Table S1. The number of protein sequences in the database that was used for the phylogenomic analysis (based on phyla).

	Grouping	Species/Strain	RefSeq	JGI	EST[1]	Independent[2]	Total
PROKARYOTES	Archaea	121	225,867	0	0	0	225,867
	Bact-Actinobacteria	282	994,292	0	0	0	994,292
	Bact-Aquificae	10	19,325	0	0	0	19,325
	Bact-BacteroidetesChlorobi	141	450,081	0	0	0	450,081
	Bact-ChlamydiaeVerrucomicrobia	38	79,759	0	0	0	79,759
	Bact-Chloroflexi	15	52,585	0	0	0	52,585
	Bact-Cyanobacteria	68	225,555	0	0	0	225,555
	Bact-Deferribacteres	2	5,338	0	0	0	5,338
	Bact-Deinococci	12	26,191	0	0	0	26,191
	Bact-Dictyoglomi	2	3,656	0	0	0	3,656
	Bact-Elusimicrobia	2	2,305	0	0	0	2,305
	Bact-Environmental	2	408	0	0	0	408
	Bact-FibrobacteresAcidobacteria	6	28,629	0	0	0	28,629
	Bact-Firmicutes	759	2,099,809	0	0	0	2,099,809
	Bact-Fusobacteria	25	59,335	0	0	0	59,335
	Bact-Gemmatimonadetes	1	3,935	0	0	0	3,935
	Bact-Nitrospirae	3	6,366	0	0	0	6,366
	Bact-Planctomycetes	6	36,794	0	0	0	36,794
	Bact-Proteobacteria	1239	4,251,165	0	0	0	4,251,165
	Bact-Spirochaetes	44	72,342	0	0	0	72,342
	Bact-Synergistetes	6	13,162	0	0	0	13,162
	Bact-Tenericutes	55	32,455	0	0	0	32,455
	Bact-Thermotogae	11	20,807	0	0	0	20,807
	Bact-Unclassified	9	17,518	0	0	0	17,518
EUKARYOTES	Alveolata	70	167,836	0	584,904	0	752,740
	Amoebozoa	22	30,550	12,410	138,624	0	181,584
	Cryptophyta	8	1,419	0	40,320	0	41,739
	Excavata	30	134,643	0	443,424	0	578,067
	Haptophyta	5	140	39,124	56,868	0	96,132
	Opisthokonta-Choanoflagellida	4	9,203	0	74,886	0	84,089
	Opisthokonta-Fungi	186	569,377	212,456	132,168	0	914,001
	Opisthokonta-Metazoa	2120	1,067,024	140,855	30,108	0	1,237,987
	Opisthokonta-Others	4	0	0	46,494	0	46,494
	Plantae-Glaucophyta	3	149	0	57,696	0	57,845
	Plantae-Rhodophyta	23	1,168	0	331,482	28,975	361,625
	Plantae-Viridiplantae	228	385,435	114,102	114,294	0	613,831
	Rhizaria	5	1,211	0	29,112	0	30,323
	Stramenopiles	47	41,980	81,762	96,078	0	219,820
	Vira	2475	84,202	0	0	0	84,202
	Others	39	1,062	0	0	0	1,062
	Total	**8,128**	**11,223,078**	**600,709**	**2,176,458**	**28,975**	**14,029,220**

[1] The actual numbers of EST contigs are the numbers in this column divided by 6 due to six-frame translations.
[2] These data represent protein models from *Cyanidioschyzon merolae* and *Calliarthron tuberculosum*.