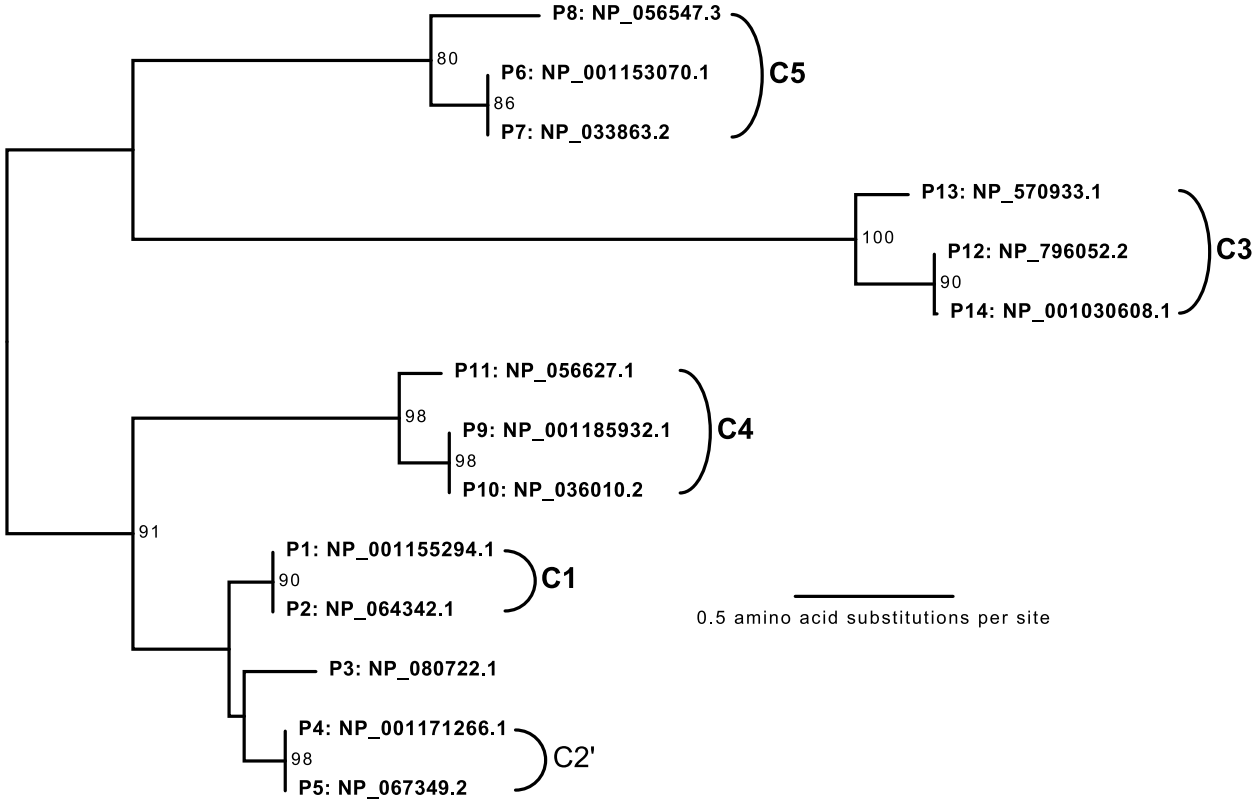

Figure S1: Maximum-likelihood phylogeny of the fourteen RGS-family proteins. A multiple sequence alignment of the commonly shared RGS domains was done using MAFFT (v7.182, Katoh and Standley, 2013) using the L-INS-i algorithm with the default parameters. The maximum likelihood phylogeny was reconstructed using PHYML (v3.1, Guindon *et al.*, 2010) using the LG amino-acid substitution model, the gamma distribution shape parameter with the maximum-likelihood estimate, and bootstrap analysis with 1,000 pseudoreplicates. Bootstrap values (%) are shown for the nodes supported by 70% or higher. Clusters are labeled C1-C5, with boldface labels when they are consistent with the ones generated by MOC.

[References]

Guindon,S. *et al*. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systems Biology*, **59**, 307-321.

Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772-780.

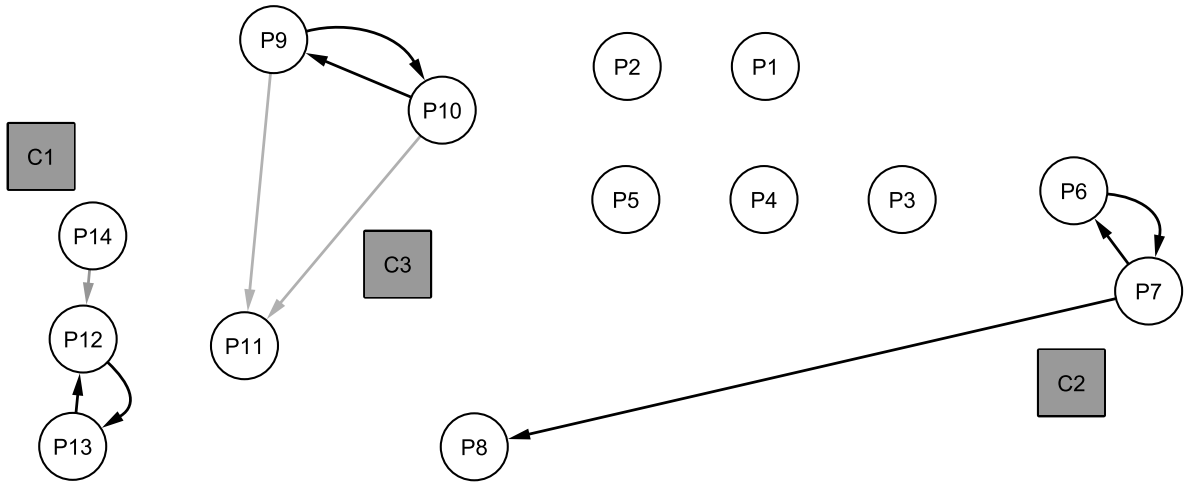

Figure S2: Secondary MOCASSIN-prot network for the fourteen RGS-family proteins. Three clusters were identified and labeled in descending order (from C1 to C3) according to their average optimal objective value. The nodes represent distinct proteins. The edges are directed so that the incoming edge weights of each node sum to 1. The edge color indicates the edge weight, from lighter (gray) to darker (black) color corresponding from lower to higher weights (edge weights are normalized to each reference protein). The optimal objective value for each protein is represented by the length of its incoming edges, with longer edges corresponding to smaller objective values (all incoming edges to a given node have the same edge length). The network was visualized using the prefuse force-directed layout in Cytoscape. Cytoscape uses the Barnes-Hut approximation algorithm to produce the “best” node layout with specified edge lengths; in some cases the incoming edge lengths are adjusted and not exactly the same.

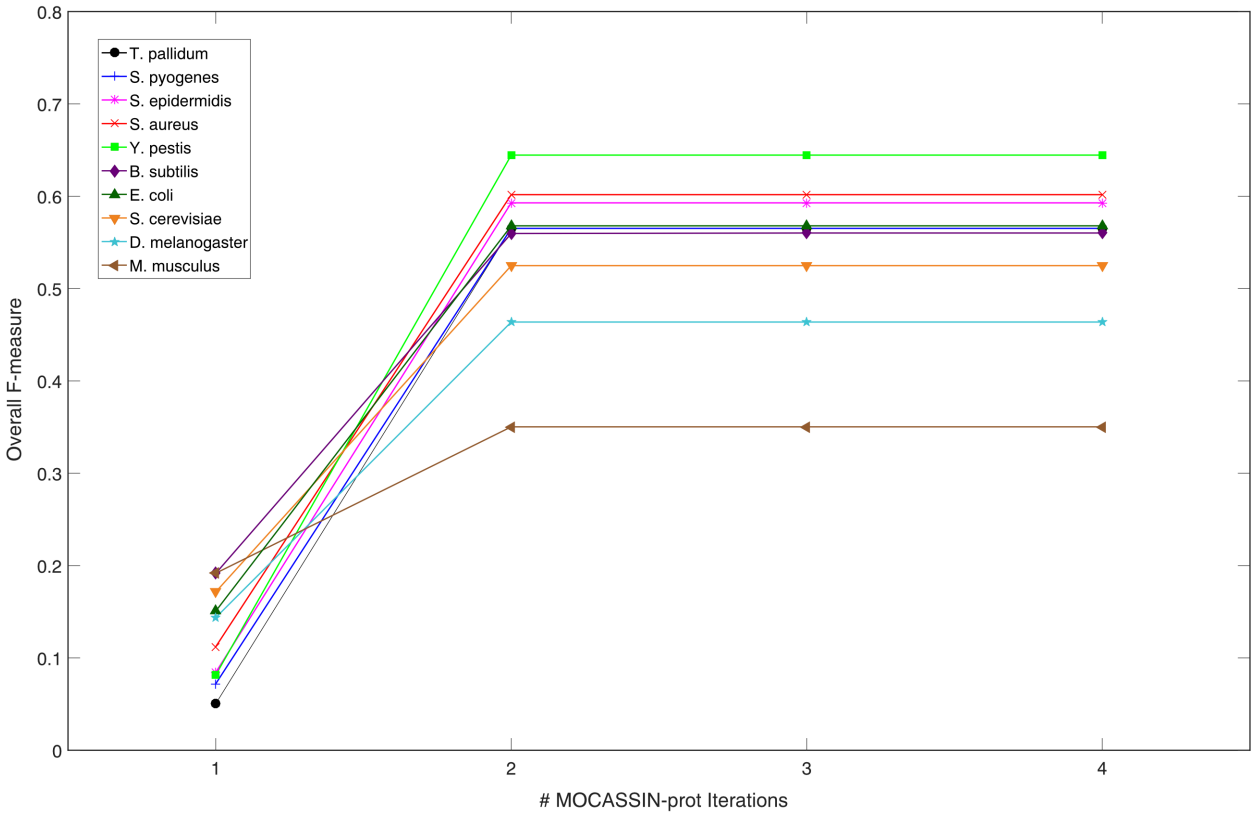

Figure S3: Clustering performance by MOCASSIN-prot with different iteration numbers. Performance was evaluated using the overall F-measure.

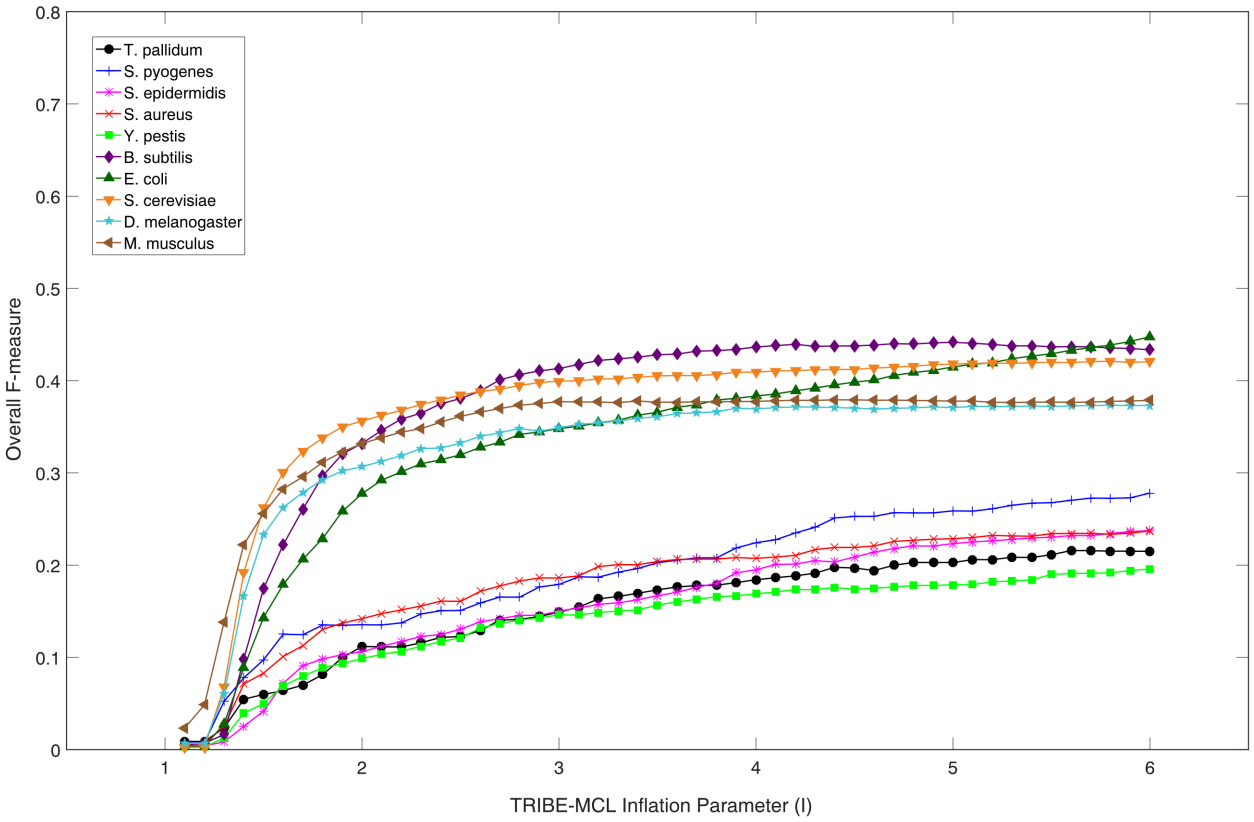

Figure S4: Clustering performance by TRIBE-MCL with different inflation values. Performance was evaluated using the overall F-measure.

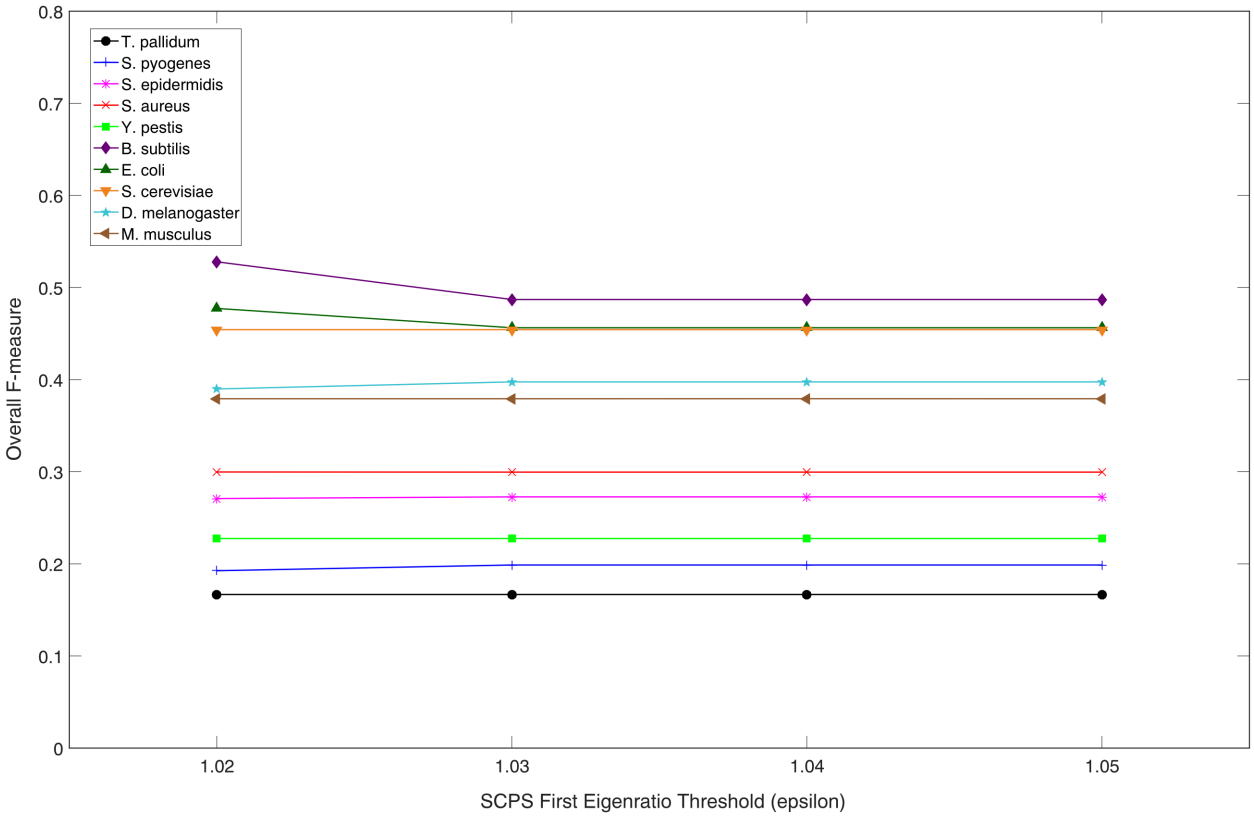

Figure S5: Clustering performance by SCPS with different *epsilon* values. Performance was evaluated using the overall F-measure.

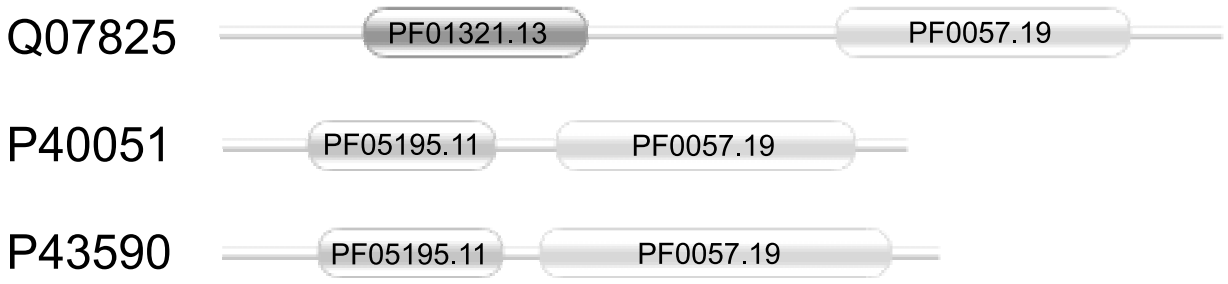

Figure S6: Domain architectures of the three *S. cerevisiae* proteins belonging to the peptidase M24B family (Q07825: the putative Xaa-Pro aminopeptidase FRA1, P40051: the intermediate cleaving peptidase 55, and P43590: the uncharacterized peptidase YFR006W). The domain architectures were identified using HMMER3 against the Pfam database. The domain structure images were generated using the domain graphic generator on the Pfam website (http://pfam.xfam.org/generate graphic).

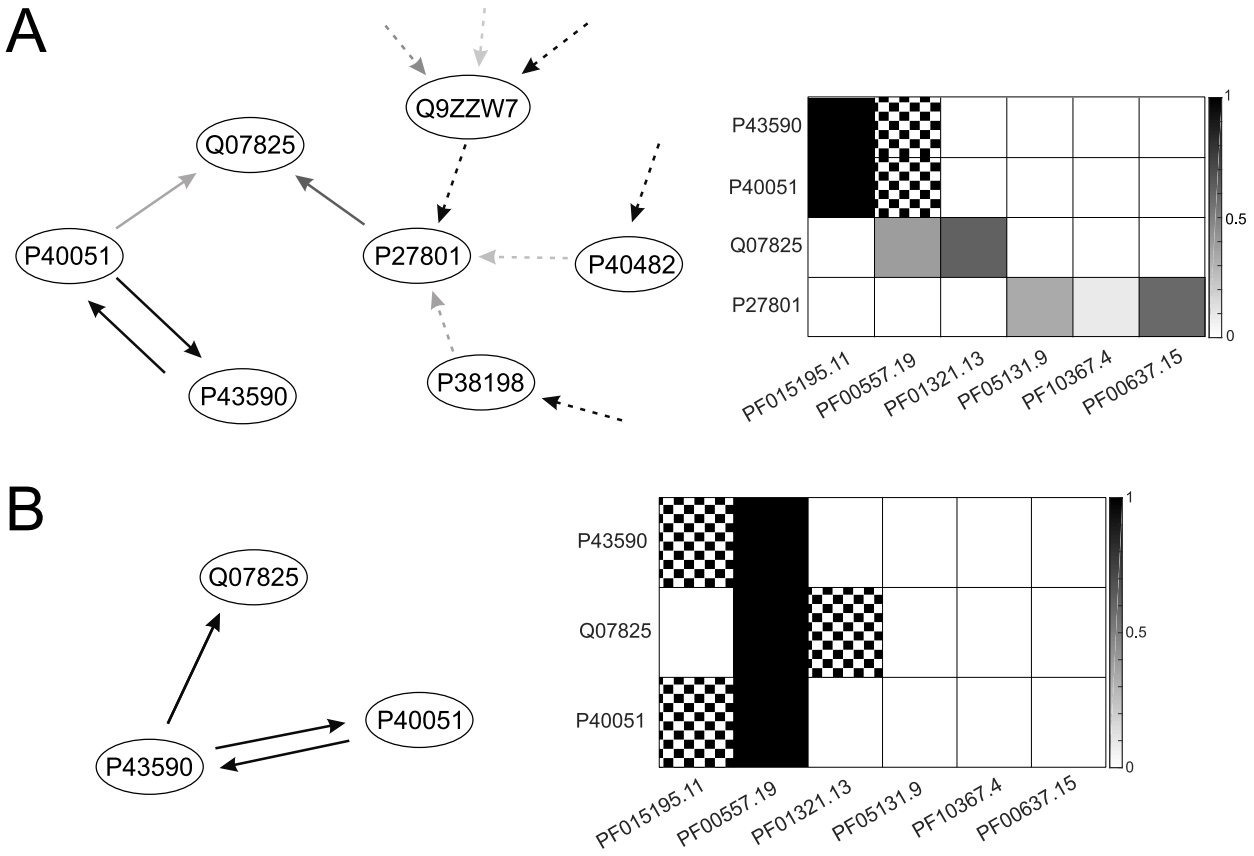

Figure S7: Network refinement and isolation of the peptidase M24B family. A. A partial view of the network and domain profile where the peptidase M24B family proteins are located in the primary network. The three proteins of the peptidase M24 family shown in Figure S6 (Q07825, P40051, and P43590) were found in a large cluster of 3,223 proteins in the initial iteration of MOCASSIN-prot. The direct similarity relationships of the three proteins within this large cluster are shown with solid edges, while dashed edges indicate non-direct relationships. **B. An isolated cluster and domain profile of the peptidase M24 family in the secondary network.** The secondary MOCASSIN-prot clustering placed all three peptidase M24B proteins into a single cluster.