

TABLE S4. List of the RCA sequences used in phylogenetic analysis. Accession numbers for GenBank entries, contig ids for MMETSP^a, and EST id for the CGP^b are listed.

	ORGANISM	SOURCE (Database)	ACCESSION NUMBER/CONTIG
Euglenozoa	<i>Euglena longa</i>	GenBank	KT818573
	<i>Euglena gracilis</i>	GenBank	EC682479; EC679240; EG565129
	<i>Eutreptiella gymnastica</i> CCMP1594	MMETSP	CAMNT_0046458117
	<i>Eutreptiella gymnastica</i> NIES-381	MMETSP	CAMNT_0000686649
Chloroplastida	<i>Chlamydomonas reinhardtii</i>	GenBank	XP_001692244
	<i>Arabidopsis thaliana</i>	GenBank	NP_565913
	<i>Oryza sativa</i>	GenBank	P93431
	<i>Ostreococcus tauri</i>	GenBank	XP_003078787
	<i>Ostreococcus lucimarinus</i>	GenBank	ABO95802
	<i>Glycine max</i>	GenBank	ADD60244
	<i>Medicago truncatula</i>	GenBank	XP_003604901
	<i>Micromonas</i> sp. RCC299	GenBank	XP_002501510
	<i>Zea mays</i>	GenBank	NP_001104921
	<i>Vitis vinifera</i>	GenBank	XP_002270571
	<i>Volvox carteri</i>	GenBank	XP_002951971
	<i>Physcomitrella patens</i>	GenBank	EDQ82463
	<i>Pycnococcus provasolii</i>	MMETSP	CAMNT_0053510249
	<i>Pyramimonas amyliifera</i>	MMETSP	CAMNT_0041893585
	<i>Pyramimonas obovata</i>	MMETSP	CAMNT_0006861995
	<i>Pyramimonas parkeae</i>	MMETSP	CAMNT_0035452761
	<i>Coccomyxa subellipsoidea</i>	GenBank	XP_005651127
	<i>Bathycoccus prasinos</i>	GenBank	XP_007512493
Glaucophyta	<i>Cyanophora paradoxa</i>	CGP	EST111910
	<i>Cyanoptyche gloeocystis</i>	MMETSP	CAMNT_0041989161
	<i>Gloeochaete witrockiana</i>	MMETSP	CAMNT_0038655913
Cyanobacteria	<i>Synechococcus</i> sp. JA-3-3Ab	GenBank	WP_011429335
	<i>Anabaena</i> sp. CA	GenBank	Q06721
	<i>Nostoc</i> sp. PCC 7120	GenBank	P58555
	<i>Arthrospira</i> sp. PCC 8005	GenBank	CDM97162
	<i>Kamptonema</i> sp.	GenBank	WP_007357073

^aMMETSP - Marine Microbial Eukaryote Transcriptome Sequencing Project (<http://marinemicroeukaryotes.org/>)

^bCGP - Cyanophora Genome Project (<http://cyanophora.rutgers.edu/cyanophora/home.php>)