

Table S3. PRP8 extein and intein sequences used in this study. SAR indicates the Stramenopiles-Alveolata-Rhizaria hypothetical supergroup.

Organism	Eukaryotic supergroup	Protein identifier	Database source*	Intein presence
<i>Wild Bathycoccus</i>	Plantae	AFUW01000096	GenBank	yes
<i>Bathycoccus</i> strain BBan7	Plantae	F Bathy18g01340	Bogas	no
<i>Micromonas pusilla</i> CCMP1545	Plantae	XP_003061371	RefSeq nr	no
<i>Micromonas</i> sp. RCC299	Plantae	XP_002509397	RefSeq nr	no
<i>Ostreococcus tauri</i>	Plantae	CAL55631	EMBL	no
<i>Ostreococcus lucimarinus</i>	Plantae	XP_001420401	RefSeq nr	no
<i>Ostreococcus</i> sp. RCC809	Plantae	59290	DoE-JGI	no
<i>Coccomyxa subellipsoidea</i>	Plantae	EIE25534	RefSeq nr	no
<i>Volvox carteri</i>	Plantae	XP_002956689	RefSeq nr	no
<i>Chlamydomonas reinhardtii</i>	Plantae	XP_001689471	RefSeq nr	no
<i>Chlorella variabilis</i>	Plantae	EFN52246	RefSeq nr	no
<i>Populus trichocarpa</i>	Plantae	XP_002327417	RefSeq nr	no
<i>Vitis vinifera</i>	Plantae	CBI36339	RefSeq nr	no
<i>Arabidopsis thaliana</i>	Plantae	NP_178124	RefSeq nr	no
<i>Physcomitrella patens</i>	Plantae	XP_001785044	RefSeq nr	no
<i>Phaeodactylum tricornutum</i>	SAR	XP_002179646	RefSeq nr	no
<i>Thalassiosira pseudonana</i>	SAR	XP_002287554	RefSeq nr	no
<i>Phytophthora infestans</i>	SAR	XP_002901763	RefSeq nr	no
<i>Ectocarpus siliculosus</i>	SAR	CBN74629	RefSeq nr	no
<i>Aureococcus anophagefferens</i>	SAR	EGB05047	RefSeq nr	no
<i>Tetrahymena thermophila</i>	SAR	XP_001015771	RefSeq nr	no
<i>Paramecium tetraurelia</i>	SAR	XP_001434030	RefSeq nr	no
<i>Perkinsus marinus</i>	SAR	XP_002774058	RefSeq nr	no
<i>Plasmodium knowlesi</i>	SAR	XP_002261014	RefSeq nr	no
<i>Theileria annulata</i>	SAR	XP_954981	RefSeq nr	no
<i>Toxoplasma gondii</i>	SAR	XP_002368068	RefSeq nr	no
<i>Capsaspora owczarzaki</i>	Opisthokonta	EFW41973	Broad OoM	yes
<i>Salpingoeca rosetta</i>	Opisthokonta	PTSG_00100	Broad OoM	yes
<i>Monosiga brevicollis</i>	Opisthokonta	XP_001747121	RefSeq nr	no
<i>Mus musculus</i>	Opisthokonta	NP_619600	RefSeq nr	no
<i>Daphnia pulex</i>	Opisthokonta	EFX85628	RefSeq nr	no
<i>Drosophila melanogaster</i>	Opisthokonta	NP_610735	RefSeq nr	no
<i>Caenorhabditis elegans</i>	Opisthokonta	NP_498785	RefSeq nr	no
<i>Nematostella vectensis</i>	Opisthokonta	XP_001633987	RefSeq nr	no
<i>Spizellomyces punctatus</i>	Opisthokonta	SPPG_02894	Broad OoM	yes
<i>Batrachochytrium dendrobatidis</i>	Opisthokonta	EGF81600	Broad OoM	yes
<i>Phycomyces blakesleeanus</i>	Opisthokonta	Phybl1_79621	Broad OoM	yes
<i>Rhizopus oryzae</i>	Opisthokonta	EIE76368	RefSeq nr	no
<i>Cryptococcus neoformans</i>	Opisthokonta	XP_566595	RefSeq nr	yes
<i>Cryptococcus gattii</i>	Opisthokonta	XP_003191948	RefSeq nr	yes
<i>Lacaria bicolor</i>	Opisthokonta	XP_001875909	RefSeq nr	no
<i>Ustilago maydis</i>	Opisthokonta	XP_757683	RefSeq nr	no
<i>Schizosaccharomyces pombe</i>	Opisthokonta	NP_593861	RefSeq nr	no
<i>Saccharomyces cerevisiae</i>	Opisthokonta	EDV09210	RefSeq nr	no

<i>Neosartorya fischeri</i>	Opisthokonta	EAW23970	RefSeq nr	yes
<i>Paracoccidioides brasiliensis</i>	Opisthokonta	XP_002791246	RefSeq nr	yes

*: RefSeq nr: NCBI reference sequence non-redundant database; Broad OoM: Broad Institute Origins of Multicellularity Database.