UNIVERSITY OF CALIFORNIA, SAN DIEGO

Bioinformatic Characterization of the 4-Toluene Sulfonate Uptake Permease (TSUP) Family of Transmembrane Proteins: Identification of the Microbial Rhodopsin Superfamily and Evidence for an Ancestral Transmembrane Hairpin Structure

A thesis submitted in partial satisfaction of the requirements for the degree

Master of Science

in

Biology

by

Maksim Aleksandrovich Shlykov

Committee in charge:

Professor Milton H. Saier, Jr., Chair
Professor James W. Golden
Professor Emeritus Immo E. Scheffler

2011

The Thesis of Maksim Aleksandrovich Shlykov is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Chair

University of California, San Diego

2011

DEDICATION

I dedicate this thesis to my family, friends, Dr. Saier and all of the previous and current members of Saier Lab who I have had the pleasure of meeting and working with and those whose work made this project possible. This thesis is also dedicated to all of the professors across multiple disciplines, especially biology, who prepared me for this undertaking. A huge thank you is owed to Dorjee Tamang, Abe Silverio, Wei Hao Zheng, Jonathan Chen, Ankur Malhotra, Erik Clarke, Luis Felipe Patino-Cuadrado, Vamsee Reddy and Ming-Ren Yen for their support and guidance.

Above all, I would like to thank Dr. Milton H. Saier, Jr., for his guidance, sense of humor, friendship, as well as his passion and enthusiasm for scientific research, which he has instilled into many students, including myself. In the two years that I have been a part of his lab, I have gotten to know Dr. Saier extremely well. From weekends where we were the only ones in lab to dealing with an unjust lab closure together, I found in Dr. Saier a friend and a mentor, as well someone who works hard and pushes others to do the same. Few can match his devotion to the lab and his students. Perhaps just as importantly, Dr. Saier's work outside the lab makes him a role model to many students and individuals across the world. I sincerely thank Dr. Saier and all of the members of Saier Lab.

EPIGRAPH

Never use epigraphs, they kill the mystery in the work!

Adli

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Milton H. Saier, Jr., for his support and for serving as the chair of my committee. Similarly, I wish to acknowledge Dr. James W. Golden and Dr. Immo E. Scheffler for taking the time and making the effort to serve on my committee. Having taken two engaging courses with Dr. Scheffler previously that I thoroughly enjoyed, it is an honor to have him evaluate my performance once again, in the context of a thesis defense.

Aside from Dr. Saier, Abe Silverio and Wei Hao Zheng were my primary sources of mentorship and were invaluable in helping me get over the learning curve for the research performed in our lab. I am grateful to both Abe and Wei Hao, who deserve an honorary mention.

I would like to acknowledge Ankur Malhotra, Erik Clarke, Jonathan S. Chen, Dorjee Tamang and Luis Felipe Patino-Cuadrado for all of their help on the technical side of this project. I would like to mention Vamsee Reddy and Dr. Ming-Ren Yen, whose programs allowed me to complete this project. I thank Andrew Lukosus for all of his help on the administrative side and Mark Whelan for his work as a TA coordinator, which allowed me to experience teaching both lab and lecture courses.

ABSTRACT OF THE THESIS

Bioinformatic Characterization of the 4-Toluene Sulfonate Uptake Permease (TSUP) Family of Transmembrane Proteins: Identification of the Microbial Rhodopsin Superfamily and Evidence for an Ancestral Transmembrane Hairpin Structure

by

Maksim Aleksandrovich Shlykov

Master of Science in Biology

University of California, San Diego, 2011

Professor Milton H. Saier, Jr., Chair

The sequence diverse and ubiquitous 4-toluene sulfonate uptake permease (TSUP) family contains few characterized members and is believed to catalyze the transport of sulfur-based compounds. Our analyses revealed that prokaryotic members of the TSUP family outnumber the eukaryotic members substantially and that extensive lateral gene transfer occurred during the evolution of the TSUP family. Despite unequal representation, both taxonomic domains share well-conserved motifs. We show that the prototypical eight TMS topology arose from an intragenic duplication of a four TMS unit. Sequence similarity and homology between TSUP

and known secondary carrier families (1) supports a secondary active transport mechanism for the TSUP family, (2) necessitates the creation of the novel Microbial Rhodopsin (MR) superfamily and (3) suggests a common primordial 2 α-helical hairpin structure for multiple families and superfamilies, similarly to what has been suggested for outer membrane β-barrels. The MR superfamily consists of six currently recognized families. Our suggestion of a Super-superfamily may, in the future, group many superfamilies together, generating a new TC hyperlink. Finally, genome context analyses confirm the proposal of a sulfur-based compound transport role for many TSUP homologues, but functional outliers appear to be prevalent as well.

Introduction

Using functional, phylogenetic, and membrane topology information derived from over 10,000 publications on functional data and novel transport systems has allowed our lab to classify over 5,000 transport proteins into over 600 families. Our work is summarized in the IUBMB approved Transporter Classification Database (http://www.tcdb.org), a curated database employing the TC system, which is analogous to the function-only based Enzyme Commission (EC) system (http://www.chem.qmul.ac.uk/iubmb/enzyme/; Saier et al. 2005; 2009). Our study focuses on the putative 4-toluene sulfonate uptake permease (TSUP) family (TCID 9.A.29) of transmembrane proteins such that the gap of knowledge for this protein family is bridged and TCDB can be updated and expanded.

Transport systems play a crucial role in all processes associated with life. Specific examples of the important roles of transport systems include nutrient transport, metabolite excretion, drug and toxin efflux, establishment of electrochemical gradients, macromolecular export, and transport of signaling molecules (Busch et al. 2002). However, their effectiveness can be utilized in ways which are detrimental to humans and other organisms.

Multi-drug resistant (MDR) pathogenic microbial strains, arising partially because of excessive use of antibiotics, have had a heavy clinical impact. Gram-positive bacteria and recently identified Gram-negative strains express antibiotic resistance due to increased synthesis of active drug efflux pumps and less commonly due to decreases in uptake or porin loss. Additional contributions to resistance are

made by non-transport effects such as inactivation of the antibiotic and drug target modification (Hancock, 2005; Thomson et al., 2005). Transporters in multicellular organisms affect the absorption, distribution, accumulation and excretion of drugs (Thomas, 2004). In the case of cancer, decreased levels of chemotherapeutic drugs leads to partial drug insensitivity and subsequent resistance. The same relationship is observed between numerous other diseases and their respective drugs, including those used to treat HIV/AIDS. Further proof for this is provided by the observation that increased expression of specific drug efflux pumps correlates with cancer progression and aggressiveness (Fletcher, 2010). This makes the already difficult task of treating cancer, HIV/AIDS, and other diseases especially daunting. Characterizing transporters can pave the way for computational modes of drug discovery, which would allow us to more effectively target various MDR pathogens and diseases (Saier et al., 2006). The importance of transport proteins, constituting roughly 10% of the proteome, on average, and the processes that they control, cannot be understated.

The 4-toluene sulfonate uptake permease (TSUP) family (TC # 9.A.29), also designated as the Domain of Unknown Function 81 (DUF81), COG0730, or TauE/SafE family, is comprised of over 1000 members spanning the prokaryotic, eukaryotic, and archaeal domains. Within the prokaryotic domain, these proteins have been identified in both Gram-positive and Gram-negative bacteria. The occurrence of multiple organismal sources within phylogenetic clusters implies extensive horizontal transfer of genes encoding the homologues (Yen et al., 2009). The majority of prokaryotic protein members range in size from 240 amino acids to 280 amino acids

with few exceptions. The archaeal members are similar in size to the prokaryotic members, but the eukaryotic members are typically 40-50% larger and range from 400-500 amino acyl residues in size (Chung et al., 2001). Some eukaryotic members possess N- and C-terminal extensions, which may play regulatory roles (Saier, 2000a; Barabote et al., 2006). When only the prokaryotic members were analyzed, an 8 transmembrane segment (TMS) topology was discovered with little deviation (Yen et al., 2009). Eukaryotes and archaea have between 6 to 12 TMSs. It was shown that some of the prokaryotic members have undergone intragenic duplication of a 4 TMS unit yielding 8 TMS proteins (Saier et al., 2006; 2009).

Functions for most of the TSUP family members have not been assigned and cannot be assumed due to the great sequence divergence among homologues (Saier et al., 2006; 2009). In fact, the few analyses that have been performed with TSUP homologues suggest differing functionality. A single TSUP member (TC # 9.A.29.1.1) of 239 amino acids in length was proposed to be a carrier for 4-toluene sulfonate uptake and to employ an inducible secondary active proton symport mechanism (Locher et al., 1993). Another group named the protein identified by Locher et al. (1993) as TsaS because it is part of the *tsa* operon in *Comamonas testosteroni*. They proposed that TsaS localizes to the inner membrane of the bacterial envelope, functioning as part of a two-component system along with an anion-specific pore named TsaT, possibly a putative outer membrane β-barrel (Mampel et al., 2004). TsaT is only expressed in the presence of 4-toluene sulfonate, whereas TsaS is expressed following growth in the presence of a number of compounds (Mampel et al.,

2004). TsaS was predicted to contain 6 TMSs, and the tight expression control implies that TsaT confers specificity whereas TsaS completes the transport process. The broad substrate specificity of TsaS is consistent with the great sequence divergence and the inability to assign related functions to the TSUP family. Mampel et al. (2004) agreed with previous findings suggesting that secondary active proton symport was the energizing mechanism.

Another TSUP homologue, TauE (TC # 9.A.29.2.1) was predicted to be a sulfite exporter and to possess 8 TMSs, but its mechanism of action had not yet been investigated (Weinitschke et al., 2007). A sulfate uptake porter termed CysZ (TC # 9.A.29.6.1) was also shown to belong in the TSUP family. As for TauE, its mechanism of action is still unknown (Rückert et al., 2005). Yet another homologue, SafE1 (TC # 9.A.29.2.2), was proposed to be a sulfoacetate exporter (Krejcík et al., 2008). A recent study has identified PmpC, a TSUP family member, to be part of the PigP regulon in *Serratia* sp. strain ATCC 39006. It was predicted to be inner-membrane localized along with the DUF395 family proteins PmpA and PmpB. Pmp A, B and C were all predicted to transport components of sulfur-containing compounds (Gristwood et al., 2011).

Although the TSUP family has been identified as a family of transporters, not much is known about this large and sequence divergent family. It has not been definitively shown that the primordial peptide unit that duplicated to give 8 TMS proteins did in fact consist of 4 TMSs. Nor is it known how the putative 4 TMS unit arose. Moreover, relationships to other transporter families have not been

investigated. No conserved motifs have been identified for this family. Additionally, structural characteristics such as sidedness and rigorous determination of TMS number have not been performed.

The hierarchical classification approach used, in combination with a lack of abundant data, placed the TSUP family of transmembrane proteins into the "recognized transporters of unknown biochemical mechanism category". Therefore, the goal of our comprehensive study is to establish the evolutionary appearance of the family, to look for other families to which it may be related, identify possible functional domains or signature motifs, provide further structural and topological data, and if possible, suggest related functions for members of the family. The questions posed are important and need to be answered for a full bioinformatic characterization of this family. Although it is unlikely that characterization of the TSUP family will lead to disease and pathogen related breakthroughs, it is likely that in the long term, continuing the characterization of the TSUP family and transporter families like it will prove to yield data relevant to bacterial pathogenesis and the use of microorganisms for purposes such as biomedication.

To achieve our goal, bioinformatic tools, techniques, and principles will be used. Establishing the evolutionary appearance of entire families has become feasible due to the availability of a large number of sequenced genomes in databases such as NCBI (http://www.ncbi.nlm.nih.gov), advances in related software, and the advent of the superfamily principle (Altschul et al., 1990; 1997; Devereux, 1984; Zhai et al., 2002; Pearson, 1998; Doolittle, 1986; Saier et al., 2009). In this approach, homology

will be determined between related proteins throughout most their lengths. To confirm homology, internal repeat elements will be compared using various programs, statistical approaches, and the superfamily principle. Comparing the TSUP family to other possibly related families will be done in a similar way. Moreover, the proposed pathway for transport system evolution is: peptides → channel proteins → secondary carriers → primary active transporters and group translocators (Saier et al., 2000b). If the mode of transport that TSUP utilizes is indeed secondary active transport, then the identification of other families with different modes of transport to which this family is related can place it into a larger and more diverse superfamily.

Any possible functional domains present in the TSUP family will be identified using the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/cdd) and signature motifs will be identified using MEME and programs like it (Marchler-Bauer et al., 2009; Bailey et al., 1994; 1998). WHAT, AveHAS, TMHMM 2.0, HMMTOP, and the positive inside rule will allow for topological determinations such as a more accurate TMS count and protein sidedness (Zhai and Saier, 2001a; 2001b; Tusnády et al., 1998; 2001; Sonnhammer et al., 1998; Krogh et al., 2001; von Heijne et al., 1998; Möller et al., 2001). Genome context analyses using TheSeed (http://www.theseed.org), and transcription factor binding site analyses, using RegPrecise and RegPredict (http://regprecise.lbl.gov/RegPrecise/; http://regpredict.lbl.gov/regpredict/) will be performed in order to predict possible related functions (Overbeek et al., 2005; Novichkov et al., 2010a; 2010b).

Methods

The BLAST function of TCDB was used to identify putative TSUP family members, Orf of *Pyrococcus abyssi* (gi# 74545625; TCID 9.A.29.4.1), YfcA of *Escherichia coli* (gi# 82592533; TCID 9.A.29.3.1.) and Orf of *Oryza sativa* (gi# 75252893; TCID 9.A.29.5.1). The TSUP members exhibited high negative e-values, indicating a high level of similarity. The first iteration of the PSI-BLAST operation (NCBI) was then performed with all 3 proteins using normal settings, with the output set to 1000 sequences, and with a cutoff of e^{-4}. A second iteration was performed in the same manner, but with a stricter cutoff of e^{-6} to minimize false positives (Altschul et al., 1990; 1997). Due to program restrictions, the corresponding TinySeq XML files were input into the MakeTable5 program separately and a 70% cutoff was used in order to remove sequence fragments, redundancies, and sequences having greater than 70% identity (Yen et al., 2009). An exceedingly high homologue count dictated the use of the CD-HIT program at 45% cutoff on the file containing all of the combined sequences (Li and Godzik, 2006). 214 sequences remained, and using MakeTable5 at a cutoff of 100% created a FASTA file for the sequences as well as a table, which included the corresponding abbreviation, sequence description, organismal source, size, gi number, organismal group or phylum, and organismal domain for each protein.

Throughout our study, the CLUSTAL X program was used to create multiple alignments of homologous proteins and the creation of phylogenetic tree files to be visualized using the FigTree program (Thompson et al., 1997; Larkin et al., 2007; Rambaut, 2009). Based on the multiple alignment, 27 sequences that introduced large

gaps, which could have possibly impeded the identification of conserved residues, were removed from the study, bringing the total number of TSUP homologues to 187.

For topological analyses of single protein sequences the WHAT, TMHMM 2.0, and HMMTOP programs were used (Zhai and Saier, 2001a; Tusnády, et al., 1998; 2001; Bailey et al., 1994; 1998). Based on a study describing their effectiveness, the TMHMM 2.0 program was used in TMS count determinations while the HMMTOP program was used for determining protein sidedness (Möller et al., 2001; Sonnhammer et al., 1998). For cases where the TMS count was in agreement, but protein sidedness differed between the two programs, the positive inside rule was used to make educated guesses concerning protein sidedness (von Heijne and Gavel, 1988). The results are in line with the claims made in the paper describing the programs' areas of effectiveness. Inputting the multiple alignment file generated by CLUSTAL X into the Average Hydropathy, Amphipathicity, and Similarity (AveHAS) program facilitated topological assessments across multiple proteins or entire families (Zhai and Saier, 2001b).

Based on visual analysis of AveHAS plots, internal repeats were examined using the IntraCompare (IC) program (Zhai and Saier, 2002). The best comparison scores, represented as standard deviation (S.D.) values, were confirmed and analyzed further using the GAP program (Devereux et al., 1984). The GAP program was set at the default settings with a gap creation penalty of 8 and a gap extension penalty of 2; it was instructed to perform 500 random shuffles. A length of 60 amino acyl residues, the average size of a prototypical protein domain, and 10 S.D., corresponding to a

probability of 10^{-24} that the level of similarity arose by chance, is considered sufficient to prove homology between two proteins or internal repeat units (Dayhoff et al., 1983; Saier, 1994; Saier et al., 2009, Yen et al., 2009). The ever-increasing number of sequences available in databases may in the future require that the 10 S.D. be shifted upwards to 11 or 12 S.D. in order to establish homology. Optimization of the GAP alignment was performed on sequences by maximizing the number of identities, minimizing gaps, and removing non-aligned sequences at the ends. Optimization yields a higher comparison score that better represents the level of similarity between two shorter internal sequences.

The initial BLAST operation conducted on TCDB did not identify any members of possibly related families with a negative e-value of greater than -4. Instead, a large screen was performed comparing the TSUP family against all families of the TC 2.A and 9.A classes. The relevant sequences were input into the Protocol1 program (V. Reddy, unpublished), which automatically obtained homologous sequences from NCBI and ran MakeTable5 using a 70% identity cut-off (Yen et al., 2009). The SSearch program feature of Protocol2 (V. Reddy, unpublished) was then run in order to compare each family to the TSUP family (Pearson, 1998). Any promising results with a S.D. of greater than 10 were automatically compared using the GSAT program feature of Protocol2. GSAT is similar to GAP, which is also an option for Protocol2, but is recommended over GAP when using Protocol2. Confirmation and optimization using GAP was performed as described previously.

Our results led to the creation of a superfamily tree using the SuperfamilyTree 1 program (Chen et al., 2011 (in preparation); Yen et al., 2009; 2010).

A search for functional domains within TSUP members was performed using the conserved domain database (CDD) of NCBI (Marchler-Bauer et al., 2009). Protein sequence motifs were identified using the MEME and MAST programs in 2 separate runs due to program restrictions (Bailey et al., 1994; 1998). The most conserved motifs across the 2 runs were analyzed and blended into a single conserved motif based on individual amino acid conservation. The appearance of duplicates of conserved motifs was noted as further proof of internal repeat elements.

To propose a possible related function, genome context analyses were performed using The SEED-Viewer, which allowed the exploration of over 4,000 curated genomes in order to find homologous genes, their operon context, and consequently their known or putative roles in other organisms (Overbeek et al., 2005). This was done alongside RegPrecise and RegPredict, which allow for the identification of transcription factor binding sites (Novichkov et al., 2010a; 2010b).

Chapter 1: Phylogenetic Analysis of TSUP Family Members and Orthologous Relationships within the Phylogenetic Clusters

The 189 TSUP family members were divided into 15 clusters based on their positions in the phylogenetic tree and the corresponding dendogram (Fig. 1 and 2). The phylogenetic tree for the genuses represented in this study is shown in Figure 3 and the corresponding dendogram in Figure 4. The proteins were listed in a clockwise fashion in Table 1, based on the phylogenetic tree. The Conserved Domain Database (CDD) was used to search for functional domains (Marchler-Bauer et al., 2011). The equivalent DUF81, TauE and COG0730 domains, described in CDD as predicted permeases, are interchangeable and characteristic of the TSUP family (Krejcík et al. 2008; Weinitschke et al., 2007). Almost all TSUP homologues studied, even those having internal, N-, or C-terminal extensions, were found to have the entire TauE domain.

Cluster 1 (43 proteins) includes proteins derived from plants as well as many types of unicellular eukaryotes. This group of proteins is extremely diverse as revealed in the phylogenetic tree shown in Figure 1. All of these proteins are from eukaryotes, but a large number of phyla are represented. The primary phyla include: plants, Apicomplexa and ciliates, but many other phyla are represented as well (See Table 1). Considering the diversity of cluster 1 organismal origins, it is not surprising that these sequences show such great sequence diversity. The average size for these proteins is 572 ± 312 amino acids (aas), but five of these homologues are much larger than the others. These proteins include Tth4, Cre1, Gla2, Cre2, and Tgo1. Excluding

these five proteins, the average size is 476 ± 71 aas. Examination of the individual proteins reveals that their typical sizes range from 385 to about 570 aas. Topologies for these proteins are of 2 and 5 to 11 putative TMSs. The 2 TMS protein, Tgo2, appears to be a fragment, but the remaining may be full length. For the larger proteins, extra hydrophilic domains can be found at the N-termini and between TMSs, but not at the C-termini. None of these regions proved homologous to other proteins of the NCBI database, and none represents a functional domain recognized by CDD. Some of these may be artifactual due to intron translation.

The two proteins that prove to be most distantly related to the other members of cluster 1 are from 2 *Cryptosporidium* species within the Apicomplexa phylum. However, Apicomplexa proteins occur at four additional positions within this cluster. Similarly, plant homologues occur in three distinct positions, again suggesting a lack of orthology. The Bacillariophyta phylum, the Codonosigidae phylum, and also the Oligohymenophorea phylum are each represented in two positions in cluster 1. For example, eleven homologues from *Paramecium* and *Tetrahymena* cluster together while a twelfth is encoded distantly from the others. Three possibilities can be considered for this one protein: it could be a pseudogene, it could be an early arising paralogue, or it could have arisen by horizontal gene transfer. In any case, the configuration of this protein cluster clearly does not follow expectation for orthologous relationships. Similarly in plants, we find a single large cluster with 8 homologues including paralogues and probable orthologues. However, two other

small clusters of plant homologues were identified, again showing that orthology is not generally observed in this cluster.

Cluster 2 proteins, consisting of 40 homologues, are much smaller than those in cluster 1. Most have sizes of about 260 aas, and the average size for all of these proteins is 264 ± 21 aas. The largest of these is a protein from *Franscisella tularensis*, a γ-Proteobacterium, with 366 aas. The size of this protein reflects the presence of an N-terminal 110 aa hydrophilic extension that was not homologous to anything in the NCBI database. Although most proteins within this cluster are predicted to have from 5 to 8 TMSs, based on TMHMM 2.0 (Möller et al., 2001; Sonnhammer et al., 1998), an AveHAS plot suggested the presence of 8 or 9 TMSs for the entirety of cluster 2, where the 9th TMS was substantially less well conserved than the other 8. It should be noted that all 8 putative TMSs occur in pairs of 2 where TMSs 1 and 2 lie close together, and the same is true for TMSs 3 and 4, 5 and 6, and 7 and 8 (See Figs. 5a-b).

Cluster 2 proteins derive from many phylogenetic groups of bacteria and also from a single plant, *Ricinus communis*. This last mentioned protein could be a chloroplast protein, explaining its phylogenetic position in cluster 2. However, horizontal gene transfer is another possibility.

The following bacterial phyla are represented: all 5 sub-phyla within the proteobacteria, Actinobacteria, Verrucomicrobia, Fusobacteria, Firmicutes, Spirochetes and Thermotogae. This observation suggests that extensive horizontal gene transfer has occurred during the evolution of this cluster.

Cluster 3 (5 proteins) derives from 2 δ-Proteobacteria, 1 Chloroflexi, 1 bacterium of unknown phylum, and 1 Euryarchaeota. Predicted topologies for these proteins range from 5 to 8 TMSs. The average size of all of the proteins in cluster 3 is 358 ± 174 aas and 252 ± 2 aas when the larger Dde1 and Orf6 proteins were excluded. Dde1 (379 aas) from *Desulfovibrio desulfuricans* contains an extension between the first and second TMSs that does not represent a CDD-recognized functional domain.

A search for functional domains in CDD revealed that the larger Orf6 protein contains an N-terminal extension including an approximately 160 aa N-terminal fragment of the degP_htrA_DO domain (Kroger et al., 2002). The catalytic characteristic of this functional domain suggests that Orf6 may have serine protease activity and/or chaperone function, possibly protecting the bacteria from stress. The degP_htrA_DO fragment observed in Orf6 comprises about a third of the full domain and half of the protease sub-domain, which possibly suggests limited endopeptidase activity (Lipinska et al., 1990; Spiess et al., 1999). The N-terminal TMS preceding the degP_htrA_DO domain may target this domain to the cell surface. Members of this serine protease family usually reside in the periplasm, and the degP_htrA_DO domain observed in Orf6 is probably localized to the periplasm. It is possible that Orf6 possesses group-translocator-like properties, transporting and hydrolyzing peptides. Although a phylum assignment for Orf6 has not been made, the corresponding 16S rRNA clusters with the Verrucomicrobia.

The average protein size of **cluster 4** (6 proteins) was calculated to be 432 ± 214 aas when all proteins were included. When the larger Bsp1, Cgl1, and Gla1

proteins were excluded from the calculation, the average protein size was 260 ± 18 aas. Cluster 4 members possess 7, 8 and 10 TMS predicted topologies. Out of the 6 proteins in cluster 4, only Bsp1, which posses a C-terminal extension, was found to have an extra functional domain. This extension possesses two tandem copies of the USP_like domain. Universal Stress Protein family members are upregulated in response to stress agents, which helps the cell tolerate stressful states (Sousa and McKay, 2001). All proteins within the USP family possess the ATP binding alpha/beta fold motif, but not all are able to bind ATP. The existence of a potential ATP binding or hydrolyzing domain introduces the possibility that this transporter can either be regulated by ATP or be energized by ATP hydrolysis and thus function as a primary active transporter rather than a secondary carrier.

The 9 proteins comprising **cluster 5** derive from all domains of life. The 7 bacterial proteins derive from Actinobacteria, Cyanobacteria and δ-Proteobacteria, while the two eukaryotic and one archaeal proteins derive from Bacillariophyta and Crenarchaeota, respectively. The average size for all of these proteins is 561 ± 839 aas; however, when Ptr3, the largest of the TSUP proteins included in this study, was excluded, an average value of 282 ± 25 aas was calculated. Furthermore, although Ptr3 was predicted to contain 4 TMSs by the TMHMM 2.0 program, the WHAT program (Zhai and Saier, 2001a) predicted 8 TMSs. A value of 10 TMSs is likely to be more accurate based on inspection of the hydropathy profile and the 8 TMS topologies of the proteins with which it clusters. This protein of 2798 aas has an N-terminal hydrophobic domain followed by a hydrophilic domain of over 2000 residues

that shows sequence similarity with a hydrophilic protein of 2409 aas from *Thalassiosira pseudonana*. Based on the 16S rRNA tree, the three Actinobacterial homologues and the three Cyanobacterial homologues may all be orthologous.

Cluster 6 (14 proteins) derives from bacteria and archaea. A diverse set of bacterial phyla is represented as is true for the archaeal. In the latter domain, members are derived from the Euryarchaeota, Korarchaeota and Crenarchaeota. This fact, plus the observation that the bacterial and archaeal homologues are interspersed, clearly suggests that extensive lateral gene transfer has occurred during the evolution of this cluster. The sizes of these proteins range from 250 aas to 333 aas, with the average being 277 ± 20 aas. Pto1 from *Picrophilus torridus*, a Euryarchaeota, is the largest protein in the cluster with 333 aas. The size of this protein reflected a 60 aa hydrophilic insert between TMSs 5 and 6 that was not homologous to anything in the NCBI database. Cluster 6 proteins are of 7 or 8 putative TMS topologies.

Cluster 7 is one of the smaller clusters, containing only 2 proteins derived from Actinobacteria. Bad1 from *Bifidobacterium adolescentis* and Gva1 from *Gardnerella vaginalis* are 292 and 267 aas in size respectively, which results in an average size of 280 ± 18 aas for cluster 7. Both proteins possess 8 putative TMSs.

Cluster 8 (30 proteins) is the third largest cluster and includes proteins solely from bacteria. All of the proteins cluster closely together; however, two sub-clusters are present. The proteins from Asa1 to Cbu2 comprise the first sub-cluster, while proteins from Lsp1 to Iba1 form the second sub-cluster. The phyla represented in the first sub-cluster are restricted to the α- and γ-Proteobacteria. In this sub-cluster, the 7

α-proteobacterial homologues are flanked by two γ-proteobacterial homologues. The second sub-cluster consists of proteins derived primarily from β- and γ-Proteobacteria. These proteins are interspersed, suggestive of lateral gene transfer. Two proteins from Bacteroidetes and Cyanobacteria are also represented, and these are sandwiched in between the proteobacterial homologues, again suggestive of lateral gene transfer. Very little size variation is observed in cluster 8, with the average size of these proteins calculated to be 275 ± 11 aas and individual proteins ranging from 264 aas to 314 aas in size. Topologies of 6 to 9 putative TMSs are represented in this cluster, with the majority possessing 8 TMSs.

Clusters 9 and **10** each consists of a single protein from a γ-Proteobacterium. Epe1 (287 aas) from *Endoriftia persephone* and Pal1 (271 aas) from *Providencia alcalifaciens* (clusters 9 and 10, respectively) have sizes similar to those observed for proteins in cluster 8. Epe1 was predicted to have 5 TMSs by TMHMM 2.0, while WHAT predicted 6 TMSs, with the hydropathy profile indicating between 6 and 8 TMSs. Pal1 possesses 8 putative TMSs.

Cluster 11 (3 proteins) is derived solely from archaea. Euryarchaeota represents the phylum for 2 of these proteins, while Orf3 from an uncultured archaeon, is most likely to also be in the Euryarchaeota phylum given its 8 TMS topology and close clustering with the other 2. The average size of these 3 proteins is 272 ± 3 aas, with sizes ranging from 270 to 276 aas.

Like cluster 11, **cluster 12** (4 proteins) derives from archaea. However, both the Euryarchaeota and Crenarchaeota phyla are represented. These proteins possess 7

to 8 TMSs and range in size from 251 to 269 aas, with an average size of 260 ± 9 aas. Their 7 to 8 TMS topologies are the same as observed for the cluster 11 archaeal proteins. Comparison with the 16S rRNA tree reveals that these proteins cannot be orthologous.

Cluster 13 (10 proteins) derives almost exclusively from Firmicutes, and most of these proteins may be orthologues. However, homologues from a single Crenarchaeota and a bacterium of unknown phylum are also represented. These proteins show very little size variation, ranging from 255 aas to 300 aas, with the average of 273 ± 13 aas. The uniformity in size is accompanied by a uniform 7 to 8 putative TMS topology for all 10 proteins. Although a phylum assignment for Tte1 has not been made, the corresponding 16S rRNA clusters with the Chloroflexi.

Cluster 14 (7 proteins) derives from the α-Proteobacteria, Firmicutes, Euryarchaeota, and Spirochaetae phyla. Considerable size variation is observed in this cluster, with the Bja1, Dau1 and Dre4 proteins representing the larger of the seven proteins. The average size for this cluster is 340 ± 61 aas and 296 ± 30 aas when the larger proteins were removed. Despite the variation in size, and the great phylogenetic distances between these homologues, the topologies are fairly uniform with members possessing 7 to 9 putative TMSs.

Cluster 15 (14 proteins) derives from eight different phyla. Two archaeal proteins derive from different phyla, and within the Proteobacteria we find representation within the beta-, gamma- and delta- subphyla. The three Firmicute homologues are sandwiched in between those derive from other phyla. In spite of the

tremendous sequence divergence of these homologues, their size variations (from 246 aas to 569 aas) is only slightly greater than two fold. Since their predicted topologies range from 7 to 9 TMSs with the majority exhibiting 8 TMSs, it is possible that they exhibit a uniform topology. The average size for the proteins in this cluster is 316 ± 86 aas and 280 ± 26 aas when the larger Mmu1, Cph2 and Dre3 proteins were excluded.

Based on the CLUSTAL X alignment, a total of 17 proteins that introduced large gaps into the alignment were removed from further analysis (Thompson et al., 1997; Larkin et al., 2007). A subsequent examination of the proteins revealed that six of them did not possess the TauE or any other CDD-recognized functional domain. These proteins possessed 2, 3, 6, 7 or 9 putative TMSs. The remaining 11, while possessing the TauE domain, had 5, 7, 8, 9 or 10 TMSs. These predicted topologies were observed in the 189 TSUP homologues included in this study, so their removal may have been due to certain unique sequence features not observed in the rest of the proteins.

Chapter 2: Topological Analyses of the TSUP Family

The average hydropathy (top, dark lines), amphipathicity (top, light lines) and similarity (bottom) plots for the 189 TSUP family members included in this study are presented in Figures 5a and b. The AveHas plot reveals 8 major peaks of hydrophobicity that correlate with the 8 major peaks of similarity. The peaks of amphipathicity suggest that the TMSs are of moderately amphipathic nature. TMSs 1 to 4 cluster together; separated from TMSs 5 to 8, which also cluster together. Upon initial inspection, this symmetry is suggestive of intragenic duplication in which case TMSs 1 to 4 would be repeated in TMSs 5 to 8.

TMSs 1 and 2 cluster closely together, as do TMSs 3 and 4. The clustering pattern of the first 4 TMSs further suggested the duplication of a primordial 2 TMS peptide to create a 4 TMS peptide that underwent subsequent duplication to give rise to the prototypical 8 TMS proteins present today. The area on the plot corresponding to TMS 1 divides into one large peak corresponding to a sufficient number of amino acyl residues to form an α-helix and one smaller peak corresponding to 9 amino acyl residues. The second smaller peak is possibly due to misalignment of some sequences relative to others.

TMSs 5 to 8 do not cluster as closely as TMSs 1 to 4. TMSs 5 to 7 are spaced about evenly from each other and cluster closely with one another. TMS 8 is separated from TMSs 5 to 7 by a hydrophilic loop of substantial size. Furthermore, both TMS 5 and 8 display a problem of misalignment, similar to what is observed for TMS 1. In TMS 5, the larger peak precedes the smaller one, as is true for TMS 1. The

smaller peaks in TMSs 1 and 5 display a well-conserved proline residue, which is a further qualitative suggestion of a duplication event. The misalignment observed for TMS 8 resembles that for TMSs 1 and 5, except that the smaller peak precedes the larger one.

The first four TMSs are separated from the last four by a large hydrophilic loop. A search using the CDD did not reveal a conserved domain in this central region. Two significant peaks of hydrophobicity and two corresponding small peaks of similarity are observed within the loop region prior to TMS 5. These peaks most likely represent the 9, 10 and 11 TMS proteins most commonly associated with the eukaryotic members of the TSUP family. 2, 4, 5, 6 and 7 TMS topologies have also been observed, but these are most commonly associated with the bacterial and archaeal domains. At least some are due to artifactual truncations due to inappropriate choices of initiation codons or to sequencing errors. The large majority of the proteins have an 8 TMS topology.

Chapter 3: Establishing Internal Repeats Within TSUP Family Members

As previously discussed, the majority of TSUP family members possess 8 predicted TMSs, while the majority of proteins that deviate from the prototypical topology possess 7 or 9 TMSs, as a result of apparent deletions or insertions. To establish the evolutionary appearance of this family, the IC and GAP programs were used to compare putative repeat elements. A comparison of 60 residues and a comparison score of greater than 10 S.D. is considered sufficient to establish homology (Dayhoff et al., 1983; Devereux et al., 1984; Saier, 1994; Saier et al., 2009, Yen et al., 2009; Zhai and Saier, 2002).

Initial visual inspection of the AveHAS plot indicated the presence of a 4 TMS repeat unit. Comparing TMSs 1-4 of all TSUP homologues with TMSs 5-8 of the same homologues, the IC program identified many repeat units with comparison scores in excess of 10 S.D.

As a representative comparison, Tko1 (7 TMSs predicted with TMHMM 2.0; 8 TMSs predicted with HMMTOP) and Mch1 (8 TMS) were chosen. Comparing TMSs 1 - 4 with TMSs 5-8 of Tko1 resulted in a comparison score of 26.3 S.D. (Fig. 6b). Similarly, comparing TMSs 1 - 4 with TMSs 5 - 8 of Mch1 resulted in a comparison score of 17.0 S.D. (Fig. 6c). After optimization, comparing TMSs 1 - 4 of Tko1 with TMSs 5 - 8 of Mch1 gave 16.0 S.D. (Fig. 6d). Thus, application of the superfamily principle (Doolittle, 1986) strongly indicates that an intragenic duplication event occurred in the evolution of the TSUP family, wherein TMSs 1 - 4 were duplicated to give TMSs 5 - 8 (Fig. 6a). Furthermore, the large comparison scores, far exceeding

the requirement to prove homology, indicate that the duplication event occurred fairly recently in evolutionary time.

We considered the possibility that a 2 TMS precursor peptide might have duplicated to give the primordial 4 TMS peptide (Gomolplitinant and Saier, 2011). Comparisons of TMSs 1- 2 with TMSs 3 - 4, 5 - 6 with 7 - 8, 1 - 2 with 7 - 8, and 3 - 4 with 5 - 6 were conducted. The maximal comparison scores were far less than 10 S.D. The 2 TMS fragments compared were usually only about 45 aas in length, so even if sufficient comparison scores had been attained, a 2 TMS duplication event could not be established. In contrast to the 4 TMS duplication event, which we believe to be an evolutionarily recent occurrence, the 2 TMS duplication event could have occurred, but it may no longer be detectable because it was an ancient event. Also, similarly to how paralogues may gain functions distinct from the protein products of the original gene from which they arose, TMSs 3 and 4 may have diverged in sequence substantially after the duplication event.

There are numerous other possibilities to explain the 4 TMS topology. For example, a 3 TMS peptide may have gained 1 TMS through an insertion to generate the 4 TMS topology. A 6 TMS protein which lost 2 TMSs or a 5 TMS protein which lost 1 TMS provide other possible explanations for a 4 TMS topology. All of these have been observed in evolutionary studies of transport protein families (Zheng et al., 2011 (in preparation)).

Chapter 4: The Microbial Rhodopsin Superfamily

All protein families within TCDB belonging to class 2.A consist of electrochemical potential-driven uniporters, symporters and antiporters. Using a modified SS Search program (Pearson, 1998; V. Reddy and M.H. Saier, unpublished program) to compare TSUP homologues with all other secondary carriers, we could identify potential superfamily relationships between TSUP family members and other secondary carriers. Comparisons to the 9.A class proteins were also performed. Our results led to the creation of the novel "Microbial Rhodopsin" (MR) superfamily for which a tree was generated using the SuperfamilyTree 1 and 2 programs (Fig. 7a; Chen et al., 2011; Yen et al., 2009; 2010). The MR superfamily consists of six protein families with members of 6, 7 and 8 putative TMSs (Table 2). An evolutionary pathway for the appearance of the different members of the MR superfamily was also proposed (Fig. 7b). Our comparisons of superfamilies of differing topological types will be discussed in Chapter 5.

The Lysosomal Cystine Transporter (LCT) Family (TC# 2.A.43)

The evolutionary pathway of the 7 TMS LCT family has been elucidated (Zhai et al., 2001). LCT family members were found to be homologous to members of the Ion-Translocating Microbial Rhodopsin (MR) family (TC# 3.E.1). All such transporters are light-driven ion pumps. This led us to include the MR family in the superfamily and to construct the superfamily tree (Fig. 7a-b; Table 2). Because the MR family has been extensively characterized from structural, functional and

mechanistic standpoints, we have chosen to name this new superfamily after this family.

For the LCT family, it was shown that TMSs 1-3 duplicated to give rise to TMSs 5-7, with TMS 4 showing insignificant sequence similarity to any one of the other six TMSs. The precursor may have been an 8 TMS protein which generated the present-day 7 TMS protein by loss of TMS 1 or 8. The 8 TMS Bco1 protein of the TSUP family is homologous to the 7 TMS Aca2 protein of the LCT family (Fig. 8). The region of homology demonstrated includes TMSs 5-7 of Bco1 and 5-7 of Aca2. Thus, TMS 8 in the TSUP family was lost to generate the 7 TMS topology of the LCT family. A comparison score of 11.8 S.D. was obtained (Fig. 8).

The Ni^{2+}-Co^{2+} Transporter (NiCoT) Family (TC# 2.A.52)

Members of subfamily 1 within the ubiquitous NiCoT family are typically 300 to 380 aas in size and possess 6-8 putative TMSs (Saier et al., 1999). NiCoT subfamily 2 is comprised of distant homologues of great size, sequence and topological variation. NiCoT transporters catalyze the uptake of Ni^{2+} and Co^{2+} using a pmf-dependent mechanism; however, a Ni^{2+} and Co^{2+} resistance protein that is believed to export the two metals to the external environment has been reported (Iwig et al., 2006; Rodrique et al., 2005). Smaller members of subfamily 2 exhibit larger topological variation of 4-8 TMSs.

Comparing TMSs 1-3 of TSUP Bja1 (8 TMSs) with TMSs 4-6 of NiCoT Pla1 (6 TMSs) yielded a comparison score of 12.8 S.D. (Fig. 9). This comparison establishes homology between members of these two families and serves to confirm

our proposed evolutionary pathway for the appearance of the NiCoT family within the MR superfamily (Fig. 7b). Based on alignments, it is likely that the 6 TMS NiCoT proteins arose from the loss of TMSs 1 and 8 after the 4 TMS intragenic duplication event.

The Branched Chain Amino Acid Exporter (LIV-E) Family (TC# 2.A.78)

Members of the LIV-E family are restricted to the bacterial and archaeal domains. LIV-E family members utilize H^+ as the antiported cation and two integral membrane proteins to catalyze export of branched chain amino acids and methionine. It has been postulated that these systems arose as a result of a rare intragenic triplication event starting with a 4 TMS primordial peptide, which resulted in the present day 8 and 4 TMS-encoding elements (Kennerknecht et al., 2002). The larger components of this family are typically around 250 aas long and possess 8 or 7 TMSs in a 4 + 4 or 3 + 1 + 3 arrangement, respectively.

Comparing TMSs 2-4 of TSUP Cba1 (8 putative TMSs) with TMSs 4-6 of LIV-E Arpr1 (7 putative TMSs) yielded a comparison score of 12.2 S.D. (Fig. 10a). The alignment explains the 3 + 1 + 3 topology by demonstrating that LIV-E most likely arose in an unusual manner in which TSUP TMSs 2-5 (post-duplication) were duplicated, to given an 8 TMS protein, and in some homologues, TMS 2 was lost at the N-terminus to yield 7 TMS homologues (Fig. 7b). Therefore, TSUP TMS 2 corresponds to TMS 4 in the 7 TMS LIV-E proteins, and TSUP TMSs 3-5 correspond to TMSs 1-3 and 5-7 in the same LIV-E homologues. This suggestion is supported by a separate alignment in which TMSs 3-5 of TSUP Cce1 (8 putative TMSs) align with

TMSs 1-3 of LIV-E Enfa2 (7 putative TMSs) to give a comparison score of 11.1 S.D. (Fig. 10b). Thus, our comparisons support the superfamily assignment and propose a unique evolutionary path for the LIV-E family. Because of its unique evolutionary pathway, structural similarity between the TSUP and LIV-E families is not predicted.

The Organic Solute Transporter (OST) Family (TC# 2.A.82)

Members of the OST family are almost exclusive to animals and are known to transport organic anions, estrone-3-sulfate, bile acids, taurocholate, digoxin and prostaglandin E1 (Dawson et al., 2005; Dawson et al., 2010; Seward et al., 2003; Wang et al., 2001). Distant homologues of the α-subunits in plants, fungi and bacteria are brought up in NCBI searches, but their scores usually border, or fall below our threshold cutoff. Furthermore, each well characterized transporter within this family functions as part of a two-component system utilizing the α- and β-subunits. The α-subunit generally contains seven TMSs, whereas the β-subunit contains only one. So far, neither subunit can function without the other (Dawson et al., 2005; Dawson et al., 2010).

Comparing TMSs 2-3 of TSUP Tsp1 (8 putative TMSs) with TMSs 1-2 of OST Cre2 (7 putative TMSs) yielded a comparison score of 12.1 S.D. (Fig. 11). Of note, the two segments compared are 59 and 61 residues in length, respectively. In a separate comparison, TMSs 5-6 of TSUP Ddi1 (8 putative TMSs) of *Dictyostelium discoideum* AX4 (gi 66825573) aligned with TMSs 5-6 of OST Dre4 (8 putative TMSs) of *Danio rerio* (gi 52218944) to give a sufficient score (11.5 S.D.) to establish homology (unpublished data). Whereas the first comparison demonstrates the loss of

TMS 1 in OST transporters, the second establishes homology between the second halves of the two families. Therefore, the loss of TMS 1 generated the 7 TMS topology of the OST family. Further studies will be required to elucidate the evolutionary route taken by this protein family. This route may have involved the precursors of the TSUP family.

Chapter 5: Tying Together Superfamilies? Evidence for an Ancestral Transmembrane Hairpin Structure

Based on sequence similarity data, we herein provide evidence for and suggest that an ancestral 2 α-helical hairpin structure gave rise to many families and superfamilies of differing sizes, topologies and functions.

The Cation Diffusion Facilitator (CDF) Family (TC # 2.A.4)

Most members of the ubiquitous CDF family possess 6 TMSs; however certain eukaryotic and mammalian homologues possess from 12 to 15 TMSs (Cousins et al., 2006; Cragg et al., 2002; Paulsen and Saier, 1997). This family is diverse in both sequence and size (300-750 aas). YiiP (TC # 2.A.4.7.1) of *E. coli* functions as a homodimer, and, as for other members of the CDF family, it possesses highly conserved aspartyl residues (D49 and D147 in YiiP), similar to those present in monovalent cation secondary active efflux pumps (Wei and Fu, 2006). CDF family members are believed to cluster according to their divalent cation specificities, based on the limited amount of functional data currently available (Montanini et al., 2007; Matias et al., 2010).

CzcD of *Bacillus subtilis* (TC # 2.A.4.1.1) may couple H^+/K^+ uptake to the efflux of Zn^{2+}, Cd^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, or Pb^{2+} (Guffanti et al., 2002). This possibly suggests electroneutrality for monovalent cation uptake coupled to divalent heavy metal ion export, but the exact stoichiometry has not been established. Surprisingly, evidence has been presented leading to the conclusion that the ancient, ubiquitous, 6 TMS CDF carriers gave rise through evolution to 4 TMS Ca^{2+} release-activated

(CRAC; TC# 1.A.52) Ca^{2+} channels in animals via the loss of TMSs 1 and 2 (Matias et al., 2010). This seems to be a rare instance of "reverse evolution" where a complex protein was the precursor of a structurally and functionally simpler protein.

Comparing TMSs 1-6 of TSUP Min1 (8 putative TMSs) with TMSs 1-6 of CDF Dede2 (6 putative TMSs) yielded a comparison score of 11.8 S.D. (Fig. 12). Given that the CDF family arose from the triplication of a 2 TMS unit, and taking into account the homology observed between the TSUP and CDF families as per our alignment and comparison score, we propose that the 4 TMS TSUP primordial peptide arose from the duplication of a 2 TMS unit. Possibly due to extensive sequence divergence, we were unable to prove this directly, but our initial hunch of a 2 TMS duplication now seems much more likely. As will be demonstrated, throughout evolution, the 4 TMS unit was subject to TMS loss at its N- and C-terminal ends, with TMSs 2 and 3 usually being retained within repeat units.

The ATP:ADP Antiporter (AAA) Family (TC # 2.A.12)

Members of the AAA family are mostly found in bacteria and plants, possibly in chloroplasts. However, a few other eukaryotic homologues exist (Winkler and Neuhaus, 1999). AAA family proteins typically range in size from 430 to 450 amino acyl residues and generally possess 6 to 13 TMSs, with the 12 TMS topology representing the prototypical arrangement. Various nucleotides have been shown to be targets of AAA transport (Winkler et al., 1999). *Rickettsia prowazekii*, an intracellular bacterial parasite that causes the human epidemic typhus, encodes five AAA family paralogues (Alexeyev et al., 1999; Bachah et al., 2010). One of the

paralogues takes up ATP from the animal cell cytoplasm in exchange for ADP, thus providing energy for itself via the gain of a pyrophosphate bond.

Comparing TMSs 2-7 of TSUP Lgr1 (8 putative TMSs) with TMSs 6-11 of AAA Ptr6 (14 putative TMSs) yielded a comparison score of 11.2 S.D. (Fig. 13a). The two programs disagree on the existence of TMSs 3 and 12 as predicted by HMMTOP. The evolutionary pathway by which AAA proteins arose is unclear; however, the post-duplication 8 TMS TSUP proteins appear to have contributed to the structure and function of AAA proteins. Although the hydropathy plot for AAA Ptr6 suggests a possible 5 or 6 TMS intragenic duplication event, our results implicate a possible gene fusion event wherein TSUP TMSs 2-7 became an internal topological feature of AAA proteins (Fig. 13b).

The Ca^{2+}:Cation Antiporter (CaCA) Family (TC # 2.A.19)

The ubiquitous CaCA family consists of members varying in size from 302 to 1199 amino acyl residues and of varying topologies between 10 and 13 TMSs (Saier et al., 1999). Previous studies have demonstrated that the 10 TMS topology arose from an early tandem duplication of a 5 TMS peptide, resulting in topological inversion of TMSs 6-10 (Sääf et al., 2001; Saier et al., 1999). Most proteins of this family function in Ca^{2+} extrusion using H^+ or Na^+ as the antiported cation. However efflux of K^+, Mn^{2+}, Cd^{2+} and Zn^{2+} has also been observed (Dipolo and Beaugé, 2006).

Comparing TMSs 4-9 of TSUP Dsp3 (9 putative TMSs) with TMSs 1-6 of CaCA Ani2 (11 putative TMSs) yielded a comparison score of 11.3 S.D. (Fig 14). TMS 1 of Dsp3 appears to be an insertion, while TMS 1 in Ani2 appears to be an

insertion and is not demonstrably homologous to TSUP TMS 3 or any other TMS within TSUP. The alignment suggests that the TSUP family is older than the CaCA family and demonstrates an unusual route by which CaCA family members may have arisen from TMSs 3-8 of the 8 TMS TSUP topology, possibly via the loss of TMSs 1 and 2.

The Inorganic Phosphate Transporter (PiT) Family (TC # 2.A.20)

Members of the ubiquitous PiT family range in size from 350 to 690 amino acyl residues. They typically possess 6 to 12 TMSs and catalyze the uptake of inorganic phosphate or sulfate by utilizing H^+ or Na^+ gradients (Saier et al., 1999). In addition to their aforementioned functions, proteins of this family have been observed to transport Mg^{2+}, Ca^{2+}, Zn^{2+}, MoO_4^{2-} and other divalent ions (Aguilar-Barajas et al., 2011; Harris et al., 2001). The mammalian PiT-2 protein functions as a viral receptor, while retaining its phosphate transport function (Salaün et al., 2001). PiT family members arose via a tandem duplication event of a 6 TMS unit (Persson et al., 1998; 1999).

Comparing TMSs 3-7 of TSUP Aeh1 (8 putative TMSs) with TMSs 2-6 of PiT Kol1 (6 putative TMSs) yielded a comparison score of 11.9 S.D. (Fig. 15). Kol1, a 6 TMS protein, appears to be fragmentary and aligns with TMSs 1-6 of the 12 TMS Cje2 protein (gi 68535688). Comparing Kol1 with TMSs 1-6 and 7-12 of Cje2 gave comparable comparison scores of 11-12 S.D., which clearly highlighted the 6 TMS repeat within PiT proteins (unpublished data). The fact that TSUP TMSs 3-7 align with TMSs 2-6 in PiT proteins suggests that the PiT repeat unit arose after the 8 TMS

TSUP topology was generated. TMS 1 of the 6 TMS repeat unit may have been an insertion, or more likely it may have diverged in sequence significantly such that its similarity to TMS 2 of TSUP is undetectable. In the latter case, TMSs 1 and 8 were lost, leaving TMSs 2-7.

The Solute:Sodium Symporter (SSS) Family (TC # 2.A.21)

Most SSS family transporters utilize Na^+ as the symported cation in order to catalyze specific uptake of sugars, amino acids, organo cations, nucleosides, inositols, vitamins, anions or urea (Reizer et al., 1994). However, a previous study showed that the Na^+ concentration has no effect on the function of MctP of *Rhizobium leguminosarum*, a monocarboxylate symporter, while the disruption of the proton gradient, via the use of the CCCP uncoupling agent, leads to strong inhibition of its function (Hosie et al., 2002). SSS proteins have a size range of approximately 400 to 700 aas and a topological range of 10-14 TMSs. The crystal structure of the 14 TMS *Vibrio parahaemolyticus* sodium/galactose symporter (vSGLT) revealed a 5 TMS unit (TMSs 2-6) repeated and topologically inverted in TMSs 7-11 (Faham et al., 2008).

Comparing TMSs 2-4 of TSUP Rjo3 (8 putative TMSs) with TMSs 10-12 of SSS Apme2 (13 putative TMSs) yielded a comparison score of 13.8 S.D. (Fig. 16a). This comparison suggests that TMSs 2-4 of TSUP were an integral part of the 5 TMS primordial peptide, which may have arisen from the duplication of the 3 TMS unit and a loss of the N-terminal TMS (Fig. 16b).

The 2-Hydroxycarboxylate Transporter (2-HCT) Family (TC # 2.A.24)

Members of the 2-HCT family utilize Na^+, lactate, or H^+ gradients to catalyze uptake of citrate and malate (Bandell et al., 1997; Bandell and Lolkema, 2000; Kästner et al., 2002; Kawai et al., 1997; Sobczak and Lolkema, 2005). Citrate/acetate, sodium citrate/OH^-, malate/lactate and citrate/lactate antiporters have also been identified. 2-HCT proteins are restricted to Gram-negative and Gram-positive bacteria, are approximately 450 amino acyl residues in size and possess 9-13 putative TMSs. A 5 or 6 TMS repeat is found in proteins of this family (Sobczak and Lolkema, 2004).

Comparing TMSs 1-4 of TSUP Pte9 (8 putative TMSs) with TMSs 5-7 of 2-HCT Cas9 (13 putative TMSs) gave a comparison score of 11.9 S.D.s (Fig. 17a). Similarly to the SSS family, TMSs 2-4 from the 4 TMS TSUP primordial peptide appear to be part of the repeat unit within members of the 2-HCT family (Fig. 17b).

The Nucleobase:Cation Symporter-2 (NCS2) Family (TC # 2.A.40)

Members of the ubiquitous NCS2 or Nucleobase/Ascorbate Transporter (NAT) family mediate the uptake of purines, pyrimidines and ascorbate using H^+ or Na^+ as the cotransported molecule (Daruwala et al., 1999; Karatza and Frillingos, 2006; Karatza et al., 2006). Proteins of this family range in size from 414 to 650 amino acyl residues and possess between 11 to 14 putative TMSs, with the 12 TMS topology being prototypical (de Koning and Diallinas, 2000; Saier et al., 1999). The hydropathy plots for NCS2 proteins are suggestive of an intragenic duplication of 6 TMSs.

Comparing TMSs 3-8 of TSUP Jsp1 (8 putative TMSs) with TMSs 7-12 of NCS2 Stpu2 (14 putative) yielded a comparison score of 11.2 S.D. (Fig. 18). The two

programs disagreed with respect to the existence of TMSs 8 and 13 of Stpu2 as predicted by HMMTOP. The evolutionary pathway by which NCS2 proteins arose may have involved (a) duplication of a 4 TMS element to give an 8 TMS element, as observed for the TSUP proteins followed by (b) loss of the first two N-terminal TMSs, and then (c) duplication of the resulting 6 TMS element to yield 12 TMS proteins.

The Glycerol Uptake (GUP) Family (TC # 2.A.50)

Members of the GUP family range from 450 to 610 aas in size, contain 8 to 13 putative TMSs and are nearly ubiquitous, being absent only from archaea (Bosson et al., 2006). Members of this family have been implicated in glycerol and activated D-alanine uptake, as well as possessing glycosyl phosphatetidylinositol (GPI) remodelase function (Bleve et al., 2005; Ghugtyal et al., 2007; Heaton and Neuhaus, 1992; Jaquenoud et al., 2008). The evolutionary route by which the GUP family arose has not been elucidated; however hydropathy plots indicate a 5 or 6 TMS repeat unit.

Comparing TMSs 2-4 of TSUP Bja1 (8 putative TMSs) with TMSs 2-4 of Gsp2 (12 putative TMSs) gave a comparison score of 12.4 S.D. (Fig. 19). It appears as though a portion of the 4 TMS primordial peptide may have given rise to the GUP family. A 2 TMS multiplication is a likely route by which the repeat unit within GUP family members arose.

The Sulfate Permease (SulP) Family (TC # 2.A.53)

Members of the ubiquitous SulP family typically possess 10-13 putative TMSs, with the 12 TMS topology being the most likely prototypical precursor of them all (Saier et al., 1999). Hydropathy plots indicate two-fold symmetry. Members of

this family are known to mediate transport of substrates by mechanisms including anion:anion exchange and sulfate:H^+ co-transport (Jiang et al., 2002). Several members can function as channels or as both carriers and channels (Ohana et al., 2011). A SulP homologue has been found fused to rhodanase, suggestive of a sulfate transport role.

When comparing the functionally similar but topologically different TSUP and SulP families, a score of 10.7 S.D. was obtained when comparing TMSs 5-8 of TSUP Bma1 (8 putative TMSs) with TMSs 1-4 of SulP Clu1 (12 putative TMSs; Fig. 20). TMS 3 within Clu1 appears to be a relatively hydrophilic putative TMS.

The Monovalent Cation (K^+ or Na^+): Proton Antiporter-3 (CPA3) Family (TC # 2.A.63)

The CPA3 family is restricted to the bacterial and archaeal domains. Efflux pumps of the CPA3 family function as large multi-component systems and usually transport Na^+ or K^+ (Fukaya et al., 2009; Hiramatsu et al., 1998; Kosono et al., 1999). Gram-positive bacterial systems have been observed to transport Na^+ or Li^+, but K^+, Ca^{2+} and Mg^{2+} are also exported to a small extent (Swartz et al., 2007). The multi-component systems typically consist of seven subunits encoded within an operon. The systems usually contain two larger subunits of 20-24 and 12-16 TMSs and five smaller subunits of 2-4 TMSs.

Comparing TMSs 1-6 in TSUP Oba1 (8 putative TMSs) with TMSs 8-13 of CPA3 Nsp1 (15 putative TMSs) yielded a comparison score of 11.3 S.D. (Fig. 21).

Our comparison suggests that CPA3 arose through the duplication of TMSs 1-6 and possibly 7 of an ancestral TSUP-like protein.

The K^+ Uptake Permease (KUP) Family (TC # 2.A.72)

Transporters of the KUP family are restricted to bacteria, mosses, fungi and plants (Grabov, 2007). Evidence suggests that KUP transporters function via a secondary active proton symport mechanism (Trchounian and Kobayashi, 1999; Zakharyan and Trchounian, 2001). KUP transporters range in size from 600 to 900 aas and contain 10-15 putative TMSs. Based on the average hydropathy plot for these proteins, two-fold symmetry is possible wherein a 6 or 7 TMS unit is repeated.

Comparing TMSs 3-8 of TSUP Cje2 (8 putative TMSs) with TMSs 7-11 of KUP Sbi4 (14 putative TMSs) yielded a comparison score of 10.8 S.D. (Fig. 22). HMMTOP and TMHMM 2.0 were in agreement on the existence and relative locations of TMSs 1-7 in Sbi4, but disagree on the presence of TMSs 8, 12, 13 and 14 (in the C-terminal half) as predicted by HMMTOP. The peaks under question appear strongly hydrophobic and may in fact prove to be TMSs. This comparison suggests that TMSs 1 and 2 of the 8 TMS TSUP topology may have been lost and that the 6 TMS unit duplicated to give rise to the prototypical 12-14 TMS topology within the KUP family.

The Threonine/Serine Exporter (ThrE) Family (TC # 2.A.79)

ThrE family members are ubiquitous, diverse in sequence, approximately 400-600 amino acyl residues in size and possess 9-11 putative TMSs. The few members that have been characterized, catalyze efflux of threonine and serine using the pmf

(Simic et al., 2001). Some distant homologues found in a range of organisms may function as parts of "spliced" two-component systems (Ziegler et al., 2000). The hydropathy profiles of ThrE proteins strongly suggest a 5 TMS repeat unit.

Comparing TMSs 1-3 of TSUP Tko1 (8 putative TMSs) with TMSs 2-4 of ThrE Aod2 (9 putative TMSs) gave a comparison score of 11.4 S.D. (Fig. 23). This comparison, as well as others, suggests that TMSs 1-3 of the 4 TSUP primordial peptide TMSs participated in generating the likely 5 TMS repeat unit within the ThrE family.

Vitamin Uptake Transporter (VUT or ECF) Family (TC # 2.A.88)

Members comprising the VUT (or ECF) family generally possess between four and seven putative TMSs, and range in size from 160-230 aas. They are known to transport vitamins such as biotin, niacin and thiamin. Many proteins of this family are homologous to and function similarly to ABC-2 porters, which arose by a 3 TMS duplication (Rodionov et al., 2002; 2006; 2009; Wang et al., 2009a). When their function is energized with ABC-type ATP-hydrolyzing subunits, these proteins are placed in the ABC superfamily of primary active transporters. However, there is little evidence for the association of many VUT family proteins with ABC-type ATP-hydrolyzing subunits, which leads to their placement in the secondary active, proton gradient utilizing, VUT family (TC # 2.A.88). Biotin and thiamin transporters have been shown to be capable of functioning as both ATP and proton driven systems (Hebbeln et al., 2007; Sun and Saier, unpublished data).

Comparing TMSs 1-3 of TSUP Csy1 (9 putative TMSs) with TMSs 3-6 of VUT Syba1 (6 putative TMSs) yielded a comparison score of 12.5 S.D. (Fig. 24). TMS 9 within Csy1 is likely the result of a gene fusion event. It is unclear how the VUT and TSUP families are related considering their 3 versus 4 TMS origins.

The Vacuolar Iron Transporter (VIT) Family (TC # 2.A.89)

Members of the VIT (DUF125) family are found in all domains of life, but they have been characterized only from plants and fungi (Kim et al., 2006). Their sizes typically range from 200 to 400 aas, and they generally possess between four and six TMSs, with a 5 TMS topology being the most common.

Comparing TMSs 2-8 of TSUP Kcr1 (8 putative TMSs) with TMSs 1-5 of VIT Suac2 (5 putative TMSs) yielded a comparison score of 12.3 S.D. (Fig. 25). Based on the alignment, where TSUP TMSs 4 and 8 do not align with Suac2, it is possible that the 6 TMS topology was generated from the duplication of a 3 TMS element containing TMSs 1-3 of TSUP. In the case of TSUP, the 4 TMS primordial peptide duplicated to give rise to the 8 TMS proteins.

The Choline Transporter-like (CTL) Family (TC # 2.A.92)

The CTL/solute carrier 44/XYPPX repeat family is represented only by two proteins in TCDB. Most proteins of this family are 600-700 aas in length and possess 8-12 putative TMSs. NCBI searches revealed that most, if not all, members of this family are restricted to eukaryotes, including humans and other vertebrates. It is unclear what the repeat unit within this family is.

Comparing TMSs 2-6 of TSUP Rsp1 (8 putative TMSs) with TMSs 2-6 of CTL Ppa3 (10 putative TMSs) yielded a comparison score of 12.5 S.D. (Fig. 26). This comparison, within the 6 TMS internal region of Ppa3, suggests that TMSs 1, 7 and 8 were lost to give rise to the inner 6 TMS portion of CTL proteins. It is possible that TMS 7 in Ppa3 represents and arose from TMS 7 of TSUP, but diverged in sequence and is no longer detectable.

The HlyC/CorC (HCC) Family of Putative Transporters (TC # 9.A.40)

The HCC family consists mostly of putative transporters for divalent cations including Ba^{2+}, Co^{2+}, Cu^{2+}, Fe^{2+}, Mg^{2+}, Mn^{2+} and Sr^{2+} (Goytain and Quamme, 2005). Hemolysin C, derived from *Brachyspira hyodysenteriae* is also a member of this family (ter Huurne et al., 1994). Putative transporters of this family have tremendous size (200 - 1000 aas) and topological variation (secreted and 1-7 TMSs).

Comparing TMSs 1-4 of TSUP Bav1 (8 putative TMSs) with TMSs 1-3 of HCC Ath6 (4 putative TMSs) yielded a comparison score of 11.8 S.D. (Fig. 27). This comparison suggests that TMS 2 of the 4 TMS TSUP primordial peptide may have been lost and another TMS gained at the C-terminal end to form the 3-4 TMS HCC topology. The region of Ath6 that aligned with TMS 2 of Bav1 exhibits limited identity to TMS 2, while the high number of acidic and basic amino acyl residues suggests that TMS 2 was possibly lost through non-conserved mutational means.

The Major Facilitator Superfamily (MFS; TC # 2.A.1)

The ancient and ubiquitous MFS consists of hundreds of thousands of sequenced members, which make up more than 70 currently recognized families

(Saier et al. 2005; 2009). Most proteins within the MFS are 400 to 600 amino acyl residues in length, possess 12 (usual), 14 (much less frequent) or 24 (rarely) TMSs, and catalyze solute:solute antiport, solute:cation symport or antiport, and uniport (Pao et al., 1998). The mechanisms by which proteins within the MFS function have been summarized (Law et al., 2008). Transported substrates vary tremendously and include ions, sugars, polyols, drugs, neurotransmitters, Krebs cycle metabolites, amino acids and many more. The 12 TMS topology arose from a 6 TMS intragenic duplication. The 14 TMS proteins probably arose by insertion of a cytoplasmic loop into the membrane between the two 6 TMS repeat units within the 12 TMS homologue. The 24 TMS topology is the product of a 12 TMS intragenic duplication. Despite these findings, the evolutionary origins of the 6 TMS precursor are still unclear. A 3 TMS duplication or a 2 TMS triplication have been proposed to be the most likely routes taken for the appearance of this 6 TMS unit (Heymann et al., 2001; Hirai et al., 2002; 2003; Huang et al., 2003).

Comparing TMSs 1-6 of TSUP Cup1 (8 putative TMSs) with TMSs 7-12 of MFS Gsp2 (12 putative TMSs) of the 2,4-diacetylphloroglucinol (PHL) Exporter (PHL-E) family (TC # 2.A.1.45) yielded a comparison score of 11.2 S.D. (Fig. 28a). Comparing TMSs 1-2 of PHL-E Bph5 (12 putative TMSs) with TMSs 3-4 of PHL-E Mci1 (12 putative TMSs) gave a comparison score of 9.2 S.D. (Fig. 28b). Although this comparison is not sufficient to prove a 2 TMS duplication event, it provides evidence for its occurrence. These comparisons lend support for a 2 TMS precursor of

the MFS, as well as the TSUP family (see CDF comparison). Several other families within the MFS aligned similarly to what has been described above.

The Resistance-Nodulation-Cell Division (RND) Superfamily (TC # 2.A.6)

The ubiquitous RND superfamily is divided into nine families, members of which catalyze the efflux of substrates such as heavy metals, drugs, lipooligosaccharides, lipids, sterols and various other substances from the cytoplasm or periplasm of a Gram-negative bacterial cell using the pmf as the energy source that drives transport (Tseng et al., 1999). Proteins of this superfamily range from 700 to 1300 amino acyl residues in length and generally possess 12 TMSs. The 12 TMS topology is made up of a single N-terminal TMS that is separated from a grouping of six TMSs by a large hydrophilic loop, which itself is separated from a group of five TMSs by a hydrophilic loop of substantial size. The repeat units each include 6 TMSs in a 1 + 5 arrangement separated by a large extracytoplasmic domain.

We were able to find substantial sequence similarity between the TSUP family and the (Gram-positive bacterial putative) Hydrophobe/Amphiphile Efflux-2 (HAE2) family (TC # 2.A.6.5) within the RND superfamily. Comparing TMSs 3-8 of TSUP Psp2 (9 putative TMSs) with TMSs 2-7 (six TMS grouping) of HAE2 Bli2 (12 putative TMSs) yielded a comparison score of 13.1 S.D. (Fig. 29a). Comparing TMSs 8-9 of HAE2 Cps2 (12 putative TMSs) with TMSs 11-12 of HAE2 Fsy2 (12 putative TMSs) gave a comparison score of 8.2 S.D. (Fig. 29b). As for the MFS, this comparison is not sufficient to prove a 2 TMS duplication event, but the high level of

similarity and identity it suggests that it had occurred. These comparisons suggest a 2 TMS origin for the RND superfamily and the TSUP family.

The Drug/Metabolite Transporter (DMT) Superfamily (TC # 2.A.7)

Members of the ubiquitous DMT superfamily fall into 26 presently recognized families, where each family tends to have a characteristic size, topological features and function (Jack et al., 2001). For the 26 families, the characteristic topologies are generally 4, 5 or 10 TMSs. These topologies are believed to have evolved from a 2 TMS primordial unit, which duplicated to give 4 TMSs, thus giving rise to several of the families within the DMT superfamily. The 5 TMS topology was generated thereafter via a C-terminal fusion and the 10 TMS topology arose from duplication of the 5 TMS unit (Lam et al., 2011).

Surprisingly, we were able to identify similarity between the TSUP family and the *Caenorhabditis elegans* ORF (CEO) family (TC # 2.A.7.8) within the DMT superfamily. None of the proteins within the small CEO family have been characterized. Comparing TMSs 1-4 of TSUP Dno1 (8 putative TMSs) with TMSs 1-4 of CEO Rco1 (9 putative TMSs) yielded a comparison score of 10.9 S.D. (Fig. 30). Although one large gap was introduced into the alignment, the percent identity and similarity observed is one of the highest of all comparisons. This comparison supports the data on the evolution of the DMT superfamily from a 2 TMS precursor, and additionally supports the 2 TMS origin of the TSUP family.

Chapter 6: Conserved Motifs in TSUP Homologues

The CLUSTAL X program did not reveal any fully conserved amino acyl residues (Thompson et al., 1997; Larkin et al., 2007). In order to identify motifs conserved among the 189 TSUP homologues included in the study, the MEME program of the MEME Suite: Motif-based sequence analysis tools was used (Bailey et al., 1994; 1998). The three motifs that were found to be the most conserved are presented in Figure 31. The motifs comprise a 56 residue area of conservation that mainly spans the first and second TMSs.

The best conserved motif is motif 1, which is 21 residues in length and has an accompanying statistical score of e^{-350} (Fig. 31a). Achiral glycine (G) residues at positions 1, 5, 9, 12, and 14-16 of motif 1 are the most well conserved residues. The proline (P), valine (V), and isoleucine (I) residues at positions 21, 20, and 13 respectively, are the next best conserved. Other residues like the hydrophobic alanine (A), leucine (L), and phenylalanine (F) residues and the hydrophilic serine (S) are interspersed between the highly conserved residues without any significant level of conservation. Localizing motif 1 to various proteins of the TSUP family revealed that motif 1 mainly spans TMS 1 and more specifically, the second half of TMS 1. The large number of Gs found within TMS 1 is intriguing, given that sterics often prevents Gs from participating in α-helices or β-sheets. The Rossmann fold (GxGxxG) is a consensus binding sequence for dinucleotides such as NAD/H/P/PH and FAD/H_2 (Iwaki et al., 2006). A "reverse" Rossmann fold sub-motif (GxxGxG) is observed in motif 1 (Iwaki et al., 2006). This is also interesting, given that the "reverse"

Rossmann fold may not always be restricted to the predicted area of TMS 1 and may constitute the loop region connecting TMSs 1 and 2. Thus, an inference that can be drawn from this observation is that in certain homologues, nicotinamide and/or flavin adenine dinucleotides may play a role in regulating transport.

Motif 2, the third best conserved motif, is 25 residues in length and has a statistical score of e^{-288} (Fig. 31b). The submotif A[VI][AG]TSL[AF][TM] (positions 2-9; 22-29 in motif 3) is highly conserved in motif 2 and overlaps the last 8 residues of motif 3. Residues 1-20 mainly comprise TMS 2, within which I, T, and S are well conserved at positions 10 (I), 13 (T), and 14, 16, 17 (S). Residue 21 was found to often mark the beginning of the loop region connecting TMSs 3 and 4 and had histidine (H) and tyrosine (Y) conserved, these two residues being about equally prevalent. Furthermore, a glycine residue is conserved well at position 25. On the whole, motif 2 describes TMS 2 and part of the subsequent loop. The first 9 residues of motif 2 correlate with the last 9 residues of motif 3.

Motif 3 is 29 residues in length, has an accompanying statistical score of e^{-342}, and is therefore the second best conserved motif (Fig. 31c). The [GA][IG]GGGL[IL][LT][VGL]P stretch (positions 1-10; 12-21 in motif 1) also shows up in motif 3. A highly conserved G and two less conserved Ls are found at positions 16, 12, and 13, respectively, which are often localized to the loop connecting TMSs 1 and 2. Apart from those three residues, the connecting loop is diverse in its residue make-up. The submotif A[SA][AG]T[SN]KA]F[MQ] is mainly localized to the first half of TMS 2 and is well conserved. Therefore, motif 3 is localized in between

motifs 1 and 2 and mainly describes the residues in the connecting loop region. The first 10 residues of motif 3 correlate with the last 10 residues of motif 1.

MEME identified one rather large area, but split it into 3 motifs based on statistical factors and program settings, which directed the program to identify motifs that were smaller than 50 residues, but larger than 6 residues in length. The 3 motifs correlate with each other well and highlight a great level of conservation of the TMS 1 and 2 sequences. The relative lack of conservation in the connecting loop between TMSs 1 and 2 implies that this region is under less strict evolutionary pressure when compared with the TMS regions. It may not contribute significantly to function.

Additionally, we observed that motifs 1 and 2, spanning the first and second TMSs, respectively, both appear twice in a number of proteins. In the proteins where they appear twice, like in Mch1, for example, they are found to also span TMSs 5 and 6, as expected if an intragenic duplication had occurred. Motif 3 only appears once in each protein analyzed, which further highlights the lack of conservation in the loop region. No evidence, such as that observed for a 4 TMS duplication event, was found for a 2 TMS peptide precursor duplicating to give a 4 TMS protein. It is possible that a 2 TMS duplication is undetectable because of the length of time that has passed since it occurred; coupled with decreased evolutionary pressure on the "supplementary" TMSs 3 and 4.

Chapter 7: Using Genome Context Analyses to Predict Functions

The small size, high gene density, intronless coding regions and simple, yet elegant operon organization of bacterial genomes allows for the making of fairly accurate functional predictions, such that future biochemical studies can be made to be more directed (Ochman and Davalos, 2006). In order to perform operon context analyses and to identify transcription factor regulons, the SEED database (Overbeek et al., 2005) along with RegPrecise and RegPredict (Novichkov et al., 2010a; 2010b) were used. SEED identified close homologues using the PSI-BLAST algorithm (Altschul et al., 1990; 1997). These analyses were applied to each cluster whenever possible. Our findings are summarized in Table 3.

Consistent with the topological ambiguities as well as the organismal and sequence diversity observed for TSUP homologues in **cluster 1**, only the 491 aa Ath3 from A. thaliana had its function predicted by the SEED database via slightly smaller (466-475 aas) close homologues found in *Mycoplasma penetrans*, *M. gallisepticum, Staphylococcus aureus*, and *Holdemania filiformis.* The Ath3 homologue was assigned the function of iron-sulfur (FeS) cluster assembly protein SufB. The SUF system, encoded by the suf operon (sufABCDSE), is one of the three FeS cluster assembly systems, with the other two being the iron-sulfur cluster (ISC) and nitrogen fixation (NIF) systems (Barras et al., 2005). FeS clusters serve as cofactors, mediating substrate binding and electron transfer. These systems become especially important during times of iron starvation or oxidative stress (Chahal, et al., 2009; Saini et al., 2010). SufA, which was absent in the organisms mentioned above, has been proposed

to be an iron chaperone and is essential for FeS cluster assembly under aerobic, but not anaerobic conditions (Wang et al., 2010). SufS is a cysteine desulfurase (EC # 2.8.1.7) and SufE is a scaffold protein. Surprisingly, biochemical studies have shown that SufB and the paralogous SufD, both of which are homologous to Ath3, function as part of a cytoplasmic complex along with SufC (Iwasaki, 2010). Although the Ath3 homologue may have gained this unique function, it is likely that Ath3 mediates the uptake of sulfur-based compounds (see cluster 5 analysis).

Similarly to cluster 1, SEED was unable to assign a function to the majority of **cluster 2** proteins. However, Rsp4, Pas1, Dno1, Par1, Cje1, Ade1, Tps5 and Gbe1 were assigned the function of putative membrane protein YfcA within several genomes. The E. coli YfcA protein posseses the TauE domain and is predicted to have 7 TMSs. A homologue of Rsp4 was identified to be part of an operon together with a gene encoding a phosphoserine phosphatase (EC # 3.1.3.3) in *Silibacter* sp. TM1040. Although not part of the same operon, genes encoding phosphoserine aminotransferase (EC # 2.6.1.52), D-3-phosphoglycerate dehydrogenase (EC # 1.1.1.95), serine/threonine protein phosphatase and L-threonine 3-dehydrogenase (EC # 1.1.1.103) surround the operon. Therefore, Rsp4 may function in transport of related substrates for glycine, serine and threonine synthesis, degradation or utilization. Pas1, from *Photorhabdus asymbiotica*, and its homologue from *Neisseria meningitidis*, are not part of an operon, but are surrounded by genes involved in the methycitrate cycle, acetyl-CoA generation and the propionate-CoA to succinate module. Dno1 is in an operon along with EngB, a GTP-binding protein, and is also

next to L-asparaginase (EC # 3.5.1.1) and a putative protease in *Dichelobacter nodosus*. The Dno1 homologue in *Pasteurella multocida* is part of a large cluster of genes encoding negative regulators of replication, a K^+ efflux pump, a murein endopeptidase, chorismate synthase (EC # 4.2.3.5) and the lipid A biosynthesis acyltransferase (EC # 2.3.1.-). 2 adjacent genes encode the gamma and tau subunits of DNA polymerase III and are involved in purine conversions and cAMP signaling in bacteria. In *Marinomonas* sp. MWYL1, genes involved in glycine and serine utilization and post-uptake glycerolipid metabolism surround the Dno1 homologue. Par1 from *Psychrobacter* sp. 273-4 is surrounded by a phosphate transporter of the NhaA Na^+:H^+ family (TC # 2.A.33), a glutamate symporter of the Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) family (TC # 2.A.23) and the butyryl-CoA dehydrogenase (EC # 1.3.99.2) involved in Ile/Val degradation, Lys fermentation and several other metabolic processes. Its homologue in *Mannheimia succiniciproducens* is part of an operon along with a murein endopeptidase, in an arrangement very similar to that which is observed for the Dno1 homologue in *P. multocida*.

Ade1 from *Anaeromyxobacter dehalogenans* is not localized to an operon, but is closely surrounded by genes encoding a response regulator receiver protein and the 4-hydroxybutarate coenzyme A transferase. The Ade1 homologue in A. sp. Fw109-5 is part of an operon with the above two genes. The function of Ade1 may overlap with that of Par1 and the presence of a response regulator receiver protein, possibly involved in bacterial cellular responses to environmental signals, suggests that Ade1 is

part of a global response, possibly to nitrogen deficiency and metabolism (Bent et al., 2004). Furthermore, the function of Ade1 may overlap with that of Dha1 from cluster 15, considering the nitrite/nitrate directed genomic similarities between Dha1 and the Ade1 homologue in *Streptomyces coelicolor* A3(2) (see cluster 15).

Cje1 from *Campylobacter jejuni* is localized to an operon encoding the flagellar P-ring protein FlgI, flagellar hook-associated protein FlgK and several other hypothetical proteins likely to be involved in flagellum structure and/or synthesis. A divergently transcribed operon encoding proteins necessary for the formation of the type II protein secretion system lies next to the flagellum operon. Cje1 may localize to the bacterial flagellum where its role is not clear, or facilitate protein secretion through an unknown transport function (Cianciotto, 2005).

Thermoanaerobacter pseudethanolicus-derived Tps5 and its homologue in T. sp. X514 are surrounded by genes involved in mannose/mannitol metabolism and utilization, as well as purine/pyrimidine conversions. Based on the genomic context of Tps5 and its homologues, an educated functional prediction cannot be made. The same holds true for Gbe1 from *Granulibacter bethesdensis* and its homologues.

Psp4, Taf1 and Pac1 are the three proteins in cluster 2 which are homologous to a putative membrane protein, YfcA. Psp4 from *Psychromonas* sp. CNPT3 is part of an operon along with genes encoding an ATP-dependent helicase DinG/Rad3 involved in DNA repair, primosomal replication protein N prime prime involved in DNA replication. Several genes involved in periplasm-localized nitrite/nitrate ammonification and a gene encoding the molybdenum cofactor biosynthesis protein

MoaA surround the above operon. The homologues in *Aliivibrio salmonicida* and *Vibrio fischeri* ES114 retain the same operon arrangement, but the MoaA and nitrite/nitrate related genes are replaced with 1 to 2 copies of an outer membrane protein with sequence similarity to the General Bacterial Porin (GBP) family (TC # 1.B.1). The porins may function as a system with Psp4.

Pac1 from *Propionibacterium acnes* is surrounded by genes encoding the translation initiator factor 2 and related proteins, a putative phosphodiesterase and the prolyl-tRNA synthetase (EC # 6.1.1.15). The Pac1 homologue in *Brevibacterium linens* BL2 is in an operon with the synthetase, suggestive of a possible proline transport function. The function of Taf1 from *Thermosipho africanus* is unclear.

The presence of enzymes involved in protein, amino acid and nitrite/nirate turnover suggests that cluster 2 proteins may function as transporters of nitrogen-containing compounds such as those mentioned above.

Mka1 from *M. kandleri* in **cluster 3** is part of an operon along with a predicted nucleotide-binding protein related to the universal stress protein UspA, which is upregulated by metabolic, oxidative and temperature stresses (Liu et al., 2007). Genes encoding a PP superfamily ATPase similar to lysidine synthase and a predicted permease with sequence similarity to TSUP TC # 9.A.29.4.1 are found neighboring the operon. Similarly, genes involved with oxidative stress surround the Mka1 homologue from *Thermococcus kodakarensis*. The Mka1 homologues in *Pyrococcus furiosus* and *P. horikoshii* are found in operons that appear to function in protein degradation, possibly suggesting an amino acid transport role, and/or representing a

part of the stress response. The large Dde1 protein from *Desulfovibrio desulfurican* is part of an operon with a single hypothetical protein. The Dde1 homologues in *D. vulgaris* (3 total- str. 'Miyazaki F'; subsp. vulgaris str. Hildenborough; subsp. vulgaris DP4) and *D. baculatum* are in operons with the same hypothetical protein, but a sigma-54 (σ^{54}) dependent transcriptional regulator is encoded adjacently and transcribed divergently from the complementary strand. It has been shown that σ^{54} plays a global regulatory role for genes encoding proteins involved in nitrogen metabolism, transport, stress responses, carbon metabolism and cell motility (Zhao et al., 2005). To study this further, we used RegPrecise, which identified the RpoN transcription factor, σ^{54}, family as the regulator of Dde1 transcription. RpoN recognizes the TGGCACGxxxxTTGCT motif. RegPrecise predicted that Dde1 is part of an operon along with four more genes encoding two histidine kinases and two response regulators. Based on SEED, the four genes did not appear to be part of the same operon because of the large distance between them.

Homologues of the *Roseiflexus castenholzi* Rca1 protein are found in operons that are divergently transcribed from a gene encoding a putative efflux pump in the Arsenical Resistance-3 (ACR3) family (TC # 2.A.59). Closely upstream of the operon, a single permease gene was found, which showed significant sequence similarity to proteins of the putative permease Duf318 (Duf318) family (TC # 9.B.28). The Duf318 family exhibits 2-fold symmetry within an 8 to 10 TMS topology and has been implicated in arsenate/arsenite resistance (Wang et al., 2009b). Consistent with this finding, closely upstream of the Duf318 homologue gene, a redox-active disulfide

protein-encoding gene and a gene encoding an ArsR transcriptional regulator are found within a single operon. ArsR homologues are known to regulate many transporters, and in addition to its likely role in regulating the Duf318 transporter, it may also regulate Rca1 (Castillo and Saier, 2010). The two transporters may transport arsenate/arsenite, or other stress-related substrates, providing further control of stress. Also of note, the degP_htrA_DO domain present in Orf6, which has been suggested to play a role in mediating stress responses and possess endopeptidase activity, further supports our claim that proteins of cluster 3 are involved in stress responses.

In **cluster 4**, Iho1 from *Ignicoccus hospitalis* is not part of an operon and is surrounded by genes encoding the rRNA biogenesis protein Nop5/Nop56, a tRNA/RNA cytosine-C5-methylase (EC # 2.1.1.-) and an NADH dehydrogenase (EC # 1.6.99.3), one of several of the respiratory dehydrogenases. Iho1 homologues in *P. furiosus*, *P. abyssi* and *P. horikoshii* appear in operons encoding polycistronic products coding for a D-isomer specific 2-hydroxyacid dehydrogenase, a deblocking aminopeptidase (EC # 3.4.11.-), a hypothetical protein and a dephospho-CoA kinase (EC # 2.7.1.24) involved in coenzyme A biosynthesis. In these organisms, close and divergently transcribed genes include (1) a GTP-binding and nucleic acid-binding protein YchF, (2) a periplasmic binding protein component of a multicomponent ATP-Binding Cassette (ABC) uptake system with sequence similarity to the Manganese/Zinc/Iron Chelate Uptake Transporter (MZT) family (TC # 3.A.1.15), (3) a molybdenum cofactor biosynthesis protein MoaB, as well as (4) a lactase and (5) β-galactosidase (EC # 3.2.1.23). The archaeal MoaB, bacterial MogA and eukaryotic

Cnx1 molybdenum cofactors are an integral part of oxidoreductase enzymes and often feed into the respiratory chain as electron carriers (Bevers et al., 2008). Although a pattern is not apparent, it is possible that Iho1 is involved in cofactor synthesis for enzymes and substrates involved in the respiratory chain.

SEED identified Min1 from *Methylacidiphilum infernorum* to be in an operon with DNA polymerase IV (EC # 2.7.7.7) and to be surrounded by hypothetical proteins. Min1 homologues in two *Thermus thermophilus* strains were assigned the function of putative sulfate permease and are located in operons along with genes encoding a ferredoxin-sulfite reductase (EC # 1.8.7.1), a sulfate adenylyltransferase (EC # 2.7.7.4) involved in inorganic sulfur assimilation, a phosphoadenylyl-sulfate reductase (EC #1.8.4.8)/adenylyl-sulfate reductase (EC # 1.8.4.10) involved in cysteine biosynthesis, and the uroporphyrinogen-III synthase (EC # 4.2.1.75), siroheme synthase/precorrin-2 oxidase (EC # 1.3.1.76)/sirohydrochlorin ferrochelatase (EC # 4.99.1.4) and uroporphyrinogen-III methyltransferase (EC # 2.1.1.107) genes involved in heme, siroheme and vitamin B12 biosynthesis. Located adjacently are genes encoding a quinone oxidoreductase. Min1 and its homologues lend further evidence for a cofactor synthesis role.

The last cluster 4 protein to be identified was Cbu3 from *Coxiella burnetii*, which is not part of an operon and is surround mostly by hypothetical proteins. Its homologue in *Celivibrio japonicus* is surrounded by genes encoding a ferrochelatase (EC # 4.99.1.1) and diguanylate cyclase/phosphodiesterase, both of which are implicated in heme, siroheme and bacterial hemoglobin synthesis. The presence of the

latter gene suggests a possible role in biofilm formation. Hemoglobin and hemoglobin-like structures are nearly ubiquitous, and in addition to their oxygen binding role, they have been shown to exhibit novel functions such as conferring protection against sulfide, maintaining acid-base balance, and possessing oxidase and peroxidase-like as well as superoxide dismutase activities (Wever and Vinogradov, 2001). The Cbu3 orthologue in *Vibrio splendidus* is found in an operon with two genes, one of which is known to be involved in nitrogen regulation. An adjacent operon contains genes encoding a prolipoprotein diacylglyceryl transferase (EC # 2.4.99.-) and a thymidylate synthase (EC # 2.1.1.45), which are involved in lipoprotein biosynthesis or folate biosynthesis/pyrimidine conversions, respectively.

Bsp1 contains a USP-like domain associated with stress responses. Based on the results of the SEED analysis, we propose that cluster 4 proteins are involved in the synthesis of metal-containing cofactors and heme groups. Cluster 4 proteins may contribute to cofactor synthesis by transporting sulfate, which could be a requirement for the process. The prevalence of oxidoreductase enzymes within cluster 4 gene clusters does not discount an oxidative stress response role (Lumppio et al., 2001).

Proteins of **cluster 5** appear to exhibit great functional diversity. The function of the first of two Sfu1 paralogues from *Syntrophobacter fumaroxidans* is unclear as its gene is surrounded by hypothetical proteins. The second paralogue is located adjacent to 3 genes involved in cobalt/zinc/cadmium resistance with sequence similarity to members of the Resistance-Nodulation-Cell Division (RND) superfamily (TC # 2.A.6). It is possible that the second Sfu1 paralogue may contribute to the

maintenance of the membrane voltage potential by extruding anions or mediating the uptake of cations during $Co^{2+}/Zn^{2+}/Cd^{2+}$-induced stress. The homologue in *Meiothermus ruber* is surrounded on one side by an operon encoding an acyltransferase and peptidase M19, and on the other side by an ATP-dependent protease La Type I (EC # 3.4.21.53), suggestive of a possible amino acid or peptide transport role. However, the homologue in *Meiothermus silvanus* is located near a gene encoding a scaffold protein for [4Fe-4S] cluster assembly, which makes transport of sulfur-based compounds a likely functional possibility.

Trichodesmium erythraeum-derived Ter2, like Sfu1, has a paralogue appearing to serve a different role, but neither it nor its homologues are found in operons. Ter2 is surrounded by genes involved in carbohydrate and RNA metabolism; clearly a broad range of functions from which a conclusion is difficult to draw. The Ter2 paralogue may be involved in fatty acid biosynthesis, as it is surrounded by a lipoprotein and an operon containing genes encoding 3-oxoacyl-[acyl-carrier-protein] synthase KASII (EC # 2.3.1.41) and cyclophilin-type peptidyl-prolyl cis-trans isomerase. The same is true for the *Crocosphaera watsonii* homologue, which is surrounded by a gene encoding a malonyl-CoA-acyl carrier protein transacylase (EC # 2.3.1.39) involved in fatty acid biosynthesis. Yet another Ter2 homologue in *Nostoc punctiforme* is in an operon coding for a cysteine desulfurase and the sulfur oxidation molybdopterin C protein. Thus, the Ter2 homologue in *N. punctiforme*, located next to the same cysteine desulfurase that is part of the suf operon (see cluster 1-Ath3), likely takes up sulfur-based compounds.

Ssp1 from *Synechococcus* sp. JA-3-3Ab and its homologue in *S.* sp. JA-2-3B' are not part of operons and are surrounded by genes encoding prephanate dehydratase (EC # 4.2.1.51), alanine racemase (EC # 5.1.1.1), a glycosyl transferase, and an NAD-binding oxidoreductase. Therefore, it is likely that Ssp1 and its homologue are amino acid or peptide transporters, with glycosyl transferase mediating O or N-linked glycosylation, which in bacteria is known to play a role in adhesion, protection against proteolysis and evasion of the host immune system (Faridmoayer et al., 2007). The NAD-binding oxidoreductase suggests a possible stress response role for Ssp1, with glycosylation being a coordinated response to stress. Ssp1 homologues in *Cyanothece* sp. CCY 0110 and ATCC 511 are located near cell division protein FtsH (EC # 3.4.24.-), which may mediate cell division-ribosomal stress.

The last protein to be identified by the SEED database is Pca1 from *Pyrobaculum calidifontis*. It is not operon-localized, but is rather positioned adjacently to a divergently transcribed preprotein translocase SecG subunit. *E. coli* SecG functions along with SecY, E, D, F and sometimes A, to form the protein excreting translocase complex (Dalbey and Chen, 2004). Previous studies have shown that SecG deficiency leads to a slight decrease in the ability of E. coli to export proteins, but this effect is much more pronounced when the cells are compromised or undergoing stress (Flower et al., 2000; Palomino and Mellado, 2008). Located more distantly are genes encoding a multidrug-efflux transporter with significant sequence similarity to members of the Drug:H^+ Antiporter (12 Spanner) (DHA1) family (TC # 2.A.1.2) of the MFS, a protein containing a UspA domain and a conserved signal

transduction protein with 2 regulatory CBS domains (Tuominen et al., 2010). All of the aforementioned genes are likely to be upregulated in response to stress, and the SecG subunit may play a role in regulating Pca1 activity during stress. It is likely that Pca1 transports amino acids or peptides.

Cluster 5 proteins are likely to function as transporters of sulfur-based compounds, amino acids/peptides, fat and other related compounds. Once again, a common pattern observed is the likely involvement of TSUP proteins in the stress response. Surprisingly, Ter2 and Ssp1 cluster closely together in Figure 1, but their functions are likely to be drastically different, further highlighting the great size and sequence diversity within the TSUP family.

All proteins except Ooe1, Sth1 and Dac1 of **cluster 6** were identified. Only proteins for which reliable functional predictions can be made will be discussed. Sus1, Sth2, Bsu1, Mth2 and Kcr1, or their homologues appear in operons along with transcription regulators of the GntR superfamily (Lee et al., 2003; Rigali et al., 2002; 2004; Vindal et al., 2007). Members of the GntR family respond to metabolite effector molecules and control genes involved in responding to various external stimuli (Hillerich and Westpheling, 2006; Hoskisson et al., 2006). Unlike its homologue in *Chitinophaga pinensis*, Sus1 from *Solibacter usitatus* is not part of an operon with a GntR transcription regulator, but is instead in an operon with two hypothetical proteins and rhodanase (thiosulfate sulfurtransferase; EC # 2.8.1.1; see cluster 15). Bsu1 from *Brucella suis* is in an operon with genes encoding GntR and a hypothetical protein. Bsu1 is also surrounded by operons encoding ABC transporters

of glycerol-3-phosphate and branched chain amino acids with sequence similarities to proteins of the Carbohydrate Uptake Transporter (CUT1) and the Hydrophobic Amino Acid Uptake Transporter (HAAT) families, respectively (TC #s 3.A.1.1 and 3.A.1.4). Kcr1 is part of an operon along with genes encoding a hypothetical protein and a protein of the DHA1 family. Proteins of cluster 6 that are usually found in operons with the GntR transcriptional regulator may have divergent functions. The remaining 6 proteins that did not appear in operons with GntR transcritptional regulators were either in operons with hypothetical proteins, not in operons or their genomic context and poor conservation of adjacent genes made functional predictions impossible. Cluster 6 proteins appear to have diverse functions based on scarce genomic context data, but this scarcity does not preclude a related function.

The only protein from **cluster 7** to be in SEED was Bad1 from *Bifidobacterium adolescentis* ATCC 15703. Bad1 and its homologue in *B. adolescentis* are not found in operons and have almost identical gene arrangements surrounding them. Both contain operons encoding genes for a collagen adhesin precursor and the LPXTG motif-specific sortase A enzyme. Sortase links proteins to the microbial envelope, while adhesins/invasins allow for bacterial adhesion to host cells (Marraffini et al., 2006; Pizarro-Cerdá and Cossart, 2006). Both proteins are involved in bacterial virulence and interactions with the host immune system, which are crucial for immune tolerance, their persistence in the human microflora and maturation of acquired immunity in humans when expanded to the whole of the bacterial flora (Karlsson et al., 2004). An alkaline phosphatase (EC # 3.1.3.1),

probably involved in phosphate metabolism, also surrounds Bad1, along with a likely operon containing the chaperone protein DnaK, heat shock protein GrpE, chaperone protein DnaJ and a transcriptional repressor of the DnaK operon, HspR. These proteins are upregulated in response to heat stress, allowing for bacterial survival at typically lethal temperatures (Schmidt and Zink, 2000). Bad1 may play a role in the heat stress response and may transport phosphate or amino acids/peptides to regulate bacterial virulence.

The majority of members of the mainly proteobacterial **cluster 8** has had their functions predicted. Afe1 from *Acidithiobacillus ferrooxidans* and Hne1 from *Hyphomonas neptunium* are found surrounded by or in an operon, respectively, with a gene encoding the outer-membrane TonB-dependent receptor (Noinaj et al., 2010). The closest homologue of the Ton-B dependent receptor in TCDB, FoxA(1) (TC # 1.B.14.1.12), is a receptor for ferrioxamine/desferrioxamine (Wei et al., 2007). In Hne1, a gene encoding the histidinol-phosphate aminotransferase (EC # 2.6.1.9), involved in histidine biosynthesis, is also present. Interestingly, some members like Afe1 and Hne1 of the TSUP family may function along with the TonB receptor and periplasmic binding proteins in a possible combined transport role, translocating a substrate across both membranes of the Gram-negative bacterial envelope.

Rru1 from *Rhodospirillum rubrum*, Rsp1 from *Ruegeria* sp. TM1040 (via a close homologue in *Silibacter* sp. TM1040) and Msp3 from *Magnetococcus* sp. MC-1, are found in operons or closely surrounded by genes encoding a γ-glutamyltranspeptidase (EC # 2.3.2.2) involved in glutathione and poly-gamma-

glutamate biosynthesis as well as the utilization of glutathione as a sulfur source. A GntR transcriptional regulator is found in an operon with Rsp1. A homologue of Rru1 in *Celivibrio japonicus* may also function with the TonB-dependent receptor given the presence of a TPR domain protein within the operon, a putative component of the TonB system (Galigniana et al., 2010). These three proteins and their homologues may supply sulfur to supplant the glutathione utilization pathway.

Apart from being localized next the GntR transcricriptional regulator, the genomic context of Mca1 from *Methylococcus capsulatus* and Har1 from *Herminiimonas arsenicoxydans* does not allow us to propose a function with great confidence. While the function of Swo1 from *Shewanella woodyi* is unclear, its homologue in Sulfurospirillum deleyianum is found in an operon with genes encoding a putative periplasmic protein and the Smf/DprA protein, which plays a role in natural bacterial transformation (Mortier-Barrière et al., 2007; Tadesse and Graumann, 2007). Thus, Swo1 may be involved in nucleic acid uptake.

Nmo1, Ama2, Msu1, Pne1, Rso1 and Iba1 may function in lipid or lipoprotein transport. All proteins except Pne1 are found in operons with genes encoding the prolipoprotein diacylglyceryl transferase and thymidylate synthase (see cluster 4). Pne1 is not part of an operon; however, its paralogue is located next to a divergently transcribed operon encoding the long-chain-fatty-acyl-CoA ligase (EC # 6.2.1.3) involved in fatty acid metabolism and various subunits of the Tripartite ATP-independent Periplasmic Transporter (TRAP-T) family (TC # 2.A.56; Pernil et al., 2010). Possibly this TRAP transporter takes up fatty acids. The closest homologues

of Msu1 identified by the SEED are in *Haemophilus influenzae* strains R2846, PittEE, 86-028NP and PittGG where they are in operons with adenosine (5')-pentaphospho-(5'') pyrophosphohydrolase (EC # 3.6.1._) and a tRNA-specific adenosine-34 deaminase (EC # 3.5.4._) along with the two previously mentioned genes.

The 2 proteins comprising **clusters 9** and **10** were not identified by the SEED database. In **cluster 11**, Mma2 from the *Methanococcus maripaludis* S2 strain, as well as its homologues in the C5, C6 and C7 strains, and in *M. vannieli*, are located in a dense cluster of genes that is conserved across all strains and species. Mma2 is in an operon with genes encoding FMN adenylyltransferase (EC # 2.7.7.2) and a putative membrane protein with sequence similarity to members of the Autoinducer-2 Exporter (AI-2E) family (TC # 2.A.86). A nearby operon includes genes coding for the hydrogenase Ehb protein P, 2-haloalkanoic acid dehalogenase (EC # 3.8.1.2) and cobalamin synthase involved in cobalamin/vitamin B_{12} biosynthesis. Genes encoding an enzyme involved in coenzyme F_{420} synthesis and several tRNA modification proteins also surround Mma2. Based on operon context, Mma2 most likely transports riboflavin, seeing as FMN, FAD and coenzyme F_{420} all contain flavin derivatives within their structures. Additional roles may include the transport of substrates necessary for vitamin B_{12} synthesis and tRNA amino-acylation; possibly Co^{2+} or amino acids, respectively.

Mma1 from M. mazei maps near an ABC transporter and an amino acid permease. The ABC transporter shows sequence similarity to members of the Lipoprotein Translocase (LPT) family (TC # 3.A.1.125), which extrude lipoproteins

(Taniguchi and Tokuda, 2008). The amino acid permease showed sequence similarity to several amino acid efflux families within the Amino Acid-Polyamine-Organocation (APC) superfamily (TC # 2.A.3). It is possible that Mma1 mediates the uptake of amino acids in coordination with the members of the APC superfamily. The Mma1 homologue in *M. maripaludis* C7, possibly a paralogue of Mma2, is found in an operon encoding a Ni^{2+} insertion protein, further supporting the possible metal transport role suggested for Mma2.

Tko1 of *T. kodakarensis* within **cluster 12** is part of an operon along with a gene encoding glycyl-tRNA synthetase (EC # 6.1.1.14) involved in tRNA aminoacylation of glycine. Therefore, Tko1 may function as a glycine uptake protein. Although found in separate phylogenetic clusters, the 7 putative TMS Tko1 and Iho1 (cluster 4) likely share similar functions, as Tko1 brought up the same homologues as Iho1. Both may in fact transport a substrate to allow for tRNA modification.

Mbo1 from *Candidatus Methanoregula boonei* is not found in an operon, but is surrounded by a sensory box histidine kinase regulator, cysteine dusulfurase, phosphohistidine phosphatase sixA (EC # 3.1.3.-) and an amidase involved in NAD and NADP synthesis. Possible functions of Mbo1 include cysteine, histidine, sulfur-based compound or NAD component transport.

The function of Sma2 from *Staphylothermus marinus* is more ambiguous. Sma2 is next to the divergently transcribed γ-glutamyltranspeptidase. Sma2 may contribute by transporting supplementary sulfur-based compounds.

It is highly likely that Cbe1 from *Clostridium beijerincki* in **cluster 13** functions in sulfite uptake. Although not part of an operon, it is near genes encoding an iron-sulfur-binding protein, the dissimilatory sulfite reductase (desulfoviridin) and the CoA-disulfide reductase (EC # 1.8.1.14). This suggests that it also functions in sulfur metabolism. Its homologue in *Blastopirellula marina*, located next to the divergently transcribed threonine dehydratase (EC # 4.3.1.19) gene, is likely to function in branched-chain amino acid synthesis.

Nma1 from *Nitrosopumilus maritimus* is not part of an operon, but is part of a dense gene cluster containing the iron-dependent repressor IdeR of the DtxR family. The main regulator of iron acquisition and metabolism is the ferric uptake repressor (Fur) of *E. coli* and its homologues (Touati, 2000). Deregulation of iron metabolism or superoxide dismutase deficiency can favor the Fenton reaction, which can contribute to oxidative stress, DNA damage, spontaneous mutagenesis and sensitivity to H_2O_2 (Jittawuttipoka et al., 2010). Similar observations have been made for IdeR (Rodriguez et al., 2002). Based on genomic context, Nma1 may function as an iron uptake transporter and be regulated by IdeR. In such a case, the overall contribution to iron homeostasis by Nma1 should be minimal when compared to the Fur protein. The Nma1 homologue in *Thermococcus onnurineus*, localized to an operon coding for an aminopeptidase and a dehydrogenase, may be an amino acid/peptide uptake permease.

Gka1, Bcl1, Sau1 and Oih1 are likely to be involved in purine and pyrimidine conversions, but the transported substrates cannot be inferred. However, homologues of Oih1 found in *Bacillus cereus* and *B. anthracis* are located adjacent to a two-gene

operon encoding a hypothetical protein and a low-affinity inorganic phosphate transporter with sequence similarity to members of the Inorganic Phosphate Transporter (PiT) family (TC # 2.A.20), which mediate the uptake of phosphate and/or sulfate (Mansilla and Mendoza, 2000). Extrusion of phosphate and/or sulfate may be the function for Oih1, and possibly also for Gka1, Bcl1 and Sau1.

The genomic context of **cluster 14** proteins provides few clues as to their functions. Bja1 from *Bradyrhizobium japonicum* and its homologues in *Rhodopseudomonas palustris*, *B.* sp. Bi and *B.* sp. BTAi1 are all found in an operon along with a hypothetical protein, and located adjacent to the operon, is a divergently transcribed drug efflux pump of the 10 TMS Drug/Metabolite Exporter (DME) family (TC # 2.A.7.3) within the Drug/Metabolite Transporter (DMT) superfamily (TC # 2.A.7). In *B.* sp. BTAi1, two additional operon-localized efflux pumps with significant sequence similarity to drug and heavy metal efflux exporters of the Heavy Metal Efflux (HME) and the Hydrophobe/Amphiphile Efflux-1 (HAE1) families (TC #s 2.A.6.1, 2.A.6.2) within the RND superfamily (TC # 2.A.6) are located nearby. It may be that Bja1 is also involved in drug or heavy metal or toxic ion efflux. The toxic cyanide or cyanate anions may be transport substrates given the presence of carbonic anhydrase (EC # 4.2.1.1), which is part of the cyanate hydrolysis subsystem (Anderson et al., 1990; Ford, 1971).

Pth1 from *Pelotomaculum thermopropionicum* appears in an operon with 2 hypothetical proteins and is accompanied by an adjacent transporter with greatest similarity to Sulfate/Tungstate Uptake Transporters (SulT; TC # 3.A.1.6) within the

ABC superfamily. Accordingly, Pth1 may extrude sulfate, tungstate, or vanadate, most likely sulfate as the primary substrate. The Pth1 homologue in *D. vulgaris* str. 'Miyazaki F', located at a distinct position from the *D. vulgaris* str. 'Miyazaki F' Dde1 homologue of cluster 3 (Pth1 paralogue), lacks the ABC porter, but contains two adjacent copies of the σ^{54} dependent transcriptional regulator, which suggests similar roles as the Dde1 homologues in cluster 3.

Homologues in *Magnetospirillum magnetotacticum* and *M. magneticum* are located in operons along with genes encoding a hypothetical protein and UspA, and next to an operon encoding the NifU and Bcl-2-associated X (BAX) protein. The NifU scaffold protein is part of the NIF system *(*Barras et al., 2005; see cluster 1) of FeS cluster assembly and is known to colocalize with the Fe-S center-containing rubrerythrin, a peroxidase involved in hydroperoxide detoxification (Lumppio et al., 2001; Maralikova et al., 2010). The BAX protein is pro-apoptotic in mammalian cells and when expressed in bacterial cells, it induces apoptosis as well. A previous study had demonstrated that E. coli BAX-resistant mutants, when exposed to a superoxide generating antibiotic, survived due to a unique anti-oxidant pathway employing the non-catalytic N-terminal end of the RNase E protein *(*Nanbu-Wakao et al., 2000*)*. NifS, a cysteine desulfurase, which does not localize to the same operon, supplies inorganic sulfide needed for Fe-S cluster formation. Homologues of Pth1 may serve as means for the uptake of inorganic sulfide (Kiyasu et al., 2000; Yuvaniyama et al., 2000).

At least in certain organisms, like the various *Desulforudis vulgaris* subspecies and strains, TSUP family members are present in more than one copy. While some paralogues appear to serve similar functions, we have identified instances where their functions are likely to be divergent.

Rme1 from *Ralstonia metallidurans* in **cluster 15** was not identified in SEED, but its close homologue in *Cupriavidus metallidurans* and several others were. In *C. metallidurans*, the Rme1 homologue does not appear to be part of an operon, but is likely to be co-regulated with the two operons closely surrounding it. The first operon is very large and encodes proteins and enzymes involved in cytochrome c biogenesis. It is particularly enriched in genes encoding enzymes having sulfur-based compounds as their target substrate, including thioredoxin, the thiol:disulfide interchange protein and a protein-disulfide reductase (EC # 1.8.1.8; Missiakas et al., 1995). The second operon encodes a hypothetical protein, a LysR family transcriptional regulator and a dihydrolipoamide dehydrogenase (EC # 1.8.1.4), which is part of the leucine degradation, HMG-CoA metabolism, glycine cleavage, TCA cycle, photorespiration and the 5-formyltetrahydrofolate cycloligase-like (5-FCL) protein subsystems (Roje et al., 2002). Homologues found in *Xanthomonas campestris* and *Acidovorax* sp. JS42 occur in an operon with a hypothetical protein, but the overall gene arrangement is retained. Rme1 and its homologues are likely to transport sulfur-based compounds and play a role in metabolic pathways.

The *Sulfolobus solfataricus*-derived Sso1 is not part of an operon, and its genomic context is not conducive for making functional predictions. Its homologue in

S. acidocaldarius, however, is located closely upstream of an operon encoding the various subunits of the CoB-CoM heterodisulfide reductase (EC # 1.8.98.1), further supporting a possible sulfur-based compound transport role.

Aau1 from *Arthrobacter aurescens* is part of an operon with a gene encoding a protein containing a rhodanase-like domain. It is surrounded by a several other smaller proteins containing a rhodanase-like domain as well as a thioredoxin. Rhodanase catalyzes the transfer of a sulfur atom from sulfane sulfur-containing compounds (sulfur atoms at oxidation state 0 or -1) to sulfur acceptors like cyanide and thiols in order to generate molecules that are less toxic to the cell (Wróbel et al., 2009). An example of a reaction which rhodanase catalyzes is as follows:

Thiosulfate + Cyanide $\xrightarrow{\text{Rhodanase}}$ Sulfite + Thiocyanate

Aau1 as well as its homologues in *Corynebacterium glutamicum*, *C. efficiens*, *Salinispora tropica* and *Mycobacterium* sp. JLS all contain a gene encoding hydroxyacylglutathione hydrolase (EC # 3.1.2.6), which may also serve as a polysulfide binding protein. In *C. efficiens* and *S. tropica*, hydroxyacylglutathione hydrolase may be in an operon along with the Aau1 homologue. However, these organisms encoding the Aau1 homologues lack the smaller rhodanase-like domain proteins.

Bsp2 from *Bacillus* B-14905 and its homologues in *B. cereus*, *Geobacillus kaustophilus* and *Exiguobacterium sibiricum* have essentially the same gene arrangement as Aau1 and its respective homologues, with various rhodanase-like domain proteins and the hydroxyacylglutathione hydrolase joining them in operons.

In addition, Bsp2 and its homologues are in an operon with, or are surrounded by a putative sulfide reductase, protein disulfide isomerase (S-S rearrangase; EC # 5.3.4.1) and/or a putative pyridine nucleotide-disulfide oxidoreductase. Aau1 and Bsp2, as well as their respective homologues, may therefore function in sulfur-based compound uptake, or more likely, sulfite export, consistent with the function of TauE (TC # 9.A.29.2.1).

Dha1 from *Desulfitobacterium hafniense* and its homologues in *D.* sp. Y51 and *Desolfotomaculum reducens* is part of an operon along with a gene encoding the protein chain release factor A. Located adjacently is another operon encoding the 4Fe-4S and NAD(P)H subunits of nitrite reductase ((EC # 1.7.1.4) involved in nitrate and nitrite ammonification), formate dehydrogenase H (EC # 1.2.1.2) and the anaerobic dimethyl sulfoxide reductase chain B (EC # 1.8.99.-). It is possible that Dha1 functions as a nitrite/nitrate/formate transporter, similar to characterized members of the Formate-Nitrite Transporter (FNT) family (TC # 2.A.44), although the FNT and TSUP families do not appear to be homologous.

Discussion

Members of the ubiquitous TSUP family appear to function as secondary carriers for sulfur-based compounds. We have characterized the TSUP family structurally, functionally and evolutionarily. Our analyses led to establishment of the novel Microbial Rhodopsin (MR) superfamily, the 21st superfamily to be included in TCDB (Saier et al. 2006; 2009). We have also presented evidence for an ancestral 2 α-helical hairpin structure that may have given rise to several families of integral membrane transport proteins. An ancestral ββ hairpin has been proposed to be the precursor of all outer membrane β-barrels in Gram-negative bacteria, mitochondria and plastids, a suggestion that partially parallels our own (Remmert et al., 2010).

The vast majority of homologues within the TSUP family possess eight putative TMSs, with some predicted to have seven or nine TMSs, possibly as a result of N- or C-terminal insertions or deletions. Conserved motifs were identified, and their presence in multiple copies within TSUP homologues support two-fold symmetry within the proteins (Fig. 31a-c).

The greatest topological variation was observed in the sequence and source organism diverse cluster 1, which consists solely of eukaryotic members that are 40-50% larger than their prokaryotic counterparts (Chung et al., 2001). However, large homologues were also identified in prokaryotic clusters. These may have been the products of gene fusion events where hydrophilic domains were introduced during their evolutionary histories. Most hydrophilic domains proved to be non-homologous to anything found in the Conserved Domain Database (CDD), but the degP_htrA_DO

domain of Orf6 and the USP_like domain of Bsp1 were identified. Their presence suggested a possible group-translocator-like function for Orf6 and a stress response role for Bsp1. Surprisingly, the known functions of these domains correlate with the predicted functions of the phylogenetic clusters in which they reside.

Comparative analysis of the phylogenetic and 16S/18S rRNA trees revealed that lateral gene transfer was common within the bacterial and archaeal domains, and less common within the eukaryotic domain. Lateral gene transfer between bacteria and archaea was found to be relatively frequent and exceptionally rare between bacteria and eukaryotes. As a result, orthology was generally not observed within the bacterial domain, with a notable exception of the Actinobacterial and Cyanobacterial homologues in cluster 5.

Following up on studies dealing with bacterial members of the TSUP family, we were able to demonstrate a 4 TMS repeat in bacteria, eukaryotes and archaea (Saier et al., 2006; 2009; Yen et al., 2009). 2 TMS repeat units have been found in several families of transport proteins, including the Oligopeptide Transporter (OPT; TC# 2.A.67), CRAC channel and CDF families (Gomolplitinant and Saier, 2011; Matias et al., 2010). However, our sensitive methods were unable to detect a 2 TMS repeat unit within TSUP homologues. Sequence divergence may have accounted for our difficulties in identifying the 2 TMS repeat unit. However, another obstacle is the criterion for proving homology between two small repeats is our self-imposed requirement for a stretch of at least 60 residues with a comparison score of at least 10 S.D. Even with a sufficient score, an inadequate comparison length introduces

uncertainty regarding the suggestion of homology. With proper statistical considerations, perhaps the length requirement may, in the future, be lowered with a concomitant increase in the comparison score requirement without loss of confidence. Since we proved unsuccessful in establishing TSUP 4-fold symmetry, we decided to take another approach, comparing TSUP family members with other proteins in TC subclasses 2.A and 9.A.

We have provided evidence for homology between the TSUP family and several other families that have yet to be assigned to superfamilies. As a result, the new Microbial Rhodopsin (MR) superfamily, consisting of 6 currently recognized families of 6-8 TMS topologies, was created (Fig. 7a-b; Table 2). All families within the superfamily appear to have arisen from a 4 TMS primordial peptide, followed by subsequent N- and C-terminal loss of one or more TMSs. The most unique evolutionary route taken appears to be that of the LIV-E (TC# 2.A.78) family. After the initial 4 TMS intragenic duplication event, LIV-E proteins appear to have lost 4 TMSs and duplicated TSUP TMSs 2-5. The N-terminal TSUP TMS 2 may then have been lost thereafter to generate the present day 7 TMS topology.

Comparisons to the CDF family, as well as the CEO family within the DMT superfamily, both of which have had their 2 TMS origin demonstrated (Lam et al., 2011; Matias et al., 2010), allowed us to indirectly show that the 4 TMS ancestral unit of TSUP arose from the duplication of an ancestral transmembrane hairpin structure. Furthermore, we demonstrated sequence similarities for short stretches of TSUP family members and those of families within the MFS and RND superfamilies.

Definitive evidence for how the MFS arose has not been forthcoming. Sequence similarity between the TSUP family and the MFS supports the theory of a 2 TMS triplication giving rise to the 6 TMS precursor that then duplicated to give the standard 12 TMS topology. This would be in contrast to the competing possibility of a 3 TMS duplication (Fig. 28a). A separate GAP comparison of MFS TMSs 1-2 and 3-4 yielded a high level of identity and provided further evidence for a 2 TMS origin (Fig. 28b).

Our search for the origins of the MFS and RND superfamilies were conducted on a small scale, utilizing only the families with which TSUP family members were compared. Expanding the comparisons to include most or all of the families within the MFS and RND superfamilies may prove to be fruitful in elucidating their evolutionary origins. Sequence similarity to so many different families within TC class 2.A supports a secondary active transport mechanism for the TSUP family. Also, some degree of sequence similarity between the TSUP family and the IT, MOP, CPA, BART and APC superfamilies was observed (unpublished data), but the regions were not conserved across comparisons. Therefore, further studies will be needed.

We are aware that sequence convergence may explain sequence similarity when the regions compared are short. Additionally, the need for stable transmembrane segments along with functional requirements may dictate sequence convergence in somewhat longer sequences (Remmert et al., 2010). To establish homology in some cases, new methods of distinguishing sequence convergence from distant homology must be devised. However, if our suggestion of relatedness between

multiple superfamilies results in the identification of Super-superfamilies, their utility may be limited if their repeat units differ. Under such circumstances there would be little reason to suggest common 3-d structural folds. Thus, by going back too far, in evolutionary time, the assumption of common structure, function and mechanism may no longer be valid. If so, such information may be of minimal value. It does appear possible, however, that a hairpin structure, of α-helical and β-strand composition may have been the precursor of a diverse set of transport proteins (Remmert et al., 2010).

For the most part, genome context analyses supported the few biochemical assays that have been performed using TSUP homologues. The results suggested a sulfur-based compound transport role (Gristwood et al., 2011; Krejcík et al., 2008; Locher et al., 1993; Mampel et al., 2004; Rückert et al., 2005; Weinitschke et al., 2007). The data from clusters 1, 5, 6, 8 and 12-15 provide support for the suggestions, arising from the limited biochemical studies (Table 3). However, eclectic and overlapping transport roles were also observed. Given the apparent functional diversity of our predictions as well as the sequence diversity inherent to the TSUP family, it may be that the TSUP family members can transport a wide range of compounds.

In view of the considerations cited above, we predict that many TSUP transporters catalyze the uptake or efflux of sulfur-containing compounds. However, several functional outliers may exist. These may transport (1) nucleotides/nucleosides/nucleobases, (2) amino acids/peptides, (3) carbohydrates and (4) lipids. Many TSUP proteins may function as part of stress responses and/or play

roles in cofactor precursor transport. At least some TSUP members appear to function with outer membrane and periplasmic proteins. The elucidation of these functions, using the predictions presented here as a guide, are likely to open up new fields of study.

Appendix

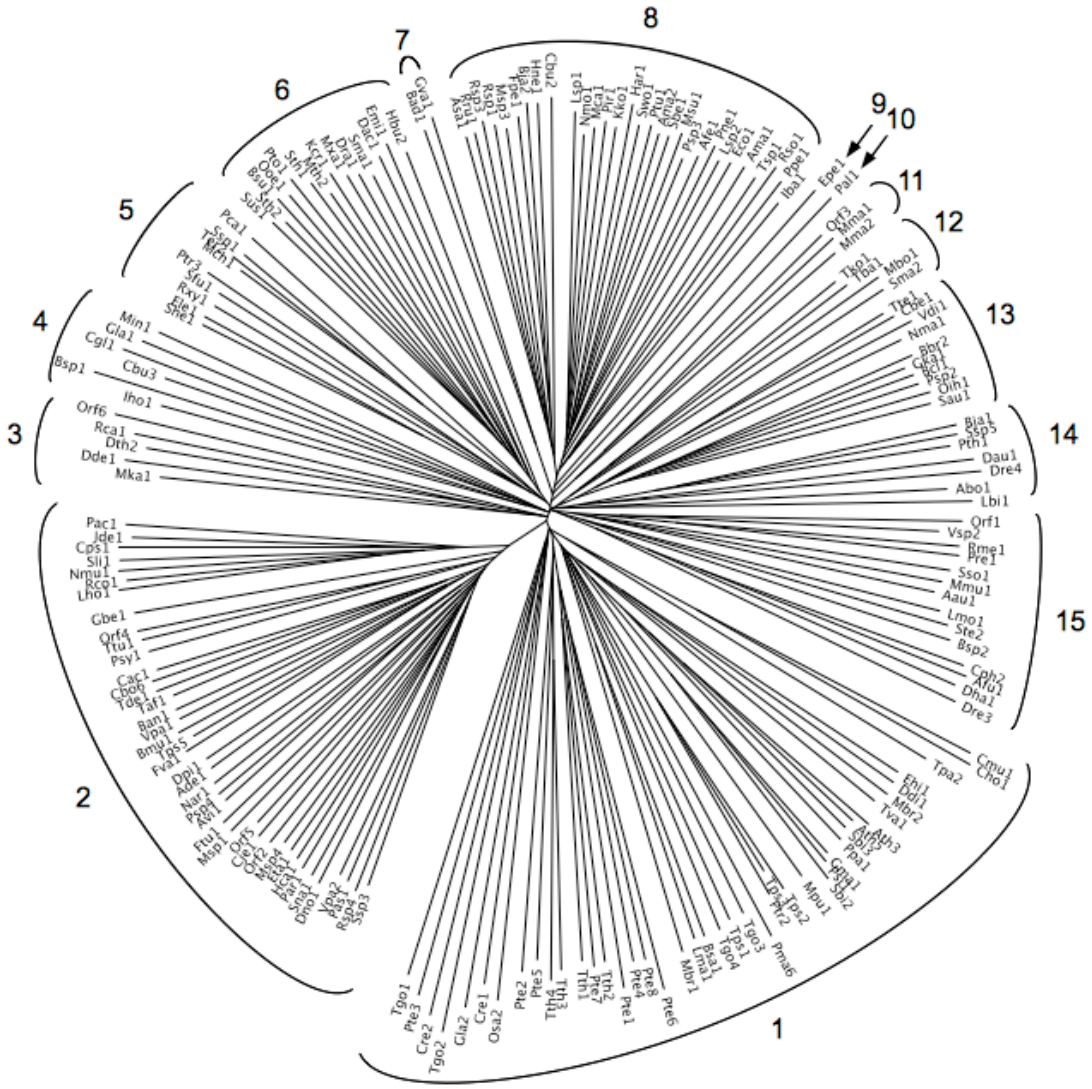

Figure 1: Phylogenetic tree of the 187 TSUP family proteins included in this study. The tree was generated using the ClustalX multiple alignment and FigTree program for visualization. Protein abbreviations and their descriptions are listed in Table 1 in a clockwise fashion starting from cluster 1. The positions of individual proteins within the phylogenetic tree are revealed in Fig. 2.

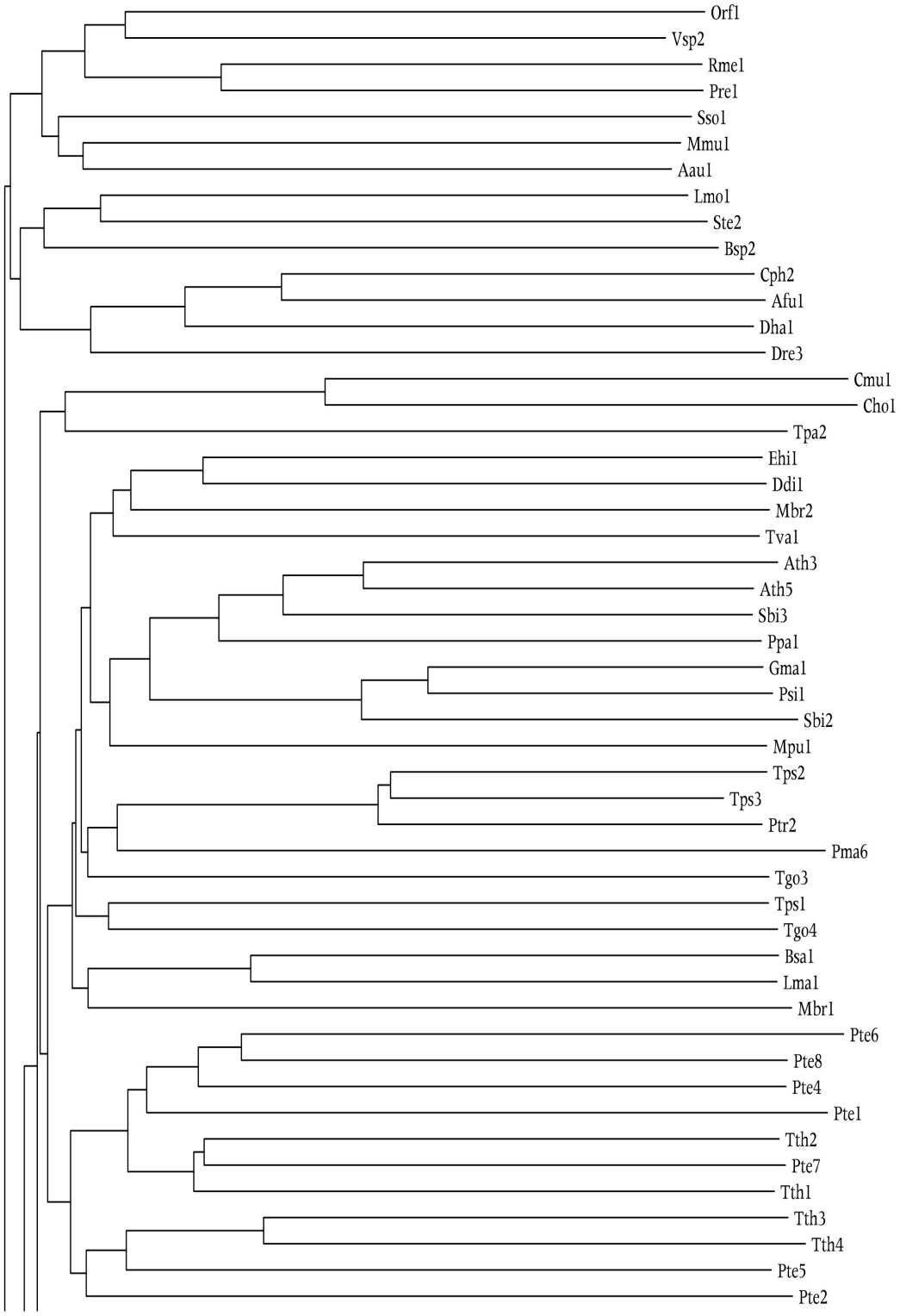

Figure 2: Dendogram of the 187 TSUP family proteins included in this study corresponding to the phylogenetic tree shown in Figure 1.

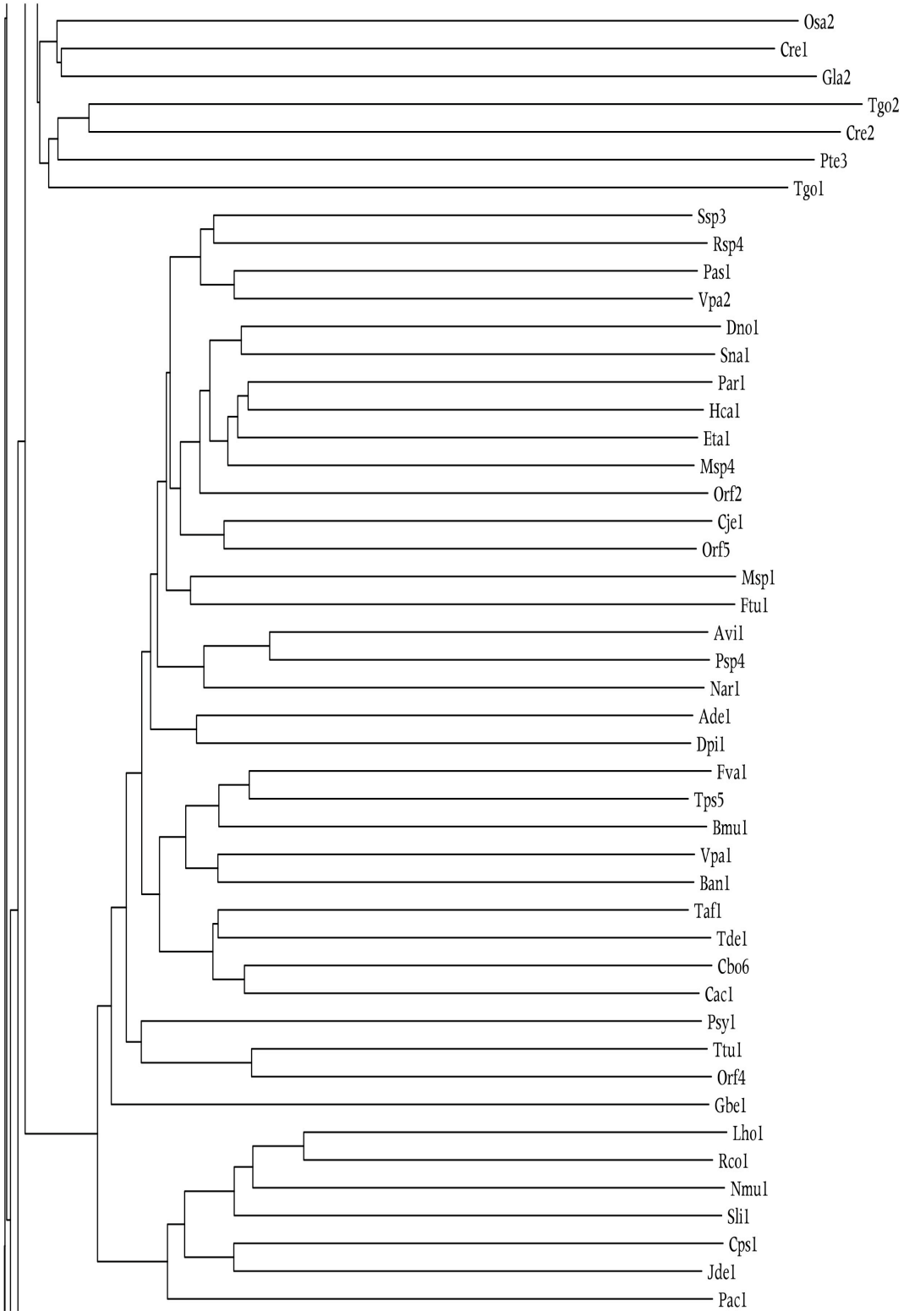

Figure 2: TSUP Dendogram Continued

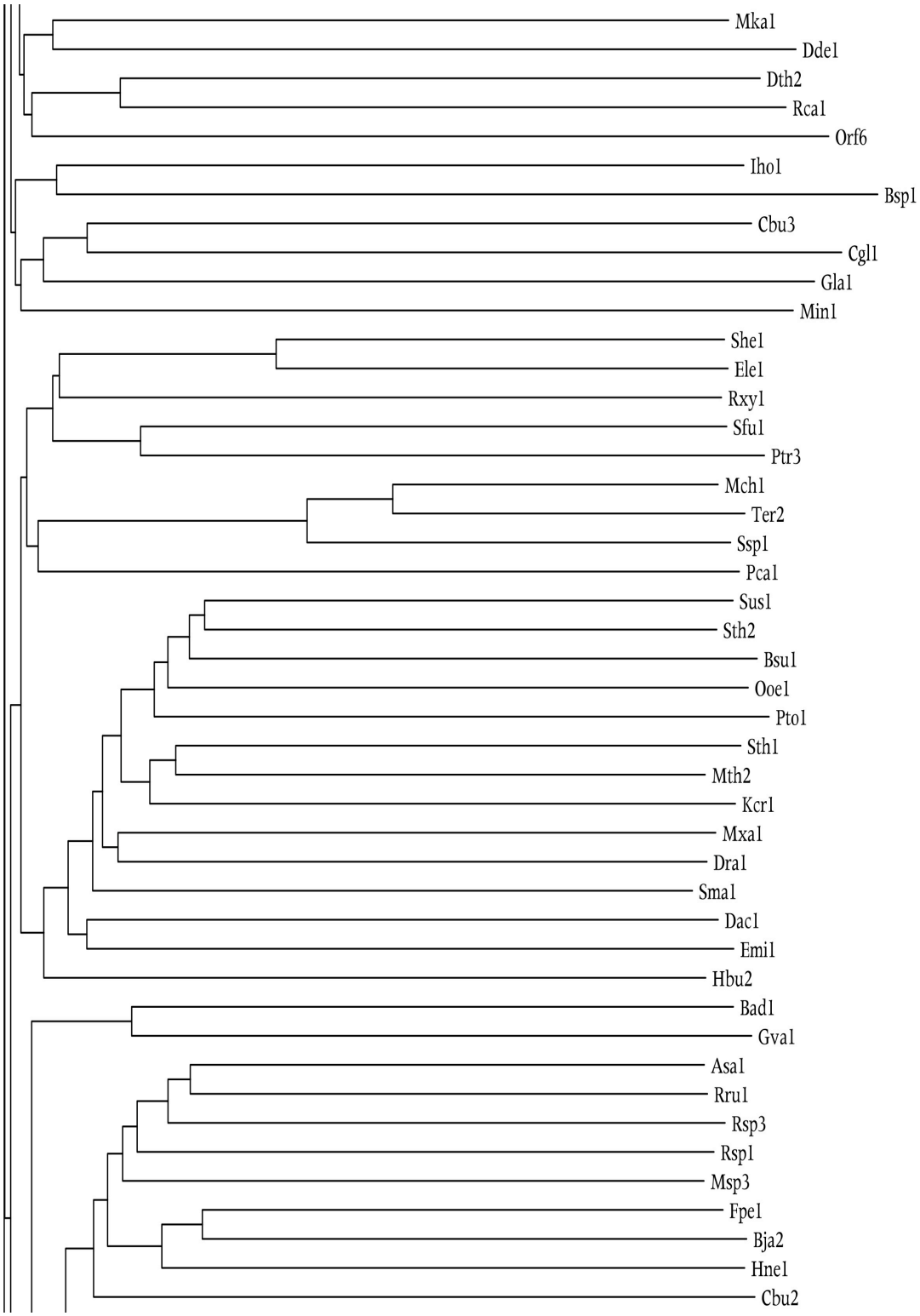

Figure 2: TSUP Dendogram Continued

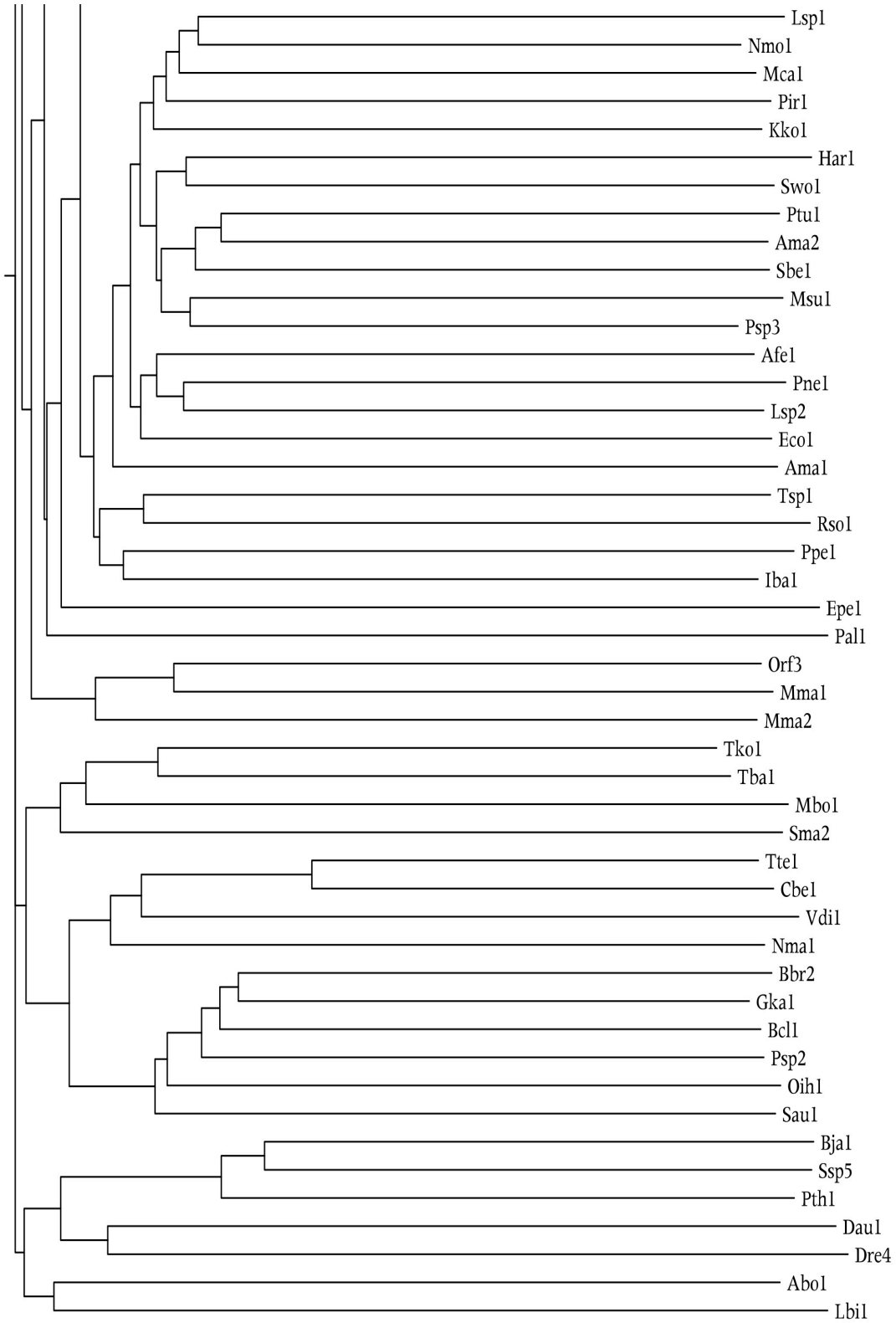

Figure 2: TSUP Dendogram Continued

Figure 3: 16S/18S rRNA phylogenetic tree of genuses represented in the study. The *Cloacamonas*, *Symbiobacterium*, *Oenococcus*, *Endoriftia* and *Desulforudis* genera were excluded due to unreliable sequence data.

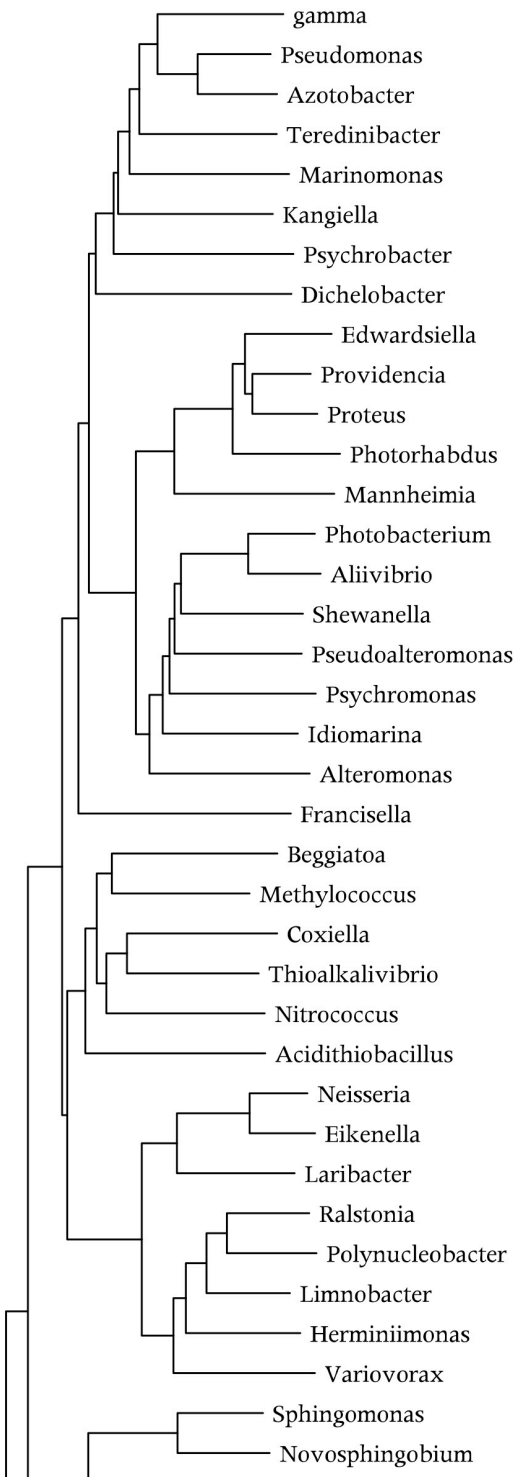

Figure 4: Dendogram corresponding to the 16S/18S rRNA phylogenetic tree presented in Figure 3.

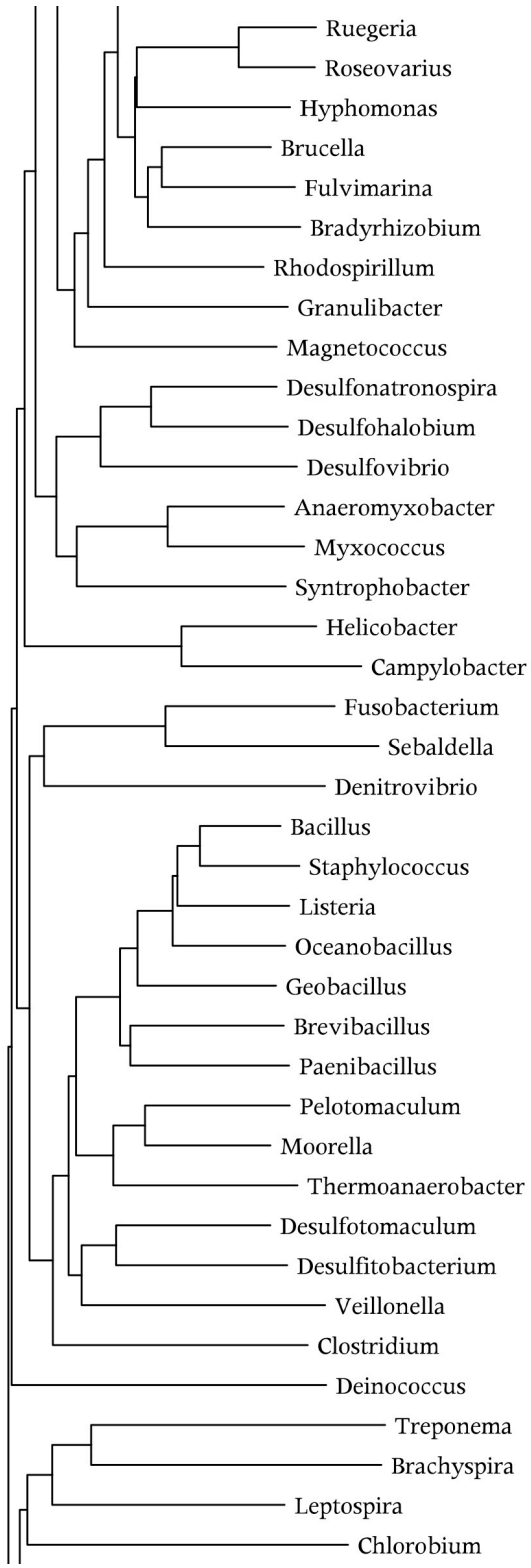

Figure 4: TSUP 16S/18S rRNA Dendogram Continued

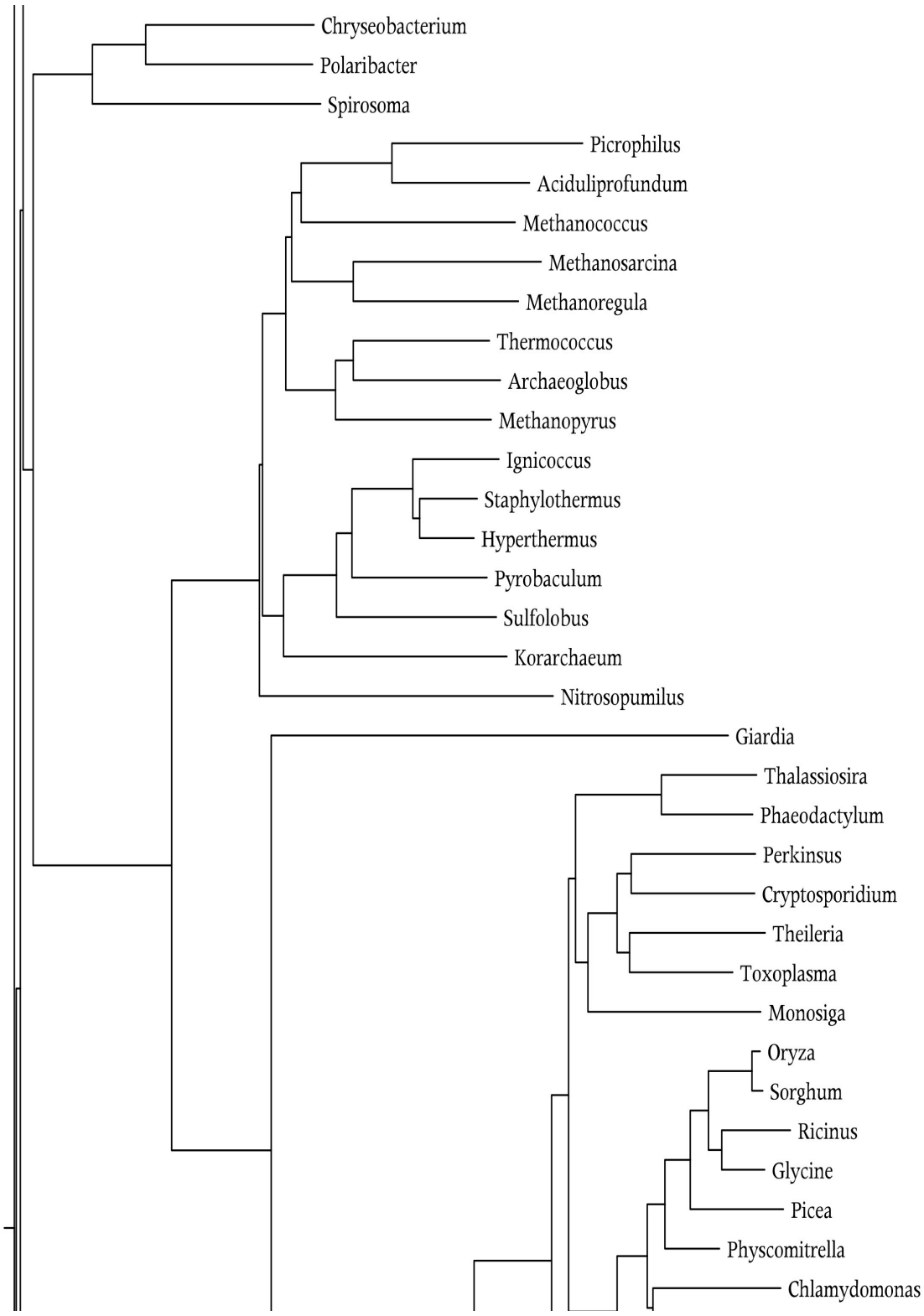

Figure 4: TSUP 16S/18S rRNA Dendogram Continued

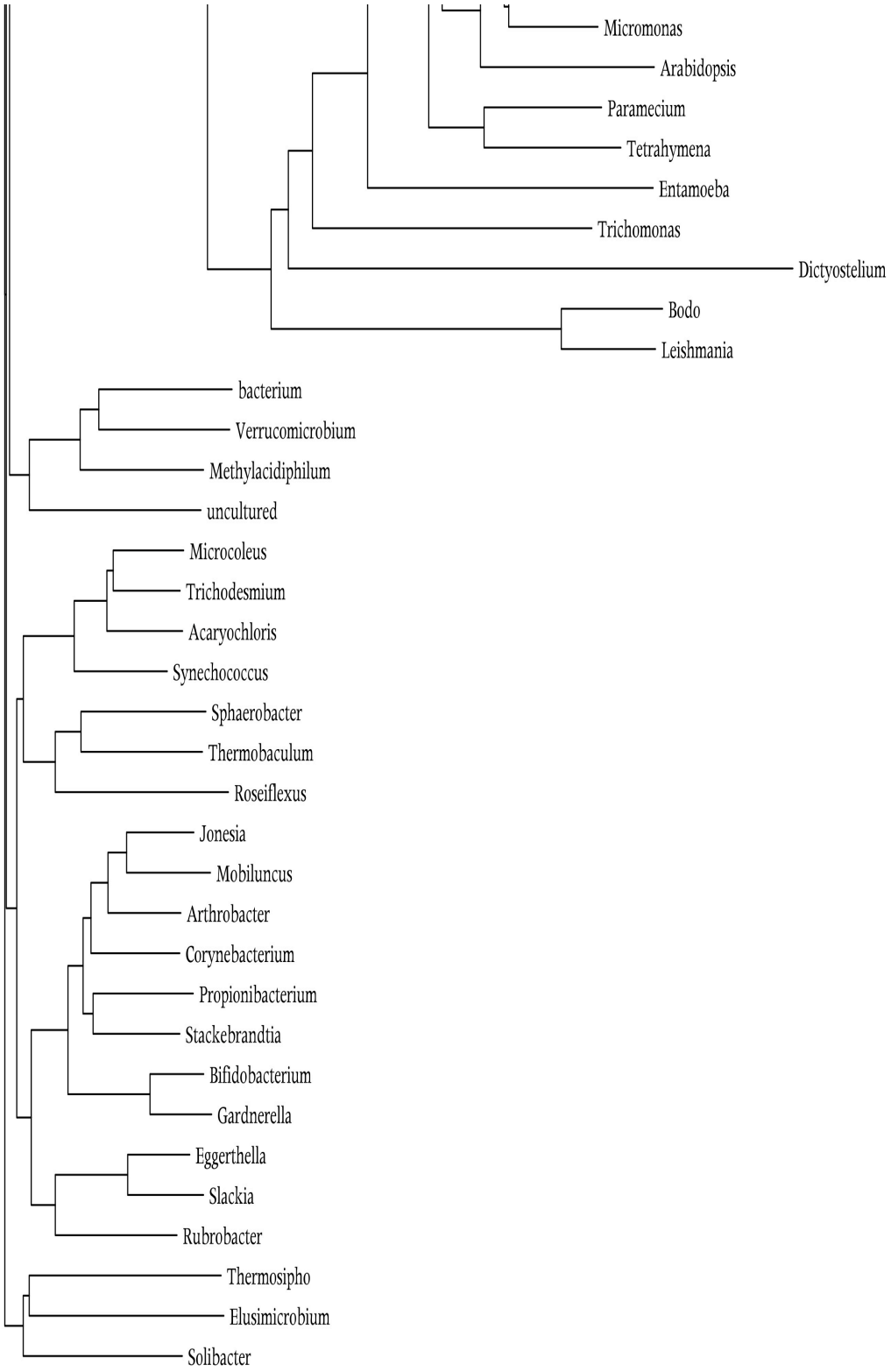

Figure 4: TSUP 16S/18S rRNA Dendogram Continued

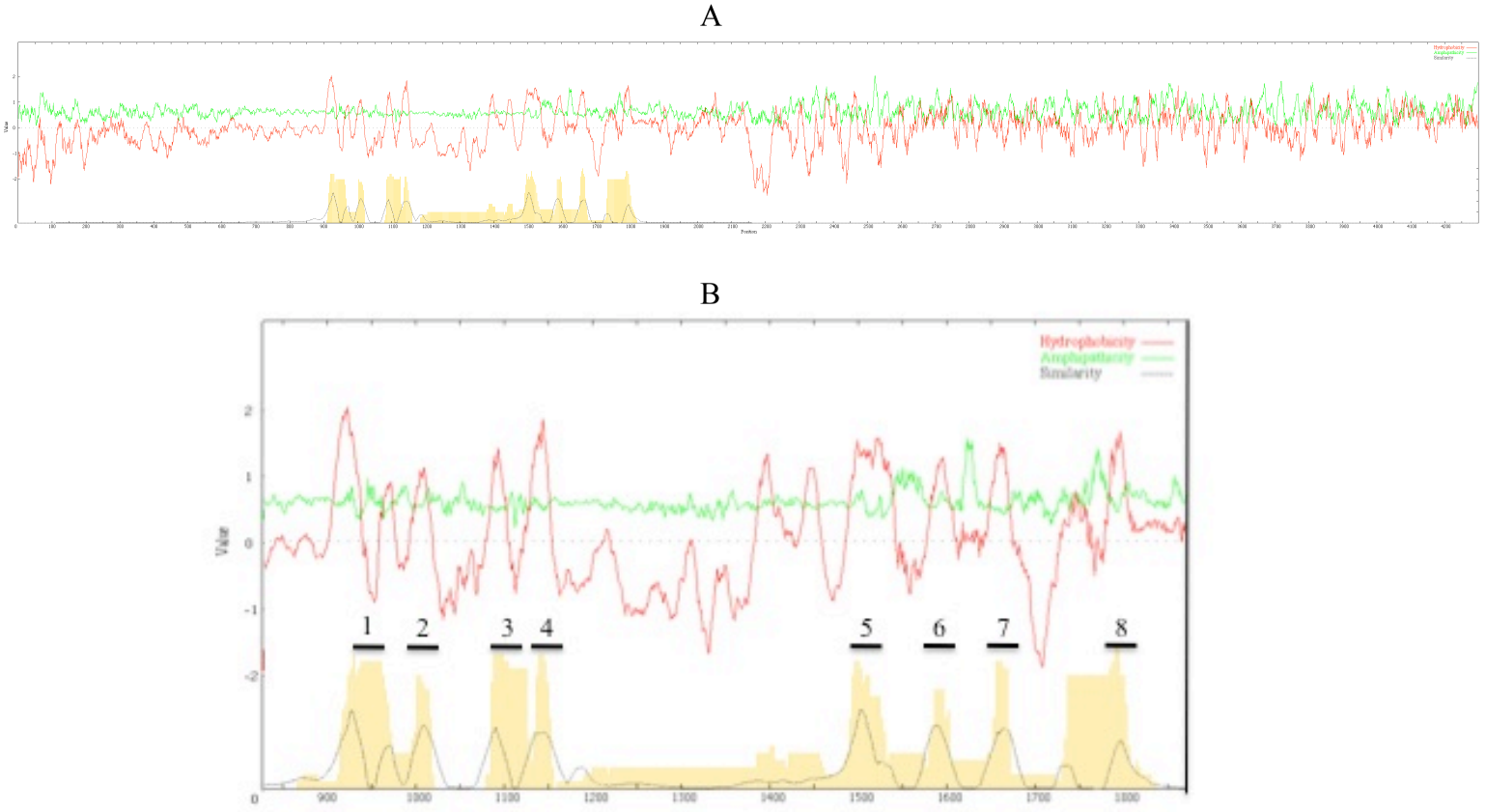

Figure 5: (A) Average hydropathy, amphipathicity, and similarity (AveHAS) plots for the 189 TSUP family proteins included in the study. Proteins with N- and C-terminal hydrophilic domain extensions contributed to the large size of the plot, and their functional roles are discussed in the text. (B) Magnification of the TMS-containing region. The plot reveals an average of 8 TMSs, which matches the results of our rigorous TMS count analysis.

Figure 6: Demonstration that 8 TMS TSUP family members arose by an intragenic duplication of a primordial 4 TMS encoding genetic element. **(A)** Summary of 3 comparisons demonstrating homology as shown in figures B-D. The ≈ symbol indicates homology. **(B)** GAP alignment of TMSs 1-4 of TSUP Tko1 (*Thermococcus kodakarensis*; gi 57640914) with TMSs 5-8 of Tko1. Initial identification of repeat units was done using the IC program. The GAP program was run with default settings and 500 random shuffles in order to generate the alignment. Residue identity is signified by a vertical line, while close and more distant similarities are signified by a colon or a period, respectively. The numbers at both ends of each line signify the positions of the residues in the protein. TMS positions were predicted using the TMHMM 2.0 program; HMMTOP was used for TMS 7. The same convention was used for subsequent comparisons. A comparison score of 26.3 S.D. was obtained. The two segments compared (TMSs 1-4 and 5-8) are 107 aas long. **(C)** GAP alignment of TMSs 1-4 of TSUP Mch1 (*Microcoleus chthonoplastes*; gi 224407624) with TMSs 5-8 of Mch1. A comparison score of 17.0 S.D. was obtained from this alignment. **(D)** GAP alignment of TMSs 1-4 of TSUP Tko1 with TSUP TMSs 5-8 of Mch1. A comparison score of 16.0 S.D. was obtained from this alignment.

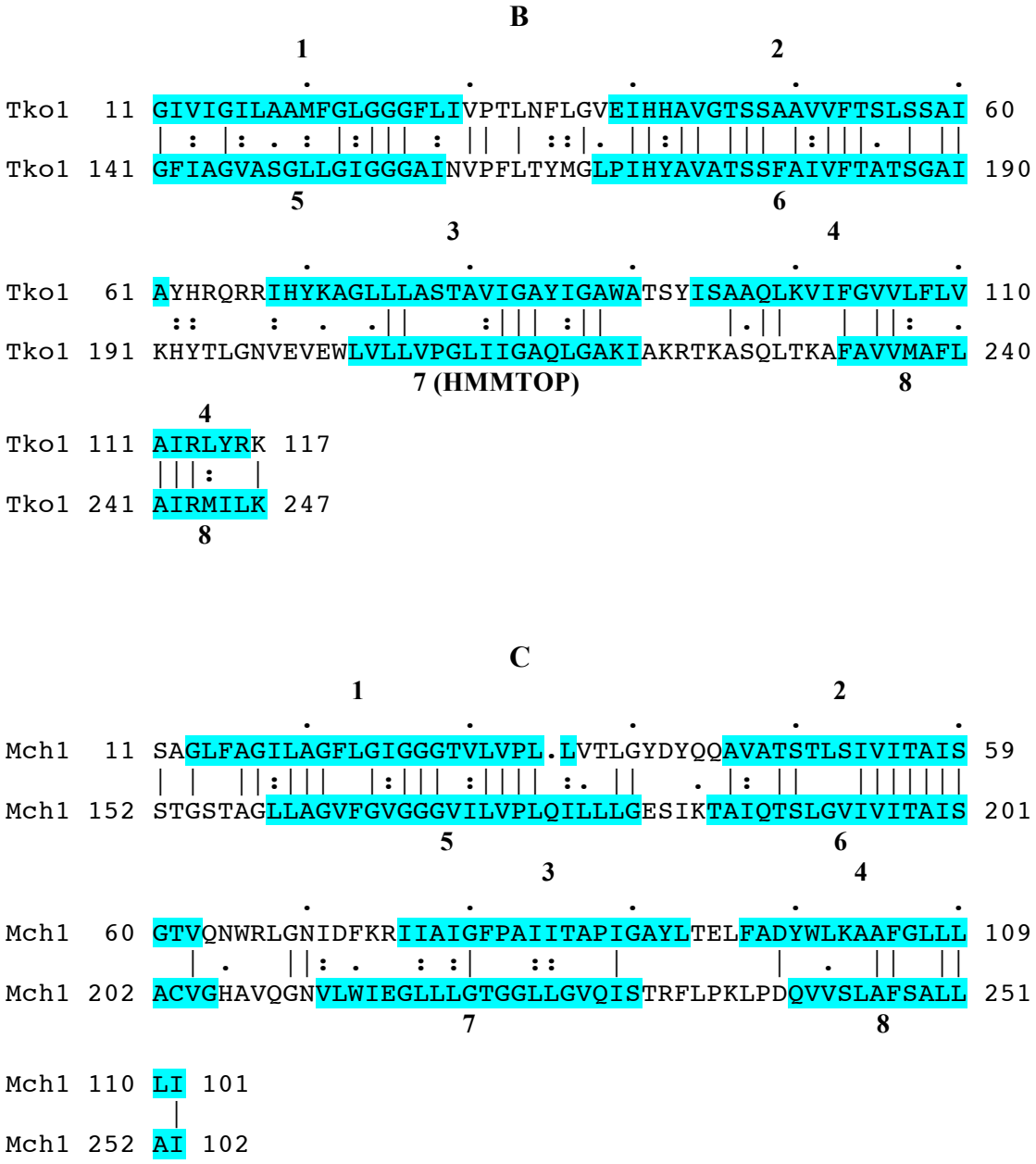

Figure 6: TSUP Internal Repeats Continued

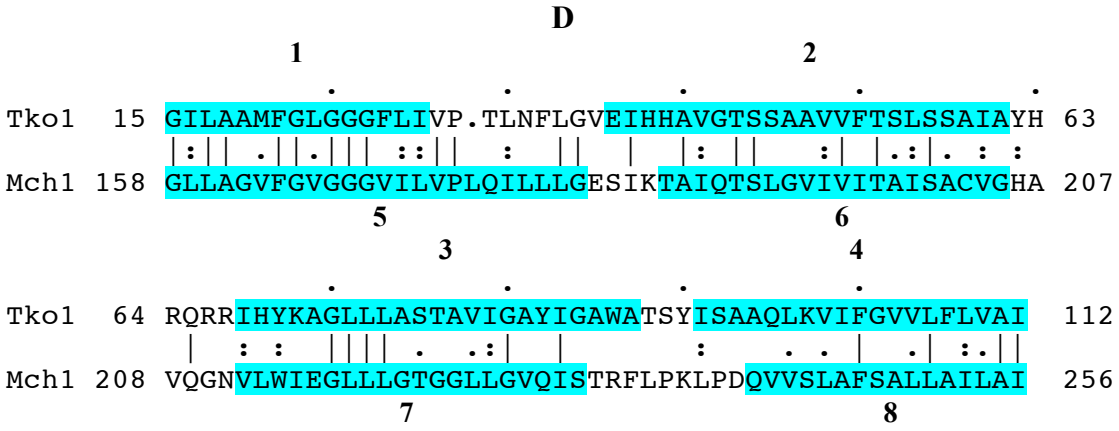

Figure 6: TSUP Internal Repeats Continued

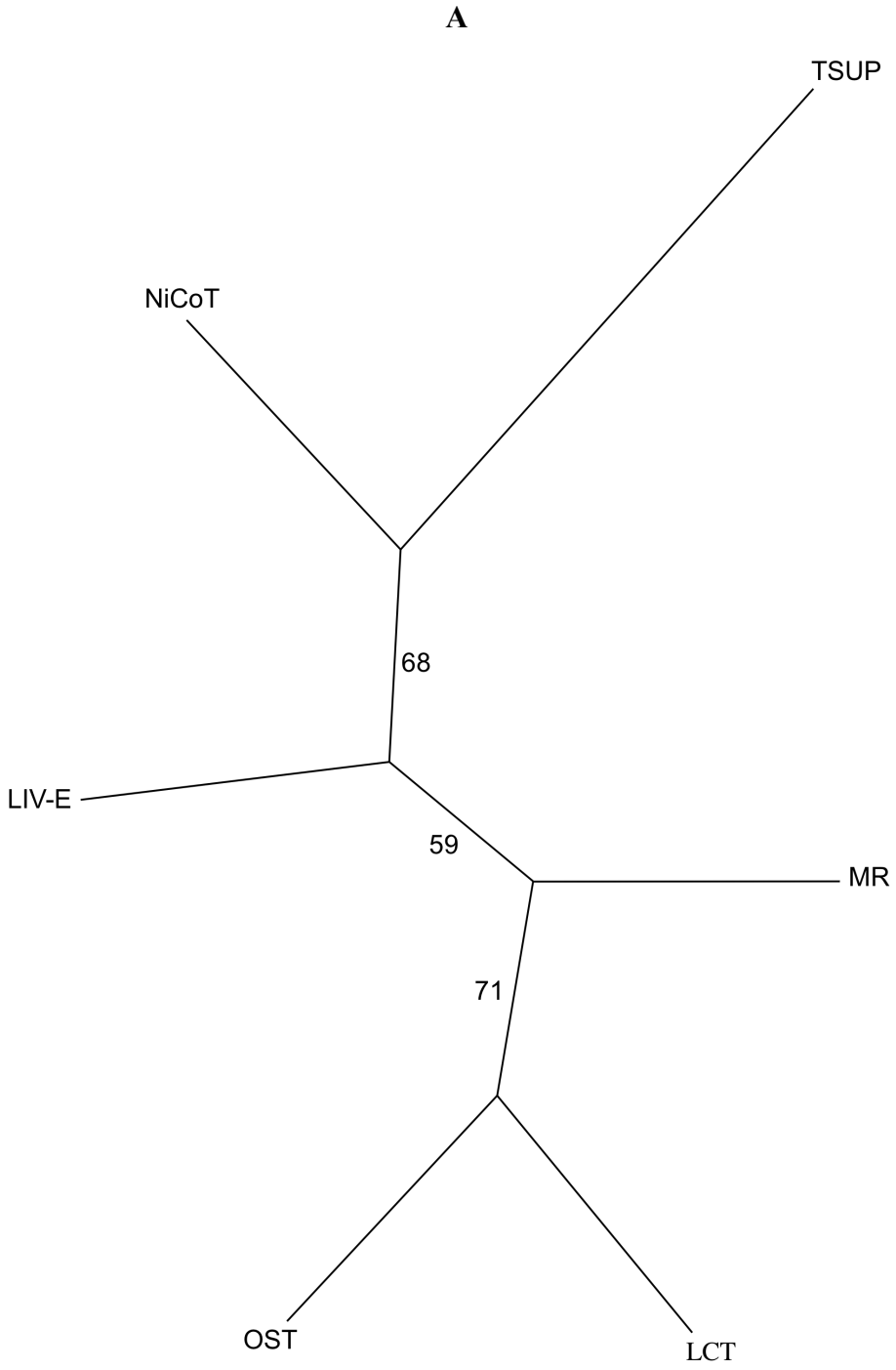

Figure 7: (A) Superfamily tree (SFT1 and 2) results for current members of the MR superfamily. The numbers indicate relative confidence levels. **(B)** Proposed evolutionary pathway for the six recognized families within the MR superfamily.

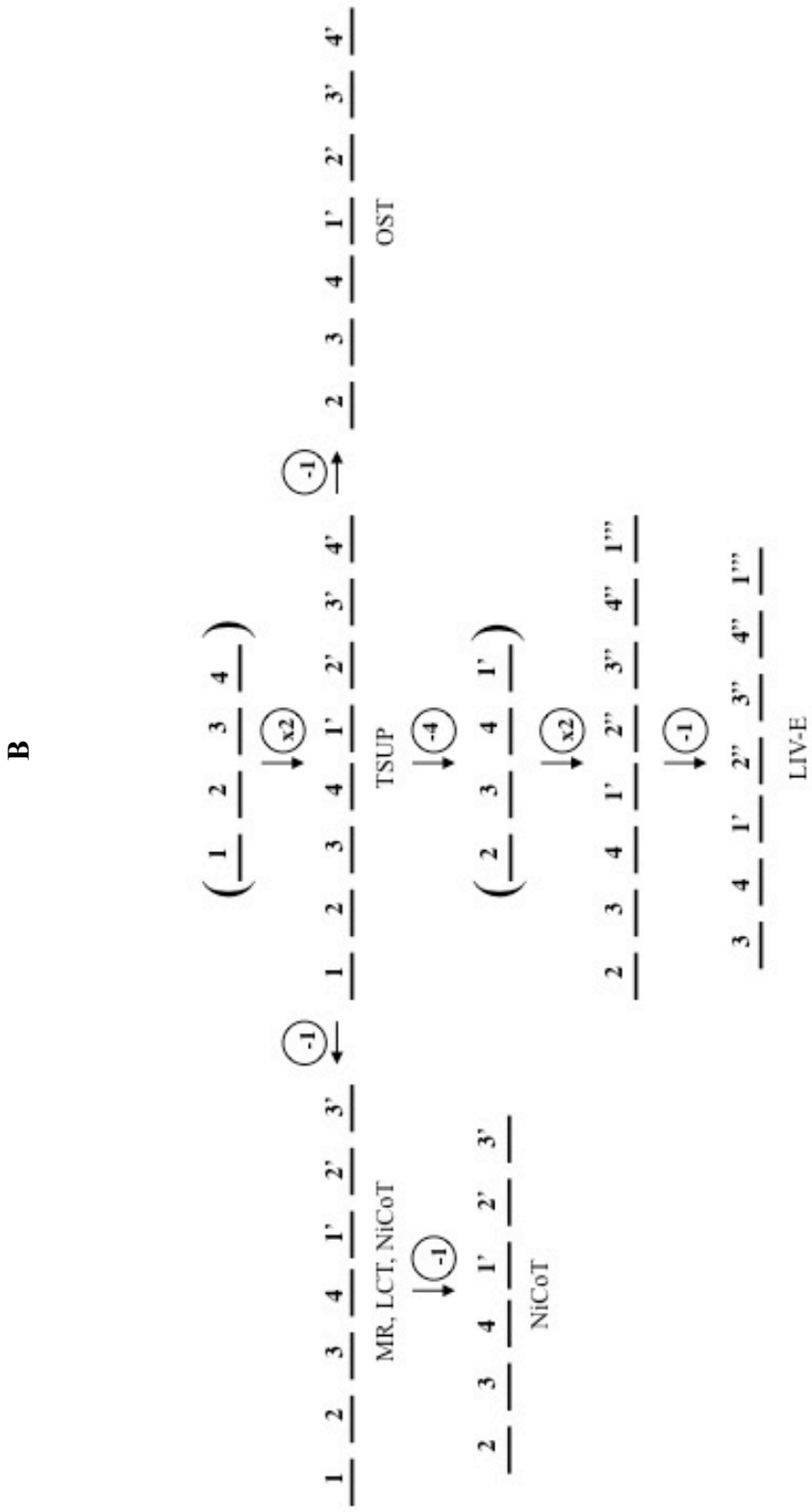

Figure 7: MR Superfamily Continued

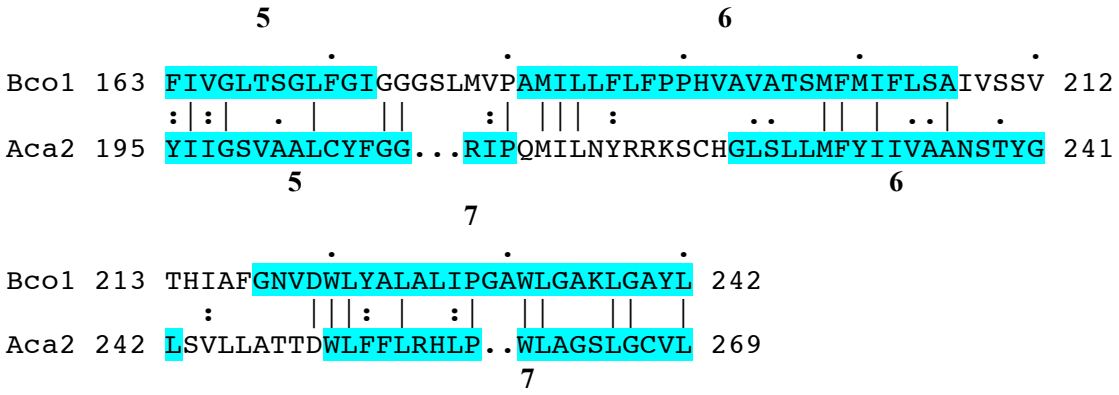

Figure 8: Homology between members of the TSUP and LCT families. GAP alignment of TMSs 5-7 of TSUP Bco1 (*Bacillus coahuilensis*; gi 283846803) with TMSs 5-7 of LCT Aca2 (*Angiostrongylus cantonensis*; gi 256016595). A comparison score of 10.8 S.D. was obtained with 40.0% similarity and 30.7% identity.

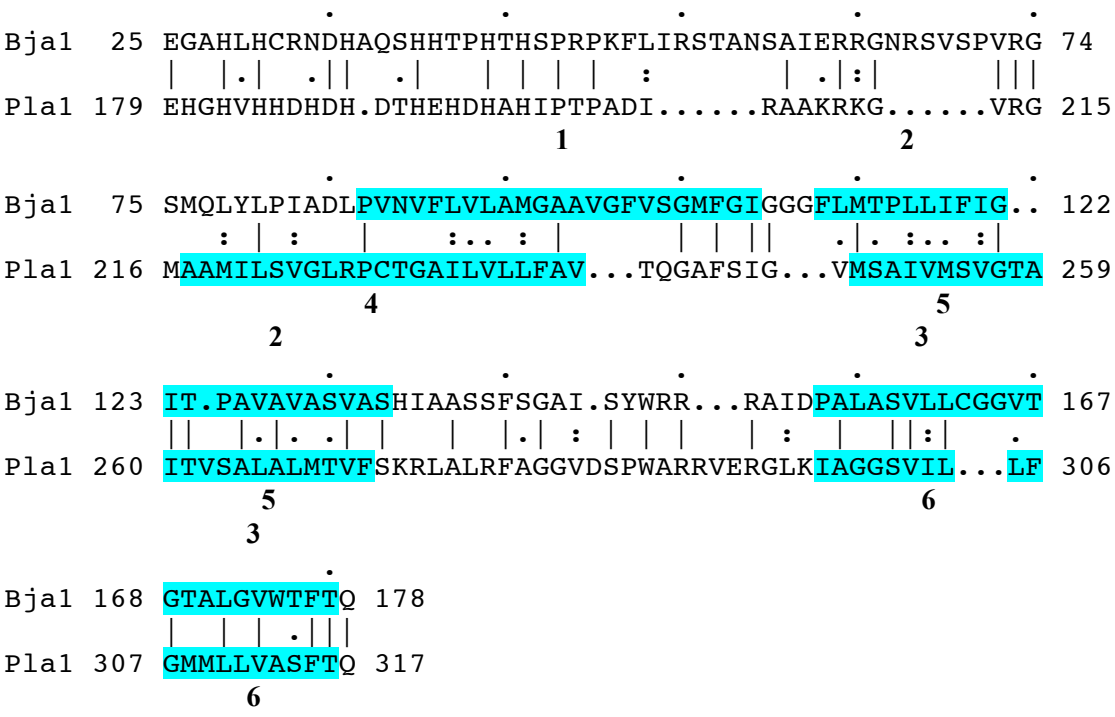

Figure 9: Homology between members of the TSUP and NiCoT families. GAP alignment of TMSs 1-3 of TSUP Bja1 (*Bradyrhizobium japonicum*; gi 27376265) with TMSs 4-6 of NiCoT Pla1 (*Parvibaculum lavamentivorans*; gi 154252649). A comparison score of 12.8 S.D. was obtained with 44.7% similarity and 36.4% identity.

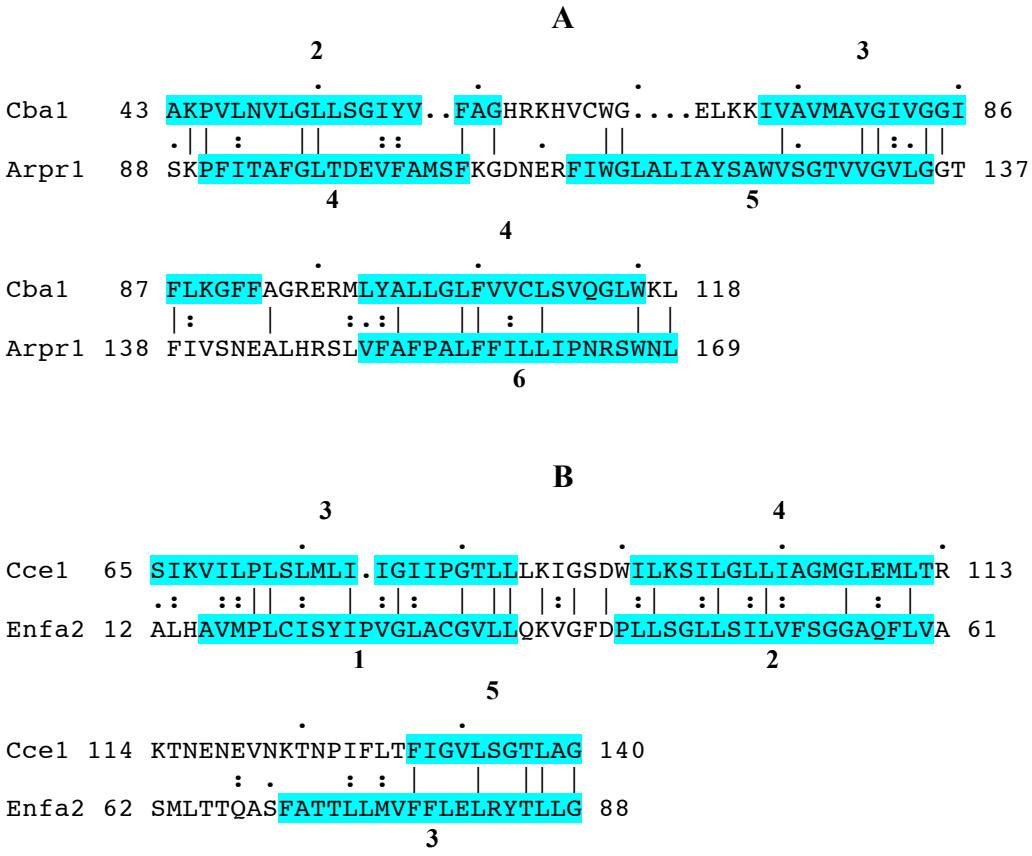

Figure 10: (A) Homology between members of the TSUP and LIV-E families. GAP alignment of TMSs 2-4 of TSUP Cba1 (*Clostridiales bacterium*; gi 239625513) with TMSs 4-6 of LIV-E Arpr1 (*Archaeoglobus profundus*; gi 284161715). A comparison score of 12.2 S.D. was obtained with 38.2% similarity and 27.6% identity. **(B)** GAP alignment of TMSs 3-5 of TSUP Cce1 (*Clostridium cellulovorans*; gi 242260426) with TMSs 1-3 of LIV-E Enfa2 (*Enterococcus faecium*; gi 227551482). A comparison score of 11.1 S.D. was obtained with 46.1% similarity and 26.3% identity.

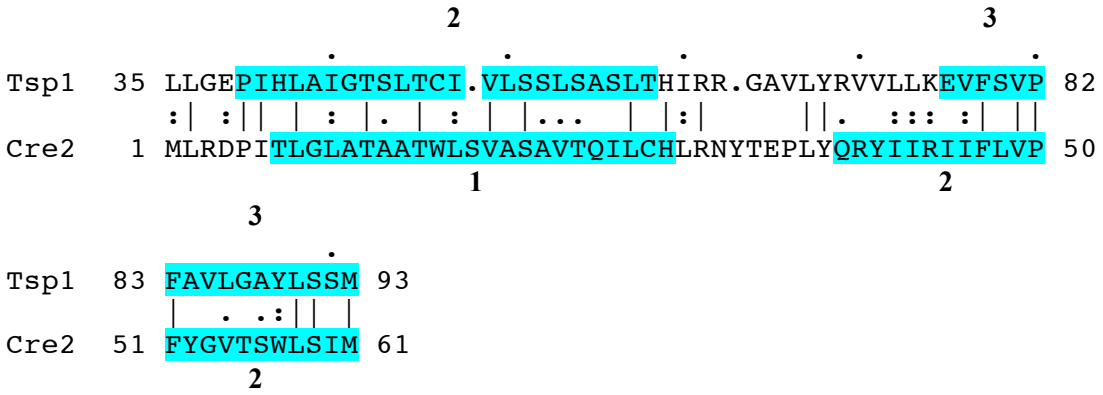

Figure 11: Homology between members of the TSUP and OST families. GAP alignment of TMSs 2-3 of TSUP Tsp1 (*Thermococcus* sp. AM4; gi 254172062) with TMSs 1-2 of OST Cre2 (*Chlamydomonas reinhardtii*; gi 159465163). A comparison score of 12.1 S.D. was obtained with 50.8% similarity and 33.9% identity.

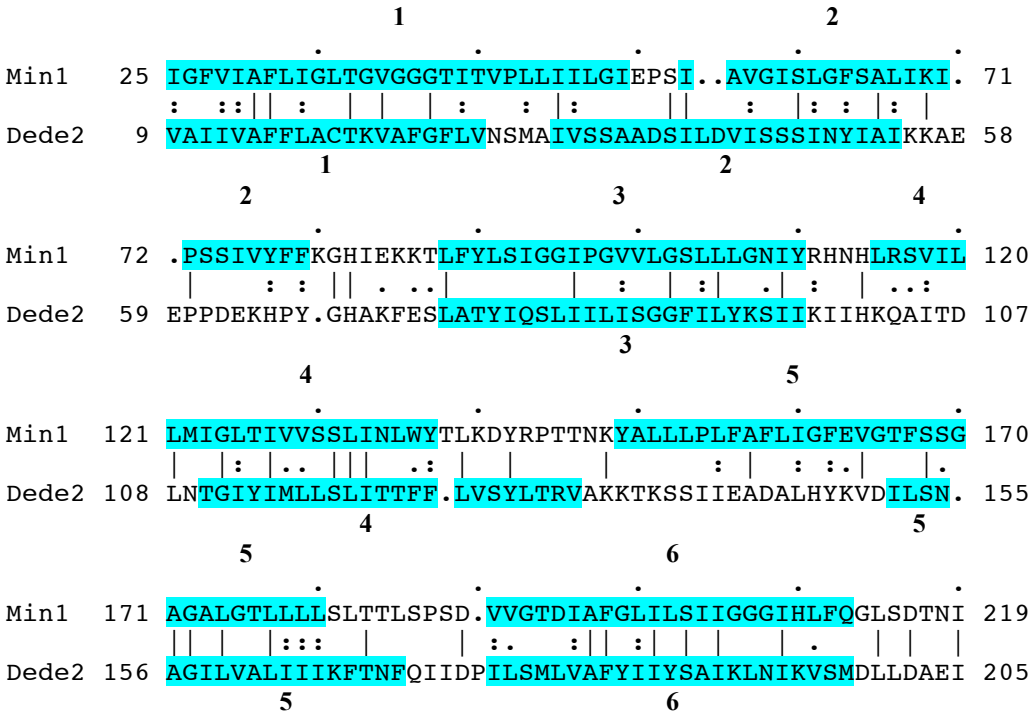

Figure 12: Sequence similarity between members of the TSUP and CDF families. GAP alignment of TMSs 1-6 of TSUP Min1 (*Methylacidiphilum infernorum*; gi 189218632) with TMSs 1-6 of CDF Dede2 (*Deferribacter desulfuricans*; gi 291280364). A comparison score of 11.8 S.D. was obtained with 39.1% similarity and 24.5% identity.

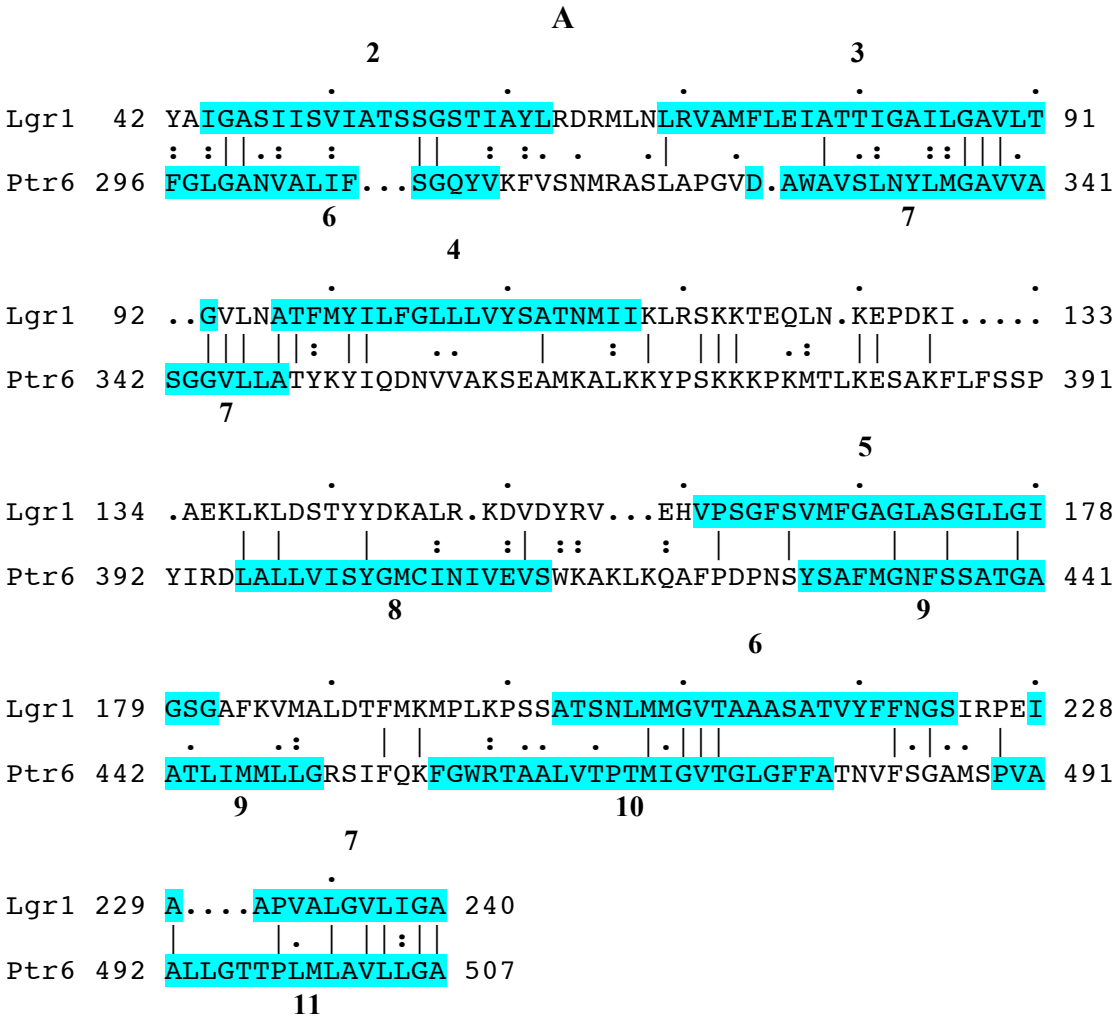

Figure 13: (A) Sequence similarity between members of the TSUP and AAA families. GAP alignment of TMSs 2-7 of TSUP Lgr1 (*Listeria grayi*; gi 299820751) with TMSs 6-11 of AAA Ptr6 (*Phaeodactylum tricornutum*; gi 219128124). A comparison score of 11.2 S.D. was obtained with 35.4% similarity and 25.1% identity. **(B)** WHAT plot for the Ptr6 transport protein.

B

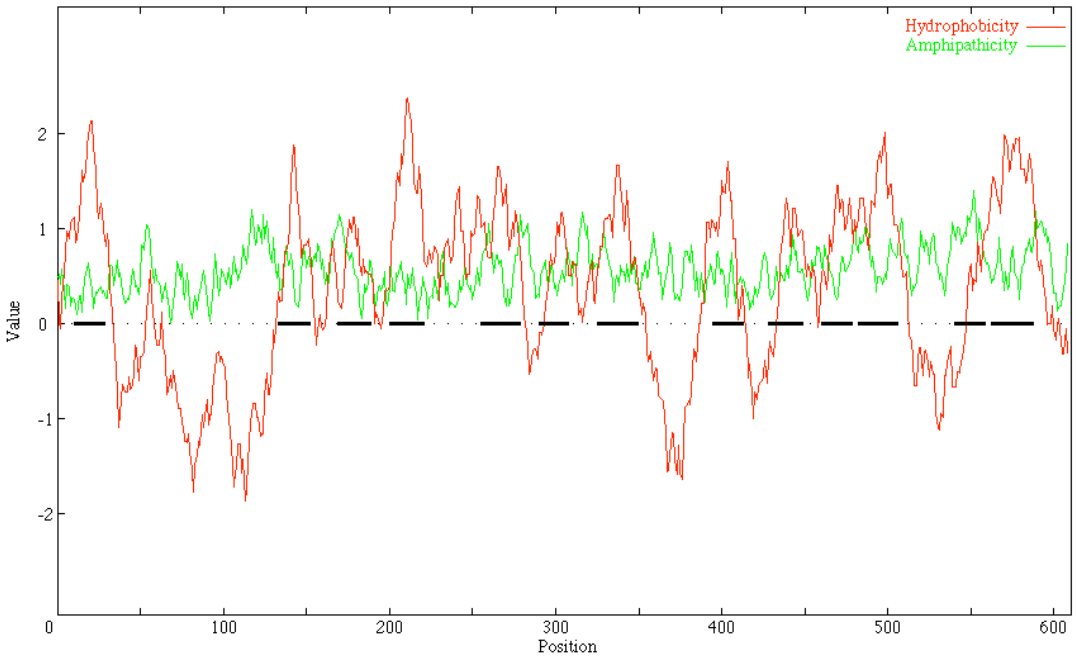

Figure 13: TSUP and AAA Comparison Continued

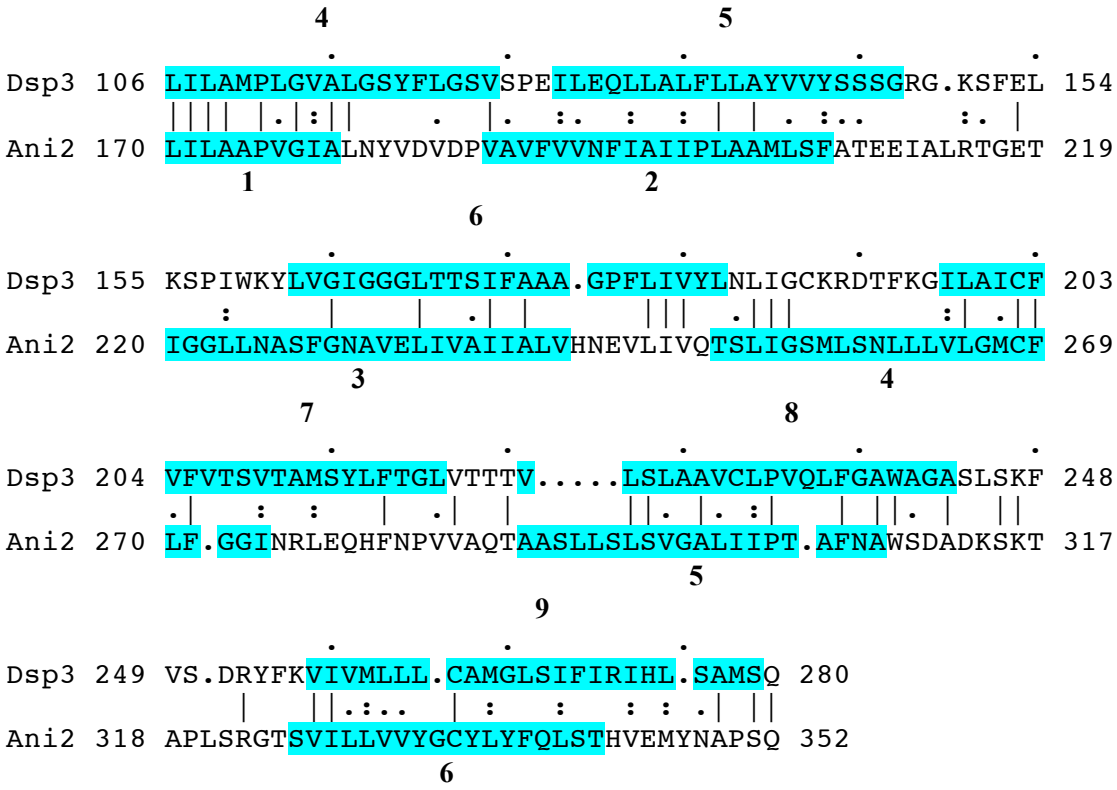

Figure 14: Sequence similarity between members of the TSUP and CaCA families. GAP alignment of TMSs 4-9 of TSUP Dsp3 (*Desulfotalea psychrophila*; gi 51245366) with TMSs 1-6 of CaCA Ani2 (*Aspergillus nidulans*; gi 67901046). A comparison score of 11.3 S.D. was obtained with 35.8% similarity and 26.6% identity.

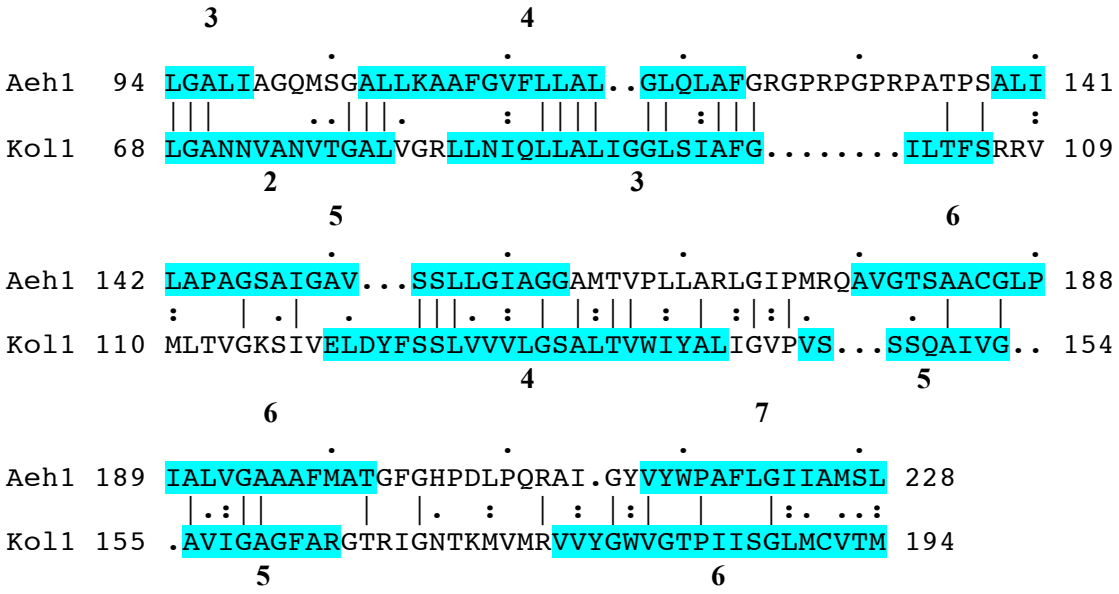

Figure 15: Sequence similarity between members of the TSUP and PiT families. GAP alignment of TMSs 3-7 of TSUP Aeh1 (*Alkalilimnicola ehrlichii*; gi 114319194) with TMSs 2-6 of PiT Kol1 (*Kosmotoga olearia*; gi 239616942). A comparison score of 11.9 S.D. was obtained with 46.3% similarity and 33.9% identity.

A

```
                  2                                3
                  .          .          .          .          .
Rjo3    50 VVNLVIDLAGRLSIVARRYADVSWG..RARLLLLGCVPGVLLGVLTRGRL 97
           .  |.  : | :||      | .  |  .| | : | | | |||. | | |
Apme2  386 LAKLLAFIYGIISIALAFLAQLLGGVLQAGLTIFGVVGGPLLGLFTLGML 435
                        10                          11
                             4
                  .          .          .          .
Rjo3    98 D...TEVIKLSAGVVTLVAVYFIARGGIKREPRKSPPAVVV 135
              ||:  ..  ::.|: .:.||  |    :|| .|| . |
Apme2  436 TETATEIGSVTGTIISLIFLFWIAFG....QPRPTPPTLPV 472
                        12
```

B

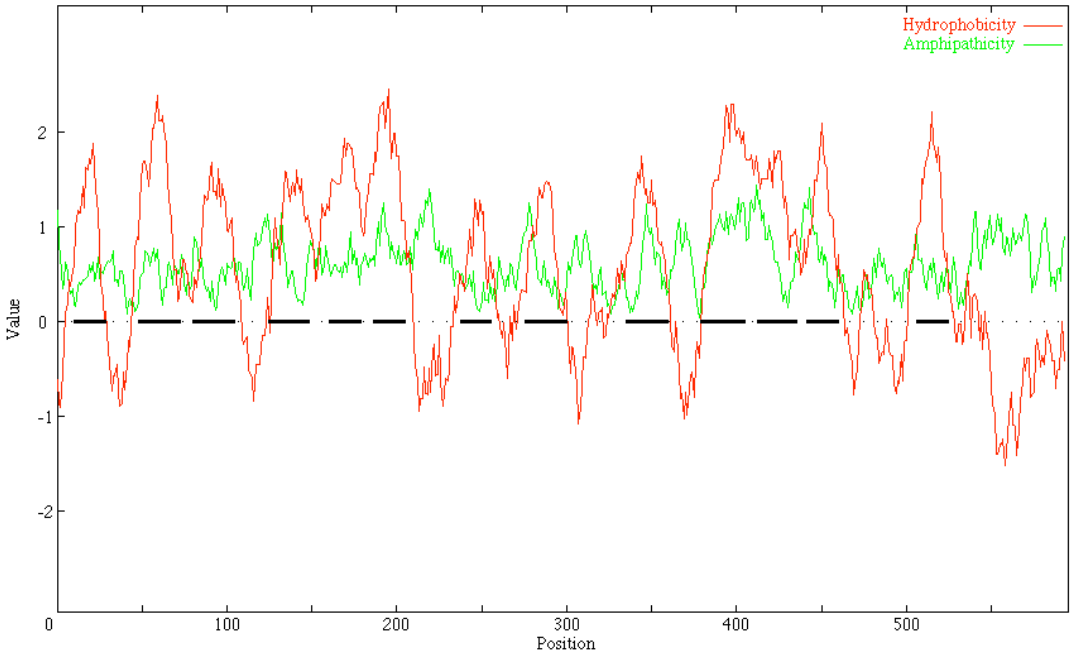

Figure 16: (A) Sequence similarity between members of the TSUP and SSS families. GAP alignment of TMSs 2-4 of TSUP Rjo3 (*Rhodococcus jostii*; gi 111025621) with TMSs 10-12 of SSS Apme2 (*Apis mellifera*; gi 110758640). A comparison score of 13.8 S.D. was obtained with 45.1% similarity and 34.1% identity. **(B)** WHAT plot for the Apme2 symporter.

A

```
                   1                                         2
                        .         .         .         .         .
Pte9   52 LAGIGGGGVVVSLMTMFFNYSQKEALLGVFLPIFGAALG.NFLNLAQQLD 100
          |||:|    . ||  :||  . :| ||  |||| |    |   . :.|  :
Cas9  150 LAGLAGAAALGSLGGLFFGKTVREVLLYYFLPIMGPGTGAGAVPMSQIYE 199
                   5                                   6

                        3                  4
                        .         .         .         .
Pte9  101 PQTKTPVVKIRSAIVACPAMIIGSMVGLILNKILPAFFLI 140
               |  .  |   | | : :|  .. ::|   :.    |:
Cas9  200 SAGMGPAAEFLS..TAVPILTLGLVMSIVLGAVVNKIGLV 237
                                  7
```

B

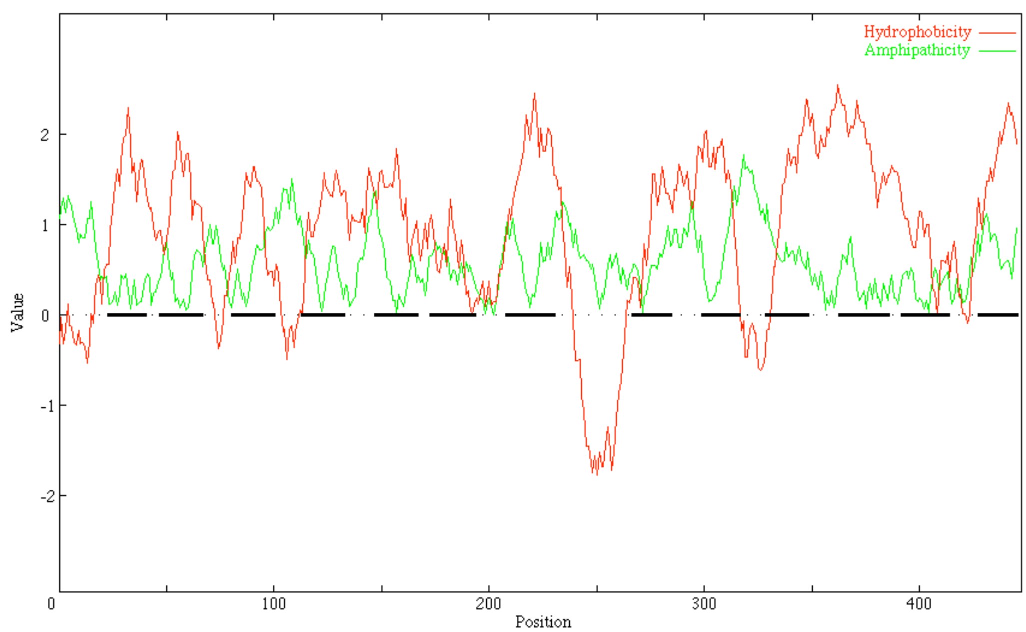

Figure 17: (A) Sequence similarity between members of the TSUP and 2-HCT families. GAP alignment of TMSs 1-4 of TSUP Pte9 (*Paramecium tetraurelia*; gi 145538953) with TMSs 5-7 of 2-HCT Cas9 (*Clostridium asparagiforme*; gi 225388638). A comparison score of 11.9 S.D. was obtained with 41.4% similarity and 28.7% identity. **(B)** WHAT plot for the Cas9 transporter.

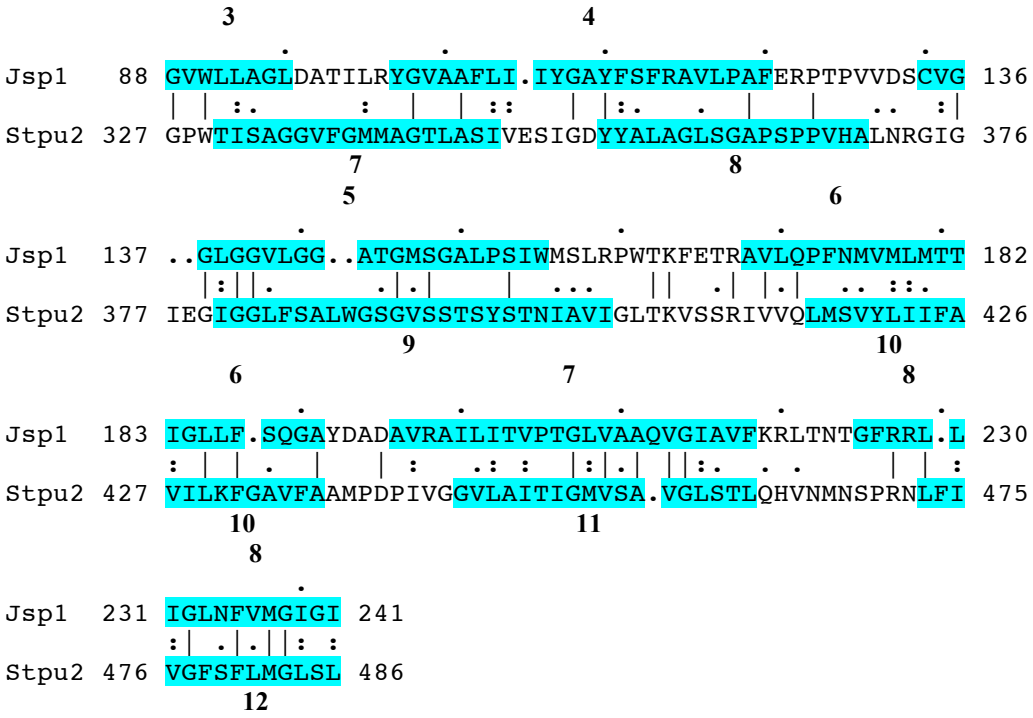

Figure 18: Sequence similarity between members of the TSUP and NCS2 families. GAP alignment of TMSs 3-8 of TSUP Jsp1 (*Jannaschia* sp. CCS1; gi 89056463) with TMSs 7-12 of NCS2 Stpu2 (*Strongylocentrotus purpuratus*; gi 115753713). A comparison score of 11.2 S.D. was obtained with 35.3% similarity and 22.9% identity.

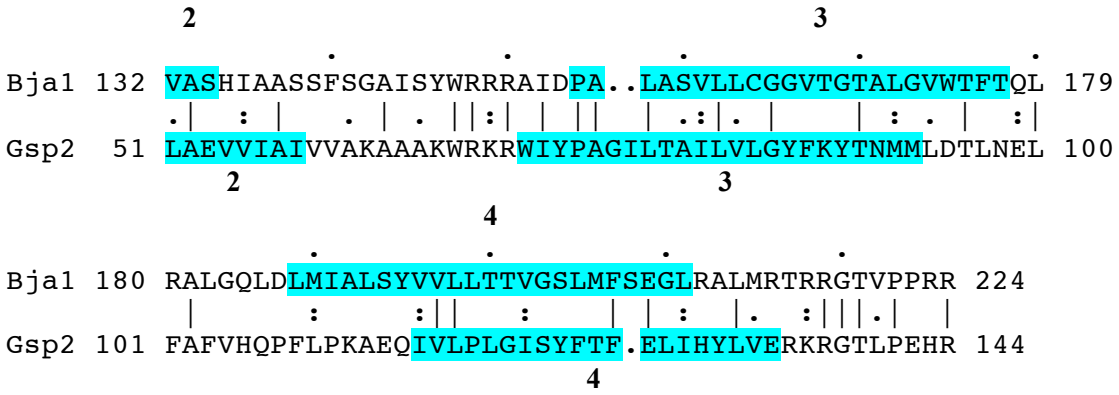

Figure 19: Sequence similarity between members of the TSUP and GUP families. GAP alignment of TMSs 2-4 of TSUP Bja1 (*Bradyrhizobium japonicum*; gi 27376265) with TMSs 2-4 of GUP Gsp2 (*Geobacillus*; gi 261404874). A comparison score of 12.4 S.D. was obtained with 39.1% similarity and 28.3% identity.

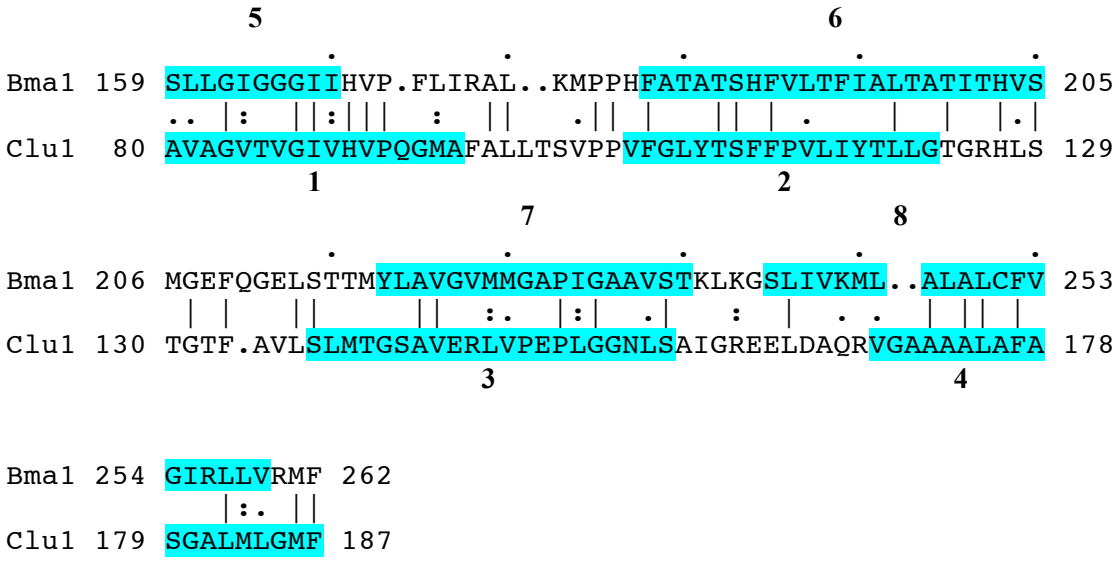

Figure 20: Sequence similarity between members of the TSUP and SulP families. GAP alignment of TMSs 5-8 of TSUP Bam1 (*Blastopirellula marina*; gi 87310558) with TMSs 1-3 of SulP Clu1 (*Canis familiaris*; gi 73968503). A comparison score of 10.7 S.D. was obtained with 40.8% similarity and 34.0% identity.

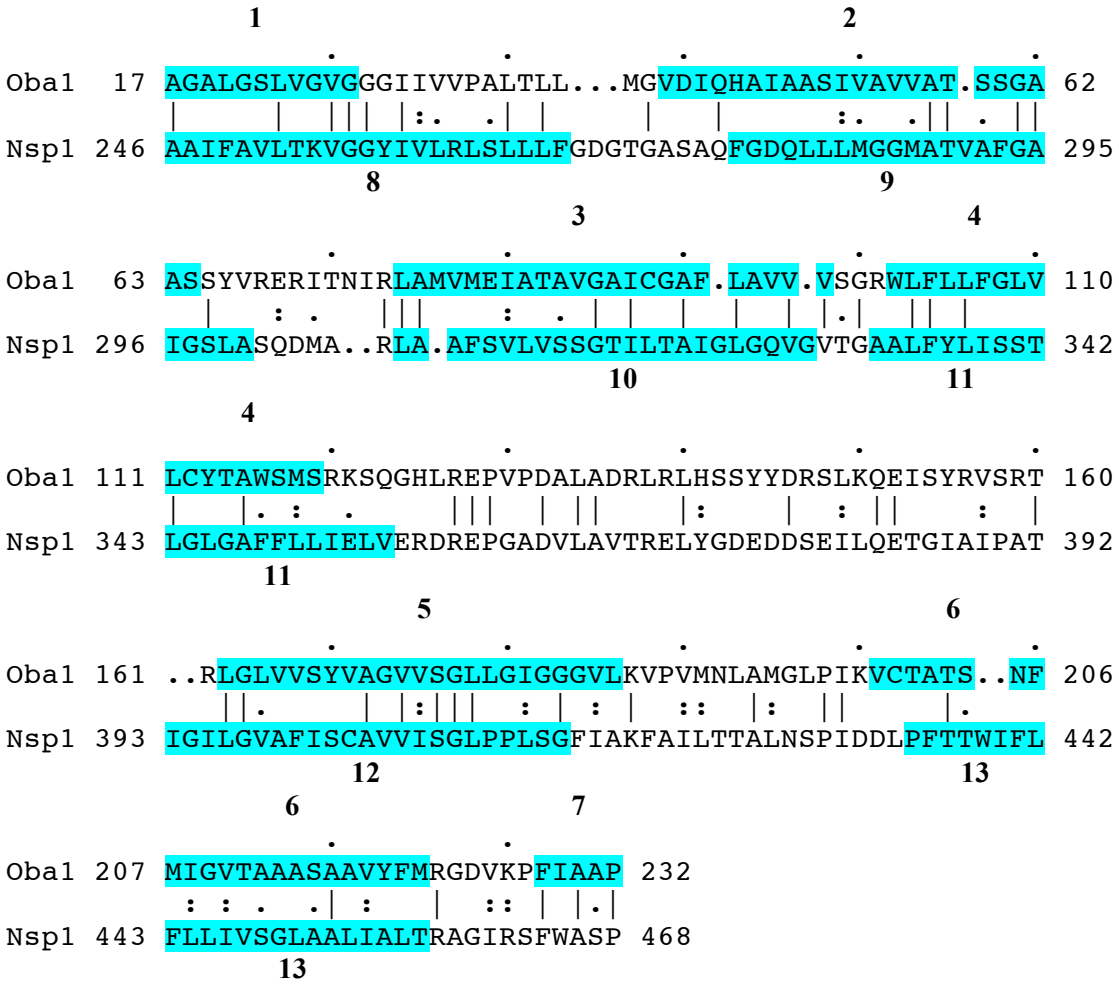

Figure 21: Sequence similarity between members of the TSUP and CPA3 families. GAP alignment of TMSs 1-6 of TSUP Oba1 (*Opitutaceae*; gi 225163757) with TMSs 8-13 of CPA3 Nsp1 (*Nitrobacter*; gi 85714313). A comparison score of 11.3 S.D. was obtained with 36.6% similarity and 27.7% identity.

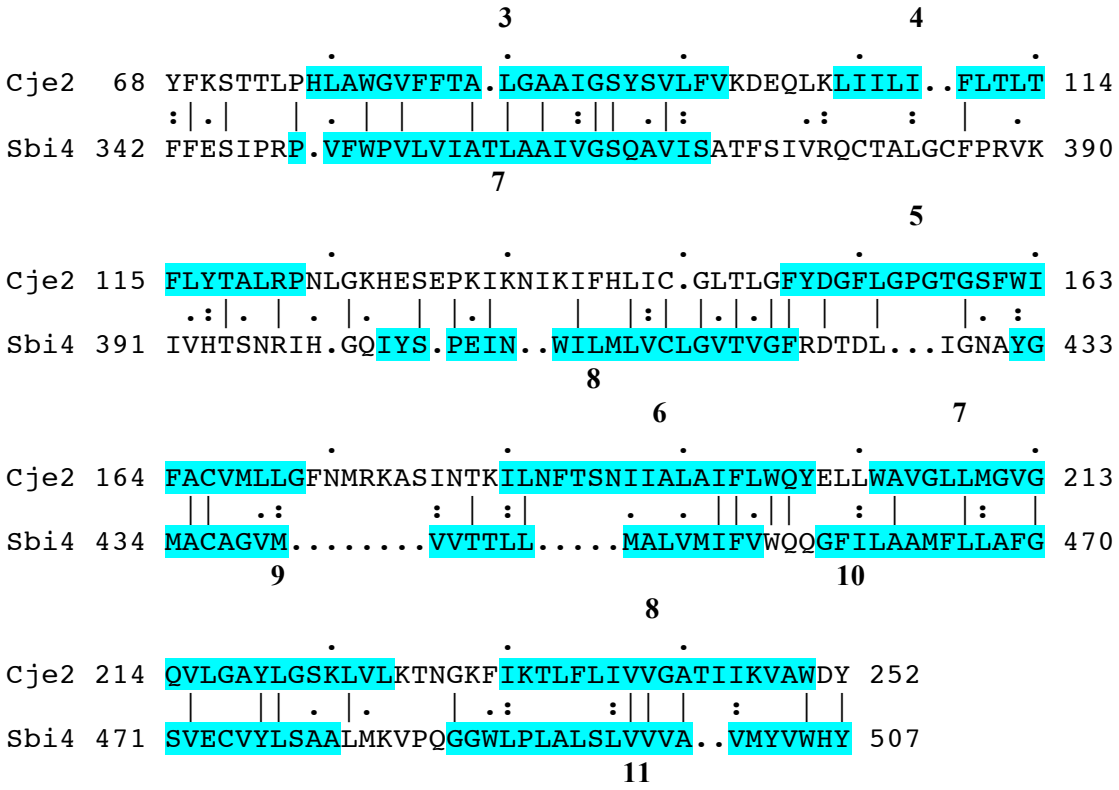

Figure 22: Sequence similarity between members of the TSUP and KUP families. GAP alignment of TMSs 3-8 of TSUP Cje2 (*Campylobacter jejuni*; gi 57238492) with TMSs 7-11 of KUP Sbi4 (*Sorghum bicolor*; gi 242057387). A comparison score of 10.8 S.D. was obtained with 40.1% similarity and 30.2% identity.

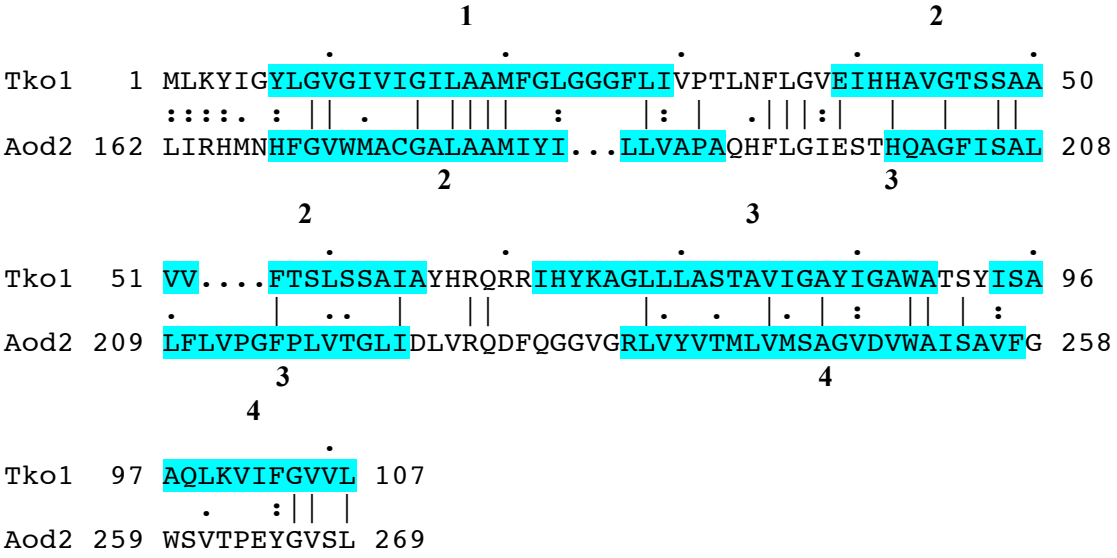

Figure 23: Sequence similarity between members of the TSUP and ThrE families. GAP alignment of TMSs 1-3 of TSUP Tko1 (*Thermococcus kodakarensis*; gi 57640914) with TMSs 2-4 of ThrE Aod2 (*Actinomyces odontolyticus*; gi 293192077). A comparison score of 11.4 S.D. was obtained with 39.4% similarity and 28.8% identity.

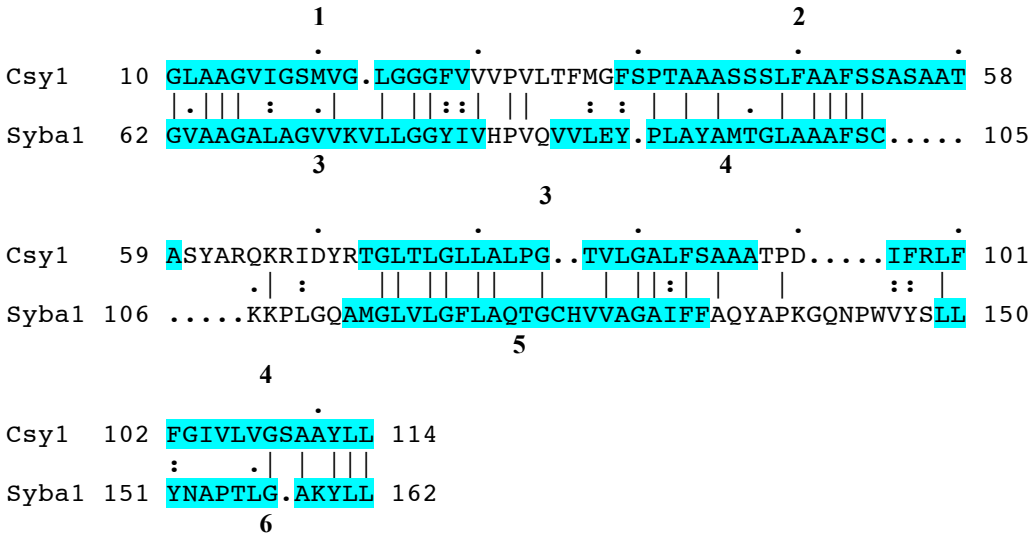

Figure 24: Sequence similarity between members of the TSUP and VUT families. GAP alignment of TMSs 1-4 of TSUP Csy1 (*Cenarchaeum symbiosum*; gi 118576383) with TMSs 3-6 of VUT Syba1 (*Synergistetes* sp. SGP1; gi 295111140). A comparison score of 12.5 S.D. was obtained with 52.7% similarity and 42.0% identity.

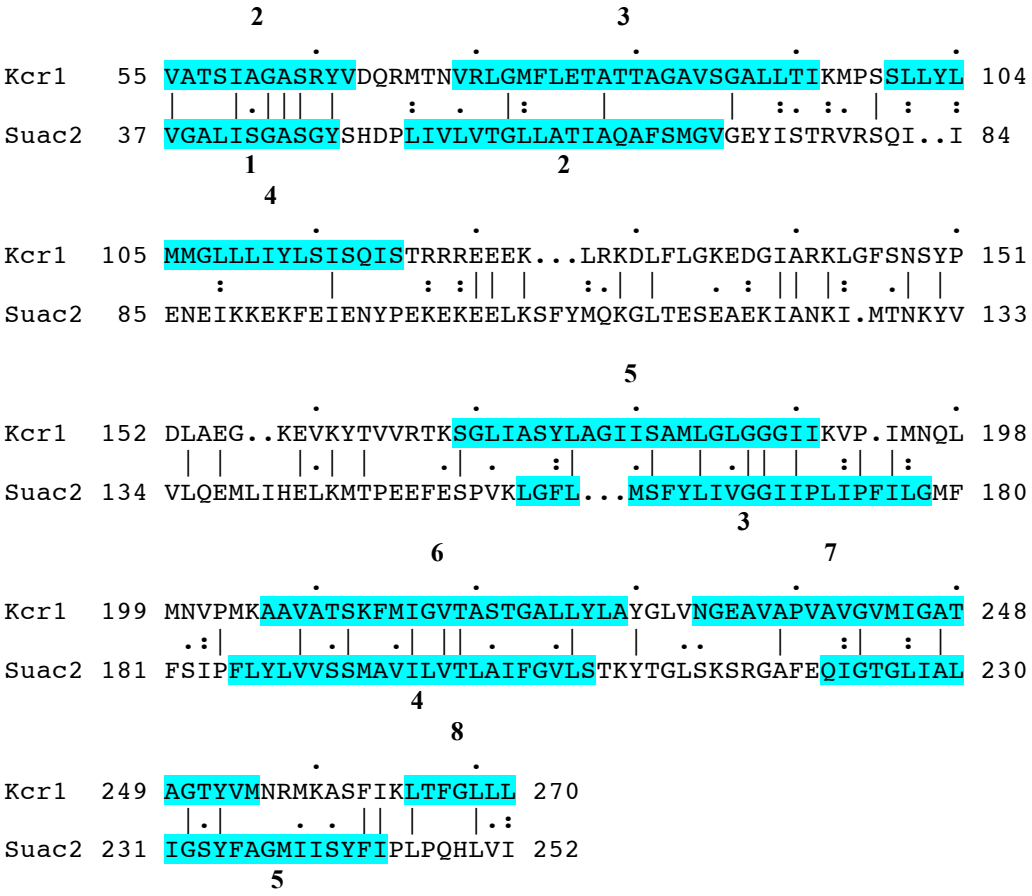

Figure 25: Sequence similarity between members of the TSUP and VIT families. GAP alignment of TMSs 2-7 of TSUP Kcr1 (*Candidatus Korarchaeum cryptofilum*; gi 170290371) with TMSs 1-5 of VIT Suac2 (*Sulfolobus acidocaldarius*; gi 70606117). A comparison score of 12.3 S.D. was obtained with 33.8% similarity and 24.8% identity.

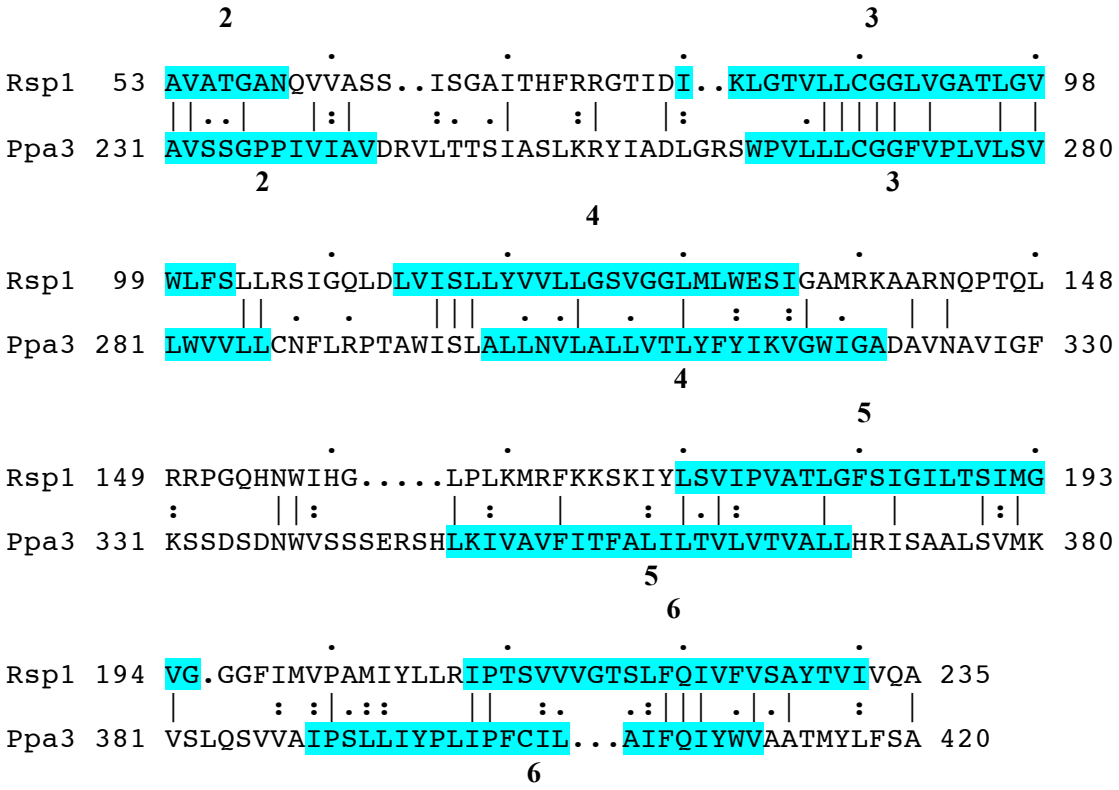

Figure 26: Sequence similarity between members of the TSUP and CTL families. GAP alignment of TMSs 2-6 of TSUP Rsp1 (8 TMSs; *Rhizobium* sp. NGR234; gi 227820754) with TMSs 2-6 of CTL Ppa3 (10 TMSs; *Physcomitrella patens*; gi 168038584). A comparison score of 11.2 S.D. was obtained with 36.1% similarity and 25.6% identity.

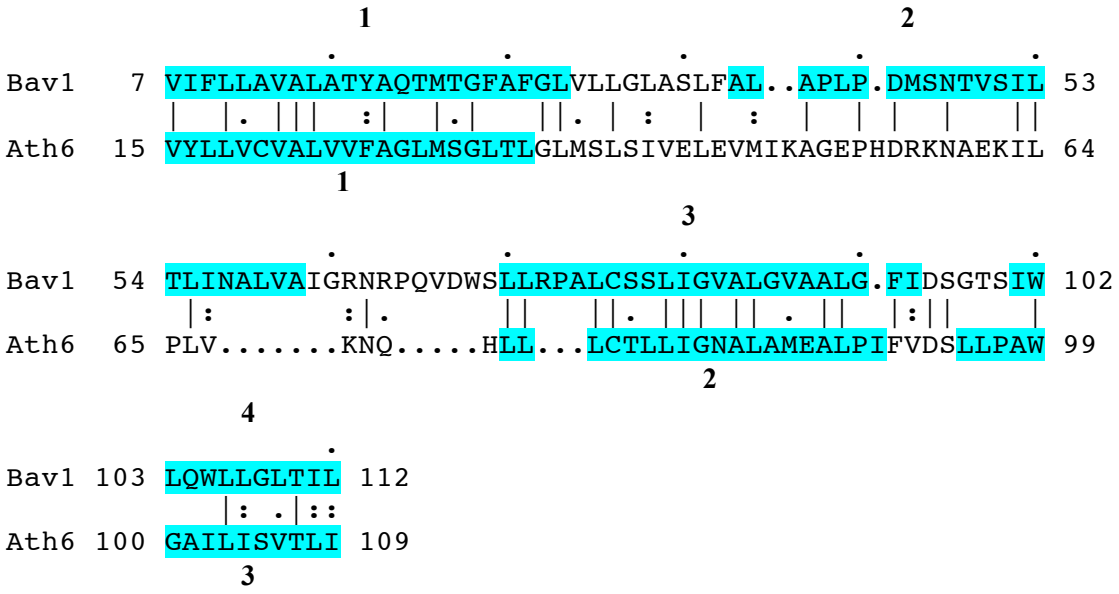

Figure 27: Sequence similarity between members of the TSUP and HCC families. GAP alignment of TMSs 1-4 of TSUP Bav1 (*Bordetella avium*; gi 187478992) with TMSs 1-3 of HCC Ath6 (*Arabidopsis thaliana*; gi 42568492). A comparison score of 11.8 S.D. was obtained with 50.5% similarity and 40.7% identity.

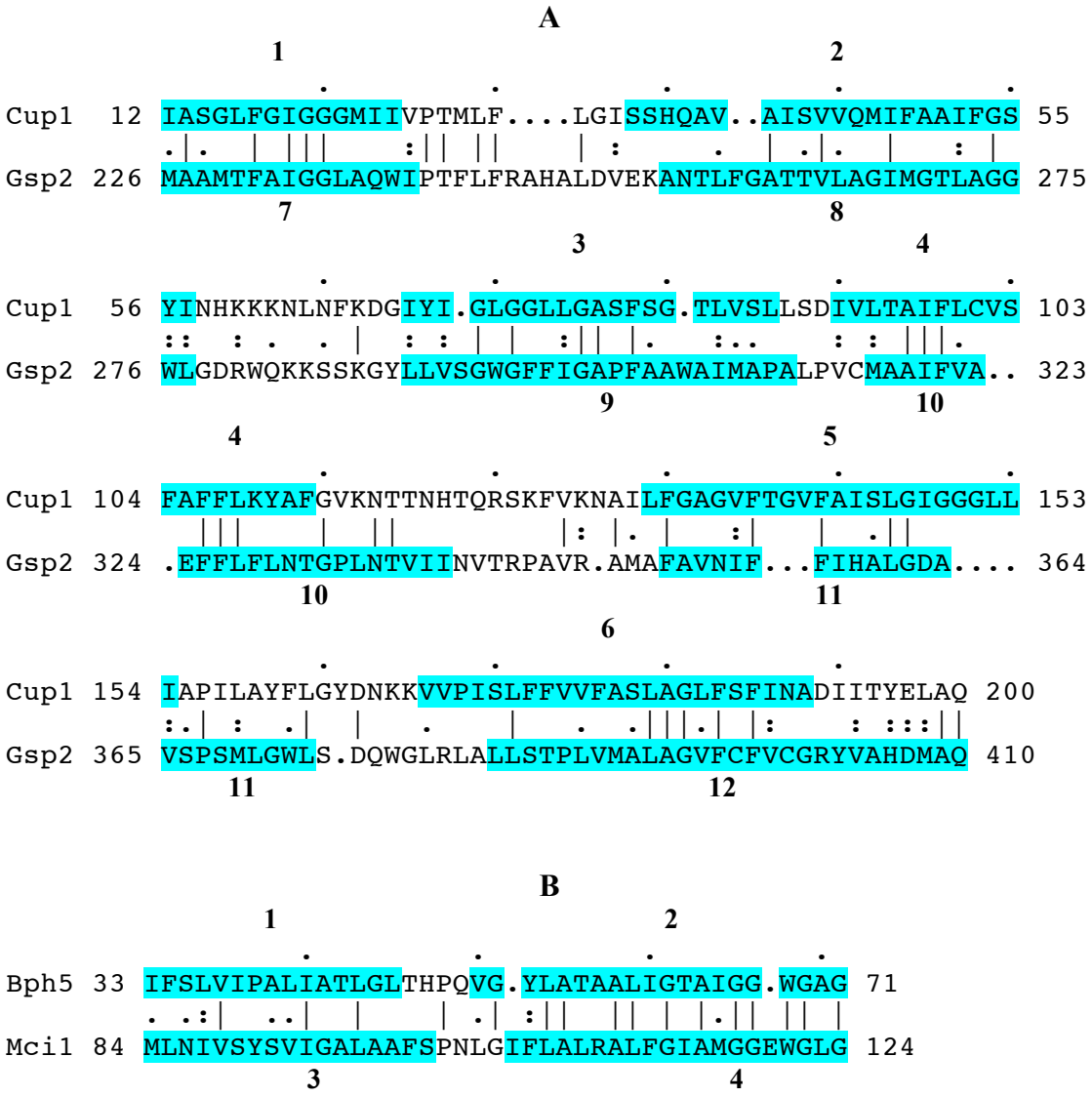

Figure 28: (A) Sequence similarity between members of the TSUP family and the PHL-E family within the MFS superfamily. GAP alignment of TMSs 1-6 of TSUP Cup1 (*Campylobacter upsaliensis*; gi 57242515) with TMSs 7-12 of PHL-E Gsp2 (*Geobacter*; gi 253700633). A comparison score of 11.2 S.D. was obtained with 38.4% similarity and 26.6% identity. **(B)** Sequence similarity between TMSs 1-2 and 3-4 of the MFS. GAP alignment of TMSs 1-2 of PHL-E Bph5 (*Burkholderia phymatum*; gi 186471805) with TMSs 3-4 of PHL-E Gsp2 (*Mesorhizobium ciceri*; gi 319781577). A comparison score of 9.2 S.D. was obtained with 46.2% similarity and 41.0% identity.

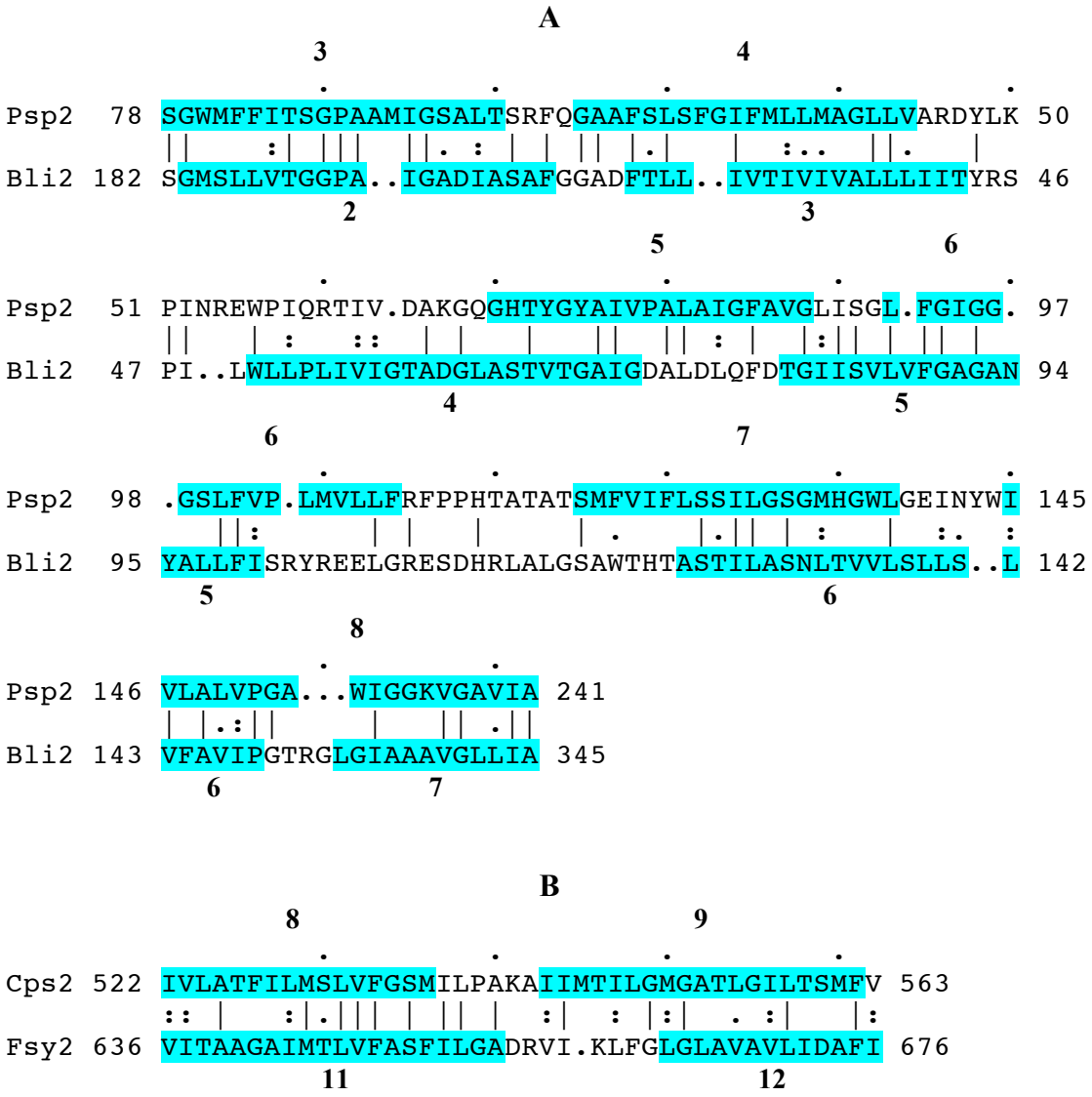

Figure 29: (A) Sequence similarity between members of the TSUP family and the HAE2 family within the RND superfamily. GAP alignment of TMSs 3-8 of TSUP Psp2 (*Paenibacillus*; gi 251794851) with TMSs 2-7 of HAE2 Bli2 (*Brevibacterium linens*; gi 260904806). A comparison score of 13.1 S.D. was obtained with 44.2% similarity and 35.9% identity. **(B)** Sequence similarity between TMSs 8-9 and 11-12 of the RND. GAP alignment of TMSs 8-9 of HAE2 Cps2 (*Corynebacterium pseudotuberculosis*; gi 300859365) with TMSs 11-12 of HAE2 Fsy2 (*Frankia* symbiont of *Datisca glomerata*; gi 289642105). A comparison score of 8.2 S.D. was obtained with 53.7% similarity and 34.1% identity.

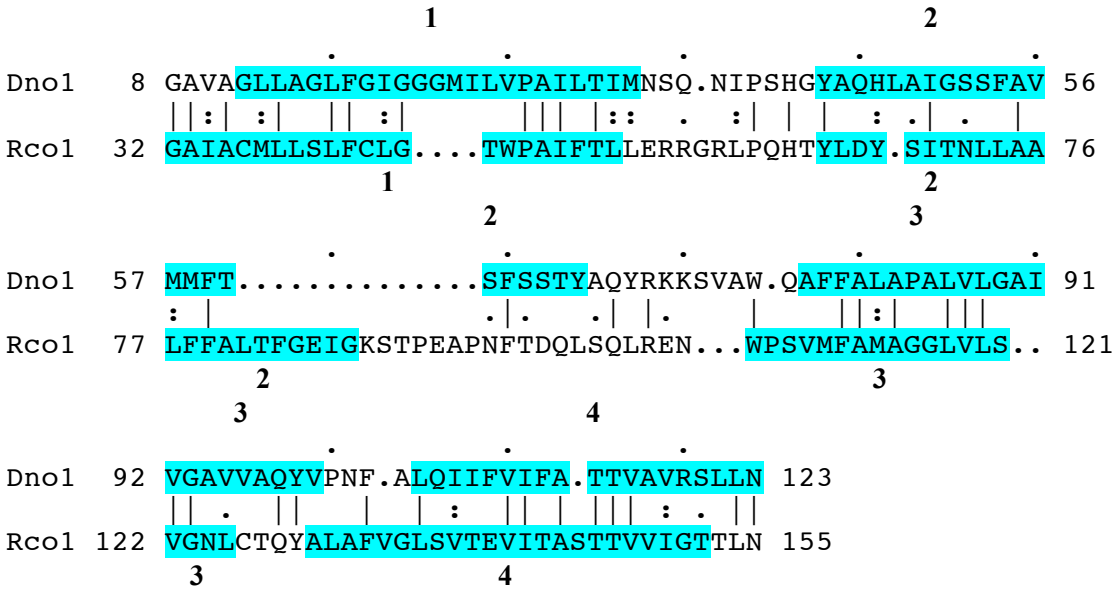

Figure 30: Sequence similarity between members of the TSUP family and the CEO family within the DMT superfamily. GAP alignment of TMSs 1-4 of TSUP Dno1 (*Dichelobacter nodosus*; gi 146329858) with TMSs 1-4 of CEO Rco1 (*Ricinus communis*; gi 255542042). A comparison score of 10.9 S.D. was obtained with 49.1% similarity and 38.7% identity.

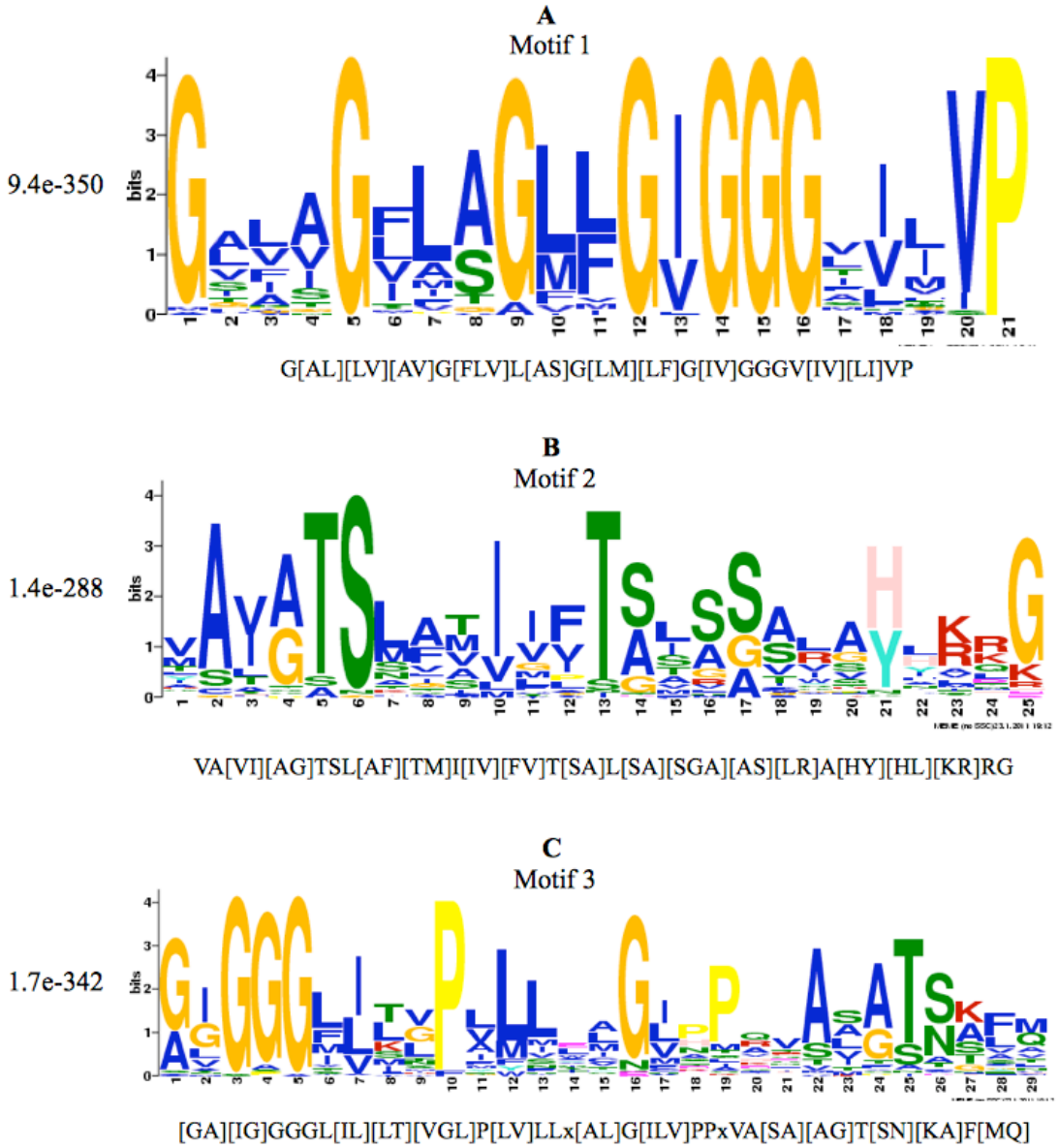

Figure 31: (A-C) The three most well conserved motifs found within the TSUP family as predicted by MEME. Corresponding statistical scores are presented on the y-axis. MAST predictions of motifs based on the MEME results are presented on the x-axis below each motif graphic.

Table 1: 189 TSUP protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig xx). Protein abbreviations, taxonomic origins, protein size (aas), gi numbers, organismal phyla, organismal domain, number of TMSs, and N-terminal orientation are included (see key below). N-terminal orientation results lacking asterisks signify agreement between programs. Average size and standard deviation values are provided for all proteins within clusters having over 1 member. Additional average size and SD values are provided after removing the specified outliers.

* TMS count differs between HMMTOP and TMHMM 2.0. TMHMM 2.0 TMS count and sidedness result used. ~70% of the time, orientation was the same using both programs. There is uncertainty in making the sidedness determination due to the fact that HMMTOP gave different TMS values.

** TMS count same using HMMTOP and TMHMM 2.0. Orientation differed between the two programs. HMMTOP result confirmed using positive inside rule.

*** TMS count same using HMMTOP and TMHMM 2.0. Orientation differed between the two programs. HMMTOP result overturned using positive inside rule.

**** TMS count same using HMMTOP and TMHMM 2.0. Orientation differed between the two programs. HMMTOP result used. Using positive inside rule yielded equal probability of protein having opposite sidedness.

Abbreviation	Organism	Protein Size	GenBank No.	Phylum	Domain	No. of TMSs	N-Term. Orientation
Cluster 1							
Cmu1	*Cryptosporidium muris* RN66	525	209881434	Apicomplexa	Eukaryota	11	Out
Cho1	*Cryptosporidium hominis* TU502	518	67601741	Apicomplexa	Eukaryota	10	In*
Tpa2	*Theileria parva* strain Muguga	409	71033393	Apicomplexa	Eukaryota	9	In*
Ehi1	*Entamoeba histolytica*	460	67466247	none	Eukaryota	10	In*
Ddi1	*Dictyostelium discoideum* AX4	549	66825573	none	Eukaryota	10	In*

Table 1: List of TSUP Homologues Continued

Mbr2	*Monosiga brevicollis* MX1	499	167525260	Codonosigidae	Eukaryota	8	Out
Tva1	*Trichomonas vaginalis* G3	448	123437805	Trichomonada	Eukaryota	9	In*
Ath3	*Arabidopsis thaliana*	491	6554197	Viridiplantae	Eukaryota	9	In*
Ath5	*Arabidopsis thaliana*	431	2911082	Viridiplantae	Eukaryota	9	In*
Sbi3	*Sorghum bicolor*	473	242058941	Viridiplantae	Eukaryota	9	In
Ppa1	*Physcomitrella patens* subsp. *patens*	405	168065030	Viridiplantae	Eukaryota	9	Out
Gma1	*Glycine max*	469	83853809	Viridiplantae	Eukaryota	9	In**
Psi1	*Picea sitchensis*	505	148906357	Viridiplantae	Eukaryota	10	In*
Sbi2	*Sorghum bicolor*	383	242044420	Viridiplantae	Eukaryota	7	In
Mpu1	*Micromonas pusilla* CCMP1545	461	226458924	Viridiplantae	Eukaryota	8	Out*
Tps2	*Thalassiosira pseudonana* CCMP1335	564	223992571	Bacillariophyta	Eukaryota	11	Out*
Tps3	*Thalassiosira pseudonana* CCMP1335	385	223998204	Bacillariophyta	Eukaryota	7	In*
Ptr2	*Phaeodactylum tricornutum* CCAP 1055/1	644	219112381	Bacillariophyta	Eukaryota	11	Out
Pma6	*Perkinsus marinus* ATCC 50983	385	239878631	Perkinsidae	Eukaryota	6	In*
Tgo3	*Toxoplasma gondii* GT1	482	221485444	Apicomplexa	Eukaryota	10	In*
Tps1	*Thalassiosira pseudonana* CCMP1335	522	224014684	Bacillariophyta	Eukaryota	9	In
Tgo4	*Toxoplasma gondii* VEG	665	221505087	Apicomplexa	Eukaryota	11	In

Table 1: List of TSUP Homologues Continued

Bsa1	*Bodo saltans*	526	206598109	Bodonidae	Eukaryota	10	Out
Lma1	*Leishmania* major strain Friedlin	511	157873729	Trypanosomatidae	Eukaryota	10	Out*
Mbr1	*Monosiga brevicollis* MX1	512	167521960	Codonosigidae	Eukaryota	9	In
Pte6	*Paramecium tetraurelia* strain d4-2	491	145483119	Oligohymenophorea	Eukaryota	9	Out**
Pte8	*Paramecium tetraurelia* strain d4-2	473	145514235	Oligohymenophorea	Eukaryota	9	Out**
Pte4	*Paramecium tetraurelia* strain d4-2	424	145501808	Oligohymenophorea	Eukaryota	10	In
Pte1	*Paramecium tetraurelia* strain d4-2	441	145493138	Oligohymenophorea	Eukaryota	10	In***
Tth2	*Tetrahymena thermophila*	570	118348626	Oligohymenophorea	Eukaryota	9	In*
Pte7	*Paramecium tetraurelia* strain d4-2	430	145531341	Oligohymenophorea	Eukaryota	10	Out*
Tth1	*Tetrahymena thermophila*	505	146183328	Oligohymenophorea	Eukaryota	11	In*
Tth3	*Tetrahymena thermophila*	503	118395416	Oligohymenophorea	Eukaryota	10	In*
Tth4	*Tetrahymena thermophila*	1325	118401229	Oligohymenophorea	Eukaryota	8	Out
Pte5	*Paramecium tetraurelia* strain d4-2	400	145528512	Oligohymenophorea	Eukaryota	10	In*
Pte2	*Paramecium tetraurelia* strain d4-2	406	145538953	Oligohymenophorea	Eukaryota	9	In*
Osa2	*Oryza sativa* Japonica Group	465	222625716	Viridiplantae	Eukaryota	5	In**
Cre1	*Chlamydomonas reinhardtii*	929	159479540	Viridiplantae	Eukaryota	6	In

Table 1: List of TSUP Homologues Continued

Gla2	*Giardia lamblia* ATCC 50803	748	159117352	Hexamitidae	Eukaryota	10	Out*
Tgo2	*Toxoplasma gondii* VEG	299	221501858	Apicomplexa	Eukaryota	2	Out*
Cre2	*Chlamydomonas reinhardtii*	1854	159469083	Viridiplantae	Eukaryota	9	In*
Pte3	*Paramecium tetraurelia* strain d4-2	454	145493226	Oligohymenophorea	Eukaryota	7	Out*
Tgo1	*Toxoplasma gondii* GT1	1659	221487433	Apicomplexa	Eukaryota	8	Out*
	Average Size= 572 +/- 312 (all)						
	Average Size= 476 +/- 71 (w/out Tth4, Cre1, Gla2, Cre2, Tgo1)						
Cluster 2							
Ssp3	*Sphingomonas sp.* SKA58	259	94497264	Alphaproteobacteria	Bacteria	7	In*
Rsp4	*Ruegeria* sp. TM1040	252	99082858	Alphaproteobacteria	Bacteria	5	Out*
Pas1	*Photorhabdus asymbiotica* subsp. *asymbiotica* ATCC 43949	260	211638062	Gammaproteobacteria	Bacteria	7	Out
Vpa2	*Variovorax paradoxus* S110	270	239813891	Betaproteobacteria	Bacteria	7	In*
Dno1	*Dichelobacter nodosus* VCS1703A	258	146329063	Gammaproteobacteria	Bacteria	6	In*
Sna1	*Stackebrandtia nassauensis* DSM 44728	256	229865833	Actinobacteria	Bacteria	6	In*
Par1	*Psychrobacter arcticus* 273-4	251	71066392	Gammaproteobacteria	Bacteria	5	Out*
Hca1	*Helicobacter canadensis* MIT 98-5491	250	224418685	Epsilonproteobacteria	Bacteria	7	Out**
Eta1	*Edwardsiella tarda*	280	158512112	Gammaproteobacteria	Bacteria	6	In*
Msp4	*Marinomonas* sp. MED121	258	87120732	Gammaproteobacteria	Bacteria	7	Out

Table 1: List of TSUP Homologues Continued

Orf2	*gamma proteobacterium*	250	90416226	Gammaproteobacteria	Bacteria	6	Out**
Cje1	*Campylobacter jejuni* RM1221	254	57238492	Epsilonproteobacteria	Bacteria	7	In*
Orf5	*bacterium Ellin514*	268	223939838	Verrucomicrobia	Bacteria	6	Out**
Msp1	*Marinomonas* sp. MED121	253	87118707	Gammaproteobacteria	Bacteria	8	In*
Ftu1	*Francisella tularensis* subsp. *holarctica* FSC200	366	167010238	Gammaproteobacteria	Bacteria	8	In*
Avi1	*Azotobacter vinelandii* DJ	289	226943937	Gammaproteobacteria	Bacteria	7	In
Psp4	*Psychromonas* sp. CNPT3	274	90407709	Gammaproteobacteria	Bacteria	8	In*
Nar1	*Novosphingobium aromaticivorans* DSM 12444	256	87200262	Alphaproteobacteria	Bacteria	6	In*
Ade1	*Anaeromyxobacter dehalogenans* 2CP-C	254	86157393	Deltaproteobacteria	Bacteria	7	Out*
Dpi1	*Desulfovibrio piger* ATCC 29098	259	212704568	Deltaproteobacteria	Bacteria	7	In
Fva1	*Fusobacterium varium* ATCC 27725	276	253583632	Fusobacteria	Bacteria	7	In*
Tps5	*Thermoanaerobacter pseudethanolicus* ATCC 33223	253	167038325	Firmicutes	Bacteria	7	In
Bmu1	*Brachyspira murdochii* DSM 12563	255	227999578	Spirochaetes	Bacteria	6	In*
Vpa1	*Veillonella parvula* DSM 2008	264	227372642	Firmicutes	Bacteria	8	In*
Ban1	*Bacillus anthracis* str. A2012	263	65318350	Firmicutes	Bacteria	8	In*
Taf1	*Thermosipho africanus* TCF52B	254	217077973	Thermotogae	Bacteria	7	Out

Table 1: List of TSUP Homologues Continued

Tde1	*Treponema denticola* ATCC 35405	262	42525707	Spirochaetes	Bacteria	7	In*
Cbo6	*Clostridium bolteae*	251	160940895	Firmicutes	Bacteria	7	Out*
Cac1	*Cloacamonas acidaminovorans*	257	218961280	none	Bacteria	7	In*
Psy1	*Pseudomonas syringae pv. oryzae*	258	237801487	Gammaproteobacteria	Bacteria	6	Out*
Ttu1	*Teredinibacter turnerae* T7901	256	237685094	Gammaproteobacteria	Bacteria	7	In
Orf4	*uncultured marine bacterium* 439	252	40062756	none	Bacteria	7	In
Gbe1	*Granulibacter bethesdensis* CGDNIH1	253	114328287	Alphaproteobacteria	Bacteria	7	Out*
Lho1	*Laribacter hongkongensis* HLHK9	310	226942144	Betaproteobacteria	Bacteria	6	In*
Rco1	*Ricinus communis*	301	223512929	Viridiplantae	Eukaryota	8	Out*
Nmu1	*Neisseria mucosa* ATCC 25996	256	225367635	Betaproteobacteria	Bacteria	7	In*
Sli1	*Spirosoma linguale* DSM 74	254	229867512	Bacteroidetes	Bacteria	6	In*
Cps1	*Corynebacterium pseudogenitalium* ATCC 33035	260	227490282	Actinobacteria	Bacteria	8	In*
Jde1	*Jonesia denitrificans* DSM 20603	269	227383462	Actinobacteria	Bacteria	7	In*
Pac1	*Propionibacterium acnes* KPA171202	255	50842975	Actinobacteria	Bacteria	6	In*
	Average Size= 264 +/- 21 (all)						
Cluster 3							
Mka1	*Methanopyrus kandleri* AV19	252	20093583	Euryarchaeota	Archaea	7	In*

Table 1: List of TSUP Homologues Continued

Dde1	*Desulfovibrio desulfuricans*	379	220904085	Deltaproteobacteria	Bacteria	9	In*
Dth2	*Desulfonatronospira thiodismutans* ASO3-1	254	225199785	Deltaproteobacteria	Bacteria	7	In*
Rca1	*Roseiflexus castenholzii*	251	156743559	Chloroflexi	Bacteria	8	In
Orf6	*uncultured bacterium*	654	239787713	none	Bacteria	9	In

Average Size= 358 +/- 174 (all)
Average Size= 252 +/- 2 (w/out Orf6, Dde1)

Cluster 4

Iho1	*Ignicoccus hospitalis* KIN4/I	240	156936864	Crenarchaeota	Archaea	7	Out*
Bsp1	*Beggiatoa* sp. PS	787	153869281	Gammaproteobacteria	Bacteria	10	In*
Cbu3	*Coxiella burnetii Dugway* 5J108-111	274	209364180	Gammaproteobacteria	Bacteria	8	In
Cgl1	*Chryseobacterium gleum* ATCC 35910	505	227369714	Bacteroidetes	Bacteria	8	In
Gla1	*Giardia lamblia* ATCC 50803	520	159115095	Hexamitidae	Eukaryota	10	Out**
Min1	*Methylacidiphilum infernorum* V4	267	189218632	Verrucomicrobia	Bacteria	8	In

Average Size= 432 +/- 214 (all)
Average Size= 260 +/- 18 (w/out Bsp1, Cgl1, Gla1)

Cluster 5

She1	*Slackia heliotrinireducens* DSM 20476	277	229879562	Actinobacteria	Bacteria	8	In
Ele1	*Eggerthella lenta* DSM 2243	307	227411437	Actinobacteria	Bacteria	8	In

Table 1: List of TSUP Homologues Continued

Rxy1	*Rubrobacter xylanophilus* DSM 9941	267	108803101	Actinobacteria	Bacteria	8	In***
Sfu1	*Syntrophobacter fumaroxidans*	269	116750841	Deltaproteobacteria	Bacteria	7	Out
Ptr3	*Phaeodactylum tricornutum*	2798	219127009	Bacillariophyta	Eukaryota	4	Out*
Mch1	*Microcoleus chthonoplastes*	267	224407624	Cyanobacteria	Bacteria	8	In
Ter2	*Trichodesmium erythraeum*	305	113475233	Cyanobacteria	Bacteria	8	In
Ssp1	*Synechococcus* sp. JA-3-3Ab	317	86606127	Cyanobacteria	Bacteria	8	In
Pca1	*Pyrobaculum calidifontis* JCM 11548	244	126458964	Crenarchaeota	Archaea	8	In*
	Average Size= 561 +/- 839 (all)						
	Average Size= 282 +/- 25 (w/out Ptr3)						
Cluster 6							
Sus1	*Candidatus Solibacter usitatus* Ellin6076	281	116624708	Acidobacteria	Bacteria	8	In***
Sth2	*Symbiobacterium thermophilum* IAM 14863	279	51892120	Firmicutes	Bacteria	8	In*
Bsu1	*Brucella suis* 1330	289	23500891	Alphaproteobacteria	Bacteria	8	In
Ooe1	*Oenococcus oeni* PSU-1	283	116491798	Firmicutes	Bacteria	8	In*
Pto1	*Picrophilus torridus* DSM 9790	333	48478318	Euryarchaeota	Archaea	8	In
Sth1	*Sphaerobacter thermophilus*	282	229877687	Chloroflexi	Bacteria	8	Out
Mth2	*Moorella thermoacetica*	271	83589239	Firmicutes	Bacteria	8	In*
Kcr1	*Candidatus Korarchaeum cryptofilum* OPF8	285	170290371	Korarchaeota	Archaea	8	In*

Table 1: List of TSUP Homologues Continued

Mxa1	*Myxococcus xanthus* DK 1622	260	108758495	Deltaproteobacteria	Bacteria	7	In
Dra1	*Deinococcus radiodurans* R1	255	15805571	Deinococcus-Thermus	Bacteria	8	In
Sma1	*Staphylothermus marinus* F1	250	126465319	Crenarchaeota	Archaea	8	In
Dac1	*Denitrovibrio acetiphilus*	274	227423788	Deferribacteres	Bacteria	7	In*
Emi1	*Elusimicrobium minutum* Pei191	275	187251557	candidate division TG1	Bacteria	7	In*
Hbu2	*Hyperthermus butylicus*	255	124028506	Crenarchaeota	Archaea	8	In*
	Average Size= 277 +/- 20 (all)						
Cluster 7							
Bad1	*Bifidobacterium adolescentis* ATCC 15703	292	119026567	Actinobacteria	Bacteria	8	In*
Gva1	*Gardnerella vaginalis* ATCC 14019	267	227507357	Actinobacteria	Bacteria	8	In
	Average Size= 280 +/- 18 (all)						
Cluster 8							
Asa1	*Aliivibrio salmonicida*	279	16605593	Gammaproteobacteria	Bacteria	8	In*
Rru1	*Rhodospirillum rubrum*	276	83592684	Alphaproteobacteria	Bacteria	8	In
Rsp3	*Roseovarius* sp. HTCC2601	274	114764120	Alphaproteobacteria	Bacteria	8	In
Rsp1	*Ruegeria* sp. TM1040	278	99080207	Alphaproteobacteria	Bacteria	7	In
Msp3	*Magnetococcus* sp. MC-1	265	117925601	Proteobacteria	Bacteria	6	In*
Fpe1	*Fulvimarina pelagi* HTCC2506	275	114707272	Alphaproteobacteria	Bacteria	7	In

Table 1: List of TSUP Homologues Continued

Bja2	*Bradyrhizobium japonicum* USDA 110	287	27375621	Alphaproteobacteria	Bacteria	6	In
Hne1	*Hyphomonas neptunium* ATCC 15444	314	114797241	Alphaproteobacteria	Bacteria	9	In
Cbu2	*Coxiella burnetii* RSA 331	275	161831015	Gammaproteobacteria	Bacteria	7	In
Lsp1	*Limnobacter* sp. MED105	278	149925520	Betaproteobacteria	Bacteria	7	In*
Nmo1	*Nitrococcus mobilis* Nb-231	266	88811005	Gammaproteobacteria	Bacteria	8	Out
Mca1	*Methylococcus capsulatus* str. Bath	294	53802665	Gammaproteobacteria	Bacteria	6	In
Pir1	*Polaribacter irgensii* 23-P	281	88803086	Bacteroidetes	Bacteria	7	In
Kko1	*Kangiella koreensis* DSM 16069	268	227997603	Gammaproteobacteria	Bacteria	8	Out****
Har1	*Herminiimonas arsenicoxydans*	287	134096092	Betaproteobacteria	Bacteria	7	In*
Swo1	*Shewanella woodyi* ATCC 51908	268	170728324	Gammaproteobacteria	Bacteria	8	Out**
Ptu1	*Pseudoalteromonas tunicata* D2	269	88860323	Gammaproteobacteria	Bacteria	8	In***
Ama2	*Alteromonas macleodii 'Deep ecotype'*	268	196158505	Gammaproteobacteria	Bacteria	7	In*
Sbe1	*Shewanella benthica* KT99	267	163752420	Gammaproteobacteria	Bacteria	8	Out**
Msu1	*Mannheimia succiniciproducens* MBEL55E	266	52424462	Gammaproteobacteria	Bacteria	8	Out
Psp3	*Photobacterium* sp. SKA34	267	89072545	Gammaproteobacteria	Bacteria	8	Out**
Afe1	*Acidithiobacillus ferrooxidans* ATCC 23270	264	218665563	Gammaproteobacteria	Bacteria	8	In*

Table 1: List of TSUP Homologues Continued

Pne1	*Polynucleobacter necessarius* subsp. *asymbioticus* QLW-P1DMWA-1	272	145589361	Betaproteobacteria	Bacteria	8	In*
Lsp2	*Limnobacter* sp. MED105	289	149926219	Betaproteobacteria	Bacteria	8	In*
Eco1	*Eikenella corrodens* ATCC 23834	270	225024689	Betaproteobacteria	Bacteria	7	In**
Ama1	*Acaryochloris marina*	278	158336922	Cyanobacteria	Bacteria	9	In
Tsp1	*Thioalkalivibrio* sp. K90mix	268	224818668	Gammaproteobacteria	Bacteria	8	In
Rso1	*Ralstonia solanacearum*	273	17549483	Betaproteobacteria	Bacteria	8	In***
Ppe1	*Proteus penneri* ATCC 35198	271	226330327	Gammaproteobacteria	Bacteria	8	Out
Iba1	*Idiomarina baltica* OS145	264	85713215	Gammaproteobacteria	Bacteria	8	In
	Average Size= 275 +/- 11 (all)						
Cluster 9							
Epe1	*Endoriftia persephone 'Hot96_1+Hot96_2'*	287	167948520	Gammaproteobacteria	Bacteria	5	In*
Cluster 10							
Pal1	*Providencia alcalifaciens* DSM 30120	271	212712467	Gammaproteobacteria	Bacteria	8	Out**
Cluster 11							
Orf3	*uncultured archaeon* GZfos34A6	276	52549977	none	Archaea	8	Out**

Table 1: List of TSUP Homologues Continued

Mma1	*Methanosarcina mazei* Go1	270	21228951	Euryarchaeota	Archaea	8	In*
Mma2	*Methanococcus maripaludis* S2	270	45358505	Euryarchaeota	Archaea	8	Out**
	Average Size= 272 +/- 3 (all)						
Cluster12							
Tko1	*Thermococcus kodakarensis* KOD1	254	57640914	Euryarchaeota	Archaea	7	In*
Tba1	*Thermococcus barophilus* MP	251	223475524	Euryarchaeota	Archaea	8	In
Mbo1	*Methanoregula boonei 6A8*	269	154149849	Euryarchaeota	Archaea	8	In
Sma2	*Staphylothermus marinus* F1	265	126466107	Crenarchaeota	Archaea	8	In
	Average Size= 260 +/- 9 (all)						
Cluster 13							
Tte1	*Thermobaculum terrenum* ATCC BAA-798	255	227375491	none	Bacteria	8	In
Cbe1	*Clostridium beijerinckii* NCIMB 8052	272	150017843	Firmicutes	Bacteria	8	In*
Vdi1	*Veillonella dispar* ATCC 17748	264	238018311	Firmicutes	Bacteria	8	In*
Nma1	*Nitrosopumilus maritimus* SCM1	257	161528556	Crenarchaeota	Archaea	7	In*
Bbr2	*Brevibacillus brevis*	274	226314422	Firmicutes	Bacteria	7	In*
Gka1	*Geobacillus kaustophilus* HTA426	300	56421519	Firmicutes	Bacteria	8	In*
Bcl1	*Bacillus clausii* KSM-K16	272	56964722	Firmicutes	Bacteria	8	In*

Table 1: List of TSUP Homologues Continued

Psp2	*Paenibacillus* sp. JDR-2	272	251794851	Firmicutes	Bacteria	8	In*
Oih1	*Oceanobacillus iheyensis*	285	23099829	Firmicutes	Bacteria	8	In*
Sau1	*Staphylococcus aureus*	275	15923912	Firmicutes	Bacteria	8	In
	Average Size= 273 +/- 13 (all)						
Cluster 14							
Bja1	*Bradyrhizobium japonicum* USDA 110	380	27376265	Alphaproteobacteria	Bacteria	8	In
Ssp5	*Sphingomonas* sp. SKA58	304	94498747	Alphaproteobacteria	Bacteria	7	Out*
Pth1	*Pelotomaculum thermopropionicum* SI	299	147678596	Firmicutes	Bacteria	7	In*
Dau1	*Desulforudis audaxviator*	394	169832116	Firmicutes	Bacteria	8	In*
Dre4	*Desulfotomaculum reducens* MI-1	426	134299284	Firmicutes	Bacteria	9	In*
Abo1	*Aciduliprofundum boonei* T469	254	223473124	Euryarchaeota	Archaea	8	Out**
Lbi1	*Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'*	325	183219704	Spirochaetes	Bacteria	8	In*
	Average Size= 340 +/- 61 (all)						
	Average Size= 296 +/- 30 (w/out Bja1, Dau1, Dre4)						
Cluster 15							
Orf1	*synthetic construct*	284	62258462	none	Unclassified	8	In*
Vsp2	*Verrucomicrobium spinosum* DSM 4136	264	171915322	Verrucomicrobia	Bacteria	8	Out*

Table 1: List of TSUP Homologues Continued

Rme1	*Ralstonia metallidurans* CH34	268	94311333	Betaproteobacteria	Bacteria	8	Out
Pre1	*Providencia rettgeri* DSM 1131	264	223992411	Gammaproteobacteria	Bacteria	8	In*
Sso1	*Sulfolobus solfataricus* P2	293	15899038	Crenarchaeota	Archaea	8	In*
Mmu1	*Mobiluncus mulieris* 35243	361	227876711	Actinobacteria	Bacteria	9	In
Aau1	*Arthrobacter aurescens* TC1	300	119952309	Actinobacteria	Bacteria	8	In
Lmo1	*Listeria monocytogenes* EGD-e	246	16802663	Firmicutes	Bacteria	8	In*
Ste2	*Sebaldella termitidis* ATCC 33386	246	229881273	Fusobacteria	Bacteria	7	In
Bsp2	*Bacillus* sp. B14905	282	126650500	Firmicutes	Bacteria	8	In*
Cph2	*Chlorobium phaeobacteroides*	408	189499528	Chlorobi	Bacteria	9	Out
Afu1	*Archaeoglobus fulgidus*	325	11499708	Euryarchaeota	Archaea	7	Out**
Dha1	*Desulfitobacterium hafniense* DCB-2	312	219669180	Firmicutes	Bacteria	7	Out**
Dre3	*Desulfohalobium retbaense*	569	227420936	Deltaproteobacteria	Bacteria	7	In*

Average Size= 316 +/- 86 (all)
Average Size= 280 +/- 26 (w/out Mmu1, Cph2, Dre3)

Table 2: Summary table of Microbial Rhodopsin superfamily members. The family name, abbreviation, typical size range, dominant topology and organismal source are presented. B - bacteria. E - eukaryotes. A - archaea.

Family #	Family Name	Abbn.	Size Range (aas)	Dominant Topology	Organismal Source
1	Ion-Translocating Microbial Rhodopsin	MR	250-350	7	B E A
2	4-Toluene Sulfonate Uptake Permease	TSUP	250-600	8	B E A
3	Lysosomal Cystine Transporter	LCT	300-400	7	E
4	Ni^{2+}-Co^{2+} Transporter	NiCoT	300-380	6	B E A
5	Branched Chain Amino Acid Exporter	LIV-E	230-270	7	B E A
6	Organic Solute Transporter	OST	330-360	7	B E

Table 3: Summary of functional predictions made for each phylogenetic cluster presented in Figure 1.

Cluster #	Proposed Functions
1	FeS cluster assembly
	Transport of sulfur-based compounds
2	Transport of peptides/amino acids and nitrite/nitrate
3	Stress response: oxidative, heat and metabolic
	Nitrogen-based compound transport
	Arsenate/arsenite resistance
4	Cofactor synthesis
	Stress response: oxidative
5	FeS cluster assembly
	Transport of sulfur-based compounds
	Transport of peptides/amino acids
	Stress response
6	Unclear
	Transport of sulfur-based compounds
7	Stress response: heat
	Transport of phosphate
	Transport of peptides/amino acids and nitrite/nitrate
8	Transport of sulfur-based compounds
	Lipid or lipoprotein transport
	Nucleic acid uptake
9	None
10	None
11	Cofactor synthesis: Co^{2+}, Ni^{2+} and riboflavin transport
	Transport of amino acids
12	Substrate transport for tRNA modification
	Transport of amino acids
	Transport of sulfur-based compounds
	Transport of NAD components
13	Sulfite uptake
	Iron uptake
	Transport of peptides/amino acids
	Extrusion of sulfate and/or phosphate
14	Extrusion of cyanide/cyanate or tungstate/vanadate/sulfate
	Uptake of sulfur-based compounds
	Nitrogen-based compound transport
15	Transport of sulfur-based compounds
	Extrusion of sulfite

References

Aguilar-Barajas, E., Díaz-Pérez, M.I., Ramírez-Díaz, H., Riveros-Rosas, H., Cervantes, C. (2011) Bacterial transport of sulfate, molybdate, and related oxyanions. Biometals. [Epub: ahead of print].

Alexeyev, M.F. and Winkler, H.H. (1999) Membrane topology of the *Rickettsia prowazekii* ATP/ADP translocase revealed by novel dual *pho-lac* reporters. J. Mol. Biol. **285**: 1503-1513.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. **215**: 403-410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**: 3389-3402.

Anderson, P.M., Sung, Y., Fuchs, J.A. (1990) The cyanase operon and cyanate metabolism. FEMS Microbiol. Letters. **87**: 247-252.

Bailey, T.L., Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology, pages 28-36, AAAI Press, Menlo Park, CA, 1994.

Bailey, T.L., Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics, **14**: 48-54.

Bandell, M. and Lolkema, J.S. (2000) Arg-425 of the citrate transporter CitP is responsible for high affinity binding of di- and tricarboxylates. J. Biol. Chem. **275**: 39130-39136.

Bandell, M., Ansanay, V., Rachidi, N., Dequin, S., Lolkema, J.S. (1997) Membrane potential-generating malate (MleP) and citrate (CitP) transporters of lactic acid bacteria are homologous proteins - substrate specificity of the 2-hydroxycarboxylate transporter family. J. Biol. Chem. **272**: 18140-18146.

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. Jr. (2006) Extra domains in secondary transport carriers and channel proteins. Biochim Biophys Acta. **1758**: 1557-1579.

Barras, F., Loiseau, L., Py, B. (2005) How *Escherichia coli* and *Saccharomyces cerevisiae* build Fe/S proteins. Adv. Microb. Physiol. **50**: 41-101.

Bechah, Y., El Karkouri, K., Mediannikov, O., Leroy, Q., Pelletier, N., Robert, C., Médigue, C., Mege, J., Raoult, D. (2010) Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. Genome Res. **20**: 655-663.

Bent, C.J., Isaacs, N.W., Mitchell, T.J., Riboldi-Tunnicliffe, A. (2004) Crystal structure of the response regulator 02 receiver domain, the essential YycF two-component systems of *Streptococcus pneumoniae* in both complexed and native states. J. Bacteriol. **186**: 2872-2879.

Bevers, L.E., Hagedoorn, P., Santamaria-Araujo, J.A., Magalon, A., Hagen, W.R., Schwarz, G. (2008) Function of MoaB proteins in the biosynthesis of the molybdenum and tungsten cofactors. Biochem. **47**: 949-956.

Bleve, G., Zacheo, G., Cappello, M.S., Dellaglio, F., Grieco, F. (2005) Subcellular localization and functional expression of the glycerol uptake protein 1 (GUP1) of *Saccharomyces cerevisiae* tagged with green fluorescent protein. Biochem. J. **390**: 145-155.

Bosson, R., Jaquenoud, M., Conzelmann, A. (2006) *GUP1* of *Saccharomyces cerevisiae* encodes an *O*-acyltransferease involved in remodeling of the GPI anchor. Mol. Biol. Cell. **17**: 2636-2645.

Busch, W., Saier, M.H. Jr. (2002) The Transporter Classification (TC) System, 2002. Crit. Rev. of Biochem. and Mol. Bio. **37**: 287-337.

Castillo, R. and Saier, M.H. Jr. (2010) Functional promiscuity of homologues of the bacterial ArsA ATPases. Int. J. Microbiol. 2010: 187373 (Published online).

Chahal, H.K., Dai, Y., Saini, A., Ayala-Castro, C., Outten, F.W. (2009) The SuFBCD Fe-S scaffold complex interacts with SufA for Fe-S cluster transfer. Biochemistry. **48**: 10644-10653.

Chen, J.S., Reddy, V., Yen, M.R., Chen, J.H., Zheng, J.H., Shlykov, M.A., Saier, M.H. Jr. (2011) Phylogenetic characterization of transport protein superfamilies: superiority of SFT programs over those based on multiple-alignments. In preparation.

Chung, Y.J., Krueger, C., Metzgar, D., Saier, M.H. Jr. (2001) Size comparisons among integral membrane transport protein homologues in bacteria, Archaea, and Eucarya. J. Bacteriol. **183**: 1012-1021.

Cianciotto, N.P. (2005) Type II secretion: a protein secretion system for all seasons. Trend. in Microbiol. **13**: 581-588.

Cousins, R.J., Liuzzi, J.P., Lichten, L.A. (2006) Mammalian zinc transport, trafficking, and signals. J. Biol. Chem. **281**: 24085-24089.

Cragg, R.A., Christie, G.R., Phillips, S.R., Russi, R.M., Kury, S., Mathers, J.C., Taylor, P.M., Ford, D. (2002) A novel zinc-regulated human zinc transporter, hZTL1, is localized to the enterocyte apical membrane. J. Biol. Chem. **277**: 22789-22797.

Dalbey, R.E. and Chen, M. (2004) Sec-translocase mediated membrane protein biogenesis. Biochim. Biophys. Acta. **1694**: 37-53.

Daruwala, R., Song, J., Koh, W.S., Rumsey, S.C., Levine, M. (1999) Cloning and functional characterization of the human sodium-dependent vitamin C transporters hSVCT1 and hSVCT2. FEBS Lett. **460**: 480-484.

Dawson, P.A., Hubbert, M., Haywood, J., Craddock, A.L., Zerangue, N., Christian, W.V., Ballatori, N. (2005) The heteromeric organic solute transporter α-β, Ostα-Ostβ, is an ileal basolateral bile acid transporter. J. Biol. Chem. **280**: 6960-6968.

Dawson, P.A., Hubbert, M.L., Rao, A. (2010) Getting the most from OST: Role of organic solute transporter, OSTα-OSTβ, in bile acid and steroid metabolism. Biochim. Biophys. Acta. **1801**: 994-1004.

Dayhoff, M.O., Barker, W.C., Hunt, L.T. (1983) Establishing homologies in protein sequences. Methods Enzymol. **91**: 524-545.

de Koning, H. and Diallinas, G. (2000) Nucleobase transporters. Molec. Memb. Biol. **75**: 75-94.

Dipolo, R. and Beaugé, L. (2006) Sodium/calcium exchanger: influence of metabolic regulation on ion carrier interactions. Phys. Rev. **86**: 155-203.

Devereux, J., Haeberli, P., Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Res. **12**: 387-395.

Doolittle, R.F. (1986) Of URFS and ORFS: A primer on how to analyze derived amino acid sequences. Mill Valley, CA: University Science Books.

Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A., Hirayama, B.A.,

Wright, E.M., Abramson, J. (2008) The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. Science. **321**: 810-814.

Faridmoayer, A., Fentabil, M.A., Mills, D.C., Klassen, J.S., Feldman, M.F. (2007) Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. J. Bacteriol. **189**: 8088-8098.

Fletcher, J.I., Haber, M., Henderson, M.J., Norris, M.D. (2010) ABC transporters in cancer: more than just drug efflux pumps. Nature Rev. Cancer **10**: 147-156.

Flower, A.M., Hines, L.L., Pfennig, P.L. (2000) SecG is an auxiliary component of the protein export apparatus of *Escherichia coli*. Mol. Gen. Genet. **263**: 131-136.

Ford, P.C. (1971) Hydrolysis of coordinated cyanate ion. A comparison of the isocyanatopentaammine complexes of ruthenium (III) and of rhodium (III). Inorg. Chem. **10**: 2153-2158.

Fukaya, F., Promden, W., Hibino, T., Tanaka, Y., Nakamura, T., Takabe, T. (2009) An Mrp-like cluster in the halotolerant cyanobacterium *Aphanothece halophytica* functions as a Na^+/H^+ antiporter. Appl. Environ. Microbiol. **75**: 6626-6629.

Galigniana, M.D., Echeverría, P.C., Erlejman, A.G., Piwien-Pilipuk, G. (2010) Role of molecular chaperones and TPR-domain proteins in the cytoplasmic transport of steroid receptors and their passage through the nuclear pore. Nucleus. **1**: 299-308.

Ghugtyal, V., Vionnet, C., Roubaty, C., Conzelmann, A. (2007) CWH43 is required for the introduction of ceramides into GPI anchors in *Saccharomyces cerevisiae*. Mol. Microbiol. **65**: 1493-1502.

Gomolplitinant, K.M. and Saier, M.H. Jr. (2011) Evolution of the oligopeptide transporter family. J. Memb. Biol. **240**: 89-110.

Goytain, A. and Quamme, G.A. (2005) Functional characterization of ACDP2 (ancient conserved domain protein), a divalent metal transporter. Physiol. Genomics. **22**: 382-389.

Grabov, A. (2007) Plant KT/KUP/HAK potassium transporters: single family - multiple functions. Ann. Bot. **99**: 1035-1041.

Gristwood, T., McNeil, M.B., Clulow, J.S., Salmond, G.P.C., Fineran, P.C. (2011) PigS and PigP regulate prodigiosin biosynthesis in *Serratia* via differential control of divergent operons, which include predicted transporters of sulfur-containing molecules. J. Bacteriol. **193**: 1076-1085.

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

Hancock, R.E.W. (2005) Mechanisms of action of newer antibiotics for Gram-positive pathogens. Lancet Infect. Dis. **5**: 209-218.

Harris, R.M., Webb, D.C., Howitt, S.M., Cox, G.B. (2001) Characterization of PitA and PitB from *Escherichia coli*. J. Bacteriol. **183**: 5008-5014.

Heaton, M.P. and Neuhaus, F.C. (1992) Biosynthesis of D-alanyl-lipoteichoic acid: cloning, nucleotide sequence, and expression of the *Lactobacillus casei* gene for the D-alanine-activating enzyme. J. Bacteriol. **174**: 4707-4717.

Hebbeln, P., Rodionov, D.A., Alfandega, A., Eitinger, T. (2007) Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module. Proc. Natl. Acad. Sci. USA. **104**: 2909-2914.

Heymann, J.A.W., Sarker, R., Hirai, T., Shi, D., Milne, J.L.S., Maloney, P.C., Subramaniam, S. (2001) Projection structure and molecular architecture of OxlT, a bacterial membrane transporter. EMBO J. **20**: 4408-4413.

Hillerich, B. and Westpheling, J. (2006) A new GntR family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. J. Bacteriol. **188**: 7477-7487.

Hirai, T., Heymann, J.A.W., Shi, D., Sarker, R., Maloney, P.C., Subramaniam, S. (2002) Three-dimensional structure of a bacterial oxalate transporter. Nature Struct. Biol. **9**: 597-600.

Hirai, T., Heymann, J.A.W., Maloney, P.C., Subramaniam, S. (2003) A structural model for 12-helix transporters belonging to the major facilitator superfamily. J. Bacteriol. **185**: 1712-1718.

Hiramatsu, T., Kodama, K., Kuroda, T., Mizushima, T., Tsuchiya, T. (1998) A putative multisubunit Na^+/H^+ antiporter from *Staphylococcus aureus*. J. Bacteriol. **180**: 6642-6648.

Hosie, A.H., Allaway, D., Poole, P.S. (2002) A monocarboxylate permease of *Rhizobium leguminosarum* is the first member of a new subfamily of transporters. J. Bacteriol. **184**: 5436-5448.

Hoskisson, P.A., Rigali, S., Fowler, K., Findlay, K.C., Buttner, M.J. (2006) DevA, a GntR-like transcriptional regulator required for development in *Streptomyces coelicolor*. J. Bacteriol. **188**: 5014-5023.

Huang, Y., Lemieux, M.J., Song, J., Auer, M., Wang, D.N. (2003) Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. Science. **301**: 616-620.

Iwaki, H., Wang, S., Grosse, S., Bergeron, H., Nagahashi, A., Lertvorachon, J., Yang, J., Konishi, Y., Hasegawa, Y., Lau, P.C. (2006) Pseudomonad cyclopentadecanone monooxygenase displaying an uncommon spectrum of Baeyer-Villiger oxidations of cyclic ketones. Appl. and Environ. Microbiol. **72**: 2707-2720.

Iwasaki, T. (2010) Iron-sulfur world in aerobic and hyperthermoacidophilic archaeal *Sulfolobus*. Archaea. 2010: 842639 (Published online).

Iwig, J.S., Rowe, J.L., Chivers, P.T. (2006) Nickel homeostasis in *Escherichia coli* - the rcnR-rcnA efflux pathway and its linkage to NikR function. Mol. Microbiol. **62**: 252-262.

Jack, D.L., Yang, N.M., Saier, M.H. Jr. (2001) The drug/metabolite transporter superfamily. Eur. J. Biochem. **268**: 3620-3639.

Jaquenoud, M., Pagac, M., Signorell, A., Benghezal, M., Jelk, J., Bütikofer, P., Conzelmann, A. (2008) The Gup1 homologue of *Trypanosoma brucei* is a GPI glycosylphosphatidylinositol remodelase. Mol. Microbiol. **67**: 202-212.

Jiang, Z., Grichtchenko, I.I., Boron, W.F., Aronson, P.S. (2002) Specificity of anion exchange mediated by mouse Slc26a6. J. Biol. Chem. **277**: 33963-33967.

Jittawuttipoka, T., Sallabhan, R., Vattanaviboon, P., Fuangthong, M., Mongkolsuk, S. (2010) Mutations of ferric uptake regulator (*fur*) impair iron homeostasis, growth, oxidative stress survival, and virulence of *Xanthomonas campestris* pv. campestris. Arch. Microbiol. **192**: 331-339.

Karatza, P. and Frillingos, S. (2006) Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. Mol. Membr. Biol. **22**: 251-261.

Karatza, P., Panos, P., Georgopoulou, E., Frillingos, S. (2006) Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential, and Ile-329-Val-339 form an α-helix. J. Biol. Chem. **281**: 39881-39890.

Karlsson, H., Larsson, P., Wold, A.E., Rudin, A. (2004) Pattern of cytokine responses to gram-positive and gram-negative commensal bacteria is profoundly changed when monocytes differentiate into dendritic cells. Infect. Immun. **72**: 2671-2678.

Kästner, C.N., Schneider, K., Dimroth, P., Pos, K.M. (2002) Characterization of the citrate/acetate antiporter CitW of *Klebsiella pneumoniae*. Arch. Microbiol. **177**: 500-506.

Kawai, S., Suzuki, H., Yamamoto, K., Kumagai, H. (1997) Characterization of the L-malate permease gene (*maeP*) of *Streptococcus bovis* ATCC 15352. J. Bacteriol. **179**: 4056-4060.

Kennerknecht, N., Sahm, H., Yen, M.R., Patek, M., Saier, M.H. Jr., Eggeling, L. (2002) Export of L-isoleucine from *Corynebacterium glutamicum*: a two-gene-encoded member of a new translocator family. J. Bacteriol. **184**: 3947-3956.

Kim, S.A., Punshon, T., Lanzirotti, A., Li, L., Alonson, J.M., Ecker, J.R., Kaplan, J., Guerinot, M.L. (2006) Localization of iron in *Arabidopsis* seed requires the vacuolar membrane transporter VIT1. Science. **314**: 1295-1298.

Kiyasu, T., Asakura, A., Nagahashi, Y., Hoshino, T. (2000) Contribution of cysteine desulfurase (NifS protein) to the biotin synthase reaction of *Escherichia coli*. J. Bacteriol. **182**: 2879-2885.

Kosono, S., Morotomi, S., Kitada, M., Kudo, T. (1999) Analyses of a *Bacillus subtilis* homologue of the Na^+/H^+ antiporter gene which is important for pH homeostasis of alkaliphilic *Bacillus* sp. C-125. Biochim. Biophys. Acta. **1409**: 171-175.

Krejcík, Z., Denger, K., Winitschke, S., Hollemeyer, K., Paces, V., Cook, A.M., Smits, T.H.M. (2008) Sulfoacetate released during the assimilation of taurine-nitrogen by *Neptuniibacter caesariensis*: purification of sulfoacetaldehyde dehydrogenase. Arch. Microbiol. **190**: 159-168.

Krogh, A., Larsson, G., von Heijne, G., Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology **305**: 567-580.

Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., Clausen, T. (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. Nature. **416**: 455-459.

Lam, V.H., Lee, J., Silverio, A., Chan, H., Gomolplitinant, K.M., Povolotsky, T.L., Orlova, E., Sun, E.I., Welliver, C.H., Saier, M.H. Jr. (2011) Pathways of transport protein evolution: recent advances. Biol. Chem. **392**: 5-12.

Larkin, M.A., Blackshields, G., Brown, N.P. Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics **23**: 2947-2948.

Law, C.J., Maloney, P.C., Wang, D.N. (2008) Ins and outs of major facilitator superfamily antiporters. Annu. Rev. Microbiol. **62**: 289-305.

Lee, M.H., Scherer, M., Rigali, S., Golden, J.W. (2003) PlmA, a new member of the GntR family, has plasmid maintenance functions in *Anabaena* sp. strain PCC 7120. J. Bacteriol. **185**: 4315-4325.

Li, W., and Godzik, A. (2006) CD-Hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. **22**: 1658-1659.

Lipinska, B., Zylicz, M., Georgopoulos, C. (1990) The Htra (Degp) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. J. Bacteriol. **172**: 1791-1797.

Liu, W., Karavolos, M.H., Bulmer, D.M., Allaoui, A., Hormaeche, R.D.C.E., Lee, J.J., Khan, A.C.M. (2007) Role of the universal stress protein UspA of *Salmonella* in growth arrest, stress and virulence. Microb. Pathog. **42**: 2-10.

Locher, H.H., Poolman, B., Cook, A.M., Konings, W.N. (1993) Uptake of 4-toluene sulfonate by *Comamonas testosteroni* T-2. J. Bacteriol. **175**: 1075-1080.

Lumppio, H.L., Shenvi, N.V., Summers, A.O., Voordouw, G., Kurtz, D.M. Jr. (2001) Rubrerythrin and rubredoxin oxidoreductase in *Desulfovibrio vulgaris*: a novel oxidative stress protection system. J. Bacteriol. **183**: 101-108.

Mampel, J., Maier, E., Tralau, T., Ruff, J., Benz, R., Cook, A.M. (2004) A novel outer-membrane anion channel (porin) as part of a putatively two-component transport system for 4-toluenesulfonate in *Comamonas testosteroni* T-2. Biochem. J. **383**: 91-99.

Mansilla, M.C. and de Mendoza, D. (2000) The *Bacillus subtilis cysP* gene encodes a novel sulphate permease related to the inorganic phosphate transporter (Pit) family. Microbiology. **146**: 815-821.

Maralikova, B., Ali, V., Nakada-Tsukui, K., Nozaki, T., Gizen, M.V.D., Henze, K., Tovar, J. (2010) Bacterial-type oxygen detoxification and iron-sulfur cluster assembly in amoebal relict mitochondria. Cell. Micriobiol. **12**: 331-342.

Marchler-Bauer A. et al. (2009). CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. **37**: 205-10.

Marraffini, L.A., DeDent, A.C., Schneewind, O. (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. Microbiol. Mol. Biol. Rev. **70**: 192-221.

Matias, M.G., Gomolplitinant, K.M., Saier, M.H. Jr. (2010) Animal Ca^{2+} release-activated Ca^{2+} (CRAC) channels appear to be homologous to and derived from the ubiquitous cation diffusion facilitators. BMC Res. Notes. **3**: 158.

Missiakas, D., Schwager, F., Raina, S. (1995) Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli* EMBO J. **14**: 3415-3424.

Möller, S., Croning, M.D.R., Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics **17**: 646-653.

Montanini, B., Blaudez, D., Jeandroz, S., Sanders, D., Chalot, M. (2007) Phylogenetic and functional analysis of the Cation Diffusion Facilitator (CDF) family: improved signature and prediction of substrate specificity. BMC Genomics. **8**: 107.

Mortier-Barrière, I., Velten, M., Dupaigne, P., Mirouze, N., Piétrement, O., McGovern, S., Fichant, G., Martin, B., Noirot, P., Le Cam, E., Polard, P., Ciaverys, J. (2007) A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. Cell. **130**: 824-836.

Nanbu-Wakao, R., Asoh, S., Nishimaki, K., Tanaka, R., Ohta, S. (2000) Bacterial cell death induced by human pro-apoptotic Bax is blocked by an RNase E mutant that functions in an anti-oxidant pathway. Genes to Cells. **5**: 155-167.

Noinaj, N., Guillier, M., Barnard, T.J., Buchanan, S.K. (2010) TonB-dependent transporters: regulation, structure, and function. Microbiol. **64**: 43-60.

Novichkov, P.S., Laikova, O.N., Novichkova, E.S., Gelfand, M.S., Arkin, A.P., Dubchak, I., Rodionov, D.A. (2010a) RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. Nucleic Acids Res. **38**: D111-118.

Novichkov, P.S., Rodionov, D.A., Stavrovskaya, E.D., Novichkova, E.S., Kazarov, A.E. Gelfand, M.S., Arkin, A.P. Mironov, A.A., Dubchak, I. (2010b) RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. Nucleic Acids Res. **38**: W299-307.

Ochman, H. and Davalos, L.M. (2006) The nature and dynamics of bacterial genomes. Science. **311**: 1730-1733.

Ohana, E., Shcheynikov, N., Yang, D., So, I., Muallem, S. (2011) Determinants of coupled transport and uncoupled current by the electrogenic SLC26 transporters. J. Gen. Physiol. **137**: 239-251.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L, Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucl. Acids. Res. **33**: 5691-5702.

Palomino, C. and Mellado, R.P. (2008) Influence of a *Streptomyces lividans* SecG functional analogue on protein secretion. Int. Microbiol. **11**: 25-31.

Paulsen, I.T. and Saier, M.H. Jr. (1997) A novel family of ubiquitous heavy metal ion transport proteins. J. Membr. Biol. **156**: 99-103.

Pearson, W.R. (1998) Empirical statistical estimates for sequence similarity searches. J. Mol Biol. **276**: 71-84.

Pernil, R., Herrero, A., Flores, E. A TRAP transporter for pyruvate and other monocarboxylate 2-oxoacids in the cyanobacterium *Anabaena* sp. strain PCC 7120 (2010) J. Bacteriol. **192**: 6089-6092.

Persson, B.L., Berhe, A., Fristedt, U., Martinez, P., Pattison, J., Petersson, J., Weinander, R. (1998) Phosphate permeases of *Saccharomyces cerevisiae*. Biochim. Biophys. Acta. **1365**: 23-30.

Persson, B.L., Petersson, J., Fristedt, U., Weinander, R., Berhe, A., Pattison, J. (1999) Phosphate permeases of *Saccharomyces cerevisiae*: structure, function and regulation. Biochim. Biophys. Acta. **1422**: 255-272.

Pizarro-Cerdá, J. and Cossart, P. (2006) Bacterial adhesion and entry into host cells. Cell. **124**: 715-727.

Rambaut, A. (2009) FigTree (Version 1.3.1) [Software]. Available from http://tree.bio.ed.ac.uk/software/figtree/.

Reizer, J., Reizer, A., Saier, M.H. Jr. (1994) A functional superfamily of sodium/ solute symporters. Biochim. Biophys. Acta. **1197**: 133-166.

Remmert, M., Biegert, A., Lupas, A.N., Söding, J. (2010) Evolution of outer membrane β-barrels from an ancestral ββ hairpin. Mol. Biol. Evol. **27**: 1348-1358.

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

Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B., Titgemeyer, F. (2004) Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new cis/trans relationships. Nucl. Acids Res. **32**: 3418-3426.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., Gelfand, M.S. (2002) Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. J. Biol. Chem. **277**: 48949-48959.

Rodionov, D.A., Hebbeln, P., Gelfand, M.S., Eitinger, T. (2006) Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters. J. Bacteriol. **188**: 317-327.

Rodionov, D.A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I.A., Erkens, G.B., Slotboom, D.J., Gelfand, M.S., Osterman, A.L., Hanson, A.D., Eitinger, T. (2009) A novel class of modular transporters for vitamins in prokaryotes. J. Bacteriol. **191**: 42-51.

Rodrigue, A., Effantin, G., Mandrand-Berthelot, M.A. (2005) Identification of rcnA (yohM), a nickel and cobalt resistance gene in *Escherichia coli*. J. Bacteriol. **187**: 2912-2916.

Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I. (2002) *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. **70**: 3371-3381.

Roje, S., Janave, M.T., Ziemak, M.J., Hanson, A.D. (2002) Cloning and characterization of mitochondrial 5-formyltetrahydrofolate cycloligase from higher plants. J. Biol. Chem. **277**: 42748-42754.

Rückert, C., Koch, D.J., Rey, D.A., Albersmeier, A., Mormann, S., Pühler, A., Kalinowski, J. (2005) Functional genomics and expression analysis of the *Corynebacterium glutamicum fpr2-cysIXHDNYZ* gene cluster involved in assimilatory sulphate reduction. BMG Genomics **6**: 121.

Sääf, A., Baars, L., von Heijne, G. (2001) The internal repeats in the Na^{+}/Ca^{2+} exchanger-related *Escherichia coli* protein YrbG have opposite membrane topologies. J. Biol. Chem. **276**: 18905-18907.

Saier, M.H., Jr. (1994) Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution. **58**: 71-93.

Pao, S.S., Paulsen, I.T., Saier, M.H. Jr. (1998) Major Facilitator Superfamily. Microbiol. Mol. Biol. Rev. **62**: 1-34.

Saier, M.H. Jr., Eng, B.H., Fard, S., Garg, J., Haggerty, D.A., Hutchinson, W.J., Jack, D.L., Lai, E.C., Liu, H.J., Nusinew, D.P., Omar, A.M., Pao, S.S., Paulsen, I.T., Quan, J.A., Sliwinski, M., Tseng, T., Wachi, S., Young, G.B. (1999) Phylogenetic characterization of novel transport protein families revealed by genome analyses. Biochim. Biophy. Acta. **1422**: 1-56.

Saier, M.H., Jr. (2000b) Vectorial metabolism and the evolution of transport systems. J. of Bacteriology **182**: 5029-5035.

Saier, M.H., Jr., Tran, C.V., Barabote, R.D. (2006) TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucl. Acids Res. **34**: 181-186.

Saier, M.H., Jr., Yen, M.R., Noto, K., Tamang, D.G., Elkan, C. (2009) The Transporter Classification Database: recent advances. Nucl. Acids Res. **37**: 274-278.

Saini, A., Mapolelo, D.T., Chahal, H.K., Johnson, M.K., Outten, W.F. (2010)

SufD and SufC ATPase activity are required for iron acquisition during in vivo Fe-S cluster formation on SufB.

Salaün, C., Rodrigues, P., Heard, J.M. (2001) Transmembrane topology of PiT-2, a phosphate transporter-retrovirus receptor. J. Virol. **75**: 5584-5592.

Schmidt, G. and Zink, R. (2000) Basic features of the stress response in three species of *Bifidobacteria*: *B. longum*, *B. adolescentis*, and *B. breve*. Int. J. Food Microbiol. **55**: 41-45.

Seward, D.J., Koh, A.S., Boyer, J.L., Ballatori, N. (2003) Functional complementation between a novel mammalian polygenic transport complex and an evolutionarily ancient organic solute transporter, *OSTα-OSTβ*. J. Biol. Chem. **278**: 27473-27482.

Simic, P., Sahm, H., Eggeling, L. (2001) L-threonine export: use of peptides to identify a new translocator from *Corynebacterium glutamicum*. J. Bacteriol. **183**: 5317-5324.

Sobczak, I. and Lolkema, J.S. (2004) Alternating access and a pore-loop structure in the Na^+-citrate transporter CitS of *Klebsiella pneumoniae*. J. Biol. Chem. **279**: 31113-31120.

Sobczak, I. and Lolkema, J.S. (2005) The 2-hydroxycarboxylate transporter family: physiology, structure, and mechanism. Microbiol. Mol. Biol. Rev. **69**: 665-695.

Sonnhammer, E.L.L., von Heijne, G., Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In J. Glasgow, T. Littlejohn, F. Major, D. Sankoff, and C. Sensen, editors, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, AAAI Press, Menlo Park, CA, 1998.

Sousa, M.C. and McKay, D.B. (2001) Structure of the universal stress protein of *Haemophilus influenzae*. Structure. **9**: 1135-1141.

Spiess, C., Beil, A., Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat schock protein. Cell. **97**: 339-347.

Swartz, T.H., Ito, M., Ohira, T., Natsui, S., Hicks, D.B., Krulwich, T.A. (2007) Catalytic properties of *Staphylococcus aureus* and *Bacillus* members of the secondary cation/proton antiporter-3 (Mrp) family are revealed by an optimized assay in an *Escherichia coli* host. J. Bacteriol. **189**: 3081-3090.

Tadesse, S. and Graumann, P.L. (2007) DprA/Smf protein localizes at the DNA uptake machinery in competent *Bacillus subtilis* cells. BMC Microbiol. **7**: 105.

Taniguchi, N. and Tokuda, H. (2008) Molecular events involved in a single cycle of ligand transfer from an ATP binding cassette transporter, LolCDE, to a molecular chaperone, LolA. J. Biol. Chem. **283**: 8538-8544.

ter Huurne, A.A., Muir, S., van Houten, M., van der Zeijst, B.A., Gaastra, W., Kusters, J.G. (1994) Characterization of three putative *Serpulina hyodysenteriae* hemolysins. Microb. Pathog. **16**: 269-282.

Thomas, S.A. (2004) Drug transporters relevant to HIV therapy. J. HIV Ther. **9**: 92-96.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876-4882.

Thomson, J.M., Bonomo, R.A. (2005) The threat of antibiotic resistance in Gram-negative pathogenic bacteria: ß-lactams in peril! Curr. Opin. In Microbiol. **8**: 518-524.

Tuoati, D. (2000) Iron and oxidative stress in bacteria. Arch. Biochem. and Biophys. **373**: 1-6.

Trchounian, A. and Kobayashi, H. (1999) Kup is the major K^+ uptake system in *Escherichia coli* upon hyper-osmotic stress at a low pH. FEBS Lett. **447**: 144-148.

Tseng, T.T., Gratwick, K.S., Kollman, J., Park, D., Nies, D.H., Goffeau, A., Saier, M.H. Jr. (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. J. Mol. Microbiol. Biotechnol. **1**: 107-125.

Tuominen, H., Salminen, A., Oksanen, E., Jämsen, J., Heikkilä, O., Lehtiö, L., Magretova, N.N., Goldman, A., Baykov, A.A., Lahti, R. (2010) Crystal structures of the CBS and DRTGG domains of the regulatory region of *Clostridium perfringens* pyrophosphatase complexed with the inhibitor, AMP, and activator, diadenosine tetraphosphate. J. Mol. Biol. **398**: 400-413.

Tusnády, G.E., Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. **283**: 489-506.

Tusnády, G.E., Simon, I. (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics **17**: 849-850.

Vindal, V., Suma, K., Ranjan, A. (2007) GntR family of regulators in *Mycobacterium smegmatis*: a sequence and structure based characterization. BMC Genomics. **8**: 289.

von Heijne, G., and Gavel, Y. (1988) Topogenic signals in integral membrane proteins. Eur. J. Biochem. **174**: 671-678.

Wang, W., Seward, D.J., Li, L., Boyer, J.L., Ballatori, N. (2001) Expression cloning of two genes that together mediate organic solute and steroid transport in the liver of a marine vertebrate. Proc. Natl. Acad. Sci. USA. **98**: 9431-9436.

Wang, B., Dukarevich, M., Sun, E.I., Yen, M.R., Saier, M.H. Jr. (2009a) Membrane porters of ATP-binding cassette transport systems are polyphyletic. J. Membr. Biol. **231**: 1-10. Epub.

Wang, L., Jeon, B., Sahin, O., Zhang, Q. (2009b) Identification of an arsenic resistance and arsenic-sensing system in *Campylobacter jejuni*. Appl. Environ. Microbiol. **75**: 5064-5073.

Wang, W., Huang, H., Tan, G., Si, F., Liu, M., Landry, A.P., Lu, J., Ding, H. (2010) In vivo evidence for the iron-binding activity of an iron-sulfur cluster assembly protein IscA in *Escherichia coli*. Biochem J. **432**: 429-436.

Weber, R.E. and Vinogradov, S.N. (2001) Nonvertebrate hemoglobins: functions and molecular adaptations. Physiological Rev. **81**: 569-628.

Wei, Y., and Fu, D. (2006) Binding transport of metal ions at the dimmer interface of the *Escherichia coli* metal transporter YiiP. J. Biol. Chem. **281**: 23492-23502.

Wei, X., Sayavedra-Soto, L.A., Arp, D.J. (2007) Characterization of the ferrioxamine uptake system of *Nitromonas europaea*. Microbiol. **153**: 3963-3972.

Weinitschke, S., Denger, K., Cook, A.M., Smits, T.H. (2007) The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. Microbiology **153**: 3055-3060.

Winkler, H.H., Neuhaus, H.E. (1999) Non-mitochondrial ATP transport. Trends Biol. Sci. **24**: 64-68.

Winkler, H.H., Daugherty, R., Hu, F. (1999) *Rickettsia prowazekii* transports UMP and GMP, but not CMP, as building blocks for RNA synthesis. J. Bacteriol. **181**: 3238-3241.

Wróbel, M., Lewandowska, I., Bronowicka-Adamska, P., Paszewski, A. (2009) The level of sulfane sulfur in the fungus *Aspergillus nidulans* wild type and mutant strains. Amino Acids. **37**: 565-571.

Yen, M.R., Choi, J., Saier, M.H. Jr. (2009) Bioinformatic analyses of transmembrane transport: novel software for deducing protein phylogeny, topology, and evolution. J. of Mol. Microbiol. and Biotech. **17**: 163-176.

Yen, M.R., Chen, J.S., Marquez, J.L., Sun, E.I., Saier, M.H. (2010) Multidrug resistance: phylogenetic characterization of superfamilies of secondary carriers that include drug exporters. Methods. Mol. Biol. **637**: 47-64.

Yuvaniyama, P., Agar, J.N., Cash, V.L., Johnson, M.K., Dean, D.R. (2000) NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. PNAS. **2**: 599-604.

Zakharyan, E. and Trchounian, A. (2001) K^+ influx by Kup in *Escherichia coli* is accompanied by a decrease in H^+ efflux. FEMS Microbiol. Lett. **204**: 61-64.

Zhai, Y. and Saier, M.H. Jr. (2001a) A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence. J. Mol. Microbiol. Biotech. **3**: 501-502.

Zhai, Y. and Saier, M.H. Jr. (2001b) A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins. J. Mol. Microbiol. Biotech. **3**: 285-286.

Zhai, Y. and Saier, M.H. Jr. (2002) A simple sensitive program for detecting internal repeats in sets of multiply aligned homologous proteins. J. Mol. Microbiol. Biotech. **4**: 375-377.

Zhai, Y., Heijne, W.H.M., Smith, D.W., Saier, M.H. Jr. (2001) Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predications for a putative fungal chaperone protein. Biochimica et Biophysica Acta. **1511**: 206-223.

Zhao, B., Yeo, C.C., Poh, C.L. (2005) Proteome investiation of the global regulatory

role of sigma 54 in response to gentisate induction in *Pseudomonas alcaligenes* NCIMB 9867. Proteomics. **5**: 1868-1876.

Ziegler, P., Sahm, H., Eggeling, L. (2000) Identification of a new translocator structure functioning to export L-threonine from the cell. J. Bacteriol. submitted.