A potential cyanobacterial ancestor of Viridiplantae chloroplasts

Wriddhiman Ghosh*, Prabir Kumar Haldar, Sabyasachi Bhattacharya,
Jaideb Chatterjee, Prosenjit Pyne and Masrure Alam

Department of Microbiology, Bose Institute,
P-1/12 CIT Scheme VIIM, Kolkata - 700054, India.

* **Correspondence:** Email: wriman@bic.boseinst.ernet.in

Phone: 91-033-25693246
Fax: 91-33-23553886

Key Words : *Cyanobacteria* / chloroplast evolution / endosymbiosis

The theory envisaging the origin of plastids from endosymbiotic cyanobacteria is well-established but it is difficult to explain the evolution (spread) of plastids in phylogenetically diverse plant groups. It is widely believed that primordial endosymbiosis occurred in the last common ancestor of all algae[1], which then diverged into the three primary photosynthetic eukaryotic lineages, viz. the Rhodophyta (red algae), Glaucocystophyta (cyanelle-containing algae) and Viridiplantae (green algae plus all land plants)[2]. Members of these three groups invariably have double membrane-bound plastids[3], a property that endorses the primary endosymbiotic origin of the organelles. On the hand, the three or four membrane-bound plastids of the evolutionary complicated Chromalveolates [chromista (cryptophytes, haptophytes, and stramenopiles) and alveolata (dinoflagellates, apicomplexans, and ciliates)] are inexplicable in the light of a single endosymbiosis event, thereby necessitating the postulation of the secondary[4,5] and tertiary[6] endosymbiosis theories where a nonphotosynthetic protist supposedly engulfed a red or a green alga[7] and an alga containing a secondary plastid itself was engulfed[8] respectively. In the current state of understanding, however, there is no clue about the taxonomic identity of the cyanobacterial ancestor of chloroplasts, even though there is a wide consensus on a single primordial endosymbiosis event. During our metagenomic investigation of a photosynthetic geothermal microbial mat community we discovered a novel order-level lineage of *Cyanobacteria* that - in 16S rRNA gene sequence-based phylogeny - forms a robust monophyletic clade with chloroplast-derived sequences from diverse divisions of Viridiplantae. This cluster diverged deeply from the other major clade encompassing all hitherto known groups of *Cyanobacteria* plus the chloroplasts of Rhodophyta, Glaucocystophyceae and Chromalveolates. Since this fundamental dichotomy preceded the origin of all chloroplasts, it appears that two early-diverging cyanobacterial lineages had possibly given rise to two discrete chloroplast descents via two separate engulfment events.

During our geomicrobiological exploration of the Tapoban-Vishnugad hot spring area in the Garhwal Himalayas (India) we found lush green photosynthetic mats (GPM) growing all over a colossal sinter of geothermal minerals deposited around a cluster of hot water vents (Fig. 1A). The GPM flourished mostly on the hot water outflows (pH 7.0-7.5, temperature ~60 ^{O}C) and often got petrified under the load of precipitating minerals (Fig. 1B). Live mat samples from four discrete zones of the sinter were collected in sterile polyethylene bottles and immediately transported to the laboratory in insulated coolers.

DNA sequence-based metagenomic methods were used to identify the microbial groups, particularly the photosynthetic elements, constituting the mats. Community genomic DNA was extracted from the four samples and used as templates to PCR-amplify small subunit (*ssu*) (16S for *Bacteria* and *Archaea*, and 18S for *Eukarya*) ribosomal RNA (rRNA) genes using universal primers that separately target members of the three domains of life (Supplementary Table 1). Subsequently, three *Cyanobacteria*-specific 16S rRNA gene (rDNA) primers were also used. Microbial species were recognized on the basis of nucleotide sequence identities of the PCR amplicons. Phylogenetic species-level relatedness groups (phylotypes) were determined on the basis of ≥97% *ssu* rDNA sequence similarity[9], while nearest taxonomic affiliations of the obtained phylotypes were determined by BLAST search analysis. No PCR product was obtained using *Archaea*- or *Eukarya*-specific primers from any of the four samples. The amplified rDNA pools obtained with *Bacteria*- and *Cyanobacteria*-specific primers were cloned and assorted randomly into libraries of 1000 clones. Thus we had four 1000-clone libraries for the four samples, i.e., 4000 clones in all. We sequenced them and found that all the samples had identical and relatively simple community structure encompassing phylotypes affiliated to *Cyanobacteria*, α-, β- and γ-*proteobacteria*, *Deinococcus*-*Thermus*, *Actinobacteria* and *Firmicutes*.

All the sequenced cyanobacterial clones were found to have 99-100% similarities among them, thereby representing one coherent phylotype. As such, only a few representative sequences were deposited to the public database (EMBL accession numbers HE578050-HE578056). Interestingly enough, 16S rDNA sequences derived from chloroplasts of Dicotyledonous Angiosperms exhibited ~99% similarities with these clones, whereas homologs from all **h**itherto **k**nown **g**roups of ***C**yanobacteria* (**HKGC**) were only ≤89% similar (Fig. 2). Between these extremities, 16S rDNAs of chloroplasts of Monocotyledonous Angiosperms, Gymnosperms, Pteridophytes, Bryophytes, Streptophyta and Chlorophyta respectively showed 97%, 97-94%, ≤96%, ≤97%, 92-91%, ≤90% sequence similarities with the new phylotype. Again, among the HKGC, members of *Oscillatoriales*, *Nostocales* and *Pleurocapsales* had maximum (87-89%, or even lower) 16S rDNA sequence similarities with the new phylotype, whereas other groups like *Gloeobacterales*, *Chroococcales*, *Stigonematales* and *Prochlorophytes* had 85-87%, or lower, similarity levels. 16S rDNA from chloroplasts of **G**laucocystophyceaea, **R**hodophyta and **C**hromalveolates (here onwards referred to as **GRC**) respectively had ≤89%, ≤87% and ≤88 % sequence similarities with the homologs from the new phylotype (Fig. 2).

On the basis of its unequivocally low similarities with, and equivalent distances from, all HKGC, we considered the new phylotype to be a novel order of *Cyanobacteria*, and for that we proposed the name *Candidatus* Thermofiliformales (CT), meaning thermotolerant filamentous forms. Notably, several uncultured cyanobacterial 16S rDNA clones in the database, isolated from a wide range of habitats, were also found to have ~99% sequence similarities with the new clones (Fig. 2). This indicated a wide ecological amplitude and global distribution of this lineage, even though it does not yet have a cultured member in its rank. It is worth-mentioning at this point that all our attempts to culture cyanobacteria from the GPM samples in different variants of BG 11 medium[10] have so far been fruitless.

While molecular signature of no photosynthetic microorganism other than cyanobacteria was identified in the metagenomic analyses, laser-scanning confocal microscopy (LSCM), utilized to selectively illuminate the photosynthetic microorganisms [by imaging chlorophyll (Chl) and phycobilin autofluorescence], exclusively revealed long filamentous organisms resembling cyanobacteria (Fig. 1, G-I). Similar morphotypes were also recognized in scanning electron microscopy (SEM) and phase contrast microscopy of the GPM (Fig. 1, B-C). In order to check whether the metagenomically identified cyanobacterial phylotype corresponded to these filamentous microorganisms, fluorescent in *situ* hybridization was carried out with the biotin-labeled oligonucleotide probe 5′-GTTTAGTTGCCACCGTTGAGTTTGGAACCCTGAAC -3′ (spanning between nucleotide positions 1115 and 1149 with reference to the *E. coli* 16S rRNA gene), which is specific for the new phylotype. This probe was hybridized to GPM specimens and then detected *in situ* using streptavidin-horse raddish peroxidase (HRP) conjugate in conjunction with fluorescein-tyramide signal amplification reagent[11]. This particular dye-tyramide system was chosen because the fluorescence excitation/emission maxima (495 nm/519 nm) of fluorescein is outside the observed range of autofluorescence of the long filamentous microorganisms suspected to represent the cyanobacterial phylotype (see LSCM data below). Even though these putative targets fluoresced on excitation at 488 nm and 514 nm, none of the corresponding emissions were below 545 nm. As such, potential interference of this green autofluorescence with fluorescein emissions could be abolished completely by imaging the latter at band pass (BP) 505-530 nm. LSCM observations for the DAPI counterstain (excitation at 360 nm and detection at BP 420-480 nm), showed that the GPM (Fig. 1J) was a mesh of diverse bacterial morphotypes including the filamentous forms (Fig. 1K). On the other hand, in the same microscopic field, only the long filamentous microorganisms fluoresced when excitation (for fluorescein) was done at 488 nm and detection set at BP 505-530 nm (Fig. 1L). Images comparable to those obtained for DAPI fluorescence were also obtained when mat fragments were analyzed in the presence of the biotin-labeled *Bacteria*-specific universal primer EUB338[12] (data not shown). While the CT-specific 35-mer oligonucleotide primer did not hybridize with standard bacterial cultures (data not shown), a blank having the CT-specific probe but no-tyramide signal amplification reagent resulted in fluorescence of the filamentous cells with intensities equivalent to their autofluorescence observed in LSCM with 488 nm excitation and emission detection at long pass (LP) 560 nm or LP 615 nm (Fig. 1H & I). In this set, however, no image was obtained below 545 nm emission detection.

Phylogenetic relationship of the new order with HKGC and chloroplasts from major plant groups was reconstructed using distance matrix (DM)[13], bootstrap DNA parsimony (DP)[13,14] and maximum likelihood (ML)[14] methods. Identical branching patterns (tree topologies) (Fig. 2) obtained from all the three analyses demonstrated the close affinity (or shared ancestry) between the CT lineage and Viridiplantae chloroplasts (VC), which together formed robust super-clades supported by high boot-strap values in all the constructed trees. The three analyses also unequivocally illustrated the deep divergence of this CT-VC super-clade from the one encompassing all HKGC and chloroplasts from GRC. This deep divergence clearly anteceded the deepest-branching (most ancient) chloroplasts within the GRC as well as the Viridiplantae radiations. This implies that two distinct cyanobacterial lineages, which had already diverged prior to the origin of any chloroplast, had given rise to two distinct chloroplast lineages. As a corollary, a potential evolutionary scenario (Fig. 3) emerges where two subpopulations (or, may be, diverged lineages) of the plastid-less last common ancestor (LCA) of all plants (which, in effect, were the LCAs of GRC and Viridiplantae respectively) engulfed the ancient ancestor of all HKGC and the *Candidatus* Thermofiliformales respectively, via two endosymbiotic events separated in time and/or space. It is however worth pointing out that the evolutionarily recent *Candidatus* Thermofiliformales is still the only cyanobacterial member of the CT-VC super-clade; but it is not unlikely that continued addition of information on cyanobacterial diversity shall reveal the more deeply-branching elements of this lineage sooner than later.

Prochlorophytes are the only cyanobacterial affiliates that, like Viridiplantae, have Chl*b* and lack phycobiliproteins as light-harvesting (LH) pigments[15]. However, the origin of Chl*b* in Viridiplantae has remained an intriguing question ever since the polyphyletic Prochlorophytes were proved to be distantly related to the chloroplasts[15,16]. In view of the novel order's shared ancestry with Viridiplantae chloroplasts we considered it imperative to check the presence of Chl*b* in these organisms. The total photosynthetic pigment-content of the GPM was extracted in 7:2 v/v acetone:methanol, separated by reverse phase high-performance liquid chromatography (HPLC) with reverse phase C18 column, and absorption maxima of the eluates analyzed by an online photodiode array detector to obtain UV-visible absorption spectra[17]. Compounds were identified by comparing their HPLC retention times plus absorption maxima with standard values[17-19]. A 39 min major eluate showed absorption spectrum characteristic of Chl *a* (maxima at 430 nm, 615 nm and 663 nm) (Fig. 1D), while directly measured absorption spectrum of a 96% ethanol extract of the GPM also showed these maxima. Notably however, none of the Chl*b*-specific absorbance maxima (viz., 462 nm, 599 nm and 648 nm)[17] were detected in either analysis. 4 ^{O}C excitation spectra of the 96% ethanol extract (recorded at 680 nm emission) also showed maxima characteristic of Chl*a*, but not Chl*b*[20,21] (Fig. 1E). An intense 672 nm fluorescence emission peak was obtained by exciting the extracts at 430 nm (the peak of the excitation spectrum of Chl*a*) (Fig. 1F); while at 470 nm (where excitation of Chl*a* is minimum but that of Chl*b* is highest) fluorescence yield was ~30 times lower (data not shown). To verify whether these spectral data were attributable solely to the cyanobacterial components, we imaged the GPM by LSCM at matching excitation and detection wavelengths. As such, nothing other than filamentous microorganisms were visible on excitation at 405 nm [emissions detectable at BP 505-545 nm and LP 615 nm] (Fig. 1H) and 633 nm (detectable at LP 650 nm) (Fig. 1I). As expected from spectroscopic data, excitation at 458 nm (with emissions detected at LP 505 nm or LP 615 nm) did not produce any image, even as longer-wavelength emissions (detectable at BP 545-600 nm and LP 615 nm, but not below 545 nm) could be imaged upon exciting the samples at 488 nm, 514 nm (plausibly due to the excitation of carotenoids followed by energy transfer to Chl*a*[22,23]) and 543 nm (attributable to phycoerythrin autofluorescence[24]) (data not shown).

It has been postulated earlier that Chl*b* and its associated structural proteins either originated several times independently (in prochlorophytes and the ancestor of green chloroplasts) and evolved convergently or the primordial green chloroplast acquired Chl*b* from prochlorophyte(s) via horizontal transfer(s)[15,16]. Our observations offered no reason to believe that the remote ancestor of *Candidatus* Thermofiliformales, though closely related to the progenitor of green plant chloroplasts, was the source of Chl*b* in evolution.

METHODS SUMMARY

Molecular community analysis

To isolate community DNA[25], mat sample [~3 g (wet weight) in ethanol or formaldehyde] was centrifuged, washed with 10:1 mM TE buffer (pH 7.8), and resuspended in sucrose lysis buffer

[10% sucrose, 0.7 M NaCl, 40 mM EDTA, 50 mM Tris-HCl (pH 8.5)] to a total volume of 10 ml. Lysozyme and achromopeptidase (1 and 50 mg ml^{-1} respectively) were then added to this mixture, which was then subjected to three cycles of disruption in a French pressure cell at 20,000 lb in^{-2}. The disrupted sample was treated with proteinase K (50 mg ml^{-1}) at 55 ^{O}C for 30 min and then with 1% SDS at 55 ^{O}C for 60 min. The slurry was further treated with 1% hexadecyltrimethyl ammonium bromide at 55 ^{O}C for 30 min to remove polysaccharides and residual proteins as precipitates. The digested sample was then treated with phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v), mixed gently on a shaker, and then centrifuged. The crude DNA in the resulting aqueous layer was obtained by ethanol precipitation and centrifugation and subjected to a standard purification procedure consisting of RNase digestion, chloroform-isoamyl alcohol (24:1) treatment, and further ethanol precipitation. DNAs were further purified, desalted, and concentrated with a Centricon-100 concentrator (Millipore). Integrity of all DNA samples was checked by agarose gel electrophoresis. The same were quantified by measuring absorbance at 260 nm and diluted with appropriate volumes of 10:0.1 TE (pH 7.8) for all downstream experiments. About 10 ng of mat DNA was used as template in PCR with small subunit (*ssu*) ribosomal RNA (rRNA) gene-specific primers listed in Supplementary Table 1. Each type of PCR product obtained from each of the seven studied samples was cloned separately in TOPO TA PCR cloning vectors (Invitrogen, Life Technologies) to give rise to distinct environmental clone libraries. Plasmid extraction, purification, and insert checking from the clones were performed as earlier[26]. Finally *ssu* rRNA gene sequences were determined from the authentic recombinant plasmids using the same primers that amplified the respective inserts or amplified *ssu* rRNA gene fragments. Phylogenetic species-level relatedness groups (or phylotypes) were determined on the basis of

≥97% *ssu* rRNA gene sequence similarity [9], while the nearest taxonomic affiliations of these groups were determined by comparing their representative nucleotide sequences with those available in the public databases using BLAST search analysis.

Phylogeny Reconstruction

Multiple alignment of sequences was done by ClustalX2[27]. Evolutionary distances (expressed in estimated numbers of changes per 100 nucleotides) were calculated by pairwise comparison of the aligned sequences by the DNADIST program. Consensus Neighbor-Joining (NJ) trees[28,29] were constructed following the majority rule and strict consensus out of 100 phylogenetic trees produced using the program NEIGHBOR in PHYLIP version 3.69[13]. Bootstrap values (100 replicates) were calculated to validate the reproducibility of the branching pattern. Phylogenies based on maximum likelihood and parsimony methods were reconstructed using MEGA5[14]. The best substitution model used for likelihood analysis (general time reversible and gamma) was selected by Bayesian as well as corrected Akaike information criteria. After the starting tree was obtained automatically by applying NJ algorithm[28,29], heuristic searches for likelihood and parsimony were performed by using the Nearest-Neighbor-Interchange as well as Close-Neighbor-Interchange branch swapping algorithms.

Microscopy

Presence and localization of cyanobacteria within the live mat samples was elucidated by phase contrast, scanning electron and laser-scanning confocal microscopy. Suitably teased portions of the mats were spotted on agar-coated microscope slides, coverslips applied and gentle pressure given to minimize movement of cells. Autofluorescence of the photosynthetic components of the mats were then imaged in a Zeiss LSM 510 Meta Confocal Microscope. For SEM, portions of the mat samples were teased with 50 mM phosphate buffer (pH 7.0), vacuum dried, and observed under an FEI Quanta 200 scanning electron microscope. Samples were treated with heavy metals like osmium or lead to give contrast to biological materials against crystals of geothermal minerals intertwined with the biofilms.

Fluorescent *in situ* hybridization

Four small fragments of all the discrete GPM samples were resuspended in phosphate-buffered saline, suitably teased and mounted onto epoxycoated 4-well slides (5 mm) under aseptic conditions. When the slides were dry the materials were fixed and dehydrated using a graded ethanol-water series (3 min treatment at each grade) of 50, 80 and 100% (v/v), each of which contained 0.5% (v/v) hydrogen peroxide to quench endogenous peroxidases. Slides were dried again, following which 20 µl of a 20 mg ml^{-1} lysozyme solution in 100 mM:50 mM Tris:EDTA buffer (pH 8.0) was pipetted onto the sample wells. The slides were incubated at 37 ^{O}C for 60–90 min, rinsed in sterile Milli-Q water, dehydrated and allowed to dry.

Relevant biotin-labeled oligonucleotide primers or probes resuspended in hybridization buffer [20 mM Tris-HCl, pH 7.5; 20 mM NaCl; 0.01 % (w/v) SDS dissolved in 30% (v/v) formamide] were added to respective wells at concentrations of 100 ng ml^{-1}. Besides the biotin-labeled CT-specific 35-mer oligonucleotide primer, each mat fragment was also analyzed in the presence of the biotin-labeled *Bacteria*-specific universal primer EUB338[12] (positive control) and a blank having the CT-specific primer but no-Tyramide signal amplification reagent (to assess background fluorescence of the target cyanobacterial cells). Slides were incubated at 37

^{O}C for 120 min. After incubation, slides were washed thrice in wash buffer [56 mM sodium chloride, 5 mM EDTA, 0.01% (v/v) SDS in 20 mM Tris-HCl pH 7.5] at 37 ^{O}C for 5 min. Washed slides were transferred to a Coplin jar containing 2X SSC having pH 7.0 (3M NaCl plus 0.3M sodium citrate gives 20X SSC concentration). They were then blocked with 100 µl of 1% blocking reagent (supplied with the kit) in phosphate-buffered saline (PBS), covered with plastic coverslips and incubated in a humidified box at room temperature for 30 minutes.

Staining and detection was done using a Tyramide Signal Amplification Kit (Molecular Probes, Life Technologies) according to the manufacturer's instructions. A working solution of streptavidin–HRP was prepared by diluting the stock solution (prepared by reconstituting the kit material in 200 µl of PBS) 1:100 in the blocking reagent, following which 100 µl of the same was added to the slide by shortly removing the plastic coverslip. The slides were maintained in the humidified box for another 30 minutes at room temperature. The coverslip was then removed and the slide washed thrice by immersing in fresh PBS wash buffer at 37°C for 5 minutes on each occasion. A working solution of fluorescein (Alexa Fluor 488)-labelled tyramide was prepared by diluting the tyramide stock solution (ingredients supplied with the kit) 1:100 in amplification buffer/0.0015% H_2O_2. The latter was prepared by adding 1 µl of 30% hydrogen peroxide (supplied with the kit) to 200 µl of a supplied amplification buffer (containing thimerosal at 0.02%) and then adding 1 µl of this intermediate dilution (0.15% H_2O_2) to a further 100 µl of the amplification buffer. 100 µl of the tyramide working solution was added to the specimen, covered with a plastic coverslip, and incubated for 10 minutes at room temperature. The coverslip was again removed and the slide washed thrice by immersing in fresh PBS wash buffer at 37°C for 5 minutes on each occasion. They were then counterstained with 15 nM DAPI for 1 minute, rinsed with PBS, and mounted in PBS/50% glycerol medium for immediate examination by a Zeiss LSM 510 Meta Confocal Microscope. Excitation of DNA-bound DAPI was done with a 360 nm UV argon laser, while Alexa Fluor 488 was excited with a 488 nm blue-green argon laser; fluorescence in the former case was detected at BP 420-480 nm, whereas Alexa Fluor 488 emissions were detected at BP 505-530 nm.

Acknowledgements The authors thank Prof. Ambarish Mukherjee and Dr. Pradipta Saha for thoughtful discussions on the subject.

Author Contributions W.G. conceived the program, did wet lab experiments, interpreted the data and wrote the manuscript; P.K.H., S.B., J.C., P.P. and M.A. did the wet lab experiments.

Competing interests statement The authors declare that they have no competing financial interest.

Correspondence and requests for materials should be addressed to W.G. (wriman@bic.boseinst.ernet.in)

REFERENCES

1. Yoon, H.S., Hackett, J.D., Pinto, G. & Bhattacharya, D. The single, ancient origin of chromist plastids. *Proc Natl Acad Sci U S A* **99**, 15507-12 (2002).
2. Cavalier-Smith, T. Only six kingdoms of life. *Proc. Biol. Sci.* **271**, 1251–1262 (2004).
3. Apt, K.E. et al. In vivo characterization of diatom multipartite plastid targeting signals. *J Cell Sci* **115**, 4061-9 (2002).
4. Li, S., Nosenko, T., Hackett, J.D. & Bhattacharya, D. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. *Mol Biol Evol* **23**, 663-74 (2006).
5. Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* **46**, 347–366 (1999).
6. Yoon, H.S. et al. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol.* **22**, 1299–1308 (2005).

7. Gibbs, S.P. The evolution of algal chloroplasts. in *Origins of plastids* (ed. Lewin, R.A.) 107-121 (Chapman and Hall, New York, 1993).
8. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G. & Bhattacharya, D. A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–818 (2004).
9. Stackebrandt, E. et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043-7 (2002).
10. Vaara, T., Vaara, M. & Niemela, S. Two improved methods for obtaining axenic cultures of cyanobacteria. *Appl Environ Microbiol* **38**, 1011-4 (1979).
11. Schonhuber, W. et al. In situ identification of cyanobacteria with horseradish peroxidase-labeled, rRNA-targeted oligonucleotide probes. *Appl Environ Microbiol* **65**, 1259-67 (1999).
12. Amann, R. et al. Combination of 16S ribosomal RNAtargeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56**, 1919–1925 (1990).
13. Felsenstein, J. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166 (1989).
14. Tamura, K. et al. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**, 2731-2739 (2011).
15. Palenik, B. & Haselkorn, R. Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes. *Nature* **355**, 265-7 (1992).
16. Urbach, E., Robertson, D.L. & Chisholm, S.W. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* **355**, 267-70 (1992).
17. Frigaard, N.U., Larsen, K.L. & Cox, R.P. Spectrochromatography of photosynthetic pigments as a fingerprinting technique for microbial phototrophs. *FEMS Microbiol Ecol* **20**, 69–77 (1996).
18. Jeffrey, S.W. & Vesk, M. Introduction to marine phytoplankton and their pigment signatures. in *Phytoplankton pigments in oceanography* (eds. Jeffrey, S.W., Mantoura, R.F.C. & Wright, S.W.) 37–84 (UNESCO Publishing, Paris, 1997).
19. Jeffrey, S.W. & Wright, S.W. High-resolution HPLC system for chlorophylls and carotenoids of marine phytoplankton. in *Phytoplankton Pigments in Oceanography* (eds. Jeffrey, S.W., Mantoura, R.F.C. & Wright, S.W.) (UNESCO Publishing, Paris, 1997).
20. Oxborough, K. Imaging of chlorophyll a fluorescence: theoretical and practical aspects of an emerging technique for the monitoring of photosynthetic performance. *J Exp Bot* **55**, 1195-205 (2004).
21. Schubert, H., Schiewer, U. & Tschirner, E. Fluorescence characteristics of cyanobacteria (blue-green algae). *J. Plankton Res.* **11**, 353-359 (1989).
22. Keränen, M., Aro, E.-M. & Tyystjärvi, E. Excitation-Emission Map as a Tool in Studies of Photosynthetic Pigment-Protein Complexes. *Photosynthetica* **37**, 225-237 (1999).
23. Poryvkina, L., Babichenko, S. & Leeben, A. Analysis of phytoplankton pigments by excitation spectra of fluorescence. in *Proceedings of EARSeL-SIG-Workshop LIDAR* Vol., 224-232 (Dresden, FRG, 2000).
24. Hoffmann, L., Talarico, L. & Wilmotte, A. Presence of CU-phycoerythrin in the marine benthic blue-green alga Oscillatoria cf. Corallinae. *Phycologia* **29**, 19-26 (1990).

25. Yamamoto, H. et al. Phylogenetic Evidence for the Existence of Novel Thermophilic Bacteria in Hot Spring Sulfur-Turf Microbial Mats in Japan. *Appl Environ Microbiol* **64**, 1680–1687 (1998).
26. Dam, B., Ghosh, W. & Das Gupta, S.K. Conjugative Type 4 Secretion System of a Novel Large Plasmid from the Chemoautotroph Tetrathiobacter kashmirensis and Construction of Shuttle Vectors for Alcaligenaceae. *Appl Environ Microbiol* **75**, 4362–4373 (2009).
27. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-82 (1997).
28. Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512-26 (1993).
29. Tamura, K., Nei, M. & Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* **101**, 11030-5 (2004).
30. Bibby, T.S., Nield, J. & Barber, J. Three-dimensional model and characterization of the iron stress-induced CP43'-photosystem I supercomplex isolated from the cyanobacterium Synechocystis PCC 6803. *J Biol Chem* **276**, 43246-52 (2001).

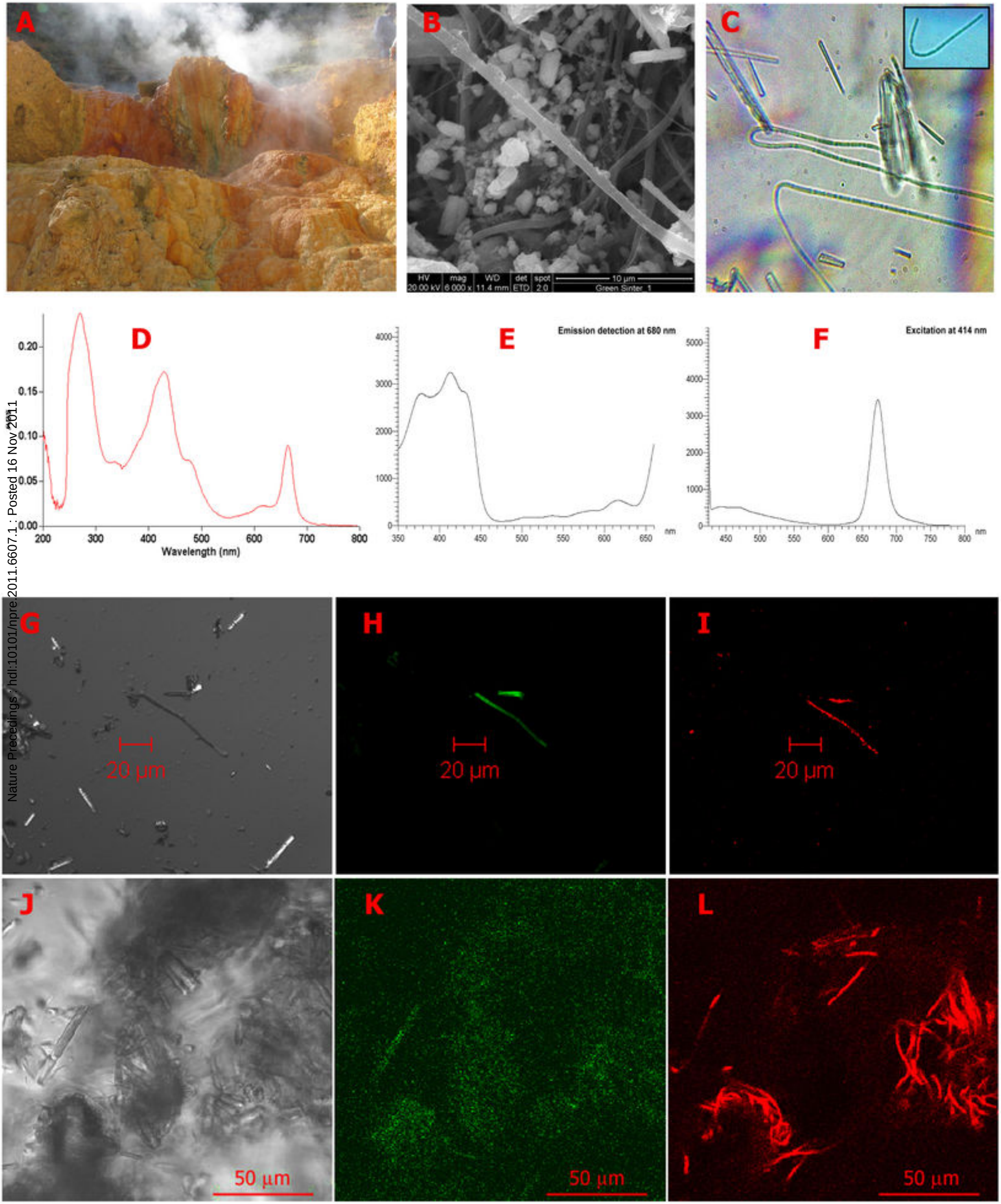
LEGENDS OF FIGURES

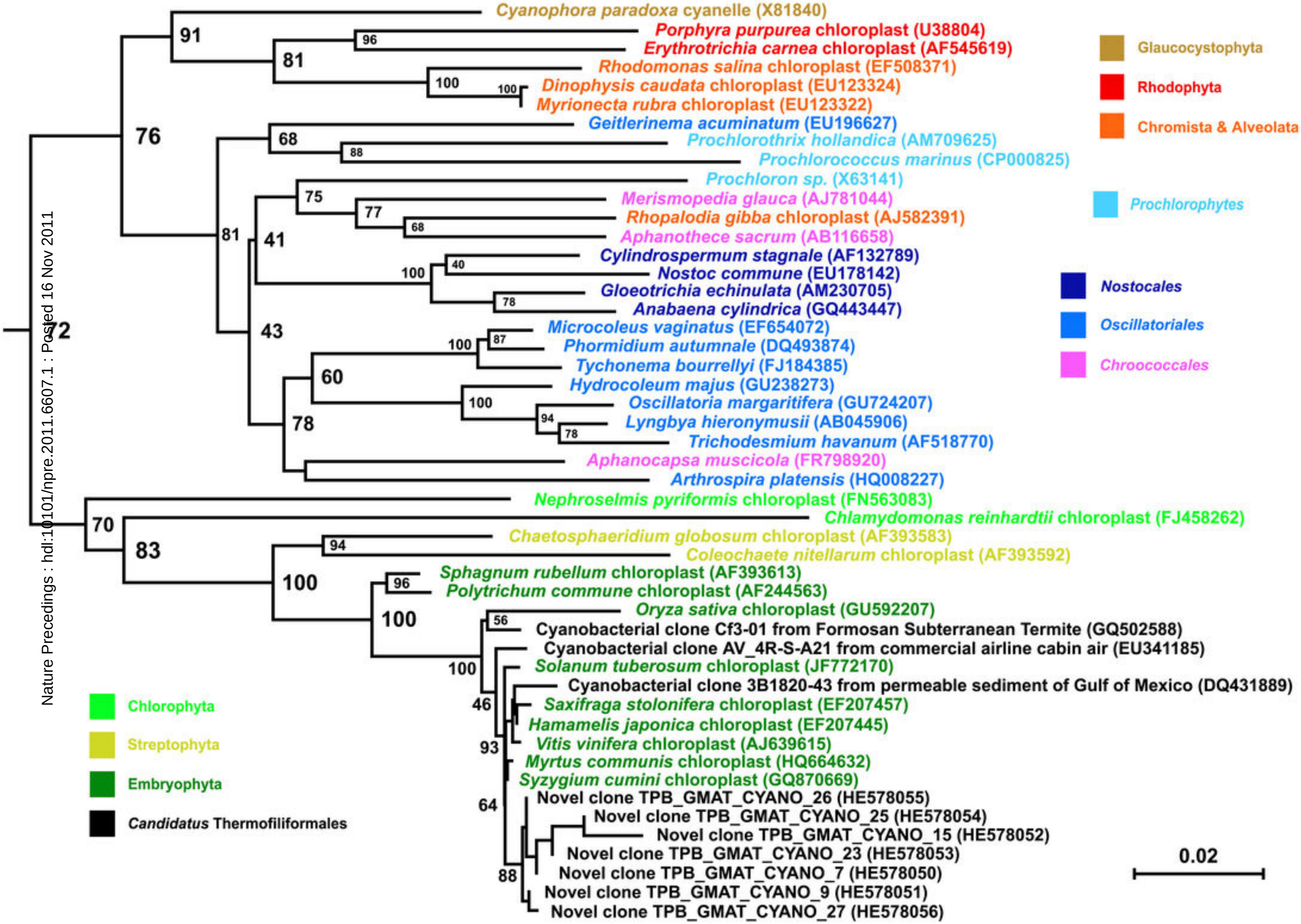
Figure 1. The green photosynthetic mats and their microscopic and spectroscopic analyses. (**A**) Lush green mats growing all over the surface of a massive geothermal sinter. (**B**) Filamentous microorganisms intertwined with mineral crystals. (**C**) Cyanobacterial filaments revealed in phase contrast microscopy. The inset shows an intact filament. Figures not to be scaled. (**D**) UV-visible absorption spectrum of the chlorophyll fraction of the acetone:methanol (7:2 v/v) extract of the green mat. Besides the 430 nm, 615 nm and 663 nm peaks, the spectrum exhibits a shoulder at 474 nm, which is characteristic of CP43-like Chl*a*-binding proteins[30]. (**E**) 4 ^{O}C excitation spectrum of the 96% ethanol extract recorded at 680 nm emission. (**F**) Fluorescence emission spectrum obtained by exciting the same extract at 430 nm. (**G**) Differential interference contrast image of a filamentous cyanobacterium from the green mat samples. (**H**) The

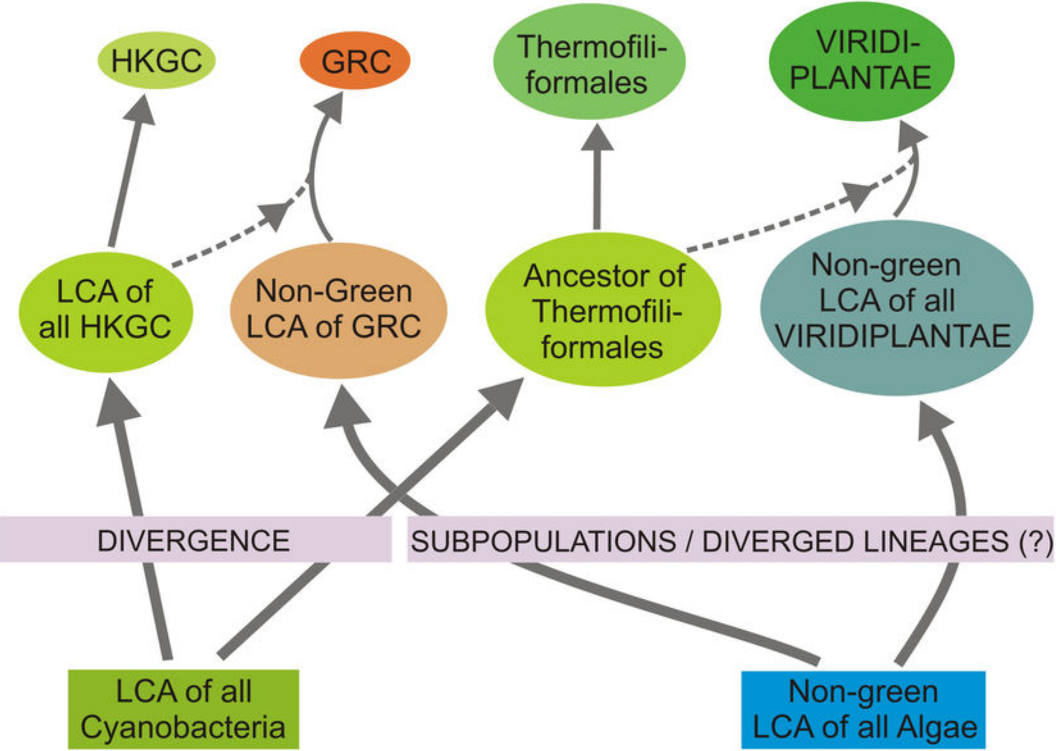
filamentous microorganism seen in ‘G’ fluoresces in LSCM at 405 nm excitation and LP 615 nm emission detection. (**I**) Fluorescence of the same organism at 633 nm excitation and LP 650 nm detection. (**J**) Differential interference contrast image of a portion of the GPM after hybridization with labeled 16S rDNA probe and DAPI counterstaining. (**K**) DAPI-specific LSCM image of the field shown in J. (**L**) Fluorescein-specific LSCM image of the same field as in J and K.

Figure 2. Rooted maximum likelihood tree constructed with 16S rRNA gene sequences from cyanobacteria closest to the new phylotype *Candidatus* Thermofiliformales and chloroplasts of some major plant groups. Different colors correspond to different taxa. Bootstrap (100 replicates) proportions supporting all major nodes are shown. The outgroup (*Aquifex aeolicus*) branch is not shown. The scale bar represents 5 substitutions per 100 sites for a unit branch length.

Figure 3. A schematic representation of the evolutionary scenario corresponding to the current phylogenetic analyses: **LCA**, **l**ast **c**ommon **a**ncestor; **HKGC**, **h**itherto **k**nown **g**roups of ***C****yanobacteria*; **GRC**, **G**laucocystophyceaea, **R**hodophyta and **C**hromalveolates. Dotted lines trace the endosymbiotic engulfment events.


A
B
HV mag WD det spot 10 µm
20.00 kV 6 000 x 11.4 mm ETD 2.0 Green Sinter_1
C
D
0.20
0.15
0.10
0.05
0.00
200 300 400 500 600 700 800
Wavelength (nm)
E
Emission detection at 680 nm
4000
3000
2000
1000
0
350 400 450 500 550 600 650
nm
F
Excitation at 414 nm
5000
4000
3000
2000
1000
0
450 500 550 600 650 700 750 800
nm
G
20 µm
H
20 µm
I
20 µm
J
50 µm
K
50 µm
L
50 µm


Cyanophora paradoxa cyanelle (X81840)
Porphyra purpurea chloroplast (U38804)
Erythrotrichia carnea chloroplast (AF545619)
Rhodomonas salina chloroplast (EF508371)
Dinophysis caudata chloroplast (EU123324)
Myrionecta rubra chloroplast (EU123322)
Geitlerinema acuminatum (EU196627)
Prochlorothrix hollandica (AM709625)
Prochlorococcus marinus (CP000825)
Prochloron sp. (X63141)
Merismopedia glauca (AJ781044)
Rhopalodia gibba chloroplast (AJ582391)
Aphanothece sacrum (AB116658)
Cylindrospermum stagnale (AF132789)
Nostoc commune (EU178142)
Gloeotrichia echinulata (AM230705)
Anabaena cylindrica (GQ443447)
Microcoleus vaginatus (EF654072)
Phormidium autumnale (DQ493874)
Tychonema bourrellyi (FJ184385)
Hydrocoleum majus (GU238273)
Oscillatoria margaritifera (GU724207)
Lyngbya hieronymusii (AB045906)
Trichodesmium havanum (AF518770)
Aphanocapsa muscicola (FR798920)
Arthrospira platensis (HQ008227)
Nephroselmis pyriformis chloroplast (FN563083)
Chlamydomonas reinhardtii chloroplast (FJ458262)
Chaetosphaeridium globosum chloroplast (AF393583)
Coleochaete nitellarum chloroplast (AF393592)
Sphagnum rubellum chloroplast (AF393613)
Polytrichum commune chloroplast (AF244563)
Oryza sativa chloroplast (GU592207)
Cyanobacterial clone Cf3-01 from Formosan Subterranean Termite (GQ502588)
Cyanobacterial clone AV_4R-S-A21 from commercial airline cabin air (EU341185)
Solanum tuberosum chloroplast (JF772170)
Cyanobacterial clone 3B1820-43 from permeable sediment of Gulf of Mexico (DQ431889)
Saxifraga stolonifera chloroplast (EF207457)
Hamamelis japonica chloroplast (EF207445)
Vitis vinifera chloroplast (AJ639615)
Myrtus communis chloroplast (HQ664632)
Syzygium cumini chloroplast (GQ870669)
Novel clone TPB_GMAT_CYANO_26 (HE578055)
Novel clone TPB_GMAT_CYANO_25 (HE578054)
Novel clone TPB_GMAT_CYANO_15 (HE578052)
Novel clone TPB_GMAT_CYANO_23 (HE578053)
Novel clone TPB_GMAT_CYANO_7 (HE578050)
Novel clone TPB_GMAT_CYANO_9 (HE578051)
Novel clone TPB_GMAT_CYANO_27 (HE578056)
72
76
91
96
81
100
100
68
88
81
75
77
68
41
100
40
78
43
100
87
60
100
94
78
78
70
83
94
100
96
100
56
100
46
93
64
88
Glaucocystophyta
Rhodophyta
Chromista & Alveolata
Prochlorophytes
Nostocales
Oscillatoriales
Chroococcales
Chlorophyta
Streptophyta
Embryophyta
Candidatus Thermofiliformales
0.02


HKGC
GRC
Thermofili-formales
VIRIDI-PLANTAE
LCA of all HKGC
Non-Green LCA of GRC
Ancestor of Thermofili-formales
Non-green LCA of all VIRIDIPLANTAE
DIVERGENCE
SUBPOPULATIONS / DIVERGED LINEAGES (?)
LCA of all Cyanobacteria
Non-green LCA of all Algae