



Sanchez-Baracaldo, P., Raven, J., Pisani, D., & Knoll, A. (2017). Early photosynthetic eukaryotes inhabited low salinity habitats. *Proceedings of the National Academy of Sciences of the United States of America*, 114(37), E7737-E7745. <https://doi.org/10.1073/pnas.1620089114>

Peer reviewed version

Link to published version (if available):
[10.1073/pnas.1620089114](https://doi.org/10.1073/pnas.1620089114)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via NAS at <http://www.pnas.org/content/114/37/E7737>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Early photosynthetic eukaryotes inhabited low salinity habitats

SUPPORTING INFORMATION

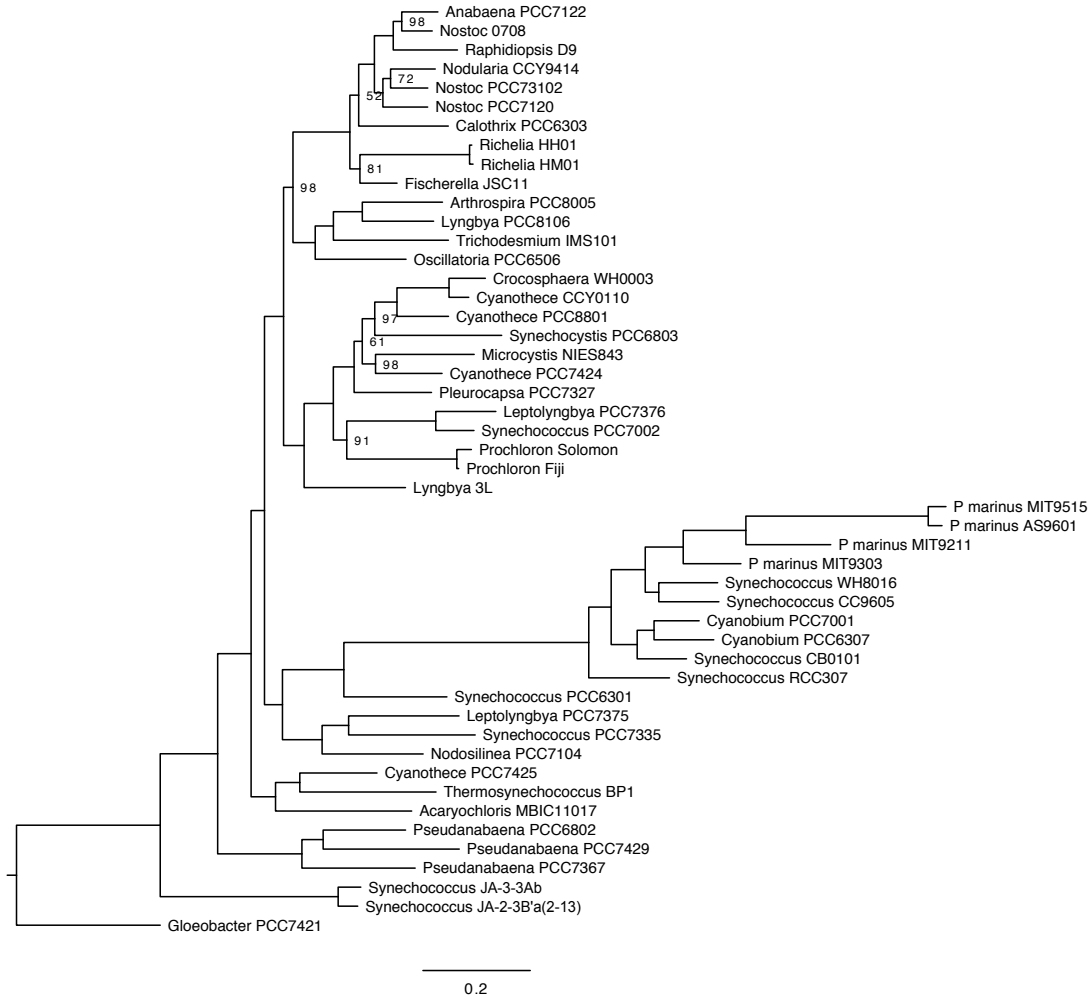
Patricia Sánchez-Baracaldo¹, John A. Raven^{2,3}, Davide Pisani⁴, and Andrew H. Knoll⁵

¹School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK; ²Division of Plant Science, University of Dundee at the James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK; ³Plant Functional Biology and Climate Change Cluster, University of Technology Sydney, Ultimo, NSW 2007, Australia; ⁴School of Biological Sciences and School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol BS8 1RJ, UK; ⁵Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138.

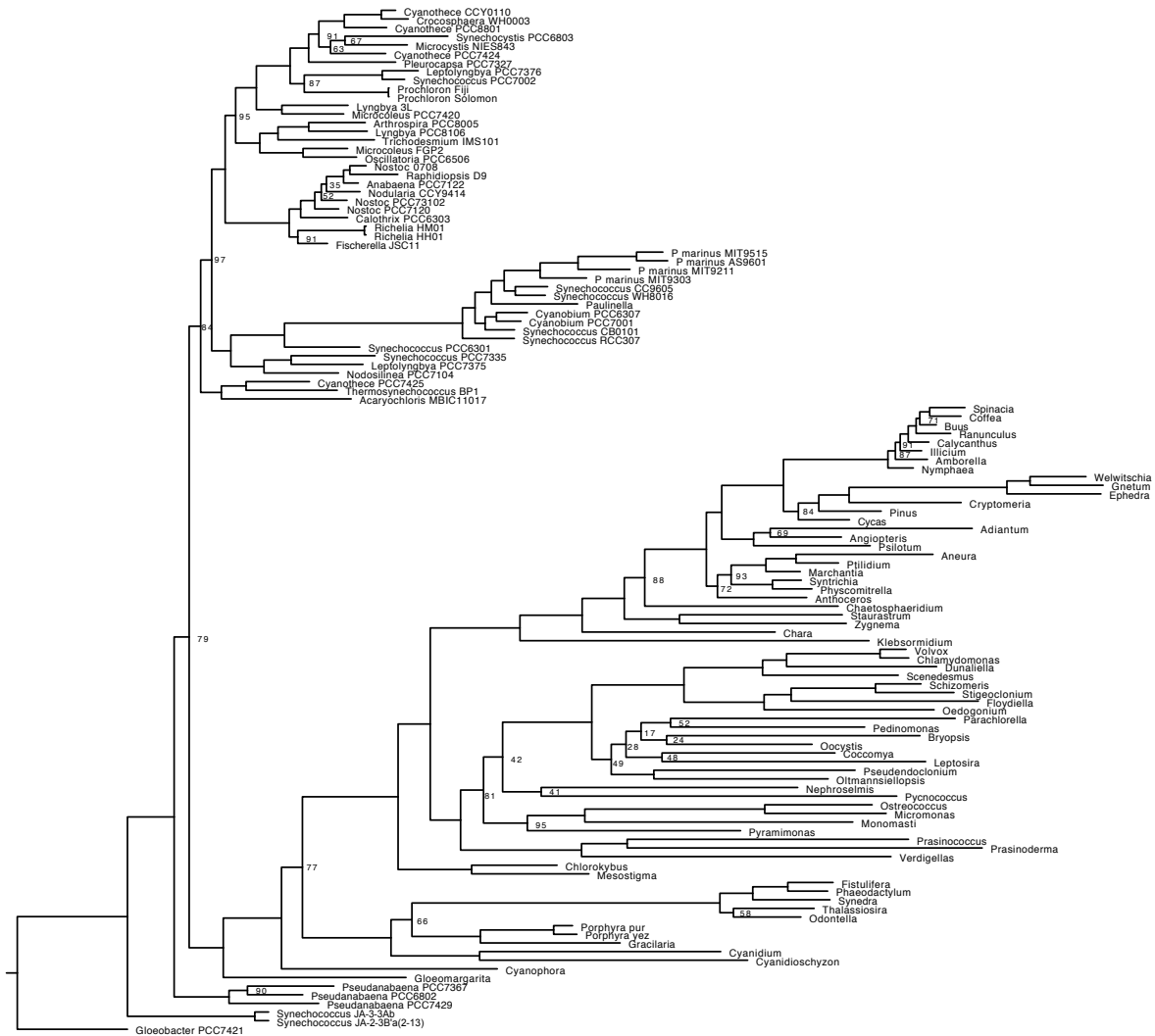
Corresponding author: (p.sanchez-baracaldo@bristol.ac.uk)

Supplementary Figures

Supplementary Figure S1. Genomic tree showing the deep branching relationships for Cyanobacteria. Topology estimated using RaxML (Maximum Likelihood). Bootstrap values are 100 for branches that show no values; values are reported for those that show relatively low values.

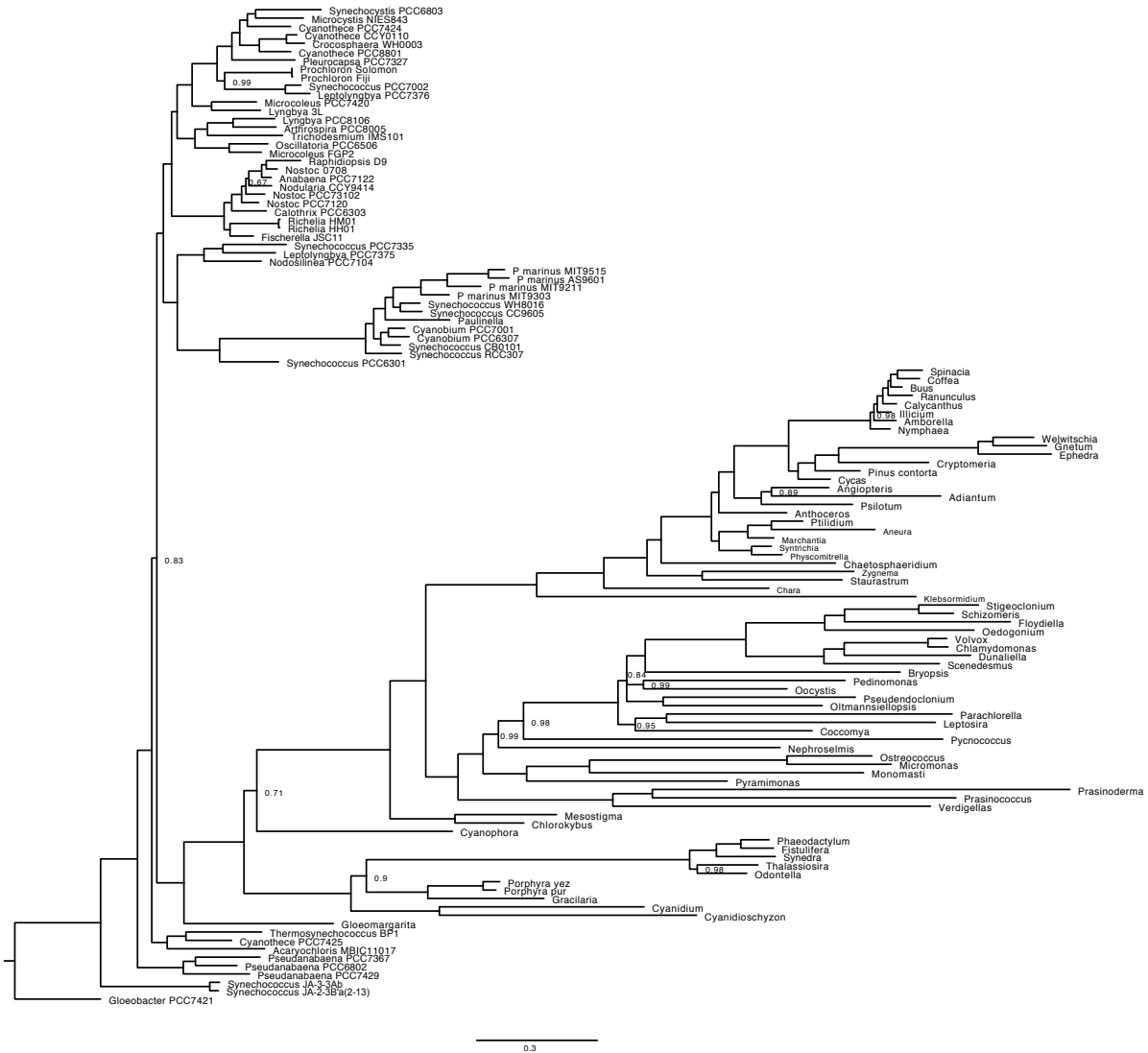


Supplementary Figure S2. Maximum Likelihood phylogeny of Cyanobacteria and photosynthetic eukaryotes. Bootstrap values are 100 for branches that show no values; values are reported for those that show lower values.

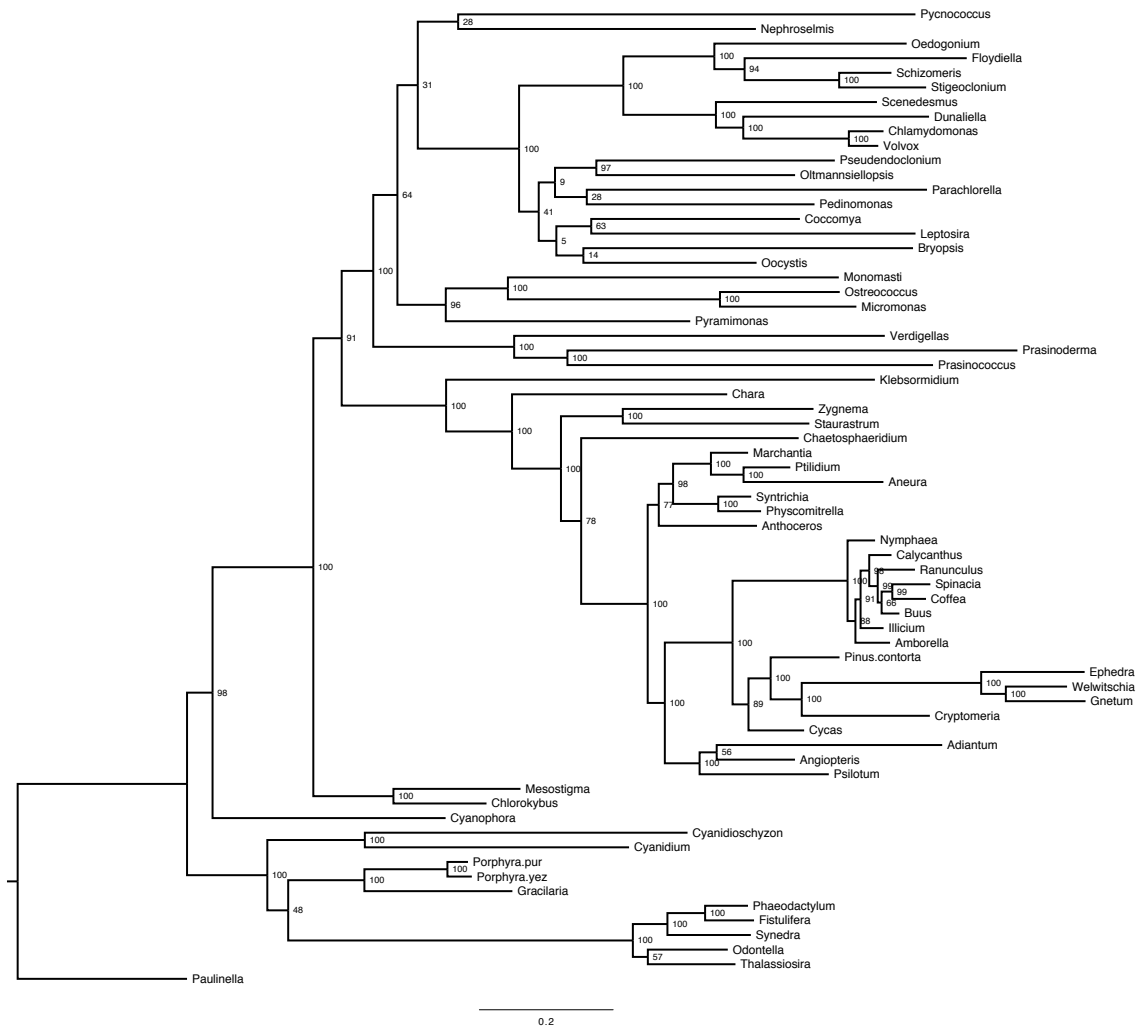


0.2

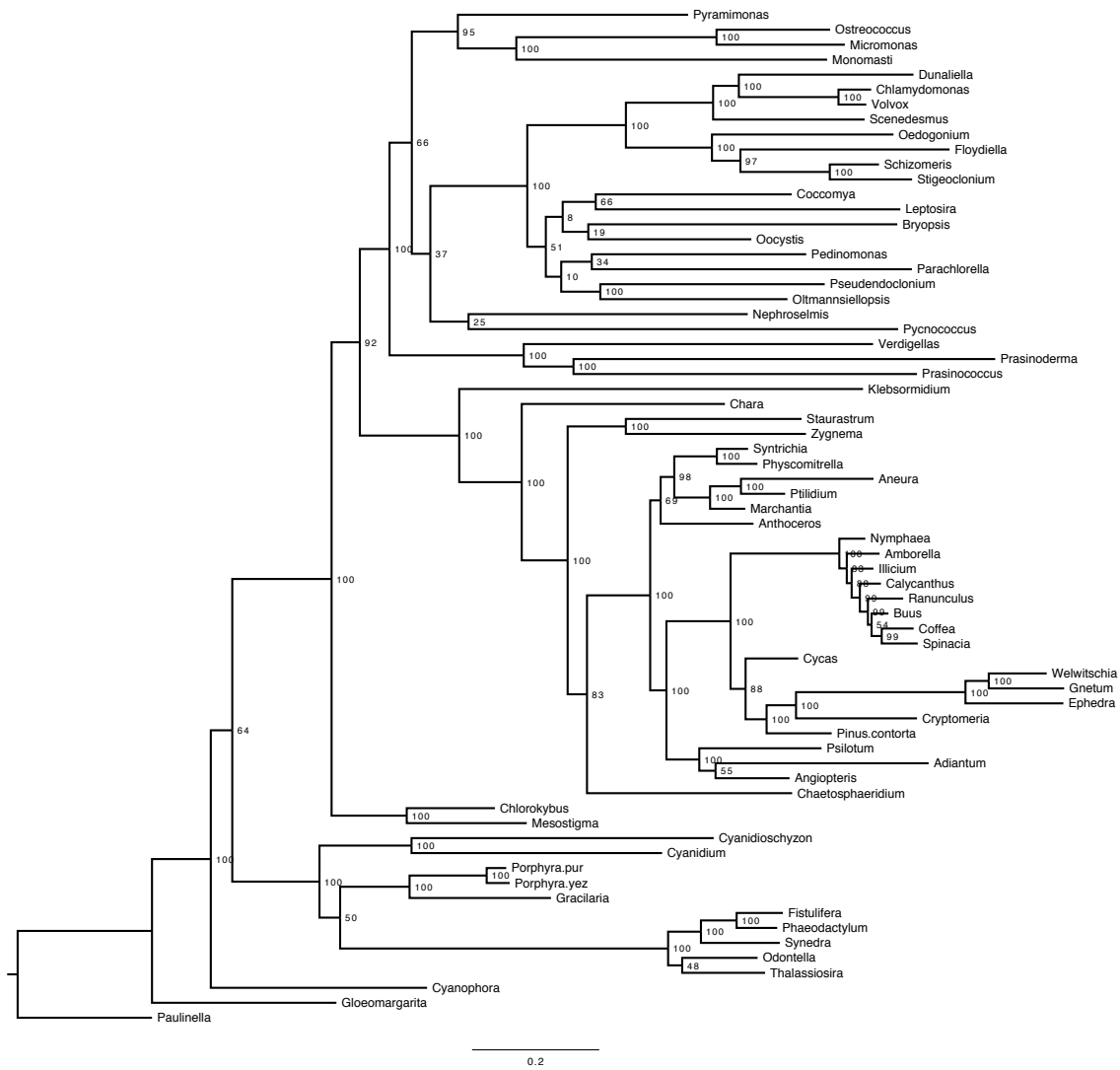
Supplementary Figure S3. Bayesian phylogeny of Cyanobacteria and photosynthetic eukaryotes. Numbers above branches indicates posterior probability. Branches with no numbers had 100% posterior probability. Convergence statistics: Number of Cycles: 82244; bpcomp maxdiff = 0.219156; Tracecomp effsize was always > 68 except for *statalpha* where effsize = 29; Tracecomp maxdiff was always < 0.53 except for *statalpha* that had a maxdiff = 0.93.



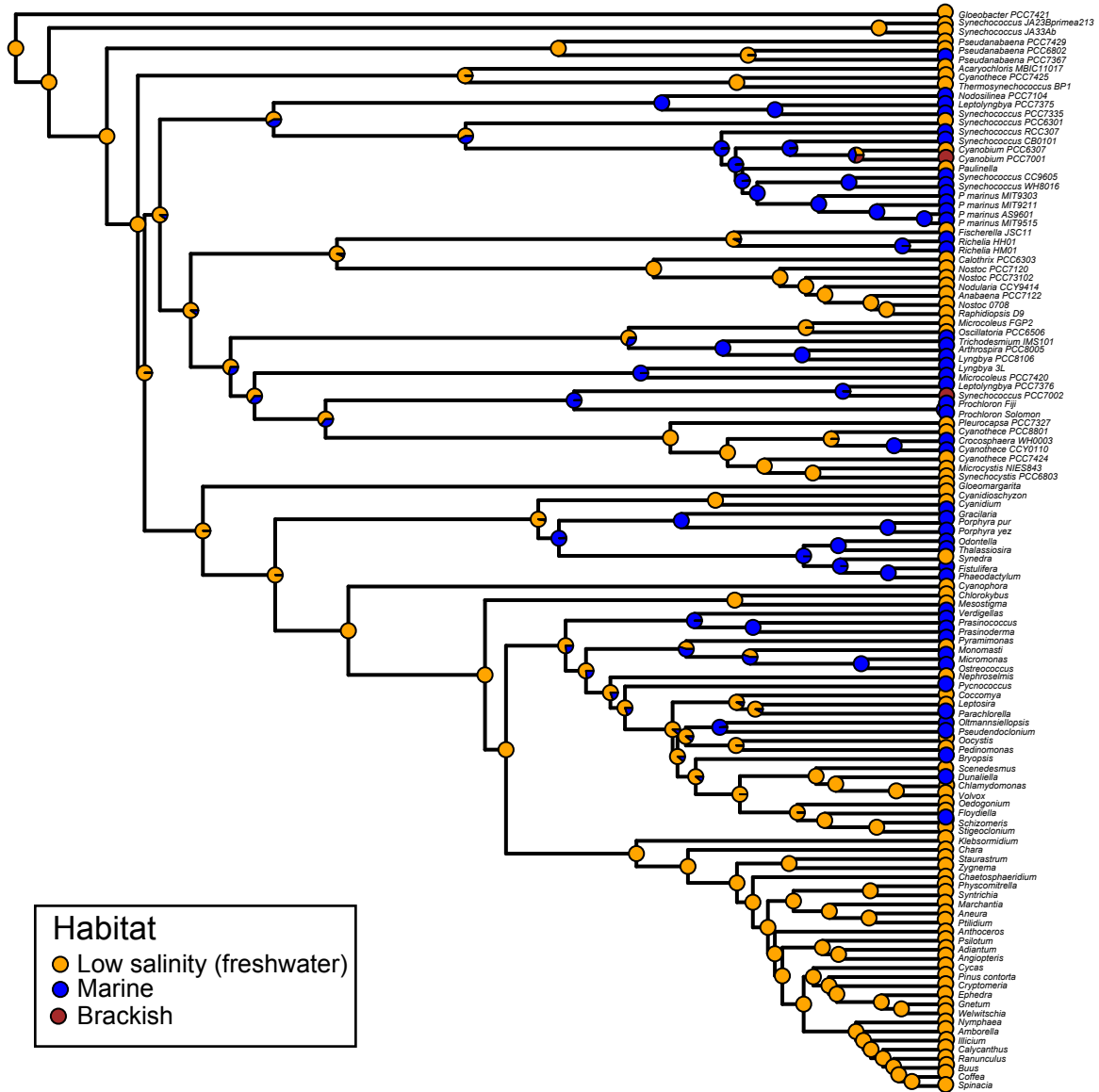
Supplementary Figure S4. Maximum Likelihood phylogeny including only eukaryotes.



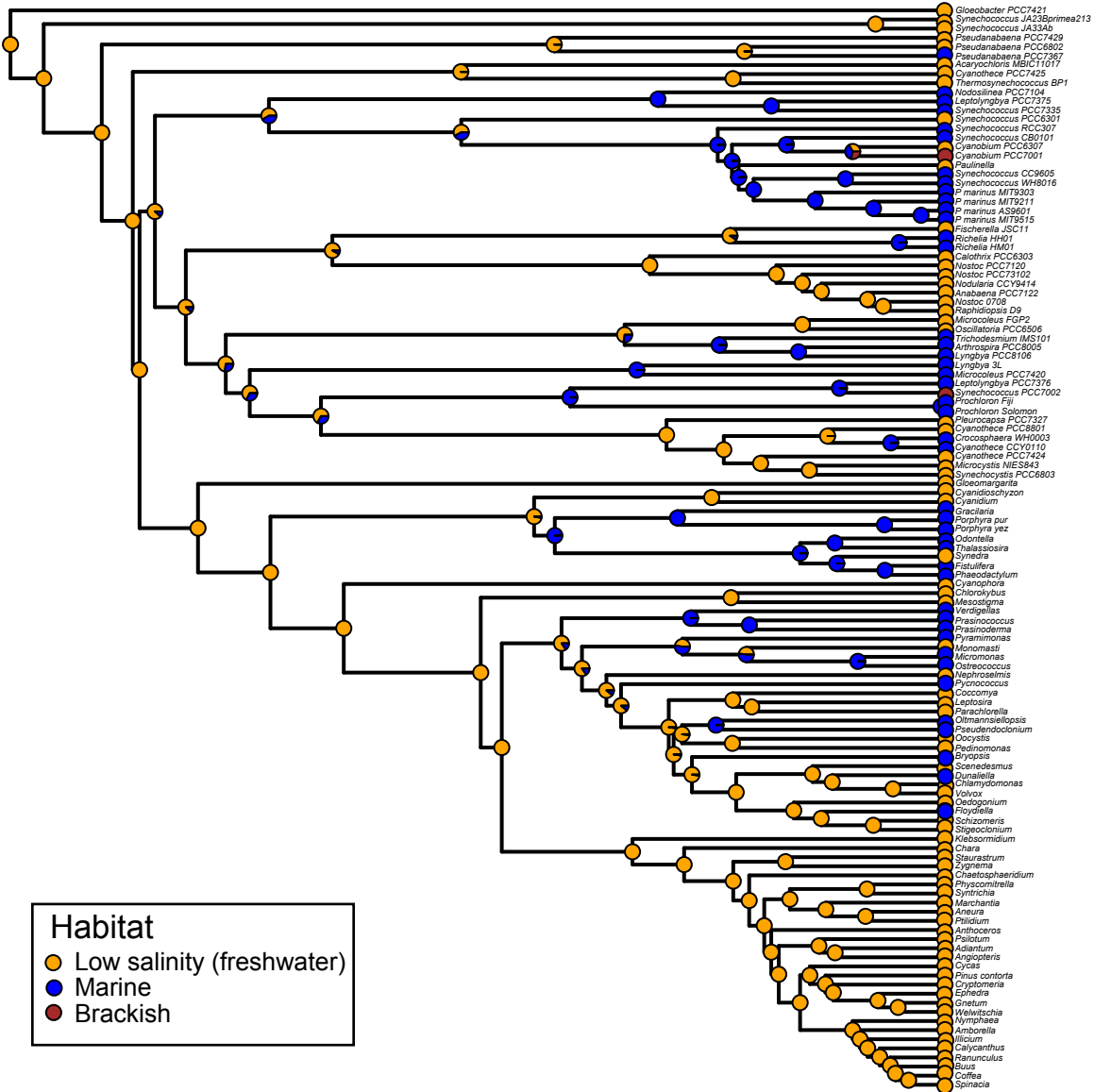
Supplementary Figure S5. Maximum Likelihood phylogeny including eukaryotes and *Gloeomargarita*.



Supplementary Figure S6. Stochastic mapping analyses of the evolution of habitat. Phylogenetic tree including 119 taxa was estimated in Phylobayes 1.7a (1). A pie chart at each node indicates the Likelihoods for three character states as follows: blue circle = marine; yellow circle = freshwater, and red = brackish. In the analyses shown here, *Parachlorella* were treated as marine and *Pedinomonas* as freshwater (Supplementary Table S2).



Supplementary Figure S7. Stochastic mapping analyses of the evolution of habitat. Phylogenetic tree including 119 taxa was estimated in Phylobayes 1.7a (1). A pie chart at each node indicates the Likelihoods for three character states as follows: blue circle = marine; yellow circle = freshwater, and red = brackish. In the analyses shown here *Parachlorella* and *Pedinomonas* were both treated as freshwater (Supplementary Table S2).



Supplementary Tables

Table S1. Gene composition of genomic dataset.

Gene	Description	COG category
CcsA	C-type cytochrome synthesis protein	CP; O
L14	Ribosomal protein L14	IP; J
L20	Ribosomal protein L20	IP; J
S11	Ribosomal protein S11	IP; J
S12	Ribosomal protein S12	IP; J
S19	Ribosomal protein S19	IP; J
S2	Ribosomal protein S2	IP; J
S3	Ribosomal protein S3	IP; J
S4	Ribosomal protein S4	IP; J
S7	Ribosomal protein S7	IP; J
RpoB	RNA polymerase beta and beta prime subunits	IP; K
RpoC1	RNA polymerase beta' subunit	IP; K
RpoC2	RNA polymerase beta' chain	IP; K
AtpA	ATP synthase CF1 subunit A	Met; C
AtpB	CF1 beta subunit of ATP synthase	Met; C
PetB	Cytochrome b6	Met; C
PetD	Cytochrome b6/f complex subunit IV	Met; C
PsaC	Photosystem I subunit VII	Met; C
RbcL	Ribulose 1,5-bisphosphate carboxylase, large subunit	Met; G
PsaA	Photosystem I p700 apoprotein A1	(Met); n.a.
PsbA	Photosystem II D1 protein	(Met); n.a.
PsbB	CP47 chlorophyll apoprotein of photosystem II	(Met); n.a.
PsbC	Photosystem II CP43 protein	(Met); n.a.
PsbD	Photosystem II protein D2	(Met); n.a.
PsbE	Cytochrome b559 alpha chain	(Met); n.a.
Ycf3	Photosystem I assembly protein	PC; R

COG functional categories (<http://www.ncbi.nlm.nih.gov/COG/>):

CP = Cellular processes

IP = Information storage and processing

Met = Metabolism

PC = Poorly Characterized

COG functional sub-categories:

C = Energy production and conversion

J = Translation, ribosomal structure and biogenesis

K = Transcription

O = Posttranslational modification, protein turnover, chaperones

R = General function prediction only

n.a. = not in COG database

Table S2. Character states for the evolution of habitat. Taxa included in this study. Abbreviated taxa names are listed for those analyses shown in Figs 2 and 3. Habitat was coded using multistate discrete character states as follows 0 = freshwater (0-0.5 ppt parts per thousand), 1 = marine (from 30 up to 50 ppt parts per thousand), and 2 = brackish (0.5 – 30 ppt parts per thousand). Character states for the genera *Parachlorella*, *Pedinomonas* were treated as marine for analyses shown in Fig. 4; most *Parachlorella* species are marine, and *Pedinomonas* have both marine and freshwater.

Taxon name	Abbreviated taxa name in Figs. 1 and 4.	Character state
<i>Acaryochloris marina</i> MBIC11017	<i>Acaryochloris</i> MBIC11017	1
<i>Adiantum capillus veneris</i>	<i>Adiantum</i>	0
<i>Amborella trichopoda</i>	<i>Amborella</i>	0
<i>Anabaena cylindrica</i> PCC7122	<i>Anabaena</i> PCC7122	0
<i>Aneura mirabilis</i>	<i>Aneura</i>	0
<i>Angiopteris evecta</i>	<i>Angiopteris</i>	0
<i>Anthoceros formosae</i>	<i>Anthoceros</i>	0
<i>Arthrospira</i> sp. PCC8005	<i>Arthrospira</i> PCC8005	1
<i>Bryopsis hypnoides</i>	<i>Bryopsis</i>	1
<i>Buus microphylla</i>	<i>Buus</i>	0
<i>Calothrix</i> sp. PCC6303	<i>Calothrix</i> PCC6303	0
<i>Calycanthus floridus</i>	<i>Calycanthus</i>	0
<i>Chaetosphaeridium globosum</i>	<i>Chaetosphaeridium</i>	0
<i>Chara vulgaris</i>	<i>Chara</i>	0
<i>Chlamydomonas reinhardtii</i>	<i>Chlamydomonas</i>	0
<i>Chlorokybus atmophyticus</i>	<i>Chlorokybus</i>	0
<i>Coccomyxa</i> sp. C169	<i>Coccomyxa</i>	0
<i>Coffea arabica</i>	<i>Coffea</i>	0
<i>Crocospaera watsonii</i> WH0003	<i>Crocospaera</i> WH0003	1
<i>Cryptomeria japonica</i>	<i>Cryptomeria</i>	0
<i>Cyanidioschyzon merolae</i> strain 10D	<i>Cyanidioschyzon</i>	0
<i>Cyanidium caldarium</i>	<i>Cyanidium</i>	0
<i>Cyanobium gracile</i> PCC6307	<i>Cyanobium</i> PCC6307	0
<i>Cyanobium</i> sp. PCC7001	<i>Cyanobium</i> PCC7001	2
<i>Cyanophora paradoxa</i>	<i>Cyanophora</i>	0
<i>Cyanothece</i> sp. CCY0110	<i>Cyanothece</i> CCY0110	1
<i>Cyanothece</i> sp. PCC7424	<i>Cyanothece</i> PCC7424	0
<i>Cyanothece</i> sp. PCC7425	<i>Cyanothece</i> PCC7425	0
<i>Cyanothece</i> sp. PCC8801	<i>Cyanothece</i> PCC8801	0
<i>Cycas taitungensis</i>	<i>Cycas</i>	0
<i>Dunaliella salina</i>	<i>Dunaliella</i>	1
<i>Ephedra equisetina</i>	<i>Ephedra</i>	0
<i>Fischerella</i> sp. JSC11	<i>Fischerella</i> JSC11	0
<i>Fistulifera</i> sp. JPCC DA0580	<i>Fistulifera</i>	1
<i>Floydiella terrestris</i>	<i>Floydiella</i>	1

<i>Gloeobacter violaceus</i> PCC7421	<i>Gloeobacter</i> PCC7421	0 ⁺
<i>Gloeomargarita lithophora</i>	<i>Gloeomargarita</i>	0
<i>Gnetum parvifolium</i>	<i>Gnetum</i>	0
<i>Gracilaria tenuistipitata</i>	<i>Gracilaria</i>	1
<i>Illicium oligandrum</i>	<i>Illicium</i>	0
<i>Klebsormidium flaccidum</i>	<i>Klebsormidium</i>	0
<i>Leptolyngbya</i> sp. PCC7375	<i>Leptolyngbya</i> PCC7375	1
<i>Leptolyngbya</i> sp. PCC7376	<i>Leptolyngbya</i> PCC7376	1
<i>Leptosira terrestris</i>	<i>Leptosira</i>	0
<i>Lyngbya majuscula</i> 3L	<i>Lyngbya</i> 3L	1
<i>Lyngbya</i> sp. PCC8106	<i>Lyngbya</i> PCC8106	1
<i>Marchantia polymorpha</i>	<i>Marchantia</i>	0
<i>Mesostigma viride</i>	<i>Mesostigma</i>	0
<i>Microcoleus vaginatus</i> FGP2	<i>Microcoleus</i> FGP2	0
<i>Microcoleus chthonoplastes</i> PCC7420	<i>Microcoleus</i> PCC7420	1
<i>Microcystis aeruginosa</i> NIES843	<i>Microcystis</i> NIES843	0
<i>Micromonas</i> sp. RCC299	<i>Micromonas</i>	1
<i>Monomasti</i> sp. OKE1	<i>Monomasti</i>	0
<i>Nephroselmis olivacea</i>	<i>Nephroselmis</i>	0
<i>Nodosilinea nodulosa</i> PCC7104	<i>Nodosilinea</i> PCC7104	1
<i>Nodularia spumigena</i> CCY9414	<i>Nodularia</i> CCY9414	0
<i>Nostoc azollae</i> 0708	<i>Nostoc</i> 0708	0
<i>Nostoc</i> sp. PCC7120	<i>Nostoc</i> PCC7120	0
<i>Nostoc punctiforme</i> PCC73102	<i>Nostoc</i> PCC73102	0
<i>Nymphaea alba</i>	<i>Nymphaea</i>	0
<i>Odontella sinensis</i>	<i>Odontella</i>	1
<i>Oedogonium cardiacum</i>	<i>Oedogonium</i>	0
<i>Oltmannsiellopsis viridis</i>	<i>Oltmannsiellopsis</i>	1
<i>Oocystis solitaria</i> SAG83 80	<i>Oocystis</i>	0
<i>Oscillatoria</i> sp. PCC6506	<i>Oscillatoria</i> PCC6506	0
<i>Ostreococcus tauri</i>	<i>Ostreococcus</i>	1
<i>Parachlorella kessleri</i>	<i>Parachlorella</i>	1 (0*)
<i>Paulinella chromatophora</i>	<i>Paulinella</i>	0
<i>Pedinomonas minor</i>	<i>Pedinomonas</i>	1 (0*)
<i>Phaeodactylum tricornutum</i>	<i>Phaeodactylum</i>	1
<i>Pinus contorta</i>	<i>Pinus contorta</i>	0
<i>Pleurocapsa</i> sp. PCC7327	<i>Pleurocapsa</i> PCC7327	0
<i>Physcomitrella patens</i> subsp <i>patens</i>	<i>Physcomitrella</i>	0
<i>Prasinococcus</i> sp. CCMP1194	<i>Prasinococcus</i>	1
<i>Prasinoderma coloniale</i>	<i>Prasinoderma</i>	1
<i>Porphyra purpurea</i>	<i>Porphyra pur</i>	1
<i>Porphyra yezoensis</i>	<i>Porphyra yez</i>	1
<i>Prochlorococcus marinus</i> str. AS9601	<i>P marinus</i> AS9601	1
<i>Prochlorococcus marinus</i> str. MIT9211	<i>P marinus</i> MIT9211	1

<i>Prochlorococcus marinus</i> str. MIT9303	<i>P marinus</i> MIT9303	1
<i>Prochlorococcus marinus</i> str. MIT9515	<i>P marinus</i> MIT9515	1
<i>Prochloron didemni</i> P2 Fiji	<i>Prochloron</i> Fiji	1
<i>Prochloron didemni</i> P3 Solomon	<i>Prochloron</i> Solomon	1
<i>Pseudanabaena</i> sp. PCC6802	<i>Pseudanabaena</i> PCC6802	0
<i>Pseudanabaena</i> sp. PCC7367	<i>Pseudanabaena</i> PCC7367	0
<i>Pseudanabaena biceps</i> PCC7429	<i>Pseudanabaena</i> PCC7429	1
<i>Pseudendoclonium akinetum</i>	<i>Pseudendoclonium</i>	1
<i>Psilotum nudum</i>	<i>Psilotum</i>	0
<i>Ptilidium pulcherrimum</i>	<i>Ptilidium</i>	0
<i>Pycnococcus provasolii</i>	<i>Pycnococcus</i>	1
<i>Pyramimonas parkeae</i>	<i>Pyramimonas</i>	1
<i>Ranunculus macranthus</i>	<i>Ranunculus</i>	0
<i>Raphidiopsis brookii</i> D9	<i>Raphidiopsis</i> D9	0
<i>Richelia intracellularis</i> HH01	<i>Richelia</i> HH01	0
<i>Richelia intracellularis</i> HM01	<i>Richelia</i> HM01	0
<i>Scenedesmus obliquus</i>	<i>Scenedesmus</i>	0
<i>Schizomeris leibleinii</i>	<i>Schizomeris</i>	1
<i>Spinacia oleracea</i>	<i>Spinacia</i>	1
<i>Staurastrum punctulatum</i>	<i>Staurastrum</i>	0
<i>Stigeoclonium helveticum</i>	<i>Stigeoclonium</i>	0
<i>Synechococcus</i> sp. CB0101	<i>Synechococcus</i> CB0101	0
<i>Synechococcus</i> sp. CC9605	<i>Synechococcus</i> CC9605	1
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	<i>Synechococcus</i> JA23Bprimea213	1
<i>Synechococcus</i> sp. JA-3-3Ab	<i>Synechococcus</i> JA33Ab	1
<i>Synechococcus elongatus</i> PCC6301	<i>Synechococcus</i> PCC6301	1
<i>Synechococcus</i> sp. PCC7002	<i>Synechococcus</i> PCC7002	2
<i>Synechococcus</i> sp. PCC7335	<i>Synechococcus</i> PCC7335	0
<i>Synechococcus</i> sp. RCC307	<i>Synechococcus</i> RCC307	0
<i>Synechococcus</i> sp. WH8016	<i>Synechococcus</i> WH8016	1
<i>Synechocystis</i> sp. PCC6803	<i>Synechocystis</i> PCC6803	0
<i>Synedra acus</i>	<i>Synedra</i>	0
<i>Syntrichia ruralis</i>	<i>Syntrichia</i>	1
<i>Thalassiosira oceanica</i> CCMP1005	<i>Thalassiosira</i>	0
<i>Thermosynechococcus elongatus</i> BP1	<i>Thermosynechococcus</i> BP1	0
<i>Trichodesmium erythraeum</i> IMS101	<i>Trichodesmium</i> IMS101	0
<i>Verdigellas peltata</i>	<i>Verdigellas</i>	1
<i>Volvox carteri nagariensis</i> UTE2908	<i>Volvox</i>	0
<i>Welwitschia mirabilis</i>	<i>Welwitschia</i>	0
<i>Zygnema circumcarinatum</i>	<i>Zygnema</i>	0

+ An epilithic or endolithic freshwater cyanobacterium

* A freshwater species

Table S3. Posterior age estimates in Myr using a Bayesian approach.

Node	Clade	UGAM	CIR	Log normal
1	<i>Gloeomargarita</i> + Archaeplastida	2,084 (2,278–1,906)	2,166 (2,334–2,015)	2,159 (2,360–1,998)
2	Archaeplastida	1,828 (2,028–1,656)	1,813 (1,962–1,642)	1,731 (1,894–1,575)
3	Cyanophora	1,828 (2,028–1,656)	1,813 (1,962–1,642)	1,731 (1,894–1,575)
4	Rhodophyta	1,132 (1,335–973)	993 (1,091–864)	968 (1,061–824)
5	Chlorophyta	1,091 (1,348–897)	1087 (1,295–914)	1,001 (1,165–838)
6	Streptophyta	890 (1,209–651)	982 (1,177–832)	939 (1,102–802)
7	Palmophyllophyceae	890 (1,008–404)	878 (1,088–714)	797 (990–651)
8	Prasinophytes	673 (1,008–459)	900 (1,080–744)	822 (1,014–679)
9	<i>Paulinella</i> + Marine <i>SynPro</i>	731 (86–415)	419 (575–298)	364 (506–247)

Node ID corresponds to those shown in Fig. 1. Age estimates are given for analyses under UGAM, CIR and log normal clock models for the topology generated by Maximum Likelihood. The CAT-GTR replacement model was implemented and the root was set with a maximum age of 2.7 Bya (2) and a minimum age of 2.32 Bya (3).

Table S4. Posterior age estimates in Myr using a Bayesian approach.

Node	Clade	UGAM	CIR	Log normal
1	<i>Gloeomargarita</i> + Archaeplastida	2,164 (2,396–1,940)	2,195 (2,390–2,022)	2,169 (2,351–1,990)
2	Archaeplastida	1,947 (2,180–1,723)	1,969 (2,157–1,797)	1,946 (2,118–1,779)
3	Cyanophora	1,737 (1,986–1,455)	1,808 (1,984–1,656)	1,781 (1,940–1,624)
4	Rhodophyta	1,172 (1,362–920)	1,063 (1,179–944)	1,065 (1,202–949)
5	Chlorophyta	1,111 (1,362–920)	1,065 (1,231–943)	1,042 (1,179–904)
6	Streptophyta	743 (1,202–664)	1,005 (1,151–881)	985 (1,123–851)
7	Palmophyllophyceae	702. (1,053–464)	883 (1,065–739)	841 (1,004–670)
8	Prasinophytes	755 (1,034–491)	883 (1,044–744)	839 (984–688)
9	<i>Paulinella</i> + Marine <i>SynPro</i>	654 (928–449)	413 (576–310)	339 (457–228)

Node ID corresponds to those shown in Fig. 1. Age estimates are given for analyses under UGAM, CIR and log normal clock models for the topology generated by Maximum Likelihood. The CAT-GTR replacement model was implemented and the root was set with a maximum age of 3 Bya (2) and a minimum age of 2.32 Bya (3).

References

1. Lartillot N, Rodrigue N, Stubbs D, & Richer J (2013) PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Syst Biol* 62(4):611-615.
2. Crowe SA, *et al.* (2013) Atmospheric oxygenation three billion years ago. *Nature* 501(7468):535-+.
3. Bekker A, *et al.* (2004) Dating the rise of atmospheric oxygen. *Nature* 427(6970):117-120.