Current Biology, Volume *33*

Supplemental Information

Evolutionary analysis of cellular reduction and anaerobicity in the hyper-prevalent gut microbe *Blastocystis*

Kristína Záhonová, Ross S. Low, Christopher J. Warren, Diego Cantoni, Emily K. Herman, Lyto Yiangou, Cláudia A. Ribeiro, Yasinee Phanprasert, Ian R. Brown, Sonja Rueckert, Nicola L. Baker, Jan Tachezy, Emma L. Betts, Eleni Gentekaki, Mark van der Giezen, C. Graham Clark, Andrew P. Jackson, Joel B. Dacks, and Anastasios D. Tsaousis

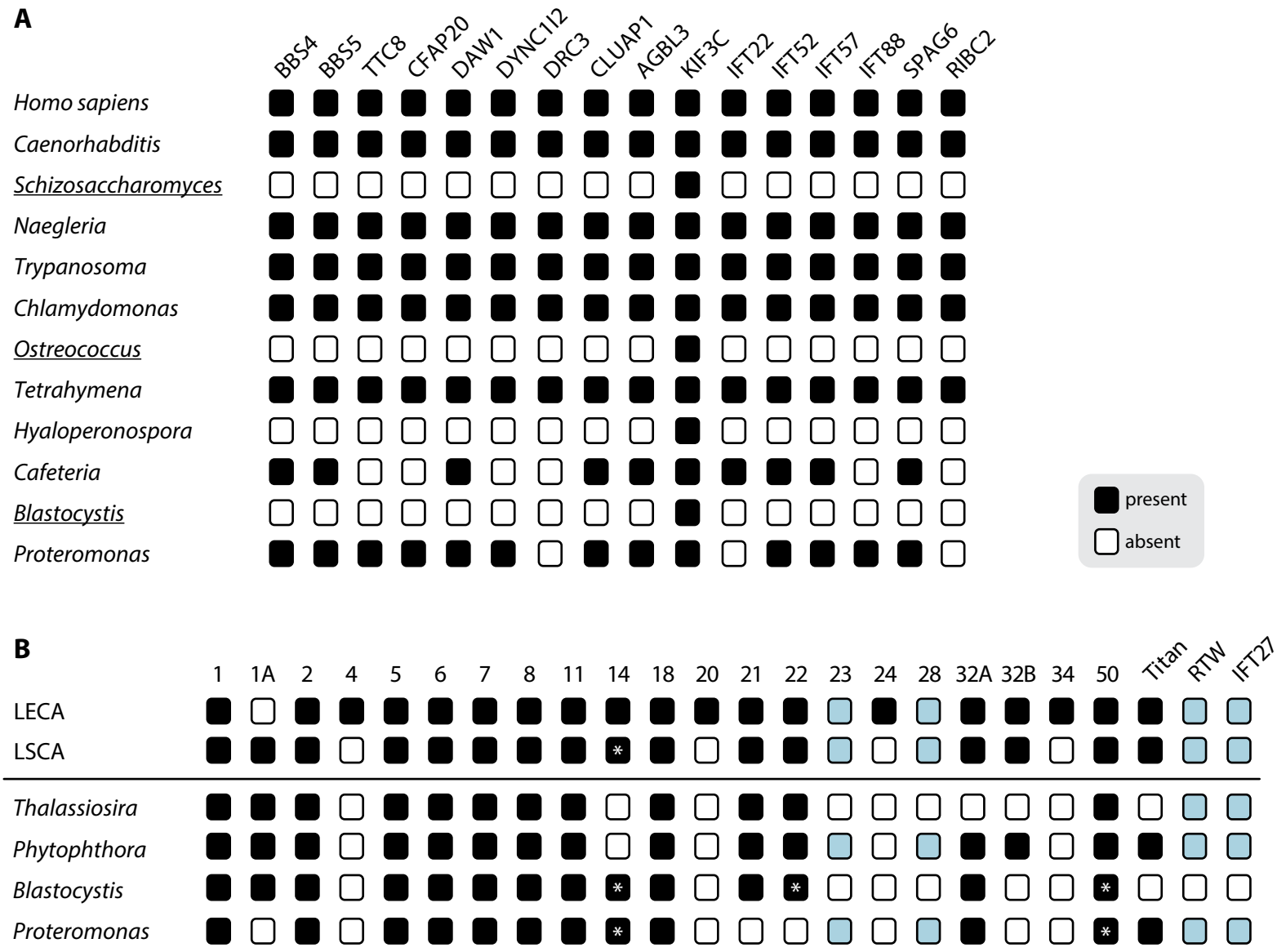

Figure S1. Presence/absence of flagellar (A) and Rab (B) proteins in P. lacerate and selected species, Related to Figure 2B, STAR methods, Data S1. (**A**) Presence or absence of 16 proteins conserved among all flagellated eukaryotes according to [S1]. These were identified in five flagellated organisms of diverse affinity (see Methods) as well as the Blastocystis ST7 genome (ASM15166v1) and the P. lacerate genome and C. burkhardae transcriptome produced in this study. A reciprocal best match by BLASTp between the human protein query and the subject protein in the non-human genome was required to confirm that an orthologue was 'present' in the latter. Non-flagellated organisms are underlined. (**B**) Repertoire of Rab GTPases in P. lacerate, and the deduced Rab set in the last stramenopile common ancestor (LSCA). Rab identity in studied species was confirmed by phylogenetic analysis. Asterisks denote Rabs that could not be unambiguously assigned. Rabs involved in flagellar function are highlighted by pale-blue. The Rab complement of LSCA was compared with that of the last eukaryotic common ancestor (LECA) [S2].

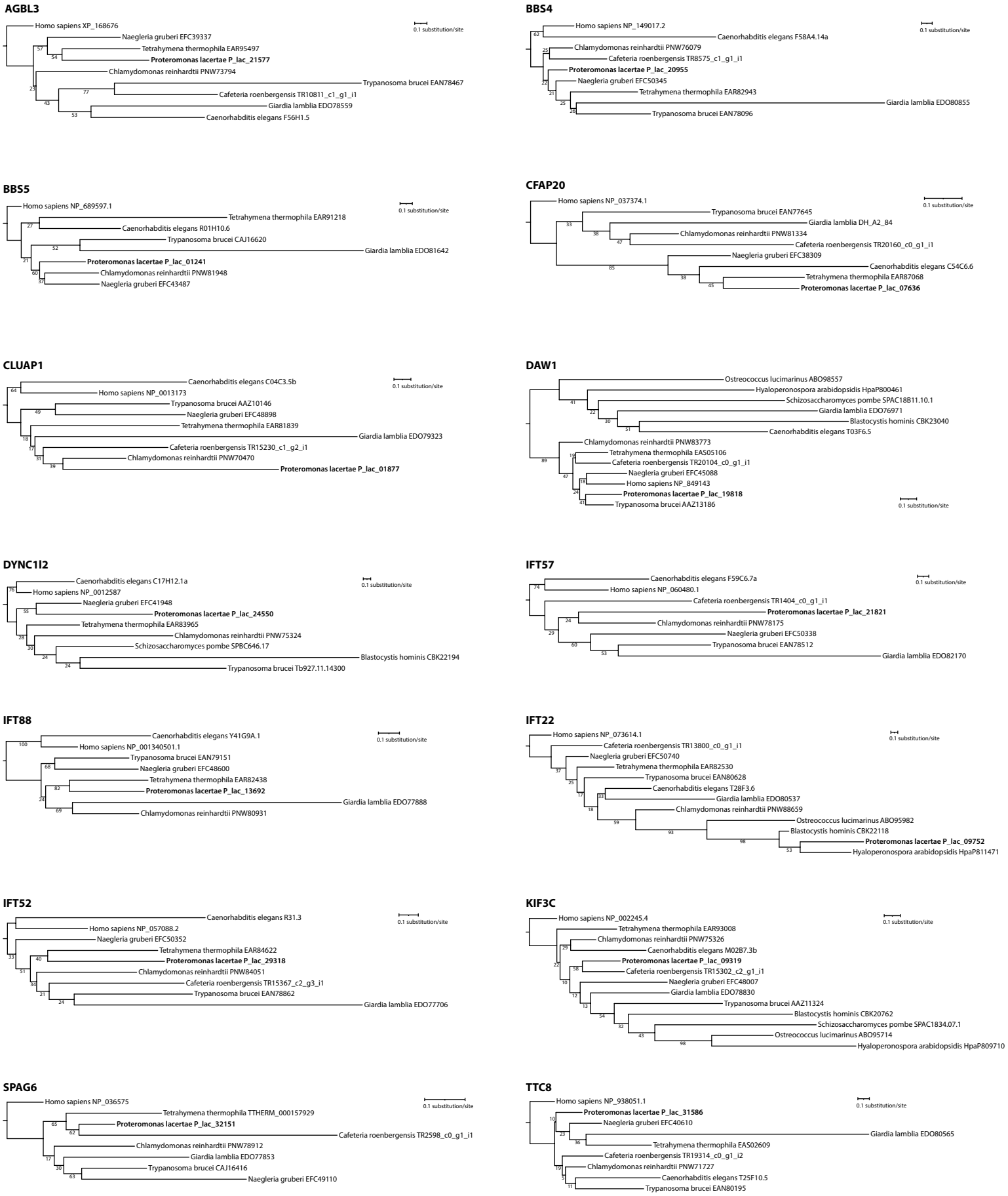

Figure S2. Phylogenetic trees of flagellar proteins, Related to STAR methods. The Maximum-Likelihood phylogenetic trees were estimated from amino acid sequence alignments under a LG+F+Γ model with 100 bootstrap replicates using RAxML [S3], except for thee CFAP21 tree, which was estimated from a nucleotide alignment using a GTR+Γ model. The trees are drawn as phylograms for convenience but are arbitrarily rooted.

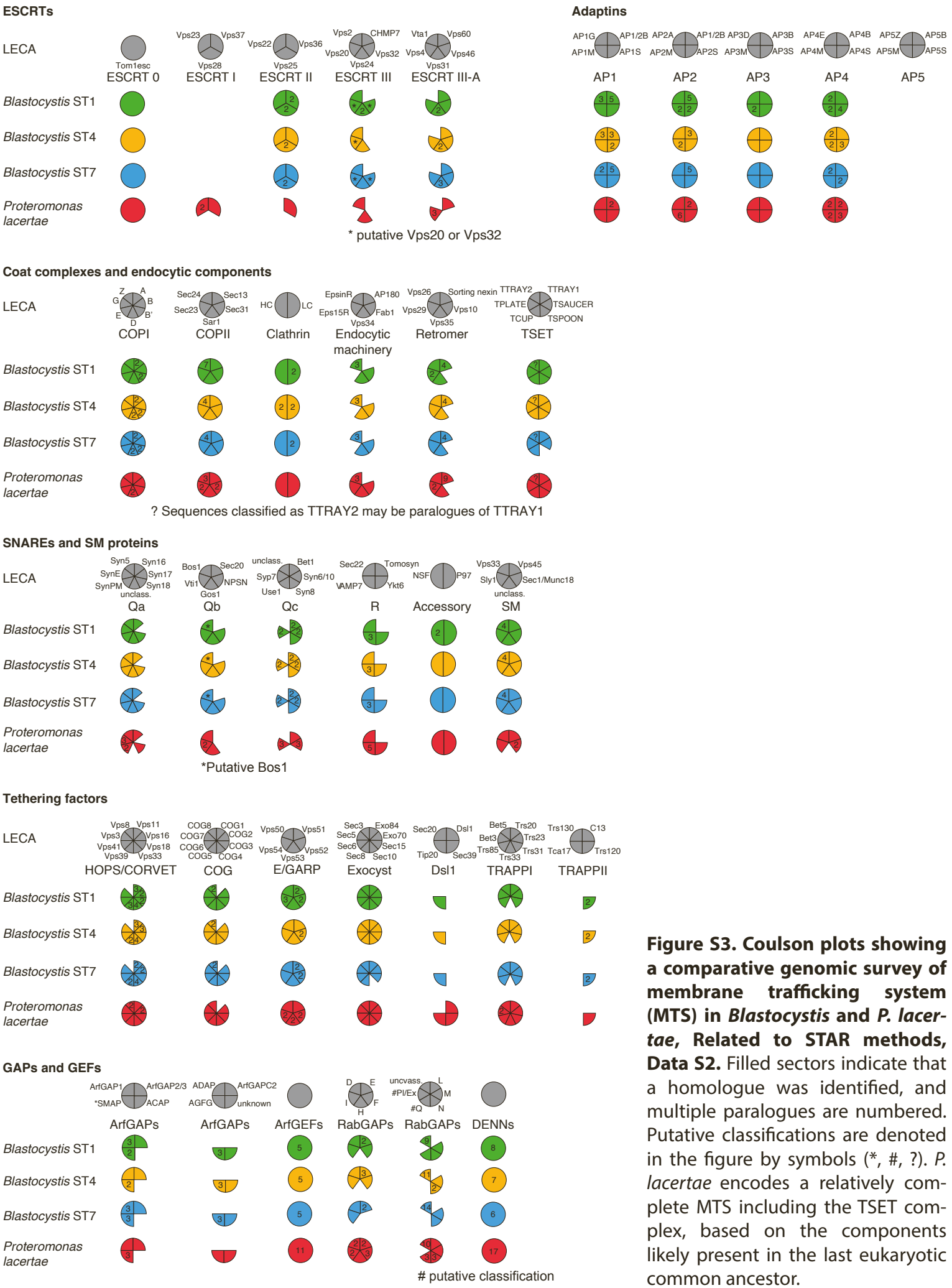

Figure S3. Coulson plots showing a comparative genomic survey of membrane trafficking system (MTS) in *Blastocystis* and *P. lacertae*, Related to STAR methods, Data S2. Filled sectors indicate that a homologue was identified, and multiple paralogues are numbered. Putative classifications are denoted in the figure by symbols (*, #, ?). *P. lacertae* encodes a relatively complete MTS including the TSET complex, based on the components likely present in the last eukaryotic common ancestor.

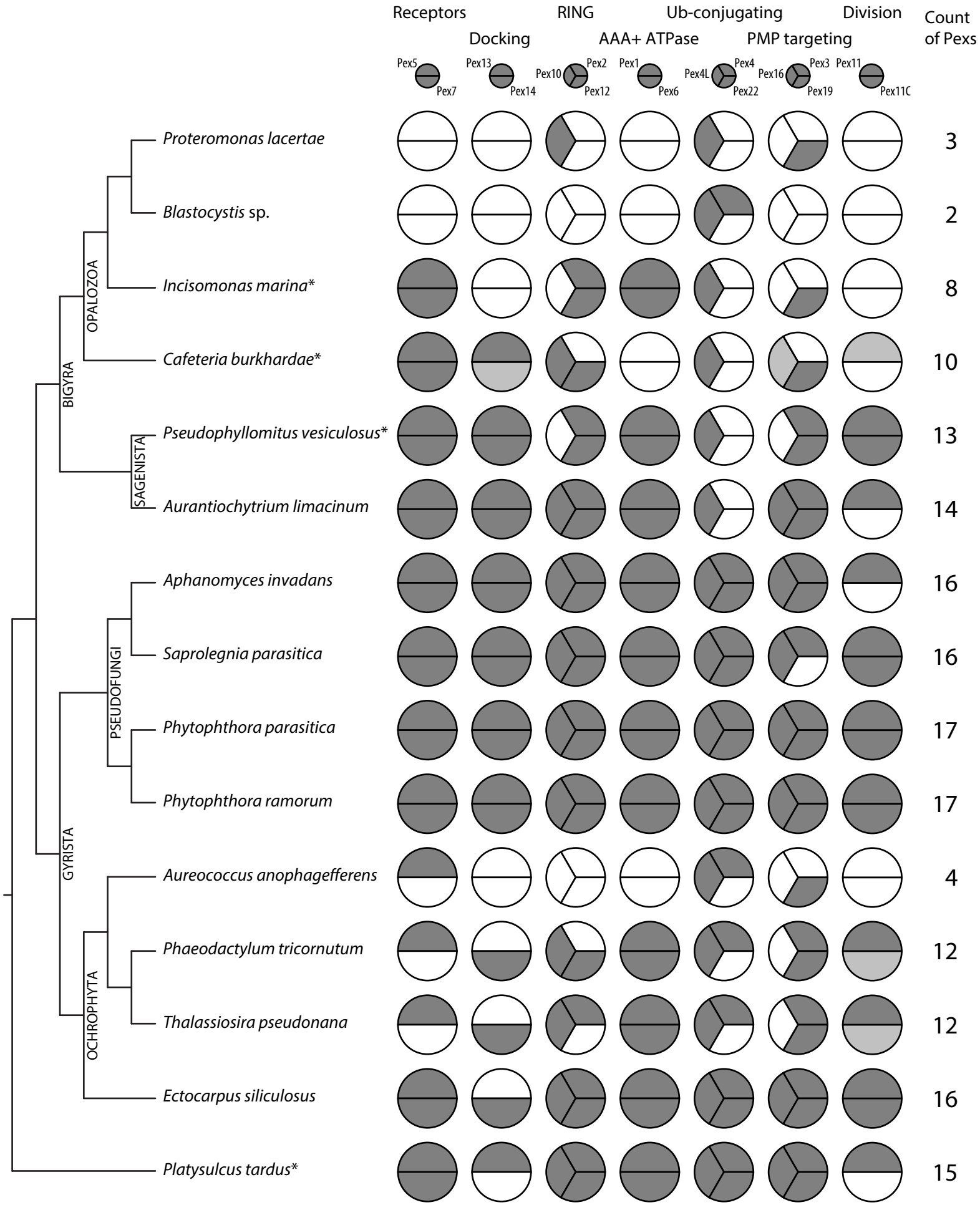

Figure S4. Comparative analysis of peroxisomal complement in stramenopiles, Related to Figure 3A. Pattern is consistent with progressive degeneration of peroxisomal complement in Bigyra, contrasted with conservation in other stramenopile lineages. Presence and absence of proteins is shown by dark grey and white color, respectively. Light grey color shows proteins that were found by manual searches. Asterisks mark species for which only transcriptomes were available.

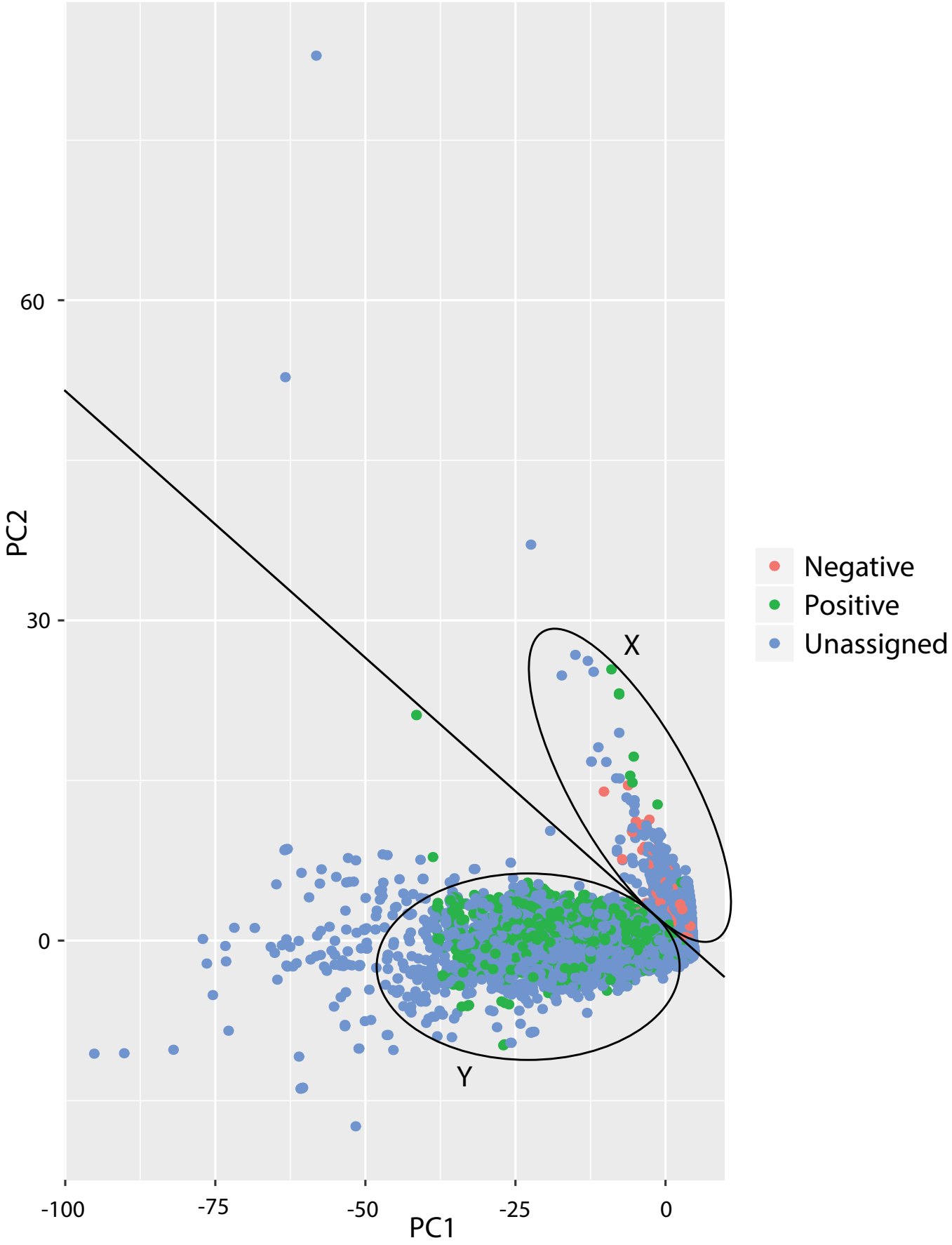

Figure S5. Principal component analysis for kmer frequencies of all transcripts, Related to STAR methods. Positive and negative control groups were clearly distinguished by the analysis, and thus the remaining transcripts were assigned as eukaryotic or prokaryotic.

Supplemental references

S1 Judelson, H.S., Shrivastava, J., and Manson, J. (2012). Decay of genes encoding the oomycete flagellar proteome in the downy mildew *Hyaloperonospora arabidopsidis*. PLoS One *7*, e47624.

S2 Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M.C., and Dacks, J.B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J. Cell Sci. *125*, 2500–2508.

S3 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.