

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Common evolutionary pathway for the appearance of the Ca^{2+} : Cation Antiporter (CaCA) family and several other transporter superfamilies

Permalink

<https://escholarship.org/uc/item/5c89x28r>

Author

Zhang, Chenchen

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Common Evolutionary Pathway
for the Appearance of
the Ca^{2+} :Cation Antiporter (CaCA) Family
and Several Other Transporter Superfamilies

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Biology

by

Chenchen Zhang

Committee in charge:

Professor Milton H. Saier, Jr., Chair
Professor Lin Chao
Professor Charles Elkan

2012

The Thesis of Chenchen Zhang is approved and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2012

TABLE OF CONTENTS

Signature Page	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Abstract	xi
Introduction	1
Methods	4
Chapter 1	
1.1: Phylogenetic Analysis of CaCA Family Homologues	6
1.2: Phylogenetic Analysis of MPE family Homologues	12
Chapter 2:	
2.1: Orthology Analysis of MPE Family Homologues.....	16
2.2: Orthology Analysis of MPE Family Homologues.....	20
Chapter 3:	
3.1: Topological Analyses of CaCA Family Homologues	24
3.2: Topological Analyses of MPE Family Homologues	27
Chapter 4:	

4.1: Establishment of Internal Repeats in CaCA Family Proteins	30
4.2: Internal Repeats in MPE Family Proteins	33
Chapter 5: Establish Homology Between Superfamilies	34
Discussion	41
Appendix	47
References	128

LIST OF TABLES

Table 1. CaCA Family Homologues.....	47
Table 2. MPE Family Homologues	56
Table 3. Paralogues in Cluster 1 (CaCA).....	67
Table 4. Paralogues in Cluster 2 (CaCA)	68
Table 5. Paralogues in Cluster 3 (CaCA).....	69
Table 6. Paralogues in Cluster 4 (CaCA).....	69
Table 7. Paralogues in Cluster 5 (CaCA).....	71
Table 8. Paralogues in Cluster 6 (CaCA).....	75
Table 9. Paralogues in Cluster 1 (MPE).	75
Table 10. Paralogues in Cluster 2 (MPE).	80
Table 11. Paralogues in Cluster 3 (MPE).	81
Table 12. Paralogues in Cluster 4 (MPE).	83
Table 13. Paralogues in Cluster 5 (MPE).	87
Table 14. Paralogues in Cluster 7 (MPE).	89
Table 15. Obvious HGT (CaCA).....	90
Table 16. Paralogues In the CaCA Family	90

Table 17. Paralogues In the MPE Family	90
Table 18. Obvious HGT (MPE).....	90
Table 19. Characteristic of Phylogenetic Clusters from the CaCA Family	91
Table 20. Characteristic of Phylogenetic Clusters from the MPE Family	91
Table 21. Comparsion Scores Between Superfamilies.	91
Table 22. Comparsion Scores Between Potential Repeat of TMSs (MPE).	92

LIST OF FIGURES

Figure 1. Protein Tree for all CaCA Family Members	93
Figure 2. Protein Tree for Cluster 1 Proteins (CaCA)	94
Figure 3. Protein Tree for Cluster 2 Proteins (CaCA)	95
Figure 4. Protein Tree for Cluster 3 Proteins (CaCA)	96
Figure 5. Protein Tree for Cluster 4 Proteins (CaCA)	97
Figure 6. Protein Tree for Cluster 5 Proteins (CaCA)	98
Figure 7. Protein Tree for Cluster 6 Proteins (CaCA)	99
Figure 8. Protein Tree for all MPE Family Members	100
Figure 9. Protein Tree for Cluster 1 Proteins (MPE)	101
Figure 10. Protein Tree for Cluster 2 Proteins (MPE)	102
Figure 11. Protein Tree for Cluster 3 Proteins (MPE)	103
Figure 12. Protein Tree for Cluster 4 Proteins (MPE)	104
Figure 13. Protein Tree for Cluster 5 Proteins (MPE)	105
Figure 14. Protein Tree for Cluster 7 Proteins (MPE)	106
Figure 15. 16S rRNA Tree (CaCA)	107
Figure 16. 16S rRNA Tree (MPE)	108
Figure 17. AveHAS Plots for all CaCA Family Members	109
Figure 18. AveHAS Plots for Cluster 1 Proteins (CaCA)	109

Figure 19. AveHAS Plot for Cluster 2 Proteins (CaCA)	109
Figure 20. AveHAS Plot for Cluster 3 Proteins (CaCA)	110
Figure 21. AveHAS Plot for Cluster 4 Proteins (CaCA)	110
Figure 22. AveHAS Plots for Cluster 5 Proteins (CaCA)	110
Figure 23. AveHAS Plots for Cluster 6 Proteins (CaCA)	111
Figure 24. AveHAS Plots for all MPE Family Members	111
Figure 25. AveHAS Plots for Cluster 1 Proteins (MPE)	112
Figure 26. AveHAS Plots for Cluster 2 Proteins (MPE)	112
Figure 27. AveHAS Plots for Cluster 3 Proteins (MPE)	113
Figure 28. AveHAS Plots for Cluster 4 Proteins (MPE)	113
Figure 29. AveHAS Plots for Cluster 5 Proteins (MPE)	114
Figure 30. AveHAS Plots for Cluster 6 Proteins (MPE)	114
Figure 31. AveHAS Plots for Cluster 7 Proteins (MPE)	115
Figure 32. Alignment of TMSs 2-6 vs TMSs 7-11 (CaCA)	115
Figure 33. Alignment of Mla1 and Dac1(CaCA)	116
Figure 34. Alignment of TMSs 1-5 vs TMSs 7-11 (CaCA)	117
Figure 35. Alignment of TMSs 1-5 vs TMSs 6-10 (CaCA)	117
Figure 36. Alignment of TMSs 1-2 vs TMSs 3-4 (CaCA)	117
Figure 37. Alignment of TMSs 7-8 vs TMSs 9-10 (CaCA)	118

Figure 38. Alignment of MPE with CaCA	118
Figure 39. Alignment of CDF with CaCA.....	119
Figure 40. Alignment of CDF with MPE	119
Figure 41. Alignment of APC with CDF	119
Figure 42. Alignment of CDF with DMT	120
Figure 43. Alignment of TDT with CDF	120
Figure 44. Alignment of CDF with RND	120
Figure 45. Alignment of TMSs 9-10 vs TMSs 11-12 (RND).....	121
Figure 46. Alignment of CaCA with RND	121
Figure 47. Alignment of CaCA with AAAP.....	121
Figure 48. Alignment of DMT with CaCA.....	122
Figure 49. Alignment of CaCA with TDT	122
Figure 50. Alignment of RND with MPE.....	122
Figure 51. Alignment of MPE with DMT	122
Figure 52. Alignment of MPE with TDT.....	123
Figure 53. Alignment of TDT with RND	123
Figure 54. Alignment of RND with DMT	123
Figure 55. Alignment of AAAP with DMT.....	124
Figure 56. Alignment of TDT with DMT.....	124

Figure 57. Alignment of TDT with AAAP	125
Figure 58. Alignment of RND with AAAP	125
Figure 59. Alignment of AAAP with MPE	126
Figure 60. Alignment of AAAP with CDF	126
Figure 61. Alignment of CDF with DMT	127
Figure 61. Proposed Evolutionary Pathway for the CaCA Family	127

ABSTRACT OF THE THESIS

Common Evolutionary Pathway
for the Appearance of
the Ca^{2+} :Cation Antiporter (CaCA) Family
and Several Other Transporter Superfamilies

by

Chenchen Zhang

Master of Science in Biology

University of California, San Diego, 2012

Professor Milton H. Saier, Jr., Chair

Ca^{2+} :Cation Antiporter (CaCA) family homologues are found in both prokaryotes and eukaryotes, while bacterial Murein Precursor Exporter (MPE) family homologues are found only in prokaryotes. CaCA family homologues have 9- 13 transmembrane spanning α -helical segments (TMSs). We demonstrated that the 10 TMSs topology arose via three sequential duplication events as following: 2 TMSs $\xrightarrow{\times 2}$ 4 TMSs $\xrightarrow{+1}$ 5 TMSs $\xrightarrow{\times 2}$ 10 TMSs. Homologues fall into 6 phylogenetic clusters that correlate with function. Comparisons between proteins and 16S rRNA trees revealed that horizontal gene transfer was frequent in prokaryotes, but rare in eukaryotes. Early gene duplication of a gene encoding common ancestral protein probably gave rise to the major cluster.

These paralogues were then transmitted vertically to give rise to sub-families of different substrate specificities and characteristics.

MPE family members are homologous to CaCA proteins. These proteins have 9-10 transmembrane spanning α -helical segments and probably evolved via the same pathway as CaCA family members although this could not be demonstrated. MPE family proteins fall into 7 clusters. Early gene duplication events may have given rise to these different clusters with Cluster 1 and 4 homologues diverging early, and Cluster 3 diverging more recently. Homology was also established between several ancient transporter superfamilies, all which appear to have evolved from a common 2 TMSs hairpin structure.

Introduction

In the last twenty-five years, over six hundred membrane transport protein families have been identified and classified in the Transporter Classification Database, TCBD (<http://www.tcdb.org>). Protein families are classified on the basis of structural, functional and evolutionary characteristics (Saier et al., 2006; 2009).

Ca^{2+} :Cation Antiporter Family (CaCA; TC#2.A.19) proteins are found in both eukaryotes and prokaryotes. Proteins exhibit divergent sequences and have arisen by a tandem intragenic duplication event (Saier *et al.*, 1999). Based on 3-D structures, these proteins contains 10 core transmembrane helices with four ion-binding sites in the center of the protein, one for Ca^{2+} , and the other three for Na^+ (Liao *et al.*, 2012). The *E. coli* ChaB (YrbG) homologue has 10 TMSs with both N- and C-termini located in the periplasm. Two 5 TMS internal repeats have been demonstrated (Saaf *et al.*, 2001). The two 5 TMS duplication units show opposite orientation in the membrane, stabilized by positively charged residues in the cytoplasmic loops. CaCA family proteins exhibit various protein sizes from 302 to 1199 residues. Prokaryotic proteins generally exhibit shorter sequences than the eukaryotic proteins (Chung et al., 2001).

The CaCA family primarily functions in Ca^{2+} extrusion (DiPolo and Beauge, 2006). These proteins may also function in K^+ extrusion; for example, *E. coli* cells expressing ChaA can mediate K^+ efflux against a K^+ concentration gradient (Radchenko *et al.*, 2006). Another family member Vnx1p does not mediate Ca^{2+} transport due to the absence of key amino acids. Vnx1p catalyzes Na^+/H^+ and K^+/H^+ but not $\text{Ca}^{2+}/\text{H}^+$ exchange (Cagnac *et al.*, 2007).

Another CaCA family protein, YfkE from *Bacillus subtilis*, exhibits robust $\text{Ca}^{2+}/\text{H}^{+}$ antiport activity, and Ca^{2+} signaling is critical for sporulation and germination. Neither Na^{+} nor K^{+} has been shown to serve as a substrate in this antiporter (Fujisawa *et al.*, 2009).

CaCA family proteins are essential for human neuronal activity by exporting calcium out of the cytosol upon stimulation. Defects in $\text{Na}^{+}/\text{Ca}^{2+}$ exchangers can lead to neurodegenerative disorders (Gomez-Villafuertes *et al.*, 2007). One $\text{Na}^{+}/\text{Ca}^{2+}$ exchanger maintains Ca^{2+} homeostasis for cardiac contractility (Iwamoto *et al.*, 1999). Calcium ions bind to CBD1 and CBD2 domains, located in large intracellular loops of some CaCA family members, and binding activates Ca^{2+} transport across the plasma membrane (Besserer *et al.*, 2007). CaCA family proteins probably arose by gene duplication before the three domains of life arose (Saier *et al.*, 1999).

MPE family proteins are observed only in prokaryotes. They have protein sizes of 370-420 amino acid residues. MPE family homologues show 9-10 transmembrane helices with a large extra-cytoplasmic loop and both N- and C-termini in the cytoplasm (Gérard *et al.*, 2002). These proteins include the FtsW cell division protein, the RodA rod shape determining protein (both of *E. coli*) and the SpoVE sporulation protein of *B. subtilis* (Boyle *et al.* 1997; Errington, 2003; Matsuzawa *et al.*, 1989; Sato *et al.*, 1990). These proteins export Lipid II, precursors of peptidoglycan across the cytoplasmic membrane (Mohammadi *et al.* 2011).

In *E. coli*, the class B PBP3 domain is concerned with septal peptidoglycan synthesis during cell division, which required FtsW, a flippase for lipid II localized at the

division site. *E. coli* FtsW and PBP form a sub-complex via a PBP dimer domain in which periplasmic loop 9/10 of FtsW (Fraipont *et al.*, 2011) plays a role.

In this study, we establish homology of CaCA and MPE family members. To establish homology between repeat elements in the transmembrane domains of the CaCA and MPE family proteins, statistics and the superfamily principle were used. We also extended the observation of homology to other families of transport proteins (Doolittle, 1981, Saier 2003; Tran and Saier 2004; Yen et al. 2009a, b).

Method

Data Sources

FASTA sequences of transport proteins were gathered from the Transporter Classification Database (www.tcdb.org). The Protocol1 Program was used to generate tables providing CaCA and MPE family homologues with protein abbreviations, descriptions, organismal sources, sizes, GI numbers, organismal kingdoms or phyla, and organismal domains, based on PSI-BLAST search (Altschul *et al.*, 1997) against the National Center for Biotechnology Information (NCBI). A modified CD-HIT program (Li *et al.*, 2006, Saier *et al.*, 2009; Yen *et al.*, 2009) was also used to eliminate redundant and similar sequences with greater than 40% identity.

Multiple Alignment Methods

The Clustal X program was used to generate multiple alignment files for CaCA and MPE family homologues as well as 16S RNA genomic sequences of each represented genus. These multiple alignments were used to generate phylogenetic trees (Thompson, Gibson *et al.* 1997). The TreeView program (Zhai *et al.*, 2002) was used to view the phylogenetic trees by opening the 'dnd' file, created by ClustalX.

Topological Analyses

The Web-Based Hydropathy, Amphipathicity and Topology (WHAT) (Zhai and Saier, 2001a) and HMMTOP (G.E Tusnády and I. Simon 2001) were used to perform the topological analyses on single protein FASTA sequences. The AveHAS program (Zhai and Saier, 2001b) creates average hydropathy, amphipathicity and similarity plots of protein homologues based on Clustal X-based multiple sequence alignments.

Establishment of homology between families.

The Global Sequence Alignment Tool (GSAT) performs binary alignments of two protein sequences and generates a comparison score, expressed in standard deviations (S.D.); a comparison score over 11 S.D. is considered sufficient to establish homology (Reddy & Saier, 2012).

Protocol 2 (Reddy & Saier, 2012) provides confirmatory evidence for homology. The program compares two sets of FASTA sequences using the Smith Waterman algorithm to locate local alignments, and runs GSAT to generate HTML reports.

The Ancient Rep (AR) program searches for internal repeats in a set of homologous protein sequences and generates the vertical.txt file as well as the horizontal.txt file (Reddy & Saier, 2012). The horizontal.txt file is generated based on comparisons within one protein. The vertical.txt is generated based on the comparison of TMSs in one protein against an entire set of protein sequences. Conclusions of homology are then based on the superfamily principle. A 11 S.D. comparison score is considered sufficient to establish homology between selected TMSs and to establish duplication units within a protein family. HMMGAP (Reddy & Saier, 2010) visualizes the highlighted TMS regions, which are inputted in the GSAT format.

Chapter 1.1: Phylogenetic Analysis of CaCA Family Members

275 CaCA family protein members are listed in Table 1. These homologues fall into 6 Clusters (Figure 1). Clusters 1, 2, 3, 4, 5 and 6 can be subdivided into 10, 7, 14, 6, 9 and 9 sub-clusters, respectively.

Cluster 1 (50 proteins) has an average protein size of 450 ± 86 residues (Figure 2). Sub-cluster 1-4 proteins are from bacteria except for Mba1, which is from an archaeon. Clustering of a single archaeal member with bacterial proteins, suggests that Mba1 arose horizontal gene transfer (HGT) from the bacterial domain. Although sub-clusters 1-4 contain proteins from a broad range of organisms, they exhibit fairly uniform protein sizes: 362 ± 12 , 381 ± 31 , 369 ± 35 and 363 residues, respectively. No protein deviates substantially from this size range. Sub-cluster 5 (2 homologues) has an average protein size of 433 ± 199 residues. Cci1 (517 residues) is much larger than Hor1 (349 residues). Based on a multiple alignment, Cci1 contains extra residues at its N-terminus, a characteristic of several other homologues. This domain was not recognized by the Conserved Domain Database (CDD) (NCBI). Sub-cluster 6-9 proteins cluster closely together and are mostly derived from Fungi except for Zma1, a single plant protein sandwiched in between fungal homologues, which, if not due to contaminating DNA, could have arisen by HGT. These 4 sub-clusters have similar but significantly different protein sizes: 448 ± 79 , 500 ± 64 , 455 ± 49 and 573 ± 53 residues, respectively. Sub-clusters containing fungal homologues have longer sequences than sub-clusters derived from bacteria as expected (Chung et al, 2001). Based on the multiple alignment, fungal homologues often contain N-terminal extensions not recognized by CDD. Sub-Cluster 10

(7 proteins with an average size of 466 ± 93 residues) contains proteins from different organismal types. Protein sizes are fairly uniform, except for Pan1 with 654 residues. Pan1 contains a long hydrophilic extension at its N-terminus as well as a long cytoplasmic hydrophilic region between TMSs 6 and 7. In general, comparing sub-clusters 1-4 with sub-clusters 6-10, the average protein sizes of the eukaryotic homologues are substantially larger than for the bacterial proteins, and sub-cluster 9 contains the largest proteins of all. Based on the multiple alignment, the extra residues are present consistently at the N-termini, exclusively in fungal homologues.

Cluster 2 (35 proteins) has an average protein size of 996 ± 226 residues (Figure 3), much longer than Cluster 1 homologues. All sub-clusters in Cluster 2 derive from fungi except for sub-cluster 2 (8 proteins) which are from trichomonadidae, heterolobosea and metazoan. Sub-cluster 2 proteins are shorter than members of the other sub-clusters. Within each sub-cluster, no protein is outside its own average size range except for Uma1, Pgr1 and Cne1 in sub-cluster 3. These proteins have extra N-terminal extensions in the cytoplasm and a hydrophilic loop between TMS 8 and 9. From an NCBI Blast search with Uma1 as the query, there is a DUF307 conserved domain at the N-terminus (CDD). Moreover, some bacteria possess proteins that show two DUF307 domains repeats, DUF 307 may involved in dimerization of transmembrane proteins.

Cluster 2 fungal proteins have long sequences due to N-terminal extensions. Members of Clusters 1 and 2 were used as query sequences to do TC-BLAST searches. The proteins from Cluster 1 are homologous to the vacuolar $\text{Ca}^{2+}:\text{H}^{+}$ exchanger sub-family (TC#2.A.19.2.8), while proteins from Cluster 2 brought up low affinity vacuolar monovalent cation sub-family member (TC#2.A.19.7.1). Segregation between these two

clusters with significant size differences may be due to functional differences (see the Discussion Section).

Cluster 3 (30 proteins) has an average protein size of 380 ± 19 residues, and all are from bacteria. Cluster 3 is sub-divided into 14 small sub-clusters (Figure 4). Sub-clusters 1, 2, 10, 12 and 13 each contains a single protein, and these sub-clusters are deeply rooted, indicating early divergence. All of these proteins have sizes within the average size range. Sub-clusters 3-6 cluster together and have similar protein sizes: 390 ± 25 , 361 ± 25 , 374 ± 25 and 378 ± 10 residues, respectively, although most of the homologues are from taxonomically diverse organisms such as β -proteobacteria, actinobacteria, α -proteobacteria, bacteroidetes, γ -proteobacteria and δ -proteobacteria. Sub-cluster 7 (3 α -proteobacterial proteins) is separated from the other sub-clusters (average protein size of 379 ± 9 residues). Sub-Cluster 8 (4 proteins) exhibits an average protein size of 394 ± 35 residues. Bfa1 and Cef1 branch closely together and far from the tree center. Bfa1 (440 residues) has extra residues at its N-terminus. Sub-clusters 9 and 11 have similar protein sizes (~ 390 residues), and no protein is outside of the average size range, although sub-cluster 9 proteins derive from diverse organisms. Sub-cluster 14 shows late divergence between Avi2 and Psy1. The average size (381 ± 15 residues) is similar to those of the other sub-clusters.

Cluster 4 (26 proteins) has an average protein size of 896 ± 67 residues and derives from metazoans except for Mbr1, Pma1 and Hvu1, which are from codosigidae, perkinsidae and viridiplantae, respectively. Cluster 4 is sub-divided into 6 sub-clusters (Figure 5). Sub-cluster 1 (3 proteins) branches distantly from the tree center, with

paralogues Cel1 and Asu2 clustering together, suggesting a consequence of recent speciation. The protein sizes for sub-clusters 1-4 are fairly uniform (829 ± 22 , 874 ± 31 , 890 ± 46 and 825 ± 18 residues, respectively). No protein is outside of the average size range except Tad3 in sub-cluster 3 (945 residues). The extra region is between TMSs 6 and 7. Sub-clusters 5 (6 proteins) and 6 (8 proteins) cluster closely together. The average protein size for sub-cluster 5 is 913 ± 31 residues. Asu1 and Cbr1 may have arisen recently by speciation since they cluster close to each other. Sub-cluster 6 exhibits an average size of 944 ± 84 residues. Odi1 has an N terminal extension located extracellularly according to HMMTOP.

The protein sequences in Cluster 4 are significantly larger than those in the other clusters. Alignment positions 270 to 480 and 1210 to 1450 exhibit high degrees of conservation, but sequences between positions 480 to 1210 lacked obvious conservation. Odi1 was chosen for the query sequence for a BLAST search of the NCBI Database. The conserved regions were recognized as parts of Na-Ca-exchangers, and the hydrophilic regions were recognized as two cytoplasmic repeated Calx- β superfamily units. Calx- α and Calx- β motifs comprise high affinity calcium binding sites (Schwarz and Benzer, 1997).

Cluster 5 (71 proteins) exhibits an average protein size of 644 ± 183 residues with substantial size variation. It is sub-divided into 9 sub-clusters (Figure 6). Sub-clusters 1-3 from metazoans cluster closely together and have fairly uniform protein sizes: 494 ± 58 , 548 and 546 ± 84 residues, respectively. None of these protein is outside of average size range. Sub-cluster 4 is near sub-cluster 3, but it has varied protein sizes (695 ± 195

residues). Only Mmu3 has long sequence (992 residues) due to an N-terminal extension. Sub-clusters 5 and 6 cluster together, but sub-cluster 5 (617 ± 75 residues) has a larger average protein size than that of sub-cluster 6 (566 ± 104 residues) because of an C-terminal extension in Ada1 (743 residues). Sub-cluster 7 branches distantly from the other sub-clusters, and exhibits greater size variation (790 ± 226 residues). Tni2, Hsa1, Bta1, Ttr1, Mmu2 and Mdo1 have N-terminal extensions and insertions between TMSs 5 and 6 with high degrees of conservation. Using Bta1, CDD recognized the N-terminal region as a K^+ -dependent Na^+/Ca^{2+} exchanger domain. Both potassium-dependent and potassium-independent sodium calcium exchange activities can be found in brain and other tissues, which efficiently regulate the homeostasis of calcium concentrations in smooth muscle, neurons, and astrocytes (Tsoi, Rhee et al. 1998). Mdo1, Mmu2, Ttr1, Bta1 and Hsa1 cluster closely together as do Ttr1 and Bta1, Dme1 and Dwi1 Tni2 and Msa1. All sub-cluster 7 homologues are from Metazoans, except for Mpu1 and Msp1, which are from Viridiplantae, suggestive of HGT. Sub-clusters 8 (2 proteins) and 9 (12 proteins) have similar protein sizes (625 ± 3 and 644 ± 183 residues, respectively), and both sub-clusters are derived from diverse organisms. In sub-cluster 9, Ppa1 and Mma1 show significantly shorter sequences than the other homologues. Ppa1 and Mma1 are the only bacterial homologues in Cluster 5, and both of them are deeply rooted.

Cluster 6 (46 proteins) has an average protein size of 334 ± 25 residues, and all homologues are derived from bacteria except for Ton1, Htu1 and Orf7 which are from archaea. No single protein is outside of the average size range. Based on the tree (Figure 7), sub-clusters 1-4 cluster together and have similar protein sizes, although each sub-cluster is derived from a variety of organisms. Sub-clusters 5-9 cluster closely together

with similar protein sizes. These homologues are from γ -proteobacteria except for Ppa1 and Dgi1 which are from δ -proteobacteria. Sub-clusters 8 and 9 contain single homologues of similar.

Chapter 1.2: Phylogenetic Analysis of MPE Family Members

361 homologues of the MPE family are listed in Table 2. All homologues are derived from prokaryotes. Homologues are divided into 7 clusters (Figure 8). Each cluster falls into sub-clusters, which are shown in Figures 9-14. Clusters 1, 2, 3, 4, 5, 6 and 7 are sub-divided into 6, 2, 7, 6, 10, 1 and 11 sub-clusters, respectively. Sub-cluster 1 is further divided into 3 sub-sub-clusters.

Cluster 1 (129 proteins) (Figure 9) has an average protein size of 391 ± 26 residues (Table 2). All homologues in sub-sub-cluster 1a are derived from α -proteobacteria, but sub-sub-cluster 1b homologues derive from taxonomically diverse organisms such as firmicutes, dictyoglomi and spirochaetes. Sub-sub-cluster 1c proteins are from thermotogae and fusobacteria. Sub-cluster 2 has the largest average protein size in Cluster 1 (421 ± 38 residues). These homologues are derived from bacteroidetes, except for Bmu1 (from a spirochaete). Bmu1 has 364 residues, the smallest protein in sub-cluster 2. Size variation in sub-cluster 3 is minimal, and the average protein size is 390 ± 20 residues. Cae1 exhibits 460 residues, the largest protein in this sub-cluster. Cae1 has a C-terminal hydrophilic extension. Sub-cluster 4 has similar protein sizes as sub-cluster 3 (386 ± 28 residues). In sub-cluster 4, Aac1 has 467 residues due to a C-terminal extension. Sub-cluster 5 has little size variation (386 ± 24 residues). Lor1 has 460 residues, outside of the average size range, due to an N-terminal extension. Sub-cluster 4 and 5 homologues primarily derive from firmicutes, except for Hsp1, Pma1, Lef1, Hma1 and Tgr1 (two aquificae species, nitrospirae, γ -proteobacteria and elusimicrobia, respectively), suggestive of horizontal gene transfer or an early gene duplication event.

Last, sub-cluster 6 homologues are from diverse organisms, but they show uniform protein sizes (399 ± 21 residues).

Cluster 2 (containing 23 homologues) shows uniform protein sizes (401 ± 20 residues) with no homologue substantially outside of the average size range. All cluster 2 homologues are derived from firmicutes except for Hco1, deeply rooted in the tree. Cluster 2 is sub-divided into two sub-clusters (Figure 10). Sub-cluster 1 is at the top of the tree (402 ± 24 residues), and sub-cluster 2 is at the bottom (399 ± 17 residues).

Cluster 3 (containing 70 homologues) has an average protein size of 463 ± 127 residues, larger than for Clusters 1 and 2. Cluster 3 is sub-divided into 7 sub-clusters (Figure 11). Sub-cluster 1 has an average protein size of 433 ± 31 residues. The largest homologue Cle2 exhibits 523 residues with a long hydrophilic region at its C-terminus. Sub-cluster 2 (average protein size of 443 ± 67 residues) includes Eha1 (529 residues), the largest protein in sub-cluster 2 with a C-terminal extension which is responsible for its size expansion. Sub-cluster 3 shows unexpected size uniformity (360 ± 10 residues), although these sub-cluster homologues are derived from varied organisms such as *Deinococcus-Thermus*, *lentisphaerae* and *chloroflexi*. Sub-cluster 5, with two homologues, includes the largest proteins in Cluster 3 (average protein size of 1050 ± 30 residues). Sus1 and Sus2 show C-terminal extensions that were recognized as penicillin binding protein transpeptidase domains (PBP). They can form sub-complexes with FtsW and are involved in murein (peptidoglycan) biosynthesis (Derouaux *et al.*, 2008; Fraipont *et al.*, 2011; Maggi *et al.*, 2008). The Dps1 (521 residues) protein size is outside the size range because of a C-terminal extension. Sub-cluster 7 has an average protein size of

443±43 residues. Osp1 shows a larger protein size (553 residues) due to extra residues at its C- and N-termini. Sub-cluster 8, from actinobacteria, except for Mth1, Tca1, Sno1 (from firmicutes) and Kra1 and Hau1 (from chloroflexi) has an average protein size of 504±127 residues. Bad1, Bbr1, Cac1 and Ccu1 exhibit larger sizes due to C-terminal extensions. These regions were not recognized by CDD. Ccu1 (921 residues) has a large C-terminus. This region was recognized as PBP dimer and transpeptidase domains by CDD. Based on HMMTOP and the positive-inside rule, these domains are in the periplasm. Sub-clusters 1, 2 and 7 homologues are exclusively derived from firmicutes.

Cluster 4 (containing 65 homologues) and Cluster 5 (containing 34 homologues) cluster loosely together and have average protein sizes of 383±22 and 389±18 residues, respectively. Both clusters show size uniformity without exception. Cluster 4 contains 6 sub-clusters with little size variation (Figure 12), although the proteins derive from taxonomically diverse organisms such as thermodesulfobacteria, verrucomicrobia, chlamydiae, planctomycete, nitrospirae, ϵ -proteobacteria, chloroflexi, fusobacteria, lentisphaerae, firmicutes, cyanobacteria, gemmatimonadetes, actinobacteria, Deinococcus-Thermus, δ -proteobacteria, chlorobi and acidobacteria.

The majorities of Cluster 5 homologues are derived from firmicutes and subdivided into 10 small sub-clusters (Figure 13). Sub-cluster 1-6, 8 and 9 homologues are exclusively derived from firmicutes, and size variation is minimal. Sub-cluster 7 has 2 homologues, Fsp1 and Fnu1, both derived from fusobacteria. Sub-cluster 7 (average protein size of 415±3 residues) has larger proteins due to extra residues at their C-termini. Sub-cluster 10 (3 homologues) has an average protein size of 377±45 residues. Gha1 (427 residues) has 47 extra amino acids at its C-terminus.

Cluster 7 (containing 19 homologues) has an average protein size of 451 ± 120 residues, and contains 10 sub-clusters (Figure 14). All homologues are derived from actinobacteria except for Ppil and Ssp1 (from synergistete) and Hpy1 (from an ϵ -proteobacterium). Sub-clusters 1 and 2 have similar sizes (~ 380 residues). However, sub-cluster 3 has an average protein size of 695 ± 312 residues. Rfa1 has a periplasmic C-terminal transpeptidase domain, functional for peptidoglycan synthesis. Transpeptidase restructures peptidoglycan while the cell are growing (Fraipont, 2011). Sub-clusters 4, 5 and 7 each consist of a single protein. Sub-clusters 6, 8 and 11 proteins all have similar sizes (~ 420 residues) with no protein outside of the average size range. Sub-cluster 10 has an average protein size of 435 ± 85 residues; but one protein Bli1, has an extra N-terminal region.

Chapter 2.1: Evidence for Vertical Descent (Orthology) for Members of the CaCA

Family

The phylogenetic tree of 16S rRNAs from all genres from which CaCA family homologues included in our study are derived is shown in Figure 15. The bacterial branch locates at the top, two archaea are on the right hand side, and all eukaryotes can be found at the bottom of the tree. α -, β - γ - and δ -proteobacteria cluster closely together at the top of the tree. Firmicutes cluster next to the α -proteobacteria on the left side of the tree, and three cyanobacteria cluster closely with the firmicutes. A single member of the Deinococcus-Thermus group clusters with bacteroidetes. chloroflexi, actinobacteria, spirochetes and thermogogales cluster loosely together. In the eukaryotic branch, metazoans cluster on the right side of the branch while viridiplantae, codonosigidae and fungi cluster closely together on the left side.

Comparing proteins with the 16S rRNAs tree, Cluster 1 contains a number of non-adjacent paralogues and shows different cluster patterns between the two trees, and hence, orthology is hard to establish. In sub-clusters 1-5, prokaryotic homologues are observed. Sub-cluster 1 homologues are primarily derived from cyanobacteria, and the cluster patterns suggest vertical inheritance. However, Sub-cluster 2 homologues derive from taxonomically distinct organisms such as β -proteobacteria, δ -proteobacteria, firmicutes, actinobacteria and γ -proteobacteria. In sub-clusters 1-5, orthology is difficult to establish. In sub-clusters 6-9, vertical descent appears likely for homologues Aor1, Uma1, Cth1 and Sja1 (from *Aspergillus*, *Ustilago*, *Chaetomium*, and *Schizosaccharomyces*, respectively). Comparing the protein and 16S rRNA trees, plant Zma1 unexpectedly clusters between Pch2 and Pan2, both derived from fungi. Horizontal

gene transfer may have occurred for Zma1 (Table 15). Paralogues observed in Cluster 1 are shown in Table 3. Two homologues are from organisms also represented in Cluster 5, while three homologues are from organisms also found in Cluster 2. Two homologues are from organisms possessing paralogues in Cluster 3, and one is from an organism represented in Cluster 4. Finally, six homologues are from organisms also represented in Cluster 1 (Table 16). In prokaryotic protein sub-clusters, vertical descent could not be established; however, eukaryotic homologue sub-clusters exhibit similar clustering patterns when comparing the protein and the 16S rRNA trees, suggesting vertical inheritance.

Cluster 2 homologues have similar clustering patterns when proteins and 16S rRNA trees are compared. Vertical descent is likely for fungal homologues, Spo1, Yli1, Sce1, Zro1, Ctr1 and Dha1. A few paralogues are found in Cluster 2 (Table 4). Three homologues are from organisms also represented in Cluster 1. Paralogues Tva1 and Tva2 cluster closely together as do Ngr1 and Ngr2. This suggests that recent gene duplication events have occurred. Examples of vertical inheritance are likely in Cluster 2.

Cluster 3 homologues show little evidence of orthology, for example, Aca1 and Nsp1 (from *Azohizobium* and *Novosphingobium*, respectively) show relationships suggesting of orthology. Tsp1 and Pn1 (from *Thiomonas* and *Polynucleobacter*, respectively) also appear orthologous. Sub-cluster 7 homologues, Msp1, Xau1 and Mlo1 (from *Methylocystis*, *Xanthobacter* and *Mesorhizobium*, respectively) exhibit clustering similar to that in the 16S rRNA tree. In sub-clusters 8 and 9, these proteins derive from highly divergent organisms. In sub-clusters 10 and 11, vertical descent is likely for homologues, Mcal, Sba1 and Eco1 (from *Methylococcus*, *Shewanella* and *Escherichia*,

respectively), but sub-clusters 12-14 homologues are from divergent organismal types. Two homologues in Cluster 3 had paralogues in Cluster 1 (Table 5 & 16).

In Cluster 4, vertical descent could not be established. Hvu1 a single member of Viridiplantae protein, Hvu1 cluster with Metazoans proteins, suggesting HTG (Table 15). Several pairs of paralogues were observed of different organisms (Table 6). Ten Cluster 4 homologues had paralogues in Cluster 5 and one organism had paralogues in Clusters 1 & 2 (Table 16). Further, six homologues had paralogues in Cluster 4. Most paralogues were found in different sub-clusters, except for Dre1 and Dre2 (from *Danio rerio*), which cluster together in sub-cluster 6, suggestive of a recent gene duplication event.

In Cluster 5, vertical descent could be established. In sub-clusters 1-6, vertical descent is likely for homologues, Dwi2, Dsi1, Dps2, Ame1, Der1, Spu1 and Dps1 (from several *Drosophila* species, *Apis* and *Strongylocentrotus*). In sub-cluster 7, homologues Hsa1, Spu2, Dwi1, Dme1, Cre1 and Cbr1 (from *Homo*, *Strongylocentrotus*, two *Drosophila* species, and two *Caenorhabditis* species) show relationships suggestive of orthology. In sub-cluster 9, Pin1 and Ota1 (from *Phytophthora* and *Ostreococcus*) also appear orthologous. Fourteen homologues have paralogues within Cluster 5 (Table 7) with most present in different sub-clusters. Tps2 (from *Thalassiosira pseudonana*) was found in sub-cluster 8, and Tps1 and Tps3 were found clustering closely together in sub-cluster 9, suggestive of a recent gene duplication event. Tni1, Tni2 (in Cluster 5) and Tni3 (in Cluster 4) are from *Tetraodon nigroviridis*. Tni1 and Tni2 cluster closely together, undoubtedly the result of another recent gene duplication event. Two viridiplantae proteins, Mpu1 and Msp1 cluster with metazoan proteins and two members of bacterial proteins Mma1 and Ppa1 cluster with eukaryotic proteins, suggesting HTG

(Table 15). Paralogues Sko1 and Sko2 have similar clustering patterns as paralogues Api2 and Api3. Examples of vertical descent are likely in Cluster 5.

The Cluster 6 homologues are exclusively from prokaryotes; orthology could be established within several sub-clusters. For example, Din1 and Dac1 (from *Desulfobacterium* and *Desulfuromonas*) in sub-cluster 1. Archaeal Ton1, Htu1 and Orf7 cluster with bacterial proteins; suggestive of HGT, and 2 members of δ -proteobacterial proteins, Dgi1 and Ppa1 cluster with γ -proteobacterial proteins, suggesting HGT (Table 15). Homologues in sub-clusters 2-4 are from diverse phyla such as δ -proteobacteria, α -proteobacteria, deinococcus, actinobacteria, cyanobacteria, chloroflexi, γ -proteobacteria, planctomyce, bacteroidetes and lentisphaera. In sub-cluster 5, orthology is likely for Ybe1, Pmi1, Vha1 and Pha1 (from *Yersinia*, *Proteus*, *Vibrio*, and *Pseudoalteromonas*, respectively). In sub-cluster 6, homologues Rbl1, Mth1 and Cbu1 (from *Reinekea*, *Methylophaga* and *Coxiella*, respectively) cluster similarly to the corresponding 16S rRNAs. Only two paralogues in Cluster 6 (Table 8), Kst1 and Kst2, from *Candidatus Kuenenia stuttgartiensis*, cluster closely together in sub-cluster 4, suggesting a recent gene duplication event.

Chapter 2.2: Evidence for Vertical Descent (Orthology) for Members of the MPE

Family

The phylogenetic tree of 16S rRNAs from all genres from which CaCA family homologues included in our study were derived is shown in Figure 16. On the top of the tree, ϵ -, α -, β -, γ -, and δ -proteobacteria cluster together, while chlorobi and bacteroidetes are represented in the same cluster next to the proteobacteria. Firmicutes localize on the lower left. Two cyanobacterial members locate between firmicutes and bacteroidetes. Clusters on the right hand side of the tree include a single members each of the chlamydia, planctomycete, verrucomicrobium and fibrobacter as well as four members of spirochetes. Actinobacteria cluster on the lower right of the tree. Single member of chloroflexi, dictyglomi, and three members each of *Thermus* and *Thermotoga* cluster at the bottom of the tree.

Comparing Cluster 1 of the MPE family protein tree with the 16S rRNA tree, vertical descent is often difficult to establish, because these homologues are from highly divergent organismal types. In sub-cluster 1a, orthology is likely for homologues Sfr1, Ccr1, Ssp1 Gob1, Rca1 and Wen1 (from *Sinorhizobium*, *Caulobacter*, *Sulfitobacter*, *Gluconacetobacter*, and *Rickettsia Wolbachia*, respectively). Sub-cluster 3 homologues are from taxonomically diverse organisms such as spirochaetes, chlamydia, verrucomicrobia, γ -proteabacteria, planctomycetes, lentisphaerae, elusimicrobia, δ -proteobacteria, deferribacteres, actinobacteria, firmicutes, zetaproteobacteria, β -proteobacteria, acidobacteria, chrysiogenetes and cyanobacteria. Horizontal gene transfer was rampant during the evolution of these homologues as revealed by comparisons with the 16S rRNA tree. The majority of sub-cluster 4 and 5 homologues are from firmicutes.

These homologues show relationships suggesting orthology. Sub-cluster 6 homologues derive from diverse organisms such as chlorobi, gemmatimonadetes, fibrobacteres, firmicutes, bacteroidetes, aquafina and chloroflexi, suggestive of horizontal gene transfer.

Forty-four homologues in Cluster 1 are from organisms also represented in the other clusters (Table 9). Ceu1 and Ceu2 are full-length paralogues from organism *Coprococcus eutactus*, and clustering closely together in sub-cluster 6, as expected for a recent gene duplication event. One homologue is from *Eubacterium dolichum* also represented in Cluster 2, while eight homologues are from organisms also found in Cluster 3. Thirty homologues had paralogues in Cluster 4 and eight homologues are from organisms also represented in Cluster 5. A single homologue is from an organism also represented in Cluster 7 (Table 16). It is difficult to know if horizontal gene transfer has occurred, as all of these homologues could have arisen by vertical transmission from multiple precursor paralogues.

Cluster 2 homologues have similar clustering patterns when the protein and the 16S rRNA trees are compared. Erh1, Edo1 and Csp1 (from *Erysipelothrix*, *Eubacterium* and *Clostridium*, respectively) show relationships suggesting orthology. Lme1, Ooe1 and Wpa1 (from *Leuconostoc*, *Oenococcus*, and *Weissella*, respectively) also appear orthologous. Single member of Haloplasmataceae (Hco1) clusters with firmicutes, suggestive of HGT (Table 18). Five proteins have paralogues in the other clusters (Table 10). One homologue is from an organism also represented in Cluster 1, and four homologues had paralogues in Cluster 5 (Table 16).

Cluster 3 homologues show divergent patterns similar to those of the other clusters. In sub-cluster 1, vertical descent is likely for homologues Cha1, Dha1 and Ame2

(from *Clostridium*, *Desulfitobacterium* and *Alkaliphilus*, respectively). Dps1 and Avi1 (from *Desulfotalea* and *Allochromatium*) show clustering patterns on the 16S rRNA tree in correspondence with the protein tree. In sub-cluster 7, Dha2, Ame1, Cth1, Eve1 and Bpr1 (from *Desulfitobacterium*, *Alkaliphilus*, *Clostridium*, *Eubacterium* and *Butyribibrio*, respectively) exhibit clustering similar to that in the 16S rRNA tree. In sub-cluster 8, Twh1, Cmi1, Bad1, Mle1, Tfu1 and Afe1 (from *Tropheryma*, *Clavibacter*, *Bifidobacterium*, *Mycobacterium*, *Thermobifida* and *Acidimicrobium*, respectively) show relationships suggesting of orthology. Twenty-two proteins had paralogues (Table 11). Eight homologues are from organisms also represented in Cluster 1. Seven homologues are from organisms also represented in Cluster 4. Four homologues had paralogues in Cluster 5, and a single homologue is from an organism also represented in Cluster 7. Finally, three homologues are from organisms also represented in distinct sub-clusters of Cluster 3. Sus1 and Sus2 in sub-cluster 5 are full-length paralogues from *Candidatus Solibacter usitatus* and cluster distantly from the center of the tree, suggesting that they arose from a recent gene duplication event.

Clusters 1 and 4 have similar clustering patterns; Cluster 4 homologues are from divergent organismal types. Paralogues were observed in Cluster 4 (Table 12); thirty homologues are from organisms also represented in Cluster 1. Seven homologues are from organisms also represented in Cluster 3. One protein has a paralogue in Cluster 5. Finally, three homologues are from organisms also represented in Cluster 7. No protein has paralogues within Cluster 4, and many paralogues were observed between Clusters 1 and 4 (Table 16). Vertical decent is difficult to establish even within sub-clusters, since

Cluster 3 homologues are derived from distinct phyla such as proteobacteria, firmicutes, and actinobacteria.

In contrast to Cluster 4, vertical descent is likely for Cluster 5 homologues, for example, for Sep1 and Oih1 (from *Straphylococcus* and *Oceanobacillus*), and for Csp1 and Edo1 (from *Clostridium* and *Eubacterium*). Two members of fusobacteria (Fsp1 and Fnu1) cluster with firmicutes, suggestive of HGT (Table 18). Many proteins in Cluster 5 have paralogues (Table 13). Full length paralogues, Bcl1 and Bcl2 (from *Bacillus clausii*), and Lmo1 and Lmo4 (from *Listeria monocytogenes*) cluster closely together in sub-clusters 3 and 2, respectively, suggesting recent gene duplication events.

Cluster 6 homologues show little evidence of orthology. Cluster 7 homologues show relationships suggesting orthology. Vertical descent appears likely for actinobacterial homologues Fsp1, Rfa1, Cgl1, Ate1, Pac1, Aha1, Bli1, Twh1 and Lxyl (from *Frankia*, *Rhodococcus*, *Corynebacterium*, *Actinoplanes*, *Propionibacterium*, *Arcanobacterium*, *Brevibacterium*, *Tropheryma*, and *Lefsonia*, respectively). Five proteins had paralogues in the other clusters, while four proteins had paralogues in Cluster 4 (Tables 14 and 17).

Chapter 3.1: Topological Analyses of CaCA Family Proteins

Figure 17 shows the average hydropathy, amphipathicity and similarity plots of all CaCA family proteins included in this study. The AveHAS plot reveals 11 hydrophobic peaks corresponding 11 peaks of similarity. The first 6 TMSs cluster together followed by a long hydrophilic region, followed by 5 more TMSs near the end of the alignment. Possibly, TMSs 2-6 are repeated in TMSs 7-11, especially, since the first hydrophobic peak shows a low degree of conservation.

By examining the AveHAS plots for each individual sub-cluster, the topologies were found to be different. The AveHAS plots for Cluster 1 (Figure 18) reveal 11 putative TMSs in a 6+5 arrangement. Comparing the AveHAS plots for Cluster 6 proteins versus those Cluster 1 proteins, the latter shows an extra hydrophobic peak at their N-termini. TMSs 2-6 show greater conservation than TMSs 7-11.

The AveHAS plots for Cluster 2 (Figure 19) shows 13 hydrophobic peaks. A long hydrophilic region is presented at the C-termini of these proteins, and TMS 1 clusters distantly from TMS 2. TMS 2 clusters far from TMS 3. TMSs 3-8 cluster closely together although TMSs 7 and 8 are separated by a small hydrophilic region. TMSs 9-13 cluster closely together while a hydrophilic region separates the sequences into two halves. When comparing the AveHAS plots for all CaCA homologues, Cluster 2 shows 2 more hydrophobic peaks near the C-termini, located at positions 700 and 950, respectively, and both show a high degree of conservation. TMS 3 presents a large hydrophobic peak with a low degree of conservation. Based on the multiple alignment, positions 1150 to 1300 (TMSs 4-8) and 1650 to 1800 (TMSs 9-13) are highly conserved.

The AveHAS plots for Clusters 1 and 3 (Figure 20) exhibit 11 hydrophobic peaks with 6 peaks clustering in the first half, and 5 clustering in the second half. A small hydrophilic region divides the sequences into their 2 halves. Each hydrophobic peak is separated from the others. The first TMS has a low degree of conservation. The highest degree of conservation is seen for positions 380 to 550 (TMSs 7-11).

The AveHAS plots for Clusters 4 (Figure 21) and 5 (Figure 22) exhibit similar patterns with 11 hydrophobic peaks. Peaks 1-5 cluster closely together in the first half, while 6 peaks cluster closely together in the second half. A large hydrophilic region separates the sequences into these 2 halves. However, Cluster 4 and 5 homologues exhibit 10 TMSs instead of 11. The hydrophobic peak 6 is placed in the cytoplasm, since hydrophobicity of this segment is significantly less than those of other TMSs and moreover, some algorithms for prediction of transmembrane helices and topology predict it to be intracellular (Iwamoto et al., 1999). Moreover, AveHAS plots for Cluster 5 show long N-terminal extensions with a low degree of conservation that is caused by the first (TMS 0) of 11 transmembrane segments that is thought to be cleaved by a putative signal peptidase (Visser et al., 2007).

The large hydrophilic regions in Clusters 4 and 5 show a low degree of conservation. There is a hydrophilic domain belonging to the Calx- β superfamily fused to the transporter for some homologues, and examples can be found in metazoans. In Cluster 4, TMSs 1-5 and TMSs 6-10 show high degrees of conservation. In Cluster 5, TMSs 6-10 show the highest degree of conservation.

The AveHAS plots for Cluster 6 (Figure 23) exhibit 10 well conserved hydrophobic peaks. 5 cluster closely together in the first, and 5 more TMSs cluster

closely together in the second half. TMSs 6-8 show greater conservation. Comparing the AveHAS plots for Clusters 1 and 3 with Cluster 6, Clusters 1 and 3 exhibit an extra TMS in their first halves. Comparing the AveHAS plots of Cluster 2 with those of Cluster 6, Cluster 2 contains an extra hydrophilic region and 3 extra peaks at its N-terminus.

Cluster 2, 3, and 5 eukaryotic proteins have larger protein as expected (Table 18). Cluster 1, 2 and 5 proteins have extra residues in their N-termini. Cluster 4 and 5 proteins have extra residues located in the middle of the sequences. No cluster has C-terminal extension.

Chapter 3.2: Topological Analyses of MPE Family Proteins

Figure 24 shows average hydropathy, amphipathicity and similarity (AveHAS) plots for all MPE family homologues included in this study. The AveHAS plots showed up to 11 peaks of hydrophobicity, but several of them were poorly conserved. In fact, low degrees of conservation at positions 100 to 200, 500 to 650 and 950 to 1500 in the multiple alignment are observed. By contrast, the WHAT program predicted from 9 to 12 TMSs for individual proteins. Inaccurate TMS predications are possibly for both the WHAT and AveHAS program. Many homologues were retrieved using Protocol 1 with a 40% cut off to eliminate redundancies and close homologues. Therefore, homologues included in this study are quite distantly related to each other. Because of the poor reliability of the AveHAS and WHAT predication, homologues were analyzed by individual cluster.

The AveHAS plots for Cluster 1 (Figure 25) reveal 10 hydrophobic peaks with a high degree of conservation between positions 100 and 600. Large hydrophilic regions with a low degree of conservation were observed at both C- and N-termini. Peak 1 is a large and distant from peaks 2-3 which cluster closely together. Peak 4, a large hydrophobic peak, shows a higher degree of conservation compared to TMSs 2-3. Following TMS 4 are 3 small peaks, which may represent TMSs 5, 6 and 7. Peaks 8-10 cluster together near the C-termini. Peak 8 is a large and hydrophobic peak; a small peak precedes it, but it could be periplasmic, since hydrophobicity of this peak is significantly less than those of other peaks, and HMMTOP predict it to be extracellular. Cluster 1 shows a high degree of conservation for peaks 1-2 and peaks 8-10.

The AveHAS plots for Clusters 2 (Figure 26) and 5 (Figure 29) exhibit similar patterns with 10 hydrophobic peaks. Compared to the AveHAS plot for Cluster 1, hydrophilic regions at both the N- and C-termini are absent in Clusters 2 and 5. Moreover, Cluster 2 shows clearer separation between peaks 6 and 7 compared to Cluster 5. TMSs 8-10 exhibit a high degree of conservation.

The AveHAS plots for Cluster 3 (Figure 27) has 10 hydrophobic peaks. In these plots, alignment positions 100 to 150 and 700 to 1200 show poor conservation while putative TMSs 8-10 are well conserved. Comparing AveHAS plots for Clusters 1 and 3, Cluster 3 shows long C-terminal extensions due to the fusion domain PBP.

The AveHAS plots for Cluster 4 (Figure 28) reveal 10 hydrophobic peaks as well. Comparing the AveHAS plots for Cluster 2 with those for Cluster 4, TMSs 1-4 and TMSs 8-10 of Cluster 4 appears similar to TMSs 1-4 and TMSs 8-10 of Cluster 2, respectively. However, the separation between 5-7 in Cluster 4 is not clear. Putative TMSs 8-10 show a high degree of conservation.

The AveHAS plots for Cluster 6 (Figure 30) exhibit 9 hydrophobic peaks. The 9 peaks are separated into 3 groups, and each containing 3 peaks. TMSs 4 of Cluster 2 are absent in Cluster 6.

The AveHAS plots for Cluster 7 (Figure 31) reveals 10 hydrophobic peaks, with clustering patterns similar to those of Clusters 1, 2 and 5. However, Cluster 7 homologues have extra regions at their C-termini with a low degree of conservation. A fusion event probably caused these C-terminal extension. TMSs 8-10 have the highest degree of conservation.

All proteins in MPE family have 10 TMSs on AveHAS plots (Table 19). The average protein size for each cluster is similar and uniform except for Clusters 3 and 7 due to the fusion domain at their C-terminus.

Chapter 4.1: Establishment of Internal Repeats in CaCA Family Proteins

The majority of homologues in the CaCA family exhibit 10 TMSs, but some of them appear to have 11 TMSs. We will try to establish the evolutionary pathway by which different members of the CaCA family arose by examining potential internal repeats.

From our analysis of the CaCA family, two duplicated units can be recognized by the Conserved Domain Database (CDD) as Na⁺/Ca²⁺ exchange family. Two repeat domains separate the sequence to into two halves (Saier et al., 1999). Of the 10 TMS homologues, we expected repeats of 5 TMSs. By using the Ancient Rep program, 5 TMS repeats could be established. Although 5 TMSs repeats were observed for all CaCA homologues, individual phylogenetic clusters display different AveHAS plots. Thus, the topologies were found to be different in each clusters due to extra TMSs in some of them.

Cluster 1 has 11 TMSs on the AveHAS plots in a 6+5 arrangement. By using the Ancient Rep and HMMGap program, 5 TMS repeats were confirmed, and an extra TMS was observed at the N-termini. Comparing TMSs 2-6 of Dac1 (gi: 328954050) with TMSs 7-11 of Cre1 (gi: 212374355) gave a binary alignment showing 23% identity, 39.6% similarity and a comparison score of 12 S.D. (Figure 32). (Both proteins exhibit 11 TMSs based on the WHAT program).

Cluster 2 has 13 peaks of average hydrophobicity in the AveHAS plots. TMSs 1 and 2 are located in the N-terminal region far from other hydrophobic peaks. TMSs 3-8 are shown in the first half of the sequence with TMSs 9-13 in the second half. Comparing Dac1 (exhibits 11 TMSs) in Cluster 1 with Mla1 in Cluster 2 (gi: 328858470) gave a comparison score of 34 S.D, and TMSs 1-11 of Dac1 aligned with TMSs 3-13 of Mla1

(Figure 33). Based on the superfamily principle, TMSs 4-8 is a repeat unit of TMSs 9-13 (of Mla1).

The AveHAS plot for Cluster 3 shows 11 TMSs in a 6+5 arrangement. Dac1 (Cluster 1) exhibits two with 5 TMS repeats, and an extra N-terminal TMS. Comparing Dac1 and Chu1 (Cluster 3) gave a comparison score of 16 S.D. with TMSs 1-11 of Dac1 aligning with TMSs 1-11 of Chu1. Based on the superfamily principle, Cluster 3 homologues have 5 TMS repeats and an extra TMS at their N-termini.

The AveHAS plots for Clusters 4, 5 and 6 reveal 10 TMSs. Mth1 (gi: 254490075) and Ppa1 (gi: 149925192) exhibit 10 TMSs (WHAT program). TMSs 1-5 of Ppa1 aligned with TMSs 6-10 of Mth1 with 28.7% identity, 48.5% similarity and a comparison score of 18 S.D. (Figure 35).

Next we will examined the probable evolutionary origin of the 5 TMSs repeat unit and its. Comparing TMSs 2-3 of Ada5 (10 TMSs) with TMSs 4-5 of Dps6 (10TMS), the binary alignment showed 30.8% identity, 63.5% similarity and a comparison score of 15 S.D. (Figure 36). The first TMS of the repeat unit did not give a good score to any internal region of CaCA family proteins. Following the superfamily principle, duplication between TMSs 2-3 and TMSs 4-5 suggests duplication between TMSs 7-8 and TMSs 9-10. Furthermore, TMSs 7-8 of Ada5 (gi: 312381030) aligned with TMSs 9-10 of Hma2 (gi: 221122333) with 25.7% identity, 48.6% similarity, and a comparison score of 11S.D. (Figure 37). Our results show that the 5 TMSs repeat unit likely arose from a 2-TMSs duplication.

In conclusion, two evolutionary pathways are possible for CaCA family proteins. First, homologues could have arisen in the following three steps: 1) Duplication of the

primordial 2 TMSs to give a 4 TMSs, 2) the addition of one more TMS at the N-terminus of the 4 TMSs to give 5 TMSs, and 3) duplication of the 5 TMSs to give 10 TMSs. The second possible pathway is as follows, 1) Triplication of the primordial 2 TMSs could have given 6 TMSs, 2) lose one TMS at the N-terminus of the 6 TMSs could then have given 5 TMSs total, and 3) duplication of the 5 TMS unit could have produced the 10 TMSs proteins. The additional TMSs in a some homologues may be due to late gene fusion events.

Chapter 4.2: Internal Repeats in MPE Family Proteins

In Table 22, we have listed possible combinations of repeat units. TMSs 2-3 of Nsa2 (gi: 319957313) aligned with TMSs 4-5 of Mtu2 (gi: 289760114) with 39.6% identity, 47.9% similarity, and a comparison score of 11S.D. (Figure 39), suggesting 2 TMSs may have given rise to the other TMSs.

The comparison scores for 3 TMS and 4 TMS repeats (around 6 S.D.) are insufficient to establish homology. Comparing TMSs 1-5 with TMSs 6-10 gave a comparison score of 9 S.D. Although this score is not sufficient to establish homology, it exhibits the highest comparison scores for these combination, moreover, TMSs 1-2 aligning with TMSs 6-7 gave a comparison score of 9 S.D., which supports the 5 TMS repeat unit as demonstrated from the CaCA family.

Chapter 5: Establishing Homology Between Superfamilies

The program Protocol 2 can be used to detect the similarity between two proteins. Comparison scores between superfamilies studied and defined previously are presented in Table 20. A comparison score > 11 S.D. is considered sufficient to establish homology.

First, homology between the MPE and CaCA families was established. The comparison score between Aex2 (gi: 315498735) of the CaCA family and Shbe1 (gi: 163751823) of the MPE family is 12 S.D. According to Protocol 2 (Figure 38), Aex2 TMSs 3-5 aligned with Shbe1 TMSs 3-5, and Aex2 TMSs 8-10 aligned with Shbe1 TMSs 8-10. Both proteins exhibit 10 TMSs. Moreover, TMSs 1-4 of a MPE family protein (gi: 296101252) aligned with TMSs 6-9 of a CaCA family protein (gi: 62181822) with a GSAT score of 10 S.D, suggesting 5 TMS repeats for MPE family proteins.

Additional superfamilies from TCDB were examined with Protocol 2. The highest comparison scores are presented in Table 20.

CDF with CACA

The Cation Diffusion Facilitator (CDF) Family (TC#: 2.A.4) showed homology with other superfamilies included in this study. CDF proteins have 6 TMSs with three 2 TMSs repeats (Matias *et al.*, 2010). Homology between the CDF families confirms the presence of the 2 TMS repeats in members of the CaCA family. For example, TMSs 1-5 of a CDF family protein (gi: 21242070) aligned with TMSs 11-15 of a CaCA family protein (gi: 320036119) with a GSAT score of 11 S.D. (Figure 39); suggesting that TMSs 12-13 represent a repeat unit of TMSs 14-15. The extra 5 TMSs are found at its N-terminus compared to 10 TMS CaCA family proteins. Moreover, TMSs 11-15 can be recognized as the Na-Ca-exchange domain by CDD. Since we already have shown that

the 5 TMSs Na-Ca exchange domain arose from 2 TMSs duplication with one more TMS at the N-termini, homology can be inferred.

CDF and MPE

Comparing TMSs 1-4 of Asp14 of the CDF family (gi: 359428766) with TMSs 2-5 of Dha10 (gi: 361855429) of the MPE family gave a comparison score of 13 S.D. (Figure 40), sufficient to establish homology. Asp14 has 6 TMSs and Dha10 has 10 TMSs. CDF family proteins have already been shown to have a triplication of a primordial hairpin structure to give rise to 6 TMSs (Matias et al., 2010). Based on the superfamily principle, MPE family proteins may also have TMSs 2-3 repeated in 4-5; TMSs 2-3 and 4-5 aligned with a comparison score of 11 S.D. (Figure 39).

CDF with AAAP (APC)

The Amino Acid-Polyamine-Organocation (APC) Family (TC#: 2.A.3) has been shown to consist of members, many of which have 5 TMS repeats to generate 10 TMSs with 2 extra TMSs at their C-termini to make a total of 12 (or more) TMSs (Wong et al., 2009). Comparing TMSs 2-5 of Dfa1 (gi: 328874879) of the AAAP family with TMSs 1-4 of Bju4 (gi: 32274627) of the CDF family gave a GSAT score of 13 S.D. (Figure 41), indicating homology between these two families. The comparison score of 13 S.D. between TMSs 2-5 of an AAAP protein and TMSs 1-4 of a CDF protein suggests 2 TMSs repeats consisting of TMSs 2-3 and 4-5 of Dfa1 (Wong et al., 2012). Therefore, the 2 TMSs may have duplicated to give 4 TMSs, and by adding one more TMS at the N-termini may have given rise to the 5 TMS repeat unit in the APC family as suggested pathway (Wong et al., 2012). Alternatively, triplication of the hairpin structure may have

given rise to a 6 TMS protein, and the 5 TMS unit resulted from loss of the N-terminal TMS.

CDF with DMT

The Drug/Metabolite Transporter (DMT) Superfamily (TC#: 2.A.7) has also been shown to have arisen from a transmembrane 2 TMSs hairpin structure by duplication followed by addition of one more TMS to generate a 5 TMS unit (Lam *et al.*, 2011). This was then duplicated to give 10 TMS proteins (Lam *et al.*, 2011). Comparing TMSs 3-6 of Cps1 of the CDF family (gi: 71279891) with TMSs 7-10 of Pnel of the DMT superfamily (gi: 145589545) gave a comparison score of 11 S.D. (Figure 42), suggesting homology between the CDF and DMT families. CDF proteins exhibit 6 TMSs (Matias *et al.*, 2008) while DMT family members can have 10 TMSs (Lam *et al.*, 2011). These results confirm the proposed 2 TMSs precursor of TMSs 7-8 and 9-10 in the latter family.

TDT with CDF

The Telurite-resistance/Dicarboxylate Transporter (TDT) Family (TC#: 2.A.16), has been shown to have an internal repeats of 5 TMSs (Saier *et al.*, 1999). Comparing TMSs 1-5 of Lsm1 of the TDT family (gi: 312216773) with TMSs 2-6 of Rlt1 of the CDF family (gi: 309778699) gave a comparison score of 11 S.D. (Figure 43), suggesting homology.

RND with CDF

The Resistance-Nodulation-Cell Division (RND) Superfamily (TC#: 2.A.6) appears to exhibit homology with the CDF family. TMSs 1-4 of Etb1 of the CDF family (gi: 146313689) aligned with TMSs 9-12 of Msz1 of the RND family (gi: 336477699) to give a comparison score of 12 S.D. (Figure 44). Furthermore, the 2 TMSs duplication

could be demonstrated in the RND family. Comparing TMSs 9-10 in Min2 (gi: 254821662) with TMSs 11-12 in Aful (gi: 11498828) (both of 12 TMSs) gave a comparison score of 10 S.D. (Figure 45). RND family proteins have a single TMS at their N-termini followed by a large extracytoplasmic domain, followed by six additional TMSs, a second large extracytoplasmic domain, and five final C-terminal TMSs. The first 6 TMSs halves of RND family proteins (TMSs 1-6) are homologous to the second 6TMSs halves (TMSs 7-12) (Tseng et al, 1999).

CaCA with RND

TMSs 3-6 of a CaCA family protein (gi: 281347334) aligned with TMSs 3-6 of an RND family protein (gi: 187927409) with a comparison score of 12 S.D. (Figure 46). TMSs 2-6 represent the first coherent 5 TMSs repeat unit of these RND proteins. The first TMS in RND proteins is distant from the other TMSs as noted above. Therefore, TMSs 2-6 of an RND family protein could be homologous to TMSs 1-5 of a CaCA family protein. Since TMSs 3-4 represent a repeat of TMSs 5-6 of a CaCA family protein, TMSs 3-4 may be homologous to TMSs 5-6 in an RND family protein.

CaCA with APC (AAP)

Comparing TMSs 2-8 of Ppa10 (gi: 168052695) of the CaCA family with TMSs 2-8 of Rop1 (gi: 167999963) of the AAP family gave a comparison score of 13 S.D. (Figure 47). Both Ppa10 and Rop1 have 10 TMSs.

CaCA with DMT

TMSs 6-10 of GlcU (gi: 73661966) of the DMT superfamily were compared to TMSs 6-10 of Ptc1 of the CaCA family (gi: 224131110) yielding a comparison score of 11 S.D. (Figure 48). This comparison suggests homology between the DMT and CaCA

families and supports an evolutionary pathway involving a 5 TMS duplication. Protein GlcU and Ptc1 both have 10 TMSs.

CaCA with TDT

TMSs 7-10 of Slp1 of the CaCA family (gi: 115709819) aligned with TMSs 7-10 of Ntf1 of the TDT family (gi: 119484216) with a GSAT score of 10 S.D. suggestive, but not proof of homology (Figure 49). Both proteins exhibit 10 transmembrane α -helical spanners.

RND with MPE

Comparing TMSs 3-6 of Ntf2 of the RND family (gi: 254787166) with TMSs 2-5 of Etb2 of the MPE family (gi: 149186203) gave a comparison score of 11 S.D. (figure 50). RND has 12 TMSs and MPE has 10 putative TMSs.

MPE with DMT

Comparing TMSs 1-3 of Vms1 of the MPE family (gi: 171914172) with TMSs 1-3 of Lbh1 of the DMT family (gi: 161506918) gave a comparison score of 11 S.D. (Figure 51). Both proteins exhibit 10 TMSs.

MPE with TDT

Comparing TMSs 1-6 of Vat1 of the MPE family (gi: 91228513) with TMSs 1-6 of Arb1 of the TDT family (gi: 116662088) gave a comparison score of 11 S.D. (Figure 52). Both proteins have 10 TMSs.

TDT with RND

Comparing TMSs 7-10 of Rer1 (gi: 226305474) of the TDT family (10 TMSs) with TMSs 3-6 of Ttu2 (gi: 254787166) of the RND family (12 TMSs) gave a comparison score of 13 S.D. (Figure 53).

RND with DMT

Comparing TMSs 8-12 of Cot1 of the RND family (gi: 344923545) with TMSs 6-10 of Ptm1 of the DMT family (gi: 359436732) gave a comparison score of 10 S.D. (Figure 54). RND family proteins present 12 TMSs and DMT family proteins have 10 TMSs.

APC with DMT

Comparing TMSs 7-10 of Afl15 (gi: 238507023) of the AAAP family with TMSs 7-10 of Ami2 (gi: 223042098) of the DMT family gave a comparison score of 14 S.D. (Figure 55), and both proteins exhibit 10 TMSs.

TDT with DMT

Comparing TMSs 2-6 of Ylt1 (gi: 50553858) of the TDT family with TMSs 2-6 of Rcm1 (gi: 255581410) of the DMT family gave a comparison score of 12 S.D. (Figure 56). Both proteins have 10 TMSs.

TDT with APC

Comparing TMSs 1-7 of Cmi2 (gi: 346321507) of the TDT family with TMSs 1-7 of Hvu2 (gi: 326519116) of the AAAP family gave a comparison score of 15 S.D. (Figure 57). Cmi2 has 10 TMSs while Hvu2 has 11 TMSs. APC family proteins have two 5 TMS internal repeats with additional TMSs at their C-termini. Therefore, the alignment makes evolutionary sense.

RND with AAAP (APC)

Comparing TMSs 4-6 of Ple1 (gi: 330446251) of the RND family with TMSs 8-10 of Ggr3 (gi: 310799985) of the AAAP family gave a comparison score of 11 S.D. (Figure 58). Ple1 has 12 TMSs, while Ggr3 has 11 TMSs.

MPE with AAAP (APC)

Comparing TMSs 7-10 of Afu3 (gi: 146324801) of the AAAP family with TMSs 7-10 of Asp17 (gi: 320093976) of the MPE family gave a comparison score of 12 S.D. (Figure 59). Afu3 has 11 TMSs and Asp17 exhibits 10 TMSs

CDF and AAAP (APC)

Comparing TMSs 2-5 of Dfa1 (gi: 328874879) of the AAAP family with TMSs 1-4 of Bju4 (gi: 32274627) of the CDF family gave a comparison score of 13 S.D. (Figure 60). Dfa1 presents 11 TMSs, and Bju4 has 6 TMSs.

CDF and DMT

Comparing TMSs 3-6 of Cps1 of the CDF family (gi: 71279891) with TMSs 7-10 of Pne1 of the DMT family (gi: 145589545) gave a comparison score of 11 S.D. (Figure 61). The comparison score of 11 S.D. suggests homology between the CDF and DMT families. Cps1 has 6 TMSs, and Pne1 has 10 TMSs.

Discussion

In this paper, we describe CaCA family homologues and define the evolutionary pathway by which these proteins arose. Two pathways are suggested. First, a genetic element encoding a 2 TMSs precursor duplicated to give 4 TMSs; this element added one TMS at its N-terminus to give 5 TMSs, and 5 TMSs duplicated to give the final 10 TMSs topology. Second, triplication of a 2 TMSs precursor gave 6 TMSs, 6 TMSs lost one TMS at its N-terminus, leaving 5 TMSs, 5 TMSs duplicated to give a 10 TMSs proteins (Figure 62). Several families have been shown to have taken a similar evolutionary pathway by arising from the 2 TMSs precursor. These families all possibly related because they derived from the same 2 TMSs element are presented in Figure 62 with their most likely evolutionary pathway.

Establishing homology between families is based on 3 criteria. First, the percentage of similarity and identity, when expressed in a comparison score, must be equal to or higher than 11 S.D. Second, the compared protein sequences of the homologues from both families must exhibit at least 60 residues and 3 transmembrane domains. And last, the alignment between two segments must make evolutionary sense; repeat units should show consistency between the two family homologues.

The putative 5 TMS repeats in MPE family homologues were insufficiently similar to allow establishment of homology. Comparing TMSs 1-5 and TMSs 6-10 of MPE family proteins gave a comparison score of 9 S.D, but comparing TMSs 1-3 and TMSs 8-10 gave a highest comparison score of 7 (Table 22). MPE family proteins are likely to have 5 TMS repeats. Homology between CaCA and MPE family members could be established, and based on the superfamily principle, the MPE family homologues are

likely to exhibit the same duplications as CaCA family homologues. MPE family homologues show an average of 10 TMSs, again suggesting a 5 TMSs duplication. From the Protocol 2 results, TMSs 3-5 of a homologue of the CaCA family aligned with TMSs 3-5 of a homologue of the MPE family, and TMSs 8-10 of these two homologues aligned as well. TMSs 1-5 and TMSs 6-10 proved to be duplication units in CaCA family homologues. Therefore, TMSs 1-5 are presumed to be duplication units of TMSs 6-10 in both families.

Homologues of the CaCA family exhibit 9 to 13 putative TMSs. The AveHAS plots for Clusters 1 and 3 present 11 TMSs; TMSs 2-6 and 7-11 proved to be 5 TMS repeat units. Cluster 2 homologues exhibit 13 probable TMSs, and we have found internal repeats encompassing TMSs 4-8 and 9-13. Cluster 4, 5 and 6 homologues have 10 TMSs with repeat units between TMSs 1-5 and 6-10. Individual homologues in different clusters may serve different functions.

CaCA family homologues are found in both eukaryotes and prokaryotes. The majority of eukaryotic homologues are from fungi and metazoans.

In Cluster 1 of the CaCA family, all homologues are from the CAXs sub-family of cation exchangers (TC#: 2.A.19.2) and exhibit an 11 TMS topology. Homologues are from both bacteria and eukaryotes. Protein sizes displayed unexpected uniformity, but homologues are derived from distantly related organisms. Cluster 1 homologues facilitate the exchange between calcium and protons in vacuoles. An AveHAS plot confirmed the 11 TMSs topology. Due to the odd number of TMSs, the two termini should localize to opposite sides of the membrane. The N-termini proved to be in the cytoplasm (Segarra,

2008). 5 TMSs internal repeats for homologues were found, suggesting that the first TMS is the extra one.

Vertical decent was often difficult to establish for prokaryotic homologues when comparing the proteins and the 16S rRNA trees, but eukaryotic homologues usually showed similar patterns between the two trees, suggesting a preponderance of vertical inheritance. Both eukaryotic and prokaryotic proteins are deeply rooted, suggesting that they arose early to perform dissimilar central cellular functions or participate in transport complexes composed of multiple subunits. Cluster 1 homologues have paralogues in Clusters 2-5, suggests early divergence.

Cluster 2 homologues are found in the sub-family of low affinity vacuolar monovalent cation H^+ antiporter (TC#: 2.A.19.7). Most of homologues are from fungi. Homologues in this sub-family are unable to mediate Ca^{2+} transport but catalyze exchange between Na^+ , H^+ and K^+ . Moreover, sequence analyses revealed the absence of key amino acids shown to be essential for Ca^{2+}/H^+ exchange (Cagnac *et al.*, 2007). These sub-family homologues are responsible for vacuolar Na^+/H^+ and K^+/H^+ , but not Ca^{2+} exchange. There is no evidence of horizontal gene transfer within this cluster, and orthology appears likely when comparing proteins with the 16S rRNAs. The AveHAS plot for Cluster 2 proteins shows 13 probable TMSs, with the three extra TMSs at the N-termini. Fusion domain DUF307 is responsible for the long N-terminal regions.

Cluster 3 homologues show a close relationship with the $Ca^{2+}:H^+$ antiporter of TC#: 2.A.19.1, and all these homologues are of prokaryotic origin. Based on a multiple alignment, greatest conservation is observed for TMSs 7-11. Some of these sub-family

proteins can also mediate K^+ efflux against a K^+ concentration gradient to maintain K^+ homeostasis (Radchenko et al., 2006).

Orthologous relationship could be established within several sub-clusters, and paralogous relationships are observed between Clusters 1 and 3 which exhibit the same topology (11 TMSs). Moreover, Cluster 1 contains both prokaryotic and eukaryotic proteins, but Cluster 3 contains only bacterial proteins. Proteins from Cluster 1 can only mediate the exchange between calcium and proton, but proteins from Cluster 3 can facilitate K^+ efflux as well. It is likely that an early gene duplication event in a CAX ancestral protein occurred that is common to proteins in Clusters 1 and 3. The paralogues were then transmitted vertically to give a rise to members of these two clusters.

All Cluster 4 homologues appear to be from the sub-family of K^+ -independent Na^+/Ca^{2+} exchangers (NCXs) (TC#: 2.A.19.3), and all of these homologues are from eukaryotes. They are plasma membrane proteins, regulating intracellular Ca^{2+} levels by extrusion of calcium to in exchange for sodium (Ren, 2006). Based on the protein tree, Cluster 4 is distant from Clusters 3 and 5. Several paralogues are observed in Cluster 4, and a recent gene duplication gave rise to Dre1 and Dre2. The AveHAS plot shows 11 TMSs and a long hydrophilic region between TMSs 5 and 6. The hydrophilic region can be recognized as tandem units of *Calx*- β motifs, which are located in the cytoplasm, and responsible for the high affinity binding of intracellular calcium (Wu *et al.*, 2011).

In Cluster 5, homologues include metazoan cation/ Ca^{2+} exchangers (CCXs) (TC#: 2.A.19.4). The N-terminal extension was recognized by CDD as the “ K^+ -dependent Na^+/Ca^{2+} exchanger domain”. However, the long hydrophilic region that separates these

two hydrophobic halves was not recognized by CDD. Several paralogues were observed. When compared to bacteria and archaea, eukaryotes show the most proteins paralogues (Makarova, Wolf et al. 2009). Orthology appears likely for Cluster 5 homologues based on 16S rRNA analysis. The clustering pattern in the protein tree suggests early divergence of a common ancestral protein to give a rise to Clusters 4 and 5.

Cluster 6 homologues are recognized as YBRG transporters (TC#: 2.A.19.5). The AveHAS plot for Cluster 6 shows 10 hydrophobic peaks and a small hydrophilic loop between TMSs 5 and 6. These sub-family proteins are essential to maintain cytosolic homeostasis for cell signaling. The 3D structure demonstrated the 5 TMSs repeat units (Liao J, 2012).

In contrast to CaCA family homologues, all MPE family homologues are from bacteria. The AveHAS plots of all clusters show similar patterns. Based on a multiple alignment, the last 4 TMSs show the highest degree of conservation. The conserved region may play a critical functional structural role. Protein size for each clusters are fairly uniform. Clusters 3 and 7 exhibit larger protein sizes due to the presence of the fusion domain, PBP, a penicillin binding protein transpeptidase domain. PBP is located in the periplasm where it can play a role in peptidoglycan synthesis during elongation (Fraipont *et al.*, 2011). The branches in the protein tree are deeply rooted except for cluster 3 which may have resulted from a more recent gene duplication event.

MPE family homologues were found only in bacteria and archaea, and horizontal gene transfer between bacteria was rampant. For example, Clusters 1 and 4 include proteins from diverse phyla. We also observed large numbers of paralogues between

Clusters 1 and 4. An early gene duplication event of a common ancestral protein may have given rise to Cluster 1 and 4 homologues. Horizontal gene transfer events can also be observed in Cluster 2; for example, the single Haloplasmataceae protein clusters with the Firmicutes proteins.

It is interesting that MPE family homologues are found in organisms that lack a peptidoglycan cell wall, organisms such as the planctomycetes, verrucomicrobia, chlamydia and mycoplasma. It will be interesting to know if these proteins play a functional role or are remnants of an earlier age at least the plantctomycetes and verrcomicrobia also possess incomplete sets of the peptidoglycan biosynthesis enzymes (Reynaud & Devos, 2011; Fuerst & Sagulenko, 2011)

Appendix

Table 1. All CaCA family proteins included in this study are listed according to the position of the protein in the phylogenetic tree (clockwise direction) (Figure 1). The average protein sizes provided for each sub-cluster. The Table presents the protein abbreviation, sequence description, organism, protein size, gi number, phylum or kingdom and organismal domain and TC number of closest hit in TCDB.

Abbreviation	Sequence	Description	Organism	ProteinSize	GenBankIndk	Group	Kingdom	TC#
Cluster1								
Sub-Cluster 1								
Ssp1	calcium/proton exchanger		Synechococcus sp. PCC 7335	376	254421487	Cyanobacteri	Bacteria	2.A.19.2.1
Mae1	unnamed protein product		Microcystis aeruginosa PCC 7806	369	159026549	Cyanobacteri	Bacteria	2.A.19.2.1
Mba1	H ⁺ /Ca ²⁺ exchanging protein		Methanosarcina barkeri str. Fusaro	354	73671133	Euryarchaeol	Archaea	2.A.19.2.1
Tsa1	calcium/proton exchanger		Turicibacter sanguinis PC909	348	293374840	Firmicutes	Bacteria	2.A.19.2.1
			Average	362				
			Standard Derivation	13				
Sub-Cluster2								
Ssp2	calcium/proton exchanger		Synechococcus sp. CC9311	381	113954240	Cyanobacteri	Bacteria	2.A.19.2.1
Bja1	unnamed protein product		Bradyrhizobium japonicum USDA 110	445	27377853	Alphaproteob	Bacteria	2.A.19.2.2
Cne1	unnamed protein product		Cupriavidus necator N-1	356	339322741	Betaproteob	Bacteria	2.A.19.2.1
Dac1	CaCA family calcium/proton antiporter		Desulfohalobac acetoxidans DSM 11109	360	328954050	Deltaproteob	Bacteria	2.A.19.2.1
Pvo1	calcium/proton antiporter, CaCA family protein		Paenibacillus vortex V453	357	315646155	Firmicutes	Bacteria	2.A.19.2.7
Cwo1	CaCA family calcium/proton antiporter		Conexibacter woesei DSM 14684	379	284043117	Actinobacter	Bacteria	2.A.19.2.1
Mca1	calcium/proton exchanger		Methylobacterium capsulatus str. Bath	361	53804065	Gammaaprote	Bacteria	2.A.19.2.1
Dra1	cation exchanger		Deinococcus radiodurans R1	406	15805402	Deinococcus	Bacteria	2.A.19.2.2
			Average	381				
			Standard Derivation	31				
Sub-Cluster3								
Rca1	calcium/cation antiporter		Roseiflexus castenholzii DSM 13941	374	156743710	Chloroflexi	Bacteria	2.A.19.2.1
Rxy1	calcium/proton exchanger		Rubrobacter xylanophilus DSM 9941	331	108805365	Actinobacter	Bacteria	2.A.19.2.1
Kra1	calcium/proton antiporter, CaCA family		Ktedonobacter racemifer DSM 44963	401	298247603	Chloroflexi	Bacteria	2.A.19.2.1
			Average	369				
			Standard Derivation	35				
Sub-Cluster4								
Aac1	calcium/proton antiporter, CaCA family		Alcyclobacillus acidocaldarius subsp. acidocaldarius Tc	363	339289772	Firmicutes	Bacteria	2.A.19.2.7
Sub-Cluster5								
Cci1	hypothetical protein CC1G_01313		Coprinopsis cinerea okayama7#130	517	299748017	Fungi	Eukaryota	2.A.19.2.2
Hor1	calcium/proton antiporter		Halothermothrix orenii H 168	349	220932971	Firmicutes	Bacteria	2.A.19.2.1
			Average	433				
			Standard Derivation	119				
Sub-Cluster6								
Man1	vacuolar calcium ion transporter, putative		Metarhizium anisopliae ARSEF 23	404	322703647	Fungi	Eukaryota	2.A.19.2.2
Aor1	vacuolar calcium ion transporter		Aspergillus oryzae RIB40	358	317150008	Fungi	Eukaryota	2.A.19.2.2
Pch2	Pc21g02760		Penicillium chrysogenum Wisconsin 54-1255	421	255953175	Fungi	Eukaryota	2.A.19.2.8
Zma1	unknown		Zea mays	506	194689038	Viridiplantae	Eukaryota	2.A.19.2.5
Pan2	hypothetical protein		Podospora anserina S mat+	551	171688494	Fungi	Eukaryota	2.A.19.2.2
			Average	448				
			Standard Derivation	79				
Sub-Cluster7								
Pch1	Pc12g14430		Penicillium chrysogenum Wisconsin 54-1255	477	255933738	Fungi	Eukaryota	2.A.19.2.8
Trc1	Ca ²⁺ transporter		Trichoderma reesei QM6a	381	340521134	Fungi	Eukaryota	2.A.19.2.2
Pgr2	hypothetical protein PGTG_14272		Puccinia graminis f. sp. tritici CRL 75-36-700-3	541	331239911	Fungi	Eukaryota	2.A.19.2.2
Sci1	hypothetical protein SCHCODRAFT_76678		Schizophyllum commune H4-8	480	302685091	Fungi	Eukaryota	2.A.19.2.8
Cci2	calcium ion transporter		Coprinopsis cinerea okayama7#130	590	169856345	Fungi	Eukaryota	2.A.19.2.2
Sci2	hypothetical protein SCHCODRAFT_70941		Schizophyllum commune H4-8	519	302674539	Fungi	Eukaryota	2.A.19.2.2
Uma1	hypothetical protein UM05132.1		Ustilago maydis 521	547	71022097	Fungi	Eukaryota	2.A.19.2.2
Pgr1	vacuolar calcium ion transporter		Puccinia graminis f. sp. tritici CRL 75-36-700-3	461	331221453	Fungi	Eukaryota	2.A.19.2.2
			Average	500				

Table 1, continued

Sub-Cluster8		Standard Derivation	64
Bde1	hypothetical protein BATDEDRAFT_33427	Batrachochytrium dendrobatidis JAM81	401 328769551 Fungi Eukaryota 2.A.19.2.2
Sma1	hypothetical protein SMAC_03870	Sordaria macrospora k-hell	516 336266004 Fungi Eukaryota 2.A.19.2.8
Cth1	hypothetical protein CTHT_0041470	Chaetomium thermophilum var. thermophilum DSM 1	493 340939046 Fungi Eukaryota 2.A.19.2.8
Nca1	hypothetical protein NCAS_0E03600	Naumovozyma castellii CBS 4309	419 342302654 Fungi Eukaryota 2.A.19.2.2
Aca1	hypothetical protein HCAG_06034	Ajellomyces capsulatus NAM1	487 154275156 Fungi Eukaryota 2.A.19.2.8
Sja1	vacuolar calcium ion transporter	Schizosaccharomyces japonicus Y5275	416 2134110026 Fungi Eukaryota 2.A.19.2.2
Average		Average	455
Standard Derivation		Standard Derivation	49
Sub-Cluster9			
Aor2	unnamed protein product	Aspergillus oryzae RIB40	606 83766993 Fungi Eukaryota 2.A.19.2.2
Aot1	Ca2+/H ⁺ antiporter	Arthroderma otae CBS 113480	559 296815490 Fungi Eukaryota 2.A.19.2.2
Ade1	calcium/proton exchanger	Ajellomyces dermatitidis ATCC 18188	609 327356001 Fungi Eukaryota 2.A.19.2.8
Pte1	hypothetical protein PTT_14927	Pyrenophora teres f. teres 0-1	608 330930188 Fungi Eukaryota 2.A.19.2.2
Tme1	hypothetical protein	Tuber melanosporum Mei28	580 296411643 Fungi Eukaryota 2.A.19.2.2
Average		Average	573
Standard Derivation		Standard Derivation	53
Sub-Cluster10			
Cpa1	calcium antiporter, Na/Ca exchange associated n	Cryptosporidium parvum Iowa II	473 66362248 Apicomplexa Eukaryota 2.A.19.2.8
Pvo1	calcium/proton exchanger	Plasmodium yoelii yoelii 17XNL	440 82593891 Apicomplexa Eukaryota 2.A.19.2.2
Nca2	calcium/proton antiporter, related	Neospora caninum Liverpool	499 325114238 Apicomplexa Eukaryota 2.A.19.2.2
Pma1	vacuolar calcium ion transporter, putative	Perkinsus marinus ATCC 50983	426 294891166 Perkinsida Eukaryota 2.A.19.2.2
Tps1	calcium/proton exchanger, calcium antiporter	Thalassiosira pseudonana CCMP1335	352 224006648 Bacillariophy Eukaryota 2.A.19.2.6
Est1	conserved unknown protein	Ectocarpus siliculosus	441 298705602 Phaeophyceae Eukaryota 2.A.19.2.2
Cre1	cation exchanger 1	Chlamydomonas reinhardtii	447 212374355 Viridiplantae Eukaryota 2.A.19.2.6
Pan1	hypothetical protein	Podospira anserina S mat+	654 171690580 Fungi Eukaryota 2.A.19.2.8
Average		Average	466
Standard Derivation		Standard Derivation	93
Cluster2			
Sub-Cluster1			
Pma1	calcium permease family membrane transporter	Penicillium marneffei ATCC 18224	1117 212535940 Fungi Eukaryota 2.A.19.7.1
Aca1	calcium permease family membrane transporter	Ajellomyces capsulatus H88	1130 325092218 Fungi Eukaryota 2.A.19.7.1
Gcl1	calcium permease family membrane transporter	Grosmanina clavigera kw1407	1155 320594127 Fungi Eukaryota 2.A.19.7.1
Mor1	hypothetical protein MGG_01381	Magnaporthe oryzae 70-15	1158 39951477 Fungi Eukaryota 2.A.19.7.1
Average		Average	1140
Standard Derivation		Standard Derivation	20
Sub-Cluster2			
Tva1	Sodium/calcium exchanger protein	Trichomonas vaginalis G3	611 123423586 Trichomonad Eukaryota 2.A.19.7.1
Tva2	Sodium/calcium exchanger protein	Trichomonas vaginalis G3	614 154419784 Trichomonad Eukaryota 2.A.19.7.1
Ddi1	DUF307 family protein	Dictyostelium discoideum AX4	791 66815529 none Eukaryota 2.A.19.7.1
Dpu1	hypothetical protein DICPUDRAFT_45117	Dictyostelium purpureum	798 330791652 none Eukaryota 2.A.19.7.1
Ppa1	DUF307 family protein	Polysphondylium pallidum PN500	758 281211163 none Eukaryota 2.A.19.7.1
Ngr1	predicted protein	Naegleria gruberi strain NEG-M	730 290990046 Heterolobos Eukaryota 2.A.19.7.1
Ngr2	predicted protein	Naegleria gruberi strain NEG-M	809 290975515 Heterolobos Eukaryota 2.A.19.7.1
Dre1	cation/H ⁺ exchanger protein 1	Danio rerio	764 110005804 Metazoa Eukaryota 2.A.19.7.1
Average		Average	734
Standard Derivation		Standard Derivation	79
Sub-Cluster3			
Mgl1	hypothetical protein MGL_3739	Malassezia globosa CBS 7966	1090 164656288 Fungi Eukaryota 2.A.19.7.1
Uma1	hypothetical protein UM04326.1	Ustilago maydis 521	1648 71020485 Fungi Eukaryota 2.A.19.7.1
Mla1	hypothetical protein MELLADRAFT_47989	Melampsora larici-populina 98AG31	806 328858470 Fungi Eukaryota 2.A.19.7.1
Pgr1	hypothetical protein PGTG_14516	Puccinia graminis f. sp. tritici CRL 75-36-700-3	1405 331239133 Fungi Eukaryota 2.A.19.7.1

Table 1, continued

Sla1	hypothetical protein SERLADRAFT_438928	Serpula lacrymans var. lacrymans 57.9	1124	336382463 Fungi	Eukaryota	2.A.19.7.1
Cci1	hypothetical protein CCIIG_02972	Coprinopsis cinerea okayama7#130	1059	299743625 Fungi	Eukaryota	2.A.19.7.1
Lbi1	predicted protein	Laccaria bicolor 5238N-H82	1128	170089165 Fungi	Eukaryota	2.A.19.7.1
Cne1	hypothetical protein CNBC1680	Cryptococcus neoformans var. neoformans B-3501A	1428	134109125 Fungi	Eukaryota	2.A.19.7.1
		Average	1211			
		Standard Derivation	265			
Sub-Cluster4						
Spo1	calcium permease (predicted)	Schizosaccharomyces pombe 972h-	881	191140111 Fungi	Eukaryota	2.A.19.7.1
Sub-Cluster5						
Yli1	YALI0024233p	Yarrowia lipolytica CLIB122	1145	50551499 Fungi	Eukaryota	2.A.19.7.1
Sub-Cluster6						
Vpo1	hypothetical protein Kpo1_1014p23	Vanderwaltozyma polyspora DSM 70294	829	156841377 Fungi	Eukaryota	2.A.19.7.1
Sub-Cluster7						
Sce1	vacuolar Na ⁺ /H ⁺ exchanger	Saccharomyces cerevisiae YJM789	908	151944229 Fungi	Eukaryota	2.A.19.7.1
Nca1	hypothetical protein NCAS_0E01330	Naumovozyma castellii CBS 4309	1027	342302430 Fungi	Eukaryota	2.A.19.7.1
Zro1	ZYROOC03256p	Zygosaccharomyces rouxii CBS 732	981	254579623 Fungi	Eukaryota	2.A.19.7.1
Lth1	KLTHOF19404p	Lachancea thermotolerans CBS G340	802	255717498 Fungi	Eukaryota	2.A.19.7.1
Kla1	hypothetical protein	Kluyveromyces lactis NRRL Y-1140	981	50309777 Fungi	Eukaryota	2.A.19.7.1
Ago1	ADL189Wp	Ashbya gossypii ATCC 10895	855	302307281 Fungi	Eukaryota	2.A.19.7.1
		Average	926			
		Standard Derivation	86			
Sub-Cluster8						
Opz1	Low affinity vacuolar membrane localized monovalent cation transporter	Ogataea parapolymorpha DL-1	1039	320582515 Fungi	Eukaryota	2.A.19.7.1
Kpa1	Low affinity vacuolar membrane localized monovalent cation transporter	Komagataella pastoris GS115	1081	254573950 Fungi	Eukaryota	2.A.19.7.1
Ctr1	hypothetical protein CTRG_05792	Candida tropicalis MYA-3404	1022	255722760 Fungi	Eukaryota	2.A.19.7.1
Le1	hypothetical protein LELG_05252	Lodderomyces elongisporus NRRL YB-4239	1103	149235055 Fungi	Eukaryota	2.A.19.7.1
Dha1	DEHA2A13904p	Debaryomyces hansenii CBS767	1082	294654849 Fungi	Eukaryota	2.A.19.7.1
Clu1	hypothetical protein CLUG_03472	Clavispora lusitanae ATCC 42720	1009	260943866 Fungi	Eukaryota	2.A.19.7.1
		Average	1056			
		Standard Derivation	38			
Cluster3						
Sub-Cluster1						
Aca1	Ca ²⁺ /H ⁺ antiporter	Azorhizobium caulinodans ORS 571	393	158424660 Alphaproteobacteria	Bacteria	2.A.19.1.1
Sub-Cluster2						
Nsp1	sodium/calcium exchanger membrane region	Novosphingobium sp. PP1Y	371	334142986 Alphaproteobacteria	Bacteria	2.A.19.1.1
Sub-Cluster3						
Tsp1	putative Ca ²⁺ /Na ⁺ antiporter	Thiomonas sp. 3As	372	294338940 Betaproteobacteria	Bacteria	2.A.19.1.1
Kfi1	sodium/calcium exchanger membrane region	Kribbella flavida DSM 17836	407	284034771 Actinobacteria	Bacteria	2.A.19.1.1
		Average	390			
		Standard Derivation	25			
Sub-Cluster4						
Pne1	unnamed protein product	Polynucleobacter necessarius subsp. symbioticus QLV	343	145588526 Betaproteobacteria	Bacteria	2.A.19.1.1
Rie1	sodium/calcium exchanger membrane region	Rhizobium leguminosarum bv. trifolii WSM1325	379	241258675 Alphaproteobacteria	Bacteria	2.A.19.1.1
		Average	361			
		Standard Derivation	25			
Sub-Cluster5						
Chu1	calcium/proton antiporter	Cytophaga hutchinsonii ATCC 33406	382	110638378 Bacteroidetes	Bacteria	2.A.19.1.1
Cgl1	calcium/proton antiporter membrane protein Ch	Chryseobacterium gleum ATCC 35910	365	300779118 Bacteroidetes	Bacteria	2.A.19.1.1

Table 1, continued

Sub-Cluster2	Mbr1	hypothetical protein sodium/calcium exchanger, putative	Average	829			
			Standard Derivation	22			
			Monosiga brevicollis MX1	852	167537060 Codonosigidi	Eukaryota	2.A.19.3.3
			Perkinsus marinus ATCC 50983	896	294953203 Perkinsidae	Eukaryota	2.A.19.3.3
Sub-Cluster3	Pma1	hypothetical protein sodium/calcium exchanger, putative	Average	874			
			Standard Derivation	31			
			Nematostella vectensis	837	156397050 Metazoa	Eukaryota	2.A.19.3.3
			PREDICTED: similar to solute carrier family 8 (sod Ciona intestinalis	907	198426482 Metazoa	Eukaryota	2.A.19.3.3
Sub-Cluster4	Cin1	hypothetical protein TRIADRAFT_78683	Branchiostoma floridae	871	260829777 Metazoa	Eukaryota	2.A.19.3.3
			Trichoplax adhaerens	945	196001263 Metazoa	Eukaryota	2.A.19.3.3
			Average	890			
			Standard Derivation	46			
Sub-Cluster5	Tad2	hypothetical protein TRIADRAFT_53617	Pediculus humanus corporis	840	242016779 Metazoa	Eukaryota	2.A.19.3.2
			Trichoplax adhaerens	805	195999284 Metazoa	Eukaryota	2.A.19.3.3
			Oikopleura dioica	831	313235239 Metazoa	Eukaryota	2.A.19.3.1
			Average	825			
Sub-Cluster6	Odi2	unnamed protein product	Standard Derivation	18			
			Hydra magnipapillata	914	221121648 Metazoa	Eukaryota	2.A.19.3.1
			Ascaris suum	860	324504975 Metazoa	Eukaryota	2.A.19.3.3
			Caenorhabditis briggsae AF16	922	309360090 Metazoa	Eukaryota	2.A.19.3.1
Sub-Cluster7	Tsp2	putative sodium/calcium exchanger 1	Trichinella spiralis	904	339235315 Metazoa	Eukaryota	2.A.19.3.3
			Loa loa	923	312066399 Metazoa	Eukaryota	2.A.19.3.3
			Hordeum vulgare subsp. vulgare	954	326500784 Viridiplantae	Eukaryota	2.A.19.3.1
			Average	913			
Sub-Cluster8	Hvu1	predicted protein	Standard Derivation	31			
			Trichoplax adhaerens	927	196014843 Metazoa	Eukaryota	2.A.19.3.3
			Danio rerio	961	181339524 Metazoa	Eukaryota	2.A.19.3.3
			solute carrier family 8 (sodium-calcium exchange	891	181339493 Metazoa	Eukaryota	2.A.19.3.3
Sub-Cluster9	Tni1	unnamed protein product	Tetradon nigroviridis	793	47213382 Metazoa	Eukaryota	2.A.19.3.1
			Equus caballus	991	149727965 Metazoa	Eukaryota	2.A.19.3.1
			Branchiostoma floridae	949	260836357 Metazoa	Eukaryota	2.A.19.3.1
			Oikopleura dioica	1089	313230968 Metazoa	Eukaryota	2.A.19.3.3
Sub-Cluster10	Dya1	unnamed protein product	Drosophila yakuba	950	195498363 Metazoa	Eukaryota	2.A.19.3.2
			Average	944			
			Standard Derivation	84			
Sub-Cluster11	Bfi1	hypothetical protein TRIADRAFT_32198	Trichoplax adhaerens	927	196014843 Metazoa	Eukaryota	2.A.19.3.3
			Danio rerio	961	181339524 Metazoa	Eukaryota	2.A.19.3.3
			solute carrier family 8 (sodium-calcium exchange	891	181339493 Metazoa	Eukaryota	2.A.19.3.3
			unnamed protein product	793	47213382 Metazoa	Eukaryota	2.A.19.3.1
Sub-Cluster12	Eca1	PREDICTED: sodium/calcium exchanger 1	Equus caballus	991	149727965 Metazoa	Eukaryota	2.A.19.3.1
			Branchiostoma floridae	949	260836357 Metazoa	Eukaryota	2.A.19.3.1
			Oikopleura dioica	1089	313230968 Metazoa	Eukaryota	2.A.19.3.3
			Average	950			
Sub-Cluster13	Bfi1	hypothetical protein TRIADRAFT_120267	Standard Derivation	84			
			Branchiostoma floridae	560	260797070 Metazoa	Eukaryota	2.A.19.4.6
			PREDICTED: solute carrier family 24 (sodium/pot	451	291243690 Metazoa	Eukaryota	2.A.19.4.6
			Saccoglossus kowalevskii	472	291221728 Metazoa	Eukaryota	2.A.19.4.6
Sub-Cluster14	Sko1	PREDICTED: sodium/potassium/calcium exchang	Average	494			
			Standard Derivation	58			
			Hydra magnipapillata	548	221122333 Metazoa	Eukaryota	2.A.19.4.6
Sub-Cluster15	Hma1	PREDICTED: similar to predicted protein					
Sub-Cluster16	Api1	PREDICTED: sodium/potassium/calcium exchang					
Sub-Cluster17	Aae1	potassium-dependent sodium-calcium exchange					

Table 1, continued

Aae2	potassium-dependent sodium-calcium exchange	Aedes aegypti	520	157107608	Metazoa	Eukaryota	2.A.19.4.6
Dwi2	GK25593	Drosophila willistoni	722	195447092	Metazoa	Eukaryota	2.A.19.4.6
Dsi1	GD15468	Drosophila simulans	744	195555698	Metazoa	Eukaryota	2.A.19.4.6
Tca2	hypothetical protein TcasGA2_TC013320	Tribolium castaneum	572	270006895	Metazoa	Eukaryota	2.A.19.4.6
Phu1	sodium/potassium/calcium exchanger 4 precursor	Pedicular humanus corporis	601	242020994	Metazoa	Eukaryota	2.A.19.4.5
Api2	PREDICTED: probable sodium/potassium/calcium Acyrthosiphon pisum	Acyrthosiphon pisum	545	328701889	Metazoa	Eukaryota	2.A.19.4.6
Api3	PREDICTED: sodium/potassium/calcium exchange	Acyrthosiphon pisum	511	328706570	Metazoa	Eukaryota	2.A.19.4.6
Cfi2	Sodium/potassium/calcium exchanger 5	Camponotus floridanus	546	307179269	Metazoa	Eukaryota	2.A.19.4.6
Dp52	GA27524	Drosophila pseudoobscura	465	198467881	Metazoa	Eukaryota	2.A.19.4.6
Agai1	AGAP010975-PA	Anopheles gambiae str. PEST	563	158287819	Metazoa	Eukaryota	2.A.19.4.6
Ame1	PREDICTED: sodium/potassium/calcium exchanger	Apis mellifera	450	328793656	Metazoa	Eukaryota	2.A.19.4.5
Cfi1	Sodium/potassium/calcium exchanger 5	Camponotus floridanus	468	307179272	Metazoa	Eukaryota	2.A.19.4.6
Der1	GG19760	Drosophila erecta	513	194897957	Metazoa	Eukaryota	2.A.19.4.5
Tca3	hypothetical protein TcasGA2_TC014488	Tribolium castaneum	582	270007787	Metazoa	Eukaryota	2.A.19.4.6
Dpu1	hypothetical protein DAPPUDRAFT_207878	Daphnia pulex	458	321476184	Metazoa	Eukaryota	2.A.19.4.6
	Average		546				
	Standard Derivation		84				
Sub-Cluster4							
Aca1	PREDICTED: sodium/potassium/calcium exchange	Anolis carolinensis	610	327259250	Metazoa	Eukaryota	2.A.19.4.5
Mmu3	PREDICTED: sodium/potassium/calcium exchange	Macaca mulatta	992	297260387	Metazoa	Eukaryota	2.A.19.4.5
Spu1	PREDICTED: similar to potassium-dependent Na/	Strongylocentrotus purpuratus	570	115871986	Metazoa	Eukaryota	2.A.19.4.5
Tad1	hypothetical protein TRIADDRAFT_33154	Trichoplax adhaerens	512	196016029	Metazoa	Eukaryota	2.A.19.4.5
Bf2	hypothetical protein BRAFLDRAFT_209452	Branchiostoma floridae	563	260822461	Metazoa	Eukaryota	2.A.19.4.5
	Average		694				
	Standard Derivation		195				
Sub-Cluster5							
Sma1	hypothetical protein	Schistosoma mansoni	599	256087501	Metazoa	Eukaryota	2.A.19.4.5
Ada1	hypothetical protein AND_11865	Anopheles darlingi	743	312376962	Metazoa	Eukaryota	2.A.19.4.5
Dp51	GA26275	Drosophila pseudoobscura	651	198454486	Metazoa	Eukaryota	2.A.19.4.6
Bte1	PREDICTED: probable sodium/potassium/calcium	Bombus terrestris	666	340717092	Metazoa	Eukaryota	2.A.19.4.6
Dpu2	hypothetical protein DAPPUDRAFT_309468	Daphnia pulex	542	321459376	Metazoa	Eukaryota	2.A.19.4.6
Aca2	PREDICTED: sodium/potassium/calcium exchange	Anolis carolinensis	588	327286665	Metazoa	Eukaryota	2.A.19.4.6
Bf3	hypothetical protein BRAFLDRAFT_82602	Branchiostoma floridae	528	260814626	Metazoa	Eukaryota	2.A.19.4.6
	Average		617				
	Standard Derivation		75				
Sub-Cluster6							
Nve2	predicted protein	Nematostella vectensis	639	156403746	Metazoa	Eukaryota	2.A.19.4.2
Hma2	PREDICTED: similar to predicted protein	Hydra magnipapillata	492	221101518	Metazoa	Eukaryota	2.A.19.4.6
	Average		566				
	Standard Derivation		104				
Sub-Cluster7							
Msa1	NCKX1	Morone saxatilis	749	118139446	Metazoa	Eukaryota	2.A.19.4.2
Tni2	unnamed protein product	Tetraodon nigroviridis	1043	47214497	Metazoa	Eukaryota	2.A.19.4.2
Tni1	unnamed protein product	Tetraodon nigroviridis	647	47255868	Metazoa	Eukaryota	2.A.19.4.2
Msa2	NCKX2.2	Morone saxatilis	638	118139452	Metazoa	Eukaryota	2.A.19.4.2
Mmu1	solute carrier family 24 (sodium/potassium/calcium)	Mus musculus	711	159032066	Metazoa	Eukaryota	2.A.19.4.2
Hsa1	sodium/potassium/calcium exchanger 1 isoform	Homo sapiens	1099	4759128	Metazoa	Eukaryota	2.A.19.4.1
Bta1	RetName: Full-Sodium/potassium/calcium exchanger	Bos taurus	1216	17865507	Metazoa	Eukaryota	2.A.19.4.1
Ttr1	retinal rod Na/Ca-K exchanger	Tursiops truncatus	1014	3170603	Metazoa	Eukaryota	2.A.19.4.1
Mmu2	sodium/potassium/calcium exchanger 1	Mus musculus	1130	21450183	Metazoa	Eukaryota	2.A.19.4.1
Mdo1	PREDICTED: sodium/potassium/calcium exchange	Monodelphis domestica	1194	334314460	Metazoa	Eukaryota	2.A.19.4.1
Spu2	PREDICTED: similar to cone sodium-calcium pota	Strongylocentrotus purpuratus	651	115709825	Metazoa	Eukaryota	2.A.19.4.2
Nve1	predicted protein	Nematostella vectensis	643	156385129	Metazoa	Eukaryota	2.A.19.4.2

Table 1, continued

Mpu1	Ca2+-Cation antiporter family	Micromonas pusilla CCMP1545	510	303279510	Viridiplantae	Eukaryota	2.A.19.4.2
Msp1	Ca2+-Cation antiporter family	Micromonas sp. RCC299	538	255082267	Viridiplantae	Eukaryota	2.A.19.4.2
Dwl1	GK24823	Drosophila willistoni	893	195438385	Metazoa	Eukaryota	2.A.19.4.2
Dme1	RecName: Full= Sodium/potassium/calcium exch; Drosophila melanogaster		888	126302573	Metazoa	Eukaryota	2.A.19.4.2
Tca1	hypothetical protein TcasGA2_TC010208	Tribolium castaneum	671	270010753	Metazoa	Eukaryota	2.A.19.4.2
Cre1	CRE-NCX-4 protein	Caenorhabditis remanei	609	308494414	Metazoa	Eukaryota	2.A.19.4.2
Bma1	Na/Ca exchangers family member (ncx-4)	Brugia malayi	650	170592429	Metazoa	Eukaryota	2.A.19.4.9
Asu1	Sodium/potassium/calcium exchanger 2	Ascaris suum	638	324509213	Metazoa	Eukaryota	2.A.19.4.2
Bma2	K+-dependent Na+/Ca+ exchanger related-protein	Brugia malayi	591	170593899	Metazoa	Eukaryota	2.A.19.4.5
Cbr1	CBR-NCX-5 protein	Caenorhabditis briggsae AF16	660	309361787	Metazoa	Eukaryota	2.A.19.4.2
	Average		790				
	Standard Deviation		226				
Sub-Cluster8							
Aan1	hypothetical protein AURANDRAFT_36185	Aureococcus anophagefferens	623	323456360	Pelagophyceae	Eukaryota	2.A.19.4.2
Tps2	sodium/potassium/calcium exchange protein	Thalassiosira pseudonana CCMP1335	627	224015509	Bacillariophyta	Eukaryota	2.A.19.4.5
	Average		625				
	Standard Deviation		3				
Sub-Cluster9							
Cin1	PREDICTED: similar to predicted protein	Ciona intestinalis	857	198422506	Metazoa	Eukaryota	2.A.19.4.3
Aam1	Na+/Ca2+ exchanger	Asterias amurensis	616	51491497	Metazoa	Eukaryota	2.A.19.4.3
Spu3	testis potassium dependent sodium/calcium	Strongylocentrotus purpuratus	652	47551219	Metazoa	Eukaryota	2.A.19.4.3
Est1	conserved unknown protein	Ectocarpus siliculosus	604	298713622	Phaeophyceae	Eukaryota	2.A.19.4.5
Pin1	Ca2 :Cation Antiporter (CaCA) Family	Phytophthora infestans T30-4	550	301095294	none	Eukaryota	2.A.19.4.5
Ota1	Solute carrier family 24 member 4, isoform 1 pre	Ostreococcus tauri	545	308809513	Viridiplantae	Eukaryota	2.A.19.4.5
Mpu2	Ca2+-Cation antiporter family	Micromonas pusilla CCMP1545	611	303285051	Viridiplantae	Eukaryota	2.A.19.4.2
Pma1	conserved hypothetical protein	Perkinsus marinus ATCC 50983	568	29494237	Perkinsidae	Eukaryota	2.A.19.4.3
Mma1	putative K+-dependent Na+/Ca+ exchanger-like	Microscilla marina ATCC 23134	389	124003363	Bacteroidetes	Bacteria	2.A.19.4.6
Ppa1	Ca2+/Na+ antiporter	Plesiocystis pacifica SIR-1	361	149923824	Deltaproteobacteria	Bacteria	2.A.19.4.3
Tps1	predicted protein	Thalassiosira pseudonana CCMP1335	691	224010691	Bacillariophyta	Eukaryota	2.A.19.4.5
Tps3	predicted protein	Thalassiosira pseudonana CCMP1335	632	224001622	Bacillariophyta	Eukaryota	2.A.19.4.1
	Average		590				
	Standard Deviation		130				
Cluster6							
Sub-Cluster1							
Dau1	kncA gene product	Desulfobacterium autotrophicum HRM2	310	224369605	Deltaproteobacteria	Bacteria	2.A.19.5.1
Din1	Na+/Ca+ antiporter, CaCA family	Desulfurispirillum indicum S5	313	317051802	Chrysiogenetia	Bacteria	2.A.19.5.2
Ton1	Ca2+/Na+ antiporter	Thermococcus onnurineus NA1	311	212224113	Euryarchaeota	Archaea	2.A.19.5.2
Dac1	K+-dependent Na+/Ca+ exchanger related-protein	Desulfuromonas acetoxidans DSM 684	314	95931468	Deltaproteobacteria	Bacteria	2.A.19.5.1
	Average		312				
	Standard Deviation		2				
Sub-Cluster2							
Hba1	CaCA family Na+/Ca+ antiporter	Hirschia baltica ATCC 49814	318	254295181	Alphaproteobacteria	Bacteria	2.A.19.5.2
Hph1	K+-dependent Na+/Ca+ exchanger related-protein	Hoeftia phototrophica DFL-43	312	163759156	Alphaproteobacteria	Bacteria	2.A.19.5.2
Orf9	K+-dependent Na+/Ca+ exchanger related-protein	alpha proteobacterium BAL199	328	163794444	Alphaproteobacteria	Bacteria	2.A.19.5.2
Opr1	Na+/Ca+ antiporter, caca family	Oceanithermus profundus DSM 14977	312	313679390	Deinococcus	Bacteria	2.A.19.5.2
Dci1	CaCA family Na(+)/Ca(+) antiporter	Dietzia cinnamomea P4	397	319951035	Actinobacteria	Bacteria	2.A.19.5.2
	Average		333				
	Standard Deviation		36				
Sub-Cluster3							
Ssp1	putative CaCA family, sodium/calcium exchanger	Synechococcus sp. WH 5701	360	87302538	Cyanobacteria	Bacteria	2.A.19.5.1
Htu1	CaCA family Na+/Ca+ antiporter	Haloterrigena turkmenica DSM 5511	324	284165688	Euryarchaeota	Archaea	2.A.19.5.1
Orf7	putative Sodium/calcium exchanger protein	uncultured marine crenarchaeote HF4000	310	167042586	Thaumarchaeota	Archaea	2.A.19.5.2
	Average		331				

Table 1, continued

Sub-Cluster4		Standard Derivation	26
Orf1	unnamed protein product	gamma proteobacterium HdN1	359
Ssm1	CaCA family Na ⁺ /Ca ⁺ antiporter	Spirochaeta smaragdinae DSM 11293	379
Vmi1	YrbG protein	Vibrio mimicus MB451	360
Ssp2	hypothetical protein sir0681	Synechocystis sp. PCC 6803	433
Cag1	CaCA family Na ⁺ /Ca ⁺ antiporter	Chloroflexus aggregans DSM 9485	369
Avi1	K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger related-protein	Azotobacter vinelandii DJ	368
Kst1	conserved hypothetical protein	Candidatus Kuenenia stuttgartiensis	321
Kst2	conserved hypothetical protein	Candidatus Kuenenia stuttgartiensis	318
Pjo1	hypothetical protein PRABACT1_OH_N_01295	Parabacteroides johnsonii DSM 18315	334
Vva1	Na ⁺ /Ca ⁺ antiporter, CaCA family	Victivallis vadensis ATCC BAA-548	318
Average		Average	356
Standard Derivation		Standard Derivation	36
Sub-Cluster5			
Ybe1	Sodium/calcium exchanger protein	Yersinia bercovieri ATCC 43970	354
Pmi1	CaCA family calcium (Ca ²⁺)-cation antiporter	Proteus mirabilis ATCC 29906	349
Vha1	conserved hypothetical protein	Vibrio Harveyi 1DA3	349
Pha1	Na/Ca transporter (CaCA)	Pseudomonas aeruginosa ATCC 27852	355
Average		Average	352
Standard Derivation		Standard Derivation	3
Sub-Cluster6			
Orf2	putative Ca ²⁺ /Na ⁺ antiporter	gamma proteobacterium HTCC2207	330
Orf8	putative K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger	gamma proteobacterium NOR51-B	322
Orf3	Inner membrane protein	gamma proteobacterium IMCC3088	325
Ttu1	K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger family protein	Teredinibacter turnerae T7901	331
Orf6	Ca ²⁺ /Na ⁺ antiporter	marine gamma proteobacterium HTCC2143	331
Orf4	Inner membrane protein YrbG2C predicted calc	gamma proteobacterium IMCC2047	321
Rbl1	K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger related-protein	Reinekea blandensis MED297	321
Ppa1	hypothetical protein PPSIR1_18417	Plesiocystis pacifica SIR-1	320
Mth1	putative K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger	Methylophaga thiooxydans DM5010	328
Cbu1	sodium-calcium exchanger	Coxiella burnetii RSA 493	322
Average		Average	325
Standard Derivation		Standard Derivation	4
Sub-Cluster7			
Dgi1	Ca ²⁺ /Na ⁺ antiporter	Desulfovibrio gigas	340
Ami1	putative sodium/calcium exchange protein	Actinobacillus minor NM305	316
Pcr1	K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger related-protein	Psychrobacter cryohalophilus K5	329
Rma1	CaCA family Na ⁺ /Ca ⁺ antiporter	Candidatus Ruthia magnifica str. Cm	317
Pin1	Na ⁺ /Ca ⁺ antiporter, CaCA family protein	Psychromonas ingrahamii 37	326
Tcr1	K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger related-protein	Thiomicrospira crunogena XCL-2	329
Orf5	putative K ⁺ -dependent Na ⁺ /Ca ⁺ exchanger	gamma proteobacterium HTCC5015	325
Kko1	Na ⁺ /Ca ⁺ antiporter	Kangia koreensis DSM 16069	327
Average		Average	326
Standard Derivation		Standard Derivation	8
Sub-Cluster8			
Aeh1	CaCA family Na ⁺ /Ca ⁺ antiporter	Alkallimnicola ehrlichii MLHE-1	327
Sub-Cluster9			
Lpn1	Ca ²⁺ /Na ⁺ antiporter	Legionella pneumophila 2300/99 Alcoy	319
Average		Average	334
Standard Derivation		Standard Derivation	25

Table 2. All MPE family proteins included in this study are listed according to the position of the protein in the phylogenetic tree (clockwise direction) (Figure 8). The average protein size $\pm$ S.D. is provided for each clusters, sub-cluster and sub-sub-cluster. The Table presents the protein abbreviation, sequence description, organism, protein size, gi number, phylum or kingdom and organismal domain.

Abbreviation	SequenceDescription	Organism	ProteinSize	GenBankIndk Group	Kingdom
Cluster 1					
Sub-cluster1					
Sub-Sub-cluster 1a					
Sfr1	cell division protein FtsW	Sinorhizobium fredii NGR234	384	227822652 Alphaproteobacteria	Bacteria
Cr1	cell division protein DivB	Caulobacter crescentus CB15	390	16126791 Alphaproteobacteria	Bacteria
Ssp1	cell division protein FtsW	Sulfitobacter sp. EE-36	389	83942744 Alphaproteobacteria	Bacteria
Gob1	cell division protein FtsW	Glucanacetobacter oboediens 1749p2	388	349685776 Alphaproteobacteria	Bacteria
Rca1	cell division protein FtsW	Rickettsia canadensis str. McKiel	379	157803834 Alphaproteobacteria	Bacteria
Esp1	cell division protein	Erythrobacter sp. SD-21	422	149186203 Alphaproteobacteria	Bacteria
Wen1	cell division protein FtsW, putative	Wolbachia endosymbiont of Muscidifurax uniraptor	371	225677017 Alphaproteobacteria	Bacteria
Pub1	cell division protein FtsW	Candidatus Pelagibacter ubique HTCC1002	374	91762844 Alphaproteobacteria	Bacteria
		Average Size	387		
		Standard Derivation	16		
Sub-Sub-cluster1b					
Tna1	cell cycle protein	Thermodesulfobium narugense DSM 14796	367	332295507 Firmicutes	Bacteria
Dth1	cell division protein FtsW	Dictyoglomus thermophilum H-6-12	361	206901982 Dictyoglomi	Bacteria
Lin1	cell division protein	Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	384	456577728 Spirochaetes	Bacteria
		Average Size	371		
		Standard Derivation	12		
Sub-Sub-cluster1c					
Tme1	cell cycle protein	Thermosipho melanesiensis B1429	362	150020835 Thermotogae	Bacteria
Fno1	cell cycle protein	Fervidobacterium nodosum RT17-B1	371	154249902 Thermotogae	Bacteria
Tsp1	cell cycle protein	Thermotoga sp. RQ2	364	170288513 Thermotogae	Bacteria
Thh1	cell cycle protein	Thermotoga thermarum DSM 5069	361	338730477 Thermotogae	Bacteria
Kol1	cell cycle protein	Kosmotoga olearia TBF 19.5.1	365	239617754 Thermotogae	Bacteria
Pmo1	cell cycle protein	Petrogaga mobilis SJ95	366	160903089 Thermotogae	Bacteria
Lbu1	cell cycle protein	Leptotrichia buccalis C-1013-b	388	257125001 Fusobacteria	Bacteria
Smo1	cell cycle protein	Streptobacillus moniliformis DSM 12112	372	269122826 Fusobacteria	Bacteria
Cac1	Essential cell division protein FtsW	Candidatus Cloacamonas acidaminovorans	379	218961096 none	Bacteria
		Average Size	370		
		Standard Derivation	9		
Sub-Cluster 2					
Fba1	transmembrane cell division protein	Flavobacteria bacterium BBEL7	401	89891768 Bacteroidetes	Bacteria
Orf3	cell division protein	unidentified eubacterium SCB49	397	149371576 Bacteroidetes	Bacteria
Cgl1	cell division protein	Chryseobacterium gleum ATCC 35910	423	300774445 Bacteroidetes	Bacteria
Chu1	cell division protein	Cytophaga hutchinsonii ATCC 33406	385	110639123 Bacteroidetes	Bacteria
Mma1	rod shape-determining protein RodA	Microscilla marina ATCC 23134	409	124009922 Bacteroidetes	Bacteria
Lby1	cell cycle protein	Leadbetterella byssophila DSM 17132	388	312130744 Bacteroidetes	Bacteria
Osp1	cell cycle protein	Odoribacter splanchnicus DSM 20712	493	325280026 Bacteroidetes	Bacteria
Bor1	rod shape-determining protein RodA	Bacteroidetes oral taxon 274 str. F0058	456	298373587 Bacteroidetes	Bacteria
Fta1	cell cycle protein	Fluvicola taffensis DSM 16823	391	327404200 Bacteroidetes	Bacteria
Apu1	hypothetical protein ALIPUT_01127	Alistipes putredinis DSM 17216	436	167752865 Bacteroidetes	Bacteria
Pgi1	cell division protein FtsW	Porphyromonas gingivalis W83	418	34540388 Bacteroidetes	Bacteria
Pue1	rod shape-determining protein RodA	Porphyromonas uenonis 60-3	465	228469416 Bacteroidetes	Bacteria
Pen1	putative cell division protein FtsW	Porphyromonas endodontalis ATCC 35406	472	229496542 Bacteroidetes	Bacteria
Bmu1	unnamed protein product	Brachyspira murdochii DSM 12563	364	296125461 Spirochaetes	Bacteria
		Average Size	421		
		Standard Derivation	38		
Sub-Cluster3					
Tvi1	cell division protein FtsW	Treponema vincentii ATCC 35580	378	257458296 Spirochaetes	Bacteria
Tpa1	cell division protein (ftsW)	Treponema pallidum subsp. pallidum str. Nichols	384	156393778 Spirochaetes	Bacteria

Table 2, continued

Tde1	cell cycle protein FtsW	Treponema denticola ATCC 35405	377	42526710	Spirochaetes	Bacteria
Bbu1	membrane protein	Borrelia burgdorferi	364	1234873	Spirochaetes	Bacteria
Sgl1	unnamed protein product	Sphaerochaeta globus str. Buddy	411	325971092	Spirochaetes	Bacteria
Cpn1	cell division protein FtsW	Chlamydomonas pneumoniae CWL029	385	15618812	Chlamydiae	Bacteria
Vba1	cell cycle protein, FtsW/RodA/SpoVE family	Verrucomicrobiae bacterium DG1235	387	254446865	Verrucomicrobia	Bacteria
Min1	Cell division protein FtsW	Methylacidiphilum infernorum V4	395	189219422	Verrucomicrobia	Bacteria
Sac1	cell division protein	Syntrophus aciditrophicus SB	390	85858525	Deltaproteobacteria	Bacteria
Dau2	ftsW gene product	Desulfobacterium autotrophicum HRM2	374	224368383	Deltaproteobacteria	Bacteria
Dol1	cell division protein FtsW	Desulfococcus oleovorans Hxd3	371	158522798	Deltaproteobacteria	Bacteria
Dvu1	cell division protein FtsW	Desulfovibrio vulgaris str. 'Miyazaki F'apost;	394	21886059	Deltaproteobacteria	Bacteria
Dvu2	cell cycle protein FtsW	Desulfovibrio vulgaris str. Hildenborough	377	46580909	Deltaproteobacteria	Bacteria
Oba1	cell division protein FtsW	uncultured candidate division OP1 bacterium	383	374856988	none	Bacteria
Kst1	similar to cell division protein FtsW	Candidatus Kuenenia stuttgartiensis	399	91200084	Planctomycetes	Bacteria
Amu1	cell cycle protein	Akkermansia muciniphila ATCC BAA-835	380	187735164	Verrucomicrobia	Bacteria
Vsp1	bacterial cell division membrane protein	Verrucomicrobium spinosum DSM 4136	389	171913319	Verrucomicrobia	Bacteria
Ppa1	cell cycle protein	Pedospira parvula Ellin514	378	223937421	Verrucomicrobia	Bacteria
Lar1	stage V sporulation protein E	Lentisphaera araneosa HTCC2155	402	149196827	Lentisphaerae	Bacteria
Vva1	cell cycle protein	Victivallis vadensis ATCC BAA-548	409	281357083	Lentisphaerae	Bacteria
Emi1	cell cycle protein	Elusimicrobium minutum Pei191	383	187251656	Elusimicrobia	Bacteria
Par1	ftsW gene product	Psychrobacter arcticus 273-4	398	71066593	Gammaaproteobacteria	Bacteria
Dac1	cell division protein FtsW	Denitrovibrio acetiphilus DSM 12809	365	291288104	Deferribacteres	Bacteria
Cae1	Hypothetical protein COLAER_00337	Collinsella aerofaciens ATCC 25986	460	139437198	Actinobacteria	Bacteria
Gel1	cell division protein FtsW/RodA/SpoVE family	Granulicatella elegans ATCC 700633	411	260584657	Firmicutes	Bacteria
Mfe1	Cell cycle protein	Mariprofundus ferrooxydans PV-1	416	114777863	Zetaproteobacteria	Bacteria
Afe1	cell division protein FtsW	Acidithiobacillus ferrooxidans ATCC 53993	390	198282526	Gammaaproteobacteria	Bacteria
Msp1	unnamed protein product	Magnetococcus sp. MC-1	375	117924064	Proteobacteria	Bacteria
Bma1	ftsW gene product	Bacteriovorax marinus SJ	394	374286893	Deltaproteobacteria	Bacteria
Bba1	cell division protein FtsW	Bdellovibrio bacteriovorus HD100	380	42524580	Deltaproteobacteria	Bacteria
Orf1	bacterial cell division membrane protein	uncultured myxobacterium HF0010_08807	373	297170217	Deltaproteobacteria	Bacteria
Rpe1	cell division protein FtsW	Candidatus Riesia pediculicola USDA	369	295698677	Gammaaproteobacteria	Bacteria
Bap1	cell division protein FtsW	Buchnera aphidicola	379	167614197	Gammaaproteobacteria	Bacteria
Xfa1	Cell cycle protein	Xylella fastidiosa Dixon	423	71275123	Gammaaproteobacteria	Bacteria
Rma1	cell cycle protein	Candidatus Ruthia magnifica str. Cm (Calyptogenia magnifica)	377	118602512	Gammaaproteobacteria	Bacteria
Kor1	hypothetical protein GCWU000324_01932	Kingella oralis ATCC 51147	409	238022027	Betaproteobacteria	Bacteria
Orf4	hypothetical protein FTT0452	synthetic construct	436	57339628	none	Unclassified
Hfo1	cell cycle protein	Holophaga foetida DSM 6591	370	373487374	Acidobacteria	Bacteria
Tsa1	cell division protein FtsW	Terriglobus saanensis SP1PR4	363	320106153	Acidobacteria	Bacteria
Sus1	unnamed protein product	Candidatus Solibacter usitatus Ellin6076	383	116626352	Acidobacteria	Bacteria
Din2	cell cycle protein	Desulfurispirillum indicum S5	379	317050623	Chrysiogenetes	Bacteria
Pma2	cell division protein FtsW	uncultured Prochlorococcus marinus clone HF10-11D6	411	91070195	Cyanobacteria	Bacteria
Sel1	unnamed protein product	Synechococcus elongatus PCC 7942	394	81299135	Cyanobacteria	Bacteria
		Average Size	390			
		Standard Derivation	20			
Sub-Cluster 4						
Edo1	hypothetical protein EUBDOL_01810	Eubacterium dolichum DSM 3991	350	160915795	Firmicutes	Bacteria
Bmy1	SpoVE/stage V sporulation protein E	Bacillus mycoides	363	323099952	Firmicutes	Bacteria
Bsp2	stage V sporulation protein E	Bacillus sp. SG-1	366	149183845	Firmicutes	Bacteria
Bsp1	Cell division protein FtsW	Bacillus sp. B14905	359	126649727	Firmicutes	Bacteria
Dau1	stage V sporulation protein E	Candidatus Desulforhabdus audaxviator MP104C	367	169831594	Firmicutes	Bacteria
Cth1	stage V sporulation protein E	Clostridium thermocellum ATCC 27405	383	125973490	Firmicutes	Bacteria
Hpr1	cell division-specific peptidoglycan biosynthesis	Halanaerobium praevalens DSM 2228	364	309388992	Firmicutes	Bacteria
Chy1	cell division protein FtsW	Carboxydotherrmus hydrogeniformans Z-2901	375	78044656	Firmicutes	Bacteria

Table 2, continued

Dha1	hypothetical protein DSY2254	Desulfotobacterium hafniense Y51	395	89895000	Firmicutes	Bacteria
Bd1	stage V sporulation protein E	Bacillus clausii KSM-K16	380	56964564	Firmicutes	Bacteria
Hsp1	cell cycle protein	Hydrogenobaculum sp. Y04AAS1	374	195952592	Aquificae	Bacteria
Pma1	cell cycle protein	Persephonella marina EX-H1	372	225850178	Aquificae	Bacteria
Aac1	unnamed protein product	Alcycobacillus acidocaldarius subsp. acidocaldarius DSM 446	467	258511024	Firmicutes	Bacteria
Lfe1	putative cell division protein (FtsW)	Leptospirillum ferroplazotrophum	386	251772198	Nitrospirae	Bacteria
Asp1	rod shape-determining protein rodA	Acidaminococcus sp. D21	401	227498870	Firmicutes	Bacteria
Tca1	cell cycle protein	Thermosinus carboxydvorans Nor1	386	121533643	Firmicutes	Bacteria
Din1	stage V sporulation protein E	Dialister invisus DSM 15470	411	260424677	Firmicutes	Bacteria
Msp2	putative stage V sporulation protein E	Megasphaera sp. UPII 135-E	422	342218989	Firmicutes	Bacteria
Mge1	putative cell division protein FtsW	Megasphaera genomosp. type_1 str. 28L	413	290967993	Firmicutes	Bacteria
		Average Size	386			
		Standard Deviation	28			
Sub-Cluster 5						
Hku1	hypothetical protein HMPREF9709_00919	Helcococcus kunzii ATCC 51366	405	374562656	Firmicutes	Bacteria
Fma1	cell division protein	Finnegoldia magna ATCC 29328	369	169824313	Firmicutes	Bacteria
Pmi1	hypothetical protein PEPMIC_00287	Parvimonas micra ATCC 33270	367	160946327	Firmicutes	Bacteria
Psp1	cell cycle protein, FtsW/RodA/SpoVE family	Peptoniphilus sp. oral taxon 375 str. F0436	361	342216832	Firmicutes	Bacteria
Psp2	cell division protein FtsW	Peptoniphilus sp. oral taxon 386 str. F0131	369	299143956	Firmicutes	Bacteria
Ceu2	hypothetical protein COPEUT_01122	Coprococcus eutactus ATCC 27759	396	163814972	Firmicutes	Bacteria
Ceu1	hypothetical protein COPEUT_02898	Coprococcus eutactus ATCC 27759	387	163816708	Firmicutes	Bacteria
Orf2	Bacterial cell division membrane protein	butyrate-producing bacterium SM4/1	380	295116221	Firmicutes	Bacteria
Osp2	stage V sporulation protein E	Oribacterium sp. oral taxon 078 str. F0262	380	291459275	Firmicutes	Bacteria
Lor1	hypothetical protein HMPREF0491_00881	Lachnospiraceae oral taxon 107 str. F0167	460	331002501	Firmicutes	Bacteria
Esp2	stage V sporulation protein E	Eupoliscium sp. 'N.t. morphotype B'apos;	383	168333727	Firmicutes	Bacteria
Fa1	unnamed protein product	Filifactor alois ATCC 35896	378	374307642	Firmicutes	Bacteria
Sva1	cell division protein, RodA/FtsW/SpoVE family	Subdoligranulum variable DSM 15176	430	261367513	Firmicutes	Bacteria
Pca1	hypothetical protein BACCAP_04425	Pseudoflavonifractor capillosus ATCC 29799	402	154500747	Firmicutes	Bacteria
Rsp1	Bacterial cell division membrane protein	Eubacteriaceae bacterium ACC19a	384	363894041	Firmicutes	Bacteria
Esa1	cell cycle protein, FtsW/RodA/SpoVE family	Eubacterium sapenum ATCC 49989	378	255994321	Firmicutes	Bacteria
Cst1	integral membrane protein involved in stabiliz	Clostridium sticklandii DSM 519	368	310659200	Firmicutes	Bacteria
Eba1	hypothetical protein HMPREF9629_01459	Ruminococcus sp. 18P13	389	291544490	Firmicutes	Bacteria
Pal1	stage V sporulation protein E	Pseudoramibacter alactolyticus ATCC 23263	376	315924804	Firmicutes	Bacteria
Ast1	hypothetical protein ANASTE_01025	Anaerofustis stercorihominis DSM 17244	367	169334622	Firmicutes	Bacteria
Cge1	cell cycle protein, FtsW/RodA/SpoVE family	Clostridiales genomosp. BVA83 str. UPI9-5	422	289450815	Firmicutes	Bacteria
Hma1	cell cycle protein	Hippea maritima DSM 10411	365	327399666	Deltaproteobacteria	Bacteria
Tgr1	putative cell division protein FtsW	uncultured Termitite group 1 bacterium phylotype Rs-D17	369	189485384	Elusimicrobia	Bacteria
		Average Size	386			
		Standard Deviation	24			
Sub-Cluster 6						
Tsi1	bacterial cell division membrane protein	candidate division TM7 single-cell isolate TM7a	432	169835606	none	Bacteria
Cab1	cell division protein FtsW	Caldithrix abyssi DSM 13497	386	373458540	none	Bacteria
Cte1	SpSE	Chlorobaculum tepidum	400	13899160	Chlorobi	Bacteria
Gau1	cell division protein FtsW	Gemmatimonas aurantiaca T-27	398	226226833	Gemmatimonadetes	Bacteria
Fsu1	cell division protein FtsW	Fibrobacter succinogenes subsp. succinogenes S85	390	261414978	Fibrobacteres	Bacteria
Lba1	hypothetical protein HMPREF9623_00666	Lachnospiraceae bacterium ACC2	433	373106702	Firmicutes	Bacteria
Sru1	cell division protein FtsW	Salinibacter ruber DSM 13855	379	83814197	Bacteroidetes	Bacteria
Tam1	cell cycle protein	Thermovibrio ammonificans HB-1	385	319789934	Aquificae	Bacteria
Ath1	cell division protein FtsW	Anaerolinea thermophila UNI-1	415	320161738	Chloroflexi	Bacteria
Nba1	essential cell division protein (stabilizes FtsZ n	NC10 bacterium 'Dutch sediment'	371	258593037	none	Bacteria
		Average Size	399			
		Standard Deviation	21			

Table 2, continued

Cluster 2	Sub-Cluster 1			367	4096797	Firmicutes	Bacteria
	Sca1	orf2; unknown function; similar to SpoVE, Rod Staphylococcus carnosus					
Sub-Cluster2	Gha1	putative cell division protein FtsW	Gemella haemolysans ATCC 10379	405	24189569	Firmicutes	Bacteria
	Cmo1	hypothetical protein GCWU000282_01305	Catonella morbi ATCC 51271	423	229824001	Firmicutes	Bacteria
	Fla1	hypothetical protein HMPREF9708_01122	Facklamia languida CCUG 37842	379	374565630	Firmicutes	Bacteria
	Eco1	putative stage V sporulation protein E	Eremococcus coeicola ACS-139-V-Col8	420	313884697	Firmicutes	Bacteria
	Gka1	cell-division protein	Geobacillus kaustophilus HTA426	403	56419613	Firmicutes	Bacteria
	Bce1	cell division protein FtsW	Bacillus cereus G9241	367	47565823	Firmicutes	Bacteria
	Lmo1	hypothetical protein lmo2688	Listeria monocytogenes EGD-e	376	16804725	Firmicutes	Bacteria
	Hco1	stage V sporulation protein E	Haloplasma contractile SSD-178	415	335429722	Haloplasmataceae	Bacteria
	Eh1	FtsW/RodA/SpoVE family cell division protein	Erysipelothrix rhusiopathiae ATCC 19414	418	323341738	Firmicutes	Bacteria
	Edo1	hypothetical protein EUBDOL_00031	Eubacterium dolichum DSM 3991	428	160913555	Firmicutes	Bacteria
	Csp1	hypothetical protein CLOSPI_01716	Clostridium spiroforme DSM 1552	427	169350939	Firmicutes	Bacteria
		Average		402			
		Standard Deviation		24			
Sub-Cluster2	Lp1	cell division protein FtsW	Lactobacillus plantarum WCFS1	388	28378756	Firmicutes	Bacteria
	Es1	cell cycle protein	Exiguobacterium sibiricum 255-15	423	172058002	Firmicutes	Bacteria
	Mpl1	cell division protein FtsW	Melissococcus plutonius ATCC 35311	391	332686577	Firmicutes	Bacteria
	Efa1	FtsWEF	Enterococcus faecium	372	4633280	Firmicutes	Bacteria
	Seq1	cell division protein	Streptococcus equi subsp. zooepidemicus	426	225868819	Firmicutes	Bacteria
	Lru1	cell division protein FtsW	Lactobacillus ruminis ATCC 25644	402	323340646	Firmicutes	Bacteria
	Lca1	cell division membrane protein	Lactobacillus casei ATCC 334	389	116494803	Firmicutes	Bacteria
	Lme1	cell division membrane protein	Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293	390	116618688	Firmicutes	Bacteria
	Ooe1	cell division membrane protein	Oenococcus oeni PSU-1	416	116491256	Firmicutes	Bacteria
	Wpa1	cell division protein FtsW	Weissella paramesenteroides ATCC 33313	403	241896034	Firmicutes	Bacteria
	Dpi1	hypothetical protein HMPREF9703_00620	Dolosigranulum pigrum ATCC 51524	390	374563026	Firmicutes	Bacteria
		Average		399			
		Standard Deviation		17			
Cluster3	Sub-Cluster 1			443	336424830	Firmicutes	Bacteria
	Lba1	hypothetical protein HMPREF0994_00868	Lachnospiraceae bacterium 3_1_57FAA_CT1				
Sub-Cluster 1	Rla1	hypothetical protein RUM/LAC_01502	Ruminococcus lactaris ATCC 29176	447	197302773	Firmicutes	Bacteria
	Cha1	hypothetical protein HMPREF9473_02234	Clostridium hathewayi WAL-18680	447	358063562	Firmicutes	Bacteria
	Csp1	hypothetical protein HMPREF1020_04479	Clostridium sp. 7_3_54FAA	462	355629432	Firmicutes	Bacteria
	Csc1	hypothetical protein CLOSCI_01775	Clostridium scindens ATCC 35704	450	167759428	Firmicutes	Bacteria
	Dfo1	hypothetical protein DORFOR_00305	Dorea formicigenerans ATCC 27755	472	166030639	Firmicutes	Bacteria
	Dha1	hypothetical protein DSY4370	Desulfobacterium hafniense Y51	397	89897116	Firmicutes	Bacteria
	Ace1	cell division membrane protein-like protein	Acetivibrio cellulolyticus CD2	468	366165597	Firmicutes	Bacteria
	Cce1	cell cycle protein	Clostridium cellulolyticum H10	448	220930259	Firmicutes	Bacteria
	Ame2	cell cycle protein	Alkaliphilus metalliredigens QYMF	457	150389358	Firmicutes	Bacteria
	Cie2	cell cycle protein	Clostridium lentocellum DSM 5427	523	326790417	Firmicutes	Bacteria
	Cbo1	hypothetical protein CLL_A2660	Clostridium botulinum B str. Eklund 17B	431	187935779	Firmicutes	Bacteria
	Dac1	cell division membrane protein-like protein	Desulfotomaculum acetoxidans DSM 771	430	258515397	Firmicutes	Bacteria
	Cce2	cell cycle protein	Clostridium cellulovorans 743B	442	302876321	Firmicutes	Bacteria
	Cbo2	putative membrane protein	Clostridium botulinum D str. 1873	429	253681570	Firmicutes	Bacteria
	Bpu1	membrane protein, putative	Bacillus pumilus ATCC 7061	415	194017330	Firmicutes	Bacteria
Sub-Cluster 1	Dgi1	cell cycle protein	Desulfotomaculum gibsoniae DSM 7213	431	357039688	Firmicutes	Bacteria
	Psp1	hypothetical protein GYMC10_0410	Paenibacillus sp. Y412MC10	439	261404280	Firmicutes	Bacteria
	Bps1	ftsW gene product	Bacillus pseudofirmus OF4	431	288555449	Firmicutes	Bacteria
	Lmo1	hypothetical protein LmonocyFSL_06455	Listeria monocytogenes FSL12-003	358	255026078	Firmicutes	Bacteria

Table 2, continued

Lmo2	hypothetical protein lmo0421	Listeria monocytogenes EGD-e	416	16802465	Firmicutes	Bacteria
Dru1	cell cycle protein	Desulfotomaculum ruminis DSM 2154	428	334091929	Firmicutes	Bacteria
Bsp1	cell cycle protein FtsW	Bacillus sp. m3-13	411	311030543	Firmicutes	Bacteria
Bha1	hypothetical protein BH3360	Bacillus halodurans C-125	431	15615922	Firmicutes	Bacteria
Bsp2	cell division protein, FtsW/RodA/SpoVE family	Bacillus sp. NRRL B-14911	428	89096250	Firmicutes	Bacteria
Ban1	cell cycle protein FtsW	Bacillus anthracis str. Ames	392	30261157	Firmicutes	Bacteria
SsI1	bacterial cell division membrane protein	Solibacillus silvestris StLB046	428	327439536	Firmicutes	Bacteria
Bco1	cell cycle protein FtsW	Bacillus coahuilensis m4-4	374	205372608	Firmicutes	Bacteria
		Average	433			
		Standard Derivation	31			
Sub-Cluster2						
Cpr1	rod shape-determining (roda protein) transmembrane protein	Coprothermobacter proteolyticus DSM 5265	342	206895993	Firmicutes	Bacteria
Bex1	putative membrane protein	Bulleidia extracta W1219	445	283768742	Firmicutes	Bacteria
Cha1	predicted protein	Clostridiales bacterium 1_7_47FAA	466	239624988	Firmicutes	Bacteria
Eha1	hypothetical protein EUBHAL_03275	Eubacterium hallii DSM 3353	529	225028983	Firmicutes	Bacteria
Rsp1	Bacterial cell division membrane protein	Ruminococcus sp. 18P13	433	291544252	Firmicutes	Bacteria
		Average	443			
		Standard Derivation	67			
Sub-Cluster3						
Taq1	cell cycle protein	Thermus aquaticus Y51MC23	352	218296769	Deinococcus-Thermus	Bacteria
Th1	cell cycle protein FtsW	Thermus thermophilus HB8	355	55981052	Deinococcus-Thermus	Bacteria
Th2	cell cycle protein FtsW	Thermus thermophilus HB27	353	46199026	Deinococcus-Thermus	Bacteria
Tra1	cell cycle protein	Truepera radiovictrix DSM 17093	359	297623833	Deinococcus-Thermus	Bacteria
Vva1	cell cycle protein	Victivallis vadensis ATCC BAA-548	360	281355335	Lentisphaerae	Bacteria
Cba2	bacterial cell division membrane protein	uncultured Chloroflexi bacterium HF0200_09109	378	297181777	Chloroflexi	Bacteria
		Average	360			
		Standard Derivation	10			
Sub-Cluster4						
Dps1	cell division protein (FtsW)	Desulfotalea psychrophila LSV54	521	51246277	Deltaproteobacteria	Bacteria
Sub-Cluster 5						
Sus1	unnamed protein product	Candidatus Solibacter usitatus Ellin6076	1071	116626558	Acidobacteria	Bacteria
Sus2	unnamed protein product	Candidatus Solibacter usitatus Ellin6076	1029	116622382	Acidobacteria	Bacteria
		Average	1050			
		Standard Derivation	30			
Sub-Cluster6						
Avi1	cell cycle protein	Allochromatium vinosum DSM 180	452	288940362	Gammaaproteobacteria	Bacteria
Sub-Cluster7						
Dha2	hypothetical protein DSY2687	Desulfotobacterium hafniense Y51	428	89895433	Firmicutes	Bacteria
Tte1	cell division membrane protein	Thermoanaerobacter tengcongensis MB4	414	20808372	Firmicutes	Bacteria
Ame1	cell cycle protein	Alkaliphilus metalliredigens QYMF	404	150388174	Firmicutes	Bacteria
Cth1	cell cycle protein	Clostridium thermocellum ATCC 27405	422	125975525	Firmicutes	Bacteria
Eve1	hypothetical protein EUBVEN_00638	Eubacterium ventriosum ATCC 27560	482	154482940	Firmicutes	Bacteria
Bpr1	rod shape-determining protein RodA2	Butyrivibrio proteoelasticus B316	461	302671004	Firmicutes	Bacteria
Osp1	cell division protein FtsW	Orlacterium sp. oral taxon 078 str. F0262	553	291459005	Firmicutes	Bacteria
Orf1	Bacterial cell division membrane protein	butyrate-producing bacterium SS3/4	432	291563161	Firmicutes	Bacteria
Cle1	cell cycle protein	Clostridium lentocellum DSM 5427	456	326791461	Firmicutes	Bacteria

Table 2, continued

Esp1	penicillin-binding protein transpeptidase	Epulopiscium sp. 'N.t. morphotype B'	451	1683333335	Firmicutes	Bacteria
Asp1	cell cycle protein, FtsW/RodA/SpoVE family	Candidatus Arthromitus sp. SFB-mouse-Japan	402	342731758	Firmicutes	Bacteria
Cpe1	cell cycle protein, FtsW/RodA/SpoVE family	Clostridium perfringens str. 13	409	18309332	Firmicutes	Bacteria
		Average	443			
		Standard Derivation	43			
Sub-Cluster8						
Twh1	cell division protein FtsW	Tropheryma whipplei str. Twist	448	28493742	Actinobacteria	Bacteria
Cmi1	FtsW/RodA/SpoVE family cell cycle protein	Clavibacter michiganensis subsp. sepedonicus	441	170780482	Actinobacteria	Bacteria
Bdi1	protein involved in cell wall formation and sta	Bifidobacterium adolescentis ATCC 15703	527	119025059	Actinobacteria	Bacteria
Bbr1	transpeptidase	Bifidobacterium breve	535	295923900	Actinobacteria	Bacteria
Mle1	RodA	Mycobacterium leprae	465	1262364	Actinobacteria	Bacteria
Cac1	cell cycle protein	Catenulispora acidiphila DSM 44928	659	256389268	Actinobacteria	Bacteria
Tfu1	unnamed protein product	Thermobifida fusca YX	481	72160831	Actinobacteria	Bacteria
Psp2	Cell division protein FtsW	Patulibacter sp. I11	447	367467153	Actinobacteria	Bacteria
Afe1	cell cycle protein	Acidimicrobium ferrooxidans DSM 10331	444	256371790	Actinobacteria	Bacteria
Cu1	cell division membrane protein	Cryptobacterium curtum DSM 15641	921	256826781	Actinobacteria	Bacteria
Mth1	cell cycle protein	Moorella thermoacetica ATCC 39073	405	83589761	Firmicutes	Bacteria
Kra1	cell cycle protein	Ktedonobacter racemifer DSM 44963	470	298242771	Chloroflexi	Bacteria
Hau1	cell cycle protein	Herpetosiphon aurantiacus DSM 785	463	159897974	Chloroflexi	Bacteria
Tca1	cell cycle protein	Thermosinus carboxydvorans Nor1	424	121533695	Firmicutes	Bacteria
Sno1	cell division protein FtsW	Selenomonas noxia ATCC 43541	424	292670251	Firmicutes	Bacteria
		Average	504			
		Standard Derivation	127			
Cluster 4						
Sub-Cluster1						
Tin1	cell cycle protein	Thermodesulfatator indicus DSM 15286	358	337286686	Thermodesulfobacteria	Bacteria
Tsp1	cell cycle protein	Thermodesulfobacterium sp. OPB45	361	337289086	Thermodesulfobacteria	Bacteria
Vsp1	rod shape-determining protein RodA	Verrucomicrobium spinosum DSM 4136	419	171914172	Verrucomicrobia	Bacteria
Cpn1	cell shape-determining protein MrdB	Chlamydomphila pneumoniae AR39	379	16752172	Chlamydiae	Bacteria
Kst1	similar to rod shape-determining protein	Candidatus Kuenenia stuttgartiensis	363	91203000	Planctomycetes	Bacteria
Lfe1	putative rod shape-determining protein (RodA	Leptospirillum ferroazotrophum	363	251771498	Nitrospirae	Bacteria
Hpy1	unnamed protein product	Helicobacter pylori 26695	381	15645363	Epsilonproteobacteria	Bacteria
		Average	375			
		Standard Derivation	21			
Sub-Cluster2						
Oba1	cell cycle protein	Opitutaceae bacterium TAV2	403	225154882	Verrucomicrobia	Bacteria
Ppa1	rod shape-determining protein RodA	Pedospaera parvula Ellin514	406	223936565	Verrucomicrobia	Bacteria
Cba1	cell cycle protein	uncultured Chloroflexi bacterium	364	374854640	Chloroflexi	Bacteria
Hau1	cell cycle protein	Herpetosiphon aurantiacus DSM 785	380	159900553	Chloroflexi	Bacteria
Fnu1	rod shape determining protein FtsW	Fusobacterium nucleatum subsp. polymorphum ATCC 10953	366	254303144	Fusobacteria	Bacteria
Fnu2	rod shape-determining protein rodA	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	366	19704362	Fusobacteria	Bacteria
Vva1	rod shape-determining protein RodA	Victivallis vadensis ATCC BAA-548	397	281356907	Lentisphaerae	Bacteria
Hpr1	rod shape-determining protein RodA	Halanaerobium praevalens DSM 2228	379	309389486	Firmicutes	Bacteria
Fma1	cell division protein RodA-like protein	Finigoldia magna ATCC 29328	367	169824414	Firmicutes	Bacteria
Ace1	cell cycle protein	Acetivibrio cellulolyticus CD2	378	366166626	Firmicutes	Bacteria
		Average	381			
		Standard Derivation	16			
Sub-Cluster3						
Dha1	hypothetical protein DS73172	Desulfotobacterium hafniense Y51	387	89895918	Firmicutes	Bacteria
Sel1	unnamed protein product	Synechococcus elongatus PCC 7942	421	81299913	Cyanobacteria	Bacteria
Pma2	hypothetical protein PMM1580	Prochlorococcus marinus subsp. pastoris str. CCMP 1986	423	33862136	Cyanobacteria	Bacteria

Table 2, continued

Cac1	rod shape-determining protein RodA	Candidatus Cloacamonas acidaminovorans	406	218962065	none	Bacteria
Gau1	rod shape-determining protein	Gemmatimonas aurantiaca T-27	421	226227163	Gemmatimonadetes	Bacteria
Rxy1	cell cycle protein	Rubrobacter xylanophilus DSM 9941	379	108804363	Actinobacteria	Bacteria
Ccu1	rod shape-determining protein RodA	Cryptobacterium curtum DSM 15641	404	256827050	Actinobacteria	Bacteria
Cwo1	rod shape-determining protein RodA	Conexbacter woesei DSM 14684	398	284044116	Actinobacteria	Bacteria
Scw1	sfr	Streptomyces coelicolor A3(2)	372	2253083	Actinobacteria	Bacteria
Fsp1	rod shape-determining protein RodA	Frankia sp. EUN1f	411	288920836	Actinobacteria	Bacteria
Afe1	rod shape-determining protein RodA	Acidimicrobium ferrooxidans DSM 10331	382	256371744	Actinobacteria	Bacteria
Psp1	cell cycle protein	Paenibacillus sp. JDR-2	382	251798393	Firmicutes	Bacteria
Pe11	cell cycle protein	Paenibacillus elgi B69	399	357010481	Firmicutes	Bacteria
Aac1	unnamed protein product	Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446	402	258511027	Firmicutes	Bacteria
Ktu1	unnamed protein product	Kyrpidia tusciae DSM 2912	385	295695344	Firmicutes	Bacteria
Sth1	cell division protein	Symbiobacterium thermophilum IAM 14863	379	51894307	Firmicutes	Bacteria
Mru1	rod shape-determining protein RodA	Meiothermus ruber DSM 1279	360	291296298	Deinococcus-Thermus	Bacteria
Tth1	rod shape-determining protein RodA	Thermus thermophilus HB27	359	46199181	Deinococcus-Thermus	Bacteria
		Average	393			
		Standard Derivation	20			
Sub-Cluster4						
Dvu1	rod shape-determining protein RodA	Desulfovibrio vulgaris str. Hildenborough	371	46579198	Deltaproteobacteria	Bacteria
Dvu2	rod shape-determining protein RodA	Desulfovibrio vulgaris str. 'Miyazaki F';	371	218887909	Deltaproteobacteria	Bacteria
Tye1	rod shape-determining protein RodA	Thermodesulfobrio yellowstonii DSM 11347	375	206890629	Nitrospirae	Bacteria
Cte1	rod shape-determining protein RodA	Chlorobium tepidum TLS	410	21673129	Chlorobi	Bacteria
Scel1	rodA gene product	Sorangium cellulosum 'So ce 56';	378	162450684	Deltaproteobacteria	Bacteria
Ppa2	rod shape-determining protein RodA	Plesiocystis pacifica SIR-1	377	149923785	Deltaproteobacteria	Bacteria
Hoc1	rod shape-determining protein RodA	Haliangium ochraceum DSM 14365	378	262199087	Deltaproteobacteria	Bacteria
Mxc1	rod shape-determining protein RodA	Myxococcus xanthus DK 1622	375	108759095	Deltaproteobacteria	Bacteria
Hfo1	cell elongation-specific peptidoglycan biosynt	Myxococcus xanthus DK 1622	362	373485978	Acidobacteria	Bacteria
Kve1	rod shape-determining protein RodA	Candidatus Koribacter versatilis Ellin345	363	94968494	Acidobacteria	Bacteria
Cth1	cell division membrane protein	Candidatus Chloracidobacterium thermophilum B	381	347755277	Acidobacteria	Bacteria
Sru1	rod shape-determining protein RodA	Salinibacter ruber DSM 13855	420	83814857	Bacteroidetes	Bacteria
		Average	380			
		Standard Derivation	17			
Sub-Cluster5						
Wgl1	hypothetical protein WGLp173	Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis	368	32490922	Gammaaproteobacteria	Bacteria
Ana1	rod shape-determining protein	Arsenophonus nasoniae	370	284009239	Gammaaproteobacteria	Bacteria
Pan1	rod shape-determining protein RodA	Photobacterium angustum S14	363	90577940	Gammaaproteobacteria	Bacteria
Xfa1	Cell cycle protein	Xylella fastidiosa Dixon	420	71274880	Gammaaproteobacteria	Bacteria
Pub1	rod shape-determining protein RodA	Candidatus Pelagibacter ubique HTCC1002	373	91762876	Alphaproteobacteria	Bacteria
Ama1	rod shape-determining protein	Anaplasma marginale str. St. Maries	356	56416374	Alphaproteobacteria	Bacteria
Ppi1	rod shape-determining protein RodA	Pyramidobacter piscicola W545	384	282857428	Synergistetes	Bacteria
		Average	376			
		Standard Derivation	21			
Sub-Cluster6						
Chu1	rod shape-determining protein	Cytophaga hutchinsonii ATCC 33406	434	110639676	Bacteroidetes	Bacteria
Orf1	rod shape-determining protein	unidentified eubacterium SCB49	417	149369988	Bacteroidetes	Bacteria
Bsp1	cell cycle protein	Blattabacterium sp. (Periplaneta americana) str. BPLAN	413	261749119	Bacteroidetes	Bacteria
Tne1	Rod shape-determining protein RodA	Thermotoga neapolitana DSM 4359	358	222100711	Thermotogae	Bacteria
Tsp2	cell cycle protein	Thermotoga sp. RQ2	336	170287895	Thermotogae	Bacteria
Pmo1	cell cycle protein	Petrogala mobilis S195	368	160902814	Thermotogae	Bacteria
Cpr1	Rod shape-determining protein RodA	Coprothermobacter proteolyticus DSM 5265	361	206895469	Firmicutes	Bacteria
Tna1	cell cycle protein	Thermodesulfobium narugense DSM 14796	412	332295530	Firmicutes	Bacteria
Dth1	rod shape-determining protein RodA	Dictyoglomus thermophilum H-6-12	366	206901820	Dictyoglomi	Bacteria

Table 2, continued

Pma1	rod shape-determining protein RodA	Persephonella marina EX-H1	364	225849617	Aquificae	Bacteria
Hsp1	cell cycle protein	Hydrogenobaculum sp. Y04AAS1	371	195953480	Aquificae	Bacteria
		Average	382			
		Standard Derivation	31			
			383			
			22			
Cluster5						
Sub-Cluster1	1-Sep rod shape determining protein RodA	Staphylococcus epidermidis ATCC 12228	403	27468600	Firmicutes	Bacteria
	rod shape determining protein RodA	Staphylococcus aureus subsp. aureus Mu50	400	15925074	Firmicutes	Bacteria
Sau1	cell division membrane protein	Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	401	73662108	Firmicutes	Bacteria
Ssa1	cell division membrane protein	Oceanobacillus iheyensis HTE831	389	23098104	Firmicutes	Bacteria
Oih1	unnamed protein product	Average	398			
		Standard Derivation	6			
Sub-Cluster2	unnamed protein product	Enterococcus faecalis V583	382	29375869	Firmicutes	Bacteria
Efa1	putative rod shape-determining protein RodA	Streptococcus oralis	407	219564548	Firmicutes	Bacteria
Sor1	hypothetical protein lmo2427	Listeria monocytogenes EGD-e	391	16804465	Firmicutes	Bacteria
Lmo1	hypothetical protein lmo2687	Listeria monocytogenes EGD-e	369	16804724	Firmicutes	Bacteria
Lmo2		Average	387			
		Standard Derivation	16			
Sub-Cluster3	cell division protein FtsW	Bacillus clausii KSM-K16	386	56964563	Firmicutes	Bacteria
Bcl1	stage V sporulation protein E	Bacillus clausii KSM-K16	418	56965579	Firmicutes	Bacteria
Bcl2	cell cycle protein FtsW	Bacillus cereus ATCC 10987	392	42783277	Firmicutes	Bacteria
Bce1		Average	399			
		Standard Derivation	17			
Sub-Cluster4	hypothetical protein CATMIT_03016	Catenibacterium mitsuokai DSM 15897	400	224543795	Firmicutes	Bacteria
Cml1	hypothetical protein CLOSPI_00527	Clostridium spiroforme DSM 1552	398	169349789	Firmicutes	Bacteria
Csp1	hypothetical protein EUBDOL_00755	Eubacterium dolichum DSM 3991	404	160914748	Firmicutes	Bacteria
Edo1	hypothetical protein HMPREF0357_10652	Erysipelothrix rhusiopathiae ATCC 19414	403	323342240	Firmicutes	Bacteria
Erh1		Average	401			
		Standard Derivation	3			
Sub-Cluster5	hypothetical protein BACCAP_03879	Pseudoflavonifractor capillosus ATCC 29799	395	154500215	Firmicutes	Bacteria
Pca1	hypothetical protein ANACOL_03753	Anaerotruncus colihominis DSM 17241	378	167772376	Firmicutes	Bacteria
Aco1	Bacterial cell division membrane protein	Ruminococcus sp. 18P13	381	291544235	Firmicutes	Bacteria
Rsp1	cell cycle protein	Subdoligranulum variable DSM 15176	387	261366552	Firmicutes	Bacteria
Sva1	unnamed protein product	Eubacterium limosum KIST612	384	310827192	Firmicutes	Bacteria
Eli1	hypothetical protein CLOLEP_03338	Clostridium leptum DSM 753	391	160934465	Firmicutes	Bacteria
Cle1		Average	386			
		Standard Derivation	6			
Sub-Cluster6	bacterial cell division membrane protein	Eubacterium saphenum ATCC 49989	368	255994114	Firmicutes	Bacteria
Esa1	unnamed protein product	Filifactor alocis ATCC 35896	379	374307398	Firmicutes	Bacteria
Fal1		Average	374			
		Standard Derivation	8			
Sub-Cluster7	rod shape-determining protein rodA	Fusobacterium sp. 3_1_5R	413	317059362	Fusobacteria	Bacteria
Fsp1	rod shape-determining protein rodA	Fusobacterium nucleatum subsp. nucleatum ATCC 25586	417	19703394	Fusobacteria	Bacteria
Fnu1		Average	415			

Table 2, continued

Sub-Cluster8			Standard Derivation	3		
Cpe1	cell cycle protein, FtsW/RodA/SpoVE family	Clostridium perfringens str. 13		374	18311118 Firmicutes	Bacteria
Asp1	cell cycle protein FtsW	Candidatus Arthromitus sp. SFB-rat-Yit		366	347542186 Firmicutes	Bacteria
		Average		370		
		Standard Derivation		6		
Sub-Cluster9						
Esp1	rod shape-determining protein RodA	Epulopiscium sp. 'N.t. morphotype B'		373	168334681 Firmicutes	Bacteria
Osi1	FtsW/RodA/SpoVE family cell division protein	Oribacterium sinus F0268		379	227872352 Firmicutes	Bacteria
Orf3	Bacterial cell division membrane protein	butyrate-producing bacterium SS3/4		376	291563264 Firmicutes	Bacteria
Cph1	cell cycle protein	Clostridium phytofermentans ISDg		389	160880500 Firmicutes	Bacteria
		Average		379		
		Standard Derivation		7		
Sub-Cluster10						
Orf1	hypothetical protein KB13_1183	beta proteobacterium KB13		364	254468958 Betaproteobacteria	Bacteria
Orf2	bacterial cell division membrane protein	uncultured gamma proteobacterium HF0010_11B23		339	297180558 Gammaproteobacteria	Bacteria
Gha1	putative cell division protein	Gemella haemolysans ATCC 10379		427	241888881 Firmicutes	Bacteria
		Average		377		
		Standard Derivation		45		
Cluster6						
Hne1	rodA gene product	Hyphomonas neptunium ATCC 15444		375	114797109 Alphaproteobacteria	Bacteria
Pma7	Bacterial cell division membrane protein	Planctomyces maris DSM 8797		486	149177963 Planctomycetes	Bacteria
		Average		431		
		Standard Derivation		78		
Cluster7						
Sub-Cluster1						
Pp11	cell division protein FtsW	Pyrimidobacter piscicola W5455		369	282857258 Synergistetes	Bacteria
Ssp1	cell division protein FtsW	Synergistes sp. 3_1_syn1		381	365174757 Synergistetes	Bacteria
		Average		375		
		Standard Derivation		8		
Sub-Cluster2						
Hpy1	unnamed protein product	Helicobacter pylori 26695		388	15646167 Epsilonproteobacteria	Bacteria
		Average				
		Standard Derivation				
Sub-Cluster3						
Fsp1	cell division membrane protein-like protein	Frankia sp. EUN1f		473	288918399 Actinobacteria	Bacteria
Rfa1	putative dual function protein	Rhodococcus fascians D188		914	356609642 Actinobacteria	Bacteria
		Average		695		
		Standard Derivation		312		
Sub-Cluster4						
Cgl1	cell division membrane protein	Corynebacterium glucuronolyticum ATCC 51866		506	227543133 Actinobacteria	Bacteria
Sub-Cluster5						
Ate1	FtsW protein	Actinoplanes teichomyceticus		420	551631153 Actinobacteria	Bacteria
Sub-Cluster6						
Pac1	cell division protein FtsW	Propionibacterium acnes KPA171202		440	50842244 Actinobacteria	Bacteria
Afe1	cell cycle protein	Acidimicrobium ferrooxidans DSM 10331		379	256372026 Actinobacteria	Bacteria
		Average		410		
		Standard Derivation		43		

Table 2, continued

Sub-Cluster7 Pde1	cell division protein FtsW	Parascardovia denticolens F0305	458 294787086 Actinobacteria	Bacteria
Sub-Cluster8 Aod1 Aco1	hypothetical protein ACTODO_01560 stage V sporulation protein E	Actinomyces odontolyticus ATCC 17982 Actinomyces coleocanis DSM 15436 Average Standard Derivation	431 154509043 Actinobacteria 411 227494647 Actinobacteria 421 14	Bacteria Bacteria
Sub-Cluster9 Aha1	cell cycle protein	Arcanobacterium haemolyticum DSM 20595	409 297571252 Actinobacteria	Bacteria
Sub-Cluster 10 Bli1 Twh1 Lxy1	cell division protein FtsW cell division protein FtsW cell division protein	Brevibacterium linens BL2 Tropheryma whipplei str. Twist Leifsonia xyl subsp. xyl str. CTCB07 Average Standard Derivation	532 260905304 Actinobacteria 370 28493194 Actinobacteria 404 50955150 Actinobacteria 435 85	Bacteria Bacteria Bacteria
Sub-Cluster11 Mmu1 Agr1 Aur1	bacterial cell division membrane protein hypothetical protein HMPREF0045_00257 stage V sporulation protein E	Mobiluncus mulieris ATCC 35239 Actinomyces graevenitzi C83 Actinomyces urogenitalis DSM 15434 Average Standard Derivation	436 306818517 Actinobacteria 422 365824547 Actinobacteria 421 227497589 Actinobacteria 426 8	Bacteria Bacteria Bacteria

Table 3. Paralogues observed in Cluster 1 of the CaCA family.

Organism	Homologues	Cluster	GI number
<i>Thalassiosira pseudonana</i>	Tps1	1	224006648
	Tps4	5	224015509
	Tps2	5	224001622
	Tps3	5	224010691
<i>Schizophyllum commune</i>	Sco1	1	302674539
	Sco2	1	302685091
<i>Puccinia graminis</i>	Pgr1	1	331239911
	Pgr2	1	331221453
	Pgr3	2	331239133
<i>Podospora anserina</i>	Pan1	1	171690580
	Pan2	1	171688494
<i>Perkinsus marinus</i>	Pma3	1	294891166
	Pma2	4	294953203
<i>Penicillium chrysogenum</i>	Pch2	1	255953175
	Pch1	1	255933738
<i>Naumovozya castellii</i>	Nca3	1	342302654
	Nca2	2	342302430
<i>Methylococcus capsulatus</i>	Mca2	1	53804065
	Mca1	3	53805213
<i>Ectocarpus siliculosus</i>	Esi2	1	298705602

Table 3, continued

	Esi1	5	298713622
<i>Coprinopsis cinerea</i>	Cci1	1	169856345
	Cci3	1	299748017
	Cci2	2	299743625
<i>Conexibacter woesei DSM 14684</i>	Cwo2	1	284043117
	Cwo1	3	284047105
<i>Aspergillus oryzae</i>	Aor1	1	317150008
	Aor2	1	83766993

Table 4. Paralogues observed in Cluster 2 of the CaCA family.

Organism	Homologues	Cluster	GI number
<i>Trichomonas vaginalis G3</i>	Tva1	2	123423586
	Tva2	2	154419784
<i>Puccinia graminis</i>	Pgr1	1	331239911
	Pgr2	1	331221453
	Pgr3	2	331239133
<i>Naumovozyma castellii</i>	Nca3	1	342302654
	Nca2	2	342302430
<i>Naegleria gruberi</i>	Ngr1	2	290990046

Table 4, continued

	Ngr2	2	290975515
<i>Coprinopsis cinerea</i>	Cci1	1	169856345
	Cci3	1	299748017
	Cci2	2	299743625

Table 5. Paralogues observed in Cluster 3 of the CaCA family.

Organism	Homologues	Cluster	GI number
Methylococcus capsulatus	Mca2	1	53804065
	Mca1	3	53805213
Conexibacter woesei	Cwo2	1	284043117
	Cwo1	3	284047105

Table 6. Paralogues observed in Cluster 4 of the CaCA family.

Organism	Homologues	Cluster	GI number
<i>Trichoplax adhaerens</i>	Tad2	4	195999284
	Tad1	4	196014843
	Tad3	4	196001263
	Tad4	5	196016029
<i>Trichinella spiralis</i>	Tsp1	4	339235315

Table 6, continued

	Tsp3	4	339238677
<i>Tetraodon nigroviridis</i>	Tni1	4	47213382
	Tni3	5	47214497
	Tni2	5	47225868
<i>Perkinsus marinus</i>	Pma3	1	294891166
	Pma2	4	294953203
	Pma4	5	294944237
<i>Pediculus humanus corporis</i>	Phu2	4	242016779
	Phu1	5	242020994
<i>Oikopleura dioica</i>	Odi1	4	313235239
	Odi3	4	313230968
<i>Nematostella vectensis</i>	Nve3	4	156397050
	Nve2	5	156385129
	Nve1	5	156403746
<i>Hydra magnipapillata</i>	Hma3	4	221121648
	Hma1	5	221101518
	Hma2	5	221122333
<i>Danio rerio</i>	Dre3	2	110005804
	Dre1	4	181339524
	Dre2	4	181339493

Table 6, continued

<i>Ciona intestinalis</i>	Cin2	4	198426482
	Cin1	5	198422506
<i>Caenorhabditis briggsae</i>	Cbr2	4	309360090
	Cbr1	5	309361787
<i>Branchiostoma floridae</i>	Bfl6	4	260836357
	Bfl3	4	260829777
	Bfl2	5	260814626
	Bfl4	5	260822461
	Bfl5	5	260797070
<i>Ascaris suum</i>	Asu1	4	324504975
	Asu2	4	324503600
	Asu3	5	324509213

Table 7. Paralogues observed in cluster 5 of the CaCA family.

Organism	Homologues	Cluster	GI number
<i>Trichoplax adhaerens</i>	Tad2	4	195999284
	Tad1	4	196014843
	Tad3	4	196001263
	Tad4	5	196016029

Table 7, continued

<i>Tribolium castaneum</i>	Tca2	5	270010753
	Tca1	5	270007787
	Tca3	5	270006895
<i>Thalassiosira pseudonana</i>	Tps1	1	224006648
	Tps4	5	224015509
	Tps2	5	224001622
	Tps3	5	224010691
<i>Tetraodon nigroviridis</i>	Tni3	4	47213382
	Tni1	5	47214497
	Tni2	5	47225868
<i>Strongylocentrotus purpuratus</i>	Spu2	5	47551219
	Spu1	5	115709825
	Spu3	5	115871986
<i>Saccoglossus kowalevskii</i>	Sko1	5	291221728
	Sko2	5	291243690
<i>Plesiocystis pacifica</i>	Ppa2	5	149923824
	Ppa3	6	149925192
<i>Perkinsus marinus</i>	Pma3	1	294891166
	Pma2	4	294953203
	Pma4	5	294944237

Table 7, continued

<i>Pediculus humanus</i>	Phu2	4	242016779
	Phu1	5	242020994
<i>Nematostella vectensis</i>	Nve3	4	156397050
	Nve2	5	156385129
	Nve1	5	156403746
<i>Mus musculus</i>	Mmu1	5	21450183
	Mmu2	5	159032066
<i>Morone saxatilis</i>	Msa1	5	118139452
	Msa2	5	118139446
<i>Micromonas pusilla</i>	Mpu2	5	303285051
	Mpu1	5	303279510
<i>Hydra magnipapillata</i>	Hma3	4	221121648
	Hma1	5	221101518
	Hma2	5	221122333
<i>Ectocarpus siliculosus</i>	Esi2	1	298705602
	Esi1	5	298713622
<i>Drosophila willistoni</i>	Dwi1	5	195438385
	Dwi2	5	195447092
<i>Drosophila pseudoobscura</i> <i>pseudoobscura</i>	Dps3	5	198467881

Table 7, continued

	Dps2	5	198454486
<i>Daphnia pulex</i>	Dpu1	5	321476184
	Dpu3	5	321459376
<i>Ciona intestinalis</i>	Cin2	4	198426482
	Cin1	5	198422506
<i>Camponotus floridanus</i>	Cfl1	5	307179272
	Cfl2	5	307179269
<i>Caenorhabditis briggsae</i>	Cbr2	4	309360090
	Cbr1	5	309361787
<i>Brugia malayi</i>	Bma1	5	170592429
	Bma2	5	170593899
<i>Branchiostoma floridae</i>	Bfl6	4	260836357
	Bfl3	4	260829777
	Bfl2	5	260814626
	Bfl4	5	260822461
	Bfl5	5	260797070
<i>Ascaris suum</i>	Asu1	4	324504975
	Asu2	4	324503600
	Asu3	5	324509213
<i>Anolis carolinensis</i>	Aca3	5	327286665
	Aca5	5	327259250

Table 7, continued

<i>Aedes aegypti</i>	Aae1	5	157107608
	Aae2	5	157107606
<i>Acyrtosiphon pisum</i>	Api1	5	328706570
	Api3	5	328701889
	Api2	5	328706562

Table 8. Paralogues observed in Cluster 6 of the CaCA family.

Organism	Homologues	Cluster	GI number
<i>Plesiocystis pacifica</i>	Ppa2	5	149923824
	Ppa3	6	149925192
<i>Candidatus Kuenenia stuttgartiensis</i>	Kst1	6	91202036
	Kst2	6	91202741

Table 9. Paralogues observed in Cluster 1 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Alicyclobacillus acidocaldarius</i>	Aac2	1	258511024
	Aac1	4	258511027
<i>Alistipes putredinis</i>	Apu1	1	167752865
	Apu2	4	167753023

Table 9, continued

<i>Bacillus clausii</i>	Bcl3	1	56964564
	Bcl1	5	56964563
	Bcl2	5	56965579
<i>Candidatus Kuenenia stuttgartiensis</i>	Kst1	1	91200084
	Kst2	4	91203000
<i>Candidatus Pelagibacter</i>	Pub2	1	91762844
	Pub1	4	91762876
<i>Candidatus Solibacter</i>	Sus1	1	116626352
	Sus2	3	116626558
	Sus3	3	116622382
<i>Clostridium thermocellum</i>	Cth3	1	125973490
	Cth1	3	125975525
<i>Coprococcus eutactus</i>	Ceu1	1	163816708
	Ceu2	1	163814972
<i>Cytophaga hutchinsonii</i>	Chu2	1	110639123
	Chu1	4	110639676
<i>Desulfitobacterium hafniense</i>	Dha1	1	89895000
	Dha2	3	89895433
	Dha4	3	89897116
	Dha3	4	89895918

Table 9, continued

<i>Desulfovibrio vulgaris</i>	Dvu3	1	218886059
	Dvu4	4	218887909
	Dvu2	1	46580909
	Dvu1	4	46579198
<i>Dictyoglomus thermophilum</i>	Dth1	1	206901982
	Dth2	4	206901820
<i>Elusimicrobium minutum</i>	Emi2	1	187251656
	Emi1	4	187251782
<i>Epulopiscium</i>	Esp3	1	168333727
	Esp1	3	168333335
	Esp4	5	168334681
<i>Eubacterium dolichum</i>	Edo3	1	160915795
	Edo1	2	160913555
	Edo2	5	160914748
<i>Eubacterium saphenum</i>	Esa1	1	255994321
	Esa2	5	255994114
<i>Filifactor alocis</i>	Fal1	1	374307642
	Fal2	5	374307398
<i>Finegoldia magna</i>	Fma2	1	169824313
	Fma1	4	169824414

Table 9, continued

<i>Fluviicola taffensis</i>	Fta2	1	327404200
	Fta1	4	327403876
<i>Gemmatimonas aurantiaca</i>	Gau1	1	226226833
	Gau2	4	226227163
<i>Halanaerobium praevalens</i>	Hpr1	1	309388992
	Hpr2	4	309389486
<i>Holophaga foetida</i>	Hfo1	1	373487374
	Hfo2	4	373485978
<i>Hydrogenobaculum</i>	Hsp1	1	195952592
	Hsp2	4	195953480
<i>Leptospirillum ferrodiazotrophum</i>	Lfe2	1	251772198
	Lfe1	4	251771498
<i>Microscilla marina</i>	Mma1	1	124009922
	Mma2	4	124006885
<i>Oribacterium</i>	Osp1	1	291459275
	Osp2	3	291459005
<i>Pedosphaera parvula</i>	Ppa3	1	223937421
	Ppa1	4	223936565
<i>Persephonella marina</i>	Pma1	1	225850178
	Pma6	4	225849617

Table 9, continued

<i>Petrotoga mobilis</i>	Pmo1	1	160903089
	Pmo2	4	160902814
<i>Pseudoflavonifractor capillosus</i>	Pca2	1	154500747
	Pca1	5	154500215
<i>Ruminococcus</i>	Rsp1	1	291544490
	Rsp2	3	291544252
	Rsp3	5	291544235
<i>Salinibacter ruber</i>	Sru1	1	83814197
	Sru2	4	83814857
<i>Subdoligranulum variabile</i>	Sva2	1	261367513
	Sva1	5	261366552
<i>Synechococcus elongatus</i>	Sel2	1	81299135
	Sel1	4	81299913
<i>Thermodesulfobium narugense</i>	Tna1	1	332295507
	Tna2	4	332295530
<i>Thermosinus carboxydivorans</i>	Tca1	1	121533643
	Tca2	3	121533695
<i>Thermotoga</i>	Tsp3	1	170288513
	Tsp2	4	170287895
<i>Treponema denticola</i>	Tde1	1	42526710

Table 9, continued

	Tde2	4	42526861
<i>unidentified eubacterium</i>	Orf8	1	149371576
	Orf9	4	149369988
<i>Verrucomicrobium spinosum</i>	Vsp2	1	171913319
	Vsp1	4	171914172
<i>Victivallis vadensis</i>	Vva2	1	281357083
	Vva1	3	281355335
	Vva3	4	281356907
<i>Xylella fastidiosa</i> Dixon	Xfa1	1	71275123
	Xfa2	4	71274880

Table 10. Paralogues observed in cluster 2 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Clostridium spiroforme</i>	Csp1	2	169350939
	Csp4	5	169349789
<i>Erysipelothrix rhusiopathiae</i>	Erh1	2	323341738
	Erh2	5	323342240
<i>Eubacterium dolichum</i>	Edo3	1	160915795
	Edo1	2	160913555
<i>Gemella haemolysans</i>	Gha1	2	241889569

Table 10, continued

	Gha2	5	241888881
<i>Listeria monocytogenes</i>	Lmo1	2	16804725
	Lmo3	5	16804465
	Lmo4	5	16804724

Table 11. Paralogues observed in Cluster 3 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Acetivibrio cellulolyticus</i>	Ace1	3	366165597
	Ace2	4	366166626
<i>Acidimicrobium ferrooxidans</i>	Afe2	3	256371790
	Afe3	4	256371744
<i>Alkaliphilus metalliredigens</i>	Ame1	3	150388174
	Ame2	3	150389358
<i>butyrate-producing bacterium</i>	Orf7	3	291563161
	Orf6	5	291563264
<i>Candidatus Solibacter usitatus</i>	Sus2	3	116626558
	Sus3	3	116622382
<i>Clostridium lentocellum</i>	Cle2	3	326791461
	Cle3	3	326790417
<i>Clostridium perfringens</i>	Cpe2	3	18309332

Table 11, continued

	Cpe1	5	18311118
<i>Coprothermobacter proteolyticus</i>	Cpr2	3	206895993
	Cpr1	4	206895469
<i>Cryptobacterium curtum</i>	Ccu1	3	256826781
	Ccu2	4	256827050
<i>Desulfitobacterium hafniense</i>	Dha1	1	89895000
	Dha2	3	89895433
	Dha4	3	89897116
	Dha3	4	89895918
<i>Epulopiscium</i>	Esp1	3	168333335
	Esp4	5	168334681
<i>Herpetosiphon aurantiacus</i>	Hau1	3	159897974
	Hau2	4	159900553
<i>Oribacterium</i>	Osp1	1	291459275
	Osp2	3	291459005
<i>Ruminococcus</i>	Rsp1	1	291544490
	Rsp2	3	291544252
	Rsp3	5	291544235
<i>Thermosinus carboxydivorans</i>	Tca1	1	121533643
	Tca2	3	121533695
<i>Thermus thermophilus</i>	Tth2	3	46199026

Table 11, continued

	Tth1	4	46199181
<i>Tropheryma whipplei</i>	Twh1	3	28493742
	Twh2	7	28493194
<i>Victivallis vadensis</i>	Vva2	1	281357083
	Vva1	3	281355335
	Vva3	4	281356907

Table 12. Paralogues observed in Cluster 4 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Acetivibrio cellulolyticus</i>	Ace1	3	366165597
	Ace2	4	366166626
<i>Acidimicrobium ferrooxidans</i>	Afe2	3	256371790
	Afe3	4	256371744
	Afe4	7	256372026
<i>Alicyclobacillus acidocaldarius</i>	Aac2	1	258511024
	Aac1	4	258511027
<i>Alistipes putredinis</i>	Apu1	1	167752865
	Apu2	4	167753023
<i>Candidatus Kuenenia stuttgartiensis</i>	Kst1	1	91200084

Table 12, continued

	Kst2	4	91203000
<i>Candidatus Pelagibacter ubique</i>	Pub2	1	91762844
	Pub1	4	91762876
<i>Coprothermobacter proteolyticus</i>	Cpr2	3	206895993
	Cpr1	4	206895469
<i>Cryptobacterium curtum</i>	Ccu1	3	256826781
	Ccu2	4	256827050
<i>Cytophaga hutchinsonii</i>	Chu2	1	110639123
	Chu1	4	110639676
<i>Desulfitobacterium hafniense</i>	Dha1	1	89895000
	Dha2	3	89895433
	Dha4	3	89897116
	Dha3	4	89895918
<i>Desulfovibrio vulgaris</i>	Dvu3	1	218886059
	Dvu4	4	218887909
	Dvu2	1	46580909
	Dvu1	4	46579198
<i>Dictyoglomus thermophilum</i>	Dth1	1	206901982
	Dth2	4	206901820
<i>Elusimicrobium minutum</i>	Emi2	1	187251656
	Emi1	4	187251782

Table 12, continued

<i>Finegoldia magna</i>	Fma2	1	169824313
	Fma1	4	169824414
<i>Fluviicola taffensis</i>	Fta2	1	327404200
	Fta1	4	327403876
<i>Frankia</i>	Fsp4	4	288920836
	Fsp5	7	288918399
<i>Fusobacterium nucleatum</i>	Fnu1	4	19704362
	Fnu3	5	19703394
<i>Gemmatimonas aurantiaca</i>	Gau1	1	226226833
	Gau2	4	226227163
<i>Halanaerobium praevalens</i>	Hpr1	1	309388992
	Hpr2	4	309389486
<i>Helicobacter pylori</i>	Hpy2	4	15645363
	Hpy3	7	15646167
<i>Herpetosiphon aurantiacus</i>	Hau1	3	159897974
	Hau2	4	159900553
<i>Holophaga foetida</i>	Hfo1	1	373487374
	Hfo2	4	373485978
<i>Hydrogenobaculum</i>	Hsp1	1	195952592
	Hsp2	4	195953480
<i>Leptospirillum ferro-diazotrophum</i>	Lfe2	1	251772198

Table 12, continued

	Lfe1	4	251771498
<i>Microscilla marina</i>	Mma1	1	124009922
	Mma2	4	124006885
<i>Pedosphaera parvula</i>	Ppa3	1	223937421
	Ppa1	4	223936565
<i>Persephonella marina</i>	Pma1	1	225850178
	Pma6	4	225849617
<i>Petrotoga mobilis</i>	Pmo1	1	160903089
	Pmo2	4	160902814
<i>Pyramidobacter piscolens</i>	Ppi1	4	282857428
	Ppi2	7	282857258
<i>Salinibacter ruber</i>	Sru1	1	83814197
	Sru2	4	83814857
<i>Synechococcus elongatus</i>	Sel2	1	81299135
	Sel1	4	81299913
<i>Thermodesulfobium narugense</i>	Tna1	1	332295507
	Tna2	4	332295530
<i>Thermotoga</i>	Tsp3	1	170288513
	Tsp2	4	170287895
<i>Thermus thermophilus</i>	Tth2	3	46199026
	Tth1	4	46199181

Table 12, continued

<i>Treponema denticola</i>	Tde1	1	42526710
	Tde2	4	42526861
<i>unidentified eubacterium</i>	Orf8	1	149371576
	Orf9	4	149369988
<i>Verrucomicrobium spinosum</i>	Vsp2	1	171913319
	Vsp1	4	171914172
<i>Victivallis vadensis</i>	Vva2	1	281357083
	Vva1	3	281355335
	Vva3	4	281356907
<i>Xylella fastidiosa</i>	Xfa1	1	71275123
	Xfa2	4	71274880

Table 13. Paralogues observed in Cluster 5 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Bacillus clausii</i>	Bcl1	5	56964563
	Bcl2	5	56965579
<i>butyrate-producing bacterium</i>	Orf7	3	291563161
	Orf6	5	291563264
<i>Clostridium perfringens</i>	Cpe2	3	18309332
	Cpe1	5	18311118

Table 13, continued

<i>Clostridium spiroforme</i>	Csp1	2	169350939
	Csp4	5	169349789
<i>Epulopiscium</i>	Esp3	1	168333727
	Esp1	3	168333335
	Esp4	5	168334681
<i>Erysipelothrix rhusiopathiae</i>	Erh1	2	323341738
	Erh2	5	323342240
<i>Eubacterium dolichum</i>	Edo3	1	160915795
	Edo1	2	160913555
	Edo2	5	160914748
<i>Eubacterium saphenum</i>	Esa1	1	255994321
	Esa2	5	255994114
<i>Filifactor alocis</i>	Fal1	1	374307642
	Fal2	5	374307398
<i>Fusobacterium nucleatum</i>	Fnu1	4	19704362
	Fnu3	5	19703394
<i>Gemella haemolysans</i>	Gha1	2	241889569
	Gha2	5	241888881
<i>Listeria monocytogenes</i>	Lmo5	2	16804725
	Lmo2	2	16802465
	Lmo1	5	16804465

Table 13 continue

	Lmo4	5	16804724
<i>Pseudoflavonifractor capillosus</i>	Pca2	1	154500747
	Pca1	5	154500215
<i>Ruminococcus</i>	Rsp1	1	291544490
	Rsp2	3	291544252
	Rsp3	5	291544235
<i>Subdoligranulum variabile</i>	Sva2	1	261367513
	Sva1	5	261366552

Table 14. Paralogues observed in Cluster 7 of the MPE family.

Organism	Homologues	Cluster	GI number
<i>Acidimicrobium ferrooxidans</i>	Afe2	3	256371790
	Afe3	4	256371744
	Afe4	7	256372026
<i>Frankia</i>	Fsp4	4	288920836
	Fsp5	7	288918399
<i>Helicobacter pylori</i>	Hpy2	4	15645363
	Hpy3	7	15646167
<i>Pyramidobacter piscolens</i>	Ppi1	4	282857428
	Ppi2	7	282857258
<i>Tropheryma whipplei</i>	Twh1	3	28493742
	Twh2	7	28493194

Table 15. Obvious horizontal gene transfer (CaCA family)

Cluster 1			
Proteins	Organisms	Recipients	Sources
Zma1	<i>Zea</i>	Viridiplantae	Fungi
Mba1	<i>Methanosarcina</i>	Archaea	Bacteria
Cluster 4			
Hvu1	Hordeum	Viridiplantae	Metazoa
Cluster 5			
Mpu1	<i>Micromonas</i>	Viridiplantae	Metazoa
Msp1	<i>Micromonas</i>	Viridiplantae	Metazoa
Mma1	Microscilla	Bacteria	Eukaryota
Ppa1	Plesiocysti	Bacteria	Eukaryota
Cluster 6			
Ton1	<i>Thermococcus</i>	Archaea	Bacteria
Htu1	Haloterrigena	Archaea	Bacteria
Orf7	Crenarchaeote	Archaea	Bacteria
Dgi1	Desulfovibrio	δ -proteobacteria	γ -proteobacteria
Ppa1	Plesiocystis	δ -proteobacteria	γ -proteobacteria

Table 16. Paralogues within and between each cluster of the CaCA family

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Cluster 1	6	3	2	1	2	0
Cluster 2		0	0	0	0	0
Cluster 3			0	0	0	0
Cluster 4				6	10	0
Cluster 5					14	0
Cluster 6						2

Table 17. Paralogues within and between each cluster of the MPE family

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Cluster 1		1	8	30	8	0	1
Cluster 2		0	0	0	4	0	0
Cluster 3			3	7	4	0	1
Cluster 4				0	1	0	3
Cluster 5					1	0	0
Cluster 6						0	0
Cluster 7							0

Table 18. Obvious horizontal gene transfer (MPE family)

Cluster 2			
Proteins	Organism	Recipients	Sources
Hco1	Haloplasma	Haloplasmataceae	Firmicutes
Cluster 5			
Fsp1	Fusobacterium	Fusobacteria	Firmicutes
Fnu1	Fusobacterium	Fusobacteria	Firmicutes

Table 19. Characteristic of phylogenetic clusters from the CaCA family

CaCA	Average protein size	Topology	Extra residues in the N-terminus	Extra residues in the middle	Extra residues in the C-terminus	Dominate organismal types
Cluster 1	450±86	11 TMSs	33%	11%	5%	Bacteria + Fungi
Cluster 2	996±226	13 TMSs	55%	12%	5%	Fungi
Cluster 3	380±19	11 TMSs	9%	18%	5%	Bacteria
Cluster 4	896±67	10 TMSs	17%	52%	1%	Metazoans
Cluster 5	644±183	10 TMSs	37%	35%	6%	Metazoans
Cluster 6	334±25	10 TMSs	1%	1%	1%	Bacteria

Table 20. Characteristic of phylogenetic clusters from the MPE family

MPE	Average Protein Size	Topology	Dominate Organismal Types
Cluster 1	391±26	10 TMSs	Diverse Organism
Cluster 2	401±20	10 TMSs	Firmicutes
Cluster 3	463±127	10 TMSs	Firmicutes+ Actinobacteria
Cluster 4	383±22	10 TMSs	Diverse Organism
Cluster 5	389±18	10 TMSs	Firmicutes
Cluster 6	431±78	9 TMSs	Planctomycetes α-proteobacteria
Cluster 7	451±120	10 TMSs	Firmicutes

Table 21. Comparison scores between superfamilies expressed in standard deviations.

	MPE	RND	CDF	APC	DMT	TDT
CaCA	12	12	11	13	11	10
MPE		11	13	12	11	11
RND			12	11	10	13
CDF				13	11	11
APC					14	15
DMT						12

Table 22. Comparison scores between potential repeats of 2, 3, 4, 5 TMSs of MPE family proteins

	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
1-2			7		6	9		5	4
2-3				11	6	5	4		
4-5									7

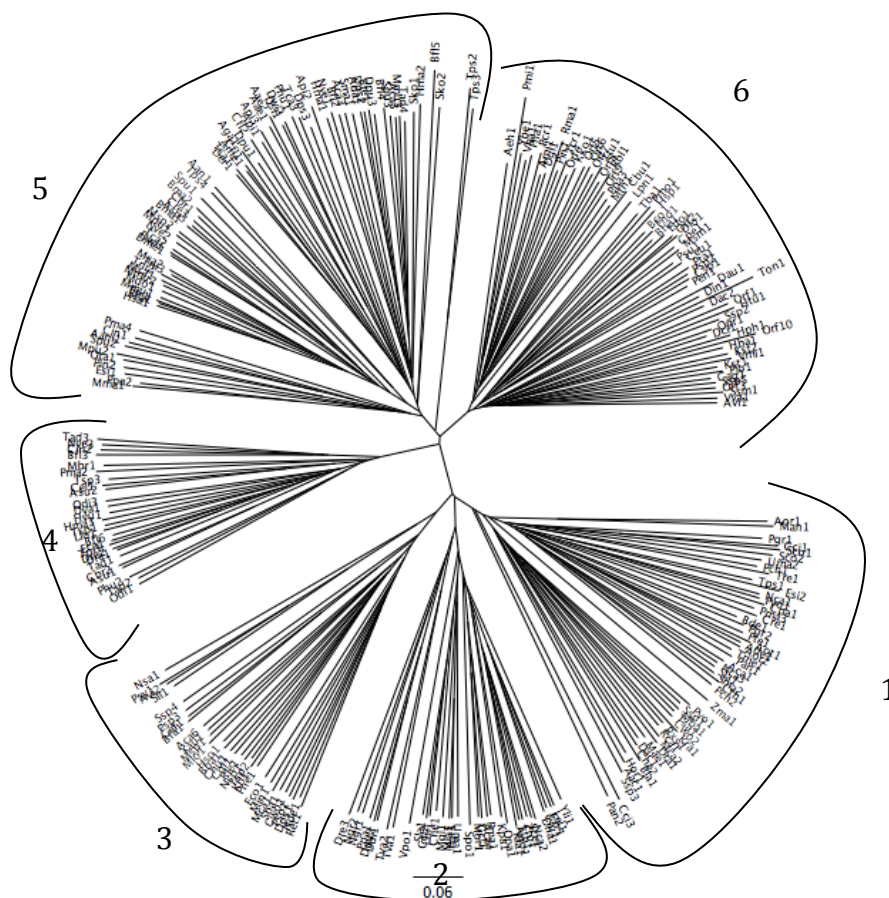
	4-6	5-7	6-8	7-9	8-10
1-3	5	7	6	5	7
2-4		7	6		7
3-5			7	6	
4-6				6	7

	5-8	7-10
1-4	5	7

	6-9
2-4	7

	7-10
3-6	6

	6-10
1-5	9



A circular phylogenetic tree illustrating the evolutionary relationships among various species. The tree is divided into ten sectors, each labeled with a number from 1 to 10. Each sector contains several species names, such as *Sip1*, *Mae1*, *Mbs1*, *Tst1*, *Sip2*, *Ble1*, *Coe1*, *Pot1*, *Che1*, *Lce1*, *Rag1*, *Rus1*, *Aes1*, *Kra1*, *Cos1*, *Hun1*, *Man1*, *Aor1*, *Pob1*, *Zma1*, *Pan1*, *Tre1*, *Pgr1*, *Scot*, *Col2*, *Sci2*, *Ums1*, *Ppr1*, *Ble1*, *Smc1*, *Coh1*, *Iff6*, *Lyn1*, *Nuc1*, *Adet*, *Pot1*, *Twe1*, *Gsa1*, *Ppo1*, *Moz1*, *Pms1*, *Eot1*, *Crs1*, *Pas1*. A scale bar at the bottom indicates a distance of 0.05.

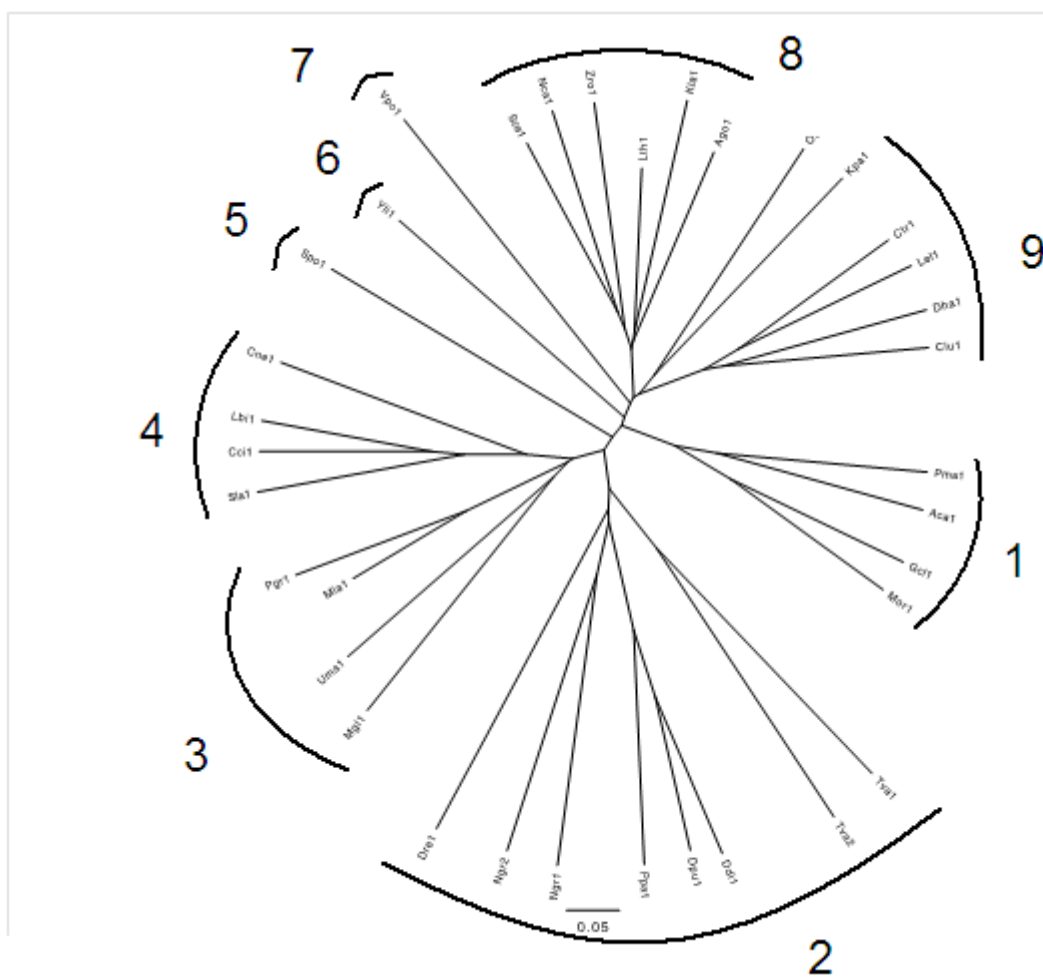


Figure 3. Protein tree for 35 homologues in Cluster 2 of the CaCA family. Cluster 2 is sub-divided into 9 sub-clusters as indicated (1-9).

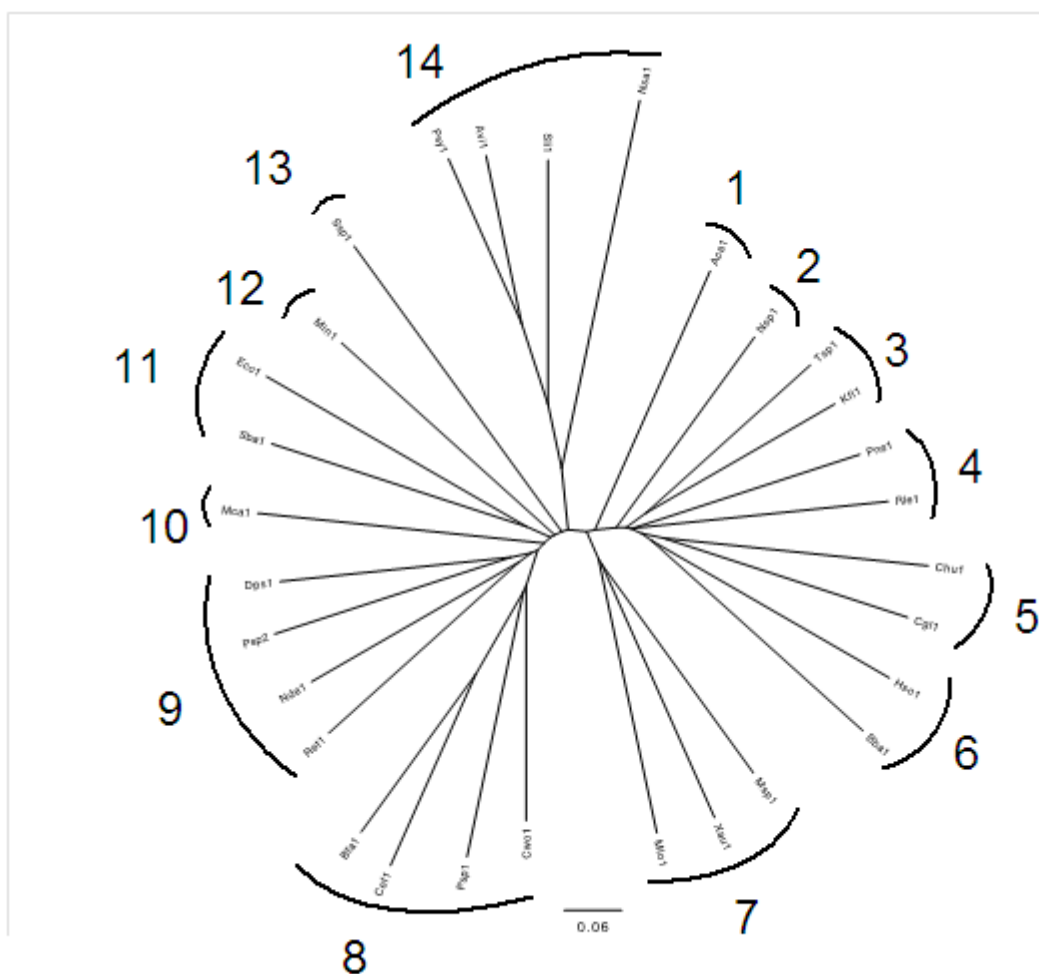


Figure 4. Protein tree for 30 homologues in Cluster 3 of the CaCA family. Cluster 3 is sub-divided into 14 sub-clusters as indicated (1-14).

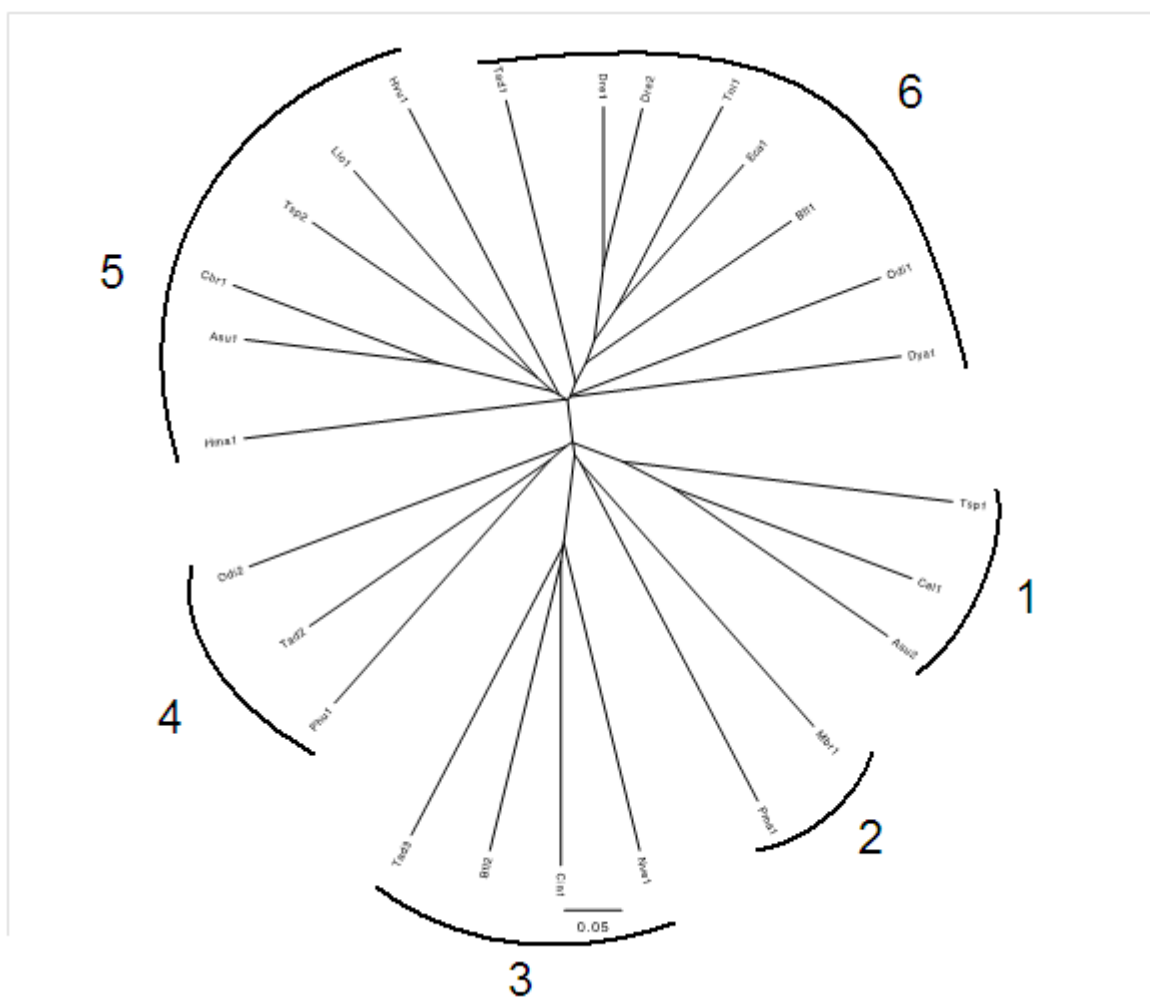
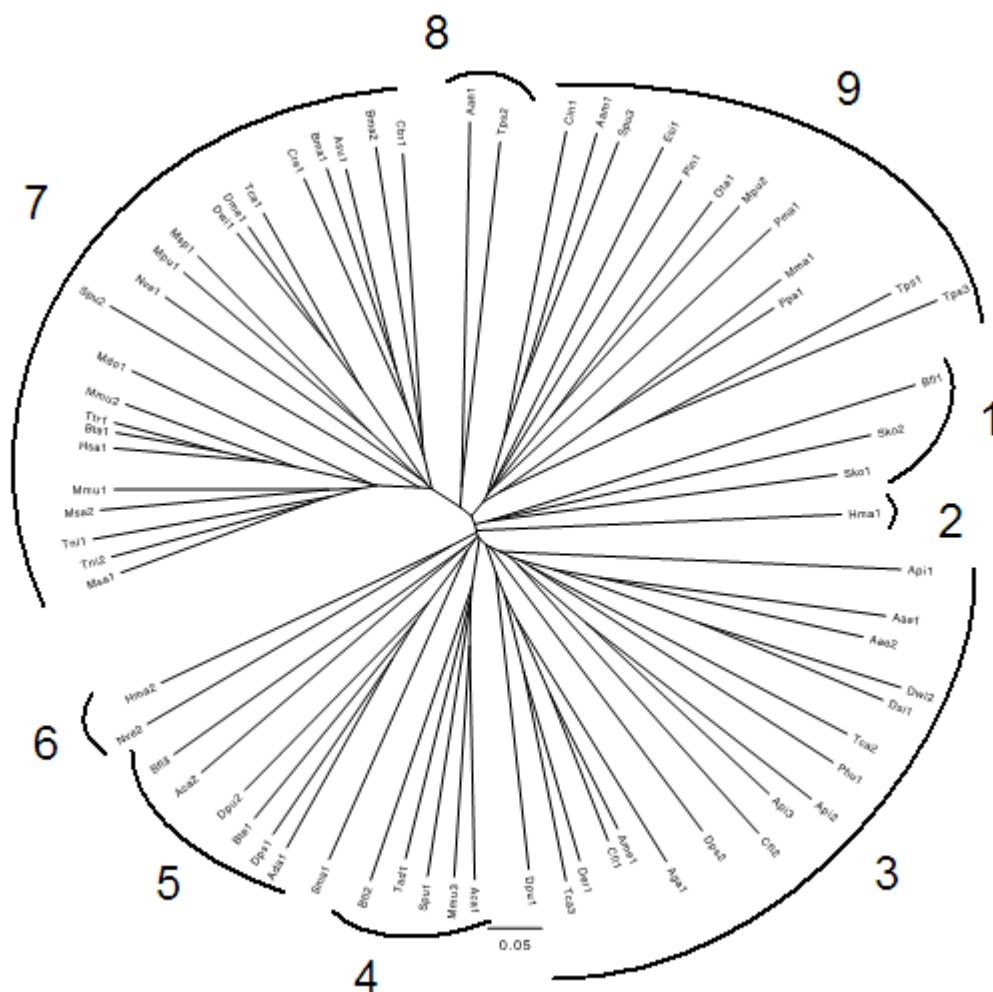


Figure 5. Protein tree for 26 homologues in Cluster 4 of the CaCA family. Cluster 4 is sub-divided into 6 sub-clusters as indicated (1-6).



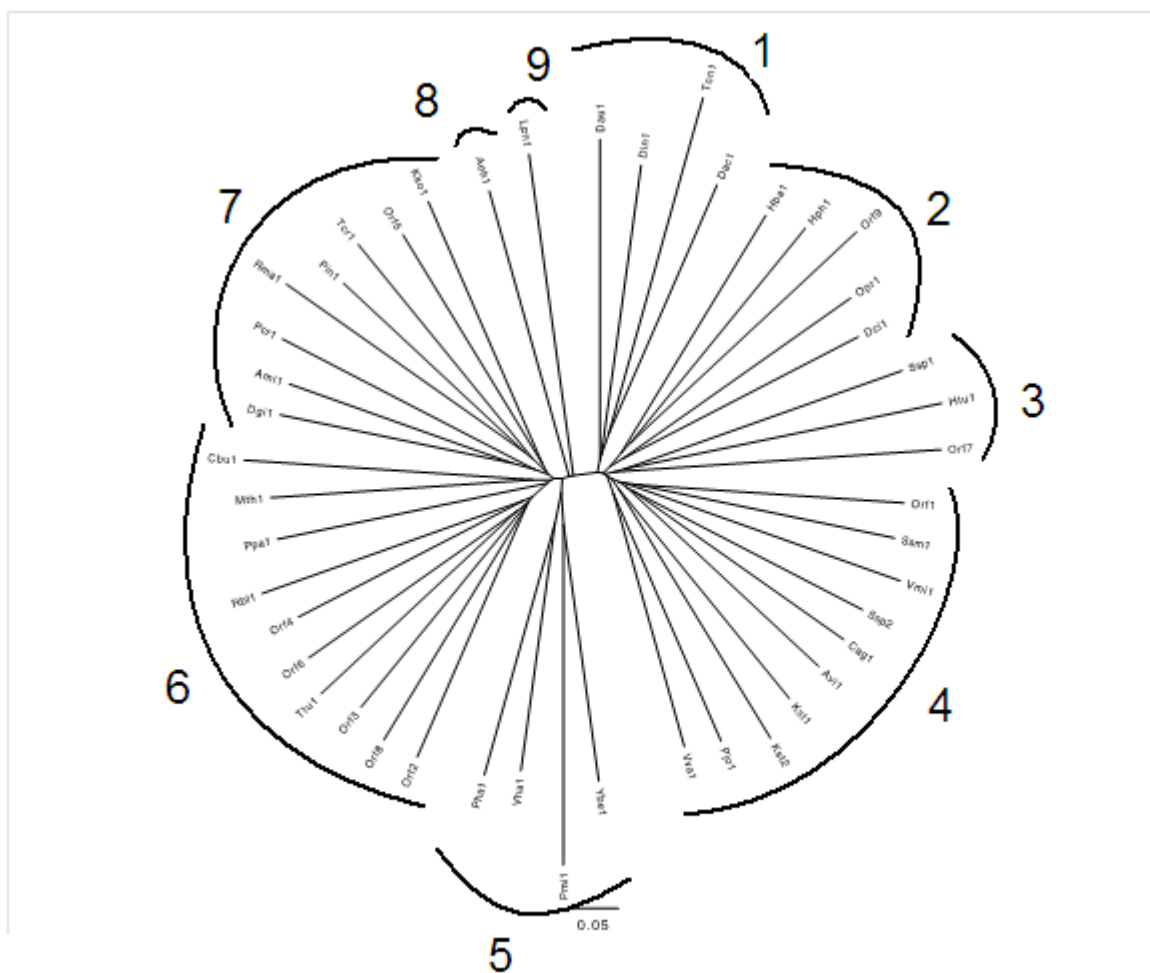


Figure 7. Protein tree for 46 homologues in Cluster 6 of the CaCA family. Cluster 6 is sub-divided into 9 sub-clusters as indicated (1-9).

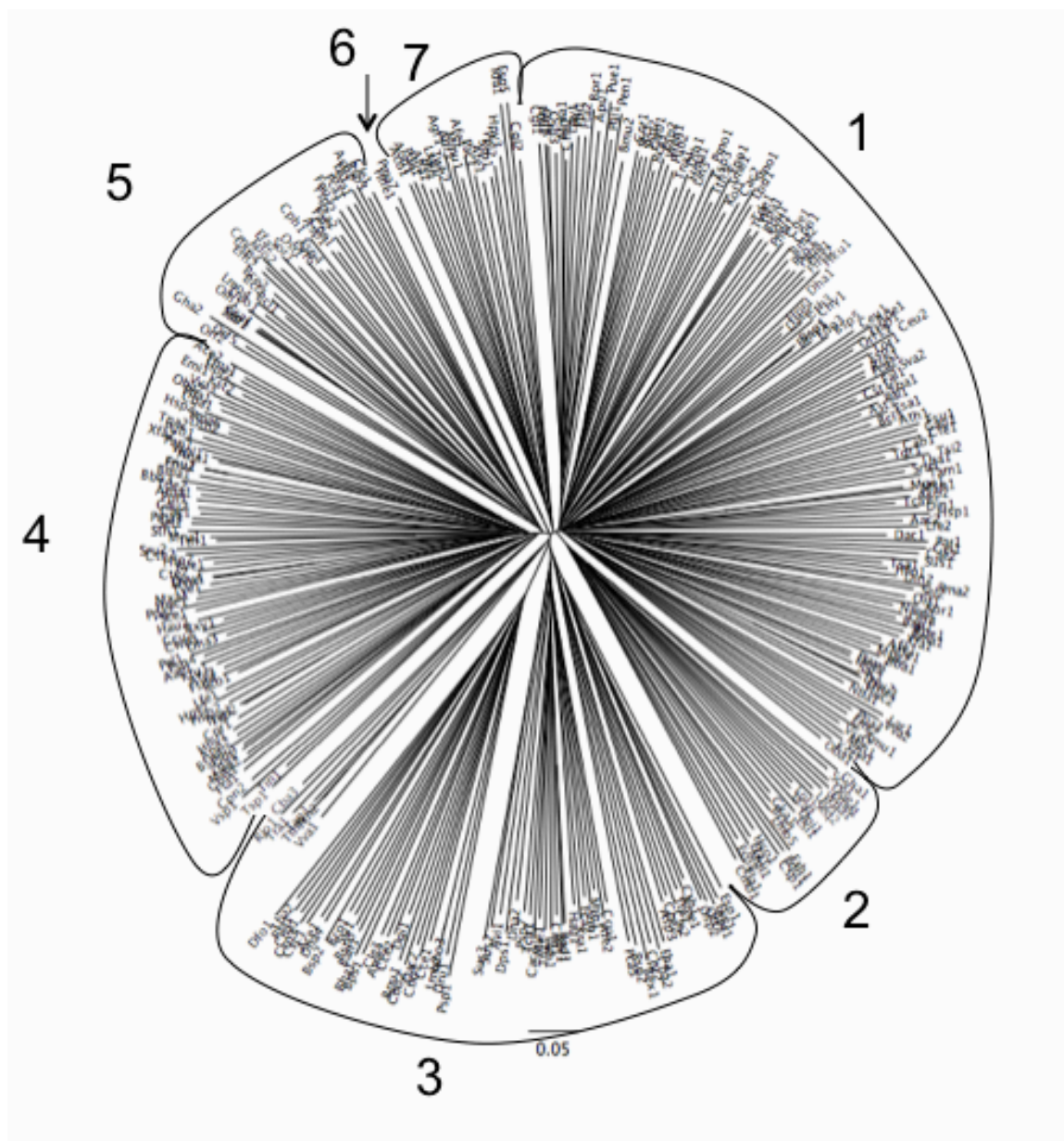


Figure 8. Protein tree for 361 MPE family homologues based on the multiple alignment generated with ClustalX, and drawn using FigTree. The Tree is divided into 7 clusters as indicated (1-7).

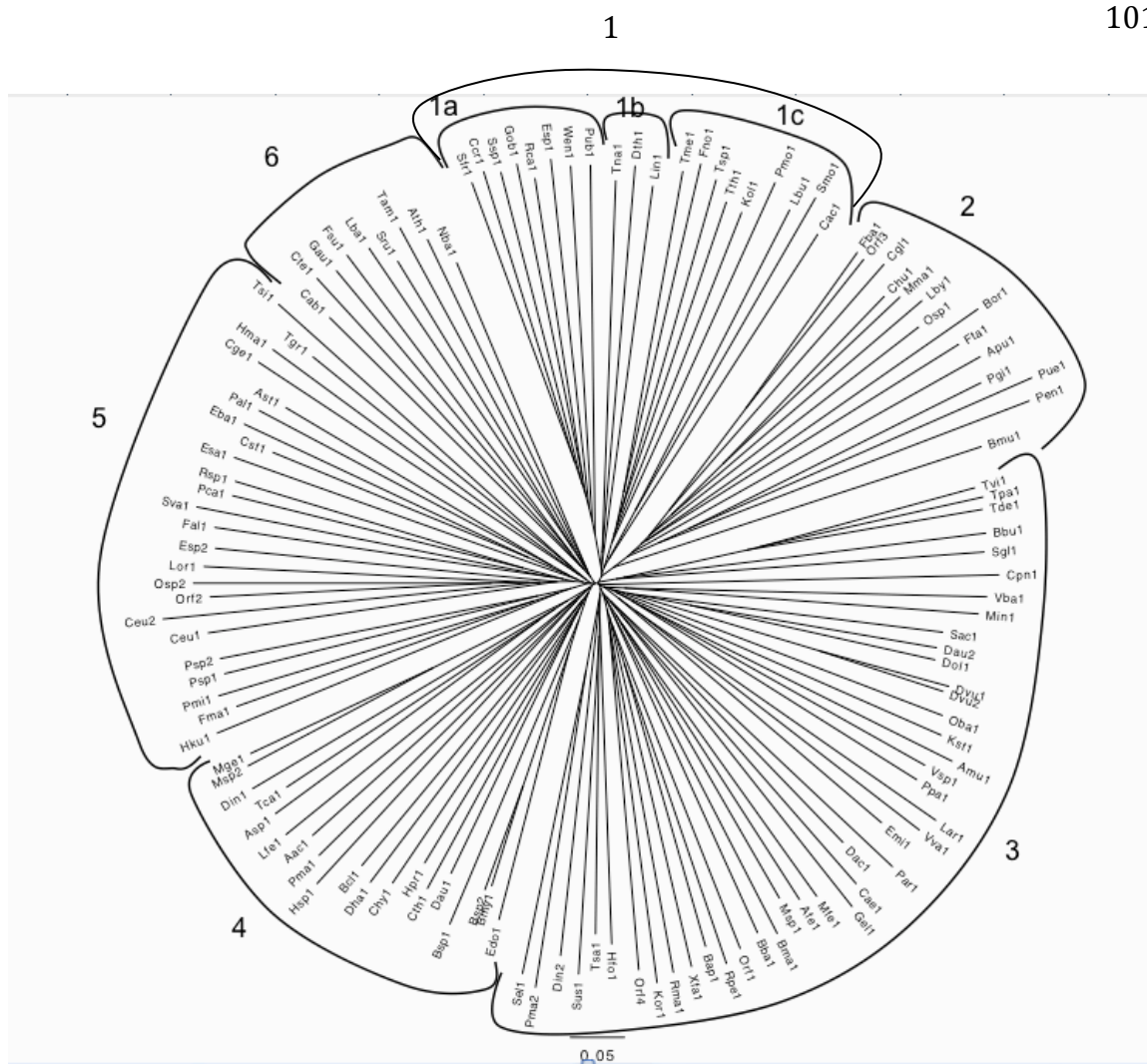


Figure 9. Protein tree for 129 homologues in Cluster 1 of the MPE family. Cluster 1 is sub-divided into 7 sub-clusters as indicated (1-7), and sub-cluster 1 is further sub-sub-divided into 3 sub-sub-clusters as indicated (1a-1c).

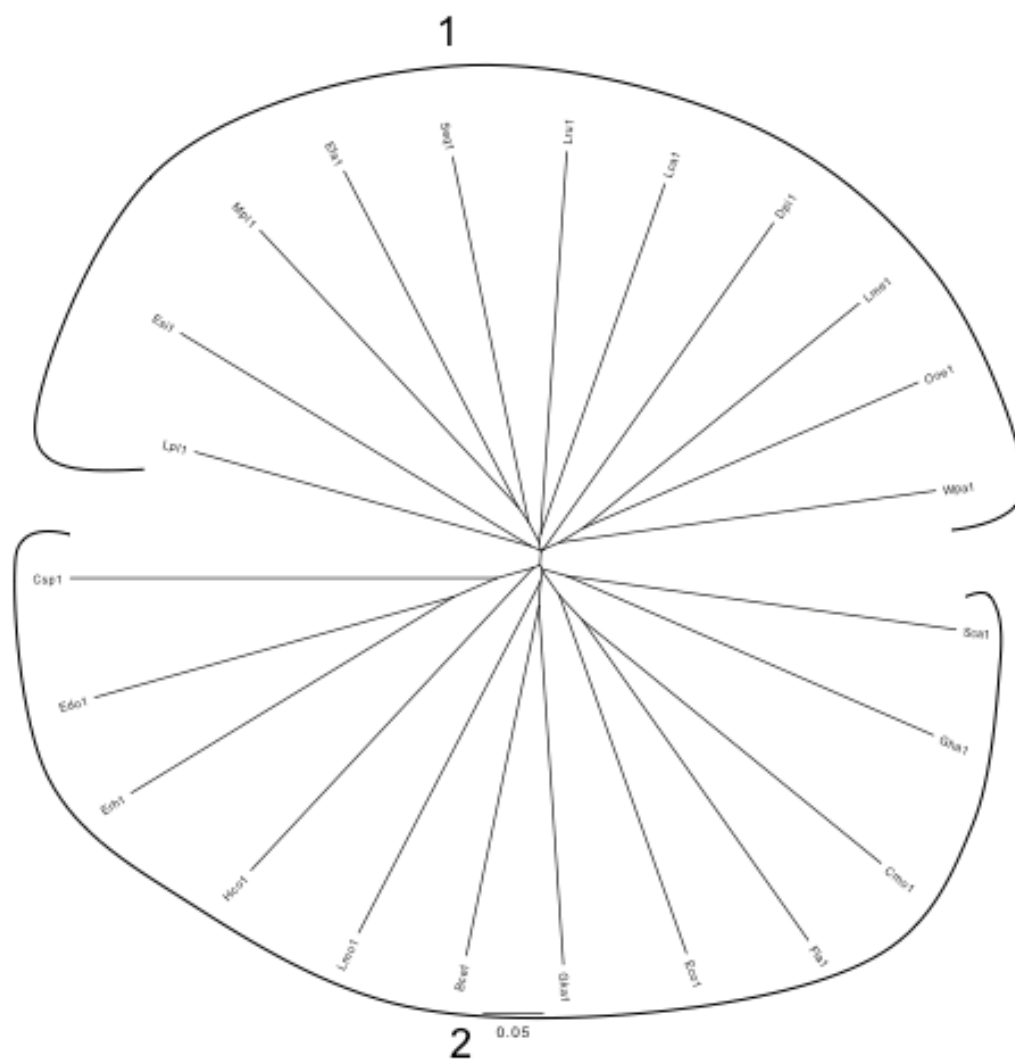


Figure 10. Protein tree for 23 homologues in Cluster 2 of the MPE family. Cluster 2 is sub-divided into 2 sub-clusters as indicated (1-2).

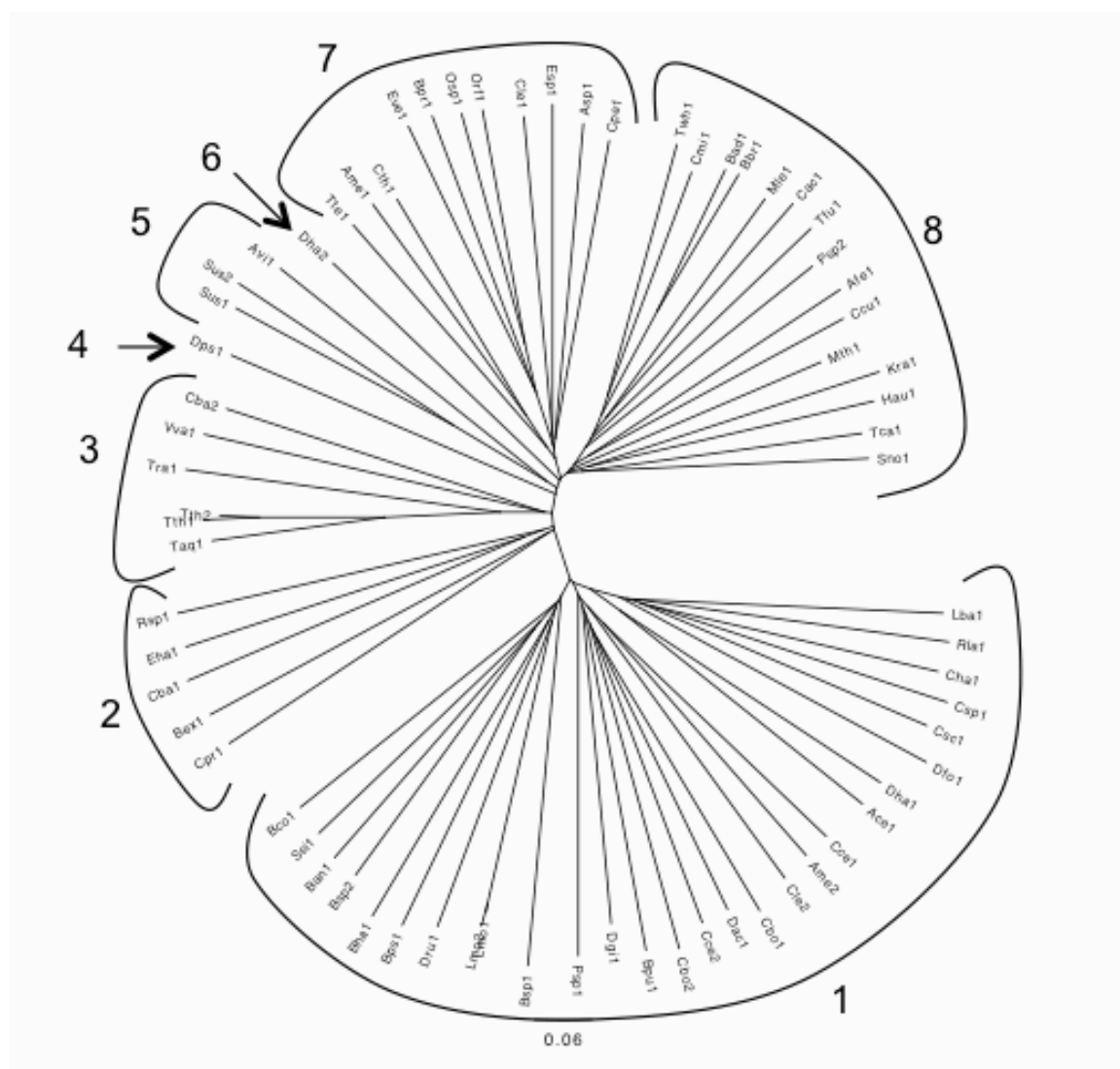


Figure 11. Protein tree for 70 homologues in Cluster 3 of the MPE family. Cluster 3 is sub-divided into 8 sub-clusters as indicated (1-8).

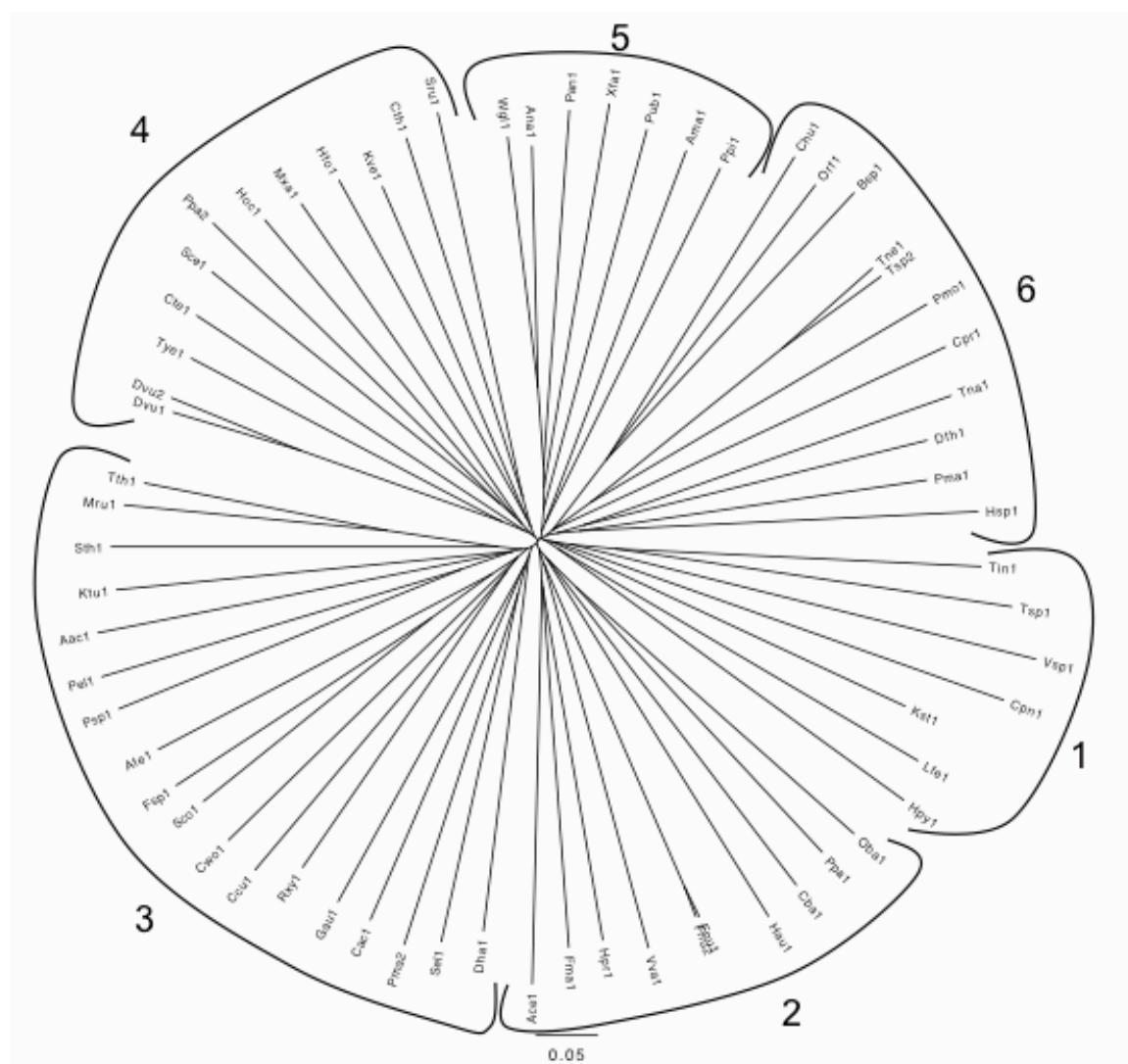


Figure 12. Protein tree for 65 homologues in Cluster 4 of the MPE family. Cluster 4 is sub-divided into 6 sub-clusters as indicated (1-6).

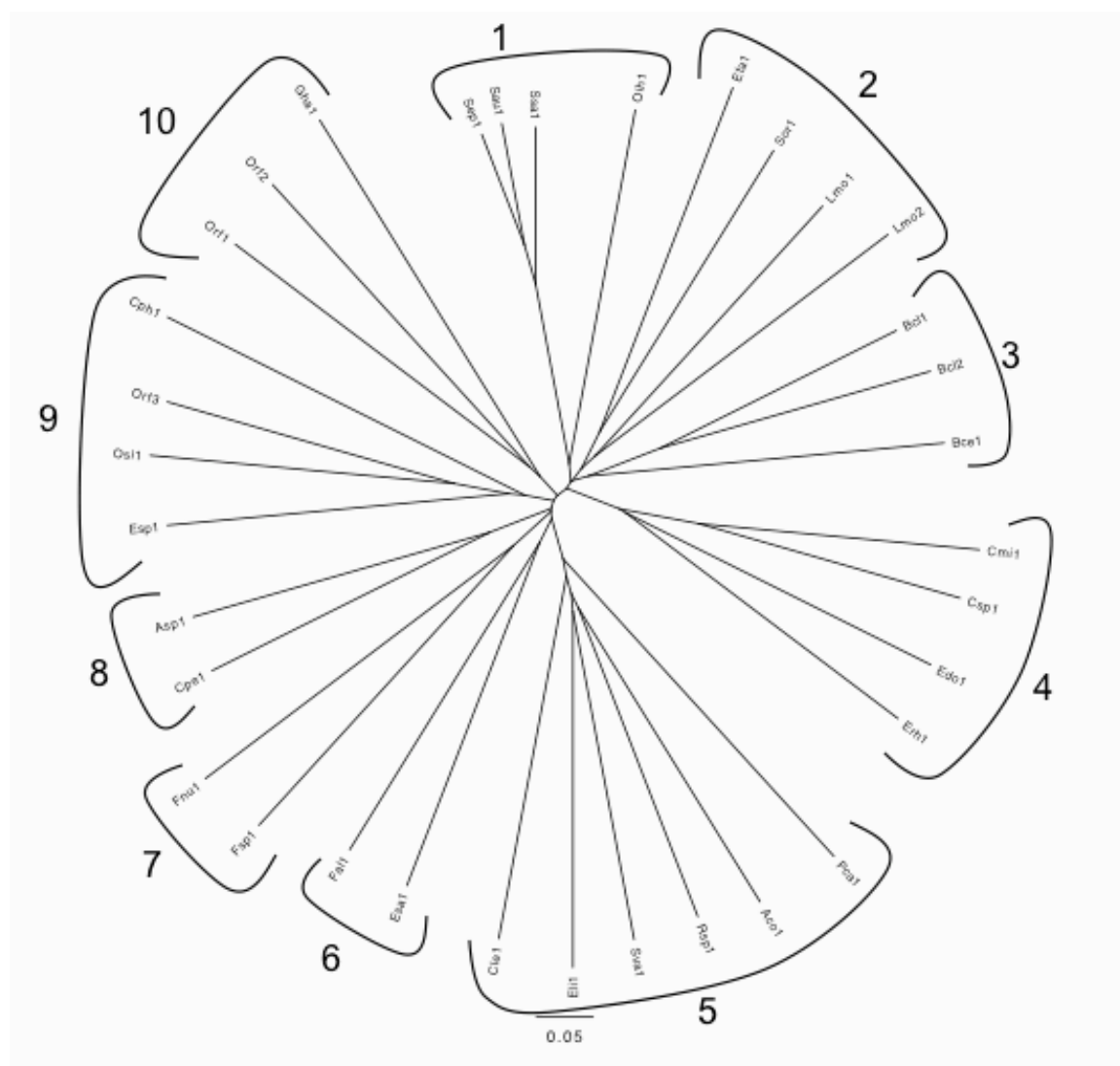


Figure 13. Protein tree for 34 homologues in Cluster 5 of the MPE family. Cluster 5 is sub-divided into 10 sub-clusters as indicated (1-10).

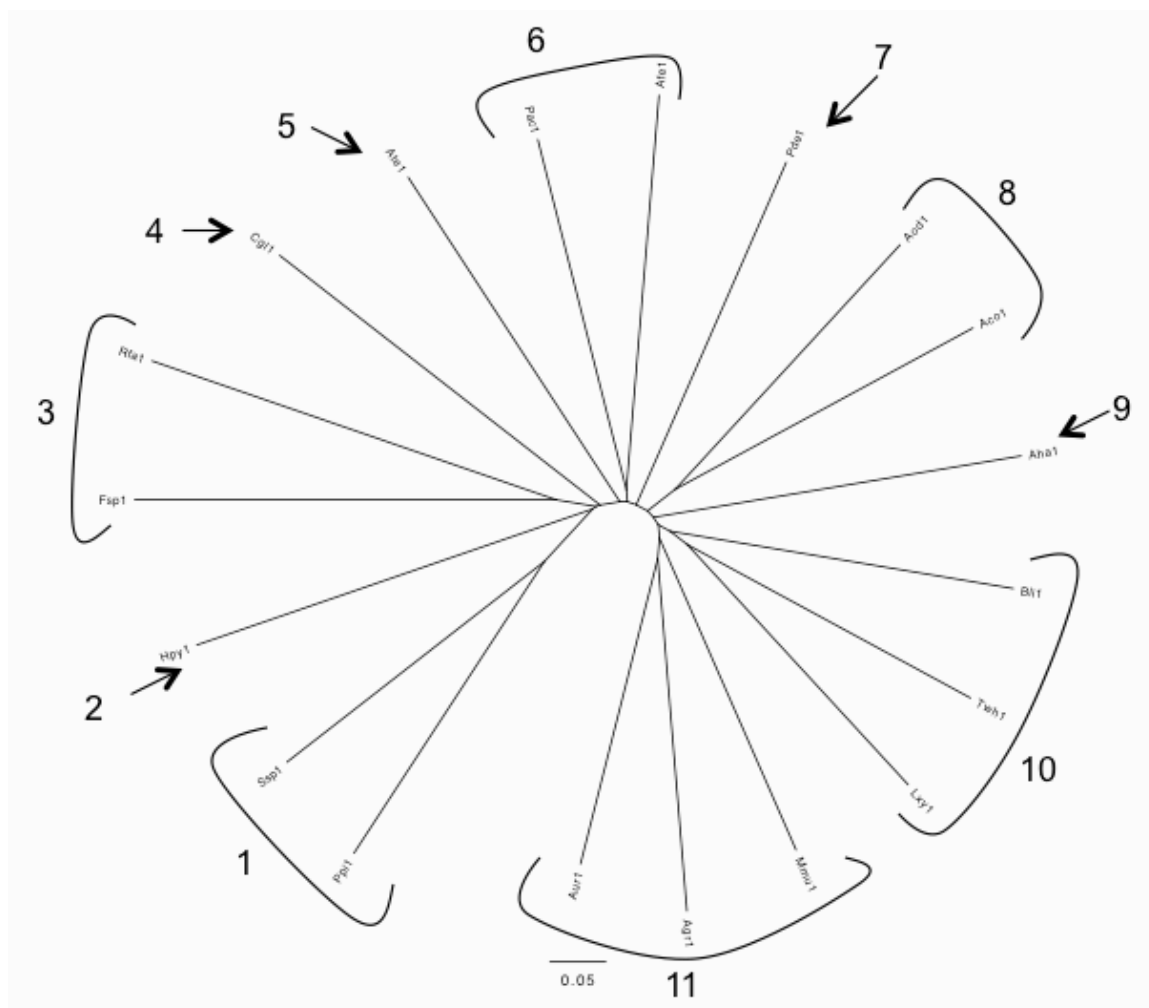


Figure 14. Protein tree for 19 homologues in Cluster 7 of the MPE family. Cluster 7 is sub-divided into 11 sub-clusters as indicated (1-11).

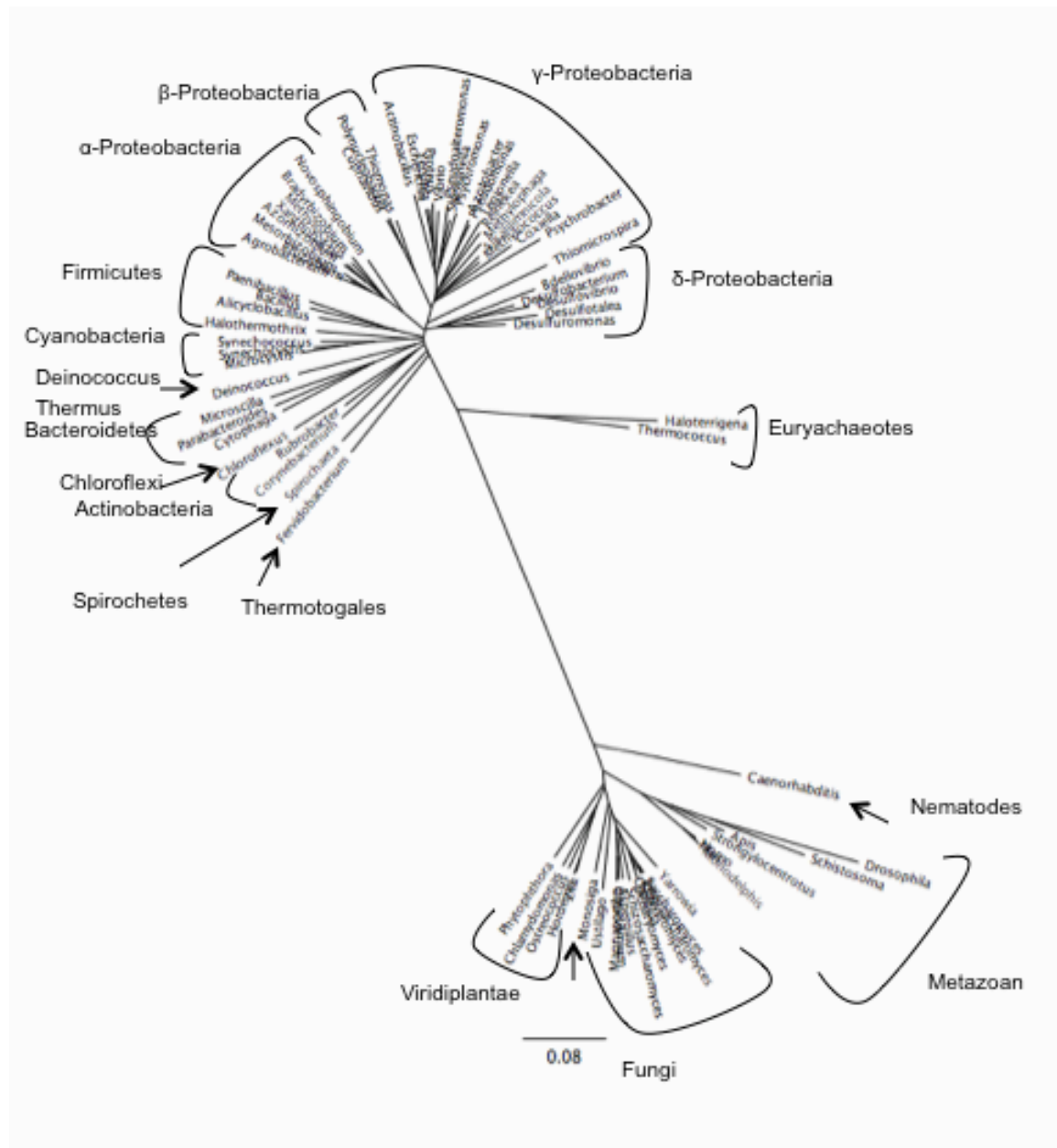


Figure 15. Phylogenetic tree of 16S rRNAs from all genera from which CaCA family homologues included in study were derived. The genus is indicated in small print inside of the pluyum brackets while the phylum/kingdom is indicated in large print outside of the phylum brackets.

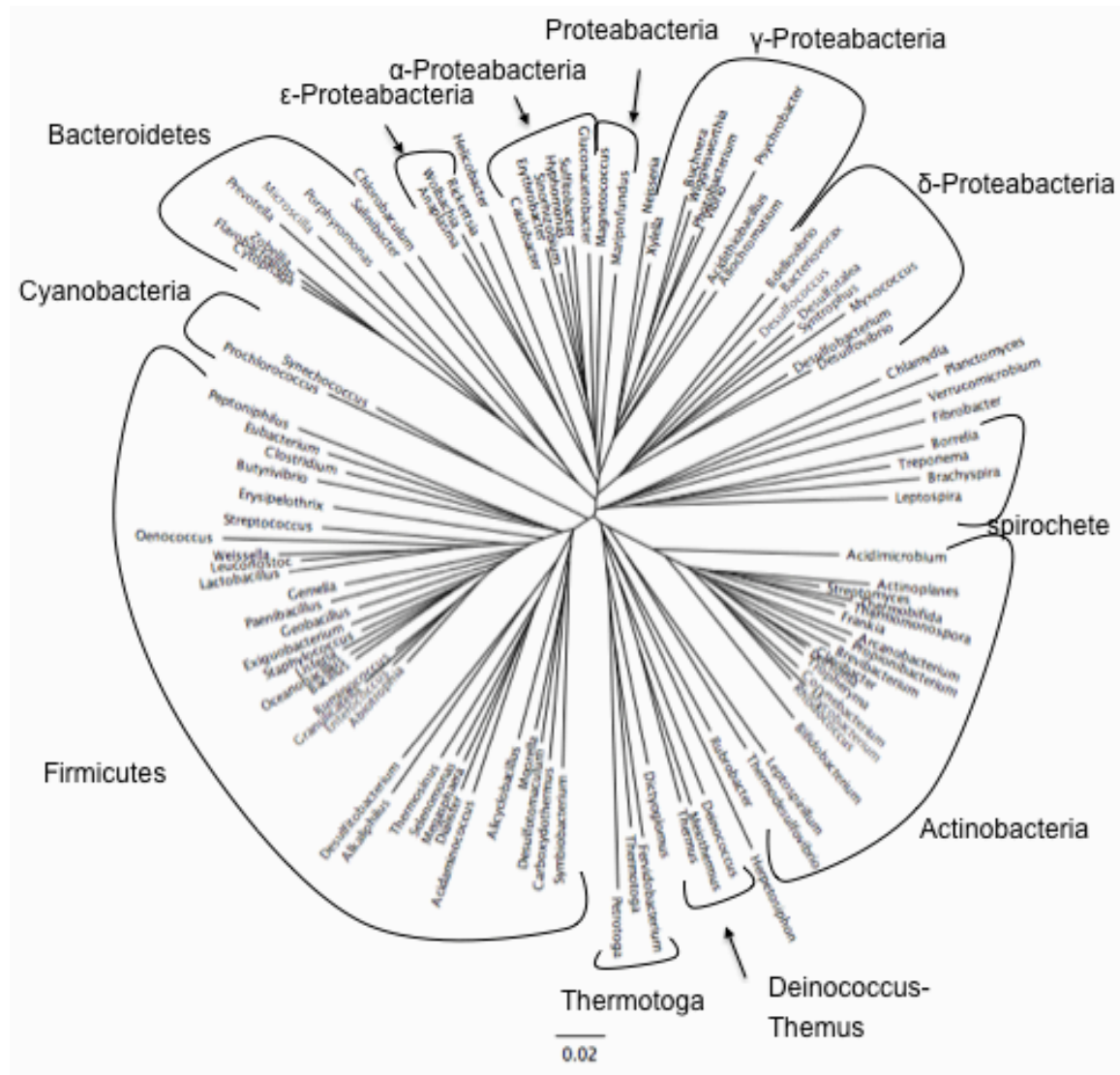


Figure 16. Phylogenetic tree of 16S rRNAs from all genera from which MPE family homologues included in study were derived. The genus is indicated in small print inside of the phylum brackets, which the phylum/kingdom is indicated in large print outside of the phylum brackets.

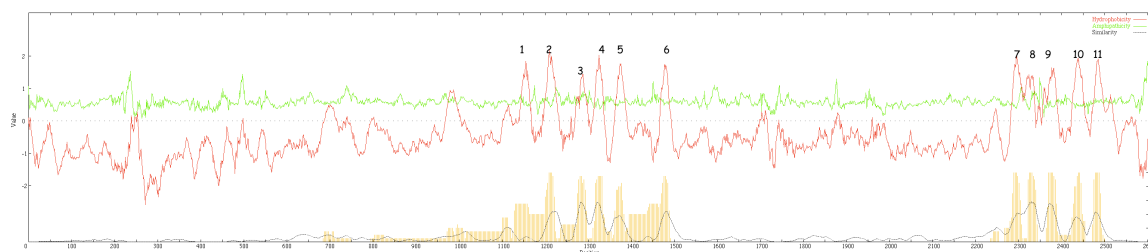


Figure 17. The AveHAS plots for all homologues of the CaCA family included in this study. The 11 hydrophobic peaks are numbered 1-11.

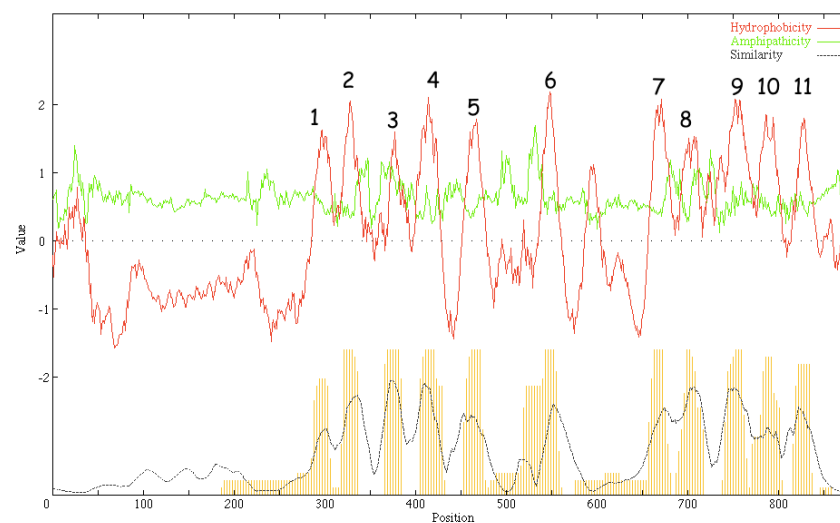


Figure 18. The AveHAS plots for homologues in Cluster 1 of the CaCA family shows 11 hydrophobic peaks are numbered 1-11.

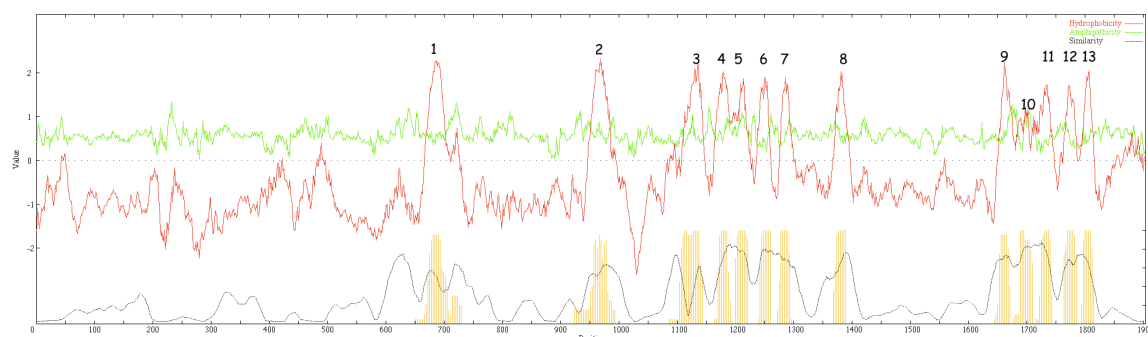


Figure 19. The AveHAS plots for homologues in Cluster 2 of the CaCA family shows 13 hydrophobic peaks as numbered 1-13.

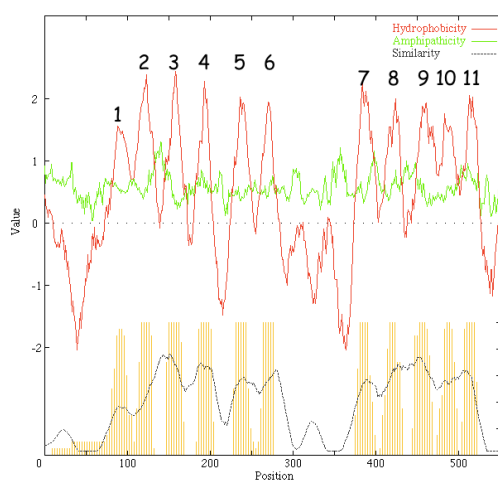


Figure 20. The AveHAS plots for homologues in Cluster 3 of the CaCA family shows 11 hydrophobic peaks as numbered 1-11.

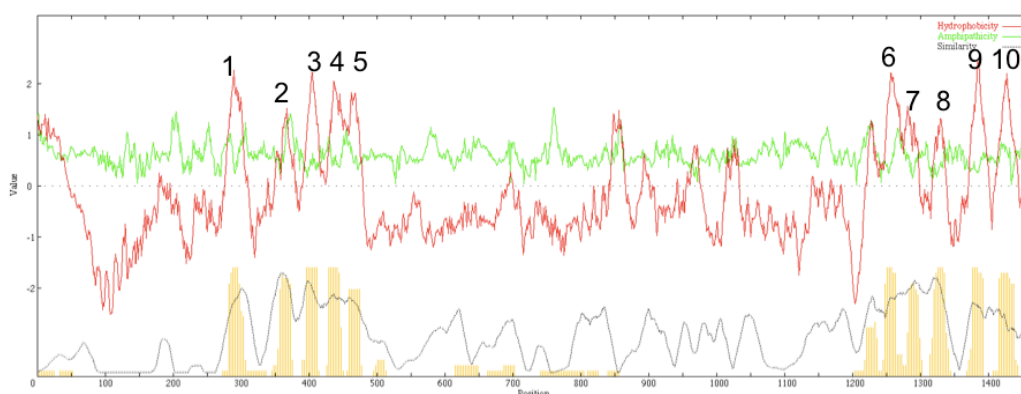


Figure 21. The AveHAS plots for homologues in Cluster 4 of the CaCA family shows 10 hydrophobic peaks as numbered 1-10.

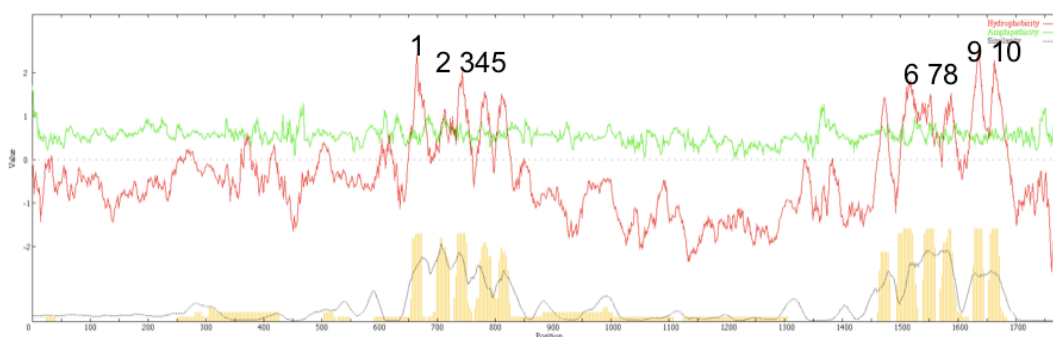


Figure 22. The AveHAS plots for homologues in Cluster 5 of the CaCA family shows 10 hydrophobic peaks as numbered 1-10.

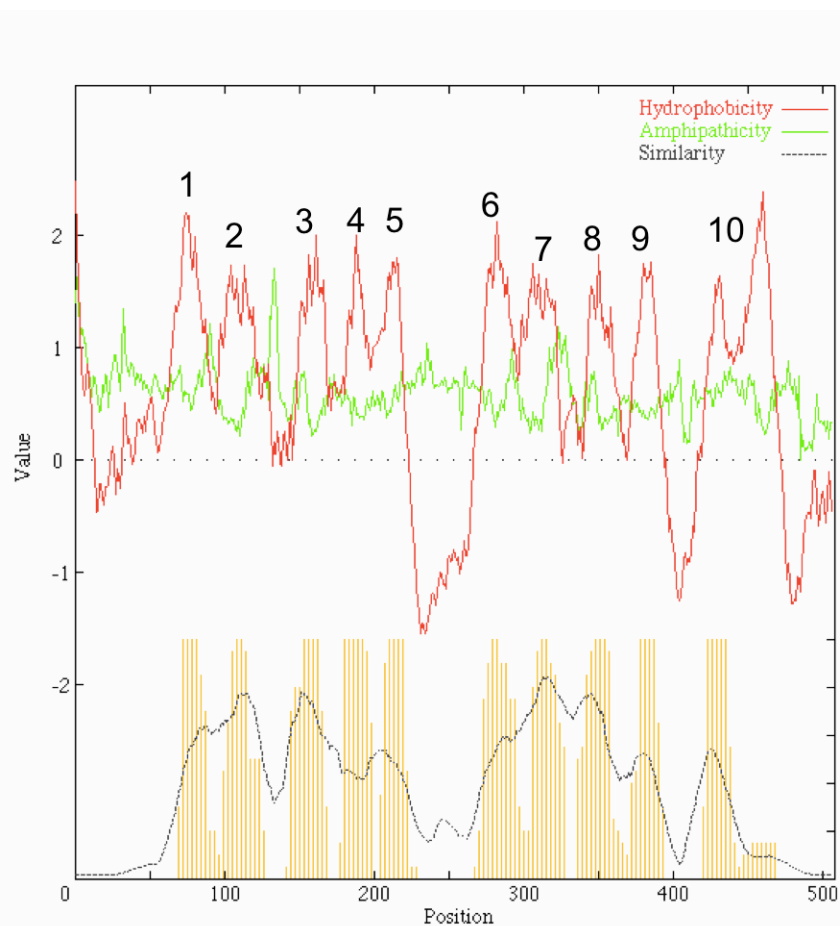


Figure 23. The AveHAS plots for homologues in Cluster 6 of the CaCA family shows 10 hydrophobic peaks as numbered 1-10.

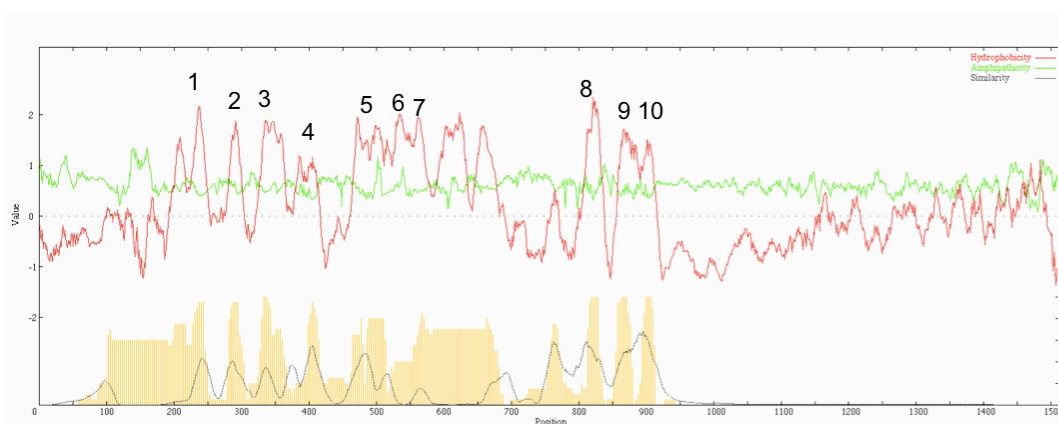


Figure 24. The AveHAS plots for all homologues of the MPE family included in this study shows 10 hydrophobic peaks as numbered 1-10.

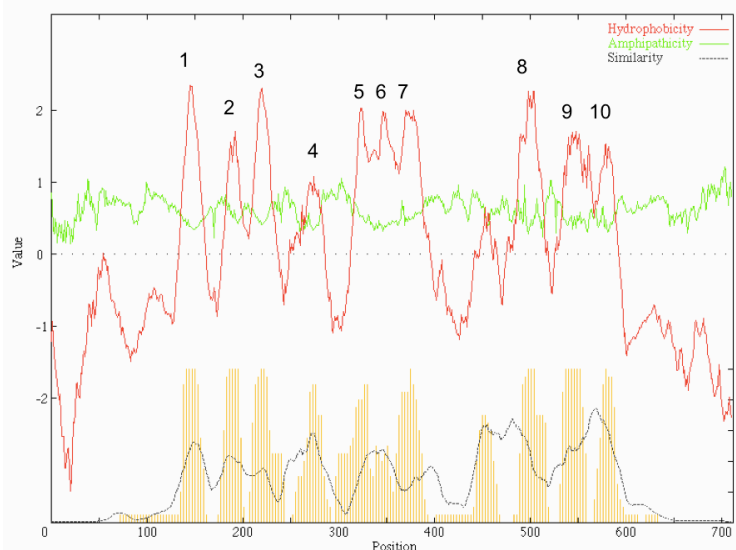


Figure 25. The AveHAS plots for homologues in Cluster 1 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

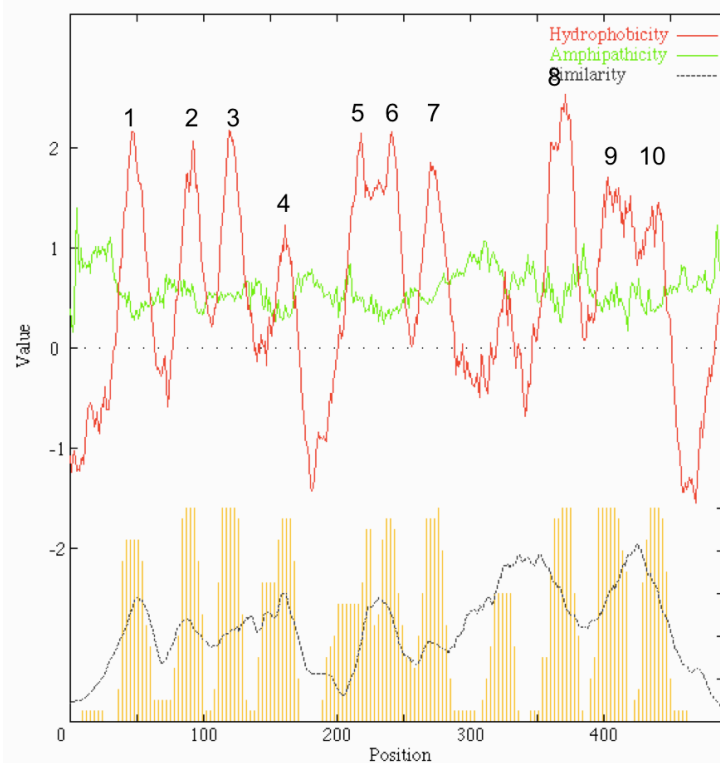


Figure 26. The AveHAS plots for homologues in Cluster 2 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

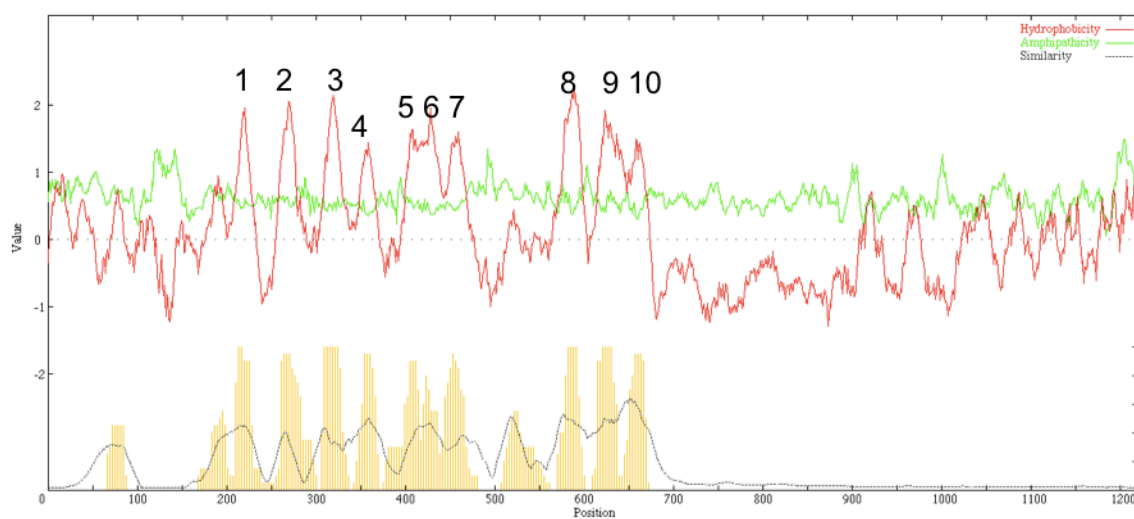


Figure 27. The AveHAS plots for homologues in Cluster 3 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

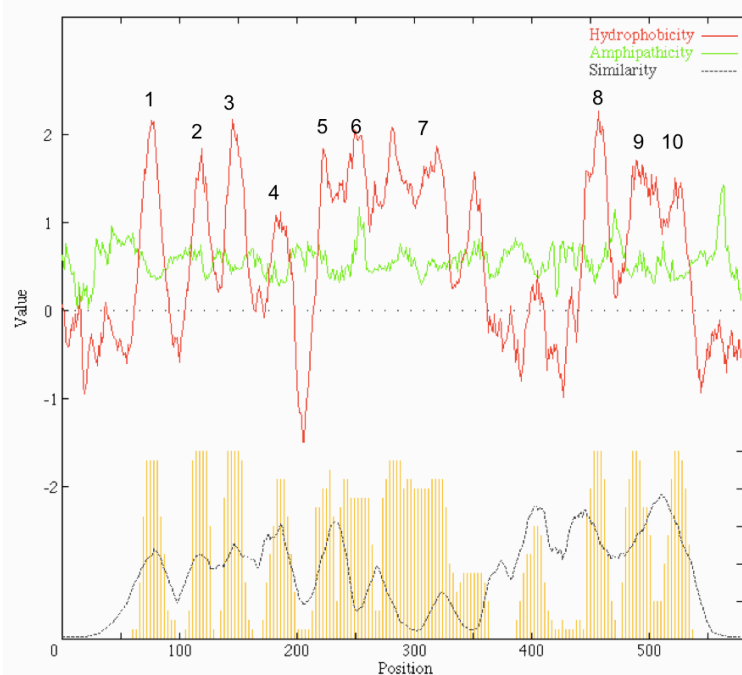


Figure 28. The AveHAS plot for homologues in Cluster 4 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

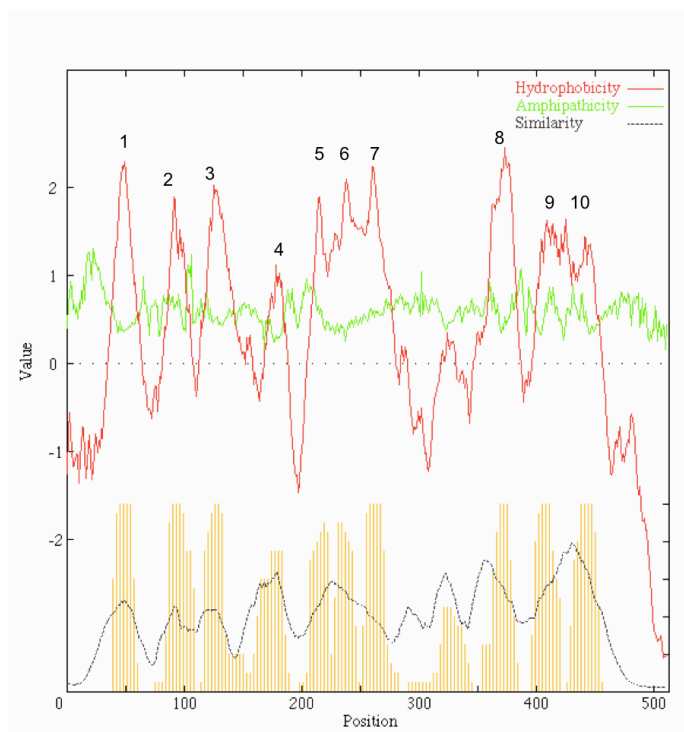


Figure 29. The AveHAS plot for homologues in Cluster 5 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

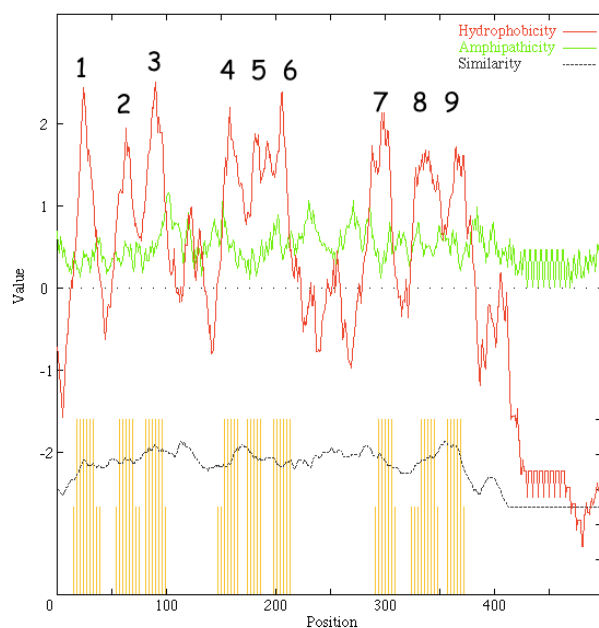


Figure 30. The AveHAS plot for homologues in Cluster 6 of the MPE family shows 9 hydrophobic peaks as numbered 1-9.

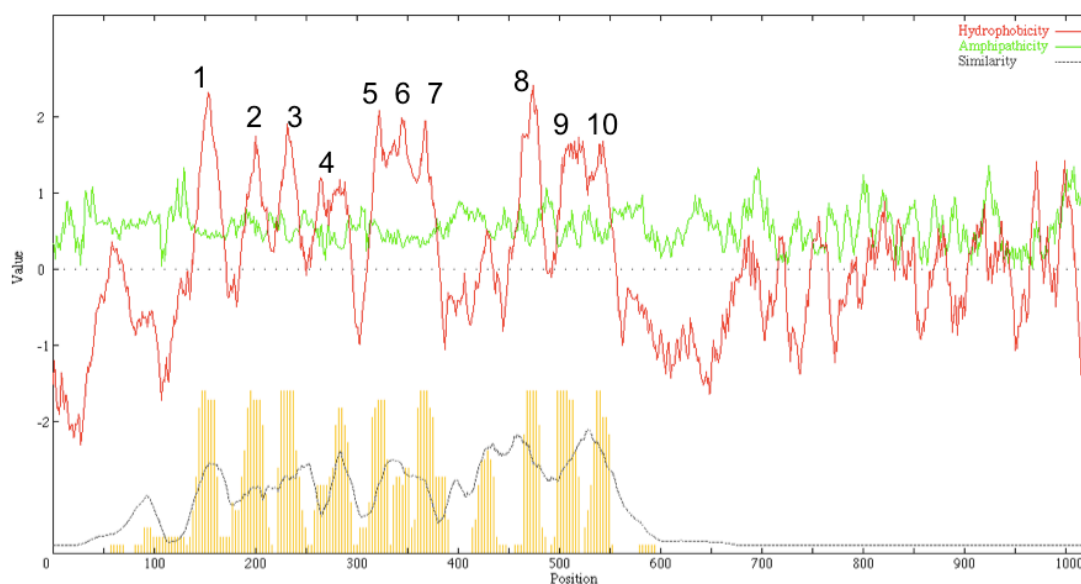


Figure 31. The AveHAS plot for homologues in Cluster 7 of the MPE family shows 10 hydrophobic peaks as numbered 1-10.

Dac1	1	-----LLFFLAAVAIFPLAALLVHATEQLASYTG--D	30
		: : : : : :	
Cre1	1	GDDDDAVPMMTVGTSIGALTIVITAIVAV--CSEFLTDSIEEVSAHTGISE	48
Dac1	31	T-IGLLNATFGNAPELIIALVALKAGLYDMVKASIIGAILANMLLGLGL	79
		: : : : : : :	
Cre1	49	TFIGLILLPIAGNACEHLTACIVAMRNKMDLSMGVAVGSSIQIALFAIPF	98
Dac1	80	AFLLGGYRRH--TQNFNPVAARNFMTMMLLAVISLAVPSTFH-NFVTADA	126
		: : : : : : :	
Cre1	99	AVVVGWATGHDFSLDFSPFSA-----MVLAV-----CVIHANFVTADA	136
Dac1	127	IQHEQYVNLAVAVV--LLATYGLSLFFMLQTHPDFFKS	162
		: : : : : : :	
Cre1	137	TSH--WL-LGVQLIALYCVIAVAYFYR-----	160

Figure 32. TMSs 2-6 of Dac1 (gi: 328954050) aligned with TMSs 7-11 of Cre1 (gi: 212374355), both of the CaCA family, with 23.0 % identity, 39.6% similarity and a comparison score of 12 S.D.

Mla1	351	KAKVLLCTYKALGIQYYKYTV	DGVNIIIFIILMPFVIFVIFD-AFFLVPLH	399	
			: : : :		
Dac1	1	-----	MKILQKIFYGLTLFFIPLT	19	
Mla1	400	-LTGFLGVISSQAVIFVMALLSVVPLSYFIGMAVASISAQSSIGMGAVIN		448	
		: : : : : : : : : : : : : : : :			
Dac1	20	IILDRIGGV-PQPLFFLAAVAIFPLAALLVHATEQLASYTGD	TIGGLN	68	
Mla1	449	ATFGSIIEILLYGIALTEGKDLVEGSIIGSVLAGVLLMPGMSMIGGSVR		498	
		: : : : : : : : : : : :			
Dac1	69	ATFGNAPELIIALVALKAGLYDMVKASII	GAILANMLLGLGLAFLLGGR	118	
Mla1	499	RKEQKFNARSAGVTSTMLIMAIIGVLTPTLFYEIYGT	FQLTCSGCPDGVS	548	
		: : : : : : : :			
Dac1	119	RHTQNFNPVAARNFMTMMLLAVISLAVPSTFHN	F-----	152	
Mla1	549	TKPGEINTEALWKCQSCRYEHLSPVNDEFYQTSVKGLSYYCAALLVLSYL		598	
		: : : : : : : : : : : :			
Dac1	153	-----VTADAI-----QH-----EQYVNLA	V-----AVVLLATYG	177	
Mla1	599	IGLWFSRLRTHASQIWQTPAAAPHTADHEPGGGSGHDAPNWSRNKSTTVLL		648	
		: : : : : : : : : : :			
Dac1	178	LSLFFMLQTH-PDFFKSSA-----EDVEP-----PDEV	RWSLPRAVAILL	216	
Mla1	649	SCTLLYAVIAEILVSVDVLDGSGIPEKLLGVTLFALVPNTTEFMNAIS		698	
		: : : : : : : : : : : : :			
Dac1	217	LTSVFVAFMSEILVGAVEETGR	TLGMSQAFIGLVILAVVGGA	ESSAVA	266
Mla1	699	FAMNGNIALSMEIGSAYALQVCLIQIPAMIAISAIYNSGKYISAHHTFTL		748	
		: : : : : : : : : : :			
Dac1	267	MGMKNRMDLSLSIAIGSSI	QIALFVAPVLVLM	S-----YFIAPQPLNL	309
Mla1	749	VFPRWDVISIVF-SVFLTTYTYIEARANYHRGTILVLSYLVLVGGFY	SAP	797	
		: : : : : : : : : : : :			
Dac1	310	VFNR-VLLGIIFLTVLIGAMVAGDGKANWFKGVQLITVYIILAIMFY	FLP	358	
Mla1	798	VGGSVTEGN	806		
Dac1	359	--GS-----	360		

Figure 33. TMSs 3-13 of Mla1 (gi: 328858470) aligned with TMSs 1-11 of Dac1 (gi: 328954050), both of the CaCA family, with 13.0 % identity, 25.0% similarity and a comparison score of 31 S.D.


```

Nsa2  1  -----QLASVLLGI-GIMVTLWLDPDRWFVRLGFAIFLLSLLAMILM      42
          |:....|:|: ...:::|...|...|. |:...|..|:....
Mtu2  1  RHPSANQQMLWTLVGVAALVVTFLKDHRLARYGYICGLAGLVFLAVP      50
          |:....|:|: ...:::|...|...|. |:...|..|:....

Nsa2 43  PFLPASLAKEVLGAKRWIRLG      63
          ..|||:|: |..|||.|||
Mtu2 51  ALLPAALS-EQNGAKIWIRL-      69
          ..|||:|: |..|||.|||

```

Figure 36. TMSs 2-3 of Nsa2 (gi: 319957313) aligned with TMSs 4-5 of Mtu2 (gi: 289760114), both of the MPE family, with 39.6% identity, 47.9% similarity and a comparison score of 11S.D.(See Table 22)

```

3-----4-----5-----
PFGTLVLAIAVTIIEVSLIVSLMLSGSGDTATLARDTVFAAIMIIMNGIVGICLLVGG..VRHHEQ.RFVLKGVSSALCVLAAMASLSLVPNF'

100%=====39%=====68%=====
3-----4-----5-----
KFSPMLLLIVGIMLVAVLLVGTTVNGAKRWLTIGP...IRIQVAELAKFAFAIYMSGYLVRRHEEIRENAKGFIKPIVVFAVYAVLILLQPDLL

8-----9-----10-----
AVVGAGLPLAVGVLIAALVLAPELSAIIKAARRNRLQTSNLALGSALA.TIGLTIPAVAVLSVIMGWPLSLGLDTKSMVLLLSFVVSIFSLGTGRTTVLQ

94%=====54%=====53%=====
8-----9-----10-----
PIFAVIGEELGFVGIIVVLSVLLFVSLRAI...RLGNLCIAIDKAFEGYLAYSIGIWNICFQTVVNV..GASIGM.LPTKGLTLPFISYGGSSLWVMTAAAMILI

```

Figure 38. TMSs 3-5 and 8-10 of Aex2 of the CaCA family (gi: 315498735; 10 TMSs) aligned with TMSs 3-5 and 8-10 of Shbe1 of the MPE family (gi: 163751823; 10 TMSs), respectively with a comparison score of 12 S.D.

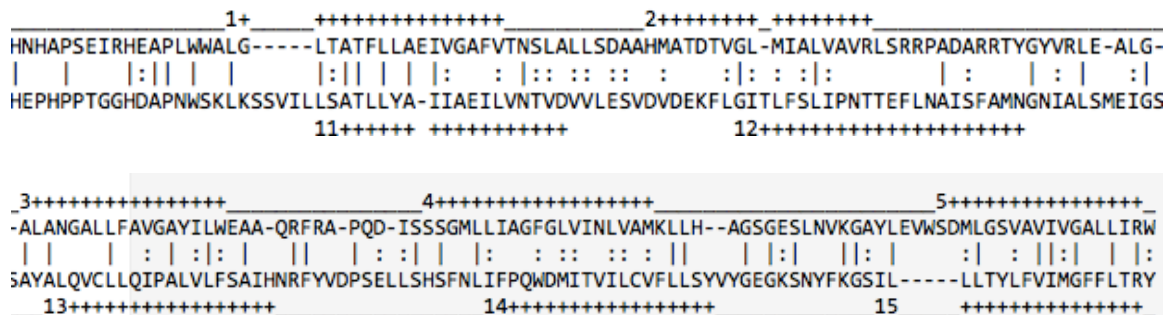


Figure 39. TMSs 1-5 of a CDF family protein (gi: 21242070; 6 TMSs) aligned with TMSs 11-15 of a CaCA family protein (gi: 320036119; 15 TMSs) with a comparison score of 11 S.D.

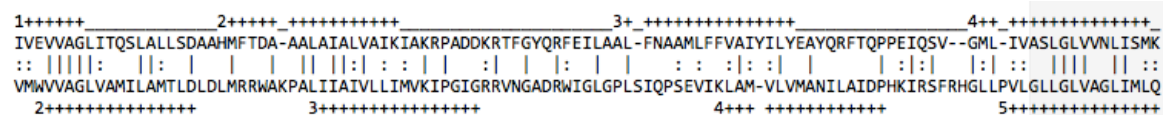


Figure 40. TMSs 1-4 of Asp14 of the CDF family (gi: 359428766; 6 TMSs) aligned with TMSs 2-5 of Dha10 of the MPE family (gi: 361855429; 10 TMSs) with a comparison score of 13 S.D.

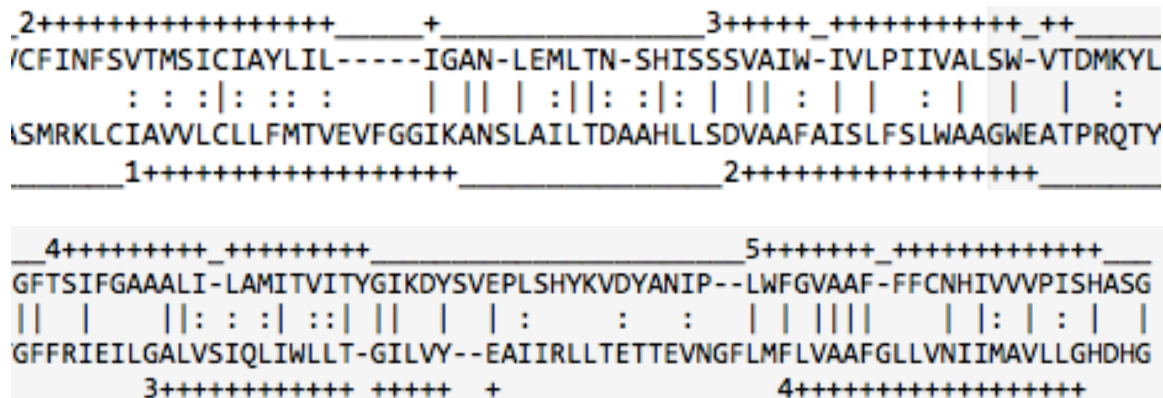


Figure 41. TMSs 2-5 of Dfa1 of the AAP family (gi: 328875879; 11 TMSs) aligned with TMSs 1-4 of Bju4 of the CDF family (gi: 32274627; 6 TMSs) with a comparison score of 13 S.D.

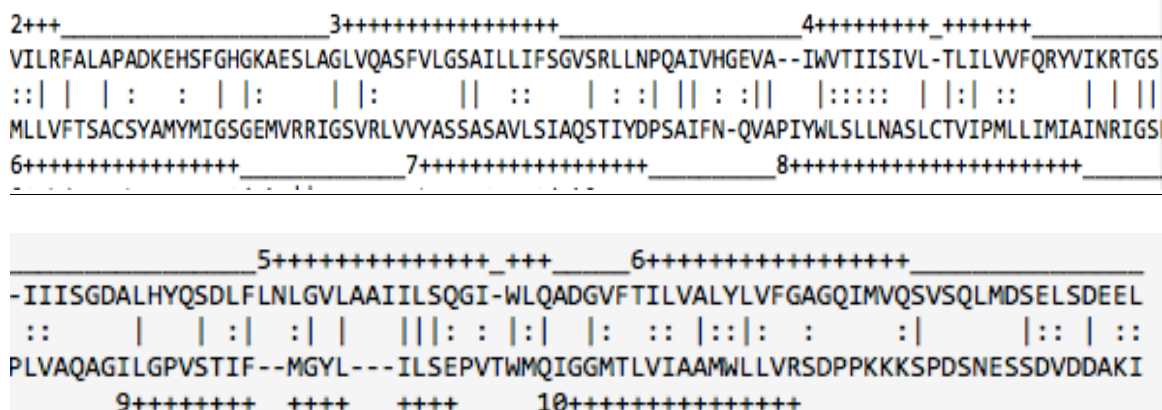


Figure 42. TMSs 3-6 of Cps1 of the CDF family (gi: 71279891; 6 TMSs) aligned with TMSs 7-10 of Pne1 of the DMT family (gi: 145589545; 10 TMSs) with a comparison score of 11 S.D.

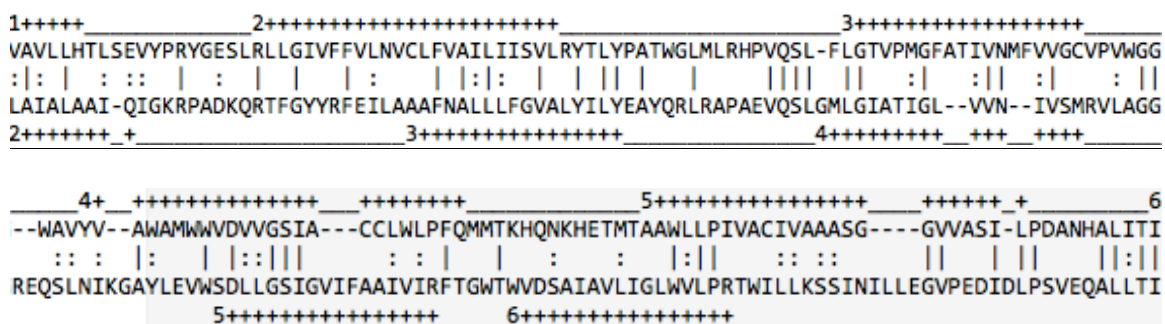


Figure 43. TMSs 1-5 of Lsm1 of the TDT family (gi: 312216773; 10 TMSs) aligned with TMSs 2-6 of Rlt1 of the CDF family (gi: 309778699; 6 TMSs) with a comparison score of 11 S.D.

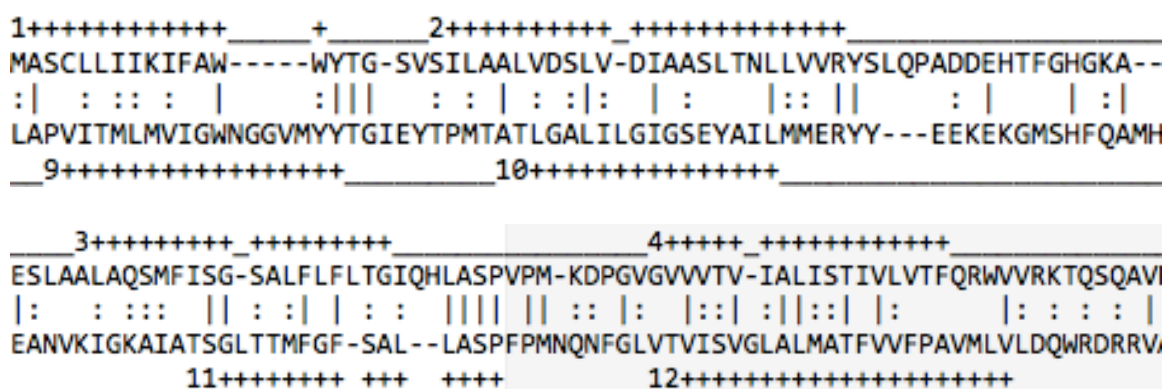


Figure 44. TMSs 1-4 of Etb1 of the CDF family (gi: 146313689; 5 TMSs) aligned with TMSs 9-12 of Msz1 of the RND family (gi: 336477699; 12 TMSs) with a comparison score of 12 S.D.

```

Min2  1  -----GVIVGTIVMSLGAAGF---FSTL-IWQHLGDFQLHWVATEFA      38
      | : | : : | | | | : | | | :
Afu1  1  VVRTIERTGKAITTSALTMAGGFGSLMFSTFPIMQNFGFIAL--VAILFS      48
      | : | : : | | | | : | | | :

Min2  39  LIVLLAVGSDYNLMLVSRIEEEIGAG      64
      | : | : : | | | | : | | | :
Afu1  49  LLAALTVVPAF-LMLTEGIEARISAA      73
      | : | : : | | | | : | | | :

```

Figure 45. TMSs 9-10 of Min2 (gi: 254821662) aligned with TMSs 11-12 of Afu1 (gi: 11498828), both of the RND family with 31.6% identity, 42.1% similarity and a comparison score of 10 S.D.

```

      3+++++ _+++++ _4+ _+++++ _+++++
:VITDKLQISDDVAGATFMAAGGSAPEL-FTSLIGVFISHSNVGIG-TIVGSAV-FNILFVIGTCALFS---
: : | : : : | : | | : | : : : : : | : | : : | : : | : : |
-MVLGLLPV---LSG---VAAGVAAVSLTFGAVHGL-----TLGFGTTLIGEAVDYSIYLFVQSARLHGATP
      3++++ _++ _+++++ _4+++++
+ _5+++++ _6+++++
REILN---LTWVPLFRDVSFYILDIMLILFFLDLIAWWSLLLLLAYAFYVFTMKWNKWLEVVWKEQL
: | : | : | : : : | : : : | : : : : : : : : : :
HDSLRAWIATYWPTIR-----LGVLTSCVGFASMLFSGFPGVLVQLGLYS--IVGLVTAALVTRYVLP
      5+++++ _+++++ _6+++++ _+++++

```

Figure 46. TMSs 3-6 of a CaCA family protein (gi: 281347334; 11 TMSs) aligning with TMSs 3-6 of an RND family protein (gi: 187927409; 12 TMSs) with a comparison score of 12 S.D.

```

      2++++ _+++++ _3+++++
QMVYLTWGP--GWWLVVSWMITLYTLWQWVEMHEMVEGKRFDRYHELQGEAFGHDLGWLWVWPQQLIVEVGVDIWVMVTGG
: : : | | : : : : : | : | : | : : : : : : : : : : :
QVLGVTWGRHIGTVWVLAWVWLVGAVLAHVHHAEEVVA-----HKVG-EPFG-SLILAWAV---TVIEVGLIVTLMISGG
1++++ _2+++++ _3+++++

      4+++++ _+++++ _5+++++
TSLQNFYKLVCSGNCFMAHHTSAWIAIF-SSVHFVLAQLPNFNSIAGVSLAAAIMSLSYSTIAWAI PASYGHSTPLVGPVNYKLPVQ
: : : : : : : : : : : : : : : : : : : : : : : : : :
PATASLARDTVFAAVMITVNGIAGLSLVGSRRYGVAL---FNA-EGTGAALATVA-SLATLSLVLP-TFTTSRP--GP-EFS-SAQ
      4+++++ _5+++++

      6+++++ _7+++++ _8+++++
SVSAHVFNALGTVAFAVAG-H-NWLEIQATIPSTKERPSKIPWNR--GWWLAYIIVAIYFPVALIGYWAYGNQVTDNIGVYGRPGVWVAMNLMVWV
: | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
LIFAAVASLV-LYGMFVLTQTGPHFFLPVIADPETGDRSSAPPTGPMAGVSLALLWVALC----AVVGLAKVESPAIEDAVEDAGLPQSIWVWIALLV
6+++++ ++++++ 7+++++ ++++++ 8+++++

```

Figure 47. TMSs 2-8 of Ppa10 of the CaCA family (gi: 168052695; 10 TMSs) aligned with TMSs 2-8 of Rop1 of the AAAP family (gi: 167999963; 11 TMSs) with a comparison score of 13 S.D.

[illegible]

Figure 48. TMSs 6-10 of GlcU of the DMT family (gi: 73661966; 10 TMSs) aligned with TMSs 6-10 of Ptc1 of the CaCA family (gi: 224131110; 10 TMSs) with a comparison score of 11 S.D.

```

7+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
MFICVG--PPAFTALALVGMAGLPEEF---KIIKDAHAIEDARILELM-AIIVGVFLWALSL-WFFLIAFVAVVRSRP
: || | : : : : : | : | || | | : : | || | | : : | : |
LLSCVALIEPSVAGITIIALGTSVPDTFASRTAAIQDQHA--DAAIGNITGSNSVNVL-GLGLPWIIVMYLYVNSDQP
7+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
8+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++

9+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
---TAFHLSWWAMVFPNTGFTLATITLGRALGSPGVLGVGSAMS---VGI-VC-MW-VFVFVYHIRA
: | : : | | : || | || | : : | : : | : | | : | :
YYVGTDNLDLAVILFTGVGTCCICIIIGRRYIVGGELG-GKKLT KYLTGLFLCLLWVYVWVASLRS
9+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++
10+++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++ ++++++

```

Figure 49. TMSs 7-10 of Slp1 of the CaCA family (gi: 115709819; 10 TMSs) aligned with TMSs 7-10 of Ntfl of the TDT family (gi: 119484216; 10 TMSs) with a comparison score of 10 S.D.

[illegible]

Figure 50. TMSs 3-6 of Ntf2 of the RND family (gi: 254787166; 12 TMSs) aligned with TMSs 2-5 of Etb2 of the MPE family (gi: 149186203; 10 TMSs) with a comparison score of 11 S.D.

1+++++ 2+++++ 3+++++

LRKFCRLHLWLLVILFALVTGVVAIYNASSFREGLHLSGMWRTQLQWAVVGMI VFFIAAFIDYKWV----RWGGWMVMYGGVIWLILPQLVH

| | | | : | | | : | : | : | : | : | : | : | :

LHKKTTRLWAFLAAL-ACVMWGISGLFAKSLFNVSSEITPMWLSQLRMVTSGIILLIVAVILKKQPIATMKNKHDAWIAYGVFGLLPLPVLFY

1+++++ 2+++++ 3+++++

Figure 51. TMSs 1-3 of Vms1 of the MPE family (gi: 171914172; 10 TMSs) aligned with TMSs 1-3 of Lbh1 of the DMT family (gi: 161506918; 10 TMSs) with a comparison score of 11 S.D.


```

7+++++_____8_+++++_____
LVIHLPAYLFIKILQGSRLTANTTIHWATW-LGCTSGIAIIAYIIASAIPVFNDLVSLV
: | :| | | : : | | |::| : || |::|:|:| :|| : || :|
VTILIPAALYFISPVDLKQIETVNEHIYFALFLLGLLSGVALISYTMASYLVFPN----VL
7+++++_____8+++++_____

9+++++_____10+++++_____
GALLGTLMFQPMGCMWLYDNWSKGKISKSPKWIFMVFWSGFVILSGTFLMVGGTYSVVSIIIDSYKK
| ||| | :|: | : :: |: | :| :: | || | |::| : ::|: ||
G-LLGYL---EPVAMMAI--SFFIGETLKPEYILLICLS-FAIL---FLIIDG----ILNLIKQKK
9 +++++ +++++ +++++ 10++++ +++++ +++++ +

```

Figure 55. TMSs 7-10 of Afl15 of the AAP family (gi: 238507023; 10 TMSs) aligned with TMSs 7-10 of Ami2 of the DMT family (gi: 223042098; 10 TMSs) with a comparison score of 14 S.D.

```

2+++++_____3+++++_____+++++_____4+++++_____+
FVLFPKQFMKMLKHPGSTFLGCIPMAF--CTTINMIHNIWQDGLACY---ILWVFNVM SLATCWGTVSMFYL---H
| | : | : | | | : : | | :| | :| :| | | | :
FPLFVSYIHRRSKNAGSTKLYYITPRLFIACAVIGV--LTGLDDFLAAYGVSLPVSTALIIATQLGFTAGFAYVLVKQ
2+++++ 3+++++ +++++ 4+++++

_____5+++++_____6+++++_____
KRTATMLNATMILPIVSVVCAATGTLFWDNVP HHLQGLQLVVCIMLWANGQCLAFGFV
| | :| | :| | :| : : | | | | | | :| |
KFTPFTVNAIFLLSIGAVVLVLHASS---DRPPHETNGQYLSGFFM--TLGASALYGFV
5+++++ 6++ +++++

```

Figure 56. TMSs 2-6 of Ylt1 of the TDT family (gi: 50553858; 10 TMSs) aligned with TMSs 2-6 of Rcm1 of the DMT family (gi: 255581410; 10 TMSs) with a comparison score of 12 S.D.

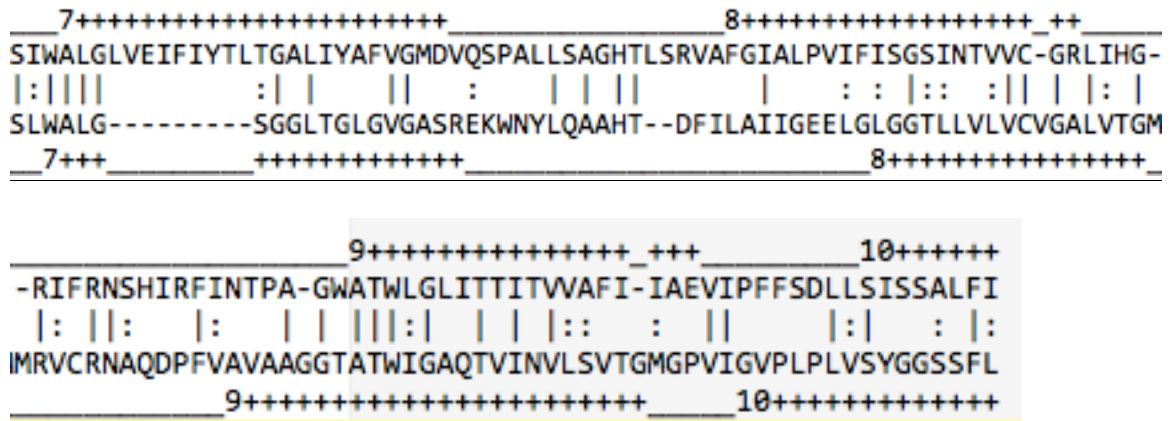


Figure 59. TMSs 7-10 of Afu3 of the AAAP family (gi: 146324801; 11 TMSs) aligned with TMSs 7-10 of Asp17 of the MPE family (gi: 320093976; 10 TMSs) with a comparison score of 12 S.D.

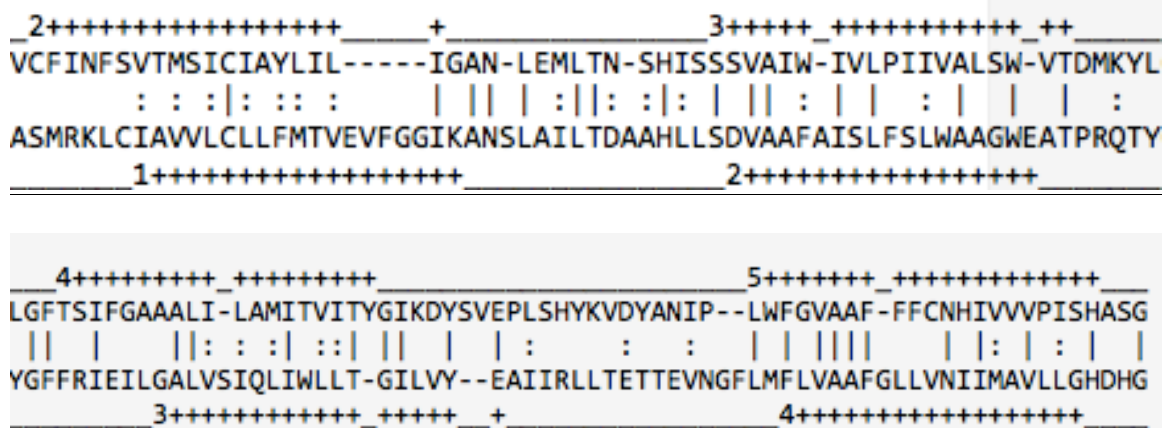


Figure 60. TMSs 2-5 of Dfa1 of the AAAP family (gi: 328874879; 11 TMSs) aligned with TMSs 1-4 of Bju4 of the CDF family (gi: 32274627; 6 TMSs) with a comparison score of 13 S.D.

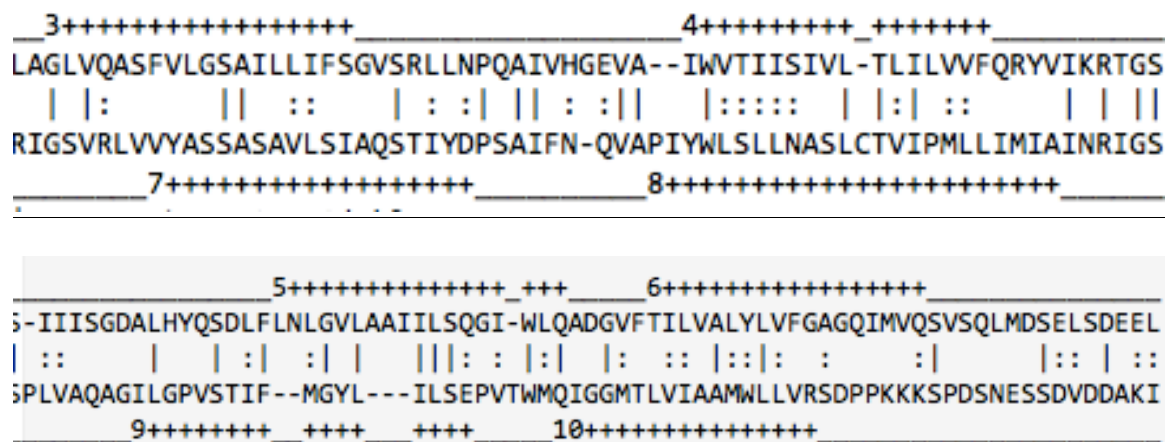


Figure 61. TMSs 3-5 of Cps1 of the CDF family (gi: 71279891; 6 TMSs) aligned with TMSs 7-10 of Pne1 of the DMT family (gi: 145589545; 10 TMSs) with a comparison score of 11 S.D.

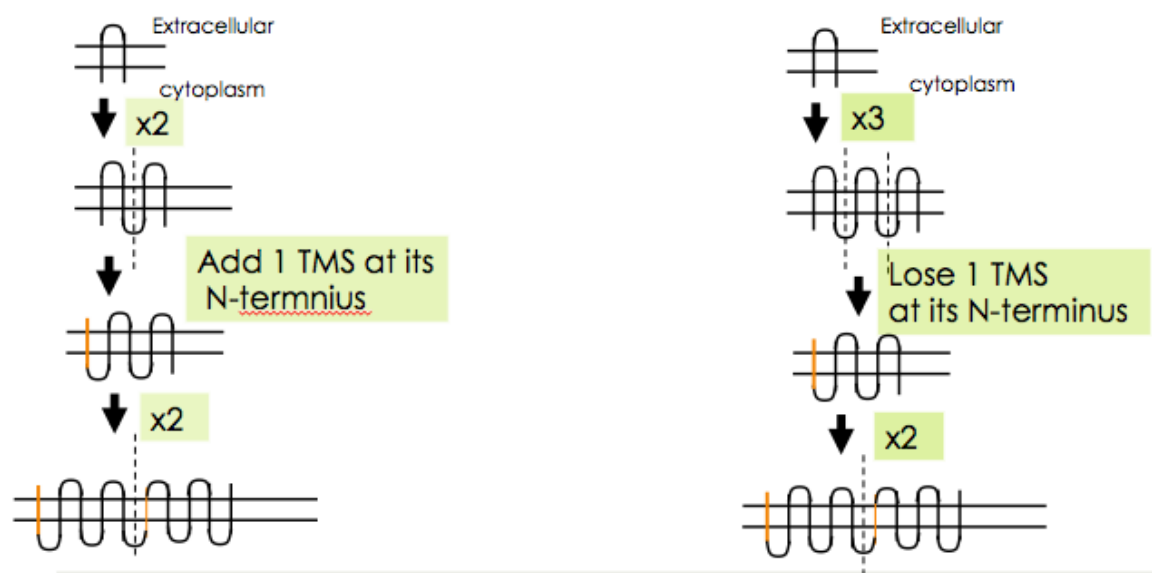


Figure 62. Proposed evolutionary pathways for CaCA family proteins.

Reference

- Altschul, S. F., T. L. Madden, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic acids research **25**(17): 3389-3402.
- Besserer, G. M., M. Ottolia, Nicoll DA, Chaptal V, Cascio D, Philipson KD, Abramson J."The second Ca²⁺-binding domain of the Na⁺ Ca²⁺ exchanger is essential for regulation: crystal structures and mutational analysis." Proceedings of the National Academy of Sciences of the United States of America **104**(47): 18467-18472.
- Boyle, D. S., M. M. Khattar, Addinall SG, Lutkenhaus J, Donachie WD. (1997). "ftsW is an essential cell-division gene in Escherichia coli." Molecular microbiology **24**(6): 1263-1273.
- Cagnac, O., M. Leterrier, Yeager M, Blumwald E. (2007). "Identification and characterization of Vnx1p, a novel type of vacuolar monovalent cation/H⁺ antiporter of Saccharomyces cerevisiae." The Journal of biological chemistry **282**(33): 24284-24293.
- Derouaux, A., B. Wolf, Claudine Fraipont, Eefjan Breukink, Martine Nguyen-Distèche, and Mohammed Terrak (2008). "The monofunctional glycosyltransferase of Escherichia coli localizes to the cell division site and interacts with penicillin-binding protein 3, FtsW, and FtsN." Journal of bacteriology **190**(5): 1831-1834.
- DiPolo, R. and L. Beauge (2006). "Sodium/calcium exchanger: influence of metabolic regulation on ion carrier interactions." Physiological reviews **86**(1): 155-203.
- Doolittle, R. F. (1981). "Protein evolution." Science **214**(4525): 1123-1124.
- Errington, J. (2003). "Regulation of endospore formation in Bacillus subtilis." Nature reviews. Microbiology **1**(2): 117-126.
- Fraipont, C., S. Alexeeva, Wolf B, van der Ploeg R, Schloesser M, den Blaauwen T, Nguyen-Distèche M. (2011). "The integral membrane FtsW protein and peptidoglycan synthase PBP3 form a subcomplex in Escherichia coli." Microbiology **157**(Pt 1): 251-259.
- Fuerst, J. A. and E. Sagulenko (2011). "Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function." Nature reviews. Microbiology **9**(6): 403-413.

- Fujisawa, M., Y. Wada, Tsuchiya T, Ito M. (2009). "Characterization of *Bacillus subtilis* YfkE (ChaA): a calcium-specific $\text{Ca}^{2+}/\text{H}^{+}$ antiporter of the CaCA family." Archives of microbiology **191**(8): 649-657.
- Gerard, P., T. Vernet, Zapun A. (2002). "Membrane topology of the *Streptococcus pneumoniae* FtsW division protein." Journal of bacteriology **184**(7): 1925-1931.
- Gomez-Villafuertes, R., B. Mellstrom, Naranjo JR. (2007). "Searching for a role of NCX/NCKX exchangers in neurodegeneration." Molecular neurobiology **35**(2): 195-202.
- Iwamoto, T., T. Y. Nakamura, Pan Y, Uehara A, Imanaga I, Shigekawa M. (1999). "Unique topology of the internal repeats in the cardiac $\text{Na}^{+}/\text{Ca}^{2+}$ exchanger." FEBS letters **446**(2-3): 264-268.
- Lam, V. H., J. H. Lee, Silverio A, Chan H, Gomolplitinant KM, Povolotsky TL, Orlova E, Sun EI, Welliver CH, Saier MH Jr. (2011). "Pathways of transport protein evolution: recent advances." Biological chemistry **392**(1-2): 5-12.
- Li, W. and A. Godzik (2006). "Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences." Bioinformatics **22**(13): 1658-1659.
- Liao, J., H. Li, et al. (2012). "Structural Insight into the Ion-Exchange Mechanism of the Sodium/Calcium Exchanger." Science **335**(6069): 686-690.
- Maggi, S., O. Massidda, Luzi G, Fadda D, Paolozzi L, Ghelardini P. (2008). "Division protein interaction web: identification of a phylogenetically conserved common interactome between *Streptococcus pneumoniae* and *Escherichia coli*." Microbiology **154**(Pt 10): 3042-3052.
- Makarova, K. S., Y. I. Wolf, van der Oost J, Koonin EV. (2009). "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements." Biology direct **4**: 29.
- Matias, M. G., K. M. Gomolplitinant, Tamang DG, Saier MH 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 research notes **3**: 158.
- Matsuzawa, H., S. Asoh, Kunai K, Muraiso K, Takasuga A, Ohta T. (1989). "Nucleotide sequence of the rodA gene, responsible for the rod shape of *Escherichia coli*: rodA and the pbpA gene, encoding penicillin-binding protein 2, constitute the rodA operon." Journal of bacteriology **171**(1): 558-560.

- Mohammadi, T., V. van Dam, Sijbrandi R, Vernet T, Zapun A, Bouhss A, Diepeveen-de Bruin M, Nguyen-Distèche M, de Kruijff B, Breukink E. (2011). "Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane." The EMBO journal **30**(8): 1425-1432.
- Radchenko, M. V., K. Tanaka, Waditee R, Oshimi S, Matsuzaki Y, Fukuhara M, Kobayashi H, Takabe T, Nakamura T. (2006). "Potassium/proton antiport system of Escherichia coli." The Journal of biological chemistry **281**(29): 19822-19829.
- Reddy, V. S. and M. H. Saier, Jr. (2012). "BioV Suite - a collection of programs for the study of transport protein evolution." The FEBS journal **279**(11): 2036-2046.
- Ren, X., D. A. Nicoll, Philipson KD. (2006). "Helix packing of the cardiac Na⁺-Ca²⁺ exchanger: proximity of transmembrane segments 1, 2, and 6." The Journal of biological chemistry **281**(32): 22808-22814.
- Reynaud, E. G. and D. P. Devos (2011). "Transitional forms between the three domains of life and evolutionary implications." Proceedings. Biological sciences / The Royal Society **278**(1723): 3321-3328.
- Saaf, A., L. Baars, von Heijne G. (2001). "The internal repeats in the Na⁺/Ca²⁺ exchanger-related Escherichia coli protein YrbG have opposite membrane topologies." Journal of Biological Chemistry **276**(22): 18905-18907.
- Saier, M. H., Jr. (2003). "Tracing pathways of transport protein evolution." Molecular microbiology **48**(5): 1145-1156.
- Saier, M. H., Jr., B. H. Eng, Fard S, Garg J, Haggerty DA, Hutchinson WJ, Jack DL, Lai EC, Liu HJ, Nusinew DP, Omar AM, Pao SS, Paulsen IT, Quan JA, Sliwinski M, Tseng TT, Wachi S, Young GB. (1999). "Phylogenetic characterization of novel transport protein families revealed by genome analyses." Biochimica et biophysica acta **1422**(1): 1-56.
- Saier, M. H., Jr. and Q. Ren (2006). "The bioinformatic study of transmembrane molecular transport." Journal of molecular microbiology and biotechnology **11**(6): 289-290.
- Saier, M. H., Jr., C. V. Tran, Barabote RD. (2006). "TCDB: the Transporter Classification Database for membrane transport protein analyses and information." Nucleic acids research **34**(Database issue): D181-186.
- Saier, M. H., Jr., M. R. Yen, Noto K, Tamang DG, Elkan C. (2009). "The Transporter Classification Database: recent advances." Nucleic acids research **37**(Database issue): D274-278.

- Sato, T., G. Theeragool, Yamamoto T, Okamoto M, Kobayashi Y. (1990). "Revised nucleotide sequence of the sporulation gene spoVE from *Bacillus subtilis*." Nucleic acids research **18**(13): 4021.
- Schwarz, E. M. and S. Benzer (1997). "Calx, a Na-Ca exchanger gene of *Drosophila melanogaster*." Proceedings of the National Academy of Sciences of the United States of America **94**(19): 10249-10254.
- Segarra, V. A. and L. Thomas (2008). "Topology mapping of the vacuolar Vcx1p Ca²⁺/H⁺ exchanger from *Saccharomyces cerevisiae*." The Biochemical journal **414**(1): 133-141.
- Shaffer, P. L., A. Goehring, Shankaranarayanan A, Gouaux E. (2009). "Structure and mechanism of a Na⁺-independent amino acid transporter." Science **325**(5943): 1010-1014.
- Thompson, J. D., T. J. Gibson, Plewniak F, Jeanmougin F, Higgins DG. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic acids research **25**(24): 4876-4882.
- Tran, C. V. and M. H. Saier, Jr. (2004). "The principal chloroquine resistance protein of *Plasmodium falciparum* is a member of the drug/metabolite transporter superfamily." Microbiology **150**(Pt 1): 1-3.
- Tseng, T. T., K. S. Gratwick Kollman J, Park D, Nies DH, Goffeau A, Saier MH Jr. (1999). "The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins." Journal of molecular microbiology and biotechnology **1**(1): 107-125.
- Tsoi, M., K. H. Rhee, Bungard D, Li XF, Lee SL, Auer RN, Lytton J. (1998). "Molecular cloning of a novel potassium-dependent sodium-calcium exchanger from rat brain." The Journal of biological chemistry **273**(7): 4155-4162.
- Tusnady, G. E. and I. Simon (2001). "The HMMTOP transmembrane topology prediction server." Bioinformatics **17**(9): 849-850.
- Visser, F., V. Valsecchi, Annunziato L, Lytton J. (2007). "Exchangers NCKX2, NCKX3, and NCKX4: identification of Thr-551 as a key residue in defining the apparent K(+) affinity of NCKX2." The Journal of biological chemistry **282**(7): 4453-4462.
- Wong, F. H., J. S. Chen, Reddy V, Day JL, Shlykov MA, Wakabayashi ST, Saier MH Jr. (2012). "The Amino Acid-Polyamine-Organocation Superfamily." Journal of molecular microbiology and biotechnology **22**(2): 105-113.

- Wu, M., S. Tong, Gonzalez J, Jayaraman V, Spudich JL, Zheng L. (2011). "Structural basis of the Ca^{2+} inhibitory mechanism of *Drosophila* $\text{Na}^{+}/\text{Ca}^{2+}$ exchanger CALX and its modification by alternative splicing." Structure **19**(10): 1509-1517.
- Yen, M. R., J. Choi, Saier MH Jr. (2009). "Bioinformatic analyses of transmembrane transport: novel software for deducing protein phylogeny, topology, and evolution." Journal of molecular microbiology and biotechnology **17**(4): 163-176.
- Zhai, Y. and M. H. Saier, Jr. (2001). "A web-based program (WHAT) for the simultaneous prediction of hydropathy, amphipathicity, secondary structure and transmembrane topology for a single protein sequence." Journal of molecular microbiology and biotechnology **3**(4): 501-502.
- Zhai, Y. and M. H. Saier, Jr. (2001). "A web-based program for the prediction of average hydropathy, average amphipathicity and average similarity of multiply aligned homologous proteins." Journal of molecular microbiology and biotechnology **3**(2): 285-286.
- Zhai, Y., J. Tchieu, Saier MH Jr. (2002). "A web-based Tree View (TV) program for the visualization of phylogenetic trees." Journal of molecular microbiology and biotechnology **4**(1): 69-70.