

1 **Supplementary information**

2

3 Table S1. Number of protein sequences, average length of protein sequences, habitat, and multicellularity in the analyzed proteomes

Taxonomic class	Species	Number of sequences		Average of protein sequence length after filtering	Habitat	Unicellular/Multicellular
		before filtering	after filtering			
Land plants	<i>Arabidopsis thaliana</i>	35386	26947	414	Terrestrial	Multicellular
	<i>Physcomitrella patens</i>	42392	25775	392	Terrestrial	Multicellular
	<i>Selaginella moellendorffii</i>	22285	20225	386	Terrestrial	Multicellular
	<i>Klebsormidium flaccidum</i>	16215	15974	536	Freshwater	Multicellular
Green algae	<i>Ostreococcus lucimarinus</i>	7773	7688	380	Marine	Unicellular
	<i>Ostreococcus tauri</i>	7725	7602	385	Marine	Unicellular
	<i>Micromonas sp. RCC299</i>	10103	9969	471	Marine	Unicellular
	<i>Micromonas pusilla</i>	10660	10557	441	Marine	Unicellular
	<i>Bathycoccus prasinos</i>	7892	7734	508	Marine	Unicellular
	<i>Coccomyxa subellipsoidea C-169</i>	9629	9374	425	Freshwater	Unicellular
	<i>Chlorella variabilis</i>	9994	9783	403	Freshwater	Unicellular
	<i>Volvox carteri</i>	14542	13925	569	Freshwater	Unicellular

	<i>Chlamydomonas reinhardtii</i>	16706	15597	500	Freshwater	Unicellular
	<i>Auxenochlorella protothecoides</i>	6927	6628	396	Freshwater	Unicellular
	<i>Monoraphidium neglectum</i>	16755	14128	367	Freshwater	Unicellular
	<i>Cyanidioschyzon merolae</i>	5014	4642	516	Freshwater	Unicellular
Red algae	<i>Galdieria sulphuraria</i>	7375	6371	419	Freshwater	Unicellular
	<i>Pyropia yezoensis</i>	10327	9731	292	Marine	Multicellular
	<i>Chondrus crispus</i>	9836	8331	380	Marine	Multicellular
	<i>Porphyridium purpureum</i>	8355	8202	476	Ubiquitous	Unicellular
Glaucophyceae	<i>Cyanophora paradoxa</i>	32167	31147	230	Freshwater	Unicellular
Oomycetes	<i>Phytophthora ramorum</i>	15743	13202	505	Terrestrial	Unicellular
	<i>Phytophthora sojae</i>	19027	16833	517	Terrestrial	Unicellular
	<i>Phytophthora infestans</i>	22658	19455	407	Terrestrial	Unicellular
	<i>Phytophthora capsici</i>	17414	16168	388	Terrestrial	Unicellular
Diatoms	<i>Phaeodactylum tricornutum</i>	10402	10063	460	Marine	Unicellular
	<i>Fragilariopsis cylindrus</i>	27143	20616	429	Marine	Unicellular
	CCMP 1102					
	<i>Thalassiosira pseudonana</i>	11776	11484	492	Marine	Unicellular

Other microalgae	<i>Aureococcus anophagefferens</i>	11501	10772	519	Marine	Unicellular
	<i>Ectocarpus siliculosus</i>	16533	15577	519	Marine	Multicellular
	<i>Symbiodinium minutum</i>	47014	42522	555	Marine	Unicellular
	<i>Emiliana huxleyi</i>	38694	21992	401	Marine	Unicellular
	<i>Guillardia theta</i>	25457	19964	432	Marine	Unicellular
	<i>Bigelowiella natans</i>	21706	21145	436	Marine	Unicellular
	Total	593126	510123	437		

4

5 The filtering method is described in “Protein sequence sets” of the “Results and Discussions” section.

6 The habitat information is retrieved from AlgaeBase (<http://www.algaebase.org/>).

7 Unicellular/Multicellular information is referred in each original paper.

8

9 Table S2. Protein annotations by KOG, Pfam, and UniProtKB^a

Taxonomic class	Species	Percentage of annotated sequences ^a (%)					Total ^b
		KOG	Pfam	UniProtKB	GO	PDB	
Land plants	<i>Arabidopsis thaliana</i>	37.4	74.7	86.2	50.8	49.8	89.0
	<i>Physcomitrella patens</i>	32.1	61.0	59.2	40.9	43.0	67.5
	<i>Selaginella moellendorffii</i>	32.5	67.8	64.7	42.0	49.9	74.2
Green algae	<i>Klebsormidium flaccidum</i>	30.0	61.5	55.8	39.1	40.7	67.9
	<i>Ostreococcus lucimarinus</i>	36.0	68.4	64.5	43.7	47.3	75.5
	<i>Ostreococcus tauri</i>	35.8	65.0	74.7	42.1	45.4	76.5
	<i>Micromonas sp. RCC299</i>	33.5	66.0	57.4	42.0	44.7	71.7
	<i>Micromonas pusilla</i>	31.0	61.9	52.1	38.4	41.3	67.5
	<i>Bathycoccus prasinus</i>	35.1	68.4	60.1	43.3	45.2	74.6
	<i>Volvox carteri</i>	27.6	54.3	46.6	34.4	36.8	59.6
	<i>Chlamydomonas reinhardtii</i>	25.3	50.6	45.6	32.1	33.8	57.0
	<i>Monoraphidium neglectum</i>	27.1	52.6	47.8	34.0	36.7	60.7
	<i>Coccomyxa subellipsoidea C-169</i>	37.6	66.4	62.4	43.9	48.0	71.5
	<i>Chlorella variabilis</i>	36.5	65.0	61.9	42.6	46.7	71.0
	<i>Auxenochlorella protothecoides</i>	40.8	67.2	65.9	44.5	48.6	73.3
	Red algae	<i>Cyanidioschyzon merolae</i>	45.7	74.3	68.4	49.0	54.1
<i>Galdieria sulphuraria</i>		41.8	71.2	70.1	47.9	52.9	77.5
<i>Pyropia yezoensis</i>		27.4	52.9	45.7	34.7	36.4	60.2
<i>Chondrus crispus</i>		30.4	52.3	47.7	36.2	39.0	56.7
<i>Porphyridium purpureum</i>		35.5	66.3	57.2	43.3	45.5	71.5
Glaucophyceae	<i>Cyanophora paradoxa</i>	14.0	31.4	25.7	19.5	19.8	37.0

Oomycetes	<i>Phytophthora ramorum</i>	33.8	64.6	53.8	42.8	44.9	69.5
	<i>Phytophthora sojae</i>	27.3	58.6	49.7	38.3	40.3	63.9
	<i>Phytophthora infestans</i>	24.5	52.2	46.0	32.8	31.1	60.4
	<i>Phytophthora capsici</i>	30.5	58.4	50.6	39.1	40.3	64.6
Diatoms	<i>Phaeodactylum tricornutum</i>	32.0	62.6	48.7	39.7	42.0	66.9
	<i>Fragilariopsis cylindrus CCMP 1102</i>	19.5	45.8	32.2	28.2	27.5	50.1
	<i>Thalassiosira pseudonana</i>	29.4	60.1	45.7	38.2	39.4	64.1
Other microalgae	<i>Aureococcus anophagefferens</i>	39.3	70.7	61.6	44.7	53.7	77.7
	<i>Ectocarpus siliculosus</i>	30.0	57.4	54.1	35.6	39.6	65.4
	<i>Symbiodinium minutum</i>	16.7	42.4	31.5	26.0	27.1	47.2
	<i>Emiliana huxleyi</i>	20.6	52.0	34.0	31.4	31.6	56.5
	<i>Guillardia theta</i>	21.5	48.7	34.4	29.2	30.6	52.6
	<i>Bigelowiella natans</i>	24.8	53.5	41.4	33.3	35.4	58.3
Total of Land plants		34.2	67.9	70.7	44.9	47.4	77.3
Total of Algae		26.9	54.6	46.2	34.7	36.6	60.4
Total of all species		28.0	56.5	49.7	36.2	38.2	62.8

10

11 ^aPoor annotations such as “poorly characterized” in KOG, “domain unknown function (DUF)” in Pfam, and “Uncharacterized protein,” “Putative uncharacterized,”

12 “Unnamed product,” and only ID in UniProtKB, were excluded from hits.

13 ^bThe values of total percentage were calculated by combining the results of KOG, Pfam, UniProtKB, GO and PDB.

14

15

16 Table S3. KOG category details

Group	Functional category
Information storage and processing	Translation, ribosomal structure and biogenesis
	RNA processing and modification
	Transcription
	Replication, recombination and repair
	Chromatin structure and dynamics
Cellular processes and signaling	Cell cycle control, cell division, chromosome partitioning
	Nuclear structure
	Defense mechanisms
	Signal transduction mechanisms
	Cell wall/membrane/envelope biogenesis
	Cell motility
	Cytoskeleton
	Extracellular structures
	Intracellular trafficking, secretion, and vesicular transport
	Posttranslational modification, protein turnover, chaperones
Metabolism	Energy production and conversion
	Carbohydrate transport and metabolism
	Amino acid transport and metabolism
	Nucleotide transport and metabolism
	Coenzyme transport and metabolism
	Lipid transport and metabolism
	Inorganic ion transport and metabolism

Secondary metabolites biosynthesis, transport, and catabolism

Poorly characterized

General function prediction only

Function unknown

17

18 This description is based on the file from “<ftp://ftp.ncbi.nih.gov/pub/COG/COG/fun.txt>.”