Supporting data

Figure S1. (A) AveHAS plots and **(B)** phylogenetic tree for 135 LCT family proteins included in this study. The tree was generated using the ClustalX multiple alignment and the FigTree program. Protein abbreviations and their descriptions are listed in Table S1 in a clockwise fashion starting from cluster 1. (**C)** Homology between members of the Sweet and LCT families. GSAT alignment of TMSs 3-7 of LCT Bba2 (*Beauveria bassiana*; gi400602576; 7 TMSs) with TMSs 3-7 of Sweet Aae2 (*Aedes aegypti*; gi157109690; 7 TMSs). A comparison score of 12.4 S.D. was obtained with 43.7% similarity and 27.2% identity. Another alignment (not shown) gave 13.6 S.D. and had 2 TMSs aligned. **(D)** Homology between members of the PNaS and LCT families. GSAT alignment of TMSs 2-4 of LCT Ago1 (*Ashbya gossypii*; gi302308260; 7 TMSs) and TMSs 6-8 of PNaS Cre1 (*Caenorhabditis remanei*; gi308511717; 11 TMSs). A comparison score of 12.8 S.D. was obtained with 48.1% similarity and 30.8% identity.

Figure S2. (A) AveHAS plots and **(B)** phylogenetic tree for 104 MR family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S2 in a clockwise fashion starting from cluster 1.

Figure S3. (A) AveHAS plots and **(B)** phylogenetic tree for 134 NiCoT family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S3 in a clockwise fashion starting from cluster 1. **(C)** Homology between members of the TSUP and NiCoT families. GSAT alignment of TMSs 4-6 of NiCoT Bja1 (*Bradyrhizobium japonicum*; gi 27376265; 6 TMSs) with TMSs 1-3 of TSUP Pla1 (*Parvibaculum lavamentivorans*; gi

154252649; 8 TMSs). A comparison score of 12.8 S.D. was obtained with 44.7% similarity and 36.4% identity.

Figure S4. (A) AveHAS plots and **(B)** phylogenetic tree for 52 OST family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S4 in a clockwise fashion starting from cluster 1. **(C)** Homology between members of the TSUP and OST families. GSAT alignment of TMSs 2-3 of TSUP Tsp1 (*Thermococcus sp. AM4*; gi254172062; 8 TMSs) with TMSs 1-2 of OST Cre2 (*Chlamydomonas reinhardtii*; gi159465163; 7 TMSs). A comparison score of 12.1 S.D. was obtained with 50.8% similarity and 33.9% identity. **(D)** Binary alignment of members of the TSUP and OST families with 4 TMSs aligned, further supporting **(C)**. GSAT alignment of TMSs 2-5 of TSUP Gfo1 (*Gramella forsetii*; gi120434413; 7 TMSs) and TMSs 2-5 of OST Dre1 (*Danio rerio*; gi169146805; 7 TMSs). The last two predicted TMSs show displacement, possibly a result of the prediction program used. A comparison score of 11.3 S.D. was obtained with 47.0% similarity and 26.2% identity.

Figure S5. (A) AveHAS plots and **(B)** phylogenetic tree for 198 Sweet family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S5 in a clockwise fashion starting from cluster 1. **(C)** Homology between members of the Sweet and OST families. GSAT alignment of TMSs 6-7 of Sweet Rco4 (*Ricinus communis*; gi255540127; 7 TMSs) with TMSs 6-7 of OST Ath8 (*Arabidopsis thaliana*; gi18411404; 7 TMSs). A comparison score of 12.3 S.D. was obtained with 44.3% similarity and 22.1% identity. **(D)** Binary alignment of members of the Sweet and OST families with 3 TMSs aligned, further supporting **(C)**. GSAT alignment of TMSs 4-6 of Sweet Asu3 (*Ascaris suum*; gi324518669; 7 TMSs) and TMSs 4-6 of

OST Ncr1 (*Neurospora crassa*; gi85106985; 7 TMSs). A comparison score of 10.3 S.D. was obtained with 49.1% similarity and 28.7% identity.

Figure S6. **(A)** AveHAS plots and **(B)** phylogenetic tree for 147 PNaS family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S6 in a clockwise fashion starting from cluster 1. **(C)** Homology between members of the Sweet and PNaS families. GSAT alignment of TMSs 4-6 of PNaS Odi8 (*Oikopleura dioica*; gi313242874; 8 TMSs) with TMSs 3-5 of MR Hwa1 (*Haloquadratum walsbyi DSM 16790*; gi110668951; 7 TMSs). A comparison score of 13.1 S.D. was obtained with 57.3% similarity and 40.0% identity.

Figure S7. **(A)** AveHAS plots and **(B)** phylogenetic tree for 74 PnuC family proteins included in this study. Protein abbreviations and their descriptions are listed in Table S7 in a clockwise fashion starting from cluster 1. **(C)** Homology between members of the PnuC and TSUP families. GSAT alignment of TMSs 2-3 of PnuC Spr1 (*Serinicoccus profundi*; gi379058816; 7 TMSs) with TMSs 3-4 of TSUP Cba4 (*Clostridiales bacterium 1_7_47FAA*; gi239625513; 8 TMSs). A comparison score of 12.4 S.D. was obtained with 49.3% similarity and 29.9% identity. **(D)** Binary alignment of members of the PnuC and TSUP families with 3 TMSs aligned, further supporting **(C)**. GSAT alignment of TMSs 3-5 of PnuC Sde2 (*Sulfurospirillum delevignum*; gi268679606; 7 TMSs) with TMSs 4-6 of TSUP Ere1 (*Eubacterium rectale*; gi238924940; 8 TMSs). A comparison score of 11.1 S.D. was obtained with 54.4% similarity and 25.2% identity.

Figure S8. **(A)** AveHAS plots and **(B)** phylogentic tree for the 41 GPCR family proteins (see Table S8A) included in this study. Due to the high similarity of homologues generated with the

Protocol 1 program, selected proteins obtained using this protocol with Dre4 as the query sequence are presented in Table S8B and not included in this tree; they all belong to the Dre4 branch (cluster 10 of the tree). Additional proteins used to produce the phylogenetic tree **(B)** were compiled from a diverse group of functional types obtained from the GPCRDB (http://www.gpcr.org/7tm/). **(C)** Homology between members of the GPCR and the microbial rhodopsin (MR) families. GSAT alignment of TMSs 5-6 of MR Mos1 (*Mixia osmundae*; gi358059197; 7 TMSs) with TMSs 5-6 of GPCR Dre1 (*Danio rerio*; gi292621989; 7 TMSs). A comparison score of 13.1 S.D. was obtained with 44.9% similarity and 32.1% identity. **(D)** Binary alignment of members of the GPCR and MR families with 4 TMSs aligned, further supporting **(C)**. GSAT alignment of TMSs 1-4 of GPCR Dre5 (*Danio rerio*; gi122064639; 7TMSs) with TMSs 1-4 of MR Cga1 (*Cryptococcus gattii*; gi321261457; 7 TMSs). A comparison score of 10.6 S.D. was obtained with 38.8% similarity and 27.9% identity. AveHAS plots and the phylogentic tree for the 56 GPCR proteins (Table S8B) that were used in finding the comparisons shown in **(C)** and **(D)** are presented in **(E)** and **(F)**, respectively. GPCR protein abbreviations and their descriptions are listed in Tables S8A and B in a clockwise fashion starting from clusters 1 of **(B)** and **(F)**, respectively.

Figure S9. (A) Schematic representation of the mean bit score of a given alignment. **(B)** Algorithm input into Mathematica 7.0 (Wolfram Research, Inc., Champaign, IL, USA*)* to determine probabilities of a given *Z* score of an alignment. **(C)** Probability of notable *Z* scores, as determined by Mathematica 7.0.

Figure S10. AveHAS plots representing all families within the TOG superfamily. The plots are similar to the plots presented in Figure 2, but due to the large N- and C-terminal extensions present in some GPCR homologues, this plot is not easy to interpret.

Figure S11. Results from the ClustalX-based neighbor-joining program, showing the TOG superfamily proteins. These proteins correspond to those shown in the SFT-derived tree shown in Fig. 3A.

Video S1. Video file of rotation of the superimposition of B-rho and M-rho.

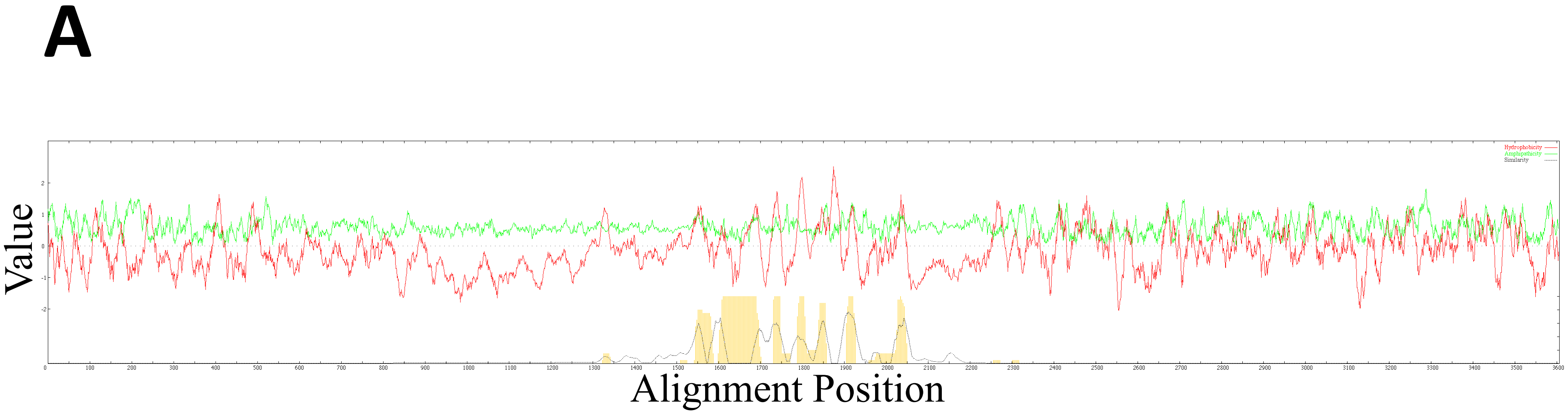

A
Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
3200
3300
3400
3500
3600
Alignment Position

Figure S1

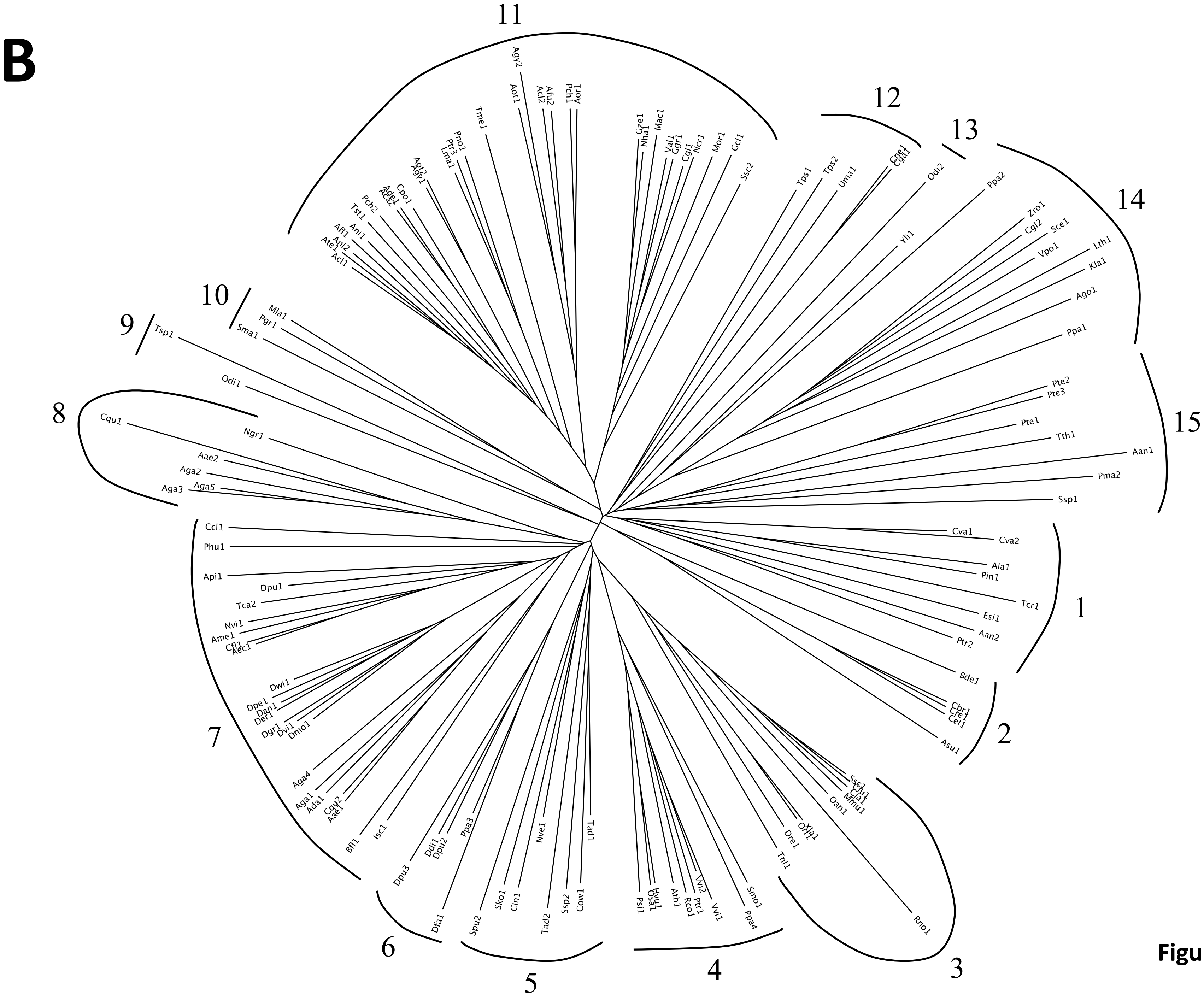

B
11
Agy2
Aot1
Acl2
Afu2
Pch1
Aor1
Tme1
Pno1
Ptr3
Lma1
Aot2
Agy1
Cpo1
Ade1
Aca2
Pch2
Tst1
Ani1
Afl1
Ani2
Ate1
Acl1
Gze1
Nha1
Mac1
Val1
Ggr1
Cgl1
Ncr1
Mor1
Gcl1
Ssc2
12
Tps1
Tps2
Uma1
Cne1
Cga1
Odi2
Yli1
13
14
Ppa2
Zro1
Cgl2
Sce1
Vpo1
Lth1
Kla1
Ago1
Ppa1
15
Pte2
Pte3
Pte1
Tth1
Aan1
Pma2
Ssp1
1
Cva1
Cva2
Ala1
Pin1
Tcr1
Esi1
Aan2
Ptr2
Bde1
2
Cbr1
Cre1
Cel1
Asu1
3
Ssc1
Clu1
Cja1
Mmu1
Oan1
Xla1
Ola1
Dre1
Tni1
Rno1
4
Smo1
Ppa4
Vvi1
Vvi2
Ptr1
Rco1
Ath1
Hvu1
Osa1
Psi1
5
Tad1
Cow1
Ssp2
Tad2
Nve1
Cin1
Sko1
Spu2
6
Ppa3
Dpu2
Ddi1
Dfa1
Dpu3
7
Isc1
Bfl1
Aae1
Cqu2
Ada1
Aga1
Aga4
Dmo1
Dvi1
Dgr1
Der1
Dan1
Dpe1
Dwi1
Aec1
Cfl1
Ame1
Nvi1
Tca2
Dpu1
Api1
Phu1
Ccl1
8
Aga3
Aga5
Aga2
Aae2
Cqu1
Ngr1
Odi1
9
Tsp1
10
Sma1
Pgr1
Mla1

Figure S1

C

```
        2                    3                          4
Aae2 68 MNDPAMIRANVVGFAI-SVVYSVFF----YLYTPRQSKGDFWKQLGIAGA 112
        |  |  :   : | || | :|| |:      || |  |   |  :
Bba2 83 MVQPNDVAYTLHGLAISSTIYSQFYPRLWGFYTTRPVKISLWCSMTFWAC 132
                         3                          4

                4                          5
Aae2 113 ITAAIVG-YAKIENPEVVEDRFGLIITVLM--LMLIAQPLFGLPEI---I 156
         | ||::|  |    |:  :    | :  |:     :   |  ||::
Bba2 133 ILAALLGTVAASAFPD-SQSWMWLDVVYLVGNFKTLLTLLKYLPQVWLNY 181
              4                         5

                                6                      7
Aae2 157 RKKSTEG---LPFAMILSGTVVGCMWLLYGIILNNTFVILQNLAAVSLSG 203
         |:|||:|   ||||: : |  :  : ||  :  ::    | |   : ||
Bba2 182 RRKSTQGLPPLPFALDIGGATLSLLQLLIDVAYSDQSAALANPVKLVLSN 231
                             6                       7

         7
Aae2 204 V 204
         :
Bba2 232 L 232
         7
```

Figure S1

D

```
              2                                       3
Ago1 44  CIGYTALVLSLVVQTVLWQNEGDTSSAVRPSVGWSDYWYAMHSWLLNGLL 93
         | ||  ::: ||| |:| |:    ||:: | || |        : |  :|
Cre1 371 CTGYLVMLVGLVV-TILIQSNSIFSSSLTPLVG-SGVVTLEQMYPL--VL 416
                   6                                      7

             3                                    4
Ago1 94  ASQVLWGTRLWG----FTGEPANRMSTAHRKMMQLFIGLFCIGT-VQYVV 138
          | :  ||   |    |: :|:      |  | |:   :  ||| : |:|
Cre1 417 GSNI--GTTFSGVLAAFSTDPSRFEKALHMAMCQVIYNI--IGTCLFYIV 461
                  7                               8

          4
Ago1 139 SITQ 142
          |:
Cre1 462 PCTR 465
         8
```

Figure S1

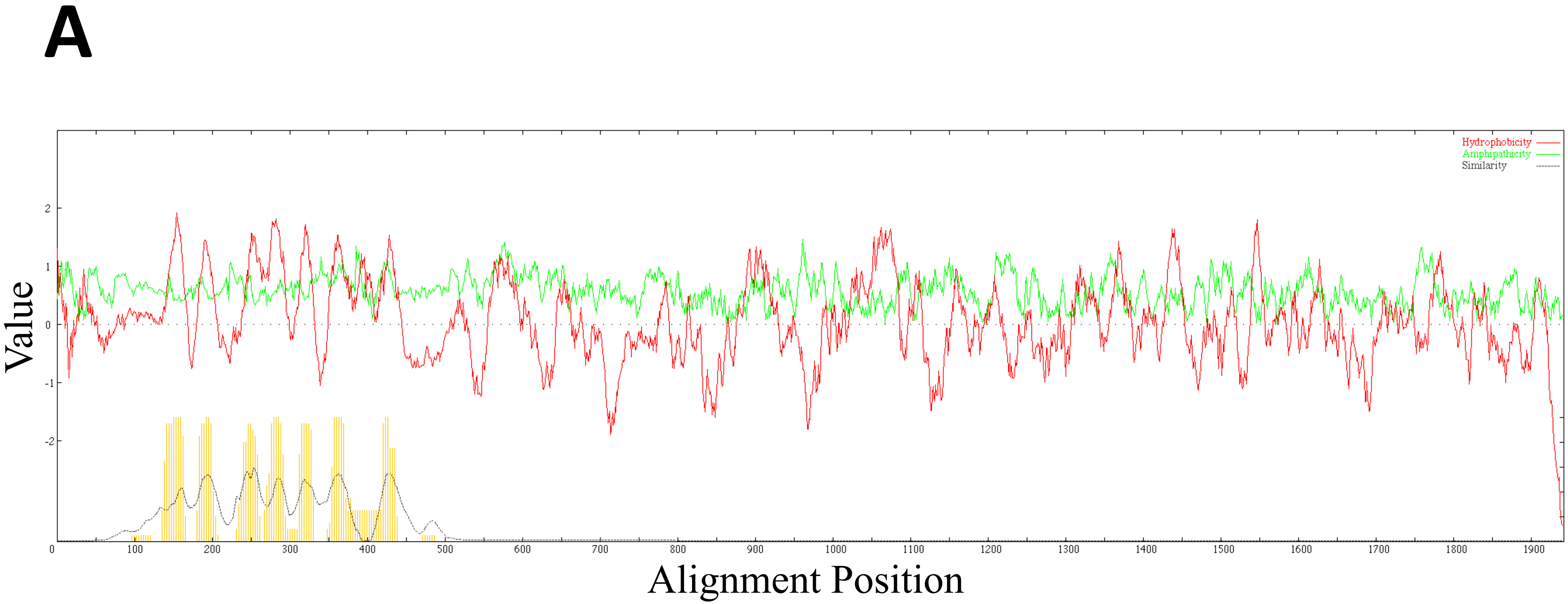

A
Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
Alignment Position
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure S2

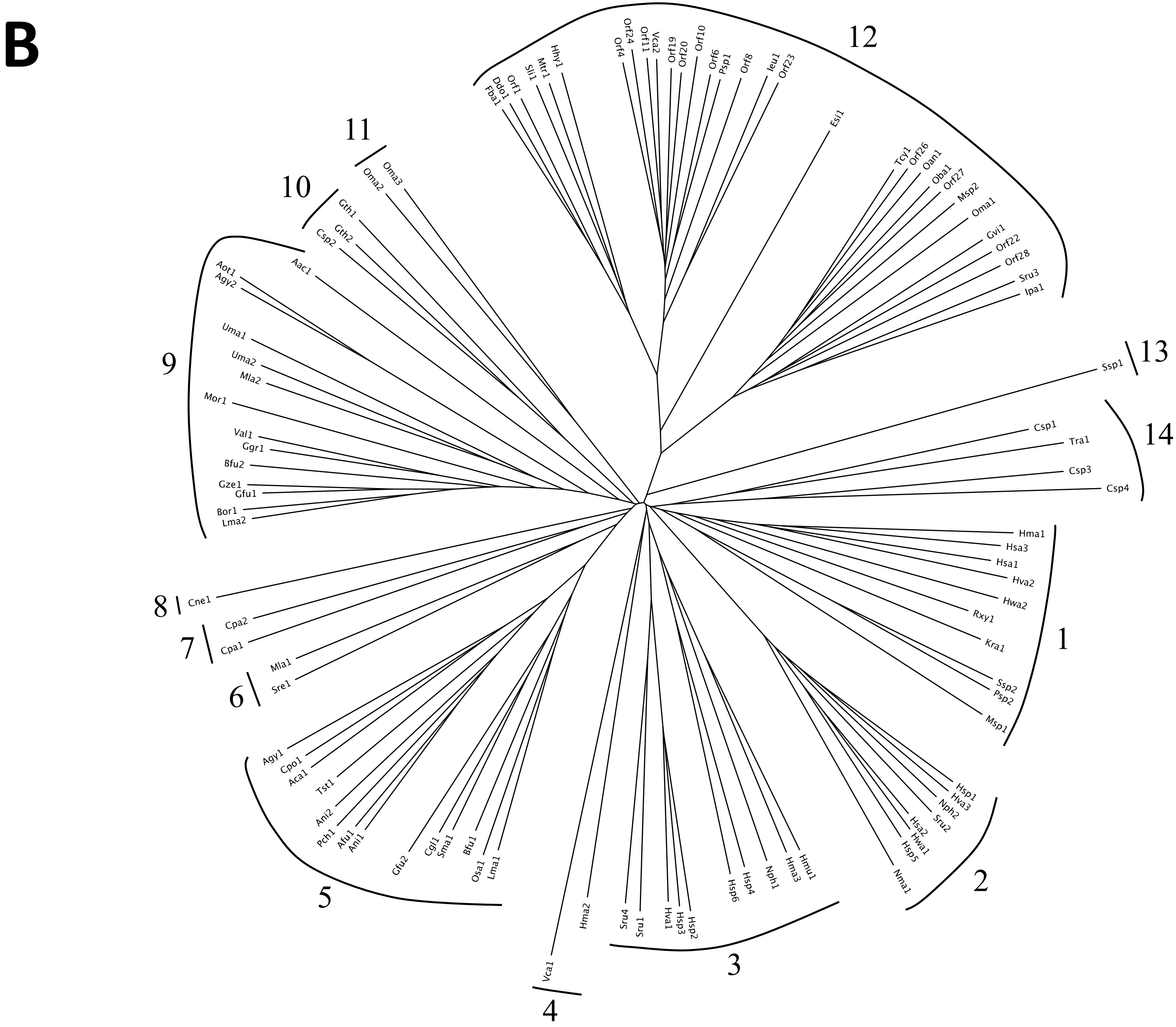
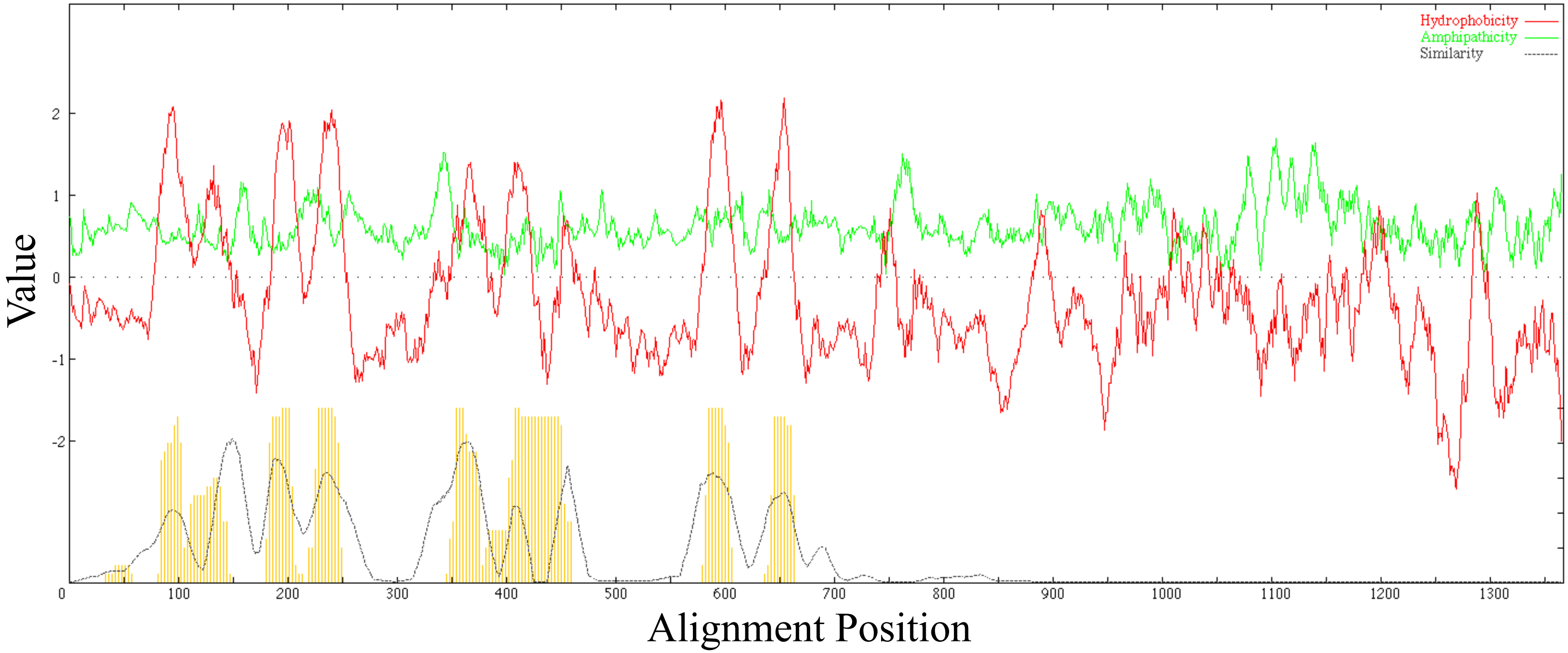

B
12
Fba1
Ddo1
Orf1
Sli1
Mtr1
Hhy1
Orf4
Orf24
Orf11
Vca2
Orf19
Orf20
Orf10
Orf6
Psp1
Orf8
Ieu1
Orf23
Esi1
Tcy1
Orf26
Oan1
Oba1
Orf27
Msp2
Oma1
Gvi1
Orf22
Orf28
Sru3
Ipa1
Ssp1
13
Csp1
Tra1
Csp3
Csp4
14
Hma1
Hsa3
Hsa1
Hva2
Hwa2
Rxy1
Kra1
Ssp2
Psp2
Msp1
1
Hsp1
Hva3
Nph2
Sru2
Hsa2
Hwa1
Hsp5
Nma1
2
Hmu1
Hma3
Nph1
Hsp4
Hsp6
Hsp2
Hsp3
Hva1
Sru1
Sru4
3
Hma2
Vca1
4
Lma1
Osa1
Bfu1
Sma1
Cgl1
Gfu2
Ani1
Afu1
Pch1
Ani2
Tst1
Aca1
Cpo1
Agy1
5
Sre1
Mla1
6
Cpa1
Cpa2
7
Cne1
8
Lma2
Bor1
Gfu1
Gze1
Bfu2
Ggr1
Val1
Mor1
Mla2
Uma2
Uma1
Agy2
Aot1
Aac1
9
Csp2
Gth2
Gth1
10
Oma2
Oma3
11

Figure S2

A



Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
Alignment Position

Figure S3

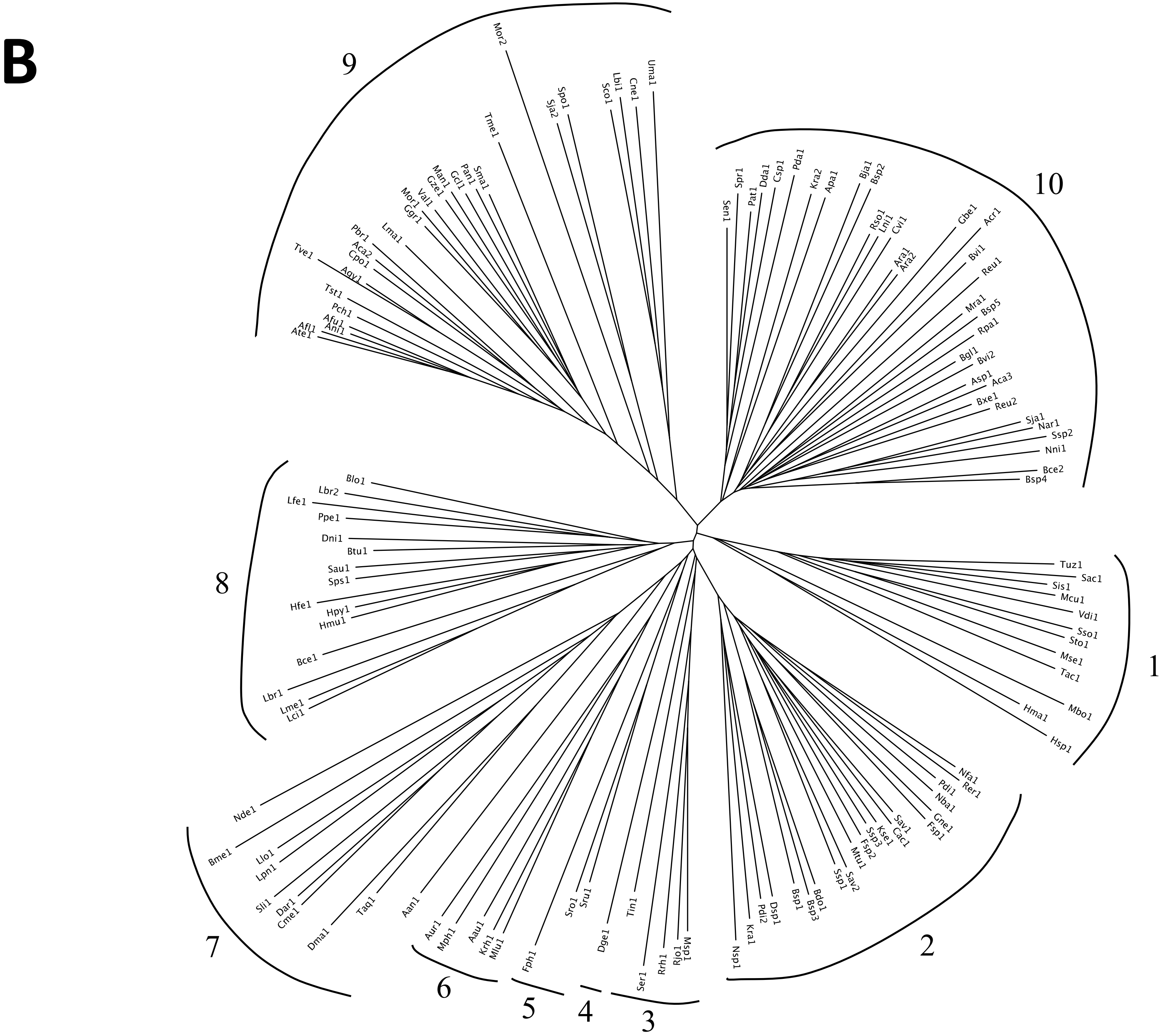

B
1
Tuz1
Sac1
Sis1
Mcu1
Vdi1
Sso1
Sto1
Mse1
Tac1
Mbo1
Hma1
Hsp1
2
Nfa1
Rer1
Pdi1
Nba1
Gne1
Fsp1
Sav1
Cac1
Kse1
Ssp3
Fsp2
Mtu1
Sav2
Ssp1
Bdo1
Bsp3
Bsp1
Dsp1
Pdl2
Kra1
Nsp1
3
Msp1
Rjo1
Rrh1
Ser1
Tin1
Dge1
4
Sru1
Sro1
5
Fph1
Mlu1
Krh1
6
Aau1
Mph1
Aur1
Aan1
7
Taq1
Dma1
Cme1
Dar1
Sli1
Lpn1
Llo1
Bme1
Nde1
8
Lci1
Lme1
Lbr1
Bce1
Hmu1
Hpy1
Hfe1
Sps1
Sau1
Btu1
Dni1
Ppe1
Lfe1
Lbr2
Blo1
9
Ate1
Afl1
Ani1
Afu1
Pch1
Tst1
Tve1
Agy1
Cpo1
Aca2
Pbr1
Lma1
Ggr1
Mor1
Val1
Gze1
Man1
Gcl1
Pan1
Sma1
Tme1
Mor2
Sja2
Spo1
Sco1
Lbi1
Cne1
Uma1
10
Sen1
Spr1
Pat1
Dda1
Csp1
Pda1
Kra2
Apa1
Bja1
Bsp2
Rso1
Lni1
Cvi1
Ara1
Ara2
Gbe1
Acr1
Bvi1
Reu1
Mra1
Bsp5
Rpa1
Bgl1
Bvi2
Asp1
Aca3
Bxe1
Reu2
Sja1
Nar1
Ssp2
Nni1
Bce2
Bsp4

Figure S3

C

```
                                     1                   2
Pla1  25 EGAHLHCRNDHAQSHHTPHTHSPRPKFLIRSTANSAIERRGNRSVSPVRG 74
         |  | |   ||   |   | | | |  :       |  |:|      |||
Bja1 179 EHGHVHHDHDH.DTHEHDHAHIPTPADI......RAAKRKG......VRG 215
                                     1                   2
Pla1  75 SMQLYLPIADLPVNVFLVLAMGAAVGFVSGMFGIGGGFLMTPLLIFIG.. 122
            : | :   |   :    : |      | | ||    |  :   :|
Bja1 216 MAAMILSVGLRPCTGAILVLLFAV...TQGAFSIG...VMSAIVMSVGTA 259
                     4                               5
                2                                  3
Pla1 123 IT.PAVAVASVASHIAASSFSGAI.SYWRR...RAIDPALASVLLCGGVT 167
         ||  | |   | |   |  | | : | | |   | :  |  ||:|
Bja1 260 ITVSALALMTVFSKRLALRFAGGVDSPWARRVERGLKIAGGSVIL...LF 306
               5                                   6
              3
Pla1 168 GTALGVWTFTQ 178
         |  | |  |||
Bja1 307 GMMLLVASFTQ 317
              6
```

Figure S3

A

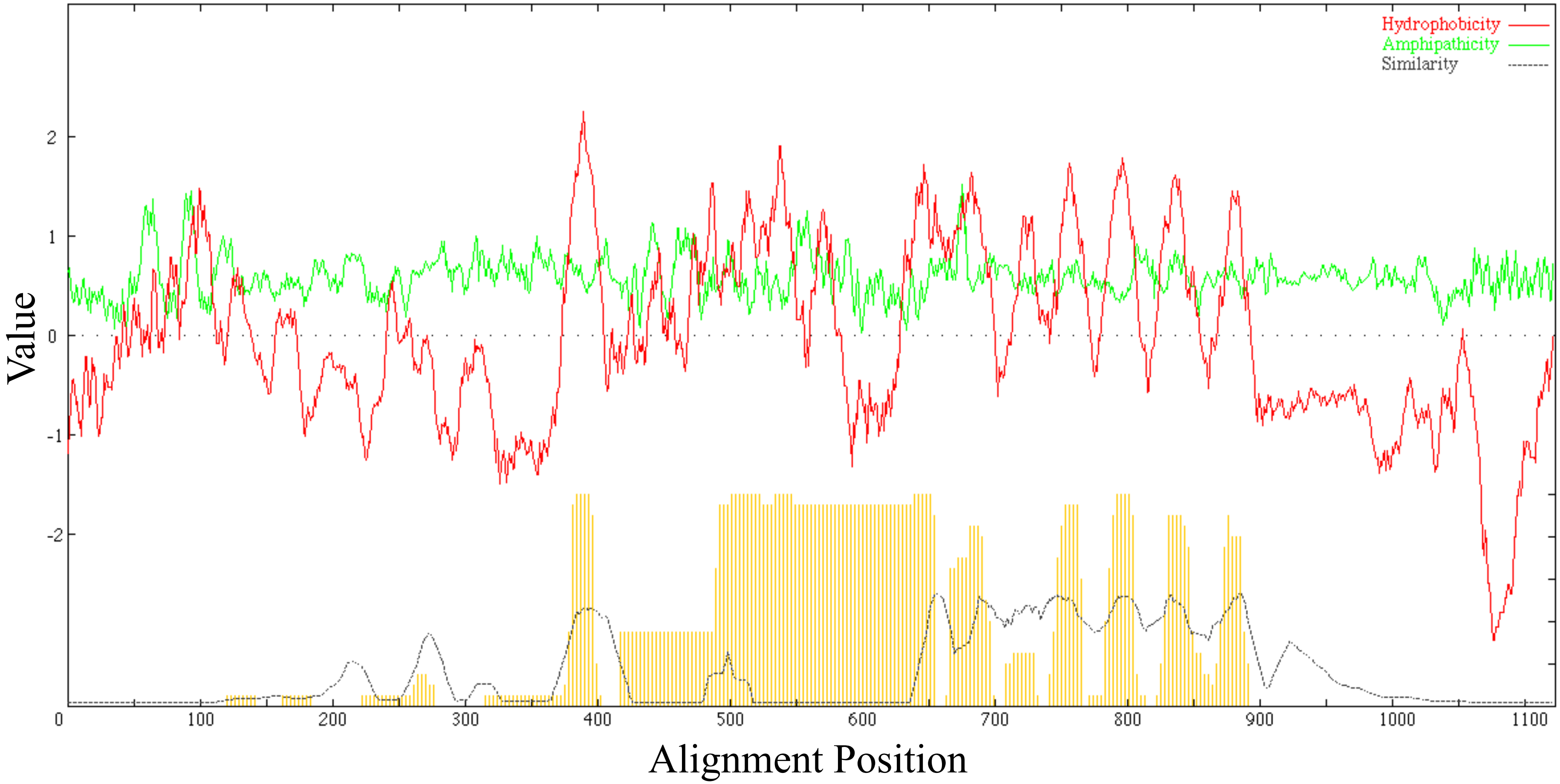

Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
Alignment Position

Figure S4

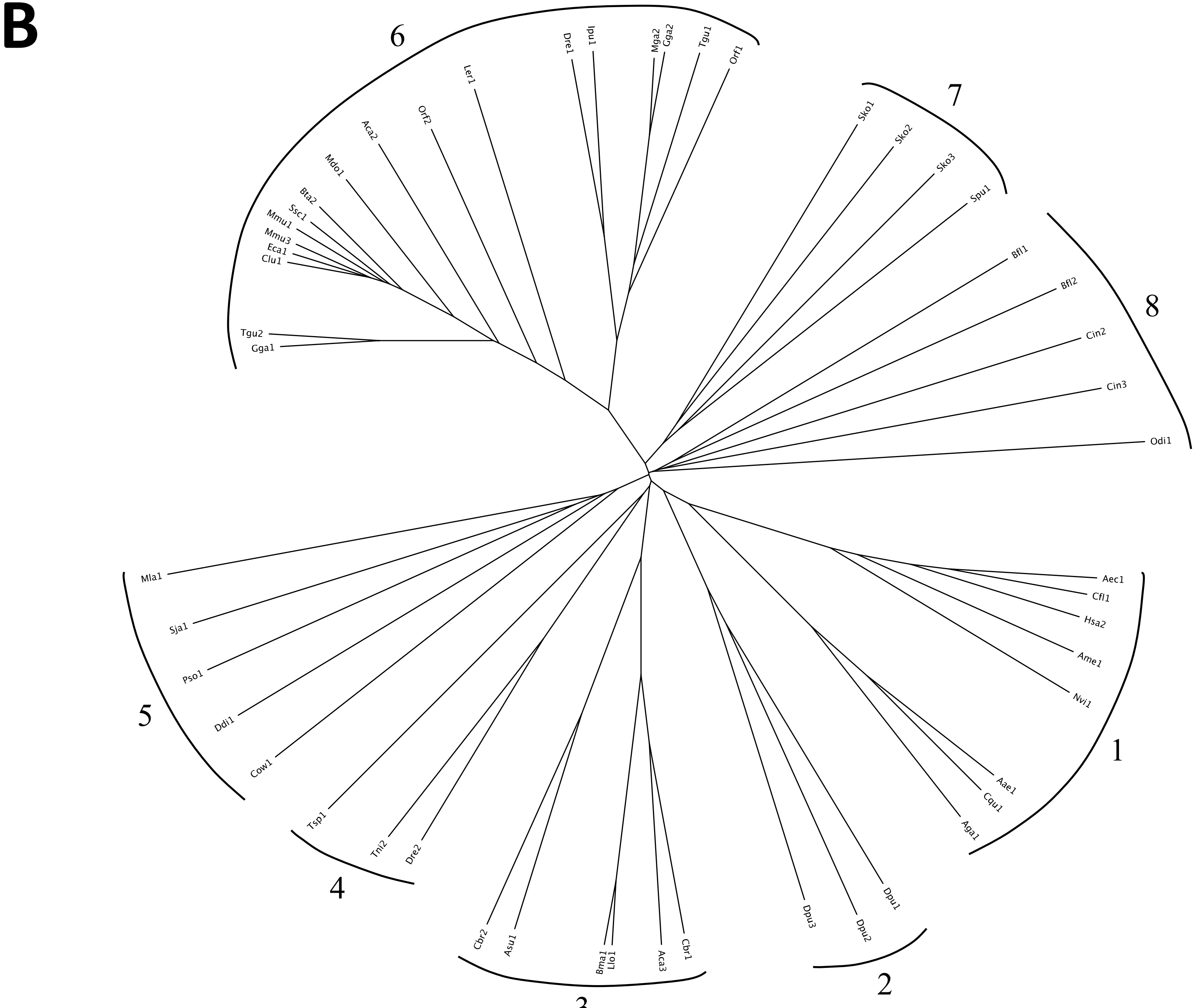

B
6
Tgu2
Gga1
Clu1
Eca1
Mmu3
Mmu1
Ssc1
Bta2
Mdo1
Aca2
Orf2
Ler1
Dre1
Ipu1
Mga2
Gga2
Tgu1
Orf1
7
Sko1
Sko2
Sko3
Spu1
8
Bfl1
Bfl2
Cin2
Cin3
Odi1
1
Aec1
Cfl1
Hsa2
Ame1
Nvi1
Aae1
Cqu1
Aga1
2
Dpu1
Dpu2
Dpu3
3
Cbr1
Aca3
Llo1
Bma1
Asu1
Cbr2
4
Dre2
Tni2
Tsp1
5
Cow1
Ddi1
Pso1
Sja1
Mla1

Figure S4

C

```
                        2                              3
Tsp1 35 LLGEPIHLAIGTSLTCI.VLSSLSASLTHIRR.GAVLYRVVLLKEVFSVP 82
        :| :|| | : |  | : | |     | |:|     ||   ::: :| ||
Cre2  1 MLRDPITLGLATAATWLSVASAVTQILCHLRNYTEPLYQRYIIRIIFLVP 50
                         1                          2

              3
Tsp1 83 FAVLGAYLSSM 93
        |     :|| |
Cre2 51 FYGVTSWLSIM 61
              2
```

Figure S4

D

```
                    2                                 3
Dre1   1 ILFIVPIYAFDSWLSLLFFTNEEYYVYFDTVRDCYEAFVIYNFLSLCYEY 50
         :| ::|: :  ::||: :| :  ||||  : |:     : :: |::
Gfo1   1 LLMVMPLPSAIAYLSIFYFISTPYYVYKE--RNNIDKLLLKK-LAIS-SL 46
                        2                                3

              3                                          4
Dre1  51 LGGESAIMAEIRGKPIQSSFVYGTCCLWGRTYSIGFLRFCKQATLQFCVV 100
         :|  : |:  |  |||      |   :    : :  ||  |:  ||   :
Gfo1  47 VGVVAGIILLIYAKPILLKKALGVFII---LFVLNSLRPKKE--LQ---L 88
                   3

                 4                                5
Dre1 101 KPLMAIITVILQAFGKYRDGDFNAAGGYLYVMIIYNVSVSLSLFALFLF 149
          | |      |   | :  | | : || |||||: | :  :  |   :|
Gfo1  89 GPKMKYTFGFL---GGFFSGVF-STGGPLYVMIVKNATPDVRTFRATMF 133
                              4                          5
```

Figure S4

A

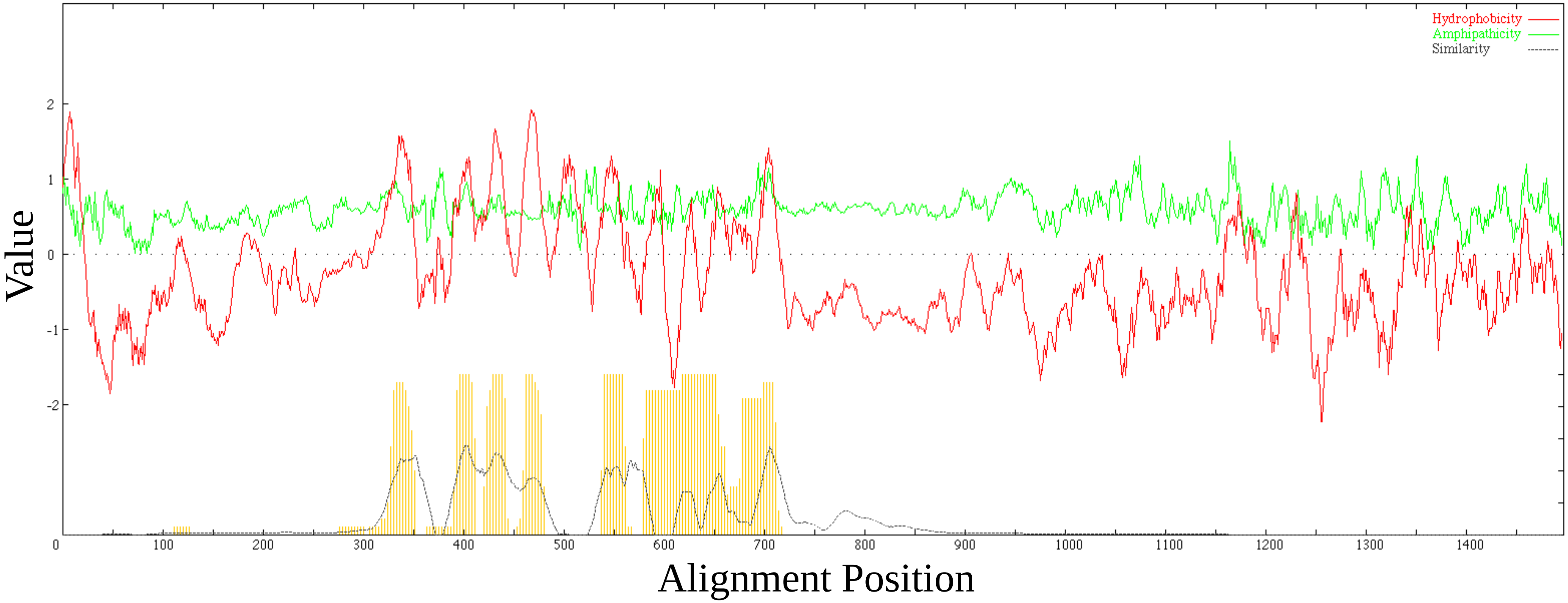

Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
Alignment Position

Figure S5

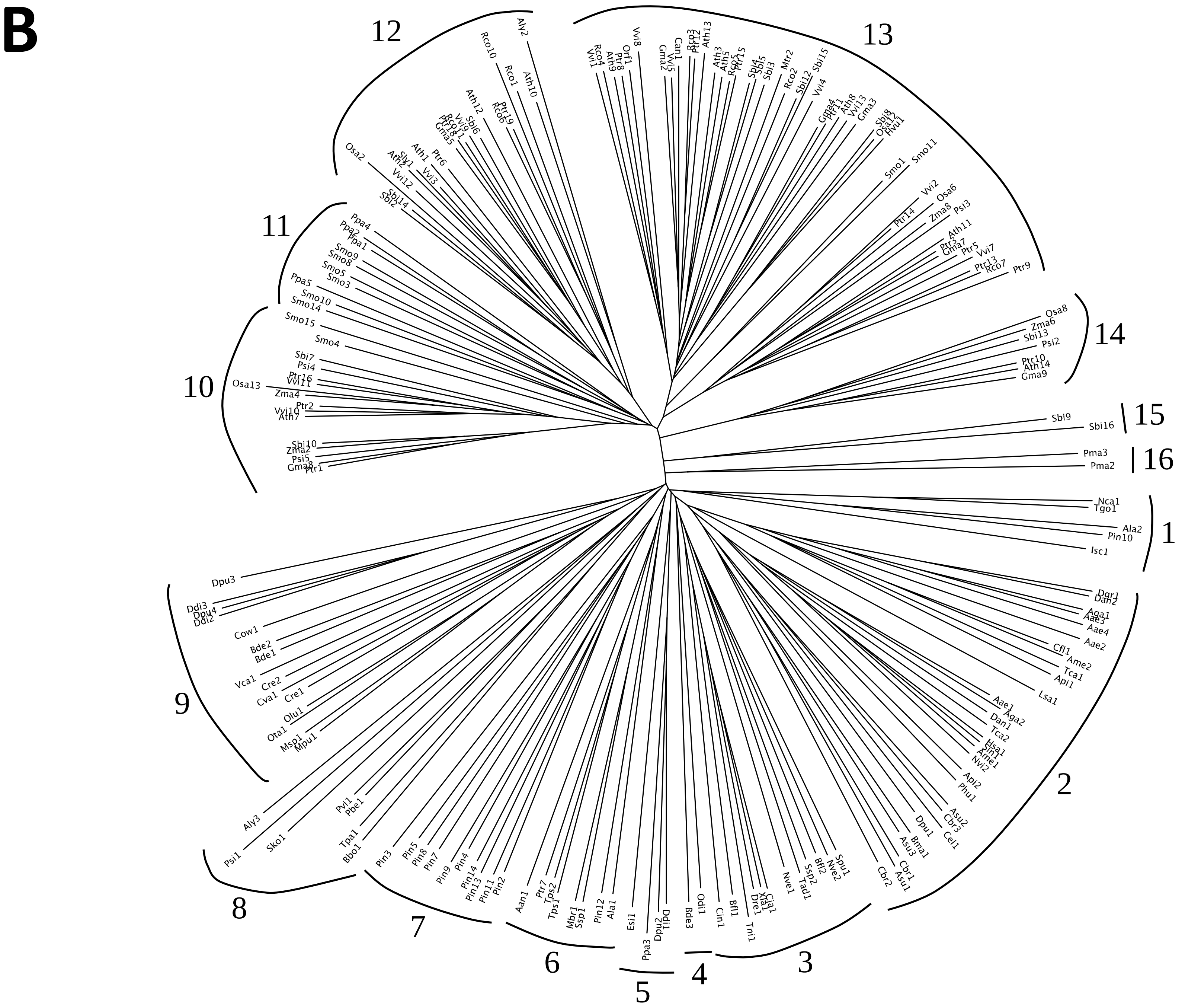

B
12
Aly2
Rco10
Rco1
Ath10
Ath12
Ptr19
Rco6
Sbi6
Vvi9
Rco11
Ptr18
Gma5
Ath1
Ptr6
Sly1
Ath2
Vvi3
Osa2
Vvi12
Sbi14
Sbi2
13
Vvi1
Rco4
Ath9
Ptr8
Orf1
Vvi8
Gma2
Vvi5
Can1
Rco3
Ptr12
Ath13
Ath3
Ath5
Rco5
Ptr15
Sbi4
Sbi5
Sbi3
Mtr2
Rco2
Sbi12
Sbi15
Vvi4
Gma4
Ptr11
Ath8
Vvi13
Gma3
Sbi8
Osa12
Hvu1
Smo1
Smo11
Vvi2
Osa6
Ptr14
Zma8
Psi3
Ath11
Ptr3
Gma7
Ptr5
Vvi7
Ptr13
Rco7
Ptr9
14
Osa8
Zma6
Sbi13
Psi2
Ptr10
Ath14
Gma9
15
Sbi9
Sbi16
16
Pma3
Pma2
1
Nca1
Tgo1
Ala2
Pin10
Isc1
2
Dgr1
Dan2
Aga1
Aae3
Aae4
Aae2
Cfl1
Ame2
Tca1
Api1
Lsa1
Aae1
Aga2
Dan1
Tca2
Hsa1
Sin1
Ame1
Nvi2
Api2
Phu1
Asu2
Cbr3
Cel1
Dpu1
Bma1
Asu3
Cbr1
Asu1
Cbr2
3
Spu1
Nve2
Bfl2
Ssp2
Tad1
Nve1
Cja1
Xla1
Dre1
Tni1
Bfl1
Cin1
Odi1
4
Bde3
5
Ddi1
Dpu2
Ppa3
6
Esi1
Ala1
Pin12
Ssp1
Mbr1
Tps1
Tps2
Ptr7
Aan1
7
Pin2
Pin11
Pin13
Pin14
Pin4
Pin9
Pin7
Pin8
Pin5
Pin3
8
Bbo1
Tpa1
Pbe1
Pvi1
Sko1
Psi1
Aly3
9
Mpu1
Msp1
Ota1
Olu1
Cre1
Cva1
Cre2
Vca1
Bde1
Bde2
Cow1
Ddi2
Dpu4
Ddi3
Dpu3
10
Ptr1
Gma8
Psi5
Zma2
Sbi10
Ath7
Vvi10
Ptr2
Zma4
Osa13
Vvi11
Ptr16
Psi4
Sbi7
Smo4
Smo15
Smo14
Smo10
11
Ppa5
Smo3
Smo5
Smo8
Smo9
Ppa1
Ppa2
Ppa4

Figure S5

C

```
              5                       6                   7
Rco4 145  VAPLAIVRKVIKTKSVEFMPFSLSFFLTLSAVMWFFYGFLKKDLYVAVPN 194
          :||:  : | :  ||: |: :     :     |  | | | |:|   : :
Ath8 249  LAPIKPLAKFLTFKSIVFLTWWQGIIVAFLFSMGLFKGSLAKELKTRIQD 298
                                  6

                   7
Rco4 195 -ILGFMFGVLQMI-LYLIYRNPKKTGDDDQKANELPNQHSIIDVAKLNTR 242
          |:    |:  :: ||:    | | |:   :   : : :: :||
Ath8 299 YIICIEMGIAAVVHLYVFPAAPYKRGERCVRNVAVMSDYASLDVPPDPEE 348
                    7

Rco4 243 VSCCEPNATTVAHSRNDREEQQ 264
         |   |   |   : |:|  |::
Ath8 349 VKDSE-RTTRTRYGRHDEREKR 369
```

Figure S5

D

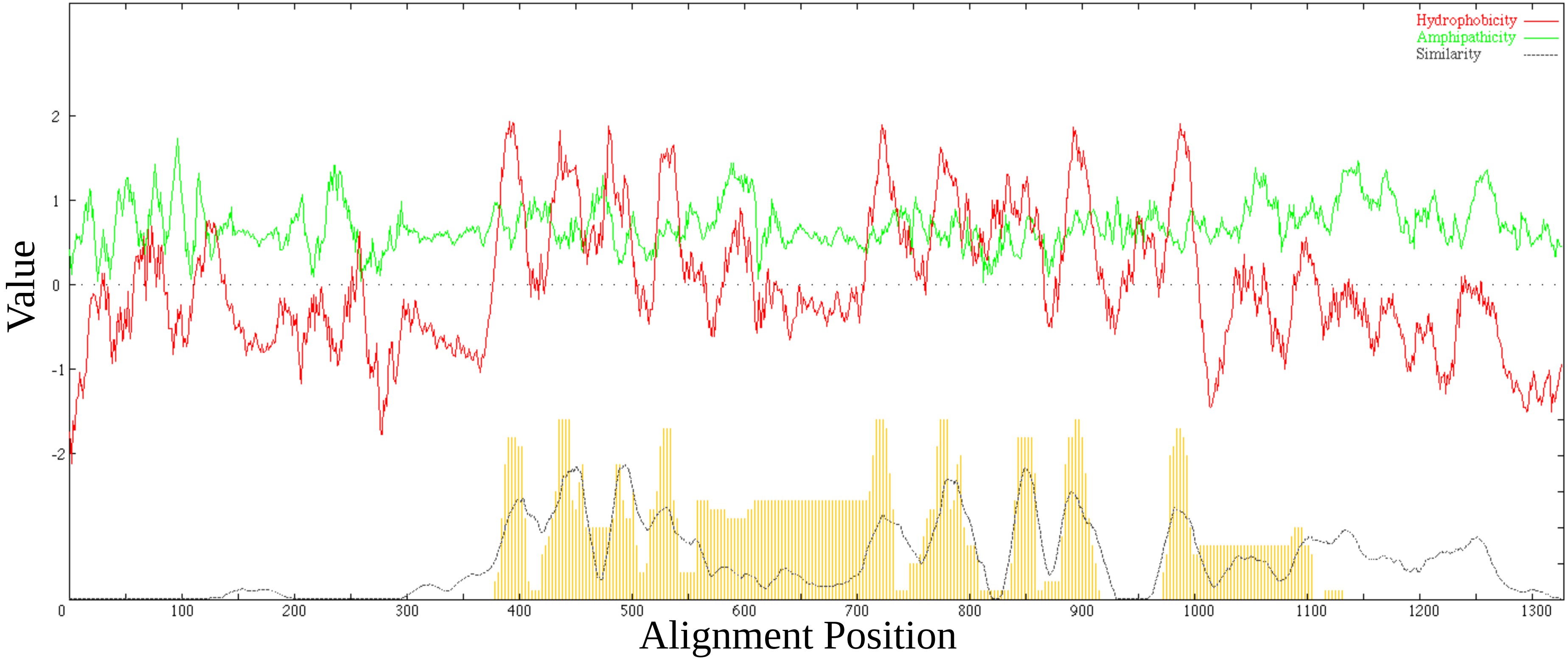
```
                          4                            5
Asu3 1   MKRRQLRFAAIE-IIFLSLIHLYVEYSQHAKEIILDHLGYIC--VAFNVA 47
         :||  |::| :: :: |: : :    | |  :| |:  ||    : :|::
Ncr1 137 IKRGILQYAWMKPVLALAAVIMKATGSYHEGDIKLNS-GYFWSGIIYNIS 185
                       4                                5

              5                                    6
Asu3 48   -TVAAPLLALGEVIRSKSTENL-PLP--LCLANLLVTSEW--LLYGFLVE 91
          ||:   |  |  |  :   :   |:|  ||:  ::  | |     | ||
Ncr1 186 VTVSLYCLGLFWVCMNNDLKPFRPMPKFLCVKAIIFASYWQGFALGILV- 234
               5                                        6

Asu3 92  DFFIKFPN 99
          |   |||
Ncr1 235 -FLGAFPN 241
             6
```

Figure S5

A



Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
Alignment Position

Figure S6

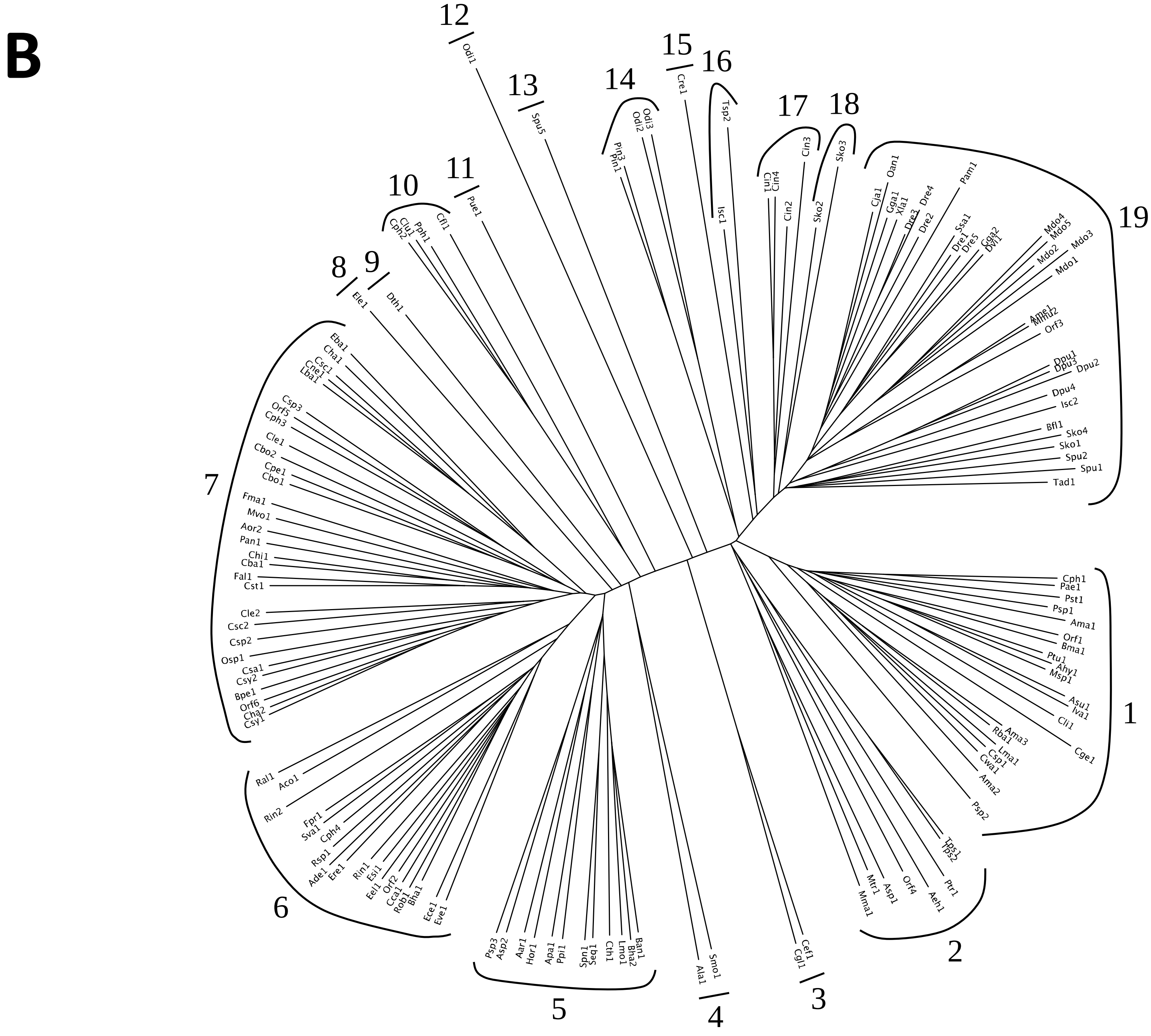

B
12
Odi1
13
Spu5
14
Pin3
Pin1
Odi2
Odi3
15
Cre1
16
Tsp2
Isc1
17
Cin1
Cin4
Cin2
Cin3
18
Sko2
Sko3
19
Cja1
Oan1
Gga1
Xla1
Dre4
Dre3
Dre2
Pam1
Ssa1
Dre1
Dre5
Gga2
Dvi1
Mdo4
Mdo5
Mdo3
Mdo2
Mdo1
Ame1
Mmu2
Orf3
Dpu1
Dpu3
Dpu2
Dpu4
Isc2
Bfl1
Sko4
Sko1
Spu2
Spu1
Tad1
1
Cph1
Pae1
Pst1
Psp1
Ama1
Orf1
Bma1
Ptu1
Ahy1
Msp1
Asu1
Iva1
Cli1
Cge1
Ama3
Rba1
Lma1
Csp1
Cwa1
Ama2
Psp2
2
Tps1
Tps2
Ptr1
Aeh1
Orf4
Asp1
Mtr1
Mma1
3
Cef1
Cgl1
4
Smo1
Ala1
5
Psp3
Asp2
Aar1
Hor1
Apa1
Ppi1
Spn1
Seq1
Cth1
Lmo1
Ban1
Bha2
6
Ral1
Aco1
Rin2
Fpr1
Sva1
Cph4
Rsp1
Ade1
Ere1
Rin1
Esi1
Eel1
Orf2
Cca1
Rob1
Bha1
Ece1
Eve1
7
Eba1
Cha1
Csc1
Cne1
Lba1
Csp3
Orf5
Cph3
Cle1
Cbo2
Cpe1
Cbo1
Fma1
Mvo1
Aor2
Pan1
Chi1
Cba1
Fal1
Cst1
Cle2
Csc2
Csp2
Osp1
Csa1
Csy2
Bpe1
Orf6
Cha2
Csy1
8
Ele1
9
Dth1
10
Cph2
Clu1
Pph1
Cfl1
11
Pue1

Figure S6

C

```
                3                         4
Hwa1 129 ALSTPMILLALGVLADVDRGSLFTVIAADIGMCVTGLGAALITSSYLFRW 177
         :: ||::  |:||:: :||    | | | ||  :||| ||   ||  ||
Odi8 239 SILTPIV--AIGVVS-IDRAYPIT-IGAHIGTTITGLIAAFSNSSGAFRK 284
                4                             5

                      5
Hwa1 178 AFYI-ISCTFFVVVLFALLFEWPVS 202
         |  : :|  || |  | | |  |::
Odi8 285 ALQVALSHVFFNVFGFTLFFLIPIT 309
                      6
```

Figure S6

A

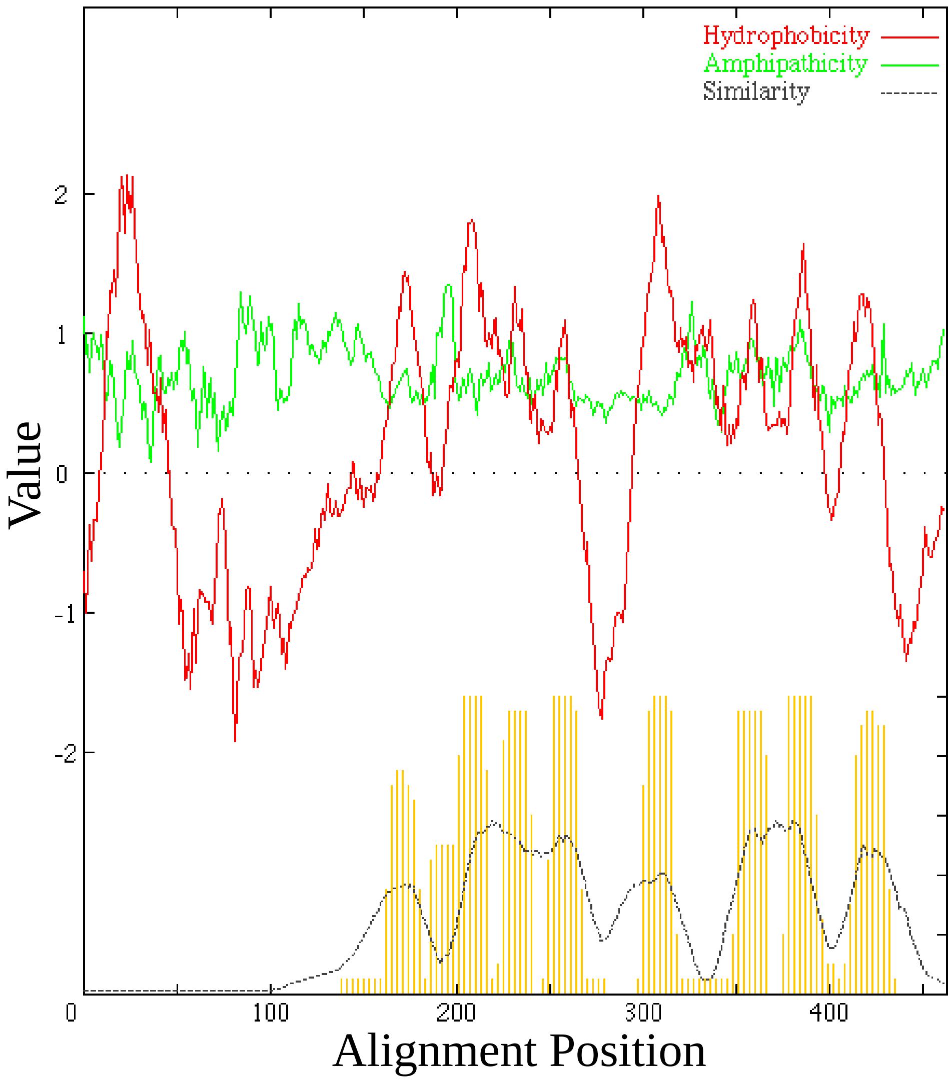

Hydrophobicity
Amphipathicity
Similarity
2
1
0
-1
-2
Value
0
100
200
300
400
Alignment Position

Figure S7

B

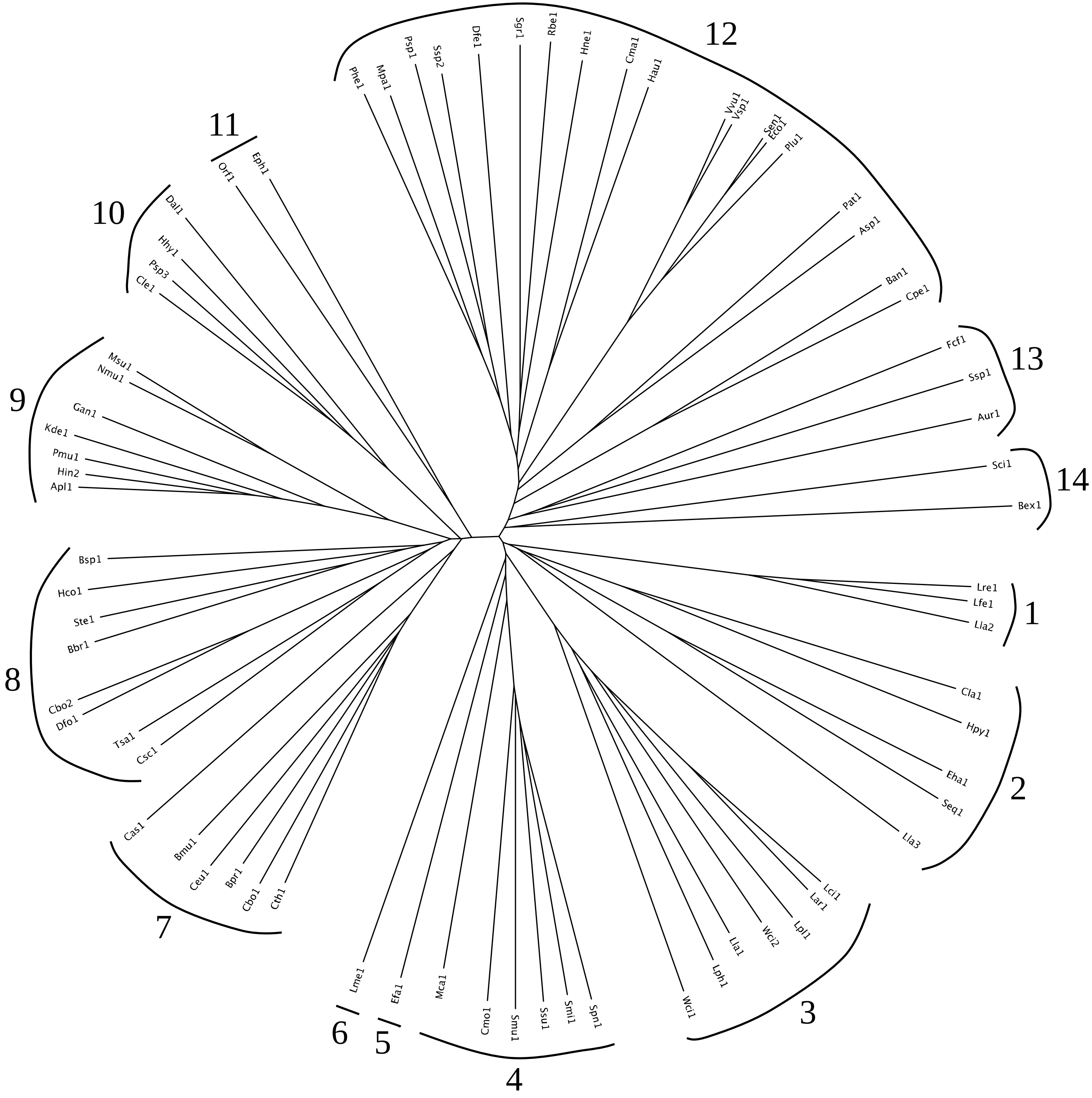

12
Phe1
Mpa1
Psp1
Ssp2
Dfe1
Sgr1
Rbe1
Hne1
Cma1
Hau1
Vvu1
Vsp1
Sen1
Eco1
Plu1
Pat1
Asp1
Ban1
Cpe1
13
Fcf1
Ssp1
Aur1
14
Sci1
Bex1
1
Lre1
Lfe1
Lla2
2
Cla1
Hpy1
Eha1
Seq1
Lla3
3
Lci1
Lar1
Lpl1
Wci2
Lla1
Lph1
Wci1
4
Spn1
Smi1
Ssu1
Smu1
Cmo1
Mca1
5
Efa1
6
Lme1
7
Cth1
Cbo1
Bpr1
Ceu1
Bmu1
Cas1
8
Csc1
Tsa1
Dfo1
Cbo2
Bbr1
Ste1
Hco1
Bsp1
9
Apl1
Hin2
Pmu1
Kde1
Gan1
Nmu1
Msu1
10
Cle1
Psp3
Hhy1
Dal1
11
Orf1
Eph1

Figure S7

C

```
        1                2                           3
Spr1 38 MRRTVWAWPVGIIGNLLLFTVFMGVWFTNPQEHSLFGQAARQVFFIATSV 87
        ::: |    |||:| : |   | |      :|  |:  |   :|  : ||
Cba4 71 LKKIVAVMAVGIVGGIFLKGFFAG------RERMLY--ALLGLFVVCLSV 112
                      3                              4

           3
Spr1 88  YGWWRWQQSRRTRAKGE 104
          | |: :  |:  |||:
Cba4 113 QGLWKLRMERQEEAKGQ 129
           4
```

Figure S7

D

```
         2                3
Sde2 1 AILFFDASLLMDSALNAYYLIMAVYGWYAWKHGKKGNEEREVSSWRMIKH 50
       |: :||  |     |  : : :|::|:|:  :|::  : :| |||: :
Ere1 1 ALAGYDAVLY--KFLGIFVIFLAIHGFYS-IYGQRQRDTKEKSSWKSMLL 47
        3                    4

                          4                      5
Sde2 51 L---KII--LFLTFLSLLIGYVMDTYTRADFAYLDSATTVFAVFTTYMLA 95
        |    |:  :|::   ||||| :    :   ::  : :||: :  : :|
Ere1 48 LIGAGIVHGMFVSGGPLLIGY-LSNEIKEKTSFRATISTVWIILNSMILV 96
                    5                             6

Sde2 96 QKI 98
        | |
Ere1 97 QDI 99
```

Figure S7

A

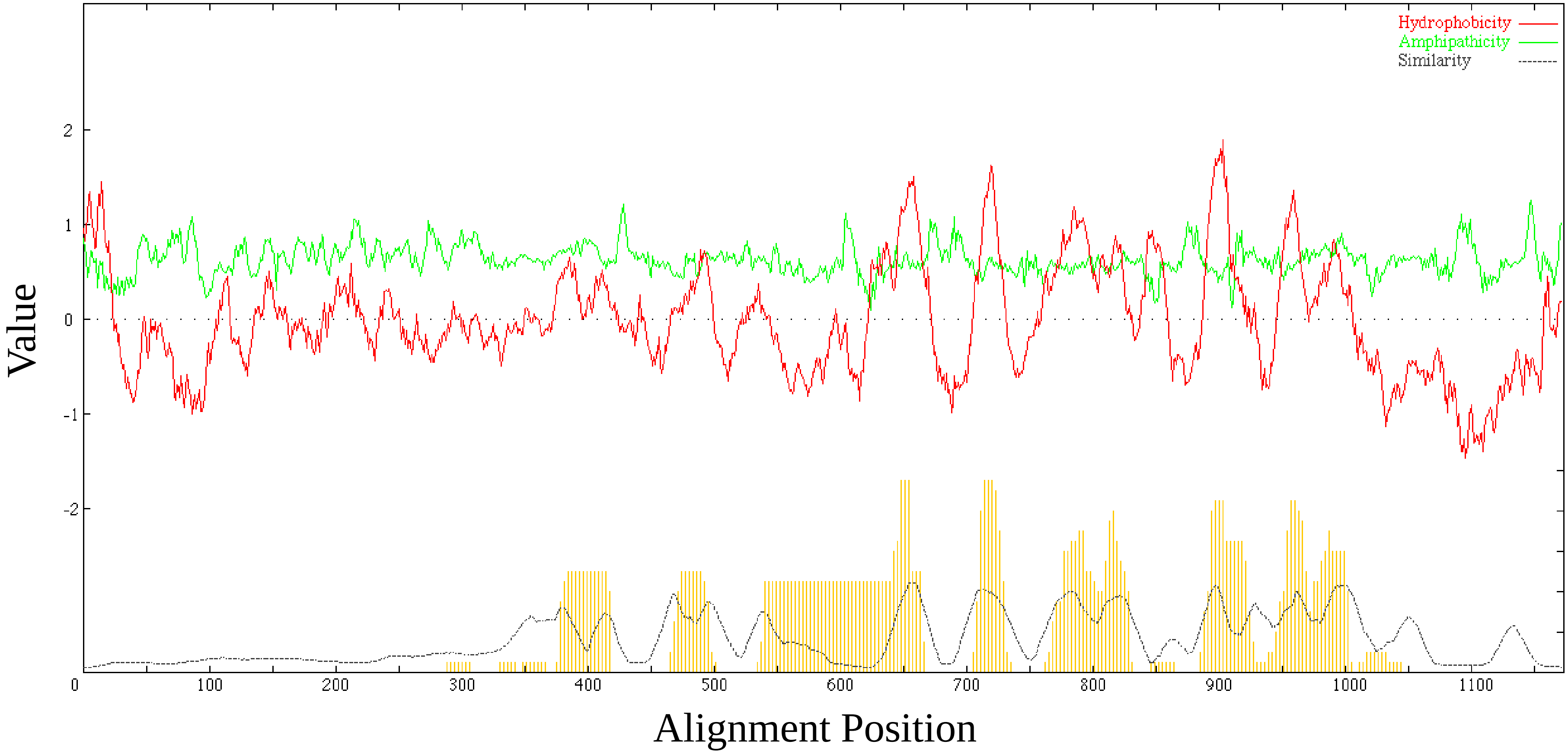

Hydrophobicity
Amphipathicity
Similarity
2
1
0
-1
-2
Value
0
100
200
300
400
500
600
700
800
900
1000
1100
Alignment Position

Figure S8

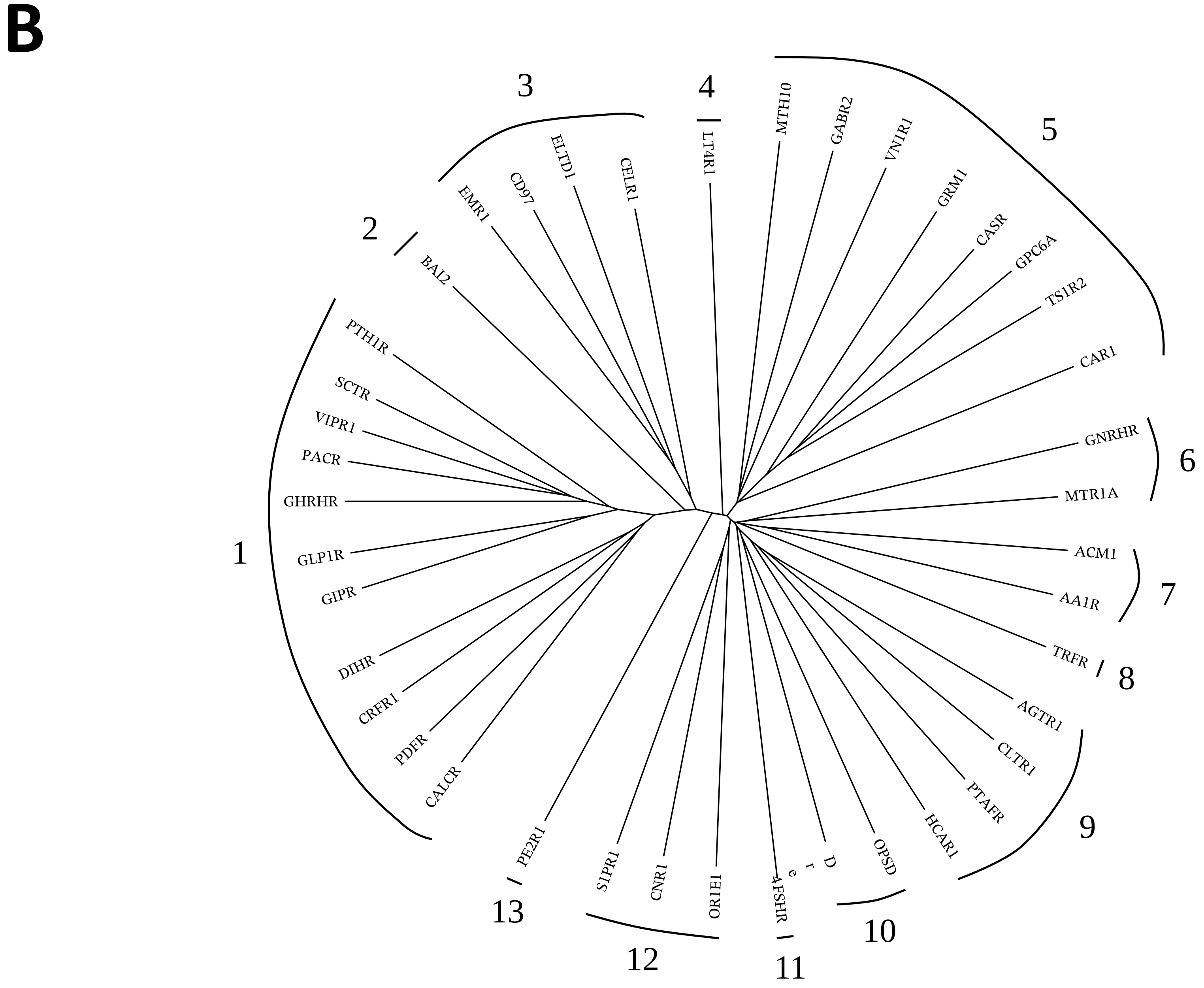

B
1
2
3
4
5
6
7
8
9
10
11
12
13
PTH1R
SCTR
VIPR1
PACR
GHRHR
GLP1R
GIPR
DIHR
CRFR1
PDFR
CALCR
BAI2
EMR1
CD97
ELTD1
CELR1
LT4R1
MTH10
GABR2
VN1R1
GRM1
CASR
GPC6A
TS1R2
CAR1
GNRHR
MTR1A
ACM1
AA1R
TRFR
AGTR1
CLTR1
PTAFR
HCAR1
OPSD
D
r
e
4FSHR
OR1E1
CNR1
S1PR1
PE2R1

Figure S8

C

```
                 5                               6
Dre1 183 GLF-ISLPIIAYCYLRVLMVVKKTKNSKKNRAIKLILGIVIMFVVFWVPY 230
         ||| ||     |    :|   : |   | :|  || ||| :   | |: |
Mos1 165 GLFTISCLFFLYVLWGLLFSARNTAFGKSDRVGKLYLGIAVYTAVLWIAY 214
                   5                               6

            6
Dre1 231 NVV-VFLKTLHEFDMLTSCEPYKIINMA 258
          ||  | :  :     |    | |:::|
Mos1 215 PVVWAFAEGTNSITGDTEVLAYAILDIA 242
           6                     7
```

Figure S8

D

```
                                                   1
Dre5  41 GPTHSHHHD-PPHPFPTVDVPDHAHYI-IGSVILIVGITGVIGNALVVYV 88
         || ||||   |  |||:   |   |   :| | : |  |      ||  :
Cga1  27 GPHHSHHLPLPTAPFPSA-TPKFLHATRVGHVSIWV-FTAFFIAGLVAVL 74
                                                    1

                                  2
Dre5  89 FCRSRTLRTAGNMFIVNLAVADFLMSVTQSPVFFAASLHRRWV--FGERP 136
         |  ||| |    :|    ||   : ::|   :  |  :   :|  :|  |
Cga1  75 FLTSRTER-KNRLFHGISAVILTVSALTY--MSLATHIGSTFVPIYGP-P 120
                                  2

                           3
Dre5  137 CE---LYAFCGALFGICSMMTLTAIAADRCLAITQPLALVSRVSRRKAG- 182
               |  |   :| |       ||     :     || ||  :   |
Cga1  121 GHHEPLVHFFRQVFSI--RYIDAAITGPLIILALSRLAGVSPATALSAAL 168
                                652
                                3

                     4
Dre5 183 AVLVVV---WLYSLGWSLPPFFGWSAYVPEGLQTSCSWDYMTFTPSVRAY 229
         | ||||   |  |:|   |    || :  :|  |  :|  :     :  :
Cga1 169 AELVVVYSAWAGSVGGGWP----WSEH-GKGAGTKWAWFAVAILAFLAVW 213
              4                                       5

            5
Dre5 230 TILL 233
         |:||
Cga1 214 TVLL 217
```

Figure S8

E

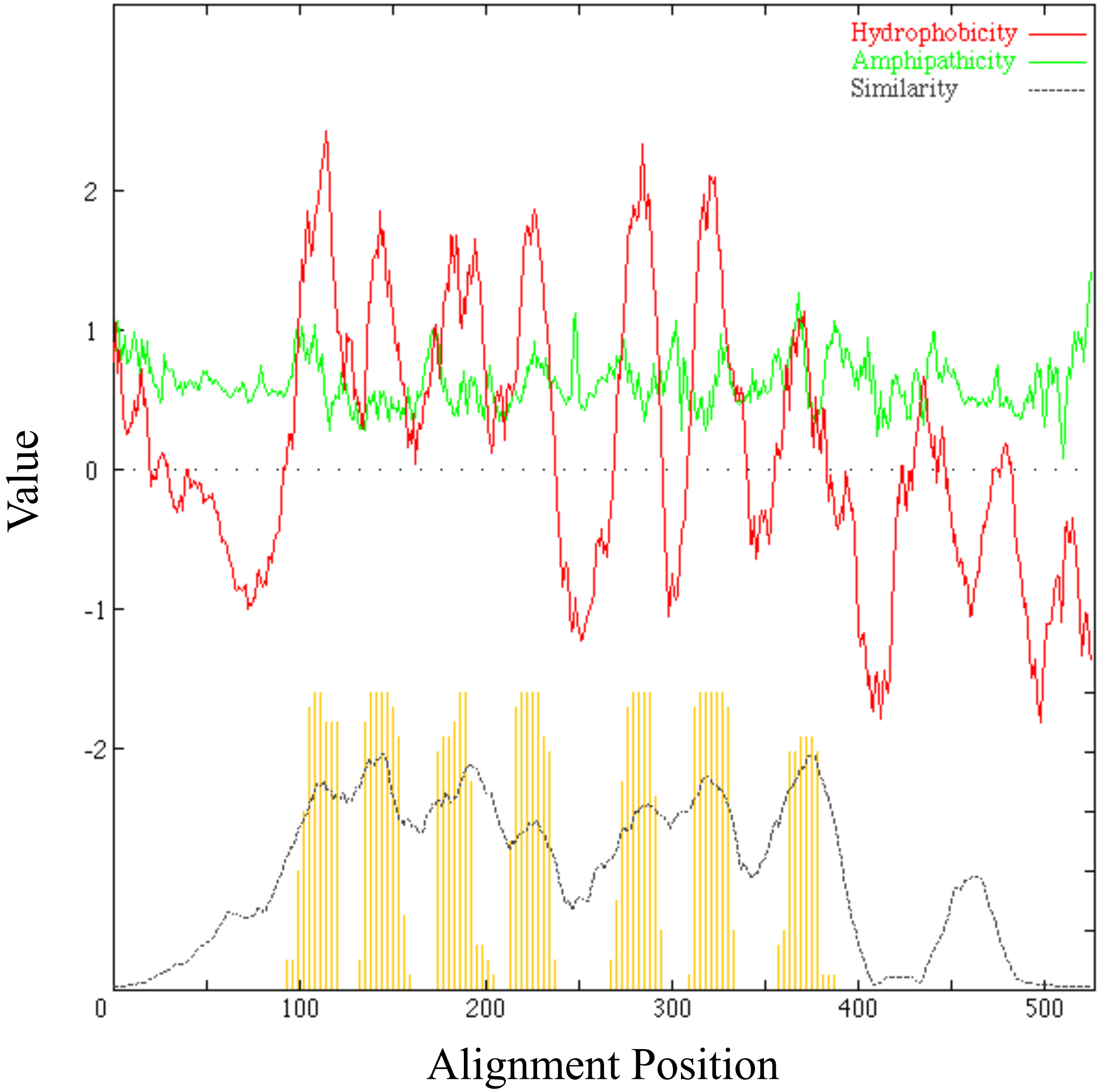
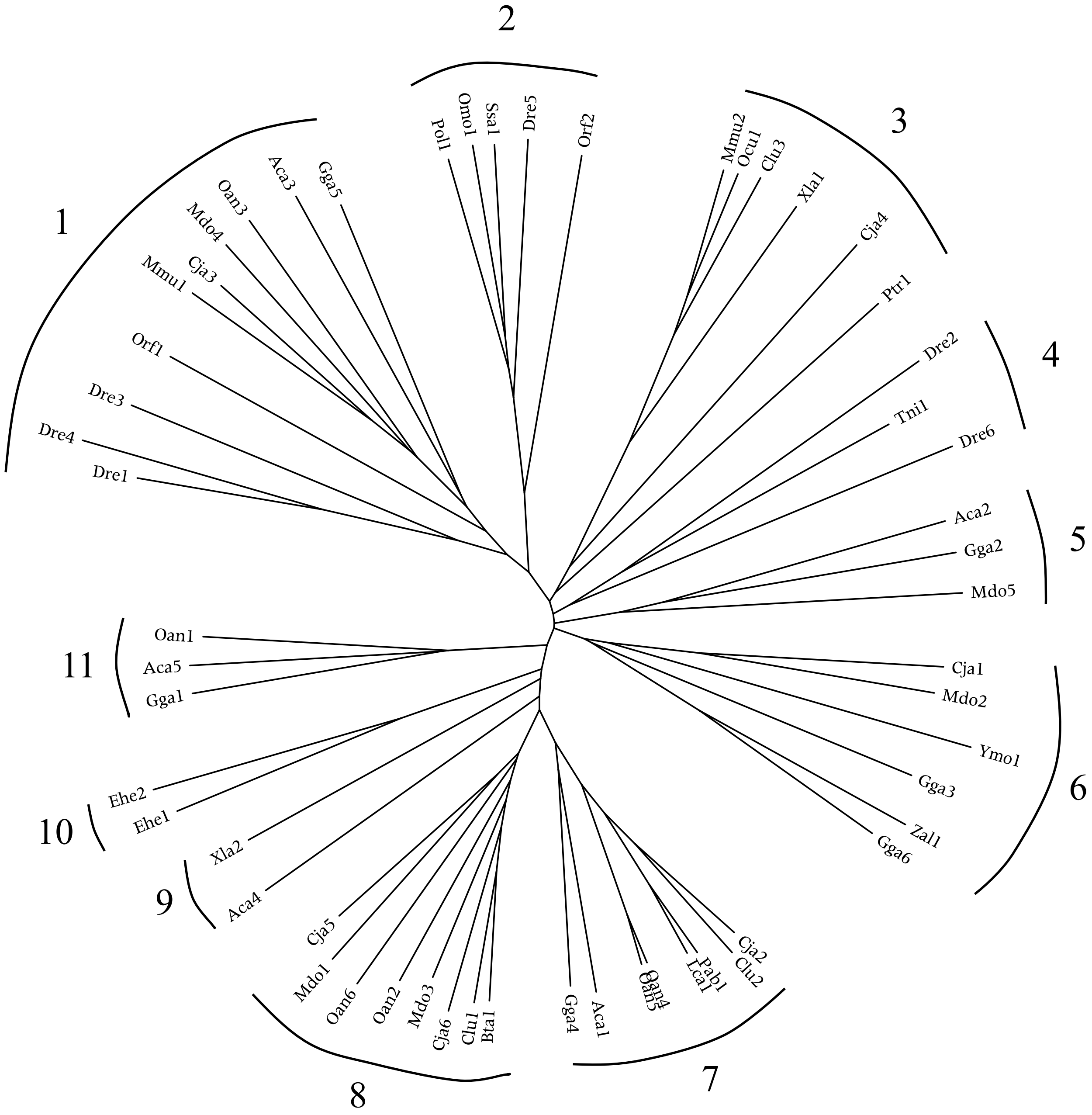

Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
Alignment Position

Figure S8

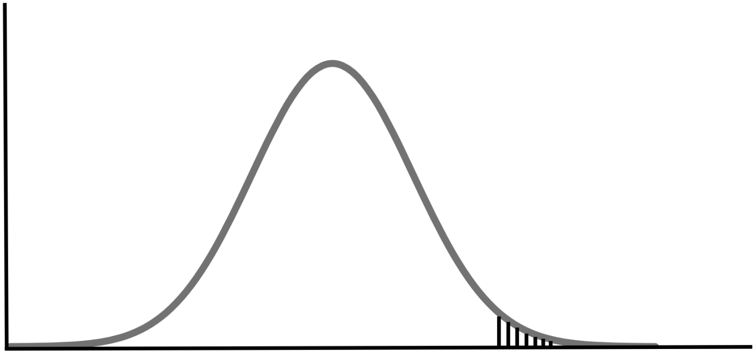
F



1
Dre1
Dre4
Dre3
Orf1
Mmu1
Cja3
Mdo4
Oan3
Aca3
Gga5
2
Pol1
Omo1
Ssa1
Dre5
Orf2
3
Mmu2
Ocu1
Clu3
Xla1
Cja4
Ptr1
4
Dre2
Tni1
Dre6
5
Aca2
Gga2
Mdo5
6
Cja1
Mdo2
Ymo1
Gga3
Zal1
Gga6
7
Cja2
Clu2
Pab1
Lca1
Oan4
Oan5
Aca1
Gga4
8
Bta1
Clu1
Cja6
Mdo3
Oan2
Oan6
Mdo1
Cja5
9
Aca4
Xla2
10
Ehe1
Ehe2
11
Oan1
Aca5
Gga1

Figure S8

A)



$\bar{x}$ = Mean bit score of the 2,000 shuffled NW (Needleman-Wunsch) scores of two proteins’ aa residues

x = Raw unshuffled NW bitscore of two proteins

σ = SD of the 2,000 shuffled NW scores of two proteins’ aa residues

$Z = \frac{x - \bar{x}}{\sigma}$ = Standard score

B) NIntegrate[(1/Sqrt[2*Pi])*E^(-.5*x^2), {x, (z score), Infinity}]

C)

$12.0 = 1.77648 * 10^{-33}$

$10.5 = 4.31901 * 10^{-26}$

$9.2 = 1.78975 * 10^{-20}$

Figure S9

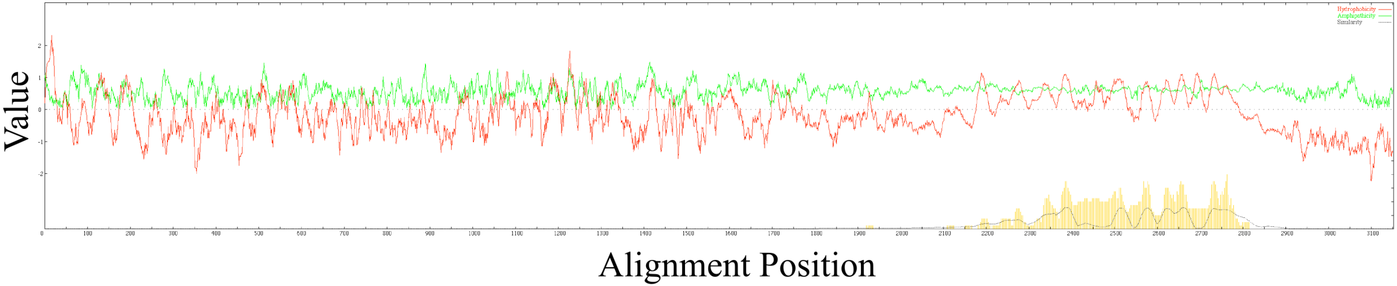

Hydrophobicity
Amphipathicity
Similarity
Value
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
Alignment Position

Figure S10

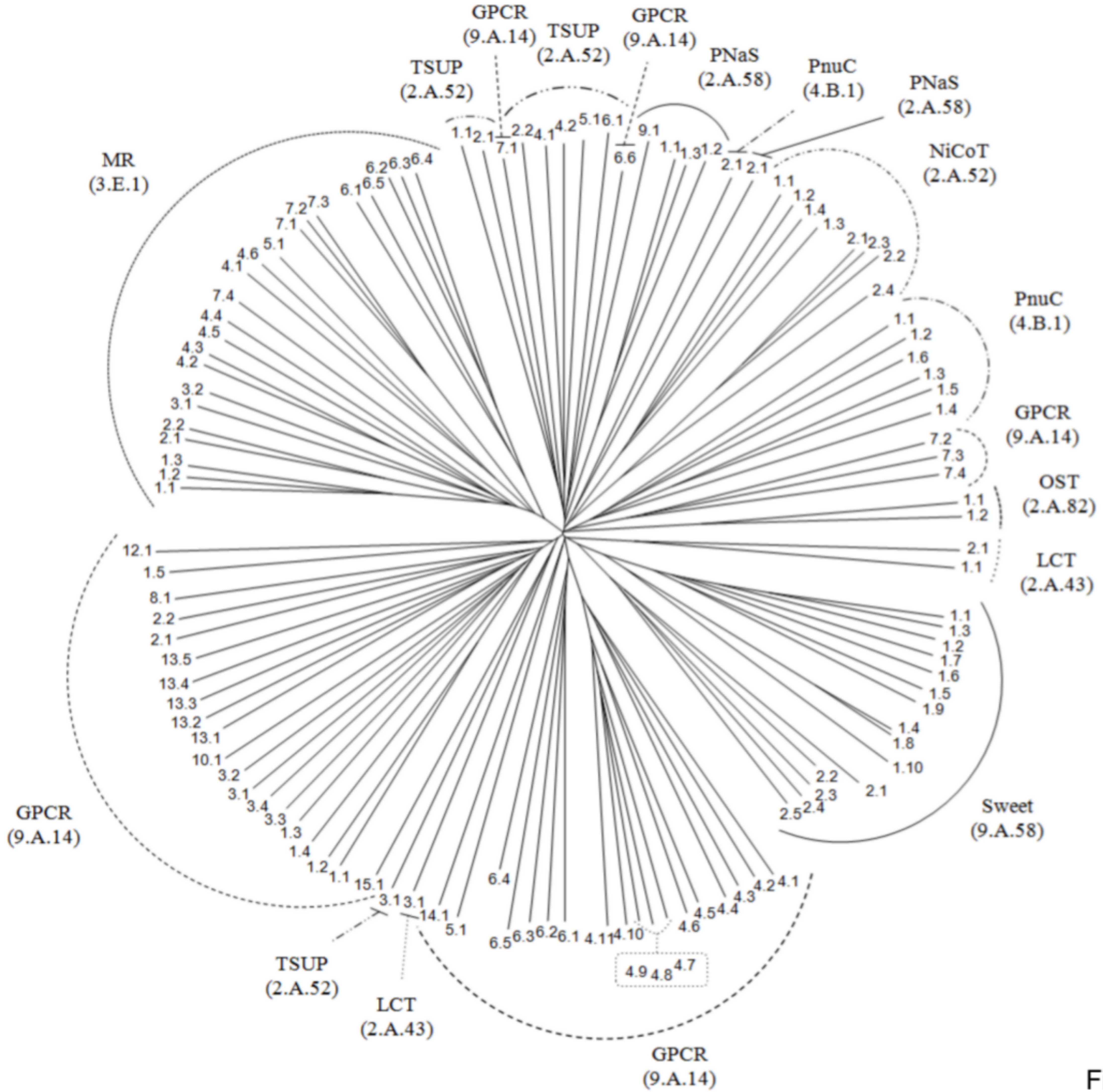

GPCR (9.A.14)
TSUP (2.A.52)
GPCR (9.A.14)
TSUP (2.A.52)
PNaS (2.A.58)
PnuC (4.B.1)
PNaS (2.A.58)
NiCoT (2.A.52)
PnuC (4.B.1)
GPCR (9.A.14)
OST (2.A.82)
LCT (2.A.43)
Sweet (9.A.58)
GPCR (9.A.14)
LCT (2.A.43)
TSUP (2.A.52)
GPCR (9.A.14)
MR (3.E.1)
1.1 2.1 7.1 2.2 4.1 4.2 5.1 6.1 6.6 9.1 1.1 1.3 1.2 2.1 2.1
1.1 1.2 1.4 1.3 2.1 2.3 2.2 2.4
1.1 1.2 1.6 1.3 1.5 1.4
7.2 7.3 7.4
1.1 1.2
2.1 1.1
1.1 1.3 1.2 1.7 1.6 1.5 1.9 1.4 1.8 1.10 2.1 2.2 2.3 2.4 2.5
4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 4.9 4.10 4.11 6.1 6.2 6.3 6.5 6.4 5.1 14.1 3.1 3.1 15.1
1.1 1.2 1.4 1.3 3.3 3.4 3.1 3.2 10.1 13.1 13.2 13.3 13.4 13.5 2.1 2.2 8.1 1.5 12.1
1.1 1.2 1.3 2.1 2.2 3.1 3.2 4.2 4.3 4.5 4.4 7.4 4.1 4.6 5.1 7.1 7.2 7.3 6.1 6.5 6.2 6.3 6.4

Figure S11

Table S1. 135 LCT protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S1). Protein abbreviations, taxonomic origins, protein sizes (aas), gi numbers, organismal phyla and organismal domains are included. Average size and standard deviation values are provided for each cluster having over 1 member. The same convention is used for subsequent tables listing family homologues.

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Cva1	*Chlorella variabilis*	307104348	Viridiplantae	Eukaryota	321
Cva2	*Chlorella variabilis*	307104349	Viridiplantae	Eukaryota	407
Ala1	*Albugo laibachii Nc14*	325182632	Albuginaceae	Eukaryota	314
Pin1	*Phytophthora infestans T30-4*	301113472	none	Eukaryota	301
Tcr1	*Trypanosoma cruzi*	322825613	Trypanosomatidae	Eukaryota	383
Esi1	*Ectocarpus siliculosus*	298705956	Phaeophyceae	Eukaryota	307
Aan2	*Aureococcus anophagefferens*	323456637	Pelagophyceae	Eukaryota	273
Ptr2	*Phaeodactylum tricornutum CCAP 1055/1*	219112913	Bacillariophyta	Eukaryota	252
Average protein size (aas)					320
SD					52
Cluster 2					
Bde1	*Batrachochytrium dendrobatidis JAM81*	328771916	Fungi	Eukaryota	294
Cbr1	*Caenorhabditis briggsae*	268530370	Metazoa	Eukaryota	403
Cre1	*Caenorhabditis remanei*	308510650	Metazoa	Eukaryota	404
Cel1	*Caenorhabditis elegans*	17532373	Metazoa	Eukaryota	404
Asu1	*Ascaris suum*	324511247	Metazoa	Eukaryota	421
Average protein size (aas)					385
SD					52
Cluster 3					
Ssc1	*Sus scrofa*	311268141	Metazoa	Eukaryota	373
Clu1	*Canis lupus familiaris*	73967293	Metazoa	Eukaryota	388
Cja1	*Callithrix jacchus*	296201066	Metazoa	Eukaryota	400
Mmu1	*Mus musculus*	13752577	Metazoa	Eukaryota	367
Oan1	*Ornithorhynchus anatinus*	149641786	Metazoa	Eukaryota	605
Rno1	*Rattus norvegicus*	109465722	Metazoa	Eukaryota	350

Xla1	*Xenopus laevis*	147899595	Metazoa	Eukaryota	375
Orf1	*Xenopus (Silurana) tropicalis*	118404708	Metazoa	Eukaryota	376
Dre1	*Danio rerio*	66472608	Metazoa	Eukaryota	384
Tni1	*Tetraodon nigroviridis*	47213130	Metazoa	Eukaryota	404
Average protein size (aas)					402
SD					73
Cluster 4					
Smo1	*Selaginella moellendorffii*	302822928	Viridiplantae	Eukaryota	280
Ppa4	*Physcomitrella patens subsp. patens*	168022702	Viridiplantae	Eukaryota	291
Vvi1	*Vitis vinifera*	225452224	Viridiplantae	Eukaryota	274
Vvi2	*Vitis vinifera*	296081320	Viridiplantae	Eukaryota	255
Ptr1	*Populus trichocarpa*	224141271	Viridiplantae	Eukaryota	274
Rco1	*Ricinus communis*	255571996	Viridiplantae	Eukaryota	244
Ath1	*Arabidopsis thaliana*	21553944	Viridiplantae	Eukaryota	270
Hvu1	*Hordeum vulgare subsp. vulgare*	326491617	Viridiplantae	Eukaryota	275
Osa1	*Oryza sativa Japonica Group*	115488162	Viridiplantae	Eukaryota	268
Psi1	*Picea sitchensis*	116784154	Viridiplantae	Eukaryota	272
Average protein size (aas)					270
SD					13
Cluster 5					
Tad1	*Trichoplax adhaerens*	196011156	Metazoa	Eukaryota	262
Cow1	*Capsaspora owczarzaki ATCC 30864*	320169718	Ichthyosporea	Eukaryota	389
Ssp2	*Salpingoeca sp. ATCC 50818*	326433282	Salpingoecidae	Eukaryota	413
Tad2	*Trichoplax adhaerens*	196001949	Metazoa	Eukaryota	335
Nve1	*Nematostella vectensis*	156400754	Metazoa	Eukaryota	259
Cin1	*Ciona intestinalis*	198418241	Metazoa	Eukaryota	399
Sko1	*Saccoglossus kowalevskii*	291224563	Metazoa	Eukaryota	384
Spu2	*Strongylocentrotus*	115676820	Metazoa	Eukaryota	369

	purpuratus				
Average protein size (aas)					351
SD					61
Cluster 6					
Dfa1	*Dictyostelium fasciculatum*	328875158	none	Eukaryota	689
Ppa3	*Polysphondylium pallidum PN500*	281212356	none	Eukaryota	271
Dpu2	*Dictyostelium purpureum*	330801850	none	Eukaryota	285
Ddi1	*Dictyostelium discoideum AX4*	66815053	none	Eukaryota	284
Dpu3	*Dictyostelium purpureum*	330801852	none	Eukaryota	288
Average protein size (aas)					363
SD					182
Cluster 7					
Isc1	*Ixodes scapularis*	240999715	Metazoa	Eukaryota	266
Bfl1	*Branchiostoma floridae*	260811948	Metazoa	Eukaryota	432
Aae1	*Aedes aegypti*	157112357	Metazoa	Eukaryota	406
Cqu2	*Culex quinquefasciatus*	170044802	Metazoa	Eukaryota	399
Ada1	*Anopheles darlingi*	312385713	Metazoa	Eukaryota	524
Aga1	*Anopheles gambiae str. PEST*	333468595	Metazoa	Eukaryota	461
Aga4	*Anopheles gambiae str. PEST*	333468594	Metazoa	Eukaryota	376
Dmo1	*Drosophila mojavensis*	195113659	Metazoa	Eukaryota	354
Dvi1	*Drosophila virilis*	195399488	Metazoa	Eukaryota	375
Dgr1	*Drosophila grimshawi*	195054092	Metazoa	Eukaryota	914
Der1	*Drosophila erecta*	194910616	Metazoa	Eukaryota	393
Dan1	*Drosophila ananassae*	194743940	Metazoa	Eukaryota	392
Dpe1	*Drosophila persimilis*	195166302	Metazoa	Eukaryota	389
Dwi1	*Drosophila willistoni*	195450813	Metazoa	Eukaryota	361
Aec1	*Acromyrmex echinatior*	332023894	Metazoa	Eukaryota	415
Cfl1	*Camponotus floridanus*	307175251	Metazoa	Eukaryota	464

Ame1	*Apis mellifera*	328776401	Metazoa	Eukaryota	463
Nvi1	*Nasonia vitripennis*	156553286	Metazoa	Eukaryota	1165
Tca2	*Tribolium castaneum*	270014202	Metazoa	Eukaryota	367
Dpu1	*Daphnia pulex*	321471968	Metazoa	Eukaryota	252
Api1	*Acyrthosiphon pisum*	328710507	Metazoa	Eukaryota	403
Phu1	*Pediculus humanus corporis*	242006380	Metazoa	Eukaryota	356
Ccl1	*Caligus clemensi*	225717886	Metazoa	Eukaryota	409
Average protein size (aas)					449
SD					199
Cluster 8					
Aga3	*Anopheles gambiae str. PEST*	333470136	Metazoa	Eukaryota	376
Aga5	*Anopheles gambiae str. PEST*	158301831	Metazoa	Eukaryota	315
Aga2	*Anopheles gambiae str. PEST*	333470137	Metazoa	Eukaryota	372
Aae2	*Aedes aegypti*	157167697	Metazoa	Eukaryota	367
Cqu1	*Culex quinquefasciatus*	170032456	Metazoa	Eukaryota	560
Ngr1	*Naegleria gruberi strain NEG-M*	290991005	Heterolobosea	Eukaryota	317
Average protein size (aas)					385
SD					90
Cluster 9					
Odi1	*Oikopleura dioica*	313239649	Metazoa	Eukaryota	419
Tsp1	*Trichinella spiralis*	316977368	Metazoa	Eukaryota	430
Average protein size (aas)					425
SD					8
Cluster 10					
Sma1	*Schistosoma mansoni*	256087309	Metazoa	Eukaryota	359
Pgr1	*Puccinia graminis f. sp. tritici CRL 75-36-700-3*	331237809	Fungi	Eukaryota	276
Mla1	*Melampsora larici-populina 98AG31*	328855007	Fungi	Eukaryota	302
Average protein size (aas)					312
SD					42
Cluster 11					

Acl1	*Aspergillus clavatus NRRL 1*	121703123	Fungi	Eukaryota	266
Ate1	*Aspergillus terreus NIH2624*	115388391	Fungi	Eukaryota	284
Ani2	*Aspergillus niger CBS 513.88*	145229229	Fungi	Eukaryota	288
Afl1	*Aspergillus flavus NRRL3357*	238483409	Fungi	Eukaryota	284
Ani1	*Aspergillus nidulans FGSC A4*	67515987	Fungi	Eukaryota	284
Tst1	*Talaromyces stipitatus ATCC 10500*	242804336	Fungi	Eukaryota	293
Pch2	*Penicillium chrysogenum Wisconsin 54-1255*	255951360	Fungi	Eukaryota	288
Aca2	*Ajellomyces capsulatus NAm1*	154280254	Fungi	Eukaryota	306
Ade1	*Ajellomyces dermatitidis ER-3*	239609412	Fungi	Eukaryota	297
Cpo1	*Coccidioides posadasii C735 delta SOWgp*	303323033	Fungi	Eukaryota	282
Agy1	*Arthroderma gypseum CBS 118893*	315044173	Fungi	Eukaryota	269
Aot2	*Arthroderma otae CBS 113480*	296811498	Fungi	Eukaryota	275
Lma1	*Leptosphaeria maculans*	312214508	Fungi	Eukaryota	282
Ptr3	*Pyrenophora tritici-repentis Pt-1C-BFP*	189202562	Fungi	Eukaryota	281
Pno1	*Phaeosphaeria nodorum SN15*	169596154	Fungi	Eukaryota	293
Tme1	*Tuber melanosporum Mel28*	296422912	Fungi	Eukaryota	375
Aot1	*Arthroderma otae CBS 113480*	296806035	Fungi	Eukaryota	279
Agy2	*Arthroderma gypseum CBS 118893*	315039707	Fungi	Eukaryota	341
Acl2	*Aspergillus clavatus NRRL 1*	121708480	Fungi	Eukaryota	274
Afu2	*Aspergillus fumigatus Af293*	70992957	Fungi	Eukaryota	280

Pch1	*Penicillium chrysogenum Wisconsin 54-1255*	255951729	Fungi	Eukaryota	272
Aor1	*Aspergillus oryzae RIB40*	317137066	Fungi	Eukaryota	272
Gze1	*Gibberella zeae PH-1*	46125503	Fungi	Eukaryota	304
Nha1	*Nectria haematococca mpVI 77-13-4*	302910912	Fungi	Eukaryota	274
Mac1	*Metarhizium acridum CQMa 102*	322701501	Fungi	Eukaryota	281
Val1	*Verticillium albo-atrum VaMs.102*	302421044	Fungi	Eukaryota	277
Ggr1	*Glomerella graminicola M1.001*	310792616	Fungi	Eukaryota	278
Cgl1	*Chaetomium globosum CBS 148.51*	116206758	Fungi	Eukaryota	278
Ncr1	*Neurospora crassa OR74A*	164426091	Fungi	Eukaryota	297
Mor1	*Magnaporthe oryzae 70-15*	39971505	Fungi	Eukaryota	279
Gcl1	*Grosmannia clavigera kw1407*	320589073	Fungi	Eukaryota	293
Ssc2	*Sclerotinia sclerotiorum 1980 UF-70*	156041170	Fungi	Eukaryota	281
Bba2	*Beauveria bassiana*	400602576	Fungi	Eukaryota	281
Average protein size (aas)					288
SD					21
Cluster 12					
Tps1	*Thalassiosira pseudonana CCMP1335*	224013618	Bacillariophyta	Eukaryota	259
Tps2	*Thalassiosira pseudonana CCMP1335*	224010223	Bacillariophyta	Eukaryota	308
Uma1	*Ustilago maydis 521*	71023203	Fungi	Eukaryota	310
Cne1	*Cryptococcus neoformans var. neoformans JEC21*	58258347	Fungi	Eukaryota	284
Cga1	*Cryptococcus gattii*	321251307	Fungi	Eukaryota	304

	WM276				
Average protein size (aas)					293
SD					22
Cluster 13					
Odi2	*Oikopleura dioica*	313233450	Metazoa	Eukaryota	561
Cluster 14					
Yli1	*Yarrowia lipolytica CLIB122*	210075509	Fungi	Eukaryota	250
Ppa2	*Pichia pastoris GS115*	254568728	Fungi	Eukaryota	311
Zro1	*Zygosaccharomyces rouxii CBS 732*	254581598	Fungi	Eukaryota	266
Cgl2	*Candida glabrata CBS 138*	50294480	Fungi	Eukaryota	249
Sce1	*Saccharomyces cerevisiae RM11-1a*	190406495	Fungi	Eukaryota	260
Vpo1	*Vanderwaltozyma polyspora DSM 70294*	156836975	Fungi	Eukaryota	259
Lth1	*Lachancea thermotolerans CBS 6340*	255710947	Fungi	Eukaryota	249
Kla1	*Kluyveromyces lactis NRRL Y-1140*	50312531	Fungi	Eukaryota	250
Ago1	*Ashbya gossypii ATCC 10895*	302308260	Fungi	Eukaryota	247
Ppa1	*Pichia pastoris GS115*	254571885	Fungi	Eukaryota	260
Average protein size (aas)					260
SD					19
Cluster 15					
Pte2	*Paramecium tetraurelia strain d4-2*	145475843	Oligohymenophorea	Eukaryota	246
Pte3	*Paramecium tetraurelia strain d4-2*	145494574	Oligohymenophorea	Eukaryota	274
Pte1	*Paramecium tetraurelia strain d4-2*	145540144	Oligohymenophorea	Eukaryota	252
Tth1	*Tetrahymena thermophila*	118398647	Oligohymenophorea	Eukaryota	267
Aan1	*Aureococcus*	323453704	Pelagophyceae	Eukaryota	3342

	anophagefferens				
Pma2	*Perkinsus marinus ATCC 50983*	294905260	Perkinsidae	Eukaryota	302
Ssp1	*Salpingoeca sp. ATCC 50818*	326433249	Salpingoecidae	Eukaryota	332
Average protein size (aas)					716
SD					1158

Table S2. 104 MR protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S2).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Hma1	*Haloarcula marismortui ATCC 43049*	55378744	Euryarchaeota	Archaea	250
Hsa3	*Halobacterium salinarum*	11992133	Euryarchaeota	Archaea	255
Hsa1	*Halobacterium salinarum*	46015682	Euryarchaeota	Archaea	227
Hva2	*Haloarcula vallismortis*	2829812	Euryarchaeota	Archaea	250
Hwa2	*Haloquadratum walsbyi DSM 16790*	110666993	Euryarchaeota	Archaea	246
Rxy1	*Rubrobacter xylanophilus DSM 9941*	108804857	Actinobacteria	Bacteria	239
Kra1	*Ktedonobacter racemifer DSM 44963*	298250763	Chloroflexi	Bacteria	242
Ssp2	*Sphingomonas sp. S17*	332185638	Alphaproteobacteria	Bacteria	234
Psp2	*Pantoea sp. aB*	304398913	Gammaproteobacteria	Bacteria	237
Msp1	*Methylobacterium sp. 4-46*	170738878	Alphaproteobacteria	Bacteria	252
Average protein size (aas)					243
SD					9
Cluster 2					
Hsp1	*Halobacterium sp. SG1*	461609	Euryarchaeota	Archaea	284

Hva3	*Haloarcula vallismortis*	2829811	Euryarchaeota	Archaea	276
Nph2	*Natronomonas pharaonis*	114809	Euryarchaeota	Archaea	291
Sru2	*Salinibacter ruber DSM 13855*	83814666	Bacteroidetes	Bacteria	249
Hsa2	*Halobacterium salinarum*	8569313	Euryarchaeota	Archaea	253
Hwa1	*Haloquadratum walsbyi DSM 16790*	110668951	Euryarchaeota	Archaea	292
Hsp5	*Haloterrigena sp. arg-4*	14194474	Euryarchaeota	Archaea	297
Nma1	*Natrialba magadii ATCC 43099*	289582237	Euryarchaeota	Archaea	279
Average protein size (aas)					278
SD					18
Cluster 3					
Hmu1	*Halomicrobium mukohataei DSM 12286*	257388720	Euryarchaeota	Archaea	243
Hma3	*Haloarcula marismortui ATCC 43049*	55376955	Euryarchaeota	Archaea	236
Nph1	*Natronomonas pharaonis*	1168615	Euryarchaeota	Archaea	239
Hsp4	*Halobacterium sp. AUS-1*	19716034	Euryarchaeota	Archaea	249
Hsp6	*Halobacterium sp. NRC-1*	15790684	Euryarchaeota	Archaea	237
Hsp2	*Halobacterium sp.*	15790610	Euryarchaeota	Archaea	239

	NRC-1				
Hsp3	*Halobacterium sp. SG1*	461613	Euryarchaeota	Archaea	247
Hva1	*Haloarcula vallismortis*	2499388	Euryarchaeota	Archaea	236
Sru1	*Salinibacter ruber DSM 13855*	83815419	Bacteroidetes	Bacteria	235
Sru4	*Salinibacter ruber DSM 13855*	83814238	Bacteroidetes	Bacteria	239
Average protein size (aas)					240
SD					5
Cluster 4					
Hma2	*Haloarcula marismortui ATCC 43049*	55377431	Euryarchaeota	Archaea	232
Vca1	*Volvox carteri f. nagariensis*	302844819	Viridiplantae	Eukaryota	1788
Average protein size (aas)					1010
SD					1100
Cluster 5					
Lma1	*Leptosphaeria maculans*	312215142	Fungi	Eukaryota	304
Osa1	*Oryza sativa Indica Group*	34559256	Viridiplantae	Eukaryota	320
Bfu1	*Botryotinia fuckeliana B05.10*	154319009	Fungi	Eukaryota	312
Sma1	*Sordaria macrospora*	289618890	Fungi	Eukaryota	307
Cgl1	*Chaetomium*	116204249	Fungi	Eukaryota	301

	globosum CBS 148.51				
Gfu2	*Gibberella fujikuroi*	209408529	Fungi	Eukaryota	293
Ani1	*Aspergillus niger CBS 513.88*	145245968	Fungi	Eukaryota	305
Afu1	*Aspergillus fumigatus Af293*	70982522	Fungi	Eukaryota	304
Pch1	*Penicillium chrysogenum Wisconsin 54-1255*	255955269	Fungi	Eukaryota	301
Ani2	*Aspergillus nidulans FGSC A4*	259485658	Fungi	Eukaryota	352
Tst1	*Talaromyces stipitatus ATCC 10500*	242780133	Fungi	Eukaryota	250
Aca1	*Ajellomyces capsulatus NAm1*	154272003	Fungi	Eukaryota	298
Cpo1	*Coccidioides posadasii*	13924710	Fungi	Eukaryota	289
Agy1	*Arthroderma gypseum CBS 118893*	315041787	Fungi	Eukaryota	286
Average protein size (aas)					302
SD					22
Cluster 6					
Sre1	*Sporisorium reilianum*	323507933	Fungi	Eukaryota	292
Mla1	*Melampsora larici-populina 98AG31*	328851545	Fungi	Eukaryota	287
Mos1	*Mixia osmundae*	358059197	Fungi	Eukaryota	323

Average protein size (aas)					290
SD					4
Cluster 7					
Cpa1	*Cyanophora paradoxa*	256681420	Glaucocystophyceae	Eukaryota	334
Cpa2	*Cyanophora paradoxa*	268376501	Glaucocystophyceae	Eukaryota	282
Average protein size (aas)					308
SD					37
Cluster 8					
Cne1	*Cryptococcus neoformans var. neoformans JEC21*	58269234	Fungi	Eukaryota	328
Cga1	*Cryptococcus gattii WM276*	321261457	Fungi	Eukaryota	328
Cluster 9					
Lma2	*Leptosphaeria maculans*	312215043	Fungi	Eukaryota	303
Bor1	*Bipolaris oryzae*	224809196	Fungi	Eukaryota	306
Gfu1	*Gibberella fujikuroi*	39725407	Fungi	Eukaryota	307
Gze1	*Gibberella zeae PH-1*	46114444	Fungi	Eukaryota	407
Bfu2	*Botryotinia fuckeliana B05.10*	154293534	Fungi	Eukaryota	338
Ggr1	*Glomerella graminicola M1.001*	310801701	Fungi	Eukaryota	293
Val1	*Verticillium albo-atrum VaMs.102*	302420879	Fungi	Eukaryota	253

Mor1	*Magnaporthe oryzae 70-15*	39955533	Fungi	Eukaryota	286
Mla2	*Melampsora larici-populina 98AG31*	328851478	Fungi	Eukaryota	295
Uma2	*Ustilago maydis 521*	71015109	Fungi	Eukaryota	335
Uma1	*Ustilago maydis 521*	71020083	Fungi	Eukaryota	334
Agy2	*Arthroderma gypseum CBS 118893*	315045588	Fungi	Eukaryota	267
Aot1	*Arthroderma otae CBS 113480*	296805307	Fungi	Eukaryota	291
Aac1	*Acetabularia acetabulum*	67906884	Viridiplantae	Eukaryota	279
Average protein size (aas)					308
SD					38
Cluster 10					
Csp2	*Cryptomonas sp. S2*	74476774	Cryptophyta	Eukaryota	261
Gth2	*Guillardia theta*	74476770	Cryptophyta	Eukaryota	249
Gth1	*Guillardia theta*	74476772	Cryptophyta	Eukaryota	254
Average protein size (aas)					255
SD					6
Cluster 11					
Oma2	*Oxyrrhis marina*	324096620	Dinophyceae	Eukaryota	252
Oma3	*Oxyrrhis marina*	327335365	Dinophyceae	Eukaryota	257
Average protein size (aas)					255
SD					4
Cluster 12					
Fba1	*Flavobacteria*	225011679	Bacteroidetes	Bacteria	241

	bacterium MS024-2A				
Ddo1	*Dokdonia donghaensis*	223452819	Bacteroidetes	Bacteria	246
Orf1	*uncultured bacterium*	47779419	none	Bacteria	254
Sli1	*Spirosoma linguale DSM 74*	284036559	Bacteroidetes	Bacteria	250
Mtr1	*Marivirga tractuosa DSM 4126*	313676873	Bacteroidetes	Bacteria	252
Hhy1	*Haliscomenobacter hydrossis DSM 1100*	332663718	Bacteroidetes	Bacteria	249
Orf4	*uncultured marine bacterium HF10_19P19*	119094158	none	Bacteria	257
Orf24	*uncultured marine bacterium HF10_25F10*	119094185	none	Bacteria	258
Orf11	*uncultured bacterium*	67483490	none	Bacteria	256
Vca2	*Vibrio campbellii*	289157785	Gammaproteobacteria	Bacteria	271
Orf19	*uncultured bacterium*	29470278	none	Bacteria	254
Orf20	*marine gamma proteobacterium HTCC2143*	119476620	Gammaproteobacteria	Bacteria	229
Orf10	*uncultured bacterium*	67906596	none	Bacteria	257
Orf6	*uncultured marine gamma proteobacterium*	45644626	Gammaproteobacteria	Bacteria	248

	EBAC20E09				
Psp1	*Photobacterium sp. SKA34*	89074634	Gammaproteobacteria	Bacteria	230
Orf8	*uncultured bacterium*	13560725	none	Bacteria	252
Ieu1	*uncultured marine group II euryarchaeote HF70_59C08*	77024964	Euryarchaeota	Archaea	280
Orf23	*uncultured organism HF10_3D09*	82548293	none	Unclassified	278
Esi1	*Exiguobacterium sibiricum 255-15*	172057442	Firmicutes	Bacteria	252
Tcy1	*Thioalkalimicrobium cyclicum ALM1*	334144590	Gammaproteobacteria	Bacteria	249
Orf26	*alpha proteobacterium BAL199*	163795412	Alphaproteobacteria	Bacteria	255
Oan1	*Octadecabacter antarcticus 238*	254449583	Alphaproteobacteria	Bacteria	257
Oba1	*Oxalobacteraceae bacterium IMCC9480*	329904644	Betaproteobacteria	Bacteria	257
Orf27	*beta proteobacterium KB13*	254468407	Betaproteobacteria	Bacteria	249
Msp2	*Marinobacter sp. ELB17*	126666904	Gammaproteobacteria	Bacteria	274
Oma1	*Oxyrrhis marina*	157093545	Dinophyceae	Eukaryota	262
Gvi1	*Gloeobacter violaceus PCC 7421*	37519767	Cyanobacteria	Bacteria	298

Orf22	*actinobacterium MWH-EgelM2-3.D6*	224384116	Actinobacteria	Bacteria	251
Orf28	*actinobacterium MWH-Uga1*	224384108	Actinobacteria	Bacteria	263
Sru3	*Salinibacter ruber DSM 13855*	83815260	Bacteroidetes	Bacteria	273
Ipa1	*Isosphaera pallida ATCC 43644*	320105020	Planctomycetes	Bacteria	308
Average protein size (aas)					258
SD					17
Cluster 13					
Ssp1	*Salpingoeca sp. ATCC 50818*	326433484	Salpingoecidae	Eukaryota	704
Cluster 14					
Csp1	*Cyanothece sp. PCC 7425*	220908815	Cyanobacteria	Bacteria	233
Tra1	*Truepera radiovictrix DSM 17093*	297625115	Deinococcus-Thermus	Bacteria	251
Csp3	*Cyanothece sp. PCC 7424*	218440842	Cyanobacteria	Bacteria	264
Csp4	*Cyanothece sp. PCC 7425*	220906310	Cyanobacteria	Bacteria	233
Average protein size (aas)					245
SD					15

Table S3. 134 NiCoT protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S3).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Tuz1	*Thermoproteus uzoniensis 768-20*	327312060	Crenarchaeota	Archaea	354
Sac1	*Sulfolobus acidocaldarius DSM 639*	70607747	Crenarchaeota	Archaea	355
Sis1	*Sulfolobus islandicus Y.N.15.51*	229583039	Crenarchaeota	Archaea	322
Mcu1	*Metallosphaera cuprina Ar-4*	330834936	Crenarchaeota	Archaea	342
Vdi1	*Vulcanisaeta distributa DSM 14429*	307596423	Crenarchaeota	Archaea	371
Sso1	*Sulfolobus solfataricus P2*	15897886	Crenarchaeota	Archaea	373
Sto1	*Sulfolobus tokodaii str. 7*	15922775	Crenarchaeota	Archaea	383
Mse1	*Metallosphaera sedula DSM 5348*	146304071	Crenarchaeota	Archaea	347
Tac1	*Thermoplasma acidophilum DSM 1728*	16082181	Euryarchaeota	Archaea	347
Mbo1	*Methanoregula boonei 6A8*	154150518	Euryarchaeota	Archaea	363
Hma1	*Hippea maritima DSM 10411*	327399247	Deltaproteobacteria	Bacteria	337
Hsp1	*Hydrogenobaculum sp. Y04AAS1*	195953769	Aquificae	Bacteria	340
Average protein size (aas)					353

SD					17
Cluster 2					
Nfa1	*Nocardia farcinica IFM 10152*	54022189	Actinobacteria	Bacteria	355
Rer1	*Rhodococcus erythropolis PR4*	226308167	Actinobacteria	Bacteria	373
Pdi1	*Pseudonocardia dioxanivorans CB1190*	331696152	Actinobacteria	Bacteria	373
Nba1	*Nocardioidaceae bacterium Broad-1*	326331536	Actinobacteria	Bacteria	368
Gne1	*Gordonia neofelifaecis NRRL B-59395*	326384692	Actinobacteria	Bacteria	386
Fsp1	*Frankia sp. EuI1c*	312195744	Actinobacteria	Bacteria	394
Sav1	*Streptomyces avermitilis MA-4680*	29827101	Actinobacteria	Bacteria	381
Cac1	*Catenulispora acidiphila DSM 44928*	256394330	Actinobacteria	Bacteria	388
Kse1	*Kitasatospora setae KM-6054*	311895361	Actinobacteria	Bacteria	363
Ssp3	*Streptomyces sp. AA4*	302530834	Actinobacteria	Bacteria	364
Fsp2	*Frankia sp. CcI3*	86739760	Actinobacteria	Bacteria	374
Mtu1	*Mycobacterium tuberculosis H37Rv*	15609993	Actinobacteria	Bacteria	372
Sav2	*Streptomyces avermitilis MA-4680*	29828875	Actinobacteria	Bacteria	377
Ssp1	*Streptomyces sp. AA4*	302527431	Actinobacteria	Bacteria	377
Bdo1	*Burkholderia dolosa*	254255317	Betaproteobacteria	Bacteria	380

	AUO158				
Bsp3	*Burkholderia sp. Ch1-1*	296156247	Betaproteobacteria	Bacteria	379
Bsp1	*Burkholderia sp. CCGE1002*	295700941	Betaproteobacteria	Bacteria	377
Dsp1	*Dermacoccus sp. Ellin185*	309810442	Actinobacteria	Bacteria	366
Pdi2	*Pseudonocardia dioxanivorans CB1190*	327309912	Actinobacteria	Bacteria	384
Kra1	*Kineococcus radiotolerans SRS30216*	152967021	Actinobacteria	Bacteria	365
Nsp1	*Nocardioides sp. JS614*	119716355	Actinobacteria	Bacteria	347
Average protein size (aas)					373
SD					11
Cluster 3					
Msp1	*Micromonospora sp. ATCC 39149*	238059430	Actinobacteria	Bacteria	387
Rjo1	*Rhodococcus jostii RHA1*	111026242	Actinobacteria	Bacteria	364
Rrh1	*Rhodococcus rhodochrous*	1850787	Actinobacteria	Bacteria	352
Ser1	*Saccharopolyspora erythraea NRRL 2338*	134102404	Actinobacteria	Bacteria	386
Tin1	*Thiomonas intermedia K12*	296137491	Betaproteobacteria	Bacteria	384
Average protein size (aas)					375

SD					16
Cluster 4					
Dge1	*Deinococcus geothermalis DSM 11300*	94971822	Deinococcus-Thermus	Bacteria	336
Cluster 5					
Sru1	*Segniliparus rugosus ATCC BAA-974*	317507819	Actinobacteria	Bacteria	279
Sro1	*Segniliparus rotundus DSM 44985*	296392871	Actinobacteria	Bacteria	273
Fph1	*Francisella philomiragia subsp. philomiragia ATCC 25017*	167627043	Gammaproteobacteria	Bacteria	309
Average protein size (aas)					287
SD					19
Cluster 6					
Mlu1	*Micrococcus luteus NCTC 2665*	239916896	Actinobacteria	Bacteria	352
Krh1	*Kocuria rhizophila DC2201*	184201812	Actinobacteria	Bacteria	347
Aau1	*Arthrobacter aurescens TC1*	119963145	Actinobacteria	Bacteria	348
Mph1	*Microlunatus phosphovorus NM-1*	334688213	Actinobacteria	Bacteria	386
Aur1	*Actinomyces urogenitalis DSM 15434*	227497026	Actinobacteria	Bacteria	319

Average protein size (aas)					350
SD					24
Cluster 7					
Aan1	*Aureococcus anophagefferens*	323450558	Pelagophyceae	Eukaryota	354
Taq1	*Thermus aquaticus Y51MC23*	218295416	Deinococcus-Thermus	Bacteria	220
Dma1	*Deinococcus maricopensis DSM 21211*	320332818	Deinococcus-Thermus	Bacteria	226
Cme1	*Cupriavidus metallidurans CH34*	94310475	Betaproteobacteria	Bacteria	278
Dar1	*Dechloromonas aromatica RCB*	71909556	Betaproteobacteria	Bacteria	269
Sli1	*Sideroxydans lithotrophicus ES-1*	291614874	Betaproteobacteria	Bacteria	271
Lpn1	*Legionella pneumophila subsp. pneumophila str. Philadelphia 1*	52842685	Gammaproteobacteria	Bacteria	255
Llo1	*Legionella longbeachae D-4968*	270159427	Gammaproteobacteria	Bacteria	226
Bme1	*Bacillus megaterium QM B1551*	294497143	Firmicutes	Bacteria	265
Nde1	*Candidatus Nitrospira defluvii*	302037429	Nitrospirae	Bacteria	267
Average protein size (aas)					263
SD					38
Cluster 8					

Lci1	*Leuconostoc citreum KM20*	170017836	Firmicutes	Bacteria	341
Lme1	*Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293*	116619095	Firmicutes	Bacteria	341
Lbr1	*Lactobacillus brevis subsp. gravesensis ATCC 27305*	227508031	Firmicutes	Bacteria	359
Bce1	*Bacillus cereus ATCC 10987*	42782706	Firmicutes	Bacteria	328
Hmu1	*Helicobacter mustelae 12198*	291276463	Epsilonproteobacteria	Bacteria	330
Hpy1	*Helicobacter pylori*	732734	Epsilonproteobacteria	Bacteria	308
Hfe1	*Helicobacter felis ATCC 49179*	315453608	Epsilonproteobacteria	Bacteria	342
Sps1	*Staphylococcus pseudintermedius HKU10-03*	319893531	Firmicutes	Bacteria	336
Sau1	*Staphylococcus aureus subsp. aureus COL*	57651089	Firmicutes	Bacteria	350
Btu1	*Bacillus tusciae DSM 2912*	295697080	Firmicutes	Bacteria	339
Dni1	*Desulfotomaculum nigrificans DSM 574*	323702331	Firmicutes	Bacteria	343
Ppe1	*Pediococcus pentosaceus ATCC 25745*	116493424	Firmicutes	Bacteria	335
Lfe1	*Lactobacillus fermentum IFO 3956*	184156204	Firmicutes	Bacteria	337

Lbr2	*Lactobacillus brevis ATCC 367*	116334421	Firmicutes	Bacteria	327
Blo1	*Bifidobacterium longum subsp. infantis ATCC 15697*	213691030	Actinobacteria	Bacteria	324
Average protein size (aas)					336
SD					12
Cluster 9					
Ate1	*Aspergillus terreus NIH2624*	115443076	Fungi	Eukaryota	1044
Afl1	*Aspergillus flavus NRRL3357*	238493689	Fungi	Eukaryota	439
Ani1	*Aspergillus niger*	134080234	Fungi	Eukaryota	443
Afu1	*Aspergillus fumigatus Af293*	71001072	Fungi	Eukaryota	474
Pch1	*Penicillium chrysogenum Wisconsin 54-1255*	255935567	Fungi	Eukaryota	475
Tst1	*Talaromyces stipitatus ATCC 10500*	242776711	Fungi	Eukaryota	501
Tve1	*Trichophyton verrucosum HKI 0517*	302654098	Fungi	Eukaryota	438
Agy1	*Arthroderma gypseum CBS 118893*	315050476	Fungi	Eukaryota	474
Cpo1	*Coccidioides posadasii C735 delta SOWgp*	303312513	Fungi	Eukaryota	370

Aca2	*Ajellomyces capsulatus G186AR*	225559374	Fungi	Eukaryota	520
Pbr1	*Paracoccidioides brasiliensis Pb03*	225684226	Fungi	Eukaryota	585
Lma1	*Leptosphaeria maculans*	312217298	Fungi	Eukaryota	428
Ggr1	*Glomerella graminicola M1.001*	310801087	Fungi	Eukaryota	414
Mor1	*Magnaporthe oryzae 70-15*	39941184	Fungi	Eukaryota	381
Val1	*Verticillium albo-atrum VaMs.102*	302417100	Fungi	Eukaryota	373
Gze1	*Gibberella zeae PH-1*	46124271	Fungi	Eukaryota	409
Man1	*Metarhizium anisopliae ARSEF 23*	322705897	Fungi	Eukaryota	423
Gcl1	*Grosmannia clavigera kw1407*	320587863	Fungi	Eukaryota	427
Pan1	*Podospora anserina S mat+*	171685454	Fungi	Eukaryota	415
Sma1	*Sordaria macrospora*	289617779	Fungi	Eukaryota	414
Tme1	*Tuber melanosporum Mel28*	296420530	Fungi	Eukaryota	436
Mor2	*Magnaporthe oryzae 70-15*	145609405	Fungi	Eukaryota	613
Sja2	*Schizosaccharomyces japonicus yFS275*	213410268	Fungi	Eukaryota	405
Spo1	*Schizosaccharomyces pombe 972h-*	63054407	Fungi	Eukaryota	405
Sco1	*Schizophyllum commune H4-8*	302680430	Fungi	Eukaryota	486

Lbi1	*Laccaria bicolor S238N-H82*	170100699	Fungi	Eukaryota	418
Cne1	*Cryptococcus neoformans var. neoformans JEC21*	58262844	Fungi	Eukaryota	536
Uma1	*Ustilago maydis 521*	71005820	Fungi	Eukaryota	496
Average protein size (aas)					473
SD					127
Cluster 10					
Sen1	*Salmonella enterica subsp. enterica serovar Typhi*	2961076	Gammaproteobacteria	Bacteria	294
Spr1	*Serratia proteamaculans 568*	157370662	Gammaproteobacteria	Bacteria	343
Pat1	*Pectobacterium atrosepticum SCRI1043*	50120191	Gammaproteobacteria	Bacteria	327
Dda1	*Dickeya dadantii 3937*	307130887	Gammaproteobacteria	Bacteria	343
Csp1	*Citrobacter sp. 30_2*	237731430	Gammaproteobacteria	Bacteria	339
Pda1	*Photobacterium damselae subsp. damselae CIP 102761*	269103122	Gammaproteobacteria	Bacteria	342
Kra2	*Ktedonobacter racemifer DSM 44963*	298242375	Chloroflexi	Bacteria	366
Apa1	*Acetobacter pasteurianus IFO 3283-01*	258543407	Alphaproteobacteria	Bacteria	344

Bja1	*Bradyrhizobium japonicum USDA 110*	27376265	Alphaproteobacteria	Bacteria	380
Bsp2	*Bradyrhizobium sp. ORS278*	146338747	Alphaproteobacteria	Bacteria	338
Rso1	*Ralstonia solanacearum GMI1000*	17548849	Betaproteobacteria	Bacteria	357
Lni1	*Lutiella nitroferrum 2002*	224825413	Betaproteobacteria	Bacteria	350
Cvi1	*Chromobacterium violaceum ATCC 12472*	34497028	Betaproteobacteria	Bacteria	378
Ara1	*Agrobacterium radiobacter K84*	222086034	Alphaproteobacteria	Bacteria	349
Ara2	*Agrobacterium radiobacter K84*	222102289	Alphaproteobacteria	Bacteria	345
Gbe1	*Granulibacter bethesdensis CGDNIH1*	114327646	Alphaproteobacteria	Bacteria	363
Acr1	*Acidiphilium cryptum JF-5*	148259367	Alphaproteobacteria	Bacteria	331
Bvi1	*Burkholderia vietnamiensis G4*	134291341	Betaproteobacteria	Bacteria	355
Reu1	*Ralstonia eutropha H16*	38637728	Betaproteobacteria	Bacteria	344
Mra1	*Methylobacterium radiotolerans JCM 2831*	170745160	Alphaproteobacteria	Bacteria	354
Bsp5	*Bradyrhizobium sp. BTAi1*	148254779	Alphaproteobacteria	Bacteria	364

Rpa1	*Rhodopseudomonas palustris CGA009*	39933801	Alphaproteobacteria	Bacteria	352
Bgl1	*Burkholderia gladioli BSR3*	330814682	Betaproteobacteria	Bacteria	353
Bvi2	*Burkholderia vietnamiensis G4*	134295707	Betaproteobacteria	Bacteria	352
Asp1	*Acidobacterium sp. MP5ACTX9*	322435659	Acidobacteria	Bacteria	346
Aca3	*Acidobacterium capsulatum ATCC 51196*	225874192	Acidobacteria	Bacteria	358
Bxe1	*Burkholderia xenovorans LB400*	91780108	Betaproteobacteria	Bacteria	354
Reu2	*Ralstonia eutropha H16*	38637691	Betaproteobacteria	Bacteria	351
Sja1	*Sphingobium japonicum UT26S*	294146734	Alphaproteobacteria	Bacteria	357
Nar1	*Novosphingobium aromaticivorans DSM 12444*	87198369	Alphaproteobacteria	Bacteria	371
Ssp2	*Sphingomonas sp. S17*	332188472	Alphaproteobacteria	Bacteria	353
Nni1	*Novosphingobium nitrogenifigens DSM 19370*	326385346	Alphaproteobacteria	Bacteria	349
Bce2	*Burkholderia cenocepacia J2315*	206561078	Betaproteobacteria	Bacteria	347
Bsp4	*Burkholderia sp. Ch1-1*	296157179	Betaproteobacteria	Bacteria	335
Average protein size (aas)					350
SD					15

Table S4. 52 OST protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S4).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Aec1	*Acromyrmex echinatior*	332025948	Metazoa	Eukaryota	331
Cfl1	*Camponotus floridanus*	307171448	Metazoa	Eukaryota	320
Hsa2	*Harpegnathos saltator*	307209024	Metazoa	Eukaryota	335
Ame1	*Apis mellifera*	328781239	Metazoa	Eukaryota	341
Nvi1	*Nasonia vitripennis*	156554046	Metazoa	Eukaryota	351
Aae1	*Aedes aegypti*	157138577	Metazoa	Eukaryota	320
Cqu1	*Culex quinquefasciatus*	170034785	Metazoa	Eukaryota	317
Aga1	*Anopheles gambiae str. PEST*	118793812	Metazoa	Eukaryota	312
Average protein size (aas)					328
SD					13
Cluster 2					
Dpu1	*Daphnia pulex*	321454526	Metazoa	Eukaryota	361
Dpu2	*Daphnia pulex*	321476811	Metazoa	Eukaryota	363
Dpu3	*Daphnia pulex*	321454527	Metazoa	Eukaryota	315
Average protein size (aas)					346
SD					27
Cluster 3					
Cbr1	*Caenorhabditis briggsae*	268530378	Metazoa	Eukaryota	348
Aca3	*Angiostrongylus cantonensis*	256016461	Metazoa	Eukaryota	370
Llo1	*Loa loa*	312068356	Metazoa	Eukaryota	349
Bma1	*Brugia malayi*	170592162	Metazoa	Eukaryota	349
Asu1	*Ascaris suum*	324510389	Metazoa	Eukaryota	332
Cbr2	*Caenorhabditis briggsae*	268531320	Metazoa	Eukaryota	342
Average protein size (aas)					348
SD					12
Cluster 4					
Dre2	*Danio rerio*	189522496	Metazoa	Eukaryota	414
Tni2	*Tetraodon nigroviridis*	47209270	Metazoa	Eukaryota	869
Tsp1	*Trichinella spiralis*	316978264	Metazoa	Eukaryota	352
Average protein size (aas)					545
SD					282

Cluster 5					
Cow1	*Capsaspora owczarzaki ATCC 30864*	320163545	Ichthyosporea	Eukaryota	342
Cre2	*Chlamydomonas reinhardtii*	159465163	Viridiplantae	Eukaryota	307
Ddi1	*Dictyostelium discoideum AX4*	66819373	none	Eukaryota	507
Pso1	*Phytophthora sojae*	210160949	none	Eukaryota	347
Ath8	*Arabidopsis thaliana*	18411404	Viridiplantae	Eukaryota	484
Sja1	*Schizosaccharomyces japonicus yFS275*	213408178	Fungi	Eukaryota	428
Mla1	*Melampsora larici-populina 98AG31*	328863521	Fungi	Eukaryota	403
Ncr1	*Neurospora crassa*	8510698	Fungi	Eukaryota	578
Average protein size (aas)					424
SD					93
Cluster 6					
Gga1	*Gallus gallus*	118095143	Metazoa	Eukaryota	335
Tgu2	*Taeniopygia guttata*	224060601	Metazoa	Eukaryota	362
Clu1	*Canis lupus familiaris*	74002938	Metazoa	Eukaryota	353
Eca1	*Equus caballus*	149731523	Metazoa	Eukaryota	340
Mmu3	*Macaca mulatta*	109054285	Metazoa	Eukaryota	340
Mmu1	*Mus musculus*	22122353	Metazoa	Eukaryota	340
Ssc1	*Sus scrofa*	311269855	Metazoa	Eukaryota	340
Bta2	*Bos taurus*	122140433	Metazoa	Eukaryota	340
Mdo1	*Monodelphis domestica*	126325983	Metazoa	Eukaryota	340
Aca2	*Anolis carolinensis*	327289371	Metazoa	Eukaryota	460
Orf2	*Xenopus (Silurana) tropicalis*	301609070	Metazoa	Eukaryota	332
Ler1	*Leucoraja erinacea*	82108802	Metazoa	Eukaryota	352
Dre1	*Danio rerio*	169146805	Metazoa	Eukaryota	389
Ipu1	*Ictalurus punctatus*	318068026	Metazoa	Eukaryota	330
Mga2	*Meleagris gallopavo*	326913536	Metazoa	Eukaryota	365
Gga2	*Gallus gallus*	118084075	Metazoa	Eukaryota	536
Tgu1	*Taeniopygia guttata*	224042671	Metazoa	Eukaryota	574
Orf1	*Xenopus (Silurana) tropicalis*	187471107	Metazoa	Eukaryota	339
Average protein size (aas)					376
SD					72
Cluster 7					

Sko1	*Saccoglossus kowalevskii*	291225656	Metazoa	Eukaryota	368
Sko2	*Saccoglossus kowalevskii*	291225660	Metazoa	Eukaryota	408
Sko3	*Saccoglossus kowalevskii*	291223152	Metazoa	Eukaryota	370
Spu1	*Strongylocentrotus purpuratus*	115955237	Metazoa	Eukaryota	377
Average protein size (aas)					381
SD					19
Cluster 8					
Bfl1	*Branchiostoma floridae*	260790111	Metazoa	Eukaryota	342
Bfl2	*Branchiostoma floridae*	260790095	Metazoa	Eukaryota	334
Cin2	*Ciona intestinalis*	198422626	Metazoa	Eukaryota	378
Cin3	*Ciona intestinalis*	198428670	Metazoa	Eukaryota	381
Odi1	*Oikopleura dioica*	313239889	Metazoa	Eukaryota	401
Average protein size (aas)					367
SD					28

Table S5. 198 Sweet protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S5).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Nca1	*Neospora caninum Liverpool*	325120066	Apicomplexa	Eukaryota	672
Tgo1	*Toxoplasma gondii GT1*	221480968	Apicomplexa	Eukaryota	666
Ala2	*Albugo laibachii Nc14*	325182587	Albuginaceae	Eukaryota	268
Pin10	*Phytophthora infestans T30-4*	301098858	none	Eukaryota	235
Isc1	*Ixodes scapularis*	241638533	Metazoa	Eukaryota	204
Average protein size (aas)					409
SD					238
Cluster 2					
Dgr1	*Drosophila grimshawi*	195079676	Metazoa	Eukaryota	232
Dan2	*Drosophila ananassae*	194751517	Metazoa	Eukaryota	228
Aga1	*Anopheles gambiae str. PEST*	118785023	Metazoa	Eukaryota	224
Aae2	*Aedes aegypti*	157109690	Metazoa	Eukaryota	222
Aae4	*Aedes aegypti*	157114191	Metazoa	Eukaryota	228
Aae3	*Aedes aegypti*	157109688	Metazoa	Eukaryota	228
Cfl1	*Camponotus floridanus*	307187728	Metazoa	Eukaryota	222
Ame2	*Apis mellifera*	66509804	Metazoa	Eukaryota	220
Tca1	*Tribolium castaneum*	270013117	Metazoa	Eukaryota	223
Api1	*Acyrthosiphon pisum*	193718421	Metazoa	Eukaryota	211
Lsa1	*Lepeophtheirus salmonis*	290561829	Metazoa	Eukaryota	229
Aae1	*Aedes aegypti*	157112765	Metazoa	Eukaryota	232
Aga2	*Anopheles gambiae str. PEST*	119113928	Metazoa	Eukaryota	229
Dan1	*Drosophila ananassae*	194757529	Metazoa	Eukaryota	226
Tca2	*Tribolium castaneum*	91082333	Metazoa	Eukaryota	213
Hsa1	*Harpegnathos saltator*	307203973	Metazoa	Eukaryota	215
Sin1	*Solenopsis invicta*	322779873	Metazoa	Eukaryota	218
Ame1	*Apis mellifera*	48099654	Metazoa	Eukaryota	220
Nvi2	*Nasonia vitripennis*	268370163	Metazoa	Eukaryota	218
Api2	*Acyrthosiphon pisum*	328717862	Metazoa	Eukaryota	271
Phu1	*Pediculus humanus*	242017704	Metazoa	Eukaryota	221

	corporis				
Asu2	*Ascaris suum*	324509744	Metazoa	Eukaryota	352
Cbr3	*Caenorhabditis briggsae*	268553953	Metazoa	Eukaryota	293
Cel1	*Caenorhabditis elegans*	17558938	Metazoa	Eukaryota	233
Dpu1	*Daphnia pulex*	321476861	Metazoa	Eukaryota	221
Bma1	*Brugia malayi*	170591735	Metazoa	Eukaryota	254
Asu3	*Ascaris suum*	324518669	Metazoa	Eukaryota	224
Cbr1	*Caenorhabditis briggsae*	268552629	Metazoa	Eukaryota	354
Asu1	*Ascaris suum*	324506374	Metazoa	Eukaryota	375
Cbr2	*Caenorhabditis briggsae*	268574740	Metazoa	Eukaryota	188
Average protein size (aas)					241
SD					45
Cluster 3					
Spu1	*Strongylocentrotus purpuratus*	72160154	Metazoa	Eukaryota	184
Nve2	*Nematostella vectensis*	156374131	Metazoa	Eukaryota	225
Bfl2	*Branchiostoma floridae*	260800273	Metazoa	Eukaryota	220
Ssp2	*Salpingoeca sp. ATCC 50818*	326431165	Salpingoecidae	Eukaryota	221
Tad1	*Trichoplax adhaerens*	196002641	Metazoa	Eukaryota	217
Nve1	*Nematostella vectensis*	156353371	Metazoa	Eukaryota	216
Cja1	*Callithrix jacchus*	296229069	Metazoa	Eukaryota	221
Xla1	*Xenopus laevis*	148235825	Metazoa	Eukaryota	216
Dre1	*Danio rerio*	206558238	Metazoa	Eukaryota	219
Tni1	*Tetraodon nigroviridis*	47225383	Metazoa	Eukaryota	219
Bfl1	*Branchiostoma floridae*	260829761	Metazoa	Eukaryota	210
Cin1	*Ciona intestinalis*	198433250	Metazoa	Eukaryota	215
Average protein size (aas)					215
SD					11
Cluster 4					
Odi1	*Oikopleura dioica*	313214716	Metazoa	Eukaryota	219
Bde3	*Batrachochytrium dendrobatidis JAM81*	328769156	Fungi	Eukaryota	224
Average protein size (aas)					222
SD					4

Cluster 5					
Ddi1	*Dictyostelium discoideum AX4*	66806629	none	Eukaryota	220
Dpu2	*Dictyostelium purpureum*	330798800	none	Eukaryota	218
Ppa3	*Polysphondylium pallidum PN500*	281212522	none	Eukaryota	673
Esi1	*Ectocarpus siliculosus*	299470883	Phaeophyceae	Eukaryota	224
Average protein size (aas)					334
SD					226
Cluster 6					
Ala1	*Albugo laibachii Nc14*	325184656	Albuginaceae	Eukaryota	239
Pin12	*Phytophthora infestans T30-4*	301101261	none	Eukaryota	239
Ssp1	*Salpingoeca sp. ATCC 50818*	326432237	Salpingoecidae	Eukaryota	348
Mbr1	*Monosiga brevicollis MX1*	167515384	Codonosigidae	Eukaryota	249
Tps1	*Thalassiosira pseudonana CCMP1335*	224011239	Bacillariophyta	Eukaryota	191
Tps2	*Thalassiosira pseudonana CCMP1335*	223994243	Bacillariophyta	Eukaryota	204
Ptr7	*Phaeodactylum tricornutum CCAP 1055/1*	219111545	Bacillariophyta	Eukaryota	229
Aan1	*Aureococcus anophagefferens*	323456250	Pelagophyceae	Eukaryota	243
Average protein size (aas)					243
SD					47
Cluster 7					
Pin2	*Phytophthora infestans T30-4*	301115910	none	Eukaryota	315
Pin11	*Phytophthora infestans T30-4*	301115904	none	Eukaryota	230
Pin13	*Phytophthora infestans T30-4*	301094617	none	Eukaryota	203
Pin14	*Phytophthora infestans T30-4*	301115906	none	Eukaryota	263
Pin4	*Phytophthora*	301115908	none	Eukaryota	247

	infestans T30-4				
Pin9	*Phytophthora infestans T30-4*	301115912	none	Eukaryota	403
Pin7	*Phytophthora infestans T30-4*	301099708	none	Eukaryota	324
Pin8	*Phytophthora infestans T30-4*	301090305	none	Eukaryota	259
Pin5	*Phytophthora infestans T30-4*	301094603	none	Eukaryota	308
Pin3	*Phytophthora infestans T30-4*	301102307	none	Eukaryota	296
Average protein size (aas)					285
SD					57
Cluster 8					
Bbo1	*Babesia bovis T2Bo*	156085796	Apicomplexa	Eukaryota	390
Tpa1	*Theileria parva strain Muguga*	71033785	Apicomplexa	Eukaryota	379
Pbe1	*Plasmodium berghei str. ANKA*	68073307	Apicomplexa	Eukaryota	487
Pvi1	*Plasmodium vivax SaI-1*	156093685	Apicomplexa	Eukaryota	661
Sko1	*Saccoglossus kowalevskii*	291221641	Metazoa	Eukaryota	1174
Psi1	*Picea sitchensis*	294461247	Viridiplantae	Eukaryota	208
Aly3	*Arabidopsis lyrata subsp. lyrata*	297792819	Viridiplantae	Eukaryota	290
Average protein size (aas)					513
SD					325
Cluster 9					
Mpu1	*Micromonas pusilla CCMP1545*	303277429	Viridiplantae	Eukaryota	209
Msp1	*Micromonas sp. RCC299*	255075637	Viridiplantae	Eukaryota	254
Ota1	*Ostreococcus tauri*	308802750	Viridiplantae	Eukaryota	250
Olu1	*Ostreococcus lucimarinus CCE9901*	145345954	Viridiplantae	Eukaryota	242
Cre1	*Chlamydomonas reinhardtii*	159476704	Viridiplantae	Eukaryota	231
Cva1	*Chlorella variabilis*	307106643	Viridiplantae	Eukaryota	266
Cre2	*Chlamydomonas reinhardtii*	159476230	Viridiplantae	Eukaryota	249
Vca1	*Volvox carteri f. nagariensis*	302840323	Viridiplantae	Eukaryota	315
Bde1	*Batrachochytrium*	328771906	Fungi	Eukaryota	233

	dendrobatidis JAM81				
Bde2	*Batrachochytrium dendrobatidis JAM81*	328771907	Fungi	Eukaryota	236
Cow1	*Capsaspora owczarzaki ATCC 30864*	320168194	Ichthyosporea	Eukaryota	283
Ddi2	*Dictyostelium discoideum AX4*	66808155	none	Eukaryota	299
Dpu4	*Dictyostelium purpureum*	330795005	none	Eukaryota	289
Ddi3	*Dictyostelium discoideum AX4*	66806795	none	Eukaryota	336
Dpu3	*Dictyostelium purpureum*	330805201	none	Eukaryota	256
Average protein size (aas)					263
SD					35
Cluster 10					
Ptr1	*Populus trichocarpa*	224085065	Viridiplantae	Eukaryota	248
Gma8	*Glycine max*	255637929	Viridiplantae	Eukaryota	247
Psi5	*Picea sitchensis*	116782985	Viridiplantae	Eukaryota	260
Zma2	*Zea mays*	226500492	Viridiplantae	Eukaryota	267
Sbi10	*Sorghum bicolor*	242090583	Viridiplantae	Eukaryota	256
Ath7	*Arabidopsis thaliana*	18400517	Viridiplantae	Eukaryota	236
Vvi10	*Vitis vinifera*	225461810	Viridiplantae	Eukaryota	235
Ptr2	*Populus trichocarpa*	224061395	Viridiplantae	Eukaryota	216
Zma4	*Zea mays*	226508060	Viridiplantae	Eukaryota	243
Osa13	*Oryza sativa Indica Group*	125549501	Viridiplantae	Eukaryota	471
Vvi11	*Vitis vinifera*	225462464	Viridiplantae	Eukaryota	232
Ptr16	*Populus trichocarpa*	224061033	Viridiplantae	Eukaryota	218
Psi4	*Picea sitchensis*	294460447	Viridiplantae	Eukaryota	231
Sbi7	*Sorghum bicolor*	242054131	Viridiplantae	Eukaryota	231
Smo4	*Selaginella moellendorffii*	302753780	Viridiplantae	Eukaryota	239
Smo15	*Selaginella moellendorffii*	302773003	Viridiplantae	Eukaryota	202
Average protein size (aas)					252
SD					61
Cluster 11					
Smo14	*Selaginella moellendorffii*	302808073	Viridiplantae	Eukaryota	498
Smo10	*Selaginella moellendorffii*	302804901	Viridiplantae	Eukaryota	362
Ppa5	*Physcomitrella patens*	168052757	Viridiplantae	Eukaryota	191

	subsp. patens				
Smo3	*Selaginella moellendorffii*	302781266	Viridiplantae	Eukaryota	254
Smo5	*Selaginella moellendorffii*	302825721	Viridiplantae	Eukaryota	246
Smo8	*Selaginella moellendorffii*	302773225	Viridiplantae	Eukaryota	244
Smo9	*Selaginella moellendorffii*	302757455	Viridiplantae	Eukaryota	211
Ppa1	*Physcomitrella patens subsp. patens*	168052158	Viridiplantae	Eukaryota	247
Ppa2	*Physcomitrella patens subsp. patens*	168036203	Viridiplantae	Eukaryota	253
Ppa4	*Physcomitrella patens subsp. patens*	168014545	Viridiplantae	Eukaryota	251
Average protein size (aas)					276
SD					90
Cluster 12					
Sbi2	*Sorghum bicolor*	242091553	Viridiplantae	Eukaryota	239
Sbi14	*Sorghum bicolor*	242053547	Viridiplantae	Eukaryota	244
Osa2	*Oryza sativa Japonica Group*	115478214	Viridiplantae	Eukaryota	375
Vvi12	*Vitis vinifera*	225452486	Viridiplantae	Eukaryota	254
Ath2	*Arabidopsis thaliana*	186532678	Viridiplantae	Eukaryota	240
Sly1	*Solanum lycopersicum*	37050896	Viridiplantae	Eukaryota	238
Vvi3	*Vitis vinifera*	225457069	Viridiplantae	Eukaryota	234
Ath1	*Arabidopsis thaliana*	21592355	Viridiplantae	Eukaryota	251
Ptr6	*Populus trichocarpa*	224133506	Viridiplantae	Eukaryota	237
Gma5	*Glycine max*	255645477	Viridiplantae	Eukaryota	247
Ptr18	*Populus trichocarpa*	224075142	Viridiplantae	Eukaryota	255
Rco11	*Ricinus communis*	255555653	Viridiplantae	Eukaryota	261
Vvi9	*Vitis vinifera*	225426236	Viridiplantae	Eukaryota	260
Sbi6	*Sorghum bicolor*	242064916	Viridiplantae	Eukaryota	250
Ath12	*Arabidopsis thaliana*	15219732	Viridiplantae	Eukaryota	261
Rco6	*Ricinus communis*	255552606	Viridiplantae	Eukaryota	242
Ptr19	*Populus trichocarpa*	224055573	Viridiplantae	Eukaryota	242
Rco10	*Ricinus communis*	255576288	Viridiplantae	Eukaryota	233
Rco1	*Ricinus communis*	255584088	Viridiplantae	Eukaryota	236
Ath10	*Arabidopsis thaliana*	18421965	Viridiplantae	Eukaryota	239
Aly2	*Arabidopsis lyrata subsp. lyrata*	297789376	Viridiplantae	Eukaryota	224
Average protein size (aas)					251
SD					30

Cluster 13

Vvi1	*Vitis vinifera*	296086628	Viridiplantae	Eukaryota	341
Rco3	*Ricinus communis*	255559318	Viridiplantae	Eukaryota	286
Ath9	*Arabidopsis thaliana*	15225014	Viridiplantae	Eukaryota	258
Ptr8	*Populus trichocarpa*	224109054	Viridiplantae	Eukaryota	259
Orf1	*Petunia x hybrida*	75172033	Viridiplantae	Eukaryota	265
Vvi8	*Vitis vinifera*	225462403	Viridiplantae	Eukaryota	278
Gma2	*Glycine max*	255638124	Viridiplantae	Eukaryota	257
Vvi5	*Vitis vinifera*	225456416	Viridiplantae	Eukaryota	270
Can1	*Capsicum annuum*	257831431	Viridiplantae	Eukaryota	301
Rco4	*Ricinus communis*	255540127	Viridiplantae	Eukaryota	279
Ptr12	*Populus trichocarpa*	224134076	Viridiplantae	Eukaryota	269
Ath13	*Arabidopsis thaliana*	15241265	Viridiplantae	Eukaryota	289
Ath3	*Arabidopsis thaliana*	15229019	Viridiplantae	Eukaryota	289
Ath5	*Arabidopsis thaliana*	15234863	Viridiplantae	Eukaryota	281
Rco5	*Ricinus communis*	255540123	Viridiplantae	Eukaryota	297
Ptr15	*Populus trichocarpa*	224136248	Viridiplantae	Eukaryota	275
Sbi4	*Sorghum bicolor*	242083388	Viridiplantae	Eukaryota	302
Sbi5	*Sorghum bicolor*	242071019	Viridiplantae	Eukaryota	291
Sbi3	*Sorghum bicolor*	242065206	Viridiplantae	Eukaryota	336
Mtr2	*Medicago truncatula*	217072936	Viridiplantae	Eukaryota	278
Rco2	*Ricinus communis*	255540121	Viridiplantae	Eukaryota	285
Sbi12	*Sorghum bicolor*	242049796	Viridiplantae	Eukaryota	273
Sbi15	*Sorghum bicolor*	242079839	Viridiplantae	Eukaryota	309
Vvi4	*Vitis vinifera*	297734468	Viridiplantae	Eukaryota	288
Gma4	*Glycine max*	255648175	Viridiplantae	Eukaryota	268
Ptr11	*Populus trichocarpa*	224072514	Viridiplantae	Eukaryota	273
Ath8	*Arabidopsis thaliana*	15240040	Viridiplantae	Eukaryota	292
Vvi13	*Vitis vinifera*	225425180	Viridiplantae	Eukaryota	289
Gma3	*Glycine max*	255636015	Viridiplantae	Eukaryota	280
Sbi8	*Sorghum bicolor*	242035771	Viridiplantae	Eukaryota	313
Osa12	*Oryza sativa Indica Group*	322967558	Viridiplantae	Eukaryota	300
Hvu1	*Hordeum vulgare subsp. vulgare*	326502880	Viridiplantae	Eukaryota	314
Smo1	*Selaginella moellendorffii*	302820242	Viridiplantae	Eukaryota	206
Smo11	*Selaginella moellendorffii*	302796894	Viridiplantae	Eukaryota	196
Vvi2	*Vitis vinifera*	225450721	Viridiplantae	Eukaryota	298
Ptr14	*Populus trichocarpa*	224123056	Viridiplantae	Eukaryota	198
Osa6	*Oryza sativa Indica Group*	218192806	Viridiplantae	Eukaryota	331
Zma8	*Zea mays*	195613130	Viridiplantae	Eukaryota	238
Psi3	*Picea sitchensis*	294462834	Viridiplantae	Eukaryota	335
Ath11	*Arabidopsis thaliana*	27754697	Viridiplantae	Eukaryota	241

Ptr3	*Populus trichocarpa*	224130420	Viridiplantae	Eukaryota	238
Gma7	*Glycine max*	255646128	Viridiplantae	Eukaryota	245
Ptr5	*Populus trichocarpa*	224123068	Viridiplantae	Eukaryota	238
Vvi7	*Vitis vinifera*	225450717	Viridiplantae	Eukaryota	278
Ptr13	*Populus trichocarpa*	224123066	Viridiplantae	Eukaryota	239
Rco7	*Ricinus communis*	255582276	Viridiplantae	Eukaryota	249
Ptr9	*Populus trichocarpa*	224123826	Viridiplantae	Eukaryota	217
Average protein size (aas)					275
SD					34
Cluster 14					
Osa8	*Oryza sativa Indica Group*	218196292	Viridiplantae	Eukaryota	248
Zma6	*Zea mays*	308081627	Viridiplantae	Eukaryota	238
Sbi13	*Sorghum bicolor*	242051563	Viridiplantae	Eukaryota	259
Psi2	*Picea sitchensis*	116791551	Viridiplantae	Eukaryota	272
Ptr10	*Populus trichocarpa*	224120614	Viridiplantae	Eukaryota	255
Ath14	*Arabidopsis thaliana*	42568507	Viridiplantae	Eukaryota	263
Gma9	*Glycine max*	255647679	Viridiplantae	Eukaryota	254
Average protein size (aas)					256
SD					11
Cluster 15					
Sbi9	*Sorghum bicolor*	242051793	Viridiplantae	Eukaryota	213
Sbi16	*Sorghum bicolor*	242051795	Viridiplantae	Eukaryota	242
Average protein size (aas)					228
SD					21
Cluster 16					
Pma3	*Perkinsus marinus ATCC 50983*	294894544	Perkinsidae	Eukaryota	231
Pma2	*Perkinsus marinus ATCC 50983*	294942416	Perkinsidae	Eukaryota	286
Average protein size (aas)					259
SD					39

Table S6. 147 PNaS protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S6).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Cph1	*Chlorobium phaeobacteroides BS1*	189500659	Chlorobi	Bacteria	389
Pae1	*Prosthecochloris aestuarii DSM 271*	194334412	Chlorobi	Bacteria	390
Pst1	*Pseudomonas stutzeri A1501*	146281974	Gammaproteobacteria	Bacteria	386
Psp1	*Photobacterium sp. SKA34*	89072547	Gammaproteobacteria	Bacteria	380
Ama1	*Alteromonas macleodii ATCC 27126*	239993177	Gammaproteobacteria	Bacteria	384
Orf1	*gamma proteobacterium HTCC5015*	254449112	Gammaproteobacteria	Bacteria	385
Bma1	*Bermanella marisrubri*	94499226	Gammaproteobacteria	Bacteria	405
Ptu1	*Pseudoalteromonas tunicata D2*	88859983	Gammaproteobacteria	Bacteria	377
Ahy1	*Aeromonas hydrophila subsp. hydrophila ATCC 7966*	117620245	Gammaproteobacteria	Bacteria	384
Msp1	*Moritella sp. PE36*	149909267	Gammaproteobacteria	Bacteria	383
Asu1	*Amycolicicoccus subflavus DQS3-9A1*	333918251	Actinobacteria	Bacteria	449
Iva1	*Isoptericola variabilis 225*	334336336	Actinobacteria	Bacteria	418
Cli1	*Congregibacter litoralis KT71*	88704233	Gammaproteobacteria	Bacteria	422
Cge1	*Corynebacterium genitalium ATCC*	300780328	Actinobacteria	Bacteria	453

	33030				
Ama3	*Acaryochloris marina MBIC11017*	158338405	Cyanobacteria	Bacteria	388
Rba1	*Rhodopirellula baltica SH 1*	32476763	Planctomycetes	Bacteria	394
Lma1	*Lyngbya majuscula 3L*	332709434	Cyanobacteria	Bacteria	379
Csp1	*Cyanothece sp. PCC 8801*	218246349	Cyanobacteria	Bacteria	388
Cwa1	*Crocosphaera watsonii WH 8501*	67921248	Cyanobacteria	Bacteria	382
Ama2	*Arthrospira maxima CS-328*	209528298	Cyanobacteria	Bacteria	362
Psp2	*Pseudovibrio sp. JE062*	254473624	Alphaproteobacteria	Bacteria	394
Average protein size (aas)					395
SD					23
Cluster 2					
Tps1	*Thalassiosira pseudonana CCMP1335*	224008354	Bacillariophyta	Eukaryota	624
Tps2	*Thalassiosira pseudonana CCMP1335*	224008010	Bacillariophyta	Eukaryota	605
Ptr1	*Phaeodactylum tricornutum CCAP 1055/1*	219129079	Bacillariophyta	Eukaryota	571
Aeh1	*Alkalilimnicola ehrlichii MLHE-1*	114320892	Gammaproteobacteria	Bacteria	413
Orf4	*uncultured gamma proteobacterium*	297172079	Gammaproteobacteria	Bacteria	413

	HF0770_07M15				
Asp1	*Algoriphagus sp. PR1*	311745175	Bacteroidetes	Bacteria	368
Mtr1	*Marivirga tractuosa DSM 4126*	313675848	Bacteroidetes	Bacteria	369
Mma1	*Microscilla marina ATCC 23134*	124006152	Bacteroidetes	Bacteria	387
Average protein size (aas)					469
SD					111
Cluster 3					
Cef1	*Corynebacterium efficiens YS-314*	259505867	Actinobacteria	Bacteria	452
Cgl1	*Corynebacterium glutamicum ATCC 13032*	62391580	Actinobacteria	Bacteria	436
Average protein size (aas)					444
SD					11
Cluster 4					
Smo1	*Solobacterium moorei F0204*	320527361	Firmicutes	Bacteria	562
Ala1	*Anaerococcus lactolyticus ATCC 51172*	227485742	Firmicutes	Bacteria	563
Average protein size (aas)					563
SD					1
Cluster 5					
Ban1	*Bacillus anthracis str. Ames*	30260929	Firmicutes	Bacteria	551
Bha2	*Bacillus halodurans C-*	15613970	Firmicutes	Bacteria	543

	125				
Lmo1	*Listeria monocytogenes serotype 4b str. F2365*	46907059	Firmicutes	Bacteria	544
Cth1	*Caldalkalibacillus thermarum TA2.A1*	335040185	Firmicutes	Bacteria	537
Seq1	*Streptococcus equi subsp. zooepidemicus*	225869078	Firmicutes	Bacteria	543
Spn1	*Streptococcus pneumoniae TIGR4*	15900410	Firmicutes	Bacteria	543
Ppi1	*Pyramidobacter piscolens W5455*	282855997	Synergistetes	Bacteria	547
Apa1	*Aminomonas paucivorans DSM 12260*	312879543	Synergistetes	Bacteria	538
Hor1	*Halothermothrix orenii H 168*	220932998	Firmicutes	Bacteria	548
Aar1	*Acetohalobium arabaticum DSM 5501*	302392317	Firmicutes	Bacteria	550
Asp2	*Acidaminococcus sp. D21*	227498921	Firmicutes	Bacteria	541
Psp3	*Phascolarctobacterium sp. YIT 12067*	323142438	Firmicutes	Bacteria	535
Average protein size (aas)					543
SD					5
Cluster 6					
Eve1	*Eubacterium ventriosum ATCC 27560*	154483404	Firmicutes	Bacteria	582
Ece1	*Eubacterium*	313905594	Firmicutes	Bacteria	580

	cellulosolvens 6				
Bha1	*Blautia hansenii DSM 20583*	260587255	Firmicutes	Bacteria	586
Rob1	*Ruminococcus obeum ATCC 29174*	153813305	Firmicutes	Bacteria	588
Cca1	*Coprococcus catus GD/7*	291521247	Firmicutes	Bacteria	596
Orf2	*butyrate-producing bacterium SS3/4*	291561247	Firmicutes	Bacteria	591
Eel1	*Eubacterium eligens ATCC 27750*	238917553	Firmicutes	Bacteria	596
Esi1	*Eubacterium siraeum DSM 15702*	167750333	Firmicutes	Bacteria	592
Rin1	*Roseburia inulinivorans DSM 16841*	225378731	Firmicutes	Bacteria	587
Ere1	*Eubacterium rectale M104/1*	291528607	Firmicutes	Bacteria	594
Ade1	*Abiotrophia defectiva ATCC 49176*	229825925	Firmicutes	Bacteria	636
Rsp1	*Ruminococcus sp. 18P13*	291543460	Firmicutes	Bacteria	583
Cph4	*Clostridium phytofermentans ISDg*	160880956	Firmicutes	Bacteria	590
Sva1	*Subdoligranulum variabile DSM 15176*	282601503	Firmicutes	Bacteria	608
Fpr1	*Faecalibacterium prausnitzii A2-165*	257439800	Firmicutes	Bacteria	600
Rin2	*Roseburia inulinivorans DSM 16841*	225375925	Firmicutes	Bacteria	598

Aco1	*Anaerotruncus colihominis DSM 17241*	167770850	Firmicutes	Bacteria	592
Ral1	*Ruminococcus albus 8*	325679893	Firmicutes	Bacteria	599
Average protein size (aas)					594
SD					13
Cluster 7					
Csy1	*Clostridium symbiosum WAL-14163*	323483835	Firmicutes	Bacteria	558
Cha2	*Clostridium hathewayi DSM 13479*	266621572	Firmicutes	Bacteria	555
Orf6	*butyrate-producing bacterium SS3/4*	291563870	Firmicutes	Bacteria	555
Bpe1	*Bacteroides pectinophilus ATCC 43243*	218133720	Bacteroidetes	Bacteria	556
Csy2	*Clostridium symbiosum WAL-14163*	323484206	Firmicutes	Bacteria	552
Csa1	*Clostridium saccharolyticum WM1*	302386531	Firmicutes	Bacteria	551
Osp1	*Oribacterium sp. oral taxon 078 str. F0262*	291459750	Firmicutes	Bacteria	563
Csp2	*Clostridium sp. SS2/1*	167767883	Firmicutes	Bacteria	559
Csc2	*Clostridium scindens ATCC 35704*	167759895	Firmicutes	Bacteria	546
Cle2	*Clostridium lentocellum DSM 5427*	326791913	Firmicutes	Bacteria	562
Cst1	*Clostridium sticklandii*	310657910	Firmicutes	Bacteria	542

	DSM 519				
Fal1	*Filifactor alocis ATCC 35896*	291166322	Firmicutes	Bacteria	538
Cba1	*Clostridium bartlettii DSM 16795*	164686723	Firmicutes	Bacteria	542
Chi1	*Clostridium hiranonis DSM 13275*	210621729	Firmicutes	Bacteria	535
Pan1	*Peptostreptococcus anaerobius 653-L*	289424100	Firmicutes	Bacteria	540
Aor2	*Alkaliphilus oremlandii OhILAs*	158321134	Firmicutes	Bacteria	541
Mvo1	*Methanococcus voltae A3*	297619354	Euryarchaeota	Archaea	545
Fma1	*Finegoldia magna ATCC 29328*	169824491	Firmicutes	Bacteria	532
Cbo1	*Clostridium botulinum B str. Eklund 17B*	187935756	Firmicutes	Bacteria	533
Cpe1	*Clostridium perfringens str. 13*	18311350	Firmicutes	Bacteria	537
Cbo2	*Clostridium botulinum D str. 1873*	253681532	Firmicutes	Bacteria	538
Cle1	*Clostridium leptum DSM 753*	160932730	Firmicutes	Bacteria	556
Cph3	*Clostridium phytofermentans ISDg*	160880899	Firmicutes	Bacteria	554
Orf5	*butyrate-producing bacterium SSC/2*	291559010	Firmicutes	Bacteria	560
Csp3	*Clostridium sp. D5*	325265388	Firmicutes	Bacteria	550
Lba1	*Lachnospiraceae bacterium 4_1_37FAA*	325662187	Firmicutes	Bacteria	542
Cne1	*Clostridium nexile*	210609822	Firmicutes	Bacteria	543

	DSM 1787				
Csc1	*Clostridium scindens ATCC 35704*	167758777	Firmicutes	Bacteria	540
Cha1	*Clostridium hathewayi DSM 13479*	266619978	Firmicutes	Bacteria	554
Eba1	*Erysipelotrichaceae bacterium 5_2_54FAA*	293402293	Firmicutes	Bacteria	558
Average protein size (aas)					548
SD					9
Cluster 8					
Ele1	*Eggerthella lenta DSM 2243*	257790704	Actinobacteria	Bacteria	587
Cluster 9					
Dth1	*Dictyoglomus thermophilum H-6-12*	206901685	Dictyoglomi	Bacteria	548
Cluster 10					
Cph2	*Chlorobium phaeovibrioides DSM 265*	145219613	Chlorobi	Bacteria	565
Clu1	*Chlorobium luteolum DSM 273*	78187008	Chlorobi	Bacteria	555
Pph1	*Pelodictyon phaeoclathratiforme BU-1*	194336325	Chlorobi	Bacteria	564
Cfl1	*Chthoniobacter flavus Ellin428*	196234084	Verrucomicrobia	Bacteria	570
Average protein size (aas)					564
SD					6

Cluster 11					
Pue1	*Porphyromonas uenonis 60-3*	228470933	Bacteroidetes	Bacteria	605
Cluster 12					
Odi1	*Oikopleura dioica*	313243884	Metazoa	Eukaryota	603
Cluster 13					
Spu5	*Strongylocentrotus purpuratus*	115663105	Metazoa	Eukaryota	561
Cluster 14					
Pin1	*Phytophthora infestans T30-4*	301105659	none	Eukaryota	516
Pin3	*Phytophthora infestans T30-4*	301123313	none	Eukaryota	513
Odi2	*Oikopleura dioica*	313229525	Metazoa	Eukaryota	654
Odi3	*Oikopleura dioica*	313242423	Metazoa	Eukaryota	652
Odi8	*Oikopleura dioica*	313242874	Metazoa	Eukaryota	654
Average protein size (aas)					596
SD					59
Cluster 15					
Cre1	*Caenorhabditis remanei*	308511717	Metazoa	Eukaryota	631
Cluster 16					
Isc1	*Ixodes scapularis*	241752738	Metazoa	Eukaryota	478
Tsp2	*Trichinella spiralis*	316974125	Metazoa	Eukaryota	676
Average protein size (aas)					577

SD					140
Cluster 17					
Cin1	*Ciona intestinalis*	198433216	Metazoa	Eukaryota	629
Cin4	*Ciona intestinalis*	198413159	Metazoa	Eukaryota	616
Cin2	*Ciona intestinalis*	198428774	Metazoa	Eukaryota	595
Cin3	*Ciona intestinalis*	198437634	Metazoa	Eukaryota	604
Average protein size (aas)					611
SD					15
Cluster 18					
Sko2	*Saccoglossus kowalevskii*	291244118	Metazoa	Eukaryota	895
Sko3	*Saccoglossus kowalevskii*	291244124	Metazoa	Eukaryota	706
Average protein size (aas)					801
SD					134
Cluster 19					
Cja1	*Callithrix jacchus*	296196772	Metazoa	Eukaryota	690
Oan1	*Ornithorhynchus anatinus*	149412321	Metazoa	Eukaryota	692
Gga1	*Gallus gallus*	46048932	Metazoa	Eukaryota	674
Xla1	*Xenopus laevis*	285403622	Metazoa	Eukaryota	674
Dre3	*Danio rerio*	33504533	Metazoa	Eukaryota	642
Dre4	*Danio rerio*	32822845	Metazoa	Eukaryota	682
Dre2	*Danio rerio*	66910279	Metazoa	Eukaryota	631
Pam1	*Pseudopleuronectes americanus*	1089908	Metazoa	Eukaryota	637
Ssa1	*Salmo salar*	213513808	Metazoa	Eukaryota	692
Dre1	*Danio rerio*	292625673	Metazoa	Eukaryota	562

Dre5	*Danio rerio*	326673248	Metazoa	Eukaryota	645
Gga2	*Gallus gallus*	118097360	Metazoa	Eukaryota	675
Dvi1	*Didelphis virginiana*	425469	Metazoa	Eukaryota	653
Mdo4	*Monodelphis domestica*	126331894	Metazoa	Eukaryota	611
Mdo5	*Monodelphis domestica*	126331888	Metazoa	Eukaryota	628
Mdo2	*Monodelphis domestica*	334331779	Metazoa	Eukaryota	553
Mdo3	*Monodelphis domestica*	334331773	Metazoa	Eukaryota	652
Mdo1	*Monodelphis domestica*	334331775	Metazoa	Eukaryota	560
Ame1	*Ailuropoda melanoleuca*	301781530	Metazoa	Eukaryota	602
Mmu2	*Macaca mulatta*	109109654	Metazoa	Eukaryota	600
Orf3	*Xenopus (Silurana) tropicalis*	301622941	Metazoa	Eukaryota	637
Dpu1	*Daphnia pulex*	321463341	Metazoa	Eukaryota	613
Dpu3	*Daphnia pulex*	321463340	Metazoa	Eukaryota	634
Dpu2	*Daphnia pulex*	321469787	Metazoa	Eukaryota	575
Dpu4	*Daphnia pulex*	321463339	Metazoa	Eukaryota	608
Isc2	*Ixodes scapularis*	242000082	Metazoa	Eukaryota	636
Bfl1	*Branchiostoma floridae*	260821227	Metazoa	Eukaryota	614
Sko4	*Saccoglossus kowalevskii*	291243895	Metazoa	Eukaryota	687
Sko1	*Saccoglossus kowalevskii*	291241855	Metazoa	Eukaryota	638
Spu2	*Strongylocentrotus purpuratus*	115741862	Metazoa	Eukaryota	721

Spu1	*Strongylocentrotus purpuratus*	72135890	Metazoa	Eukaryota	604
Tad1	*Trichoplax adhaerens*	195999576	Metazoa	Eukaryota	606
Average protein size (aas)					635
SD					42

Table S7. 74 PnuC protein sequences included in this study. Proteins are listed clockwise starting from Cluster 1 (see Fig. S7).

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Lre1	*Lactobacillus reuteri 100-23*	194467100	Firmicutes	Bacteria	241
Lfe1	*Lactobacillus fermentum IFO 3956*	184156132	Firmicutes	Bacteria	240
Lla2	*Lactococcus lactis subsp. cremoris SK11*	116511656	Firmicutes	Bacteria	237
Average protein size (aas)					239
SD					2
Cluster 2					
Cla1	*Campylobacter lari RM2100*	222823375	Epsilonproteobacteria	Bacteria	219
Hpy1	*Helicobacter pylori 26695*	15645903	Epsilonproteobacteria	Bacteria	220
Sde2	*Sulfurospirillum delevignum DSM 6946*	268679606	Epsilonproteobacteria	Bacteria	200
Eha1	*Eubacterium hallii DSM 3353*	225025810	Firmicutes	Bacteria	227
Seq1	*Streptococcus equi subsp. zooepidemicus*	225868279	Firmicutes	Bacteria	234
Lla3	*Lactococcus lactis subsp. cremoris SK11*	116511706	Firmicutes	Bacteria	241
Average protein size (aas)					223

SD					14
Cluster 3					
Lci1	*Leuconostoc citreum KM20*	170017105	Firmicutes	Bacteria	264
Lar1	*Leuconostoc argentinum KCTC 3773*	326693493	Firmicutes	Bacteria	263
Lpl1	*Lactobacillus plantarum WCFS1*	28377442	Firmicutes	Bacteria	273
Wci2	*Weissella cibaria KACC 11862*	332637196	Firmicutes	Bacteria	259
Lla1	*Lactococcus lactis*	146322280	Firmicutes	Bacteria	249
Lph1	*Lactobacillus phage LP65*	56693131	Myoviridae	Viruses	256
Wci1	*Weissella cibaria KACC 11862*	332638280	Firmicutes	Bacteria	244
Average protein size (aas)					258
SD					10
Cluster 4					
Spn1	*Streptococcus pneumoniae TIGR4*	15901687	Firmicutes	Bacteria	269
Smi1	*Streptococcus mitis NCTC 12261*	307707857	Firmicutes	Bacteria	259
Ssu1	*Streptococcus suis 05ZYH33*	146319626	Firmicutes	Bacteria	273
Smu1	*Streptococcus mutans UA159*	24379389	Firmicutes	Bacteria	259
Cmo1	*Catonella morbi ATCC 51271*	229823079	Firmicutes	Bacteria	278

Mca1	*Macrococcus caseolyticus JCSC5402*	222151695	Firmicutes	Bacteria	245
Average protein size (aas)					264
SD					12
Cluster 5					
Efa1	*Enterococcus faecalis V583*	29375333	Firmicutes	Bacteria	226
Cluster 6					
Lme1	*Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293*	116618612	Firmicutes	Bacteria	255
Cluster 7					
Cth1	*Clostridium thermocellum ATCC 27405*	125975259	Firmicutes	Bacteria	234
Cbo1	*Clostridium bolteae ATCC BAA-613*	160937996	Firmicutes	Bacteria	231
Bpr1	*Butyrivibrio proteoclasticus B316*	302669157	Firmicutes	Bacteria	236
Ceu1	*Coprococcus eutactus ATCC 27759*	163814331	Firmicutes	Bacteria	244
Bmu1	*Brachyspira*	296126138	Spirochaetes	Bacteria	230

	murdochii DSM 12563				
Cas1	*Clostridium asparagiforme DSM 15981*	225405522	Firmicutes	Bacteria	235
Average protein size (aas)					235
SD					5
Cluster 8					
Csc1	*Clostridium scindens ATCC 35704*	167758545	Firmicutes	Bacteria	230
Tsa1	*Turicibacter sanguinis PC909*	293376974	Firmicutes	Bacteria	229
Dfo1	*Dorea formicigenerans ATCC 27755*	166032332	Firmicutes	Bacteria	227
Cbo2	*Clostridium bolteae ATCC BAA-613*	160940289	Firmicutes	Bacteria	228
Bbr1	*Brevibacillus brevis NBRC 100599*	226311114	Firmicutes	Bacteria	237
Ste1	*Sebaldella termitidis ATCC 33386*	269121730	Fusobacteria	Bacteria	230
Hco1	*Haloplasma contractile SSD-17B*	334893023	Tenericutes	Bacteria	252
Bsp1	*Bacillus sp. NRRL B-14911*	89099838	Firmicutes	Bacteria	231

Average protein size (aas)					233
SD					8
Cluster 9					
Apl1	*Actinobacillus pleuropneumoniae serovar 1 str. 4074*	53729065	Gammaproteobacteria	Bacteria	230
Hin2	*Haemophilus influenzae 86-028NP*	68249634	Gammaproteobacteria	Bacteria	226
Pmu1	*Pasteurella multocida subsp. multocida str. Pm70*	15603703	Gammaproteobacteria	Bacteria	234
Kde1	*Kingella denitrificans ATCC 33394*	325266270	Betaproteobacteria	Bacteria	229
Gan1	*Gallibacterium anatis UMN179*	332289165	Gammaproteobacteria	Bacteria	227
Nmu1	*Neisseria mucosa ATCC 25996*	288576217	Betaproteobacteria	Bacteria	263
Msu1	*Mannheimia succiniciproducens MBEL55E*	52424248	Gammaproteobacteria	Bacteria	241
Average protein size (aas)					236
SD					13
Cluster 10					
Cle1	*Clostridium lentocellum DSM 5427*	326792984	Firmicutes	Bacteria	228

Psp3	*Paenibacillus sp. Y412MC10*	261410044	Firmicutes	Bacteria	225
Hhy1	*Haliscomenobacter hydrossis DSM 1100*	332662236	Bacteroidetes	Bacteria	226
Dal1	*Desulfatibacillum alkenivorans AK-01*	218781325	Deltaproteobacteria	Bacteria	230
Average protein size (aas)					227
SD					2
Cluster 11					
Orf1	*synthetic construct*	62261752	none	Unclassified	289
Eph1	*Enterobacteria phage EPS7*	182682884	Siphoviridae	Viruses	225
Average protein size (aas)					257
SD					45
Cluster 12					
Phe1	*Pedobacter heparinus DSM 2366*	255530038	Bacteroidetes	Bacteria	204
Mpa1	*Mucilaginibacter paludis DSM 18603*	312889447	Bacteroidetes	Bacteria	210
Psp1	*Pedobacter sp. BAL39*	149277331	Bacteroidetes	Bacteria	205
Ssp2	*Sphingobacterium spiritivorum ATCC 33300*	227537640	Bacteroidetes	Bacteria	199
Dfe1	*Dyadobacter*	255037445	Bacteroidetes	Bacteria	209

	fermentans DSM 18053				
Sgr1	*Streptomyces griseus subsp. griseus NBRC 13350*	182439523	Actinobacteria	Bacteria	226
Rbe1	*Rubrivivax benzoatilyticus JA2*	332526687	Betaproteobacteria	Bacteria	213
Hne1	*Hyphomonas neptunium ATCC 15444*	114798743	Alphaproteobacteria	Bacteria	218
Cma1	*Corynebacterium matruchotii ATCC 14266*	305680194	Actinobacteria	Bacteria	232
Hau1	*Herpetosiphon aurantiacus ATCC 23779*	159897043	Chloroflexi	Bacteria	210
Spr1	*Serinicoccus profundi*	379058816	Actinobacteria	Bacteria	241
Vvu1	*Vibrio vulnificus YJ016*	37675870	Gammaproteobacteria	Bacteria	250
Vsp1	*Vibrio sp. Ex25*	254227586	Gammaproteobacteria	Bacteria	243
Sen1	*Salmonella enterica subsp. enterica serovar Typhimurium str. LT2*	16764122	Gammaproteobacteria	Bacteria	239
Eco1	*Escherichia coli O157:H7 str. EDL933*	15800460	Gammaproteobacteria	Bacteria	239
Plu1	*Photorhabdus*	37525421	Gammaproteobacteria	Bacteria	241

	luminescens subsp. laumondii TTO1				
Pat1	*Pseudoalteromonas atlantica T6c*	109897078	Gammaproteobacteria	Bacteria	212
Asp1	*Acidobacterium sp. MP5ACTX9*	322435112	Acidobacteria	Bacteria	211
Ban1	*Bacillus anthracis str. A2012*	65317639	Firmicutes	Bacteria	218
Cpe1	*Clostridium perfringens str. 13*	18309381	Firmicutes	Bacteria	225
Average protein size (aas)					222
SD					15
Cluster 13					
Fcf1	*Francisella cf. novicida Fx1*	332678310	Gammaproteobacteria	Bacteria	210
Ssp1	*Sphingobacterium spiritivorum ATCC 33300*	227539605	Bacteroidetes	Bacteria	212
Aur1	*Aerococcus urinae ACS-120-V-Col10a*	326803505	Firmicutes	Bacteria	208
Average protein size (aas)					210
SD					2
Cluster 14					
Sci1	*Spiroplasma citri*	110005352	Tenericutes	Bacteria	302
Bex1	*Bulleidia extructa W1219*	283768398	Firmicutes	Bacteria	403
Average protein size (aas)					353
SD					71

Table S8A. 41 of the 97 GPCR protein sequences included in this study and used for the phylogenetic tree shown in Fig. S8B. Proteins are listed clockwise starting from Cluster 1. Proteins represented by Dre4 of cluster 10 are presented in Table S8B.

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**	**Functional Type**
Cluster 1						
CALCR	*Homo sapiens*	399180	Metazoa	Eukaryota	490	Calcitonin Receptor
PDFR	*Drosophila melanogaster*	74949071	Metazoa	Eukaryota	669	PDF Receptor
CRFR1	*Homo sapiens*	461836	Metazoa	Eukaryota	444	Corticotropin-releasing factor receptor
DIHR	*Acheta domesticus*	2495063	Metazoa	Eukaryota	441	Diuretic hormone receptor
GIPR	*Homo sapiens*	1346133	Metazoa	Eukaryota	466	Gastric Inhibitory polypeptide receptor
GLP1R	*Homo sapiens*	311033387	Metazoa	Eukaryota	463	Glucagon-like peptide receptor
GHRHR	*Homo sapiens*	3041685	Metazoa	Eukaryota	423	Gwroth hormone-releasing hormone receptor
PACR	*Homo sapiens*	1171986	Metazoa	Eukaryota	468	Pituitary adenylate cyclase-activating polypeptide receptor
VIPR1	*Homo sapiens*	418253	Metazoa	Eukaryota	457	Basoactive intestinal polypeptide receptor
SCTR	*Homo sapiens*	2506489	Metazoa	Eukaryota	440	Secretin receptor
PTH1R	*Homo sapiens*	417555	Metazoa	Eukaryota	593	Parathyroid hormone peptide receptor
Average protein size: (aas)					487	
SD					75	
Cluster 2						
BAI2	*Homo sapiens*	229462818	Metazoa	Eukaryota	1585	Brain-specific angiogenesis inhibitor
Average protein size: (aas)					1585	
SD					0	
Cluster 3						
EMR1	*Homo sapiens*	290457673	Metazoa	Eukaryota	886	EGF-like module-containing mucin-like hormone receptor-like
CD97	*Homo sapiens*	90110013	Metazoa	Eukaryota	835	CD97 antigen
ELTD1	*Homo sapiens*	212276505	Metazoa	Eukaryota	690	EGF and latrophilin domain-containing protein
CELR1	*Homo sapiens*	22095551	Metazoa	Eukaryota	3014	Cadherin EGF LAG seven-pass G-type receptor

Average protein size: (aas)					1356	
SD					1108	
Cluster 4						
LT4R1	*Homo sapiens*	3041713	Metazoa	Eukaryota	352	Leukotriene B4 receptor
Average protein size: (aas)					352	
SD					0	
Cluster 5						
MTH10	*Drosophila melanogaster*	166215065	Metazoa	Eukaryota	585	Methuselah-like receptor
GABR2	*Homo sapiens*	12643641	Metazoa	Eukaryota	941	Gamma-aminobutyric acid type B receptor
VN1R1	*Homo sapiens*	71153399	Metazoa	Eukaryota	353	Vomeronasal type-1 receptor
GRM1	*Homo sapiens*	311033443	Metazoa	Eukaryota	1194	Metabotropic glutamate receptor
CASR	*Homo sapiens*	1168781	Metazoa	Eukaryota	1078	Extracellular calcium-sensing receptor
GPC6A	*Homo sapiens*	74745292	Metazoa	Eukaryota	926	Odorant receptor
TS1R2	*Homo sapiens*	116242831	Metazoa	Eukaryota	839	Taste receptor type 1
CAR1	*Dictyostelium discoideum*	115617	None	Eukaryota	392	Cyclic AMP receptor
Average protein size: (aas)					789	
SD					312	
Cluster 6						
GNRHR	*Homo sapiens*	399777	Metazoa	Eukaryota	328	Gonadotropin-releasing hormone receptor
MTR1A	*Homo sapiens*	1346544	Metazoa	Eukaryota	350	Melatonin receptor
Average protein size: (aas)					339	
SD					16	
Cluster 7						

ACM1	*Homo sapiens*	113118	Metazoa	Eukaryota	460	Muscarinic acetylcholine receptor
AA1R	*Homo sapiens*	231473	Metazoa	Eukaryota	326	Adenosine receptor
Average protein size: (aas)					393	
SD					95	
Cluster 8						
TRFR	*Homo sapiens*	464921	Metazoa	Eukaryota	398	Thyrotropin-releasing hormone receptor
Average protein size: (aas)					398	
SD					0	
Cluster 9						
AGTR1	*Homo sapiens*	231519	Metazoa	Eukaryota	359	Angiotensin II receptor
CLTR1	*Homo sapiens*	20138087	Metazoa	Eukaryota	337	Cysteinyl leukotriene receptor
PTAFR	*Homo sapiens*	129557	Metazoa	Eukaryota	342	Platelet-activating factor receptor
HCAR1	*Homo sapiens*	47117701	Metazoa	Eukaryota	346	Hydroxycarboxylic acid receptor
Average protein size: (aas)					346	
SD					9	
Cluster 10						
OPSD	*Homo sapiens*	129207	Metazoa	Eukaryota	348	Rhodopsin
Cja1	*Callithrix jacchus*	296228403	Metazoa	Eukaryota	354	Opsin
Average protein size: (aas)					351	
SD					4	
Cluster 11						
FSHR	*Homo sapiens*	311033420	Metazoa	Eukaryota	695	Follicle-stimulating hormone receptor
Average protein size: (aas)					695	
SD					0	

Cluster 12						
OR1E1	*Homo sapiens*	400671	Metazoa	Eukaryota	314	Olfactory receptor
CNR1	*Homo sapiens*	115562	Metazoa	Eukaryota	472	Cannabinoid receptor
S1PR1	*Homo sapiens*	205371820	Metazoa	Eukaryota	382	Sphingosine 1-phosphate receptor
Average protein size: (aas)					389	
SD					79	
Cluster 13						
PE2R1	*Homo sapiens*	83287886	Metazoa	Eukaryota	402	Prostaglandin E2 receptor
Average protein size: (aas)					402	
SD					0	

Table S8B. 56 of the 97 GPCR protein sequences included in this study for the phylogenetic tree shown in Fig. S8E. Proteins are listed clockwise starting from Cluster 1.

Abbreviation	**Organism**	**GenBank Index#**	**Group**	**Domain**	**Protein Size**
Cluster 1					
Dre1	*Danio rerio*	292621989	Metazoa	Eukaryota	354
Dre4	*Danio rerio*	113682063	Metazoa	Eukaryota	355
Dre3	*Danio rerio*	326679306	Metazoa	Eukaryota	362
Orf1	*Xenopus (Silurana) tropicalis*	301607740	Metazoa	Eukaryota	339
Mmu1	*Mus musculus*	148677107	Metazoa	Eukaryota	338
Cja3	*Callithrix jacchus*	296225024	Metazoa	Eukaryota	333
Mdo4	*Monodelphis domestica*	126341584	Metazoa	Eukaryota	352
Oan3	*Ornithorhynchus anatinus*	149632071	Metazoa	Eukaryota	351
Aca3	*Anolis carolinensis*	327282149	Metazoa	Eukaryota	357
Gga5	*Gallus gallus*	124249288	Metazoa	Eukaryota	377
Average protein size: (aas)					352
SD					13
Cluster 2					
Pol1	*Paralichthys olivaceus*	34740123	Metazoa	Eukaryota	360
Omo1	*Osmerus mordax*	225706150	Metazoa	Eukaryota	356
Ssa1	*Salmo salar*	291190313	Metazoa	Eukaryota	371
Dre5	*Danio rerio*	292621989	Metazoa	Eukaryota	384
Orf2	*Xenopus (Silurana) tropicalis*	166157823	Metazoa	Eukaryota	345
Average protein size: (aas)					363
SD					15
Cluster 3					
Mmu2	*Macaca mulatta*	23305858	Metazoa	Eukaryota	355
Ocu1	*Oryctolagus cuniculus*	283806620	Metazoa	Eukaryota	358
Clu3	*Canis lupus familiaris*	52546730	Metazoa	Eukaryota	359

Xla1	*Xenopus laevis*	148238158	Metazoa	Eukaryota	352
Cja4	*Callithrix jacchus*	9587726	Metazoa	Eukaryota	347
Ptr1	*Pan troglodytes*	332816585	Metazoa	Eukaryota	436
Average protein size: (aas)					368
SD					34
Cluster 4					
Dre2	*Danio rerio*	66911140	Metazoa	Eukaryota	312
Tni1	*Tetraodon nigroviridis*	47224097	Metazoa	Eukaryota	309
Dre6	*Danio rerio*	292629502	Metazoa	Eukaryota	352
Average protein size: (aas)					324
SD					24
Cluster 5					
Aca2	*Anolis carolinensis*	327282179	Metazoa	Eukaryota	360
Gga2	*Gallus gallus*	118086158	Metazoa	Eukaryota	358
Mdo5	*Monodelphis domestica*	126341642	Metazoa	Eukaryota	354
Average protein size: (aas)					357
SD					3
Cluster 6					
Cja1	*Callithrix jacchus*	296228403	Metazoa	Eukaryota	354
Mdo2	*Monodelphis domestica*	126341638	Metazoa	Eukaryota	336
Ymo1	*Yaba monkey tumor virus*	38229175	Poxviridae	Viruses	346
Gga3	*Gallus gallus*	71896604	Metazoa	Eukaryota	377
Zal1	*Zonotrichia albicollis*	296399392	Metazoa	Eukaryota	377
Gga6	*Gallus gallus*	113951667	Metazoa	Eukaryota	335
Average protein size: (aas)					354
SD					19

Cluster 7					
Cja2	*Callithrix jacchus*	296225031	Metazoa	Eukaryota	359
Clu2	*Canis lupus familiaris*	57111007	Metazoa	Eukaryota	358
Pab1	*Pongo abelii*	207080176	Metazoa	Eukaryota	352
Lca1	*Lemur catta*	5713071	Metazoa	Eukaryota	339
Oan4	*Ornithorhynchus anatinus*	149632073	Metazoa	Eukaryota	348
Oan5	*Ornithorhynchus anatinus*	149632065	Metazoa	Eukaryota	420
Aca1	*Anolis carolinensis*	327282177	Metazoa	Eukaryota	355
Gga4	*Gallus gallus*	50732904	Metazoa	Eukaryota	354
Average protein size: (aas)					361
SD					25
Cluster 8					
Bta1	*Bos taurus*	296474745	Metazoa	Eukaryota	358
Clu1	*Canis lupus familiaris*	57101676	Metazoa	Eukaryota	372
Cja6	*Callithrix jacchus*	296225029	Metazoa	Eukaryota	373
Mdo3	*Monodelphis domestica*	126341586	Metazoa	Eukaryota	354
Oan2	*Ornithorhynchus anatinus*	149632069	Metazoa	Eukaryota	378
Oan6	*Ornithorhynchus anatinus*	149632067	Metazoa	Eukaryota	344
Mdo1	*Monodelphis domestica*	126341640	Metazoa	Eukaryota	363
Cja5	*Callithrix jacchus*	297206879	Metazoa	Eukaryota	355
Average protein size: (aas)					362
SD					12
Cluster 9					
Aca4	*Anolis carolinensis*	327282173	Metazoa	Eukaryota	344
Xla2	*Xenopus laevis*	148223850	Metazoa	Eukaryota	353
Average protein size: (aas)					349
SD					6

Cluster 10					
Ehe1	*Equid herpesvirus 2*	124738401	Herpesviridae	Viruses	387
Ehe2	*Equid herpesvirus 2*	124738361	Herpesviridae	Viruses	383
Average protein size: (aas)					385
SD					3
Cluster 11					
Gga1	*Gallus gallus*	154813804	Metazoa	Eukaryota	355
Aca5	*Anolis carolinensis*	327282147	Metazoa	Eukaryota	355
Oan1	*Ornithorhynchus anatinus*	149455250	Metazoa	Eukaryota	361
Average protein size: (aas)					357
SD					3

Table S9. The highest comparison scores between TOG superfamily (left column) and the MC and MIP families. Proteins representative of families used in these comparisons are found in tables S1-S9.

	2.A.29 **(MC) 2 TMSs**	2.A.29 **(MC) 3 TMSs**	1.A.8 **(MIP) 2 TMSs**	1.A.8 **(MIP) 3 TMSs**
9.A.14 **(GPCR)**	6.7 S.D.	N/A	5.6 S.D.	2.2 S.D.
2.A.43 **(LCT)**	8.4 S.D.	N/A	9.4 S.D.	8.8 S.D.
3.E.1 **(MR)**	8.7 S.D.	N/A	10.4 S.D.	8.6 S.D.
2.A.52 **(NiCoT)**	9.5 S.D.	9.5 S.D.	8.9 S.D.	7.4 S.D.
2.A.82 **(OST)**	9.1 S.D.	N/A	8.2 S.D.	8.2 S.D.
4.B.1 **(PnuC)**	8.8 S.D.	N/A	7.8 S.D.	7.0 S.D.
2.A.102 **(TSUP)**	9.4 S.D.	9.4 S.D.	10.2 S.D.	9.5 S.D.
2.A.58 **(PNaS)**	9.5 S.D.	N/A	10.5 S.D.	8.5 S.D.
9.A.58 **(Sweet)**	5.9 S.D.	N/A	8.8 S.D.	8.8 S.D.
Average	8.4 S.D.	9.5 S.D.	8.9 S.D.	7.7 S.D.

The highest comparison score between the MIP and MC families was 9.2 S.D. Cells shaded grey are comparison scores where the alignment contained 3 TMSs, while unshaded cells are comparison scores where the alignment contained 2 TMSs. N/A indicates that no alignment containing 3 TMSs aligned could be found.

Table S10

Table S10. The table presents HMM:HMM pairwise comparison scores indicating probabilities of homology as measured by HHsearch in the HHSuite (hhsuite-2.0.16). The similarity between Opsin (Alpha cluster of Rhodopsin GPCRs; GRAFS system of classification) and Microbial Rhodopsin (3.E.1) gave a score of 30%. In the footnote to the table, we show the consensus alignment between the HMM comparison between the Opsin cluster and the Microbial Rhodopsin family.

	2.A.102	**2.A.82**	**3.E.1**	**9.A.58**	**2.A.52**	**2.A.58**	**4.B.1**	**2.A.43**	**1.A.69**
Amine	*1%*	*0%*	*0%*	*0%*	*3%*	*0%*	*0%*	*1%*	*0%*
MECA	*0%*	*0%*	*0%*	*0%*	*2%*	*0%*	*0%*	*1%*	*0%*
Melatonin	*0%*	*6%*	*10%*	*0%*	*0%*	*1%*	*0%*	*1%*	*0%*
Opsin	*0%*	*0%*	*31%*	*0%*	*0%*	*0%*	*0%*	*0%*	*0%*
Prostaglandir	*0%*	*0%*	*0%*	*0%*	*1%*	*0%*	*0%*	*1%*	*0%*
beta	*0%*	*0%*	*0%*	*0%*	*0%*	*1%*	*0%*	*0%*	*0%*
chemokine	*0%*	*0%*	*0%*	*0%*	*1%*	*0%*	*0%*	*0%*	*0%*
MCH	*0%*	*1%*	*1%*	*1%*	*0%*	*0%*	*0%*	*1%*	*0%*
SOG	*1%*	*0%*	*0%*	*2%*	*0%*	*0%*	*1%*	*0%*	*0%*
glycoprotein	*0%*	*0%*	*1%*	*0%*	*0%*	*3%*	*0%*	*0%*	*0%*
MAS	*0%*	*1%*	*2%*	*1%*	*0%*	*1%*	*0%*	*0%*	*0%*
purin	*0%*	*1%*	*1%*	*1%*	*1%*	*1%*	*0%*	*1%*	*0%*

```
Q Consensus       226 p~~vrslf~tLr~l~vVLWl~YPIVWlLGpeG~g-vIsv~~etI~Y~ILD---LlAKv~FGfLlL~~r~~~~e~i--~i~  299 (753)
                      .++++.....|-..|.++|.=|=++=++--+|-+ .+++..++|-+-++-   +.+-++|.|.-=-.|+.+-.+.  +..
T Consensus       300 Ek~vakm~~~mi~~F~vcW~PY~~va~~~~~g~~~~~~P~ms~ipa~fAKS~tiyNPVIYv~mn~kFR~~ilql~~c~~~  379 (507)

Q Consensus       300 ~~~~~~~~~~~~g~~~e~~~~~~ee~~pet  329 (753)
                      --+|....+--+|.|+.+.-++..+.+.+.
T Consensus       380 ~c~~~~~~~~~~~~~~~~~~~~~~q~~~~~  409 (507)

Q Consensus       171 i~~LLvADI~MIVtGLiAALS~....~t~K-W~w..YtIGc--------iaflyVly~L~~~~~~aa~tkp~~--vrslf  233 (753)
                      |..+=++|++|.+++++-.++.    ..++ |.|  -..+|        .+.++-+|.|...-.|-+-..=++  ..++.
T Consensus       114 ivNlAvaDl~vs~~~~~~t~~s....~ly~~~Vf..g~~gC~~~Gf~~slfGivSm~sLt~IA~dRYivI~kP~~~~r~~  187 (507)

Q Consensus       234 --~tLr~l~vVLWl~YPIVWlLGpe  256 (753)
                        ..++.+++..|+ |-++|.+.|.
T Consensus       188 t~~a~~~i~~~iWl-ysl~walPPf  211 (507)
```