	PbrA	Klebsiella pneumoniae	801	38639700	Gamma-proteobacteria	B	2C	
71	Pme1	heavy metal translocating P-type ATPase	Pseudomonas mendocina ymp	734	118058132	Gamma-proteobacteria	B	2C	
72	Pae1	hypothetical protein PaerP_01003458	Pseudomonas aeruginosa PA7	748	94414702	Gamma-proteobacteria	B	2C	
73	Psy1	cadmium-translocating P-type ATPase	Pseudomonas syringae pv. phaseolicola 1448A	754	71736779	Gamma-proteobacteria	B	2C	
74	Pfl1	cadmium-translocating P-type ATPase	Pseudomonas fluorescens Pf-5	818	70733177	Gamma-proteobacteria	B	2C	
75	Pfl2	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Pseudomonas fluorescens PfO-1	769	77461592	Gamma-proteobacteria	B	2C	
76	Ppu2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Pseudomonas putida F1	750	82740121	Gamma-proteobacteria	B	2C	
77	Ppu1	Heavy metal translocating P-type ATPase	Pseudomonas putida W619	750	119856322	Gamma-proteobacteria	B	2C	

78	Pen1	cadmium translocating P-type ATPase	Pseudomonas entomophila L48	748	104784142	Gamma-proteobacteria	B	2C
79	Rme2	Heavy metal translocating P-type ATPase	Ralstonia metallidurans CH34	794	94313516	Beta-proteobacteria	B	2C
80	Reu1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Ralstonia eutropha JMP134	783	73537805	Beta-proteobacteria	B	2C
81	Reu2	putative heavy metal efflux P-type ATPase	Ralstonia eutropha H16	791	116695583	Beta-proteobacteria	B	2C
82	Dar1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	743	71907877	Beta-proteobacteria	B	2C
83	Bph1	heavy metal translocating P-type ATPase	Burkholderia phytofirmans PsJN	790	118043038	Beta-proteobacteria	B	2C
84	Bps2	putative heavy metal resistance membrane ATPase	Burkholderia pseudomallei K96243	836	53720986	Beta-proteobacteria	B	2C
85	Bmu1	heavy metal translocating P-type ATPase	Burkholderia multivorans ATCC 17616	862	118720004	Beta-proteobacteria	B	2C
86	Bvi1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Burkholderia vietnamiensis G4	864	67547414	Beta-proteobacteria	B	2C
87	Bsp2	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Burkholderia sp. 383	866	78064781	Beta-proteobacteria	B	2C
88	Sgl1	putative cation transport ATPase	Sodalis glossinidius str. 'morsitans'	780	85059831	Gamma-proteobacteria	B	2D
89	Bav1	cadmium-transporting ATPase	Bordetella avium 197N	753	115422714	Beta-proteobacteria	B	2D
90	Bbr1	putative membrane transport ATPase	Bordetella bronchiseptica RB50	778	33601483	Beta-proteobacteria	B	2D
91	Ppr2	heavy metal translocating P-type ATPase	Pelobacter propionicus DSM 2379	783	118579269	Delta-proteobacteria	B	2E
92	Rso1	Lead, cadmium, zinc and mercury transporting ATPase	Ralstonia solanacearum UW551	784	83749132	Beta-proteobacteria	B	2E
93	Rpi2	putative metal-transporting P-type ATPase transmembrane protein	Ralstonia pickettii 12J	783	121530307	Beta-proteobacteria	B	2E
94	Ssp2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocatingP-type ATPase:Heavy metal translocating P-type ATPase	Synechococcus sp. WH 5701	700	87302472	Cyanobacteria	B	2F
95	Aav1	heavy metal translocating P-type ATPase	Acidovorax avenae subsp. citrulli AAC00-1	629	120610495	Beta-proteobacteria	B	2F

96	Pna1	heavy metal translocating P-type ATPase	Polaromonas naphthalenivorans CJ2	745	121606885	Beta-proteobacteria	B	2F	
97	Dar2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Dechloromonas aromatica RCB	739	71908244	Beta-proteobacteria	B	2F	
98	Cte3	heavy metal translocating P-type ATPase	Comamonas testosteroni KF-1	743	118053586	Beta-proteobacteria	B	2F	
99	Asp3	heavy metal translocating P-type ATPase	Acidovorax sp. JS42	753	121593734	Beta-proteobacteria	B	2F	
100	Asp6	heavy metal translocating P-type ATPase	Acidovorax sp. JS42	673	121594932	Beta-proteobacteria	B	2F	
101	Cte4	heavy metal translocating P-type ATPase	Comamonas testosteroni KF-1	711	118053356	Beta-proteobacteria	B	2F	
102	Dac1	heavy metal translocating P-type ATPase	Delftia acidovorans SPH-1	732	118732697	Beta-proteobacteria	B	2F	
103	Cte1	heavy metal translocating P-type ATPase	Comamonas testosteroni KF-1	975	118049985	Beta-proteobacteria	B	2G	
104	Lpn1	cadmium translocating P-type ATPase CadA	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	729	52841243	Gamma-proteobacteria	B	2G	
105	Lpn2	hypothetical protein lpl1049	Legionella pneumophila str. Lens	713	54293990	Gamma-proteobacteria	B	2G	
106	Bma1	Cd/Co/Hg/Pb/Zn-translocating P-type ATPase	Blastopirellula marina DSM 3645	743	87311031	Planctomycetes	B	3	all 3: 774.78 ± 97.92
107	Ftu1	heavy metal cation transport ATPase	Francisella tularensis subsp. novicida U112	721	118497001	Gamma-proteobacteria	B	3	
108	Sfu1	heavy metal translocating P-type ATPase	Syntrophobacter fumaroxidans MPOB	737	116751190	Delta-proteobacteria	B	3	
109	Msp1	heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	1022	110347018	Alpha-proteobacteria	B	3	
110	Mma1	COG2217: Cation transport ATPase	Magnetospirillum magnetotacticum MS-1	733	46202126	Alpha-proteobacteria	B	3	
111	Mma3	Cation transport ATPase	Magnetospirillum magneticum AMB-1	684	83310111	Alpha-proteobacteria	B	3	
112	Rsp3	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Roseobacter sp. MED193	765	86136996	Alpha-proteobacteria	B	3	
113	Rsp1	cadmium-translocating P-type ATPase	Roseobacter sp. MED193	784	86139673	Alpha-proteobacteria	B	3	
114	Dsh1	heavy metal translocating P-type ATPase	Dinoroseobacter shibae DFL 12	784	118738365	Alpha-proteobacteria	B	3	
115	Chy1	cation-transporting ATPase, E1-E2 family, selenocysteine-containing	Carboxydothermus hydrogenoformans Z-2901	686	78044655	Firmicutes (LowGC G+)	B	4	663.5 ± 31.82

116	Pam1	putative cadmium-transporting ATPase	Candidatus Protochlamydia amoebophila UWE25	641	46445921	Chlamydiae	B	4	
117	Lin1	cation-transporting ATPase, P-type	Lawsonia intracellularis PHE/MN1-00	688	94987040	Delta-proteobacteria	B	5A	all 5: 695.47 ± 59.88
118	Lsa2	Zinc-transporting ATPase	Lactobacillus salivarius subsp. salivarius UCC118	643	90962608	Firmicutes (LowGC G+)	B	5B	
119	Wsu1	YVGW PROTEIN	Wolinella succinogenes DSM 1740	707	34557497	Epsilon-proteobacteria	B	5C	
120	Efa1	cadmium-translocating P-type ATPase	Enterococcus faecalis V583	700	29375967	Firmicutes (LowGC G+)	B	5D	
121	Efa2	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Enterococcus faecium DO	615	69247326	Firmicutes (LowGC G+)	B	5D	
122	Fnu1	Zinc-transporting ATPase	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	614	19703603	Fusobacteria	B	5E	
123	Cce1	cadmium-translocating P-type ATPase	Clostridium cellulolyticum H10	618	118726401	Firmicutes (LowGC G+)	B	5F	
124	Bce1	Zinc-transporting ATPase	Bacillus cereus ATCC 14579	788	30018782	Firmicutes (LowGC G+)	B	5F	
125	Bce2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bacillus cereus subsp. cytotoxis NVH 391-98	785	89201527	Firmicutes (LowGC G+)	B	5F	
126	Bsu2	hypothetical protein BSU33490	Bacillus subtilis subsp. subtilis str. 168	702	16080402	Firmicutes (LowGC G+)	B	5F	
127	Bli1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase,Heavy metal translocating P-type ATPase	Bacillus licheniformis ATCC 14580	703	52081840	Firmicutes (LowGC G+)	B	5F	
128	Cth1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium thermocellum ATCC 27405	707	67916589	Firmicutes (LowGC G+)	B	5F	
129	Csp3	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium sp. OhILAs	866	106893409	Firmicutes (LowGC G+)	B	5F	
130	Dha1	hypothetical protein DSY0829	Desulfitobacterium hafniense Y51	784	89893575	Firmicutes (LowGC G+)	B	5F	

131	Csa1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Caldicellulosiruptor saccharolyticus DSM 8903	790	82500189	Firmicutes (LowGC G+)	B	5F
132	Ame2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Alkaliphilus metalliredigenes QYMF	788	77686825	Firmicutes (LowGC G+)	B	5F
133	Csp4	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium sp. OhILAs	551	106894859	Firmicutes (LowGC G+)	B	5F
134	Cbe1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium beijerincki NCIMB 8052	809	82746499	Firmicutes (LowGC G+)	B	5F
135	Cno1	cadmium-translocating P-type ATPase	Clostridium novyi NT	711	118443354	Firmicutes (LowGC G+)	B	5F
136	Tet1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Thermoanaerobacter ethanolicus ATCC 33223	699	76796108	Firmicutes (LowGC G+)	B	5F
137	Tet2	Heavy metal translocating P-type ATPase	Thermoanaerobacter ethanolicus X514	658	114845135	Firmicutes (LowGC G+)	B	5F
138	Cdi1	putative heavy-metal-transporting ATPase	Clostridium difficile 630	795	115249322	Firmicutes (LowGC G+)	B	5F
139	Cpe1	cadmium-translocating P-type ATPase	Clostridium perfringens ATCC 13124	738	110800762	Firmicutes (LowGC G+)	B	5F
140	Cte5	Zinc-transporting ATPase	Clostridium tetani E88	650	28211582	Firmicutes (LowGC G+)	B	5F
141	Cac1	Cation transport P-type ATPase	Clostridium acetobutylicum ATCC 824	699	15895509	Firmicutes (LowGC G+)	B	5F
142	Mla1	hypothetical protein Mlab_1065	Methanocorpusculum labreanum Z	626	124363426	Euryarchaeota	A	5G
143	Esi1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Exiguobacterium sibiricum 255-15	707	68055013	Firmicutes (LowGC G+)	B	5G
144	Lsa1	Putative heavy metal-transporting P-type ATPase	Lactobacillus sakei subsp. sakei 23K	696	81428039	Firmicutes (LowGC G+)	B	5G
145	Hor1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Halothermothrix orenii H 168	745	89211379	Firmicutes (LowGC G+)	B	5H
146	Pab1	cation-transporting ATPase, P-type	Pyrococcus abyssi GE5	689	14521140	Euryarchaeota	A	5H

147	Mma4	cation transport ATPase	Methanococcus maripaludis S2	691	45357779	Euryarchaeota	A	5H
148	Sth3	putative zinc/cadmium-transporting ATPase	Symbiobacterium thermophilum IAM 14863	710	51892763	Actinobacteria (HighGC G+)	B	5I
149	Mth1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	Moorella thermoacetica ATCC 39073	700	83591026	Firmicutes (LowGC G+)	B	5I
150	Dha2	heavy metal translocating P-type ATPase	Desulfitobacterium hafniense DCB-2	800	109648787	Firmicutes (LowGC G+)	B	5I
151	Dha3	hypothetical protein DSY4631	Desulfitobacterium hafniense Y51	808	89897377	Firmicutes (LowGC G+)	B	5I
152	Npu2	COG2217: Cation transport ATPase	Nostoc punctiforme PCC 73102	664	23126943	Cyanobacteria	B	5I
153	Ssp7	hypothetical protein slr0798	Synechocystis sp. PCC 6803	721	16331908	Cyanobacteria	B	5I
154	Obr1	P type ATPase BXA1	Oscillatoria brevis	660	22506637	Cyanobacteria	B	5I
155	Lsp2	hypothetical protein L8106_12725	Lyngbya sp. PCC 8106	651	119486614	Cyanobacteria	B	5I
156	Chu1	cation-transporting ATPase; possible zinc-transporting ATPase	Cytophaga hutchinsonii ATCC 33406	664	110637914	Bacteroidetes	B	5J
157	Chu2	cation-transporting ATPase	Cytophaga hutchinsonii ATCC 33406	667	110637474	Bacteroidetes	B	5J
158	Fjo1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Flavobacterium johnsoniae UW101	663	90587529	Bacteroidetes	B	5J
159	Orh1	heavy metal transporting ATPase	Ornithobacterium rhinotracheale	675	47059343	Bacteroidetes	B	5J
160	Spu1	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Shewanella putrefaciens CN-32	682	77815133	Gamma-proteobacteria	B	5J
161	Lbl2	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Leeuwenhoekiella blandensis MED217	653	86141422	Bacteroidetes	B	5J
162	Tsp1	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Tenacibaculum sp. MED152 (Polaribacter dokdonensis MED 152)	651	86133069	Bacteroidetes	B	5J
163	Rbi1	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Robiginitalea biformata HTCC2501	654	88804161	Bacteroidetes	B	5J
164	Gfo2	heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-ty pe ATPase	Gramella forsetii KT0803	656	120434396	Bacteroidetes	B	5J

165	Csp5	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Cellulophaga sp. MED134 (Dokdonia donghaensis MED 134)	655	86130586	Bacteroidetes	B	5J	
166	Lbl3	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Leeuwenhoekiella blandensis MED217	643	86141940	Bacteroidetes	B	5J	
167	Bfr1	putative transmembrane cation transport P type ATPase	Bacteroides fragilis NCTC 9343	648	60680801	Bacteroidetes	B	5J	
168	Bth2	cation-transporting ATPase, P-type, putative zinc-transporting ATPase	Bacteroides thetaiotaomicron VPI-5482	652	29347922	Bacteroidetes	B	5J	
169	Tde1	cadmium-translocating P-type ATPase	Treponema denticola ATCC 35405	643	42526892	Spirochaetes	B	5K	
170	Csa2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Caldicellulosirupt or saccharolyticus DSM 8903	698	82499949	Firmicutes (LowGC G+)	B	5K	
171	Hhe1	cation transport ATPase	Helicobacter hepaticus ATCC 51449	695	32266085	Epsilon-proteobacteria	B	5L	
172	Hac1	heavy metal-transporting ATPase	Helicobacter acinonychis str. Sheeba	686	109947461	Epsilon-proteobacteria	B	5L	
173	Hfe1	Cadmium, zinc and cobalt-transporting ATPase	Helicobacter felis	681	10720043	Epsilon-proteobacteria	B	5L	
174	Ota1	Hma1 cadmium/zinc-transporting ATPase, putative (IC)	Ostreococcus tauri	1052	119358808	Viridiplantae	E	6	989.33 ± 175.06
175	Mtr1	Cof protein; ATPase, E1-E2 type	Medicago truncatula	829	92875650	Viridiplantae	E	6	
176	Aha2	putative cadmium/zinc-transporting ATPase 3	Arabidopsis halleri subsp. halleri	757	37665197	Viridiplantae	E	6	
177	Ath1	putative heavy-metal transporter	Arabidopsis thaliana	951	20384833	Viridiplantae	E	6	
178	Tca2	P1B-type heavy metal transporting ATPase	Thlaspi caerulescens	1186	46361991	Viridiplantae	E	6	
179	Aha1	P1B-type ATPase 4	Arabidopsis halleri subsp. gemmifera	1161	63056225	Viridiplantae	E	6	
180	Csp2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Clostridium sp. OhILAs	615	106894562	Firmicutes (LowGC G+)	B	7A	all 7: 640.14 ± 30.66
181	Lme1	Cation transport ATPase	Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	616	116326496	Firmicutes (LowGC G+)	B	7A	
182	Lsa3	Lead, cadmium, zinc and mercury transporting ATPase	Lactobacillus salivarius subsp. salivarius UCC118	634	90962758	Firmicutes (LowGC G+)	B	7A	

183	met2	cation-transporting P-type ATPase	uncultured methanogenic archaeon	647	116077934	Euryarchaeota	A	7B	
184	Hma5	cadmium transporting P-type ATPase	Haloarcula marismortui ATCC 43049	639	55376427	Euryarchaeota	A	7B	
185	Cgl2	cation transport ATPase	Corynebacterium glutamicum ATCC 13032	625	19552400	Actinobacteria (HighGC G+)	B	7C	
186	Cgl1	cation transport ATPase	Corynebacterium glutamicum ATCC 13032	625	19554164	Actinobacteria (HighGC G+)	B	7C	
187	Cef2	putative copper-transporting ATPase	Corynebacterium efficiens YS-314	578	25026633	Actinobacteria (HighGC G+)	B	7C	
188	Nsp4	heavy metal translocating P-type ATPase	Nocardioides sp. JS614	656	119715897	Actinobacteria (HighGC G+)	B	7C	
189	Bli3	COG2217: Cation transport ATPase	Brevibacterium linens BL2	666	62422868	Actinobacteria (HighGC G+)	B	7C	
190	Rxy2	Heavy metal translocating P-type ATPase	Rubrobacter xylanophilus DSM 9941	639	108804635	Actinobacteria (HighGC G+)	B	7C	
191	Pde3	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Paracoccus denitrificans PD1222	687	69937554	Alpha-proteobacteria	B	7C	
192	Ame1	ATPase, E1-E2 type:Copper-translocating P-type ATPase:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Alkaliphilus metalliredigenes QYMF	636	77683982	Firmicutes (LowGC G+)	B	7C	
193	Hau1	Heavy metal translocating P-type ATPase	Herpetosiphon aurantiacus ATCC 23779	699	113940140	Chloroflexi	B	7C	
194	Efa4	P-type ATPase cation exporter	Enterococcus faecium	645	42521343	Firmicutes (LowGC G+)	B	8A	639.46 ± 14.22
195	Esi2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Exiguobacterium sibiricum 255-15	649	68054678	Firmicutes (LowGC G+)	B	8A	
196	Bcl2	cation-transporting ATPase	Bacillus clausii KSM-K16	637	56965111	Firmicutes (LowGC G+)	B	8A	
197	Bha2	cadmium-transporting ATPase	Bacillus halodurans C-125	637	15613307	Firmicutes (LowGC G+)	B	8A	
198	Bsp6	cadmium-transporting ATPase	Bacillus sp. NRRL B-14911	678	89097328	Firmicutes (LowGC G+)	B	8A	
199	Bce3	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bacillus cereus subsp. cytotoxis NVH 391-98	641	89199965	Firmicutes (LowGC G+)	B	8A	

200	Bth1	cation-transporting ATPase, P-type	Bacillus thuringiensis str. Al Hakam	641	118476163	Firmicutes (LowGC G+)	B	8A	
201	Bsu3	hypothetical protein BSU13850	Bacillus subtilis subsp. subtilis str. 168	637	16078449	Firmicutes (LowGC G+)	B	8A	
202	Bli2	YkvW	Bacillus licheniformis ATCC 14580	635	52785356	Firmicutes (LowGC G+)	B	8A	
203	Lwe1	heavy metal-translocating P-type ATPase	Listeria welshimeri serovar 6b str. SLCC5334	625	116872028	Firmicutes (LowGC G+)	B	8A	
204	Lmo2	cadmium-translocating P-type ATPase	Listeria monocytogenes str. 4b H7858	626	47091600	Firmicutes (LowGC G+)	B	8A	
205	Efa3	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Enterococcus faecium DO	642	69247967	Firmicutes (LowGC G+)	B	8A	
206	Spy1	lead, cadmium, zinc and mercury transporting ATPase	Streptococcus pyogenes MGAS9429	620	94988846	Firmicutes (LowGC G+)	B	8A	
207	Ath2	HMA1; copper-exporting ATPase	Arabidopsis thaliana	819	15235511	Viridiplantae	E	8B	820.5 ± 2.12
208	Osa1	Os06g0690700	Oryza sativa (japonica cultivar-group)	822	115469636	Viridiplantae	E	8B	
209	Ctr1	Metal Transport P-type ATPase	Chlamydia trachomatis D/UW-3/CX	659	15605460	Chlamydiae	B	8C	664 ± 12.70
210	Cpn1	cation-transporting ATPase, E1-E2 family	Chlamydophila pneumoniae AR39	683	16752171	Chlamydiae	B	8C	
211	Cab1	putative cation transport related membrane protein	Chlamydophila abortus S26/3	657	62185469	Chlamydiae	B	8C	
212	Cfe1	cadmium/zinc cation transporting ATPase	Chlamydophila felis Fe/C-56	657	89897922	Chlamydiae	B	8C	
213	Cef1	putative cadmium-transporting atpase	Corynebacterium efficiens	695	23578015	Actinobacteria (HighGC G+)	B	9	793.72 ± 81.68
214	Lbl1	cadmium translocating P-type ATPase	Leeuwenhoekiella blandensis MED217	850	86141933	Bacteroidetes	B	9	
215	Gfo1	heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-ty pe ATPase	Gramella forsetii KT0803	661	120434636	Bacteroidetes	B	9	
216	Sru1	cadmium efflux ATPase	Salinibacter ruber DSM 13855	824	83815585	Bacteroidetes	B	9	
217	Asp5	heavy metal translocating P-type ATPase	Arthrobacter sp. FB24	851	116662145	Actinobacteria (HighGC G+)	B	9	
218	Csp1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Caulobacter sp. K31	873	113933097	Alpha-proteobacteria	B	9	
219	Ccr1	cation transporting ATPase	Caulobacter crescentus CB15	677	16126959	Alpha-proteobacteria	B	9	
220	Asp4	heavy metal translocating P-type ATPase	Acidovorax sp. JS42	880	121594240	Beta-proteobacteria	B	9	
221	Bmu2	heavy metal translocating P-type ATPase	Burkholderia multivorans ATCC 17616	861	118721130	Beta-proteobacteria	B	9	

222	Eli1	cadmium translocating P-type ATPase	Erythrobacter litoralis HTCC2594	831	85373712	Alpha-proteobacteria	B	9	
223	Ppu3	cadmium translocating P-type ATPase	Pseudomonas putida KT2440	665	26986786	Gamma-proteobacteria	B	9	
224	Sal1	Heavy metal translocating P-type ATPase	Sphingopyxis alaskensis RB2256	830	103488060	Alpha-proteobacteria	B	9	
225	Rme4	Heavy metal translocating P-type ATPase	Ralstonia metallidurans CH34	829	94152372	Beta-proteobacteria	B	9	
226	Rpa4	Heavy metal translocating P-type ATPase	Rhodopseudomonas palustris BisB5	853	91976614	Alpha-proteobacteria	B	9	
227	Pla1	heavy metal translocating P-type ATPase	Parvibaculum lavamentivorans DS-1	809	121525714	Alpha-proteobacteria	B	9	
228	Bsp7	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bradyrhizobium sp. BTAi1	840	78692489	Alpha-proteobacteria	B	9	
229	Xau2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Xanthobacter autotrophicus Py2	835	89362164	Alpha-proteobacteria	B	9	
230	Xau3	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Xanthobacter autotrophicus Py2	834	89360771	Alpha-proteobacteria	B	9	
231	Ssp4	Heavy metal translocating P-type ATPase	Sphingomonas sp. SKA58	833	94497154	Alpha-proteobacteria	B	9	
232	Msp5	heavy metal translocating P-type ATPase	Mesorhizobium sp. BNC1	833	110347229	Alpha-proteobacteria	B	9	
233	Dac2	heavy metal translocating P-type ATPase	Delftia acidovorans SPH-1	638	118730278	Beta-proteobacteria	B	9	
234	Nar1	Heavy metal translocating P-type ATPase	Novosphingobium aromaticivorans DSM 12444	826	87200161	Alpha-proteobacteria	B	9	
235	Mfl2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	870	89337470	Actinobacteria (HighGC G+)	B	9	
236	Ssp6	cadmium translocating P-type ATPase	Sulfitobacter sp. NAS-14.1	682	83956091	Alpha-proteobacteria	B	9	
237	Hne1	cadmium-translocating P-type ATPase	Hyphomonas neptunium ATCC 15444	663	114799122	Alpha-proteobacteria	B	9	
238	Ssp3	cation-transporting ATPase; E1-E2 ATPase	Synechocystis sp. PCC 6803	642	16331905	Cyanobacteria	B	10	696.2 ± 82.93

239	Lsp1	cation-transporting ATPase; E1-E2 ATPase	Lyngbya sp. PCC 8106	672	119483255	Cyanobacteria	B	10	
240	Nsp2	Cd/Co/Hg/Pb/Zn-translocating P-type ATPase	Nodularia spumigena CCY9414	641	119509788	Cyanobacteria	B	10	
241	Npu1	COG2217: Cation transport ATPase	Nostoc punctiforme PCC 73102	656	23130381	Cyanobacteria	B	10	
242	Nsp3	cadmium-transporting ATPase	Nostoc sp. PCC 7120	694	17230653	Cyanobacteria	B	10	
243	Sth2	putative cadmium-transporting ATPase	Symbiobacterium thermophilum IAM 14863	656	51894011	Actinobacteria (HighGC G+)	B	10	
244	Cau2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chloroflexus aurantiacus J-10-fl	914	76258037	Chloroflexi	B	10	
245	Cag2	Heavy metal translocating P-type ATPase	Chloroflexus aggregans DSM 9485	640	118048714	Chloroflexi	B	10	
246	Rca2	Heavy metal translocating P-type ATPase	Roseiflexus castenholzii DSM 13941	723	118063539	Chloroflexi	B	10	
247	Rsp5	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Roseiflexus sp. RS-1	724	106892616	Chloroflexi	B	10	
248	Lpn3	cadmium efflux ATPase	Legionella pneumophila subsp. pneumophila str. Philadelphia 1	635	52841244	Gamma-proteobacteria	B	11	635
249	Nfa1	putative cation-transporting ATPase	Nocardia farcinica IFM 10152	615	54027681	Actinobacteria (HighGC G+)	B	12A	653.15 ± 13.35
250	Msp6	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium sp. JLS	669	92908829	Actinobacteria (HighGC G+)	B	12A	
251	Mva1	heavy metal translocating P-type ATPase	Mycobacterium vanbaalenii PYR-1	656	120404609	Actinobacteria (HighGC G+)	B	12A	
252	Mfl4	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	655	89342038	Actinobacteria (HighGC G+)	B	12A	
253	Rer1	putative cadmium resistance protein (CadA)	Rhodococcus erythropolis	671	33867187	Actinobacteria (HighGC G+)	B	12A	
254	Mfl6	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	660	89339407	Actinobacteria (HighGC G+)	B	12A	

255	Jsp1	putative cation-transporting ATPase	Janibacter sp. HTCC2649	656	84496041	Actinobacteria (HighGC G+)	B	12A	
256	Nsp5	heavy metal translocating P-type ATPase	Nocardioides sp. JS614	656	119716436	Actinobacteria (HighGC G+)	B	12A	
257	Asp8	heavy metal translocating P-type ATPase	Arthrobacter sp. FB24	650	116662172	Actinobacteria (HighGC G+)	B	12A	
258	Mar1'	putative cation-transporting ATPase	marine actinobacterium PHSC20C1	649	88856664	Actinobacteria (HighGC G+)	B	12A	
259	Aau2	cadmium-translocating P-type ATPase	Arthrobacter aurescens TC1	652	119952482	Actinobacteria (HighGC G+)	B	12A	
260	Aau1	cadmium-translocating P-type ATPase	Arthrobacter aurescens TC1	651	119952630	Actinobacteria (HighGC G+)	B	12A	
261	Asp7	heavy metal translocating P-type ATPase	Arthrobacter sp. FB24	651	116662126	Actinobacteria (HighGC G+)	B	12A	
262	Gwe1	CadA protein	Gordonia westfalica	730	40445315	Actinobacteria (HighGC G+)	B	12B	715.83 ± 9.15
263	Rer2	putative cadmium resistance protein (CadA)	Rhodococcus erythropolis	715	33867146	Actinobacteria (HighGC G+)	B	12B	
264	Mfl7	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	705	89337462	Actinobacteria (HighGC G+)	B	12B	
265	Msp7	Heavy metal translocating P-type ATPase	Mycobacterium sp. MCS	723	108798440	Actinobacteria (HighGC G+)	B	12B	
266	Mfl3	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	712	89340253	Actinobacteria (HighGC G+)	B	12B	
267	Mfl5	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Mycobacterium flavescens PYR-GCK	710	89339439	Actinobacteria (HighGC G+)	B	12B	
268	Mar2`	putative cadmium-transporting ATPase	marine actinobacterium PHSC20C1	632	88856732	Actinobacteria (HighGC G+)	B	12C	638 ± 24.62
269	Cje1	putative cadmium-transporting ATPase	Corynebacterium jeikeium K411	630	68536524	Actinobacteria (HighGC G+)	B	12C	
270	Bli4	COG2217: Cation transport ATPase	Brevibacterium linens BL2	606	62422742	Actinobacteria (HighGC G+)	B	12C	
271	Cef3	putative cation-transporting ATPase	Corynebacterium efficiens YS-314	650	25026566	Actinobacteria (HighGC G+)	B	12C	
272	Cef4	putative cation-transporting atpase	Corynebacterium efficiens	672	23578013	Actinobacteria (HighGC G+)	B	12C	
273	Nph1	transport ATPase 2 (probable substrates zinc/cadmium)	Natronomonas pharaonis DSM 2160	730	76800784	Euryarchaeota	A	13A	all 13: 813.43 ± 60.18
274	Hsp2	ZntA	Halobacterium sp. NRC-1	757	15789464	Euryarchaeota	A	13A	
275	Hma3	zinc-transporting ATPase	Haloarcula marismortui ATCC 43049	859	55377209	Euryarchaeota	A	13A	

276	Hma1	zinc-transporting ATPase	Haloarcula marismortui ATCC 43049	894	55376485	Euryarchaeota	A	13A	
277	Hwa1	cadmium-transporting ATPase	Haloquadratum walsbyi DSM 16790	861	110668188	Euryarchaeota	A	13A	
278	Hma2	zinc-transporting ATPase	Haloarcula marismortui ATCC 43049	806	55380298	Euryarchaeota	A	13B	
279	Hma4	cation-transporting ATPase	Haloarcula marismortui ATCC 43049	787	55376530	Euryarchaeota	A	13B	
280	Mst1	predicted cation transport ATPase	Methanosphaera stadtmanae DSM 3091	731	84488951	Euryarchaeota	A	14	678 ± 65.34
281	Mma2	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Methanoculleus marisnigri JR1	698	110603920	Euryarchaeota	A	14	
282	Mth2	cadmium efflux ATPase	Methanothermobacter thermautotrophicus str. Delta H	605	15678439	Euryarchaeota	A	14	
283	Gst1	cadmium efflux pump protein	Geobacillus stearothermophilus	727	16753175	Firmicutes (LowGC G+)	B	15A	all 15: 730.67 ± 54.52
284	Gka2	cation-transporting ATPase	Geobacillus kaustophilus HTA426	708	56419285	Firmicutes (LowGC G+)	B	15A	
285	Bha1	cadmium-transporting ATPase	Bacillus halodurans C-125	707	15616598	Firmicutes (LowGC G+)	B	15A	
286	Bcl1	cadmium-transporting ATPase	Bacillus clausii KSM-K16	709	56962051	Firmicutes (LowGC G+)	B	15A	
287	Sau1	Probable cadmium-transporting ATPase (Cadmium efflux ATPase)	Staphylococcus aureus	804	584870	Firmicutes (LowGC G+)	B	15A	
288	Bps1	Probable cadmium-transporting ATPase (Cadmium efflux ATPase)	Bacillus pseudofirmus	723	231677	Firmicutes (LowGC G+)	B	15A	
289	Sep1'	cadmium resistance protein B	Staphylococcus epidermidis ATCC 12228	802	27466993	Firmicutes (LowGC G+)	B	15A	
290	Lmo1	Probable cadmium-transporting ATPase (Cadmium efflux ATPase)	Listeria monocytogenes	711	3121832	Firmicutes (LowGC G+)	B	15A	
291	Sth1	cadmium efflux ATPase	Streptococcus thermophilus	707	46019880	Firmicutes (LowGC G+)	B	15A	
292	Lla1	cadmium resistance protein	Lactococcus lactis subsp. cremoris	705	71024887	Firmicutes (LowGC G+)	B	15A	
293	Oih1	cadmium-transporting ATPase	Oceanobacillus iheyensis HTE831	711	23097731	Firmicutes (LowGC G+)	B	15A	
294	Bsp1	cadmium-transporting ATPase	Bacillus sp. NRRL B-14911	603	89097714	Firmicutes (LowGC G+)	B	15A	
295	Gka1	cadmium-transporting ATPase	Geobacillus kaustophilus HTA426	712	56419120	Firmicutes (LowGC G+)	B	15A	
296	Sag1	cation-transporting ATPase, E1-E2 family	Streptococcus agalactiae 2603V/R	709	22537407	Firmicutes (LowGC G+)	B	15A	
297	Gsu1	cadmium-translocating P-type ATPase	Geobacter sulfurreducens PCA	713	39997245	Delta-proteobacteria	B	15B	
298	Tca1	heavy metal translocating P-type ATPase	Thermosinus carboxydivorans Nor1	691	121535981	Firmicutes (LowGC G+)	B	15B	

299	Swo1	cadmium-transporting ATPase	Syntrophomonas wolfei subsp. wolfei str. Goettingen	735	114567595	Firmicutes (LowGC G+)	B	15C	
300	Dre1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Desulfotomaculum reducens MI-1	783	88944312	Firmicutes (LowGC G+)	B	15C	
301	met1	putative Cd(2+)-translocating P-type ATPase	uncultured methanogenic archaeon	708	116077933	Euryarchaeota	A	15C	
302	Bwe1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Bacillus weihenstephanensis KBAB4	833	89208250	Firmicutes (LowGC G+)	B	15D	
303	Ssa1	cadmium resistance protein	Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	843	73662912	Firmicutes (LowGC G+)	B	15D	
304	Ava1	Cd/Co/Hg/Pb/Zn-translocating P-type ATPase	Anabaena variabilis ATCC 29413	751	75907348	Cyanobacteria	B	16A	all 16: 759.29 ± 56.30
305	Nsp1	cation-transporting ATPase	Nostoc sp. PCC 7120	879	17158758	Cyanobacteria	B	16A	
306	Mfe1	cadmium-translocating P-type ATPase	Mariprofundus ferrooxydans PV-1	769	114776637	unclassified Proteobacteria	B	16B	
307	Rxy1	Heavy metal translocating P-type ATPase	Rubrobacter xylanophilus DSM 9941	711	108803439	Actinobacteria (HighGC G+)	B	16B	
308	Rca1	Heavy metal translocating P-type ATPase	Roseiflexus castenholzii DSM 13941	750	118065328	Chloroflexi	B	16C	
309	Cag1	Heavy metal translocating P-type ATPase	Chloroflexus aggregans DSM 9485	734	118045776	Chloroflexi	B	16C	
310	Cau1	ATPase, E1-E2 type:Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase:Heavy metal translocating P-type ATPase	Chloroflexus aurantiacus J-10-fl	721	76261326	Chloroflexi	B	16C	
311	Tth1	cation-transporting ATPase	Thermus thermophilus HB8	684	55980675	Deinococci	B	17	684

Table 6. Motif analyses of the seventeen clusters representing the Heavy Metal P-type ATPases

	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5	Motif 6	Motif 7	Motif 8	Motif 9
	PGD	PAD	TGES	PEGL	DKTGTKT	KGAPE	DPPR	MVTGD	VAVTGDGVNDSPALKKADIGVAM
1	PG(D)	P(A)D	TGES	PCAL	DKTGTKT	HPLG(A)	D(E)(P)R	MLTGD	(VAM)VGDGINDAPAL(AA)A(D)IGIAM
	729-331	734-736	752-755	857-860	900-906	907-9911	1067-1070	1092-1096	1137-1159
2	PG(E)	(AL)D	TGES	PCAL	DKTGTKT	(H)G(K)P(V)	D(TV)R	MLTGD	(V)GM(V)GDGINDAPALA(R)A(D)IGFAM
	572-574	577-579	595-598	700-703	742-749	750-754	889-902	921-925	972-994
3	(P)GD	PLD	TGES	PCAL	DKTGTLT	(E)G(E)P(E)	D(R)IR	MLTGD	(V)AMVGDGVNDAPAM(AR)A(DV)GIAM
	522-524	528-560	546-549	662-665	705-711	712-716	834-837	857-861	901-923
4	PGE	P(V)D	TGES	PCAL	DKT(G)TLT	KG(EP)E	D(V)IR	MLTGD	V(I)MVGDGINDAPALA(A)ADAGIAM
	220-222	225-227	243-246	348-350	391-397	398-402	522-525	544-548	588-610
5	(P)G(E)	(PL)D	TGES	PCAL	DKTGTLT	(K)GF(E)V	D(EIK)	MVTGD	(VA)F(V)GDGINDA(P)VL(AR)A(DV)GIAM
	496-498	501-503	519-522	637-640	681-687	688-692	813-816	838-840	886-908
6	AGE	PID	TGES	PCGL	DKTGTIT	RGEF(I)V	D(AC)R	MLTGD	(TA)MVGDGINDAPALA(T)ADIGISM
	479-481	484-486	502-505	607-610	650-656	657-570	794-797	816-820	876-898
7	(P)G(E)	PVD	TGES	P(G)AL	DKTGTLT	(E)G(RPE)	D(TI)R	M(L)TGD	VAMVGDGVNDAPALA(A)AD(I)GIAM
	250-252	256-268	274-277	370-373	414-420	421-425	581-584	604-608	648-670
8	(P)G(E)	(PA)D	TGE(S)	PCAL	DKTGTLT	(K)G(KPV)	D(TPR)	MVTGD	(VA)MVGDGINDAPAL(AA)A(T)VG(I)AM
	286-288	291-293	309-312	420-423	464-470	471-475	521-624	644-648	689-711
9	(V)GD	(PA)D	TGES	PCAL	DKTGTLT	(E)G(K)PR	D(TP)R	MISGD	(V)AVGDGVNDAPAMA(NT)VGIAM
	419-421	432-434	450-453	571-574	614-620	621-625	747-750	770-774	814-836
10	P(GE)	(PT)D	TGES	PCAL	DKTGTLT	(T)GK(PQ)	D(RV)R	MLTGD	VAMVGDGINDAPALA(A)ATVGIAM
	476-478	481-483	499-502	605-608	648-654	655-659	831-834	854-858	898-920
11	PGE	ATD	TGES	PCAL	DKTGTLT	QGKPK	DPLR	MLTGD	VMMVGDGVNDAPALATADIGIAM
1 seq	161-163	166-168	184-187	288-291	331-337	338-342	459-462	482-486	526-548
12	PGE	ATD	TGES	PCAL	DKTGTLT	(R)N(QPA)	DELR	MLTGD	(T)AMGDGVNDAPALATAD(I)GIAM
	193-195	198-200	216-219	319-322	362-368	369-373	483-486	505-509	548-570
13	PGE	P(T)D	TGES	PCAF	DKTGTLT	(K)G(ELA)	D(E)LR	MLTGD	VAMVGDGINDAPALA(T)A(D)VGIAM
	388-390	393-395	411-414	531-534	574-580	581-585	744-747	767-771	811-833
14	PGD	PLD	TGES	PCAL	DKTGTLT	(E)GR(L)E	D(VI)R	MLTGD	VVMGDGVNDAPALA(R)ANVGIAM
	252-254	257-259	275-278	380-383	423-429	430-434	550-553	572-576	616-638
15	PG(Q)	AMD	TGES	PCAL	DKTGTLT	(K)G(V)PV	D(E)VR	MLTGD	VAMVGDGVNDAPALA(A)S(T)VGIAM
	383-385	388-390	406-409	511-514	554-560	561-565	689-692	712-716	756-778
16	(P)GE	PVD	TGES	PCAL	DKTGTLT	(T)G(K)P(V)	D(T)IR	MLTGD	VGMVGDGINDAPALAA(DV)GIAM
	434-436	439-441	458-461	569-572	612-618	619-623	754-757	777-781	821-843
17	PGE	PAD	TGEP	PCAL	DKTGTLT	LGKPT	DTPR	LLTGD	VAMVGDGVNDAPALARATVGLAV
1 seq	219-221	224-226	242-245	346-349	389-395	396-400	509-512	530-534	576-598

Table 7. The prokaryotic organisms whose16S rRNAs were used to construct a phylogenetic tree representing each genus found among the 311 protein sequences representing the Heavy Metal P-type ATPases.

	Abb.	Description	Organism	Length	Gi#
1	Gka1	B.kaustophilus 16S ribosomal RNA	Geobacillus kaustophilus	1432	39549
2	Mma1	Mycobacterium marinum 16S rRNA gene	Mycobacterium marinum	1489	44459
3	Vch1	V.cholerae gene for 16S ribosomal RNA	Vibrio cholerae	1452	49417
4	Atu1	Agrobacterium tumefaciens 16S rRNA gene	Agrobacterium tumefaciens	1489	142272
5	Ahy1	Aeromonas hydrophila 16S ribosomal RNA	Aeromonas hydrophila	1492	173692
6	Lmo1	Listeria monocytogenes 16S ribosomal RNA	Listeria monocytogenes	1553	175140
7	Lpn1	Legionella pneumophila 16S ribosomal RNA	Legionella pneumophila	1544	175168
8	Mth1	Methanococcus thermolithotrophicus 16S ribosomal RNA	Methanothermococcus thermolithotrophicus	1452	175445
9	Pae1	P.aeruginosa 16S ribosomal RNA	Pseudomonas aeruginosa	1517	175722
10	Tpa1	Treponema pallidum 16S ribosomal RNA	Treponema pallidum	1573	176249
11	Tbr1	Thermoanaerobacter finii (DSM 3389) 16S ribosomal RNA (16S rRNA) gene	Thermoanaerobacter brockii subsp. finnii	1523	349568
12	Lma1	Leeuwenhoekiella marinoflava gene for 16S ribosomal RNA, partial sequence, strain: NCIMB 397	Leeuwenhoekiella marinoflava	1257	425727

13	Cau1	Chloroflexus aurantiacus gene for 16S ribosomal RNA, partial sequence	Chloroflexus aurantiacus J-10-fl	1401	550527
14	Bsp1	Bradyrhizobium spec. (LMG 9980) gene for 16S rRNA	Bradyrhizobium sp.	1441	563846
15	Sau1	Staphylococcus aureus 16S ribosomal RNA (16S rRNA) gene	Staphylococcus aureus	1500	576603
16	Ype1	Yersinia pestis 16S ribosomal RNA (16S rRNA) gene	Yersinia pestis	1467	576926
17	Rsp1	Rhodococcus sp. (P6) 16S rRNA gene	Rhodococcus sp.	1446	577792
18	Pfu1	Pyrococcus furiosus 16S small subunit ribosomal RNA	Pyrococcus furiosus	1495	643670
19	Pca1	Pelobacter carbinolicus 16S ribosomal RNA gene, partial sequence	Pelobacter carbinolicus	1414	727426
20	Eac1	Exiguobacterium acetylicum 16S rRNA gene, partial sequence	Exiguobacterium acetylicum	1482	893364
21	Esp1	Enterobacter sp. 16S rRNA gene, partial sequence	Enterobacter sp.	1500	1073242
22	Eco1	E.coli (ATCC 11775T) gene for 16S rRNA	Escherichia coli	1450	1240022
23	Hor1	Halothermothrix orenii ribosomal RNA sequence	Halothermothrix orenii	1534	1256954
24	Dsp1	Desulfotomaculum sp. strain T93B 16S-like ribosomal RNA	Desulfotomaculum sp.	1392	1353384
25	Lin1	Lawsonia intracellularis 16S ribosomal RNA gene, partial sequence	Lawsonia intracellularis	1468	1389678

26	Mde1	M.defluvii 16S rRNA gene	Methanothermobacter defluvii	1445	1430856
27	Ghi1	G.hirsuta 16S rRNA gene	Gordonia hirsuta	1461	1666638
28	Orh1	Ornithobacterium rhinotracheale 16S ribosomal RNA gene, complete sequence	Ornithobacterium rhinotracheale	1421	1842059
29	Sty1	Salmonella typhi 16S ribosomal RNA gene, complete sequence	Salmonella typhi	1541	1857865
30	Cps1	Chlamydophila psittaci gene for 16S rRNA, strain:6BC	Chlamydophila psittaci 6BC	1507	1902841
31	Dsp2	Desulfitobacterium sp. 16S rRNA gene, clone 2	Desulfitobacterium sp.	1616	1915888
32	Mgl1	Moorella glycerini 16S small subunit ribosomal RNA gene, complete sequence	Moorella glycerini	1513	1916225
33	Cpe1	C.perfringens 16S rRNA gene	Clostridium perfringens	1504	2058294
34	Bbr1	Bordetella bronchiseptica 16S rRNA gene	Bordetella bronchiseptica	1532	2174260
35	Kox1	Klebsiella oxytoca gene for 16S ribosomal RNA, partial sequence	Klebsiella oxytoca	1441	2209046
36	Cla1	C.lactoaceticus 16S rRNA gene	Caldicellulosiruptor lactoaceticus	1505	2222664
37	Ssp1	Syntrophobacter sp. 16S ribosomal RNA	Syntrophobacter sp.	1484	2463455
38	Ljo1	Lactobacillus johnsonii 16S rRNA gene	Lactobacillus johnsonii	1487	2597958
39	Cfu1	Caulobacter fusiformis gene for 16S rRNA, partial sequence	Caulobacter fusiformis	1385	2754583

40	Hge1	Herpetosiphon geysericola 16S ribosomal RNA gene, partial sequence	Herpetosiphon geysericola	1417	2760922
41	Nsp1	Nodularia sp. 16S rRNA gene, isolate BCNOD9427	Nodularia sp. BCNOD9427	1489	2920730
42	Ppr1	Photobacterium profundum gene for 16S ribosomal RNA, strain:SS9	Photobacterium profundum SS9	1518	2924634
43	Nsp2	Nostoc ATCC53789 16S ribosomal RNA gene, partial sequence	Nostoc sp. ATCC 53789	1481	3132717
44	Lae1	Lyngbya sp. 16S rRNA gene, strain PCC 7419	Lyngbya aestuarii PCC 7419	1451	3242226
45	Neu1	Nitrosomonas europaea 16S ribosomal RNA gene, complete sequence	Nitrosomonas europaea	1520	3414677
46	Hsp1	Haloarcula sp. gene for 16S rRNA	Haloarcula sp.	1470	4115513
47	Fnu1	Fusobacterium nucleatum subsp. nucleatum 16S ribosomal RNA gene	Fusobacterium nucleatum subsp. nucleatum	1479	4490387
48	Hsp2	Halobacterium sp. AUS-1 DNA for 16S ribosomal RNA	Halobacterium sp. AUS-1	1465	4580005
49	Sma1	Serratia marcescens 16S ribosomal RNA gene, partial sequence	Serratia marcescens	1494	4883843
50	Cfu2	Cytophaga fucicola 16S rRNA gene, type strain NN015860, partial	Cellulophaga fucicola	1472	5701823

51	Efa1	Enterococcus faecalis 16S ribosomal RNA gene, partial sequence	Enterococcus faecalis	1510	5732229
52	Osa1	Oscillatoria sancta PCC 7515 16S ribosomal RNA gene, partial sequence	Oscillatoria sancta PCC 7515	1409	5771438
53	Rxy1	Rubrobacter xylanophilus partial 16S rRNA gene	Rubrobacter xylanophilus	1509	6006622
54	Abe1	Anabaena bergii 16S ribosomal RNA gene, partial sequence	Anabaena bergii	1399	8571953
55	Mco1	Marinomonas communis 16S ribosomal RNA ribosomal RNA gene, partial sequence	Marinomonas communis	1448	9622547
56	Sth1	Symbiobacterium thermophilum DNA for 16S rRNA	Symbiobacterium thermophilum	1479	11079170
57	Lme1	Leuconostoc mesenteroides DNA for 16S ribosomal RNA, strain NCFB 529	Leuconostoc mesenteroides	1449	11602800
58	Sru1	Salinibacter ruber strain M1 16S ribosomal RNA gene, partial sequence	Salinibacter ruber	1491	12007478
59	Csp1	Cellulophaga sp. ACEM20 16S ribosomal RNA gene, partial sequence	Cellulophaga sp. ACEM20	1425	14517332
60	Ssp2	Synechocystis PCC6805 gene for 16S rRNA, partial sequence	Synechocystis sp. PCC 6805	1437	16215699
61	Acr1	Alkaliphilus crotonoxidans 16S ribosomal RNA gene, partial sequence	Alkaliphilus crotonatoxidans	1553	19072573

62	Spy1	Streptococcus pyogenes MGAS8232, complete genome	Streptococcus pyogenes MGAS8232	1895017	19745201
63	Csp2	Chromobacterium sp. 70 16S ribosomal RNA ribosomal RNA gene, complete sequence	Chromobacterium sp. 70	1455	21898816
64	Xpo1	Xanthobacter polyaromaticivorans gene for 16S ribosomal RNA, partial sequence	Xanthobacter polyaromaticivorans	1460	29335757
65	Mla1	Methanocorpusculum labreanum strain DSM 4855 16S ribosomal RNA gene, partial sequence	Methanocorpusculum labreanum	1409	29373912
66	CFr1	Candidatus Fritschea eriococci strain Elm 16S ribosomal RNA and 23S ribosomal RNA genes, complete sequence	Candidatus Fritschea eriococci	4531	31747860
67	Avi1	Azotobacter vinelandii DSM576 16S ribosomal RNA gene, partial sequence	Azotobacter vinelandii	1398	33242483
68	Mth2	Methanoculleus thermophilus gene for 16S rRNA	Methanoculleus thermophilus	1431	33342024
69	Rbi1	Robiginitalea biformata strain HTCC2501 16S ribosomal RNA gene, partial sequence	Robiginitalea biformata HTCC2501	1434	37912049
70	Jsp1	Janibacter sp. NSA5-4 gene for 16S rRNA	Janibacter sp. NSA5-4	1516	46575825

71	Wsu1	Wolinella succinogenes strain ATCC 29543 16S ribosomal RNA, partial sequence	Wolinella succinogenes	1481	47558942
72	Oih1	Oceanobacillus iheyensis strain MSU3110 16S ribosomal RNA gene, partial sequence	Oceanobacillus iheyensis	1455	50980367
73	Cgl1	Corynebacterium glutamicum strain CICC10226 16S ribosomal RNA gene, partial sequence	Corynebacterium glutamicum	1472	55735425
74	Plu1	Photorhabdus luminescens subsp. luminescens strain ATCC 29999 16S ribosomal RNA gene, partial sequence	Photorhabdus luminescens subsp. luminescens	1475	58042756
75	Sgl1	Sodalis glossinidius strain GP-SG1 16S ribosomal RNA gene, partial sequence	Sodalis glossinidius	1506	58761272
76	Bsu1	Bacillus subtilis partial 16S rRNA gene, isolate SMF7	Bacillus subtilis	1522	60098072
77	Tli1	Tenacibaculum sp. CL-TF13 16S ribosomal RNA gene, partial sequence	Tenacibaculum litoreum	1446	62131505
78	Gpo1	Gramella portivictoriae strain UST040801-001 16S ribosomal RNA gene, partial sequence	Gramella portivictoriae	1468	62903125

79	Bpy1	Burkholderia pyrrocinia isolate RG6-5 16S ribosomal RNA gene, partial sequence	Burkholderia pyrrocinia	1387	63020470
80	Pfe1	Paracoccus ferrooxidans strain BDN-1 16S ribosomal RNA gene, partial sequence	Paracoccus ferrooxidans	1418	63020950
81	Ctr1	Chlamydia trachomatis strain A/Har-1 clone 1 16S ribosomal RNA gene, partial sequence	Chlamydia trachomatis	1550	63354709
82	Dfi1	Deinococcus sp. CC FR2-10 16S ribosomal RNA gene, partial sequence	Deinococcus ficus	1453	66394748
83	Paq1	Polaromonas aquatica 16S rRNA gene, strain CCUG 39797	Polaromonas aquatica	1407	68051131
84	Msp1	Magnetospirillum sp. PM2411 gene for 16S ribosomal RNA	Magnetospirillum sp. PM2411	1455	68533196
85	Gsp1	Geobacter sp. CLFeRB 16S ribosomal RNA gene, complete sequence	Geobacter sp. CLFeRB	1485	70906126
86	Hta1	Hyphomonas taiwanensis strain HYP1 16S ribosomal RNA gene, partial sequence	Hyphomonas taiwanensis	1418	75265811
87	Psp1	Polynucleobacter sp. MWH-BledIIIW10 partial 16S rRNA gene, strain MWH-BledIIIW10	Polynucleobacter sp. MWH-BledIIIW10	1424	77917378

88	Rsp2	Rhodoferax sp. PIC-C33 16S ribosomal RNA gene, partial sequence	Rhodoferax sp. PIC-C33	1473	77994459
89	Nvu1	Nitrobacter vulgaris partial 16S rRNA gene, type strain DSM 10236T	Nitrobacter vulgaris	1441	78271519
90	Sel1	Synechococcus elongatus PCC 7942, complete genome	Synechococcus elongatus PCC 7942	2695903	81167692
91	Slu1	Sediminicola luteus gene for 16S ribosomal RNA, partial sequence, strain:CNI-1-5	Sediminicola luteus	1440	82568484
92	Fph1	Francisella sp. 2005/50/F292-6C 16S ribosomal RNA gene, partial sequence	Francisella philomiragia subsp. noatunensis	1416	83031476
93	Mfl1	Methylobacillus flagellatus strain KT 16S ribosomal RNA gene, partial sequence	Methylobacillus flagellatus	1400	83272653
94	Hwa1	Haloquadratum walsbyi clone 2B08 sequence	Haloquadratum walsbyi	38670	85680316
95	Epe1	Erwinia persicina strain GS04 16S ribosomal RNA gene, partial sequence	Erwinia persicina	1470	86161535
96	Dts1	Delftia tsuruhatensis strain P18 16S ribosomal RNA gene, partial sequence	Delftia tsuruhatensis	1406	86450325

97	Aal1	Aurantimonas altamirensis strain S21B 16S ribosomal RNA gene, partial sequence	Aurantimonas altamirensis	1366	87044947
98	Rsp3	Ralstonia sp. PHD-12 16S ribosomal RNA gene, partial sequence	Ralstonia sp. PHD-12	1493	87312602
99	Rsp4	Rhodopseudomonas sp. TUT3627 gene for 16S rRNA, partial sequence	Rhodopseudomonas sp. TUT3627	1482	88606680
100	Aav1	Acidovorax avenae strain FC-501 16S ribosomal RNA gene, partial sequence	Acidovorax avenae	1462	89214127
101	Bdo1	Bacteroides dorei gene for 16S rRNA, partial sequence, strain:JCM 13472	Bacteroides dorei	1490	89242107
102	Rsp5	Roseobacter sp. LOB-8 16S ribosomal RNA gene, partial sequence	Roseobacter sp. LOB-8	1404	89258470
103	Ssp3	Sulfitobacter sp. SPB-4 16S ribosomal RNA gene, partial sequence	Sulfitobacter sp. SPB-4	1398	89258485
104	Dsp3	Dechloromonas sp. EMB 269 16S ribosomal RNA gene, partial sequence	Dechloromonas sp. EMB 269	1458	89348121
105	Ssp4	Sphingopyxis sp. EMB 355 16S ribosomal RNA gene, partial sequence	Sphingopyxis sp. EMB 355	1407	89348125

106	Raz1	Rhodobacter azotoformans strain S3 16S ribosomal RNA gene, partial sequence	Rhodobacter azotoformans	1459	89357194
107	Osp1	Oceanicola sp. 20 16S ribosomal RNA gene, partial sequence	Oceanicola sp. 20	1427	89520386
108	Rsp6	Rhizobium sp. As-2 16S ribosomal RNA gene, partial sequence	Rhizobium sp. As-2	1391	89953754
109	Rsp7	Rhizobium sp. Lv6.1Se 16S ribosomal RNA gene, partial sequence	Rhizobium sp. Lv6.1Se	1446	89954503
110	Psp2	Psychrobacter sp. AM11 partial 16S rRNA gene, isolate AM11	Psychrobacter sp. AM11	1501	90071222
111	Mti1	Mesorhizobium tianshanense strain RCAN08 16S ribosomal RNA gene, partial sequence	Mesorhizobium tianshanense	1479	90186421
112	Hca1	Helicobacter canadensis strain L231 16S ribosomal RNA gene, partial sequence	Helicobacter canadensis	1407	90194336
113	Bce1	Brucella cetaceae partial 16S rRNA gene, strain NCTC 12891	Brucella cetaceae	1430	90818672
114	Pmi1	Proteus mirabilis strain O 16S ribosomal RNA gene, partial sequence	Proteus mirabilis	1451	90856205

115	Swo1	Syntrophomonas wolfei subsp. methybutyica strain 5J-1 16S ribosomal RNA gene, partial sequence	Syntrophomonas wolfei subsp. methylbutyratica	1602	90903520
116	Csp3	Comamonas sp. Dant 3-8 16S ribosomal RNA gene, partial sequence	Comamonas sp. Dant 3-8	1453	91701740
117	Asp1	Arthrobacter sp. m3 16S ribosomal RNA gene, partial sequence	Arthrobacter sp. m3	1484	91771955
118	Ssp5	Sphingomonas sp. B9LA 16S ribosomal RNA gene, partial sequence	Sphingomonas sp. B9LA	1369	91805180
119	Esp2	Erythrobacter sp. CNU001 16S ribosomal RNA gene, partial sequence	Erythrobacter sp. CNU001	1480	91982998
120	Nre1	Novosphingobium subarcticum gene for 16S rRNA, partial sequence, strain:T7b	Novosphingobium resinovorum	1400	92019078
121	Bsp2	Brevibacterium sp. CNJ737 PL04 16S ribosomal RNA gene, partial sequence	Brevibacterium sp. CNJ737 PL04	1493	92091011
122	Nsp3	Nocardia sp. CNS044 PL04 16S ribosomal RNA gene, partial sequence	Nocardia sp. CNS044 PL04	1484	92091036
123	Nsp4	Nocardioides sp. CNJ892 PL04 16S ribosomal RNA gene, partial sequence	Nocardioides sp. CNJ892 PL04	1482	92091040

124	Ssp6	Shewanella sp. BSi20587 16S ribosomal RNA gene, partial sequence	Shewanella sp. BSi20587	1504	93009053
125	uBl1	Uncultured Blastopirellula sp. clone 6P2-75 16S ribosomal RNA gene, partial sequence	uncultured Blastopirellula sp.	1428	126143007
126	Mfe1	Mariprofundus ferrooxydans strain JV-1 16S ribosomal RNA gene, complete sequence	Mariprofundus ferrooxydans	1528	145226686
127	Rsp8	Roseiflexus sp. RS-1, complete genome	Roseiflexus sp. RS-1	5801598	148654188
128	Dsh1	Dinoroseobacter shibae DFL 12, complete genome	Dinoroseobacter shibae DFL 12	3789584	157910316
129	Tca1	Thermosinus carboxydivorans Nor1	Thermosinus carboxydivorans Nor1	349669	121307835
130	Pla1	Parvibaculum lavamentivorans DS-1, complete genome	Parvibaculum lavamentivorans DS-1	3914745	154250456
131	Tth1	Thermus thermophilus HB27, complete genome	Thermus thermophilus HB27	1894877	46198308
132	Nph1	Natronomonas pharaonis DSM 2160 complete genome	Natronomonas pharaonis DSM 2160	2595221	76556520
133	Chy1	Carboxydothermus hydrogenoformans Z-2901, complete genome	Carboxydothermus hydrogenoformans Z-2901	2401520	78042616
134	Rru1	Rhodospirillum rubrum ATCC 11170, complete genome	Rhodospirillum rubrum ATCC 11170	4352825	83591340

135	Mst1	Methanosphaera stadtmanae DSM 3091, complete genome	Methanosphaera stadtmanae DSM 3091	1767403	84488831
136	Lla1	L.lactis ribosomal RNA operon encoding 16S, 5S, and 23S ribosomal RNA, transfer RNA-Ala and transfer RNA-Asn	Lactococcus lactis	5953	44070

Table 8: An overview of the sequence characteristics of the clusters representing Copper P-type ATPase Family.[3]

Cluster#	# of Sequences found within Cluster	Ave. AA length for cluster	Standard Deviations	Phylogenetic groups	Extra Domains
1	4	771	52	TM, EP, CR	None
2		769	53	AC	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ). COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter),
3	26	806	16	AD, DP, CB, GP, BP, EP, unclassified proteobacteria	None
4	20	706	41	FI, AQ, BP, EA, DP, D, CH	None
5	5	829	18	D, AP, AC,	None
6	47	788	52	BP, GP, AP, EP, P	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), pfam04945 (YHS domain), cd01057 (AAMH_A, Aromatic and Alkene Monooxygenease Hydroxylases, subunit A), COG3350 (uncharacterized conserved protein), smart00746 (TRASH, metallochaperone-like domain),
7	1	807	807	unclassified Proteobacteria	None
8	27	816	52	GP, BP, AP, CN, D, one unknown organism	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter),
9	24	849	84	GP	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter),

[3] **TM** = Thermotogae (B), **EP** = ε-proteobacteria (B), **AP**= α-proteobacteria (B), **DP**= Δ-proteobacteria (B), **GP** = γ-proteobacteria (B), **BP**= β-proteobacteria (B), **AC** = Actinobacteria (B), **FI** = Firmicutes (B), **F** = Fusobacteria (B), **D** = Deinococci (B), **AD**= Acidobacteria (B), **CB**= Chlorobi (B), **BC** = Bacteroidetes (B), **SP** = Spirochaetes (B), **AQ** = Aquificae (B), **EA** = Euryarchaeota (A), **CR** = Crenarchaeota (A), **CH** = Chloroflexi (B), **P** = Planctomycetes (B), **CN** = Cyanobacteria, **V** = Viridiplantae (E), **FN** = Fungi (E), **MY**= Mycetozoa (E), **MZ** = Metazoa (E), **CL**= Chlamydiae, (E) = Eukaryota, (B) = Bacteria; (A) = Archaea

10	39	813	84	CH, AC, DP, AD, BC, FI, EA, CB	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter), pfam04945 (YHS domain), cd01057 (AAMH_A, Aromatic and Alkene Monooxygenease Hydroxylases, subunit A), COG3350 (uncharacterized conserved protein), smart00746 (TRASH, metallochaperone-like domain), PRK00807 (50S ribosomal protein L24e),
11	9	831	71	AC, EA, CB, FI, DP	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter), cd5062_PTKc_IGF-1R, (Protein-Tyrosine-like Kinase Family, Insulin-like Growth Factor-1 Receptor; catalytic domain),
12	7	658	41	FI	COG4633, uncharacterized protein conserved in bacteria
13	45	1206	214	V, MY, MZ, FN	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter),
14	9	826	76	V, CN	None
15	19	786	63	FI, SP, EP, FU, DB, DP	None
16	3	806	26	BP, GP	None
17	17	757	34	CL, AP, DP, GP, BP, one unknown organism	None
18	3	725	4	FI	None
19	22	796	49	SP, FI, AC	pfam00115 (COX2, Cytochrome C oxidase subunit II, periplasmic domain), COG2131 (SufI, Putative multicopper oxidases)
20	30	812	92	GP, BP, AP	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671 (copper transporter),

Table 9: An overview of the sequence characteristics of the clusters representing the Heavy Metal P-type ATPase Family.[4]

Cluster#	# of Sequences found within Cluster	Ave AA length for cluster	Standard Deviations	Phylogenetic Groups	Extra Domains
1A	18	760	59	BP, AP, GP, D,	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671
1B	16			AP	
1C	21			GP	
1D	3			GP	
2A	1	all 2: 776	all 2: 63	BP	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ),
2B	1			BP	
2C	27			BP, DP, GP	
2D	3			GP, BP	
2E	3			DP, BP	
2F	9			CB, BP	
2G	2			GP	
3	9	all 3: 775	all 3: ± 98	P, GP, DP, AP	None
4	2	664	32	FI, CL	None
5A	1	all 5: 695	all 5: 60	DP	
5B	1			FI	
5C	1			EP	
5D	2			FI	
5E	1			F	
5F	19			FI	
5G	3			EA, FI	
5H	3			FI, EA	
5I	8			AC, FI, CN	
5J	13			BC, GP	
5K	2			SP, FI	
5L	3			EP	None
6	6	989	175	V	None
7A	3	all 7: 640	all 7: 31	FI	
7B	2			EA	
7C	9			AC, FI, AP, CH	None
8A	13	639	14	FI	
8B	2	821	2	V	
8C	4	664	13	CH	None

[4] **TM** = Thermotogae (B), **EP** = ε-proteobacteria (B), **AP**= α-proteobacteria (B), **DP**= Δ-proteobacteria (B), **GP** = γ-proteobacteria (B), **BP**= β-proteobacteria (B), **AC** = Actinobacteria (B), **FI** = Firmicutes (B), **F** = Fusobacteria (B), **D** = Deinococci (B), **AD**= Acidobacteria (B), **CB**= Chlorobi (B), **BC** = Bacteroidetes (B), **SP** = Spirochaetes (B), **AQ** = Aquificae (B), **EA** = Euryarchaeota (A), **CR** = Crenarchaeota (A), **CH** = Chloroflexi (B), **P** = Planctomycetes (B), **CN** = Cyanobacteria, **V** = Viridiplantae (E), **FN** = Fungi (E), **MY**= Mycetozoa (E), **MZ** = Metazoa (E), **CL**= Chlamydiae, (E) = Eukaryota, (B) = Bacteria; (A) = Archaea

9	25	793	82	AC, BC, AP, GP, BP,	None
10	10	696	83	CN, AC, CH	pfam019878 (TrmB), COG3355 (Predicted transcriptional regulator), COG1378 (predicted transcriptional regulators)
11	1	635	635	GP	None
12A	13	653	13	AC	
12B	6	716	9	AC	
12C	5	638	25	AC	None
13A	5	all 13: 813	all 13: 60	EA	
13B	2			EA	cd00371 (HMA), pfam00403 (HMA), COG2608 (CopZ), COG2217 (ZntA), PRK11033 (zntA), PRK10671
14	3	678	65	EA	None
15A	14	all 15: 731	all 15: 55	FI	
15B	2			DP, FI	
15C	3			FI, EA	
15D	2			FI	None
16A	2	all 16: 759.29 ± 56.30	all 16: 759.29 ± 56.30	CN	
16B	2			AC, unclassified Proteobacteria	
16C	3			CH	None
17	1	684	684	D	None

References

Anthonisen A.N., Clausen J.D., Andersen J.P. (2006). Mutational Analysis of the Conserved TGES Loop of Sacroplasmic Reticulum Ca2+ -ATPases. *Journal of Biological Chemistry*, Vol. 281, No **42** 31572-31582

Axelsen K., Palmgren M.G., (1998) Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol* **46**: 84-101

Barabote R.D., Tamang D.G., Abeywardena S.N., Fallah N.S., Fu J.Y.C., Lio J.K., Mirhosseini P., Pezeshk R., Podell S., Salampessy M.L., Thever M.D., Saier, M.H. (2006). Extra domains in secondary transport carriers and channel proteins. *Biochimica et Biophysica Acta*. **1758**: 1557 – 1579.

Benito B., Garciadeblás B, Roderíguez-Navarro A, (2000) Molecular cloning of the calcuium and sodium ATPases in *Neurospora crassa*. *Molecular Microbiology* **35**(5), 1079-1088.

Dosztányi Z., Magyar C., Tusnády G.E., Cserzõ M., Fiser A., Simon I., (2003) Servers for sequence–structure relationship analysis and prediction. *Nucleic Acids Research*, Vol. 31, No. **13** 3359–3363

Geourjon C., Deleage G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci.* Dec; **11**(6):681-4.

Kyte J., Doolittle R.F. (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*. **157**: 105-132.

Møller J.V., Juul B., le Maire M. (1995). Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochimica et Biophysica Acta*. **1286**: 1-51.

Mukherjee T., Mandal D., Bhaduri A., (2000) Leishmania plasma membrane Mg2+ -ATPase is a H+/K+ -antiporter involved in glucose symport. Studies with sealed ghosts and vesicles of opposite polarity. *J Biol Chem.* Feb 23; **276** (8):5563-9.

Thever M.D., M.S. thesis, P-type ATPase Homologues in Eukaryotes, University of California, San Diego, La Jolla, CA, 2007.

Yamaguchi A., Tamang D.G., Saier M.H. (2007). Mercury Transport in Bacteria. *Water, Air, & Soil Pollution*. **182**: 219-234.