Extensive molecular tinkering in the evolution of the membrane attachment mode of the Rheb GTPase

Kristína Záhonová[1], Romana Petrželková[1], Matus Valach[2], Euki Yazaki[3], Denis V. Tikhonenkov[4], Anzhelika Butenko[1], Jan Janouškovec[5], Štěpánka Hrdá[6], Vladimír Klimeš[1], Gertraud Burger[2], Yuji Inagaki[7], Patrick J. Keeling[8], Vladimír Hampl[6], Pavel Flegontov[1], Vyacheslav Yurchenko[1], Marek Eliáš[1*]

[1]Department of Biology and Ecology & Institute of Environmental Technologies, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

[2]Department of Biochemistry and Robert-Cedergren Centre for Bioinformatics and Genomics, Université de Montréal, Montreal, Canada

[3]Institute for Biological Sciences, University of Tsukuba, Tsukuba, Japan

[4]Laboratory of Microbiology, Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Russia

[5]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom

[6]Department of Parasitology, Faculty of Science, Charles University, Prague, Czech Republic

[7]Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan

[8]Department of Botany, University of British Columbia, Vancouver, Canada

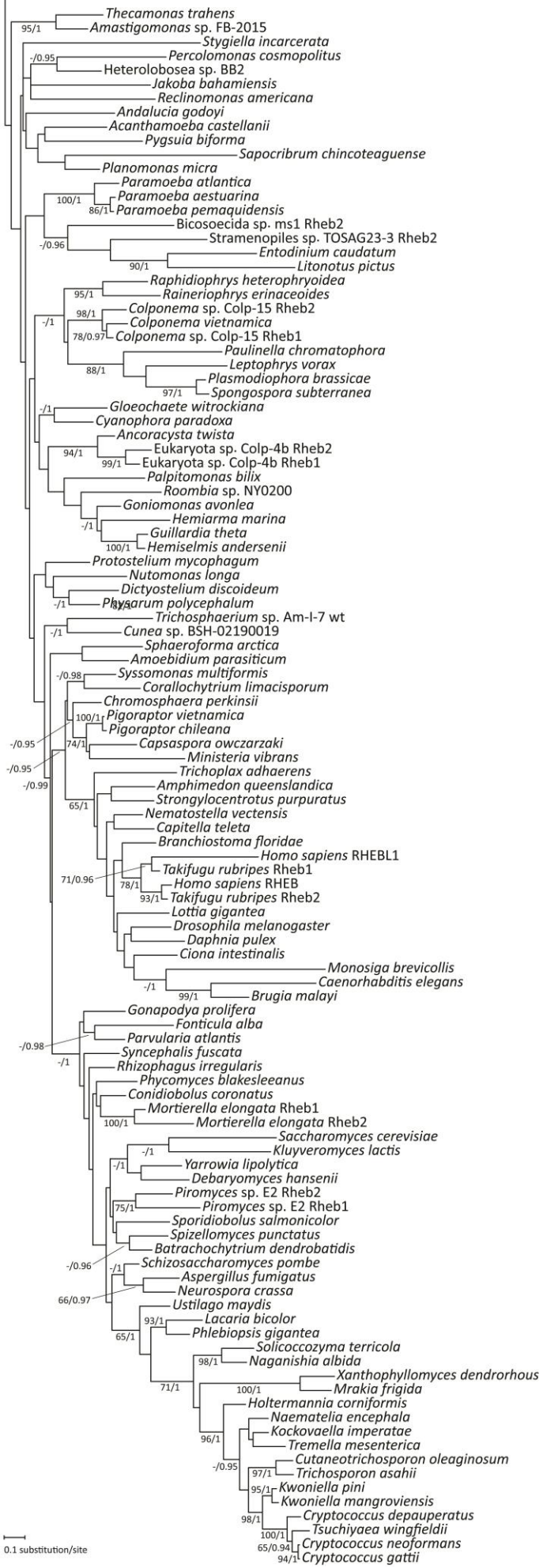
[*]Author for correspondence: Marek Eliáš, Department of Biology and Ecology, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czech Republic, tel.: +420-597092329, e-mail: marek.elias@osu.cz

Supplementary Figures S1 to S7

Fig. S1. Phylogenetic analysis confirming identification of Rheb orthologs. The maximum likelihood tree was inferred using RAxML-HPC2 (LG4X+Γ substitution model) based on a multiple alignment of the conserved GTPase domain (151 aligned amino acid position). The root is arbitrarily placed between the clade comprising bona fide Rhebs and other Ras family proteins. The numbers at branches refer to ML bootstrap values (shown only when ≥65%) and posterior probabilities (≥0.9) calculated using MrBayes (WAG+Γ+I substation model). If the species has more than one Rheb gene, these are distinguished by a specific gene label, otherwise only the species name is indicated. Sequence identifiers of the Rheb sequences included in the analysis is provided in table S1, those of other Ras family proteins are indicated in brackets (the respective databases are listed in table S1). Note the position of the *Naegleria gruberi* protein "61087" (among "other Ras family proteins"), which seems to have been misinterpreted as a Rheb ortholog by van Dam et al. (2011b).

B)

see part A), previous page



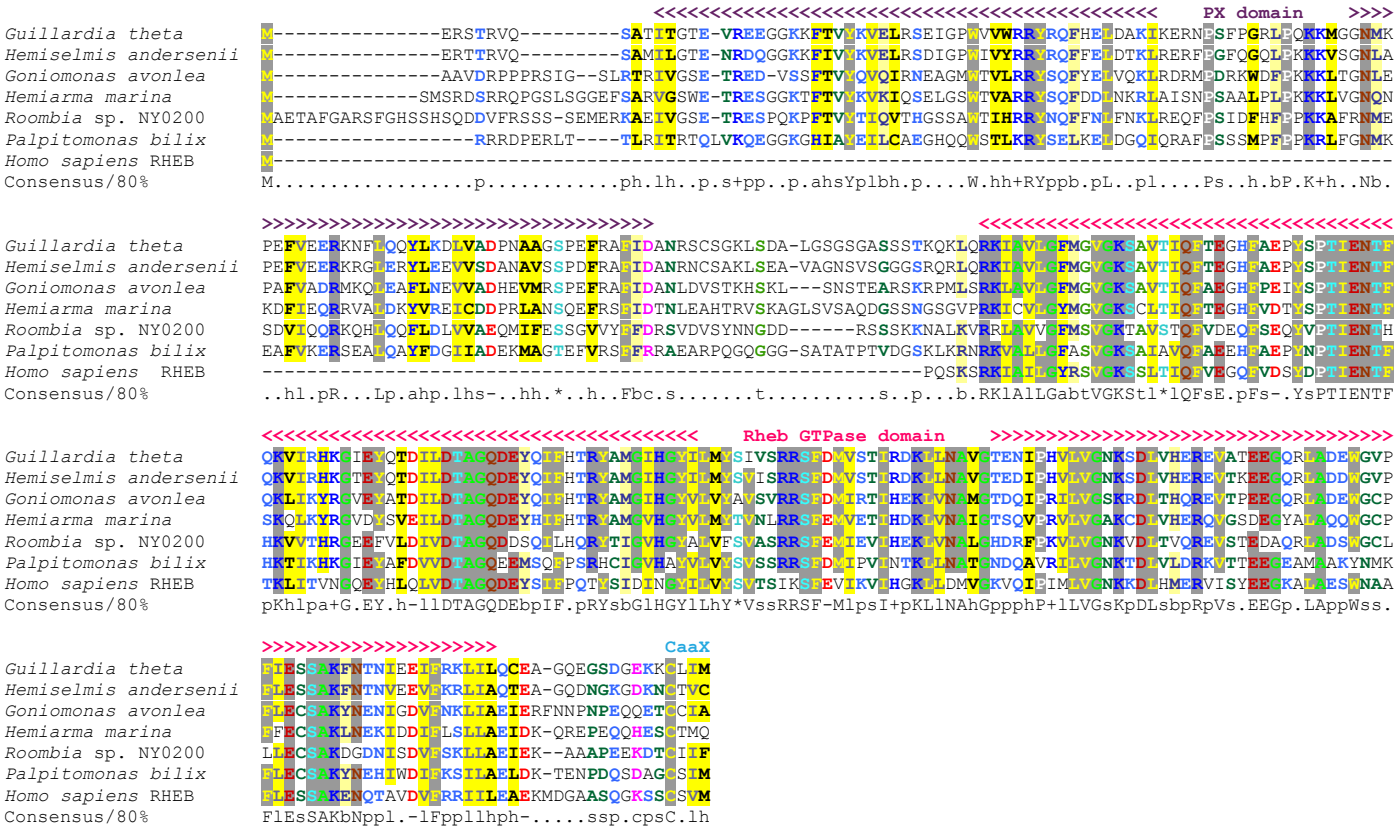
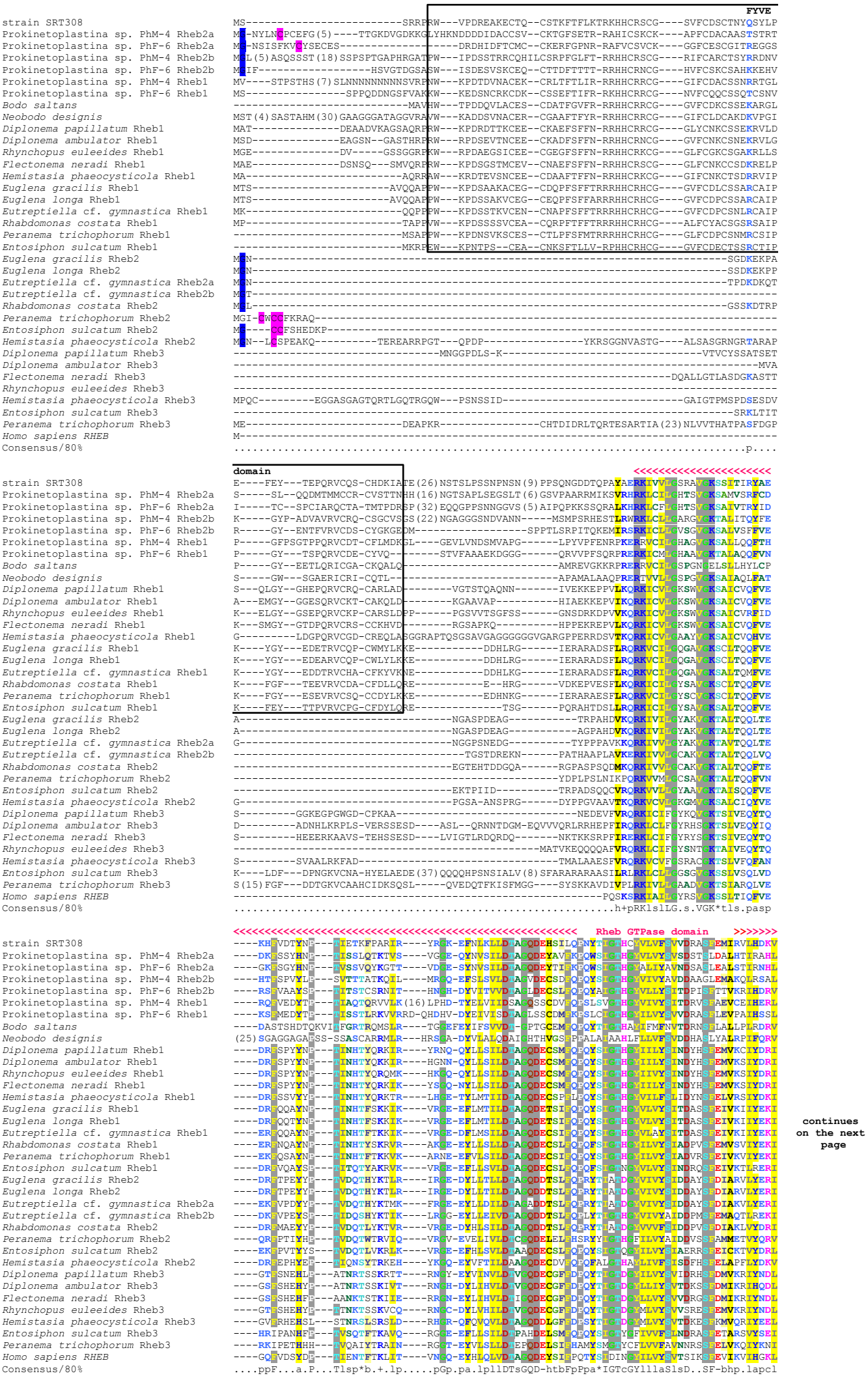


Fig. S2. Annotated multiple alignment of PX-Rheb protein sequences. Sequence accession numbers are provided in table S1. The alignment was prepared using MAFFT version 7 and processed for visualization using Chroma v1.0. The *Homo sapiens* RHEB sequence represents the conventional Rheb form lacking the N-terminal PX domain.


FYVE
domain
Rheb GTPase domain
continues on the next page

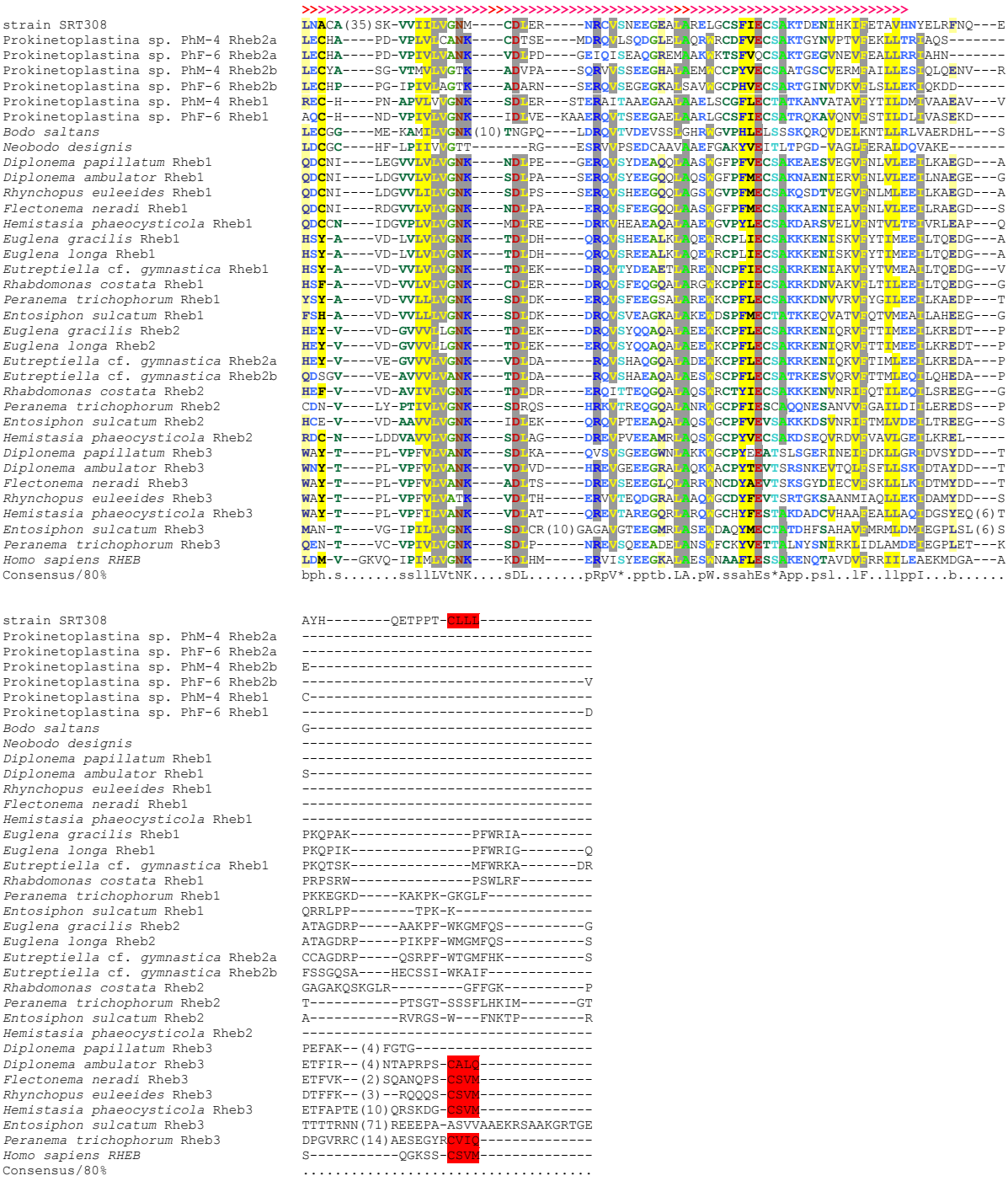


Fig. S3. Annotated multiple alignment of Rheb protein sequences from Euglenozoa. Sequence accession numbers are provided in table S1. The alignment was prepared using MAFFT v7, manually adjusted, and processed for visualization using Chroma v1.0 (numbers in brackets indicate the number of amino acid residues omitted from the display for simplicity). The *Homo sapiens* RHEB sequence represents the conventional Rheb form. Positions predicted to be N-myristoylated (see table S3) are highlighted with a blue background, positions predicted to be S-palmitoylated (see table S4) are highlighted with a pink background, the CaaX box is highlighted with a red background. Note that the sequence of the C-terminus of the Rheb3 protein from *Entosiphon sulcatum* has not been defined (the available transcript sequence is either 3'-truncated or has a retained intron), so it is unclear whether this protein has the CaaX box (as its orthologs) or lacks it. The region corresponding to the FYVE domain is boxed. Note that Pfam search (http://pfam.xfam.org/) failed to recognize the FYVE domain in the Rheb1 proteins from the two Prokinetoplastina species, but the presence of the domain (albeit possibly divergent) in these two proteins is supported by sequence conservation in the respective region and searches against the NCBI Conserved Domains database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

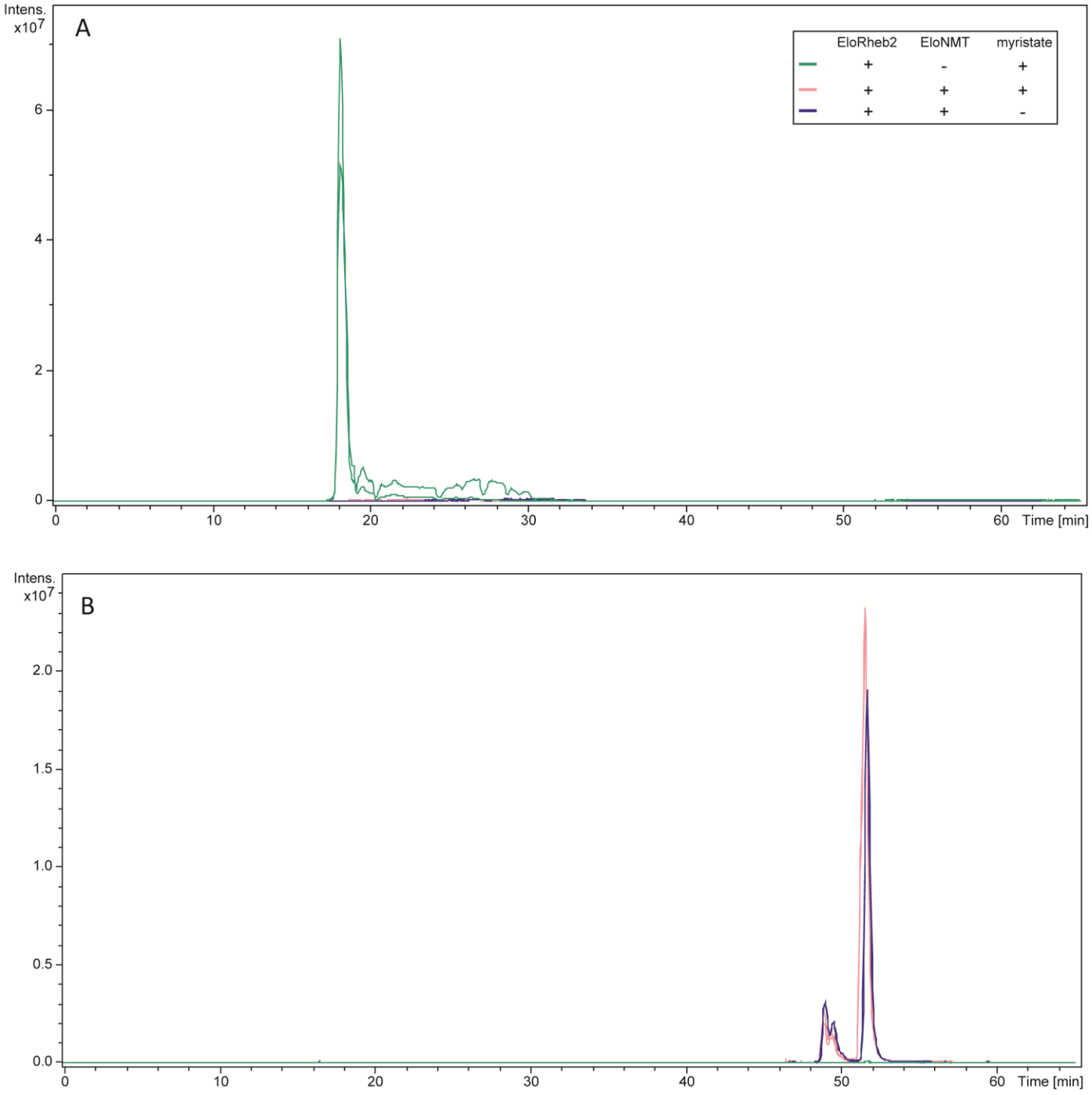


Fig. S4. N-terminal myristoylation of the *Euglena longa* Rheb2 protein detected by mass spectrometry. EloRheb2 protein was affinity purified on S-protein agarose, dissected from a gel after 12% SDS-PAGE electrophoresis, cleaved with trypsin, and the peptides were analyzed by mass spectrometry. The plots shows the signal for two different variants of the N-terminal tryptic peptide (GNSSDKEKPPANGASPDEAGAGPAHDVK) of the Rheb2 protein: A) an unmodified form (2702.25 Da, a lower LC-MS retention time); B) a myristoylated form (2912.45 Da, a higher LC-MS retention time). Both peptide variants were monitored in three different samples (each sample processed twice, the results shown as two lines of the same colour): green, samples from *E. coli* expressing Rheb2, but not N-myristoyl transferase (NMT), supplemented with myristate; pink and blue, samples from *E. coli* expressing Rheb2 and NMT in the presence or absence of external myristate, respectively. Note the lack of a green signal for the myristoylated peptide (consistent with the lack of myristoylation in the absence of NMT). The lack of a pink and a blue signal for the non-myristolated peptide indicates a high efficiency of NMT-dependent N-myristoylation in the heterologous system of *E. coli* cells, even in the absence of the external myristate.

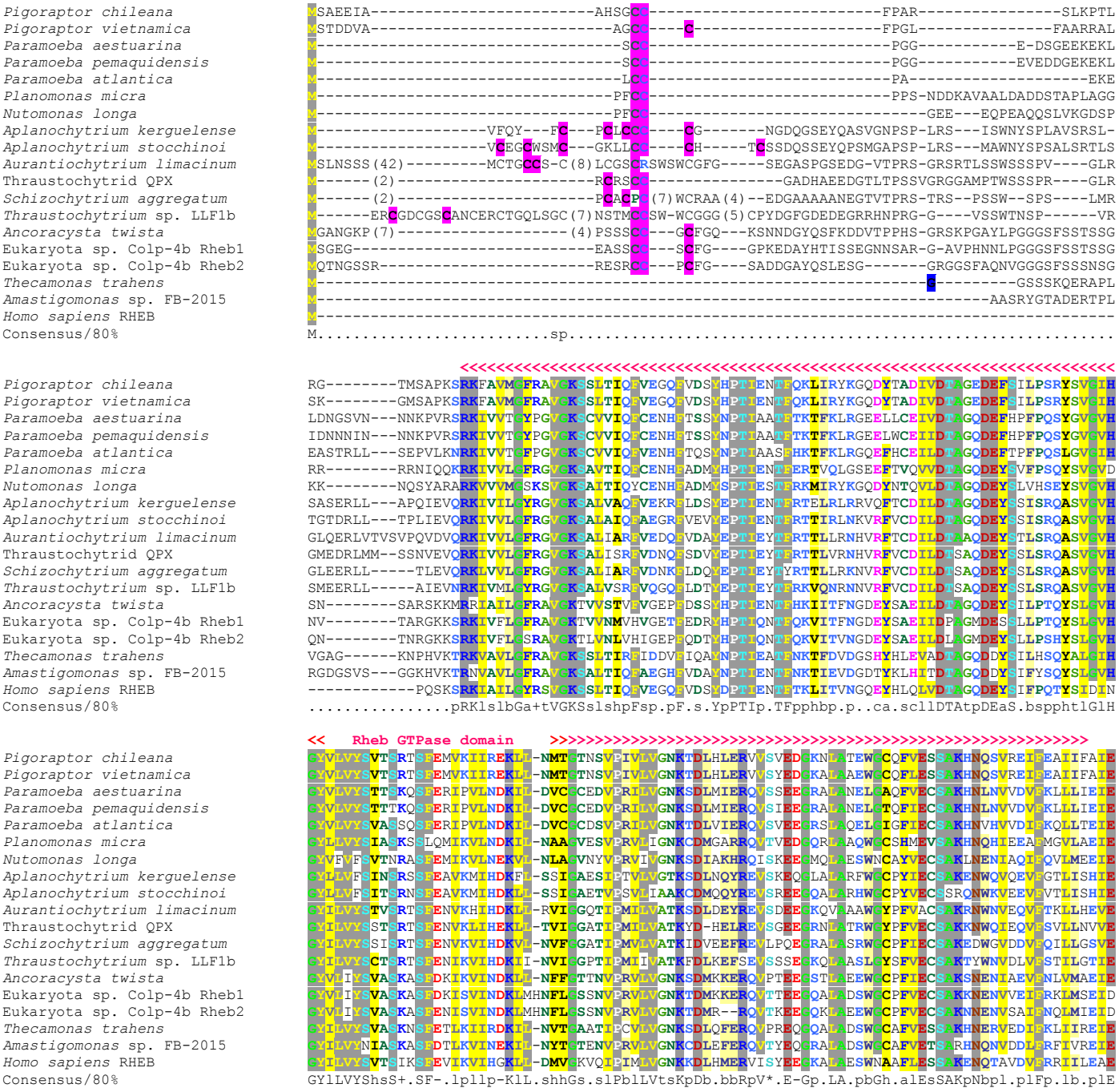


Fig. S5. Annotated multiple alignment of Rheb protein sequences with putative S-palmitoylation (excluding dually acylated Rhebs from Euglenozoa). Sequence accession numbers are provided in table S1. The alignment was prepared using MAFFT v7, manually adjusted, and processed for visualization using Chroma v1.0 (numbers in brackets indicate the number of amino acid residues omitted from the display for simplicity). The *Homo sapiens* RHEB sequence represents the conventional Rheb form lacking an acylated N-terminal extension or palmitoylated cysteines adjacent to the C-terminal CaaX box. Positions predicted to be S-palmitoylated (see table S4) are highlighted with a pink background. Note also the putatively myristoylated glycine residue (highlighted with a blue background; see also table S3) at the N-terminus of the Rheb sequence from *Thecamonas trahens*.

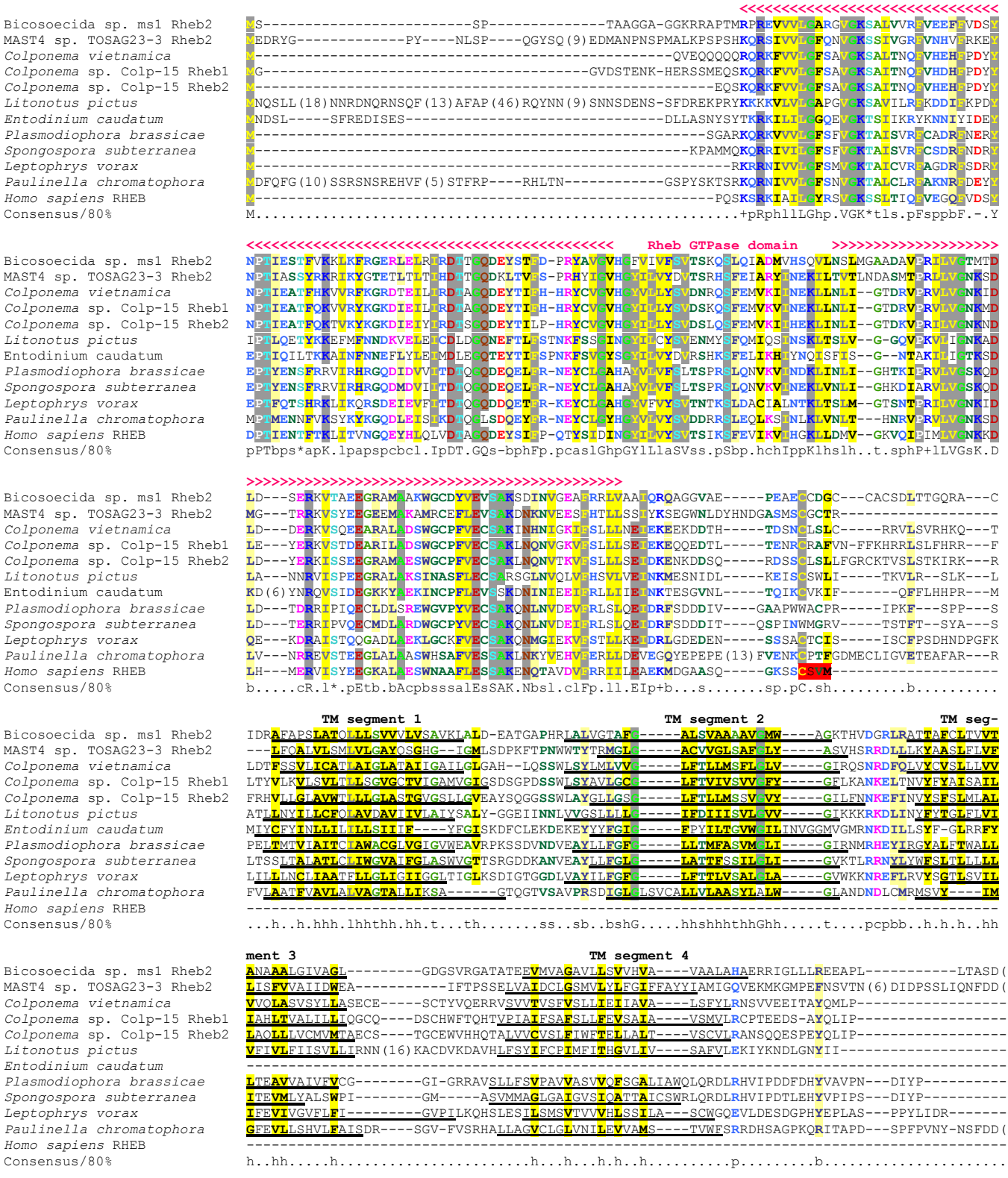


Fig. S6. Annotated multiple alignment of Rheb-TM protein sequences from the SAR clade. Sequence accession numbers are provided in table S1. The *Homo sapiens* RHEB sequence represents the conventional Rheb form with the C-terminal CaaX box (highlighted in red). The alignment was prepared using MAFFT version 7, manually adjusted, and processed for visualization using Chroma v1.0 (numbers in brackets indicate the number of amino acid residues omitted from the display for simplicity). Underlined regions correspond to trans-membrane (TM) segments as predicted using the TMHMM Server 2.0 except for the last TM segment in the Rheb sequence from *Leptophrys vorax*, which was not recognized by TMHMM but was predicted by the alternative tool TMpred. The Rheb sequence from *Entodinium caudatum* lacking the last two transmembrane segments may not be complete. Note that some of the Rheb-TM sequences still exhibit a cysteine residue at the position corresponding to the cysteine residue of the CaaX box of conventional Rheb (suggesting that Rheb-TM emerged via direct accretion of the C-terminal region with the transmembrane segments to the original C-terminus).

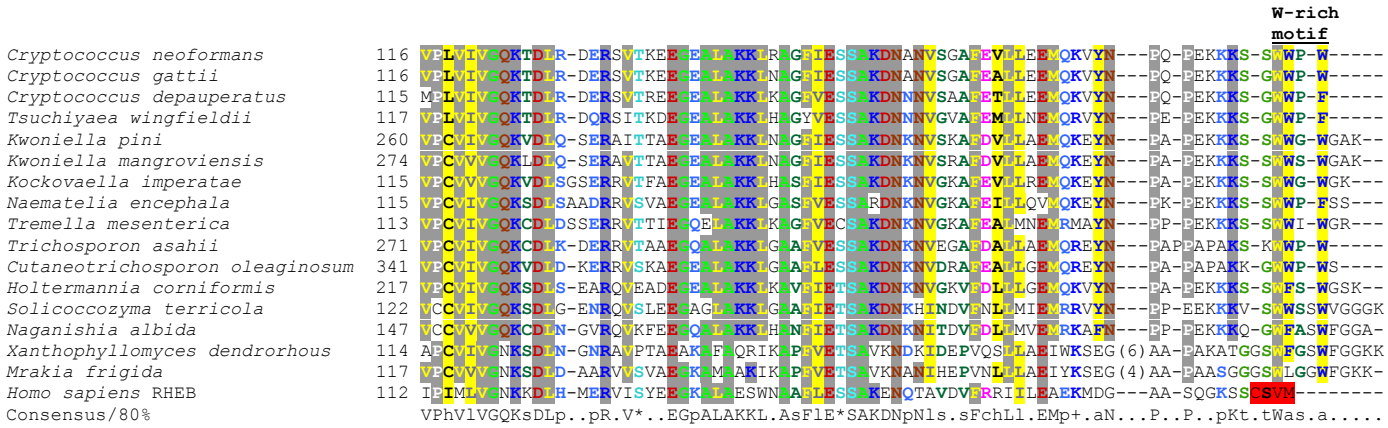


Fig. S7. Annotated multiple alignment of the C-terminal region of Rheb protein sequences from Tremellomycetes. The *Homo sapiens* RHEB sequence represents the conventional Rheb form with the CaaX box at the C-terminus (highlighted in red). The numbers on the left indicate the position of the first amino acid residue shown in the original full-length sequence (for accession numbers of the sequences see table S1). The alignment was prepared using MAFFT version 7, manually adjusted, and processed for visualization using Chroma v1.0 (numbers in brackets indicate the number of amino acid residues omitted from the display for simplicity). Note the replacement of the CaaX box by a tryptophan-rich motif in all tremellomycete sequences.