

Additional file 1 to the paper:

Comprehensive analysis of small GTPases in Microsporidia reveals new extremes in eukaryotic cell biology

Romana Vargová^{1*}, Vladislava Majnušová¹, Marek Eliáš^{1*}

Department of Biology and Ecology, Faculty of Science, University of Ostrava,
Chittussiho 10, 710 00 Ostrava, Czech Republic

*Correspondence:

Romana Vargová

romana.vargova@osu.cz

Marek Eliáš

marek.elias@osu.cz

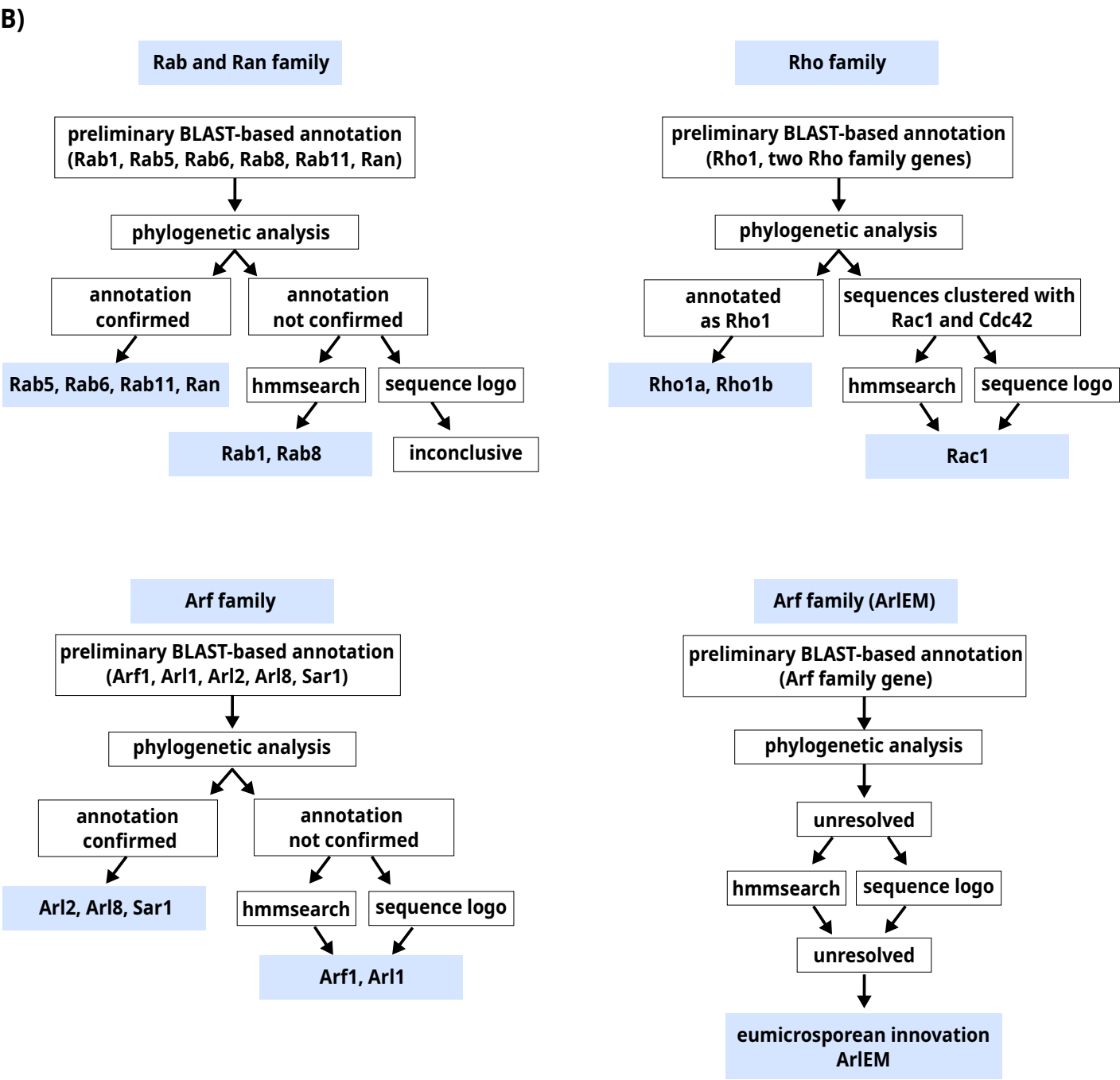
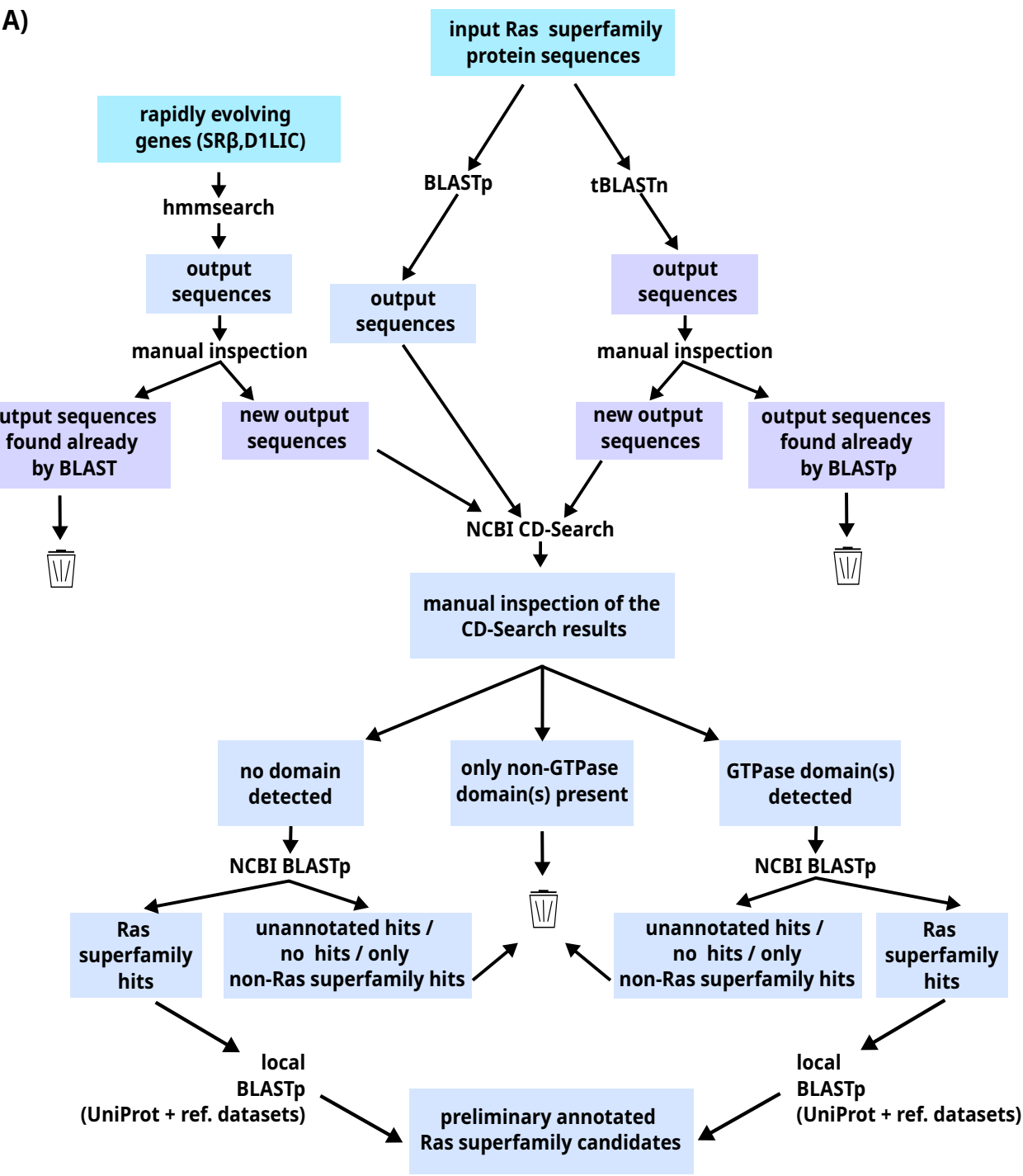


Figure S1. Flowchart of the detection and annotation of Ras superfamily protein sequences in the RM clade. A) Initial search was performed using BLASTp against the set of predicted proteins. A complementary tBLASTn search was performed against genome sequences; only novel sequences not previously detected by BLASTp were retained. For rapidly evolving divergent genes (SR β , D1LIC), hmmsearch was also used in addition to BLAST. This initial candidate set was subsequently evaluated using the NCBI CD-Search tool. Sequences containing only non-GTPase domain(s) were discarded. Sequences with either no detected domain(s) or those containing GTPase domain(s) were subjected to NCBI BLASTp searches. Results were manually inspected, and only sequences with top hits corresponding to the Ras superfamily were retained. These sequences were then preliminarily annotated based on local BLASTp searches against our custom database composed of reference eukaryotic proteomes from the UniProt database combined with reference datasets of Rab, Arf, and Rho families [45, 46, 60]. **B)** Preliminary annotations were validated through phylogenetic analysis. In cases where preliminary annotation was not confirmed or remained unresolved, hmmsearch and sequence logo analyses were employed. The eumicrosporean-specific ArLEM could not be assigned to a particular ancestral pan-eukaryotic paralog. This group likely represents a highly divergent eumicrosporean-specific innovation of unclear origin (see the main text for details).

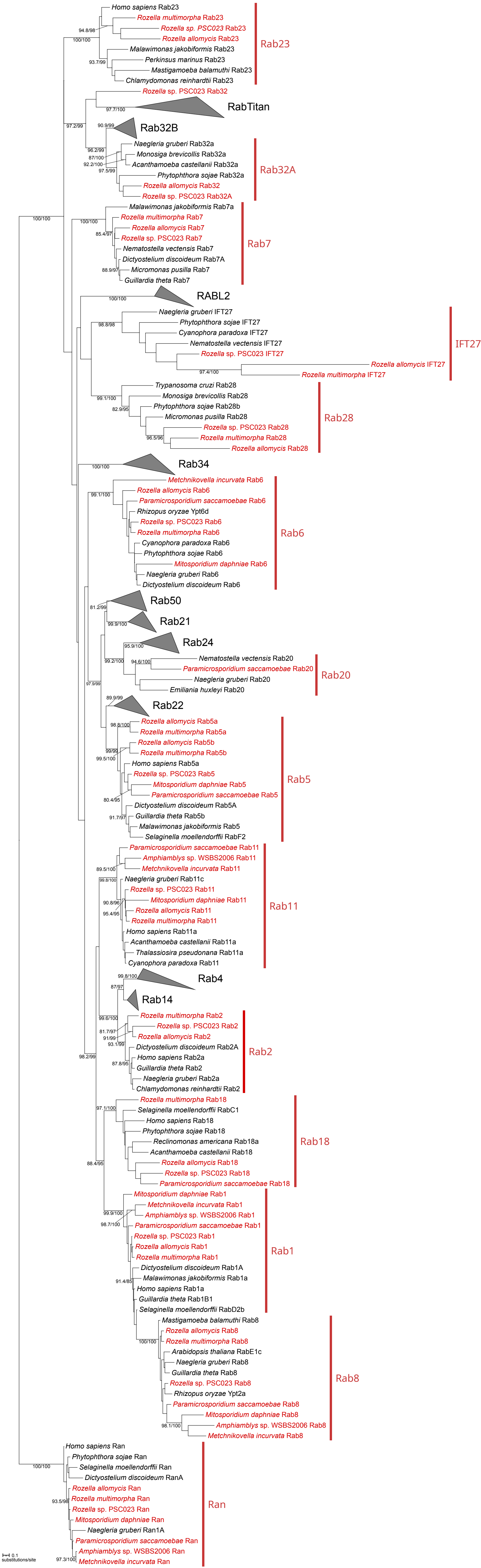


Figure S2. Phylogenetic analysis of Rab and Ran family proteins from the RM clade organisms and a selection of other eukaryotes. The tree shown is a result of a ML analysis of a reference Rab and Ran family dataset from Elias et al. [45] and Rab and Ran family sequences from three rozellid species and microsporidians *P. saccamoebae*, *M. daphniae*, *M. incurvata*, and *Amphiambllys* sp. WSBS2006 (highlighted in red), except for one incomplete and divergent Rab-like sequence from *R. multimorpha* (see Methods for details). The tree was inferred using IQ-tree with LG+I+G4 model with the ultrafast bootstrap (UFB) algorithm and the SH-aLRT test (both 10 000 replicates, only supports ≥ 80 for SH-aLRT and ≥ 95 for UFB are displayed). Branches representing ancestral paralogs absent from rozellids and Microsporidia are collapsed. Note that apart from Rab and Ran family, the reference dataset contains also Rab-like sequences RABL2 and IFT27. The paraphyly of the Rab1 group (with Rab8 nested in it) is an apparently result of the phylogenetic inference observed also in previous analyses [45]. The sequence labelled “*Rozella* sp. PSC023 Rab32”, which branches without support sister to the “RabTitan” clade, looks like a standard Rab family protein and does not resemble RabTitans in its structure (the latter are large proteins with long regions C-terminal to the GTPase domain lacking a C-terminal prenylation motif), so we consider it a divergent lineage-specific paralog derived from Rab32A or Rab32B.

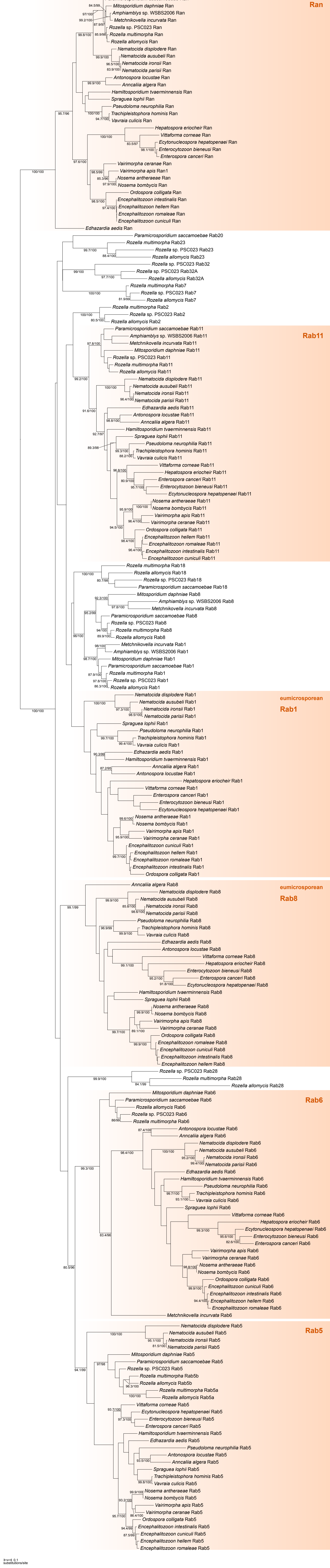


Figure S3. Phylogenetic analysis of Rab and Ran family proteins from the RM clade organisms. This is an expanded version of the tree shown in Fig. 3. The tree shows a result of a ML analysis of all Rab and Ran family sequences from rozzellids, SB-Microsporidia, metchnikovellids, and Eumicrosporea. For details regarding the phylogenetic analysis and display conventions see the legend of Fig. 3. Note that the assignment of two of the eumicrosporean Rab clades, the one annotated as Rab1 and Rab8 in the tree, is not directly supported by the phylogenetic analysis and is based on other analyses described in the main text of the paper.



Figure S4. Sequence logos of Rab family proteins. Sequence logos of putative eumicrosporean Rab1 and Rab8 (shaded in violet), together with Rab1 and Rab8 from rozellids, SB-Microsporida, and metchnikovellids, and Rab1 and Rab8 from other eukaryotes are depicted. For comparison, sequence logos of other ancestral Rab groups are included. For details on the sequences included in the analysis, see Methods and Table S12. Light pink rectangles mark positions where the majority of eumicrosporean Rab1 and Rab8 sequences share the same amino acid residue with Rab1 and Rab8, respectively, from their close relatives.

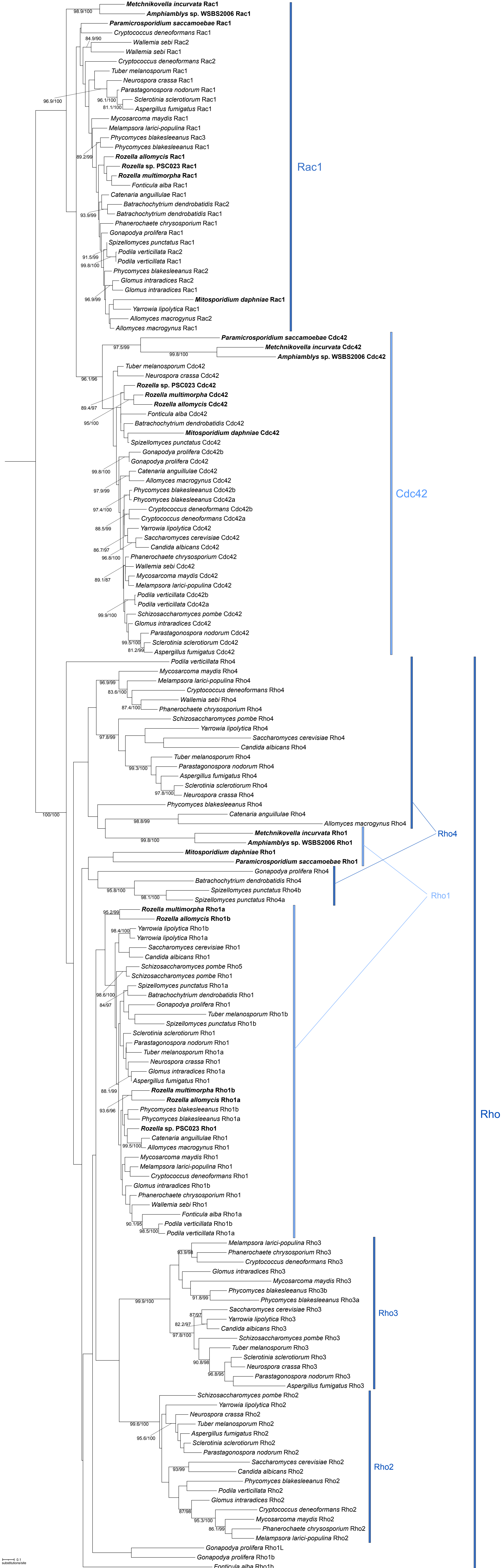


Figure S5. Phylogenetic analysis of the Rho family from rozellids, SB-Microsporidia, and metchnikovellids. The tree shown is a result of a ML analysis of a reference Holomycota Rho family dataset combining sequences from Eliáš and Klimeš [60], additional reference taxa outside the RM clade (*Fonticula alba*, *Gonapodya prolifera*, and *Catenaria anguillulae*), and rozellids, SB-Microsporidia, and metchnikovellids. The tree was inferred using IQ-tree with LG+I+G4 model with the ultrafast bootstrap (UFB) algorithm and the SH-aLRT test (both 10 000 replicates). Support values ≥ 80 for SH-aLRT and ≥ 95 for UFBoot are shown. The list of protein sequences from taxa other than the RM clade that were included in the analysis is available in Table S9A.

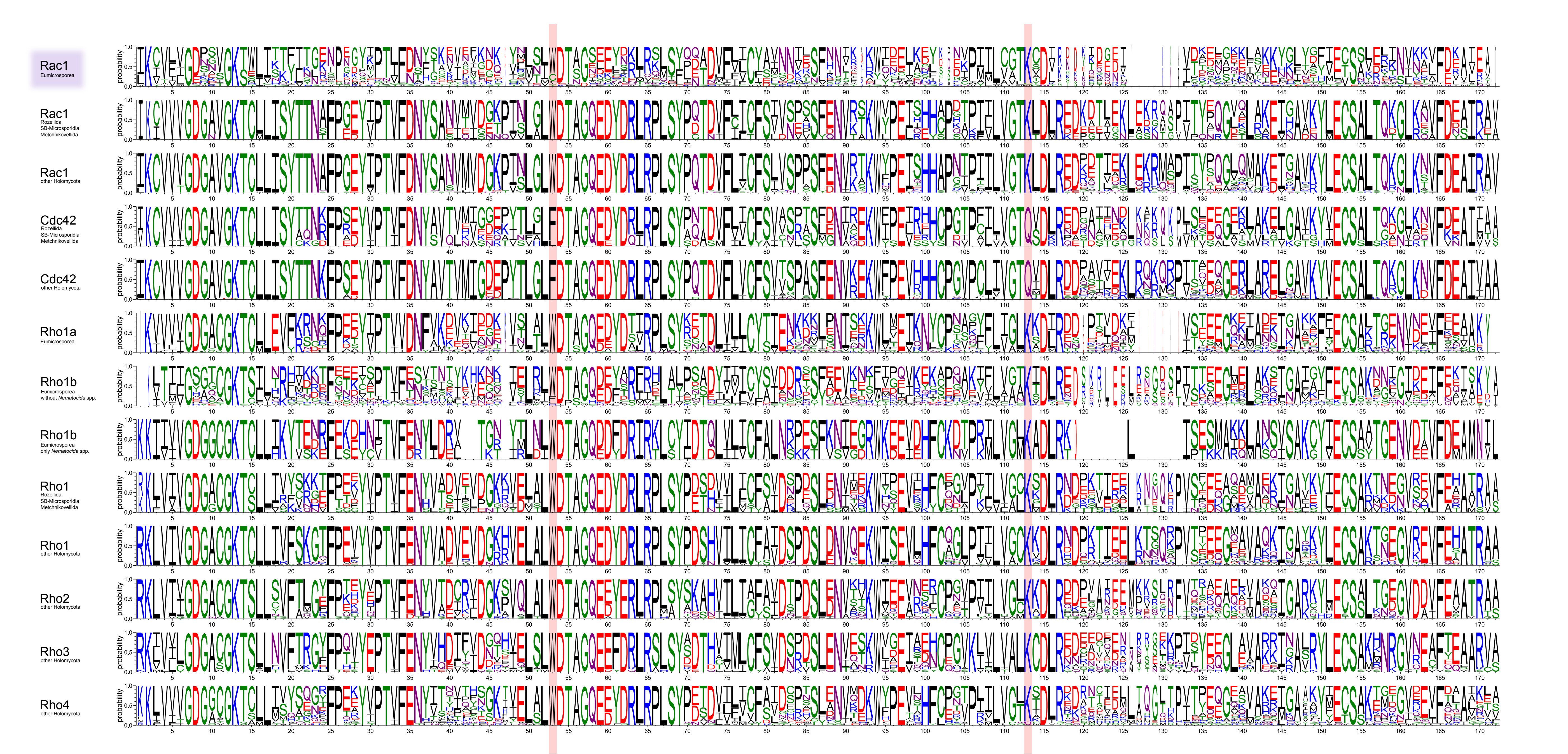


Figure S6. Sequence logos of Rho family proteins. Sequence logo of putative eumicrosporean Rac1 proteins (shaded in violet), Rac1 and Cdc42 from rozellids, SB-Microsporidia and metchnikovellids, and Rac1 and Cdc42 from other Holomycota species. Pink rectangles mark the positions where a conserved residue (tryptophan and lysine, respectively) is shared with Rac1 proteins, whereas Cdc42 proteins exhibit a different conserved residue at these positions (phenylalanine and glutamine, respectively). Note a single exception from this pattern, namely one of the two Rac1/Cdc42 paralogs from *E. aedis* that shares a phenylalanine residue with Cdc42. To provide a wider perspective on conserved amino acid sequences, sequence logos from Rho proteins are included. For details on the sequences, see Methods and Table S12.

A) Rho1 (Rho family)

<i>Rozella allomycis</i> Rho1a	KL---KKKKK-----CSLL----
<i>Rozella allomycis</i> Rho1b	RK---KNKLK-----CLLM----
<i>Rozella multimorpha</i> Rho1a	KK---KKGAK-----CMLL----
<i>Rozella multimorpha</i> Rho1b	KI---KKKKA-----CLVL----
<i>Rozella</i> sp. PSC023 Rho1	KT---RKKKA-----CVVL----
<i>Paramicrosporidium saccamoebae</i> Rho1	QR---RKKSAK-S--IN----- (6) RCLIQ----
<i>Mitosporidium daphniae</i> Rho1	EQ---SYKPEE-RK-SSLVPTLLSPA (9) HECMM----
<i>Amphiamblys</i> sp. WSBS2006 Rho1	TR---SLAGRFLSMFIG---FVTFRS-----CRTGDCCSY
<i>Metchnikovella incurvata</i> Rho1	KS---FASRIA-SGIAN---VFTEKA-----CR-----
<i>Nematocida displodere</i> Rho1a	GK---SATPKS-R--LR---FLW-----CCC-----
<i>Nematocida parisii</i> Rho1a	QK---ETSSQK-G--FK---FFY-----CC-----
<i>Nematocida ironsii</i> Rho1a	QK---ETSSQK-G--FK---FFY-----CC-----
<i>Nematocida ausubeli</i> Rho1a	QK---ETSQK-G--FK---FFY-----CC-----
<i>Spraguea lophii</i> Rho1a	KK---EETQQTCKS--LS---TLF-----CCGG-----
<i>Pseudoloma neurophilum</i> Rho1a	RK---QKKEKI-SL-FQ---KIF-----CCKFCC----
<i>Vavraia culicis</i> Rho1a	KK---RMKEQP-NW-FK---KIF-----CCGLC----
<i>Trachipleistophora hominis</i> Rho1a	KK---RIKEQP-SW-FK---RIF-----CCGFC----
<i>Hamiltosporidium tvaerminnensis</i> Rho1a	KS---TAKKEK-K--WY---FCF-----CC-----
<i>Edhazardia aedis</i> Rho1a1	RS---PKRNSN-KR-WF---FCY-----CC-----
<i>Edhazardia aedis</i> Rho1a2	IF---DDNYSG-GS-GN--RFCW-----CC-----
<i>Anncaliaia algerae</i> Rho1a	RS---RTPSES-----RFP-----CCCYS (4)
<i>Antonosporea locustae</i> Rho1a	KK---KQTSRR---IS---RCL-----CC-----
<i>Vittaforma corneae</i> Rho1a	SE---TLEPEE-TC-CG (4) WCR-----CCGK-----
<i>Hepatospora eriocheir</i> Rho1a	KM---KESTKG-GS-KG---ICG-----CCMP-----
<i>Ecytonucleospora hepatopenaei</i> Rho1a	NI---SSVDEK-GC-CG-----CLFGTK----
<i>Enterosporea canceri</i> Rho1a	DT---VPESKS-A--CA---CCS-----CLLGG----
<i>Enterocytozoon bieneusi</i> Rho1a	HT---AQDTTQ-N--GG---CCA-----CLLGG----
<i>Vairimorpha ceranae</i> Rho1a	KE---SQSIDK-G--CG---LFY-----CC-----
<i>Vairimorpha apis</i> Rho1a	KV---AENMDQ-GCF--LW-----CCK-----
<i>Nosema bombycis</i> Rho1a	KD---GVNENA-G--CG---FCF-----CC-----
<i>Nosema antheraeae</i> Rho1a	KD---GISENA-G--CG---FCF-----CC-----
<i>Ordospora colligata</i> Rho1a	KK---IASDAG-S--CG---FLPCIG-----CCYR-----
<i>Encephalitozoon cuniculi</i> Rho1a	KQ---AASDAR-S--CG---FFSCIR-----CCRYR----
<i>Encephalitozoon intestinalis</i> Rho1a	KQ---ATSDAE-S--CS---FLSCMK-----CCR-----
<i>Encephalitozoon romaleae</i> Rho1a	KG---AAESG-P--CS---FFSCMS-----CCHYR----
<i>Encephalitozoon hellem</i> Rho1a	KQ---AVSESG-P--CS---LFSCMS-----CCRYG----
<i>Nematocida displodere</i> Rho1b	KR---KRKSKK----- (2) ---CKTL----
<i>Nematocida parisii</i> Rho1b	SK---KKKDRG----- (2) ---CKTL----
<i>Nematocida ironsii</i> Rho1b	SK---KKKDRG----- (2) ---CKTL----
<i>Nematocida ausubeli</i> Rho1b	TK---KKKDRG----- (2) ---CKTL----
<i>Spraguea lophii</i> Rho1b	KG---AYKKKS-----KFL-----CC-----
<i>Pseudoloma neurophilum</i> Rho1b	LF (7) KKKTPW---FR---KLF-----CCA-----
<i>Vavraia culicis</i> Rho1b	KV (7) KKKKGF---FS---SLFQR-----CC-----
<i>Trachipleistophora hominis</i> Rho1b	KV (7) KKKKGF---FH---ALFQR-----CC-----
<i>Hamiltosporidium tvaerminnensis</i> Rho1b	HE---TKKGF-SFKKFFE---KLF-----CGICK----
<i>Anncaliaia algerae</i> Rho1b	DI (5) KKKRSF---ME---KFFGY-----CGCGK----
<i>Antonosporea locustae</i> Rho1b	MV---SRKKDG---FA--KKVLRL--TCCECKSD (7)
<i>Vittaforma corneae</i> Rho1b	SS---PKQRGL---FR--RIFF-----CCL-----
<i>Ecytonucleospora hepatopenaei</i> Rho1b	EN---PPKHKN---FF-----CC-----
<i>Enterosporea canceri</i> Rho1b	ES---VDTPW---WN-----FCCCTL----
<i>Enterocytozoon bieneusi</i> Rho1b	EK---VNKKPW---YQ-----CLCC-----
<i>Vairimorpha ceranae</i> Rho1b	SL---KKDKSW---FK---RFLSS-----FACCLIE----
<i>Vairimorpha apis</i> Rho1b	TV---HGNSHW---YD---CLL-----CCRSND-L
<i>Nosema bombycis</i> Rho1b	HF---KNTKGW---FS---QFLDV-----FNCCSKN----
<i>Nosema antheraeae</i> Rho1b	HF---KNTKGW---FS---QFLDV-----FNCCSKN----
<i>Ordospora colligata</i> Rho1b	HL---KESKSF---FG---GVWSSLC-DFICCREIN--
<i>Encephalitozoon cuniculi</i> Rho1b	QQ---KESMSF---IG---AMCASLW-NCLSCRQIN--
<i>Encephalitozoon intestinalis</i> Rho1b	YL---RENKSF---MG---RILTSLW-NFLCCRKIN--
<i>Encephalitozoon romaleae</i> Rho1b	HL---KESRSF---IG---GILASLW-DFLCCRKIN--
<i>Encephalitozoon hellem</i> Rho1b	HQ---KESRSF---IG---RILNSLW-DFLCCQIN--

C - predicted as palmitoylated

C - predicted as substrate for FTase by PrePS

C - predicted as substrate for FTase and GGTaseI by PrePS

B) Cdc42 and Rac1 (Rho family)

<i>Rozella allomycis</i> Cdc42	PKVGKSSK-----KCTLL-----
<i>Rozella multimorpha</i> Cdc42	PKT-KSSR-----KCLLL-----
<i>Rozella</i> sp. PSC023 Cdc42	PT--KKGK-----KCVLL-----
<i>Paramicrosporidium saccamoebae</i> Cdc42	KKAKASGNSGKSKKCLIV-----
<i>Mitosporidium daphniae</i> Cdc42	PKIKKR-----LCSLM-----
<i>Amphiamblys</i> sp. WSBS2006 Cdc42	TQDSQRPPS (23) TAPII (70) FFSCCR-----
<i>Metchnikovella incurvata</i> Cdc42	TSESKSNTK (20) GCOSL (76) FG-CCS-----
<i>Rozella allomycis</i> Rac1	QPKESRKK-----GCIML-----
<i>Rozella multimorpha</i> Rac1	PKKPAKNK-----GCMLL-----
<i>Rozella</i> sp. PSC023 Rac1	PRRETRKK-----NCTLL-----
<i>Paramicrosporidium saccamoebae</i> Rac1	QPVGKKGKK (9) -KCSIQ-----
<i>Mitosporidium daphniae</i> Rac1	TGKKKKKGK-----KCTIV-----
<i>Amphiamblys</i> sp. WSBS2006 Rac1	SEEPEKRG-----FCRRL-----FG-CCS-----
<i>Metchnikovella incurvata</i> Rac1	QSATESSSGGIVAWFKNL-----FA-CCG-----
<i>Spraguea lophii</i> Rac1	EEVVEKA-----RCCGW-----F--CCK-----
<i>Pseudoloma neurophilia</i> Rac1	NRTFKQ-----ML----YEYCCFCCL---
<i>Vavraia culicis</i> Rac1	DKTFIR-----LM----FEYLCWCCV---
<i>Trachipleistophora hominis</i> Rac1	DKTFFR-----LM----FEYFCWCCV---
<i>Hamiltosporidium tvaerminnensis</i> Rac1	KKRSSFFS-----SC-----FGCCCHKTAVKN
<i>Edhazardia aedis</i> Rac1a	KTQKERKNL (7) -----FLRWCMCY----
<i>Edhazardia aedis</i> Rac1b	HSLYDS-----YG-CCLSY (17)
<i>Anncalia algerae</i> Rac1	SLKGTKARRG-----CCL-----
<i>Antonospora locustae</i> Rac1	PSKRRRR-----FF-CC-----
<i>Vittaforma corneae</i> Rac1	IVLERG-----YCRRV-----FR-CC-----
<i>Hepatospora eriocheir</i> Rac1	KESIVEEV-----CCPC (5) CF
<i>Ecytonucleospora hepatopenaei</i> Rac1	DDFDKNGCV-----SC-----CCF-----
<i>Enterospora canceri</i> Rac1	VVNRGVER-----SCCG-----CC-----
<i>Enterocytozoon bieneusi</i> Rac1	QNMNTRH-----QC-----CCFY-----
<i>Vairimorpha ceranae</i> Rac1	KEYGKQ-----KCCGL-----FY-CC-----
<i>Vairimorpha apis</i> Rac1	KILKND-----RCCGI-----FN-CC-----
<i>Nosema bombycis</i> Rac1	KNGRNGYI-----SCCGL-----FT-CC-----
<i>Nosema antheraeae</i> Rac1	KNGRTGYI-----SCCGL-----FT-CC-----
<i>Ordospora colligata</i> Rac1	RRSRTE-----SI-----WLCCCKCI----
<i>Encephalitozoon cuniculi</i> Rac1	RKSKFR-----SI-----WF-CCKCG-----
<i>Encephalitozoon intestinalis</i> Rac1	RKSKFR-----SI-----WF-CCKCG-----
<i>Encephalitozoon romaleae</i> Rac1	KKSKFR-----SI-----WF-CCKCG-----
<i>Encephalitozoon hellem</i> Rac1	KKSKFR-----SM-----WF-CCKCG-----

C - predicted as palmitoylated

C - predicted as substrate for FTase by PrePS

C - predicted as substrate for FTase and GGTaseI by PrePS

C) Ras1, Rap1 and Ral (Ras family)

<i>Rozella allomycis</i> Ras1	REIRRHNNK-----TNPKKAK----KKKG C NLL-
<i>Rozella allomycis</i> Ras2	RELRRKQSK-----NEKH-----SGKFCC----
<i>Rozella allomycis</i> Ras3	KETRRIKRL-----ISDKEKMTKP---KVYSC C TIL-
<i>Rozella multimorpha</i> Ras1	SEIRSYSSS-----SKPAGGE----KRRK C LIL-
<i>Rozella multimorpha</i> Ras3	LEVRRIREI-ANADPAASDSAKKEKESMV---RKPSC C VIL-
<i>Rozella</i> sp. PSC023 Ras1	REIRRFNKA-----SNEQRPDR---RKKG C TLL--
<i>Rozella</i> sp. PSC023 Ras2	RAAKRMERVPAVPVS-----K K KK C ILL--
<i>Rozella</i> sp. PSC023 Ras3	RETRKYKSSKATASP----ATAATTATKT---SSTQCA C ILL
<i>Rozella</i> sp. PSC023 Ras4	REIQRYKGGPSGSSG (10) SSNGQQRPRPNDGRGGG C VVL-
<i>Paramicrosporidium saccamoebae</i> Ras1	KEIQKSRQE-----QAAGEKKNAPNR-KSTG C TTM--
<i>Rozella allomycis</i> Rap1	FHDLVRQINQS-----LPEKADKKKK-----K C MI-L
<i>Rozella multimorpha</i> Rap1	FFDLVRQINRS-----LPDKADKKKK-----K C LI-L
<i>Rozella</i> sp. PSC023 Rap1	FFDLVRQINKS-----LPDKVDKKKK-----K C TL-L
<i>Paramicrosporidium saccamoebae</i> Rap1	FMRLVEVALKEED (30) RKDETGKQKK (8) RQKGRK C VI-Q
<i>Mitosporidium daphniae</i> Rap1	FESI IALSASKRL----SVDTLDDKKH---QGRARR C EI-L
<i>Metchnikovella incurvata</i> Rap1	FFGLVRKEMKK-----RTKN---QRRGILCCIGN
<i>Amphiamblys</i> sp. WSBS2006 Rap1	FYGLCKSLCAK-----KKKA---S C SFFLCRH-F
<i>Rozella allomycis</i> Ral	KEDEKEEDRSVFQLYITVARMIRDYKQTHPKNDSSKSLLS C FRS C FGK
<i>Rozella multimorpha</i> Ral	RINVEE-----IYATLLRKII EQKLASTEVENAGNG C TIL----
<i>Rozella</i> sp. PSC023 Ral	RTNVEE-----AYTTLMRRVRDVKGAQAEARARHAKR C VLL----

C - predicted as palmitoylated
C - predicted as substrate for FTase by PrePS
C - predicted as substrate for GTaseI by PrePS
C - predicted as substrate for FTase and GGaseI by PrePS

Figure S7. Prenylation and palmitoylation of the Rho and Ras family proteins. The C-terminal part of the multiple sequence alignments of the Rho1 (A), Rac1 and Cdc42 (B), and Ras, Rap1, and Ral (C) protein sequences is shown. Cysteine residues predicted to be palmitoylated are shaded in grey. Cysteine residues predicted by the prenylation prediction suite PrePS to be prenylated by farnesyltransferase (FTase) and/or geranylgeranyltransferase I are shown in a different colour according to the legend.

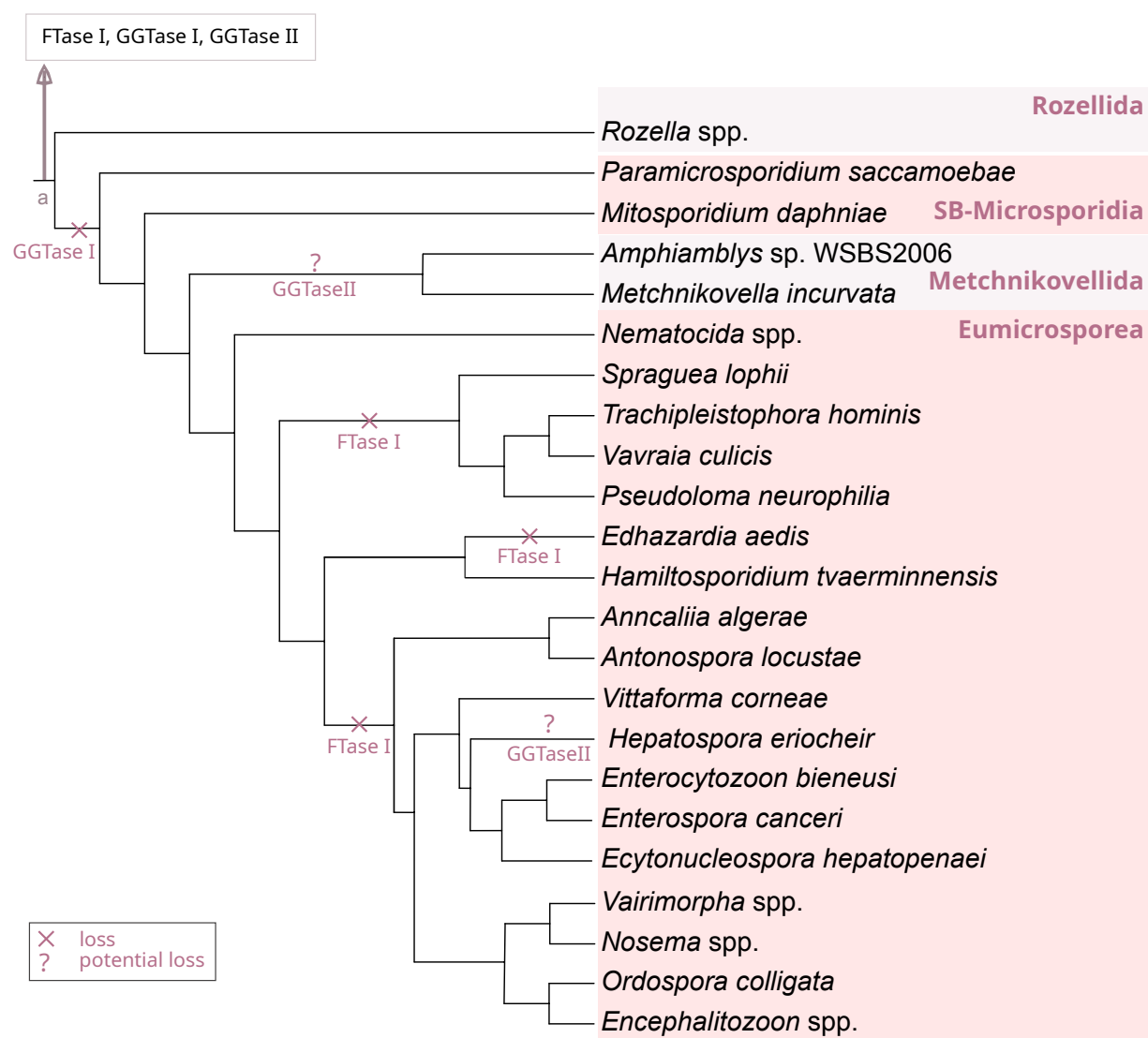


Figure S8. Loss of prenyltransferase subunits in the RM clade. The loss events were inferred from examining the distribution of the subunit homologs in extant RM clade members as described in the main text and mapped onto a schematic topology of the group (the same as the one used in Fig. 4).



Figure S9. Sequence logos of Arf family proteins. Sequence logo of putative eumicrosporean Arf1 and Arl1, Arf1 and Arl1 from rozellids, SB-Microsporidia, and metchnikovellids and Arf1 and Arl1 from other Eukaryotes. To provide pan-eukaryotic perspective on conserved amino acid sequences, sequence logos from the rest of ancestral pan-eukaryotic Arf family are included. For details on the sequences, see Methods and Table S12. The pink rectangle marks position where a cysteine residue is shared between putative eumicrosporean Arl1 and Arl1 from their close relatives.

GxxxxGKS/T

T

DxxG

G1

G2

*

G3

<i>R. allomycis</i> Arf1	-----MGLNISKL-F---SFFG K EMR IL MLGLDAAG KTT ILYKL-----KLGEI---VTIPTIG F NVET V E---Y K NIS F TVWD V GG Q
<i>R. allomycis</i> Arf6	-----MGLWFSRIFS---GFK K P IR LL MLGLD N AG KTT ILYQ-----IKLKQN---IPTIPTVG F NVESL---D F ENIK F NVWD V GG Q
<i>R. allomycis</i> Arl1	-----M G ALFSRIFS---RLFG E K YR IL LGLD N AG KTT ILYRL-----QLGEV---LSTIPTIG F NMEVLQ---Y K N L K F QVWD L GG Q
<i>R. allomycis</i> Arl2	-----MGLLTVIKKNRM--- K E R IR IL LGLD N SG KTT ILKKL-----NN-EPIDTVSPTLG F NI K TM---E H L N Y K LNIWD V GG Q
<i>R. allomycis</i> Arl8	-----M C LTS P CKTLFEWLKALI C K R E IE V T I L G LQ N SG KTS FVNSM-----ANE F Q E ISVPTVG F N V R K I---K K G Y I L R I CD V GG H
<i>R. multimorpha</i> Arl13	-----MFS C FP-RRK---IHPT K N E I V V V L GHSG V G K SH M ISTL-----KEGRGK--EVILPTKG F SIET V T (4) N G D S F L V K F Y E TGG S
<i>R. allomycis</i> Arfrp1	-----MYTLFKGLYS---LITQ R E E Y N L L I G I D N AG KTT LLEKI----K S RYSPKGGLPPEKIVPTIG L N F ANI---F F K G Y K L H FWD L GG Q
<i>R. allomycis</i> Sar1	-----MFILDWFRSILDFLG L S Q NA K L L F L G L D NAG KTT L L HVL-----KN-SRMATLQPTLHP T SEEL---I V D S V K F T T F DL G G H
<i>N. displodere</i> ArLEM	-----M-- G G-IIS-K C IN R K R S V V F I G AAG S G KTS L L IL G F-----KG-NQETSAGTVT I Y E Q S L I ---R G N G Y T ID L FD L SG N
<i>N. parisii</i> ArLEM	-----M G SAISK C L N R E K SA V F I G A K D SG K K E I I K I L-----RKDKT---D V IS N M Y E H F T V---R G K G Y S T E V Y A I S E D
<i>N. ironsii</i> ArLEM	M G SAIS-----K C L N R E K SA V F I G A K D SG K K E I I K I L-----RKDKT---D V IS N M Y E H F T V---R G K G Y S T E V Y A I S E D
<i>N. ausubeli</i> ArLEM	M G SAIS-----K C L N R E K SA V F I G P K D SG K R D I I K I L-----RKDKT---E Q AV N M Y E Q F T V---R G K G Y S T E I Y A I S E D
<i>S. lophii</i> ArLEM	-----M V F N C S -----P A I R C R M I V I G L D R P G K R R F Y S K L T G-----E H D E I E N T L V K H L E Q K I---E Y N G L Q I Y L C L D E N
<i>P. neurophilia</i> ArLEM	-----M-----R C C -----E S K I K S R I T V L S K N K Y S T I V W Q L A K S-----NDVPQ-R N S H L E K N I---Y L N L K V C L C T V D S-
<i>V. culicis</i> ArLEM	---MSL-----S C C -----K K Q S R A R I T L L G P D T K S T N T M Y TR L T K S-----T K E R N T H A R H L E T L T---S L N S L K V S L C N V E C -
<i>T. hominis</i> ArLEM	---MSI-----Q C C -----T K Q S R A R I T L L G P D T K G T N M I Y TR L T K S-----T K G R N T H A R H L E T L T---T L N S L K V S L C N V E C -
<i>H. tvaerminnensis</i> ArLEM	-----M G Q L C S D C -----K R P V C R I V I L G L D E A G M Y L Y S N L T-----N K K E E H K N E L V Y K H D E I---K L N G M S I F L C S V L G D
<i>E. aedis</i> ArLEM	-----M G N I E K C C -----G A R K C R I T I L G L D V V A A K L F E K-MA-----K I N V D S T H V Q H F S E S I ---K K Y N V S T T L V Y V P A G
<i>A. algerae</i> ArLEM	-----M G S I C S V C L-----S K K S F D I S I V E T D Q E K R E S L Y K F I T K -----G H T L ---K G Q Y Q E I K L---K N Y K Y N S L K V L N T D
<i>A. locustae</i> ArLEM	-----M G F V Y S F C Y-----K E K K Y F L V I L G I D K K F K K V F A R L T G-----S E P A L-E S K S R Y E E K K V---Y Y G G L S T I V C C L D D D
<i>V. corneae</i> ArLEM	-----M S N-PLN- C C -----C P N Q K Y T I Y I T-G S M E D K R I F C E K V F N-----I K L R D-I S F S E N L V ---H F N G T S L I L L V I A N E
<i>H. eriocheir</i> ArLEM	-----M F I S D-L C F C L N -----G K K K E T K A F L I -G K N Q Y K K V L K D F F D -----I E V N N F D D F Q T I P A ---D I N G N K I T L L M H S D T
<i>E. hepatopenaei</i> ArLEM	-----M S F W D S C C L G L -----G G E P V K K T V F V I G K D E D K M W F L E Q F G I N E S N K G Y S T K K I K K M E Q D D N S T Y Q E Y S I---C L N S T V V I F S F S S E
<i>E. canceri</i> ArLEM	-----M S F F Q-M C N- C C -----G E P P K K P V F I I-G P D K E K K W F L -E Q F G -----I T I K D-D V Y Q E F K A ---M V N G E L V L F S C L S Q A
<i>E. bieneusi</i> ArLEM	---MF-----G F C -N C C -----A E A S Y K P V F I -I G S D K D K E F L D R -----L G I V T S K E Q- Y Q E Y R Y---K A S E N V V F S C L N D K
<i>N. ceranae</i> ArLEM	-----M G N I C -S C L Q -----K D I P K Y I I S V V G- L E E D I K N V L V H I V S-----G F D-K E I K L T K F S E Y V I---T F Y D I E L I I H A H I I T
<i>N. apis</i> ArLEM	-----M G N F C -S C L D -----I P K R K L I I S V I G - L E E D V Q I V L K T I L P-----Y Y --V S N P L P K F S E H S L---V K G R Y K I I I H A H I I T
<i>N. bombycis</i> ArLEM	-----M G N F C -S C L N -----T K K P K C I V S I I G - L E E D V K Q I L N V I V P-----N F D-P H I C L S K F S E Y P I---T Y N N A E I I I N A H I N
<i>N. antheraeae</i> ArLEM	-----M G N F C -S C L N -----T K K Q K C I V S I I G - L E E D V K Q I L N V I V P-----N F D-L Q S K L S K F S E Y P I---T Y N N A E I I I N A H I N
<i>O. colligata</i> ArLEM	-----M K W I L S L C C R -----S V R K R Y A V T L L G -L Q N D Q R R V L Q Y L N P -----N G -K D N K N I R P F S E D V C ---V V H G M E L V L Q S L D D D
<i>E. cuniculi</i> ArLEM	-----M E W I L S L C C G -----V T R K Y I V A L L G- R Q E D Q R H V L K H S V S-----G Y -K E D K S L R P F S E D T C ---I A N G M E L I F Q S L D D N
<i>E. intestinalis</i> ArLEM	-----M D W V L S L C C G -----A T R K Y V V A L L G- L E E D Q R H V L K H S V S-----N Y -K E D K S L R P F S E D T C ---I A N G M E L I F Q S L D D N
<i>E. romaleae</i> ArLEM	-----M N W V L S L C C G -----T T R K Y V I T L L G- L Q E D Q R H V L K H S I S-----S Y -N M D K P L R P F S E D I C ---I A N G M E L I L Q S L D D N
<i>E. hellem</i> ArLEM	-----M S W I F T L C C G -----A T R K Y V V T L L G- L Q E D Q R H V L R H S I S-----N Y -K M D K S L R P F S E D T C ---I A N G M E L I L Q S L D D N
Consensus/80%	p.c..lhb1G..p.spp.hb.....apb..h....h.shph.h.sl..p

N/TKxD

G4

<i>R. allomycis</i> Arf1	DKIRPLWRHYFQNTQGIIFFVVD	SN-DRDRVNEARDELMRMLNEDEL--RDALLLVFANKQDL	PNAM---QAAEITDKL-----GLHSLR----QRDW	
<i>R. allomycis</i> Arf6	AKIRPLWRHYFTGTGLGLIFVIDS	-QDKERIEESRQELVKIVSDREM--KNCVILIFANKQDL	ERSM---GPTELSEQL-----QLHKFVG----DRKW	
<i>R. allomycis</i> Arl1	SSIRPYWRCCYYANTQAIIVVVD	ST-DKARLSVSKAELISMLEEEEL--KDAVLLVFANKQDL	NGAM---SVAEVSEGL-----GLSLLR----NRQW	
<i>R. allomycis</i> Arl2	KSIRAYWRNYFEQTDGLVWVVD	DS-MDERRLLDCKNELSQLQEERL--FGVSLLLIMANKQDI	PGAMV---PDKIQNALDLQS-----IKTHSW	
<i>R. allomycis</i> Arl8	QRFRPLWERYCRNAQATAFVIDL	-SDNRSLLLAKVEFQNLLE-NVN--PDVPILILANKCDL	--A---NS-T---EVDEVKNMFVCL--GTQHARL	
<i>R. multimorpha</i> Arl13	EALQRLWPNNYAESHGIIYLF	KIFDDKAINELQQKTFEELTKDERL--SGKPIILV-CINEEF	SGDD---QLTTKLDAL-----SK----EREL	
<i>R. allomycis</i> Arfrp1	EEIRRLWTTYYKECHGVFFVIDS	-TDKSRLMEVQNSFESLILHHEA--DGVPVILIANKQDK	GNM---KIVEIKEVF---NKIAMVID---ARES	
<i>R. allomycis</i> Sar1	KQARRLWRDYFPEVSGIVVMVDC	-ADVKRLNESIEEFGNLLTMEEI--SKVPIILVLGNKID	LPGAIS--EELRYRFNLQMTTGKGKVALAENIRAI	
<i>N. displodere</i> ArLEM	QDHTQFWKFFTEHSHLVVFIVDI	-SSKESVQQSKDVFEAFYTSYAI--QKDNIIIFLLNKIDL	--V-PEEARNE----LVSLFHSEFMSVVGMMKGNTQ	
<i>N. parisii</i> ArLEM	QNHVNFWKFFYTQNCNLVVFVNI	-ASESLIEESKNVFENFTTSYGT--NKESIVFMLNGR	REEETETEEKAEEKRQNILKFKEYFATISGLDKKILD	
<i>N. ironsii</i> ArLEM	QNHVNFWKFFYTQNCNLVVFVNI	-ASESLIEESKNVFENFTTSYGT--NKESIVFMLNGR	REEETETEEKAEKNRQNILKFKEYFATISGLDKKILD	
<i>N. ausubeli</i> ArLEM	QNHINFWKFFYTQNCNLVVFVNI	-SEEKQIEESKNVFENFTTSYGT--NKESIAFMLCRK	RE--GETEEKSTLEDQRNSLKFKEHFSTIAGLDKKLLE	
<i>S. lophii</i> ArLEM	EGMRKLWSCNAKLADALYVIDA	ADS-KNFDYSVSEFKTITEGEKNIEK-MNVLLFCNNA	TNMDG-----RL---NVLEKEYGGY---KNI	
<i>P. neurophilia</i> ArLEM	-TDLDLWEKNVKYSEGVLIIIE	EAYEI-NSPE-----VQNFISDLNKQIILVLCNT	TNIGDT-----N-----ID--DSPD---SQT	
<i>V. culicis</i> ArLEM	-HDKELWSLNMRFSDGVVFIVD	VDED-KLDL-----ERLLKDRMDRHVVLMCN	TRNIGDC-----V-----VESAKDWG---DRV	
<i>T. hominis</i> ArLEM	-HDKELWTLNMRFSDBGVVFVN	VDDDD-KLDL-----DRLVDRMERHVVLMCN	TRNIGDC-----V-----VDGVKDWG---DRV	
<i>H. tvaerminnensis</i> ArLEM	EKARNLWSVQIQNSDGIIFVVD	ISNE-ASYKEITEEFQRMIDVKAFTEG-MNILLCCN	NGLKTD-----EYREVGDAIFSPFLEN---PGI	
<i>E. aedis</i> ArLEM	ERLVNIWNAEALHSDGVIIYVD	DKKES-DDSLSPIPQLEDLFSNVIENG	D-NFLLVVVRNGEKLTD-----DE---LKMVCD-FNK---EKA	
<i>A. algerae</i> ArLEM	LYSADISDTLTKFCCQGVFVCD	DLTK--QSYGEVITEFENYLSHFVKN-K-CFVIVFCY	VYTYT--PA-----E-----EFEQYASL---NSF	
<i>A. locustae</i> ArLEM	NITHELWTVHVNCANACVFVCN	PNDL--ESQEHIKTFEWQFEGKKLDKQD-FCVLIIVH	SKSRFTG-----P-----ILNNYNNL---KNV	
<i>V. corneae</i> ArLEM	GDLRDVHQMHMKMANAVIL	-----L-RDRD-----ENVETDK----	KTLFVRMNGKSQ-----E-----SE-----N	
<i>H. eriocheir</i> ArLEM	VTDENLRLNIQLEHCSHLIF	-----L-TDED-----VEQFYKK----	PTILEFLMRGVSHQ-----R-----NR-----E	
<i>E. hepatopenaei</i> ArLEM	IDDPNVEEIHLEASVAVVVLNT	TDEM-V-----SDSVEKCVSLN	EEVIKT-DVVVFCK-SNVYGS-----VSDALRRSCEEIG-H---GKI	
<i>E. canceri</i> ArLEM	VDDPNVEELHTKACLGVVV	-----L-SEHD-----VPVTTNK----	PTFFIYTGGQSNQ-----L-----SN-----G	
<i>E. bieneusi</i> ArLEM	LEDPNIEEEHIKNCLGVVYL	-----SPHD-----HPITIK----	KPVFFIYTAGESKQ-----I-----S---KD---NFI	
<i>N. ceranae</i> ArLEM	PMIKTLIGLHSSVESAVIYSMD	AD-----SKESIGAIREYVE	ESEVRTKNDIVTVLCSR-GSYLDN-----PELDTVQMQFT-GNIH---DHF	
<i>N. apis</i> ArLEM	PLIKSLIKLHFPVESAVIYSID	AD-----NPTSIDHARMLM	DEQHIKND-SVTILCSR-GKYLEN-----PVLQEIKD	EIS-GGFH---SHL
<i>N. bombycis</i> ArLEM	SLIKNLINLHKSSEFASIYAIN	AD-----NKLSLESARNIM	DDQERIQQ-KLISVLCK-GKYLDS-----PGINELK	VMFQ-ETIH---DSF
<i>N. antheraeae</i> ArLEM	SLIKNLINLHKSSEFASIYAIN	AD-----NKLSLESARNIM	DDQERIQQ-KLISVLCK-GKYLDS-----PGINELK	VMFQ-ETIH---DSF
<i>O. colligata</i> ArLEM	SDLYEIRGIHVSHCAAVMFVSD	ADD-----SDSVEKCVSLN	EEVIKT-DVVVFCK-SNVYGS-----VSDALRRSCEEIG-H---GKI	
<i>E. cuciculi</i> ArLEM	ADLYEIRGIHTSLAAAVVFAVN	ADN-----KASVERCIAL	INEDATKTR-DTIVVLCSR-SDIYGS-----EAVQELK	RTCDEIA-H---GKI
<i>E. intestinalis</i> ArLEM	ADLYEIRGIHMGLSAAAVFAVD	ADN-----KTSVEKCITL	IGEDATKTK-DVIVVLCK-SNIYGN-----ESVQELK	RTCDEIS-H---GKI
<i>E. romaleae</i> ArLEM	ADLYEIRGIHMGLSAAGVIFA	VDADN-----KSSVEKCI	ALINEDTTKTK-DIIVVLCSR-SNIYGN-----ESVQELK	RTCDEIG-H---GKI
<i>E. hellem</i> ArLEM	ADLYEIRGIHMGLSAAGVIFA	VDADS-----KSSVEKCI	ALINEDTTKTK-DIIVVLCSR-SNIYGN-----ESVQELK	RTCDEIG-H---GKI
Consensus/80%	plb.hbh..s.tllahs.....hp.....lhhbhp..pb.....			

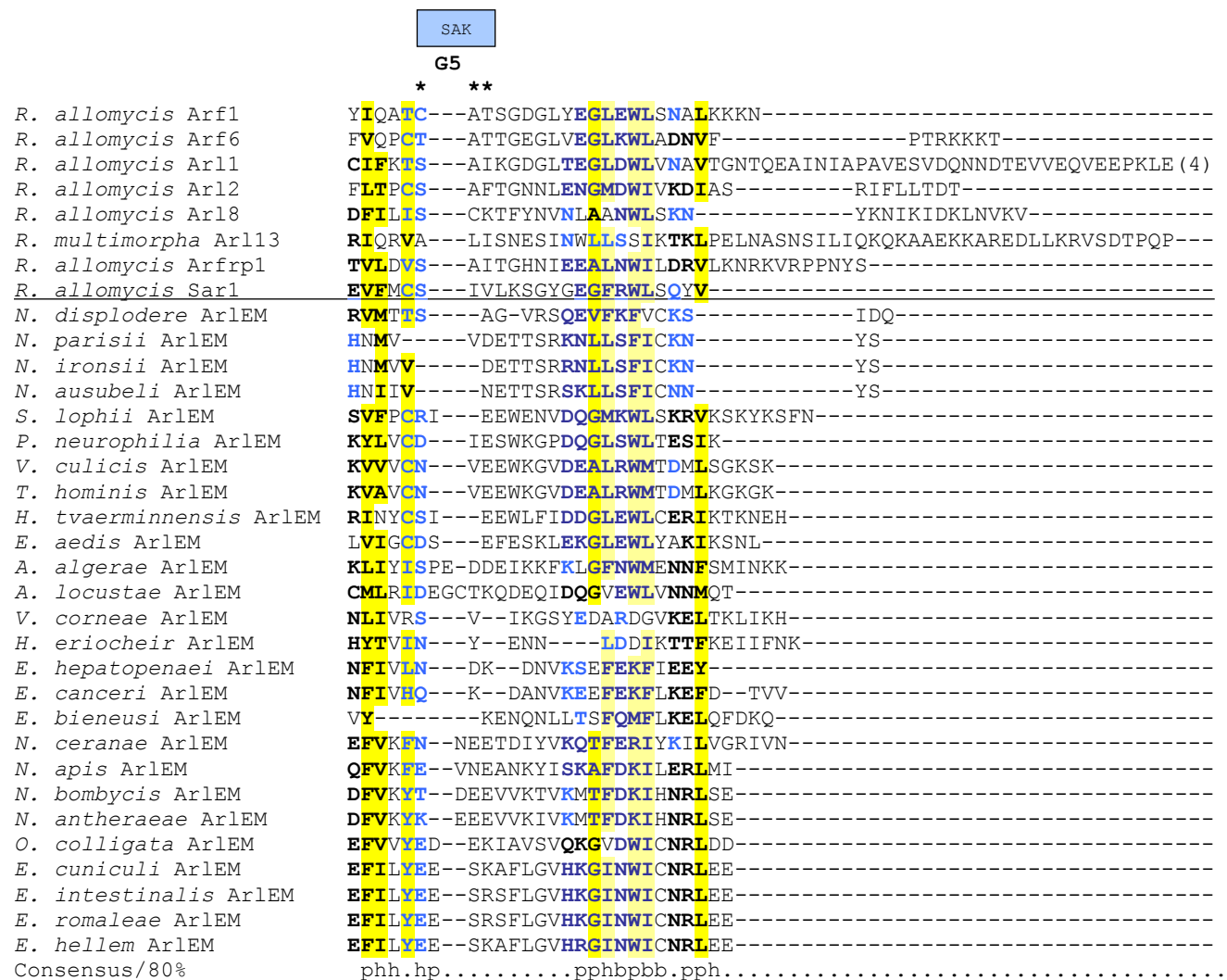


Figure S10. Multiple sequence alignment of eumicrosporean ArLEM proteins. The alignment displayed is the comprehensive version of the one presented in Fig. 8, including ArLEM sequences from the full set of eumicrosporean taxa analysed. See the legend to Fig. 8 for details.

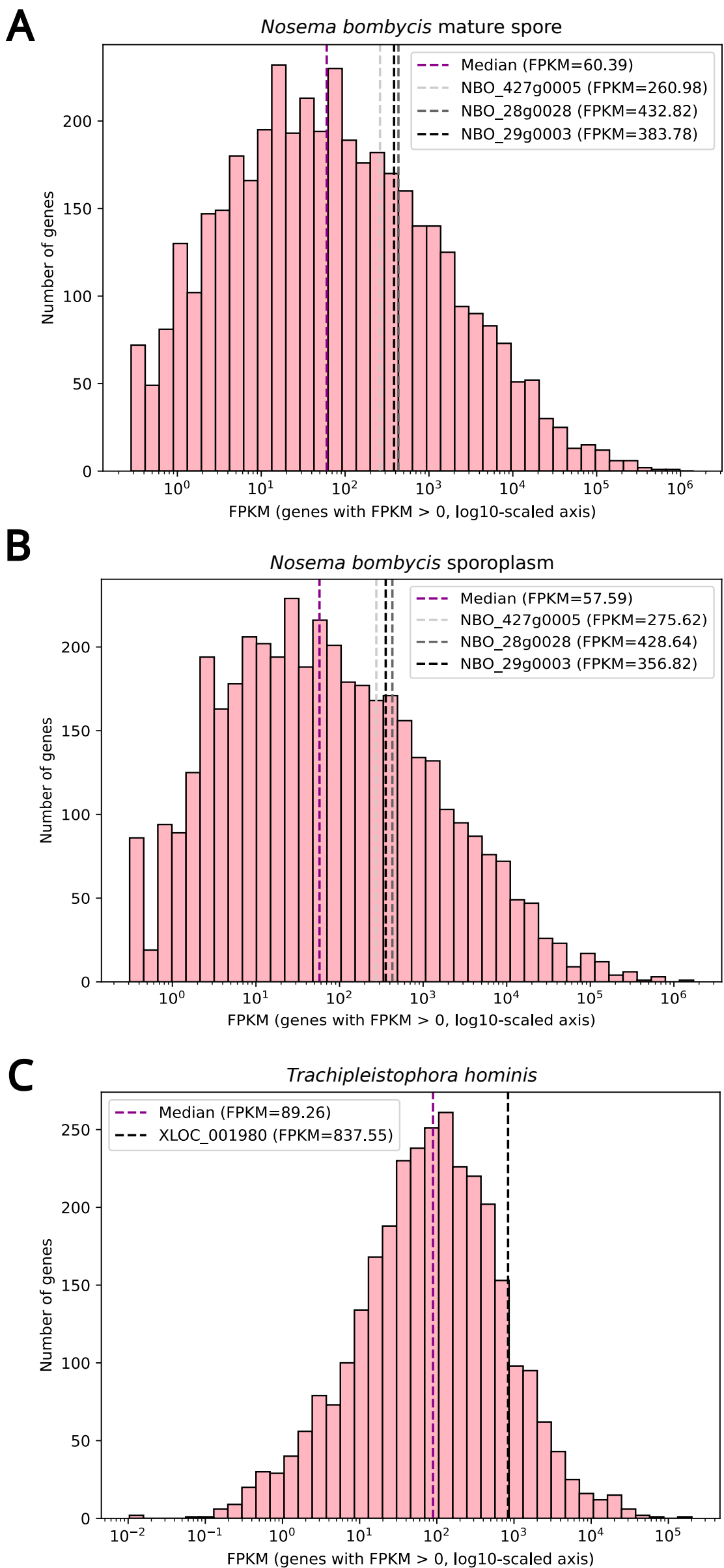


Figure S11. Transcriptomic evidence for ArlEM gene expression. Distribution of FPKM values for all expressed genes (FPKM >0) based on RNA-seq data from *Nosema bombycis* mature spores (A) and sporoplasm (B), and from *Trachipleistophora hominis* (C). The generation and mapping of the RNA-seq data onto the genome of *N. bombycis* and *T. hominis* were previously published by He et al. [148] and Watson et al. [149], respectively. Mean FPKM values across replicates were used to generate histograms with log-spaced bins on a log10-scaled x-axis. Vertical dashed lines indicate the median FPKM of all expressed genes (magenta) and the FPKM value of the ArlEM gene and its variants (black, grey). In *N. bombycis*, the three lines correspond to three slightly different ArlEM versions (allelic variants or recent gene duplications), whereas in *T. hominis*, a single black line marks its sole ArlEM ortholog. In all datasets, ArlEM and its variants are expressed at above-median levels.

N. displodere -----MSIKGVVVRVFSRGSI-----VRKEECMQAYLAHFITEMEL-TPGVQ-----TKRVNRNDYISYVLGVV-----NIYLAVFQSR-----ELETEIEQRL-----LETKERTESK
N. parisi M-----ESRERVLSLVGTGRGSI-----VLGERSNMGGYLESFLQKAQP-FSTIK-----TIKIRNRDLITYSFGTS-----NIYIVYKEA-----ETEREIEDAL-----NEAKERIEKS
N. ironsi M-----ESRERVLSLVGTGRGSI-----VLGERSNMGGYLESFLQKAQP-FSTIK-----TIKIRNRDLITYSFGTS-----NIYIVYKEA-----ETEREIEDAL-----SEAKERIEKS
N. ausubeli M-----ENKEKVLTLVGTGRGSI-----VLGETCSMAAYLESFLQKMQA-FSTIK-----SVKVRNDRITYTFGVS-----NIYLVVYKEA-----ELEKEIESAV-----MDAKDKTEKS
P. neurophilia -----MSIELRFTKHN-----VLFEQSSSEFLPNKYLEKLRLDETAINRNDTVINRNETVQGLNVKWSVDN-----EVVIMVNDQDK-----SDNNLKKNNLRTGS (11) DHSKDNKDHSK-----DIKDHSSKKND
V. culicis -----MLLSTIFDTHN-----VHHQSGTKPXYIDSIIDTVVR-KNIQHG-----EIMVQGRNVAYRMIGK-----RVFMCVGDDANA-----QIDRMAAAYQEGGT-----AGERAVDERS
T. hominis -----MLLSTIFDLHN-----VHHQSGTKPSYINNIIDTVVR-KNIKHG-----ELTVQGRNVAYRTVGK-----KVMIMCVGDDADE-----QIDRMAAEYLEDGA-----VREHGVDYDR
D. muelleri -----MITAITFTSTN-----TLKSGKREQLYLSKFINNQLV-NSLNSN-----NSPLKIQGNYHYIIVKD (25) LVYLLIDSK-----KIDNFSRK-----SEACERIEKS
H. tvaerminnensis -----MFCVYTKKSI-----LRYQSGEKEGYLDEIILRVGI-SPIQK-----TSLVRGKNITVESRGS-----DFFLIIDSKQTFD-----SFILIEKFYCNFE-----DFDKNFDEIIK (6) SKEDKQSSLNK
E. aedis -----LNFIIFNTKIV-----LIFEKSTAKFRDNFITNIDI-NKLSG-----TESVRDINIFYKINGR-----TIYTLFTEESNCE-----ALDKKEIKLF-----EKPKEEVN
A. algerae M-----LNFIIFNTKIV-----LIFEKSTAKFRDNFITNIDI-NKLSG-----TESVRDINIFYKINGR-----TIYTLFTEESNCE-----ALDKKEIKLF-----EKPKEEVN
A. locustae -----MHFVIFNKKSI-----LFEFEGEATFLNSLIRITINI-QQASG-----SRNIRDTLMEXKIEGQ-----RIYLSFGNKP-----SLDEFIKKH-----KHKDSTENK
V. corneae -----MKN--NF--FIVFETSISI-----VYFSKGNKEIFINNFIANMDPDQLKS-----SLKVRDKFVEYEVGRP-----LVYLSVGNKP-----SLNELLNKF-----DCEKNEFIEEK
H. eriocheir -----MLNQTRDMLFVFDKRGF-----VFYETGPAPYFLDEIILKLDI-TNHN-----EKKIRDKFLETKTTEGD-----IVYLCYSTTNK-----HLEEFIEYNYKTTI-----LKQK-----
E. hepatopenaei MH (9) PMANHHGNFAVFDLRF-----LWHEEGTAFYLNDFIVKMDR-TKMSE-----KKVRDMVAEWRIEGD-----LIFYLAIGAKEG-----VLDTKVVEEF-----KRN-----
E. canceri -----MVS--TE--FIVFGKNF-----LYYESGDAPYLLNNLIGKIDL-KSVYN-----TKIIRQDKFVEYKIQGM-----LIFCISNNKN-----NLKFIKDYITI-----YPNRLYNDNL
E. bieneusi -----MSFVVFENKRN-----VLYEESVVESEFVKNFVLSIDP-NILVG-----SKIVRDDKIEYEURGP-----LVYLSVGHQK-----GLHELIER-----QKEPEKEEK
V. ceranae -----MPFCVFNKKIW-----LQOESIPESYLNSSIIDL-STLSG-----SRIVRDDKVEYEURGP-----IYYSVGGRK-----TLKEMIQD-----KNEVKKEVK
V. apis -----MPFCVFNKKIW-----LQOESIPESYLNSSIIDL-STLSG-----SRIVRDDKVEYEURGP-----IYYSVGGRK-----TLKEMIQD-----KNEVKSNVK
N. bombycis -----MGFCVFSKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
N. antheraeae -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
O. colligata -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
E. cuniculi -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
E. intestinalis -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
E. romaleae -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
E. hellem -----MGFCVFNKKKL-----LVHEEGNRDNLNQMIISLDI-SKLG-----LRKVKDQIMEYEIRGP-----TVYLSVSKEA-----SLEKIKAYLGI-----NNEKPSALNL
S. lophii -----MFTCTIITDSSDIT-TIEQSTPTPEFNSFISNTSF-NNIFDQ-----QIKQVQDYHILLYKNKS-----KYNIIISKDINDL (4) TIDKKKID-----LEETTKK
C. dikerogammari ME---D---IYRAEIKLSKELQNHFEIERVKKEFYFDNLLQITA-QKISG-----FIKTNILKQNDIISYKIMDS-----MLYLVVSNRPV-----LEETTKK
Consensus/80% Y.hlhsp.G.....lb.ppG..P.abpphl.p.h.....b.lps..h.h.s.....lal.h.pp.....lc...p.....

N. displodere HSNQKSARSE-----EKRPEGLFEES-SEDKEERYVQTSYT-----DSDEEDRTGWERWKKSKSEFDL-----KQGS-----A-----ITEV-----SIRQHYYSKNVPHELSSTQ
N. parisi EN-----SRVLGYIFKNEIVGEEAGDSTTEKADAAEE-----IY--DSDEEETTGWDKWKPRNIRL-----TGKK-----E-----VNES-----GIQDHLSRNIPHLLSRK
N. ironsi EN-----SRVLGYIFKNEIAGEETDEFTTEKTDKAAE-----VY--DSDEEETTGWDKWKQPRNIRL-----TGKK-----E-----VNES-----AIQDHLSRNIPHLLSRK
N. ausubeli EG-----AKVLFGIKGGK-EDEDAVDCTESSQKEVAA-----ID--DSDEEETTGWDRWKKPRNLLNFL-----TAKK-----E-----VSEK-----SIQDHLTRNVYPVLSRK
P. neurophilia KDHKDKNKKT--GR-----KWL-GDVSQLDYSSREKVPS-----ESSDEEVHAIDNLFSENNTSF--FQRIKKTLSLTK-NNSLNK-----KNENTLLTKLE--LKQLTES--TEKNVSSDVIYNH
V. culicis AMQKQGGKHR-----HWV-GTERELDYSSSART-----VSY--VPVTKSTNLE-----DVKR-----R-----LFKTFD--REALLTHLNKNVAPVAGD
T. hominis AAVTKQEKRR-----RWG-EKEKELDYSSSTRT-----VSY--VPDTKSAFNL-----NIKR-----R-----VFKTFD--REALLALHVNKNVAPVAGD
D. muelleri -----NVLFFF-NEQESLDYSIKHEKRR-----LKEQSLIQITFLQPIIH--NFRSILD-IT--NKFYLIKIKL-----NIMI-----R-----WSSKKFN--RKSLYNQVAKNVSIPTVR
H. tvaerminnensis IKQSKVRRWS-----EEN-LDAAILDYSSSEKSIPTY-----ESKLNTI-TK--KFDPPRKSSD-----SKKI-----K-----TEEL-----KEKLSTHVDKNVSLAISQK
E. aedis -----SIE-KEKDILDYSSVSTIPN-----IEDTFE--VK--KPSFLRKNL-----SKEI-----K-----VSEL-----KEKMKMHVNKNVAPMIAES
A. algerae ENAKKMHTEI--EE-----QKC-DERDALDYSTSTVV-----QEIVDMNIS--KVIPTRSRL-----SNKI-----H-----VDDI-----YNMKTHVNKNVDPVAAAM
A. locustae QHESKGNKEI-----EKP-VVSDTLDYSETGSNI-----SSI IKT-IR--KADFKKASD-----TGKV-----S-----IEQL-----QSKMIGHVNKNVDPPTFCRI
V. corneae QS-----IEV-KEKSELDFSVSEDTS-----SFTTIKNI IASSVIKLQNKTSGL-----TGRI-----S-----LDEL-----ERIILNHVNKNVTEYANL
H. eriocheir -----DSEPEKTFNKWH-----ERIKNKTGGL-----SNKI-----K-----LSDI-----VIRIREKVNKNVIEPKYAES
E. hepatopenaei -----GRSAVKNQDGKI-----GRVTRKTLQGL-----TGKI-----E-----LGL-----RTKMVREKVNKNVTPKYAEQ
E. canceri N-----NISASKIRN-----VS--QTLFYKSSGL-----TGKI-----H-----FKEL-----KAKIIDHFRNMNTEIKYANI
E. bieneusi K-----KVV-SQEEKLNYSDVSVIR-----RNFINI-TT--KKAFKKGSGL-----SNKI-----H-----VDEL-----KTKMELHVNKNVDYNTTKI
V. ceranae -----KEVIKDKVI--DN-----SLQ-EIINPLDYSTRSSLSVSP-----SL--T-----PTHSS-----FNSKFMRKNL-----SKEI-----S-----VSDI-----RNKMELHVNKNVDYNTTKI
V. apis SNV-KSNVKS (6) DK-----QIN-EDINPLDYSTRSSSLLP-----PI--S-----NQPSLNS-----ISSKFMRKNL-----SKEI-----S-----VSDI-----RNKMELHVNKNVDYNTTKI
N. bombycis KYQKKIDREG-----KC-----EDV-EKEDELNYSETSSV-----GSFVDS-FK--KATIRKAGGL-----KNQI-----K-----VDL-----KDKMTHVNKNVSYQITNT
N. antheraeae KKADKGHKR--RW-----EDA-GDEGGLDYSTSSSV-----NYFVDA-VK--KATVKKAGGL-----KGEI-----D-----VNEL-----KDKMTTHVNKNVSYQITKT
O. colligata KRAWKKNKER--KW-----EDI-GNEEELNYSSPSSGV-----SNFVKV-IK--KATIOKAGGL-----KDQI-----Q-----VKEL-----KDKMTTHVNKNVSYQITRT
E. cuniculi KKTRNSIKGR--KW-----EDA-GNEEELNYSTSSGI-----SNLVEV-IR--KATVQKAGGL-----KDEI-----Q-----VSEL-----KDKMTTHVNKNVSSQITKT
E. intestinalis KKLGNAGAGR--KW-----EEV-GNEEELDYSTSSRI-----SNLVEV-IR--KATVQKAGGL-----RDKI-----Q-----VSEL-----KDKMTTHVNKNVSYQITKT
E. romaleae KIQTKNKEK--KR-----QWI-GDEKELNFSINIKEKKERKHKE-----KVIEEKIH--IK--NVFLKIVV-----IMQIFSFLLK-----IR-----FNKYNITINALETKEKLSNIVSK
E. hellem -----IETNKKIFDPPKPVKIKPENLKNTTF-----ETFFSIFDPLKNILETIYII--FKPLKKVFAILLINI KIIYFKIKSFYLO-----FFRSKLT--RENAIKI--LNKNFDLNVTEK
S. lophii sb*.p.....p.F.lF...p.pb.....p...h.c.....b..+lpl+Nls..hsp.
C. dikerogammari
Consensus/80%

A) Jpred 4 results for the eumicrosporean with standard SRA proteins

Net, jhm, jpssm, Lupas 14, Lupas 21, Lupas 28, Jnet_25, Jnet_5, Jnet_0, Jnet Rel

B) Jpred 4 results for the S. lophii SRA protein

Net, jhm, jpssm, Lupas 14, Lupas 21, Lupas 28, Jnet_25, Jnet_5, Jnet_0, Jnet Rel

C) Jpred 4 results for the C. dikerogammarii SRA protein

Net, jhm, jpssm, Lupas 14, Lupas 21, Lupas 28, Jnet_25, Jnet_5, Jnet_0, Jnet Rel

Legend

Secondary structure code:

alpha helix ('H')

beta sheet ('E')

not H or E ('-')

Colour code for alignment:

Blue - Complete identity at a position

Shades of red - The more red a position is, the higher the level of conservation of chemical properties of the amino acids

Jnet - Final secondary structure prediction for query

jalign - Jnet alignment prediction

jhm - Jnet hmm profile prediction

jpssm - Jnet PSIBLAST pssm profile prediction

Lupas - Lupas Coil prediction (window size of 14, 21 and 28)

Note on coiled coil predictions

- = less than 50% probability

c = between 50% and 90% probability

C = greater than 90% probability

Jnet_25 - Jnet prediction of burial, less than 25% solvent accessibility

Jnet_5 - Jnet prediction of burial, less than 5% exposure

Jnet_0 - Jnet prediction of burial, 0% exposure

Jnet Rel - Jnet reliability of prediction accuracy, ranges from 0 to 9, bigger is better.

Figure S13. Secondary structure of eumicrosporean SRA protein sequences. Results of the secondary structure prediction by the program JPred 4 are shown for SRA proteins from all Eumicrosporea except for S. lophii and C. dikerogammarii (A) and for S. lophii (B) and C. dikerogammarii (C). The legend is provided at the bottom part of the figure.

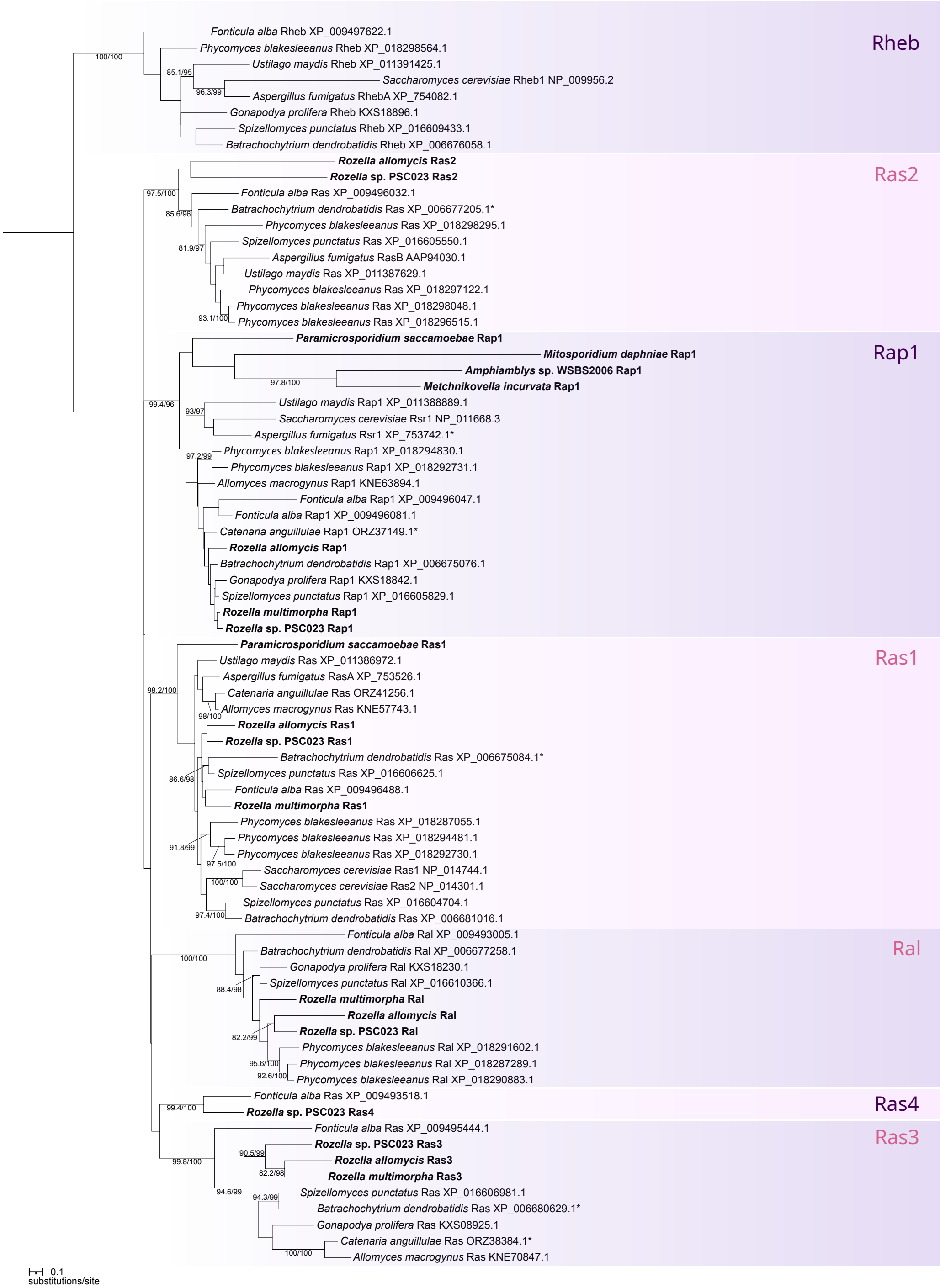


Figure S14. Phylogenetic analysis of the Ras family GTPases from the RM clade. The tree shown is a result of a ML analysis of reference Ras family sequences from ten outgroup Holomycota species (see Table S9B, sequences with corrected protein models or with protein model created de novo are denoted by * and **, respectively), *R. allomycis*, *R. multimorpha*, *Rozella* sp. PSC023, *P. saccamoebae*, *M. daphniae*, *M. incurvata*, and *Amphibamblys* WSBS2006 sp. The tree was inferred using IQ-tree with LG+I+G4 model with the ultrafast bootstrap (UFB) algorithm and the SH-aLRT test (both 10 000 replicates, only support values ≥ 80 for SH-aLRT and ≥ 95 for UFBoot are shown).

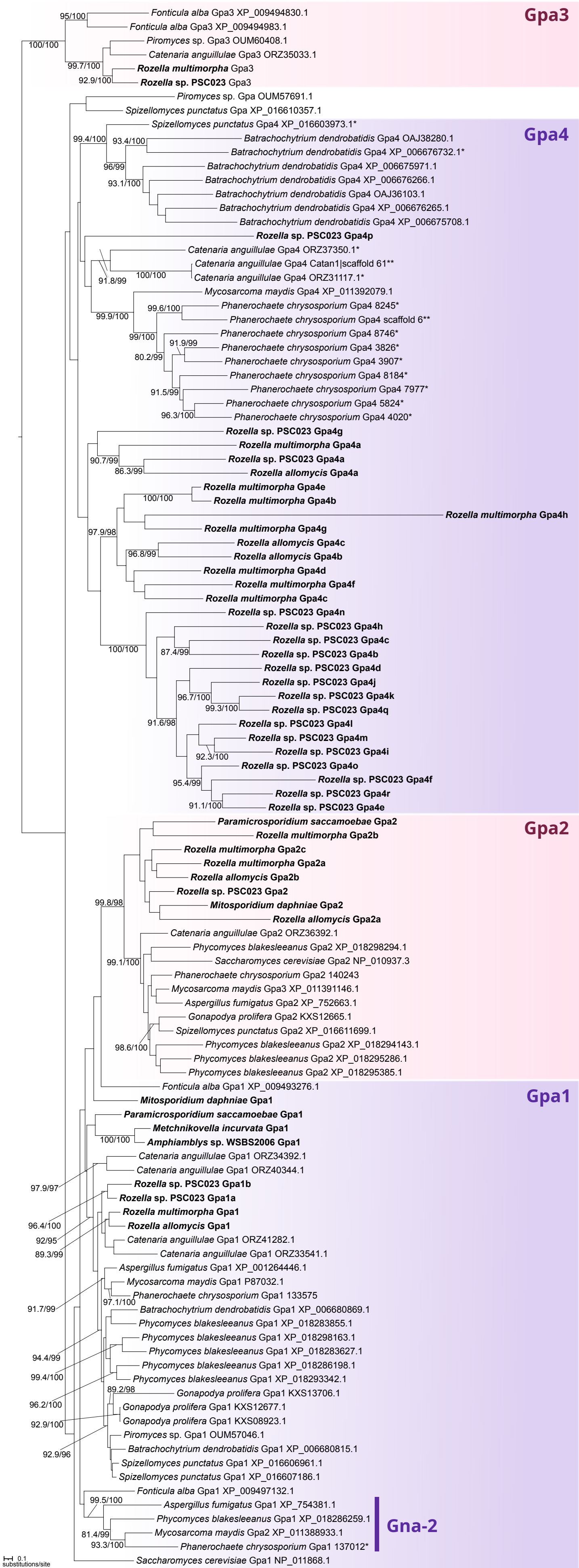


Figure S15. Phylogenetic analysis of Gα protein sequences from rozellids, SB-Microsporidia, and metchnikovellids. The tree shown is a result of a ML analysis of reference Gα protein sequences from eleven Holomycota species (see Table S9C, sequences with corrected protein models or with protein model created de novo are denoted by * and **, respectively), *R. allomycis*, *R. multimorpha*, *Rozella* sp. PSC023, *P. saccamoebae*, *M. daphniae*, *M. incurvata*, and *Amphiamblys* sp. WSBS2006. The tree was inferred using IQ-tree with Q.yeast+F+I+G4 model with the ultrafast bootstrap (UFB) algorithm and the SH-aLRT test (both 10 000 replicates, only support values ≥ 80 for SH-aLRT and ≥ 95 for UFB are shown). The fungi-specific subgroup Gna-2 is delineated within the Gpa1 group (see Methods for detailed information).

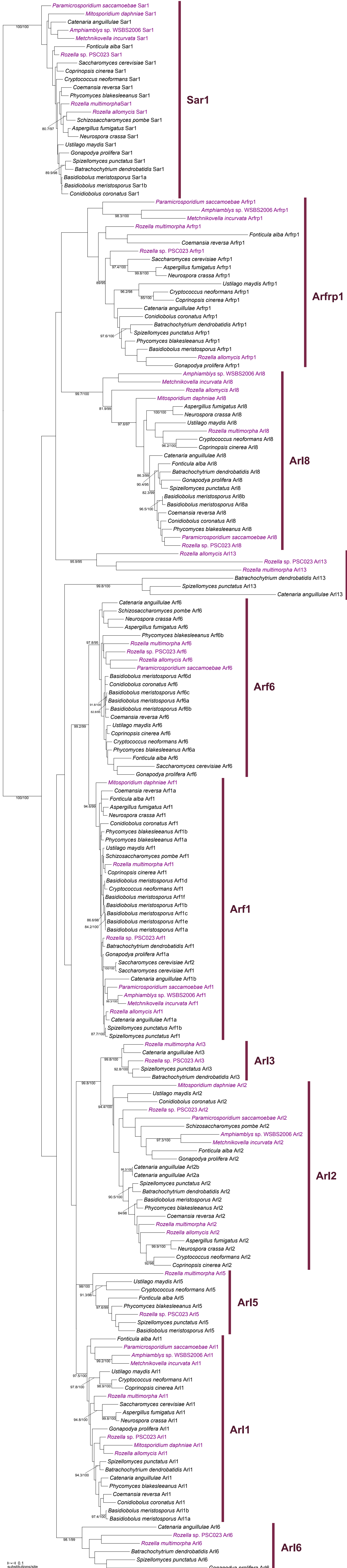


Figure S16. Phylogenetic analysis of Arf family sequences from rozellids, SB-Microsporidia, and metchnikovellids. The tree shown is a result of a ML analysis of a set of Holomycota Arf family sequences from Vargová et al. [46] and Arf family sequences from *R. allomycis*, *R. multimorpha*, *Rozella* sp. PSC023, *P. saccamoebae*, *M. daphniae*, *M. incurvata*, and *Amphiamblys* sp. WSBS2006. The tree was inferred using IQ-TREE with LG+I+G4 model with the ultrafast bootstrap (UFB) algorithm and the SH-aLRT test (both 10 000 replicates, only support values ≥ 80 for SH-aLRT and ≥ 95 for UFBoot are shown).