New Phytologist

Research

Nucleomorph phylogenomics suggests a deep and ancient origin of cryptophyte plastids within Rhodophyta

Lukáš V. F. Novák[1,2], Sergio A. Muñoz-Gómez[1,3,4], Maria Ciobanu[1], Fabian van Beveren[1], Laura Eme[1], Purificación López-García[1] and David Moreira[1]

[1]Unite d'Ecologie Systematique et Evolution, CNRS, AgroParisTech, Universite Paris-Saclay, 91190, Gif-sur-Yvette, France; [2]Univ Brest, CNRS, Ifremer, EMR 6002 BIOMEX, BEEP, F-29280, Plouzané, France; [3]Department of Biological Sciences, Purdue University, West Lafayette 47907 IN, USA; [4]Center for Plant Biology, Purdue University, West Lafayette 47907 IN, USA

Author for correspondence:
David Moreira
Email: david.moreira@universite-paris-saclay.fr

Received: *11 November 2025*
Accepted: *5 March 2026*

New Phytologist (2026)
doi: 10.1111/nph.71116

Key words: algae, Cryptophyta, evolution, nucleomorphs, plastids, protists, Rhodophyta, symbiosis.

Summary

- The evolutionary origin of red algae-derived complex plastids remains obscure. Cryptophyta, one of four eukaryotic lineages harboring these plastids, still contains nucleomorphs, which are highly reduced remnants of red algal nuclei. The genes present on nucleomorph genomes can be used for phylogenomic reconstruction in order to unravel the evolutionary origin of red complex plastids and provide data independent from previously analyzed plastid-encoded proteins.
- Here, we leverage these genes in a first phylogenomic attempt at pinpointing the position of cryptophyte nucleomorphs within a comprehensive diversity of Rhodophyta, including new sequence representatives from seven deep-branching red algae.
- Our analysis, supported by a series of rigorous topology tests, places cryptophyte nucleomorphs as sister to the extremophilic, freshwater subphylum Cyanidiophytina. This conflicts with previously published analyses based on plastidial genes that placed red complex plastids closer to the mesophilic Rhodophytina.
- While the precise sister group remains debated, our results robustly reject a nucleomorph origin from within any currently recognized class of Rhodophyta, instead suggesting an ancient origin of complex red plastids among the deepest branches of the red algal tree of life.

Introduction

Complex plastids are typically photosynthetic organelles found in multiple unrelated groups of eukaryotes, derived from a symbiosis with another photosynthetic eukaryote, as opposed to primary plastids derived directly from a symbiosis with a cyanobacterium (Irisarri *et al.*, 2022; Füssy & Oborník, 2024; Novák Vanclová & Dorrell, 2024). The origin of complex plastids of red algal (Rhodophyta) ancestry found in photosynthetic members of the Cryptista, Alveolata, Stramenopiles, and Haptista ('CASH') remains one of the great unresolved questions of eukaryotic evolution. Two competing hypotheses have tried to explain the presence of these organelles of apparently shared origin in these four distantly related lineages of hosts (Yoon *et al.*, 2002; Burki *et al.*, 2020). On the one hand, the chromalveolate hypothesis (Cavalier-Smith, 1999) posits a single origin of these plastids from a single secondary endosymbiosis event in a common ancestor of CASH, followed by its vertical inheritance and multiple independent plastid losses in non-photosynthetic CASH species. On the other hand, the serial endosymbiosis hypothesis (Sanchez-Puerta & Delwiche, 2008) postulates that secondary endosymbiosis occurred more recently in a single CASH lineage, after the four lineages had already diverged, and spread to the rest horizontally via tertiary, or even higher degree, endosymbioses. This serial endosymbiosis hypothesis has been proposed in multiple versions, positing different numbers and order of symbiotic events (Sanchez-Puerta & Delwiche, 2008; Petersen *et al.*, 2014; Stiller *et al.*, 2014; Bodył, 2018). Alternatively, there could have been two independent secondary endosymbioses followed by further spread by symbioses of a higher degree (Pietluch *et al.*, 2024).

Regardless of which particular scenario is true, the identity of the red algal donor(s) of complex plastids also remains uncertain. Phylogenomic analyses of plastid-encoded genes (Yoon *et al.*, 2002; Iida *et al.*, 2007; Janouškovec *et al.*, 2010; Stiller *et al.*, 2014; Ševčíková *et al.*, 2015; Kim *et al.*, 2017) have consistently placed the CASH complex plastids in the vicinity of the Rhodophytina, the mesophilic branch of Rhodophyta, as opposed to the ancestrally extremophilic, freshwater Cyanidiophytina (Yoon *et al.*, 2017). However, our insufficient understanding of phylogenetic relationships among red algal groups, as well as poor taxon sampling of rhodophyte genomes, have for a long time limited the usefulness of such studies in precisely pinpointing the donor taxon. Recent large-scale phylogenomic analyses of plastid- and mitochondrion-encoded genes (Muñoz-Gómez *et al.*, 2017; van Beveren *et al.*, 2022) have greatly advanced our understanding of the internal phylogeny of the

phylum Rhodophyta. These analyses support the basic division between Cyanidiophytina (class Cyanidiophyceae) and Rhodophytina, and split the latter into two monophyletic subphyla: Proteorhodophytina (classes Compsopogonophyceae, Porphyridiophyceae, Rhodellophyceae, and Stylonematophyceae), comprising red algae with single-celled or simple multicellular body plans, and Eurhodophytina (classes Bangiophyceae and Florideophyceae), with complex multicellularity.

To attempt to phylogenetically place the CASH plastids within the tree of Rhodophyta, we have used an alternative source of genetic data – the nucleomorph genomes of Cryptophyta (Cryptista). Nucleomorphs (NMs) are highly reduced nuclei of archaeplastid algae that reside in the periplastidial space of certain complex plastids. There are at least three independent examples of NMs derived from green algae (in chlorarachniophytes and in two lineages of green dinoflagellates, see Ludwig & Gibbs, 1989; Sarai *et al.*, 2020), but only one derived from red algae (Ludwig & Gibbs, 1987; Douglas *et al.*, 2001). Cryptophyte NMs have been studied for several decades; phylogenetic analyses of their rRNA genes have suggested a rhodophyte origin (Douglas *et al.*, 1991; Maier *et al.*, 1991), and they have even been included in explorations of rhodophyte genomic diversity (Wong *et al.*, 2023). However, to our knowledge, there has not been any published phylogenomic analysis attempting to locate the position of cryptophyte NMs within Rhodophyta, with a single exception where NMs were not the focus and their placement was not discussed (Strassert *et al.*, 2021).

Cryptophyte NMs contain several hundred open reading frames (ORFs) that can be divided into three categories: ORFs with homologues outside cryptophyte NMs, nucleomorph-specific ORFs with homologues only in other cryptophyte NMs, and ORFs without recognizable sequence similarity to any other known sequence. Only ORFs with homologues outside cryptophyte NMs are therefore suitable for phylogenomic analyses aimed at placing NMs in a broader eukaryotic context. With this goal in mind, we have improved the taxon sampling of red algae by sequencing seven novel nuclear genomes and performed a series of phylogenomic investigations into the position of cryptophyte NMs among Rhodophyta using a large dataset of 180 conserved ORFs. Our analyses suggest that cryptophyte NMs are not affiliated with any single class of mesophilic rhodophytes and instead are more closely related to Cyanidiophytina.

Materials and Methods

Sequencing

To improve the taxon sampling of nucleus-encoded protein sequences of Rhodophyta species available for phylogenomic analyses, we sequenced DNA of seven red algae: *Galdieria* sp. ACUF 613 (Cyanidiophyceae), *Madagascaria erythrocladioides* West et Zuccarello (Compsopogonophyceae), *Erythrolobus coxiae* Scott, Baca, Ott, et West, *Porphyridium aerugineum* Geitler, *Timspurckia oligopyrenoides* Yang, Scott, et West (Porphyridiophyceae), *Dixoniella grisea* (Geitler) Scott *et al.*, and *Rhodella violacea* (Kornmann) Wehrmeyer (Rhodellophyceae). For culturing, DNA extraction, and sequencing methods see van Beveren *et al.* (2022). Raw genomic reads for *D. grisea*, *E. coxiae*, *Galdieria* sp. ACUF 613, *M. erythrocladioides*, *P. aerugineum*, and *T. oligopyrenoides* were obtained by Illumina HiSeq 2500 and NovaSeq 6000. Raw genomic reads for *R. violacea* were obtained by Illumina HiSeq 2500 and NovaSeq 6000 plus Oxford Nanopore MinION Mk 1B. The reads were assembled using SPAdes assembler, v.3.14.1 (Prjibelski *et al.*, 2020). Nuclear contigs were separated from organellar and prokaryotic contamination-derived sequences using Whokaryote, v.1.1.2 (Pronk & Medema, 2022). Raw reads were mapped to the resulting contigs using the BWA-MEM (Li, 2013) algorithm (github.com/lh3/bwa), and BLASTn (Altschul *et al.*, 1990) was used to search the contigs against the NCBI nt database, release 240. Results of the read mapping and BLASTn search were visualized in BlobTools2 (Challis *et al.*, 2020) and used for additional manual decontamination, resulting in preliminary genomic assemblies (not shown).

Taxon sampling

Previously published transcriptomic reads for *M. erythrocladioides* (SRX554333), *T. oligopyrenoides* (SRX554192), and *R. violacea* (ERX2100208) were downloaded from the NCBI SRA database (Sayers *et al.*, 2021) and mapped to their respective genomic assemblies using the BWA-MEM algorithm and visualized in the Integrative Genomics Viewer, v.2.9.4 (Robinson *et al.*, 2011). Sets of > 200 protein-coding gene models were manually predicted for each of the three assemblies with mapped transcripts and used as training sets for the online AUGUSTUS gene predictor (Hoff & Stanke, 2013). The resulting species parameters, as well as previously available *Galdieria sulphuraria* (Galdieri) Merola (Cyanidiophyceae) species parameters, were used in parallel for AUGUSTUS automatic gene prediction based on each of the seven genomic assemblies. The completeness of the individual sets of AUGUSTUS-predicted genes was evaluated using BUSCO, v.5.4.3, with the eukaryota_odb10 dataset (Manni *et al.*, 2021) and the one with the highest estimated completeness was selected for each of the seven assemblies. These preliminary predicted proteomes were searched for protein sequences of interest for use in further analyses.

Selection of markers

In order to maximize the independence of our phylogenomic matrix from plastid data, we started with a seed dataset of 228 NM-encoded proteins (markers) from the non-photosynthetic cryptophyte *Cryptomonas paramecium* (Ehrenberg) Hoef-Emden et Melkonian, which were annotated as not being associated with the plastid or photosynthesis (Tanifuji *et al.*, 2010). The 228 *C. paramecium* NM-encoded protein sequences from the 'Conserved ORFs' dataset (Tanifuji *et al.*, 2010) were searched against a custom protein database consisting of the seven novel rhodophyte preliminary predicted proteomes plus three cryptophyte NM predicted proteomes downloaded from the NCBI GenBank database, release 245, and multiple predicted proteomes representing a broad diversity of eukaryotes downloaded from the

EukProt database, v.2 (Richter *et al.*, 2022). The two best BLASTp hits (those with the lowest *E*-values) were recovered from each proteome for each query sequence. To identify orthologs among all the retrieved hits, the sequences were aligned using MAFFT, v.7.455 (Katoh & Standley, 2013) and the resulting alignments were trimmed using BMGE, v.1.12 (Criscuolo & Gribaldo, 2010) and then used to reconstruct initial single-gene trees with FastTree, v.2.1.8 with the JTT + CAT model (Price *et al.*, 2010). The trees (Supporting Information Dataset S1) were manually inspected and used to: (1) select 180 proteins that produced trees where all the NM sequences branched together to the exclusion of all other sequences (no clear alternative patterns were observed in the trees of the discarded proteins), (2) select 54 proteomes most relevant to the studied question, that is four from cryptophyte NMs and 50 from four lineages of Archaeplastida (Rhodophyta, *Rhodelphis*, Glaucophyta, and Viridiplantae), and (3) when paralogs were present, select a single paralog per tree – determined by a shorter branch. This selection resulted in an initial phylogenomic dataset of 180 markers in 54 species. Details about the selected species are shown in Table S1.

Phylogenetic analyses

The selected protein sequences were gathered in full length into 180 single-gene multi-FASTA files. PREQUAL, v.1.01 (Whelan *et al.*, 2018) was used to identify and mask positions with non-homologous characters, and the masked sequences were aligned using MAFFT, v.7.455 (Katoh & Standley, 2013). The alignments were then trimmed using DIVVIER (-partial -mincol 28), v.1.01 (Ali *et al.*, 2019). The resulting alignments (Table S2) were used to reconstruct single-marker phylogenetic trees as well as to assemble a concatenated supermatrix using phyutility, v.2.2 (Smith & Dunn, 2008). Details about the matrix are shown in Dataset S2.

The supermatrix (Dataset S3) was analyzed with IQ-TREE, v.2.2.0 (Minh *et al.*, 2020) using the substitution model LG + C60 + G + F to account for across-site compositional heterogeneity (Si Quang *et al.*, 2008) and both ultrafast bootstrap (-bb 1000) and nonparametric bootstrap (-bo 100) for the maximum likelihood (ML) phylogenetic analysis, and with PhyloBayes-MPI, v.1.8c (Lartillot *et al.*, 2013) with the CAT-GTR substitution model (Lartillot & Philippe, 2004) for the Bayesian inference phylogenetic analysis (Fig. 1). The PhyloBayes-MPI analysis with two chains was run for more than 50 000 generations. The chains did not converge (maxdiff = 1, meandiff = 0.0190534, burn-in: 1000, and sub-sampling every 10 trees), as is often the case with large phylogenomic datasets. Trace files were inspected with Tracer (Rambaut *et al.*, 2018). Trees were visualized and edited using TreeViewer, v.2.1.0 (Bianchini & Sánchez-Baracaldo, 2024).

Single-marker trees (Dataset S4) were reconstructed using IQ-TREE (-bb 1000 -m LG + C60 + G + F) v.2.2.0. PhyKIT, v.1.11.3 (Steenwyk *et al.*, 2021) was used to extract branch lengths in the 180 single-marker trees and this information guided further tests (see below) of topology robustness by sequential removal of markers or individual sequences.

Phylogenetic tests

In test 1, all markers were sorted by the length of the root branch of cryptophyte NMs (first common ancestor to last common ancestor of NMs) from longest to shortest to generate eight additional supermatrices, by removing, in a stepwise fashion, 10% of the markers. In test 2, all markers were sorted by the average compound length of cryptophyte NM branches (first common ancestor of NMs to tips) from longest to shortest to generate eight supermatrices, by removing, in a stepwise fashion, 10% of the markers. In test 3, only individual sequences were removed, not entire markers. All NM sequences were sorted by their compound branch length (first common ancestor of NMs to tips) from longest to shortest to generate eight supermatrices, by removing, in a stepwise fashion, 10% of the NM sequences. In this test, 14 sequences (from NMs, Rhodophyta, and *Rhodelphis*), identified in the IQ-TREE-generated single-marker trees as disrupting the NM monophyly, were removed together with the first 10% longest-branch sequences. Details about the removed markers, sequences, and their sorting in tests 1–3 are shown in Table S3.

In test 4, IQ-TREE, v.2.2.0 was used to calculate site-specific evolutionary rates (--rate -mlrate -m LG + C60 + G + F) for each position in the original supermatrix (Dataset S5) and SiteStripper, v.1.02 (github.com/hverbruggen/SiteStripper) was used to generate eight supermatrices, each by removing in a stepwise fashion 10% of the positions with the highest evolutionary rates.

In test 5, the amino acid compositional bias of the species in the supermatrix was determined (Fig. 2f). Occurrences of all amino acids in all positions in the matrix were counted and the positions were sorted according to the ratio F + S + I + N + K/A + P + V + D + M in NMs divided by the ratio F + S + I + N + K/A + P + V + D + M in all other species, as described before (Muñoz-Gómez *et al.*, 2019b). In order to avoid dividing by zero, an arbitrary value of 0.01 was added to all the amino acid occurrence values before the calculation. Details about this calculation and sorting are given in Table S4. A custom Python script was used to generate eight supermatrices, each with 10% of the remaining positions removed, starting with those with NMs most enriched in F + S + I + N + K and other species most depleted in A + P + V + D + M.

The modified supermatrices generated in tests 1–5 were used to reconstruct ML trees (IQ-TREE, -bb 1000 -m LG + C60 + G + F) and the ultrafast bootstrap values supporting the original or alternative positions of NMs within the resulting topologies were plotted in Fig. 2.

To further evaluate the effects of compositional heterogeneity on phylogenetic inference, in test 6, we recoded the supermatrix into a four-character state amino acid alphabet that minimizes the compositional heterogeneity present in the data. The program minmax-chisq, which implements the methods of Susko and Roger (Susko & Roger, 2007), was used to find the best recoding scheme for our supermatrix. The best recoding scheme (S4) was found to be NQFPV, RDCLKMST, AEGHIY, and W. Phylogenetic analyses on the recoded supermatrix were performed with IQ-TREE, v.2.2.2.7 and the across-site

New Phytologist

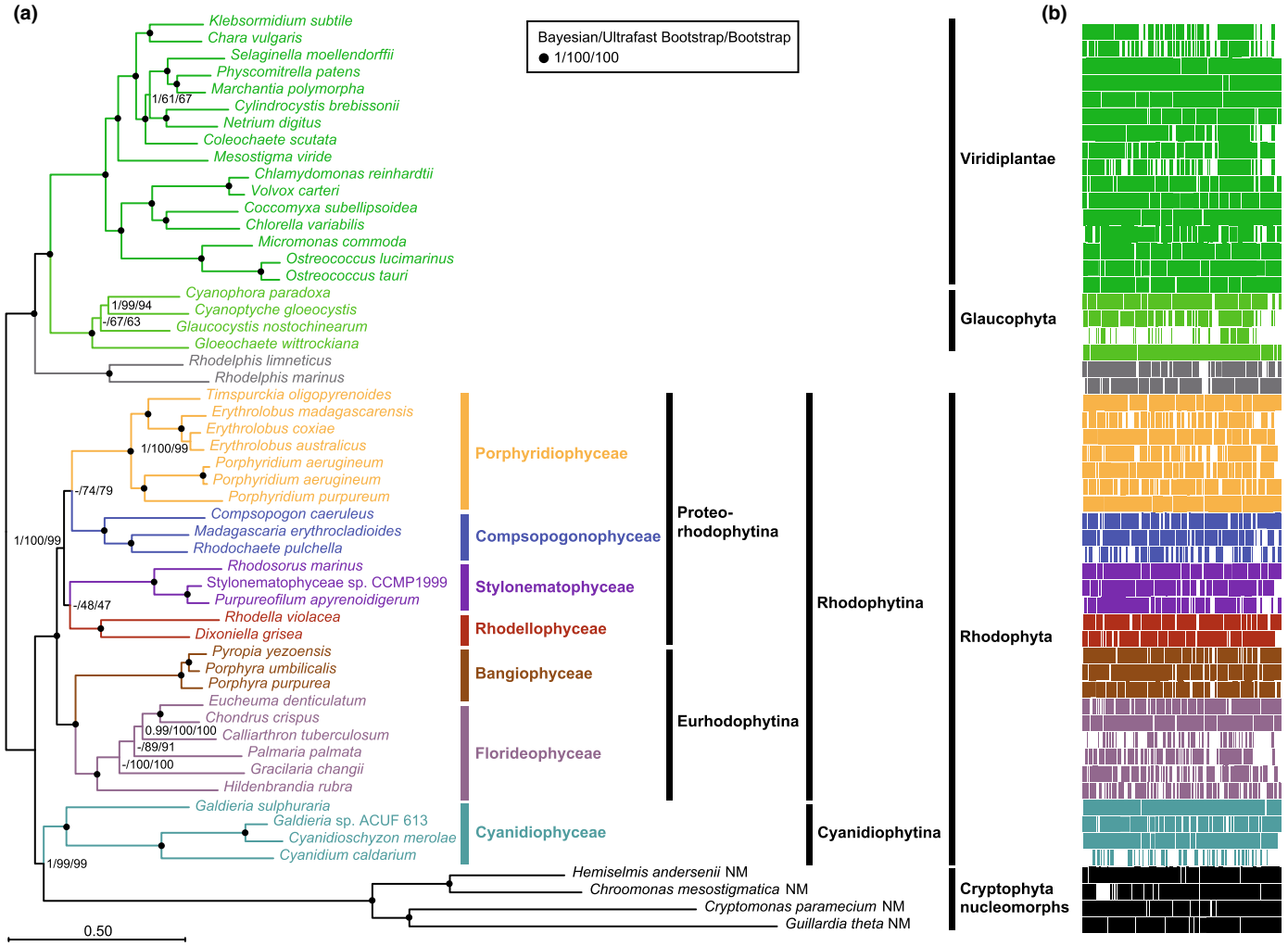


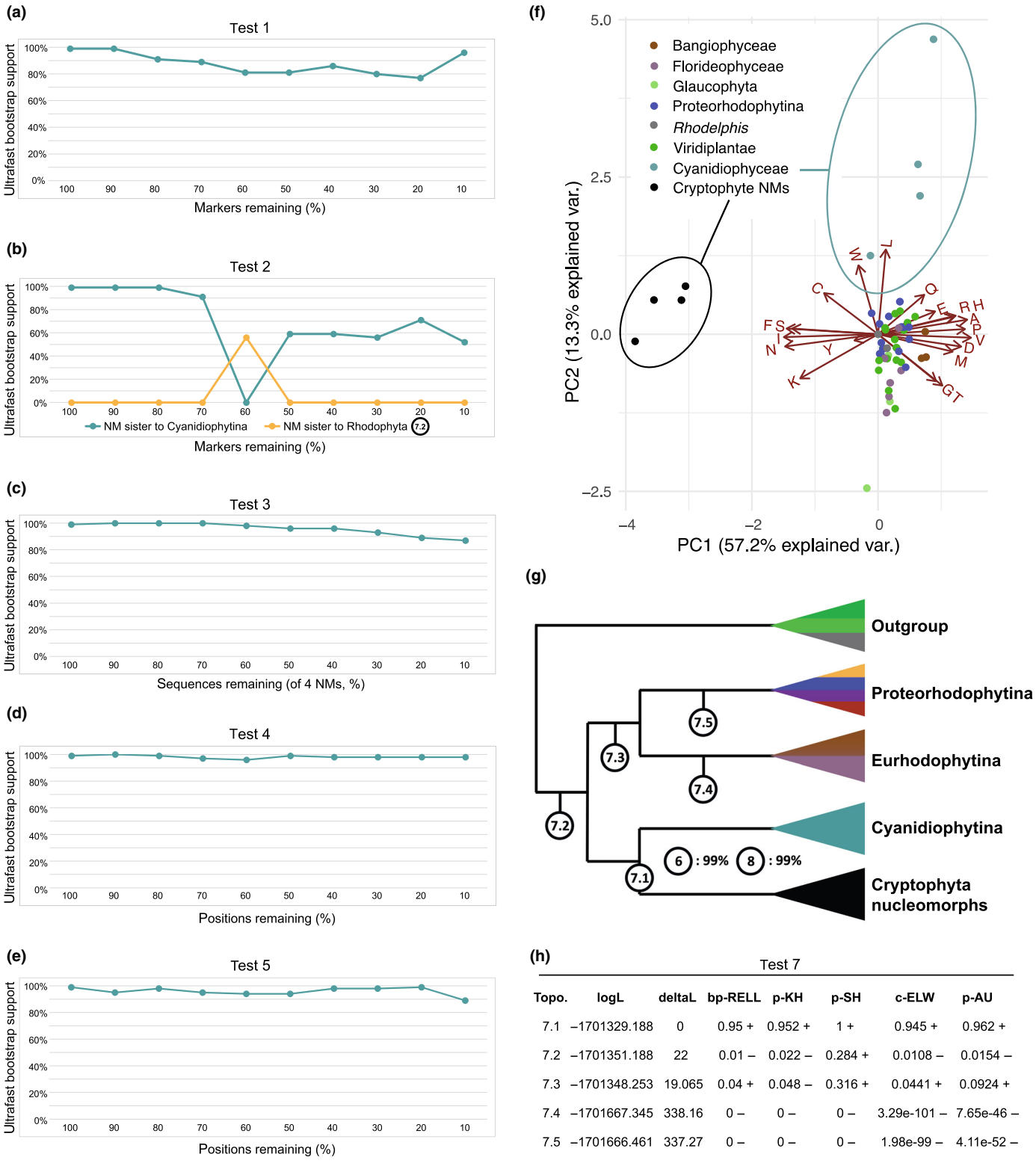
Fig. 1 Phylogenetic tree placing cryptophyte NMs among Archaeplastida. (a) Maximum likelihood (LG + C60 + G + F) phylogenetic tree constructed from a concatenation of 180 nucleus/nucleomorph-encoded proteins (51 610 positions, 54 taxa). Branch support values are derived from Bayesian posterior probability (CAT-GTR) and maximum likelihood ultrafast and nonparametric bootstrap. (b) Completeness of the individual datasets shown as a schematic representation of the complete supermatrix alignment (gaps in white, taxa with a large number of gaps correspond to genomes with low sequence coverage). The conspicuous block of missing data in *C. mesostigmatica* is related to the previously reported lineage-specific loss of proteasome-related genes in this genome (Moore *et al*., 2012).

compositionally heterogeneous GTR + C60(S4) + F + R6 model. C60(S4) is an adaptation of the C60 mixture model to four-character states. It is obtained by adding the frequencies of the amino acids that belong to each bin in the dataset-specific four-character state scheme S4.

In test 7, we calculated ML trees (IQ-TREE, -bb 1000 -m LG + C60 + G + F) with alternative topologies by constraining possible relationships under the following scenarios: (1) unconstrained original tree, that is NMs + Cyanidiophytina, (2) NMs outside Rhodophyta, (3) NMs + Rhodophytina, (4) NMs + Eurhodophytina, and (5) NMs + Proteorhodophytina. AU test (Shimodaira, 2002) was used as implemented in IQ-TREE (-m LG + C60 + G + F -n 0 -zb 1000 -au) to test the alternative hypotheses.

To evaluate the effect of functional divergence in the protein sequences, in test 8, we used the software FunDi, v.1.1 (Gaston *et al*., 2011) to estimate functionally divergent sites in the branch that separates nucleomorphs from all other nuclear genomes. About 7% of all amino acid sites were estimated to be functionally divergent with a posterior probability of > 0.5. These sites were subsequently removed using a custom Perl script in order to reduce potential artifacts related to model misspecification, since current phylogenetic models do not properly capture functional divergence in proteins (Inagaki *et al*., 2003; Muñoz-Gómez *et al*., 2022). The resulting supermatrix consisted of 47 951 amino acids. Phylogenetic analyses of this supermatrix without functionally divergent sites were performed with IQ-TREE v.2.2.2.7 and the across-site compositionally heterogeneous LG + C60 + G4 + F mixture model.

In order to test the influence of taxon sampling on the resulting topologies (Moreira *et al*., 1999), as well as to incorporate new and relevant genomic assemblies recently published, we



(h) Test 7

Topo.	logL	deltaL	bp-RELL	p-KH	p-SH	c-ELW	p-AU
7.1	−1701329.188	0	0.95 +	0.952 +	1 +	0.945 +	0.962 +
7.2	−1701351.188	22	0.01 –	0.022 –	0.284 +	0.0108 –	0.0154 –
7.3	−1701348.253	19.065	0.04 +	0.048 –	0.316 +	0.0441 +	0.0924 +
7.4	−1701667.345	338.16	0 –	0 –	0 –	3.29e-101 –	7.65e-46 –
7.5	−1701666.461	337.27	0 –	0 –	0 –	1.98e-99 –	4.11e-52 –

Fig. 2 Results of the topology robustness tests 1 to 8. (a–e) Changes in ultrafast bootstrap support for the sister relationship between cryptophyte NMs and Cyanidiophytina or alternative topologies as calculated based on reduced supermatrices in tests 1–5. (f) Principal component analysis of the taxa used in the phylogenetic analyses based on the amino acid composition of their sequences. (g) Simplified schematic representation of the alternative placements of cryptophyte NMs used in test 7 and ultrafast bootstrap support for the sister relationship between cryptophyte NMs and Cyanidiophytina as calculated based on modified supermatrices in tests 6 and 8. (h) Results of approximately unbiased topology tests (test 7).

performed tests 9 and 10 with varying numbers of included taxa. To construct datasets for these tests, we searched the newly added assemblies using HMMER, v.3.4 (Eddy, 2011), with HMM profiles constructed from the original 180 protein alignments. The best-scoring results of the HMMER searches were included in a set of 180 datasets used to reconstruct initial IQ-TREE (LG + C60 + G + F, -bb 1000) phylogenetic trees (Dataset S6) and to remove sequences forming exceptionally long branches in a process mirroring the creation of the initial phylogenomic dataset.

In test 9, we included five new cryptophyte NM genomic assemblies (Kim *et al.*, 2022; George *et al.*, 2023) and repeated the initial ML phylogenetic tree reconstruction in order to determine the relative branch lengths of all nine NMs (first common ancestor of NMs to tips). Then we constructed eight additional supermatrices by removing, in a stepwise fashion, individual NM datasets from the analyses in order of decreasing branch length. This was intended to evaluate the influence of the overall NM branch length on the topology.

In test 10, we evaluated the influence of the outgroup composition by including 68 additional taxa from the EukProt database (Richter *et al.*, 2022) covering a broad diversity of eukaryotes beyond the Archaeplastida and repeating the initial phylogenetic analysis to sort all the outgroup taxa by their relative branch length (first common ancestor of Rhodophyta + NMs to tips). Eight additional supermatrices were generated by removing, in a stepwise fashion, 10% of the outgroup taxa with the longest branches.

All alternative trees used in the tests can be found in Dataset S7. All bioinformatic methods are summarized in Fig. S1.

Results

We sequenced the genomes of seven red algae, which, together with previously published genome- and transcriptome-derived predicted proteomes available in the EukProt database (Richter *et al.*, 2022), cover all seven classes of Rhodophyta and, for the first time, allow analyzing the cryptophyte NMs phylogenetic position in relation to the full known diversity of red algae. Our final phylogenomic sequence supermatrix contained 51 610 amino acid positions belonging to 180 conserved protein markers (see annotations in Dataset S2). Our NM-based supermatrix considerably increases the number of amino acid sites compared to supermatrices based on plastid genomes (e.g. *c.* 20 000 sites in Pietluch *et al.*, 2024). In addition, our supermatrix also contained an improved taxon sampling, including 54 species of Viridiplantae, Glaucophyta, *Rhodelphis*, all seven classes of Rhodophyta (including three of the four known orders of Cyanidiophyceae (Park *et al.*, 2023)), and NMs of four distantly related cryptophyte species: *C. paramecium*, *Guillardia theta*, *Chroomonas mesostigmatica*, and *Hemiselmis andersenii* (Hoef-Emden *et al.*, 2002; Kim & Archibald, 2013; Greenwold *et al.*, 2023).

We carried out phylogenomic analyses of the sequence supermatrix by ML with the LG + C60 + G + F mixture model and by Bayesian inference with the CAT-GTR model. The resulting trees were outgroup-rooted on the non-rhodophyte Archaeplastida. Both analyses (Fig. 1) recovered the expected topology of Archaeplastida and Rhodophyta, including the monophyly of the recently recognized Proteorhodophytina (Muñoz-Gómez *et al.*, 2017), with full support (1/100/100: Bayesian posterior probability/ultrafast bootstrap/nonparametric bootstrap). Surprisingly, the analyses recovered the cryptophyte NMs as sister to Cyanidiophytina with strong support (1/99/99). This result is conflicting with all previously published phylogenomic analyses of CASH plastids (including Cryptophyta) based on plastid-encoded genes, which unanimously placed CASH as sister to the other major branch of red algae, the Rhodophytina. However, the relationship of cryptophyte NMs and Cyanidiophytina had already been observed in a few studies, including SSU rRNA gene phylogenies (Sarai *et al.*, 2020), a multi-marker analysis of nucleus- and NM-encoded genes, but containing a single cryptophyte NM representative (Baurain *et al.*, 2010), and a similar multi-marker analysis containing more taxa but that retrieved very low statistical support (Strassert *et al.*, 2021). Using our updated markers and taxon sampling, we tested the robustness of this topology in a series of analyses (tests 1–8, see the Materials and Methods section). These tests are designed to uncover potential influence on the ML results of long-branch attraction due to the fast-evolving character of NM sequences or to the sequence compositional bias observed in the NM-encoded proteins (Baurain *et al.*, 2010). If the Cryptophyte + Cyanidiophytina topology was a result of these phylogenetic artifacts, the progressive removal of divergent data (individual positions or marker genes) from the alignment would likely result in a clear trend of lowering statistical support and, possibly, toward an alternative topology (see Fig. 2).

In tests 1–4, designed to test the effect of the NM long branches, we inferred ML trees from supermatrices generated by gradually removing divergent data in increments of 10% (Fig. 2a–d), only in test 2 (sequential removal of markers by the average compound length of cryptophyte NM branches) did we observe a substantial drop in the support for the monophyly of cryptophyte NMs and Cyanidiophytina. An alternative topology (NMs sister to all Rhodophyta) rose over 50% support in a single tree (with 60% of markers remaining). None of the other tests showed this alternative topology, and the support for NMs sister to Cyanidiophytina always remained above 90%.

Tests 5 and 6 addressed the possible influence of sequence compositional biases (Fig. 2e–g). In test 5, we inferred ML trees from supermatrices generated by gradually removing compositionally biased sites in increments of 10%. The support for NMs sister to Cyanidiophytina in these trees always remained above 90% (Fig. 2e). The ML tree from test 6, where we recoded the amino acid supermatrix into a four-character state alphabet, an approach known to alleviate sequence compositional biases (Susko & Roger, 2007), such as those detected in NM-encoded proteins (Fig. 2f), presented the same topology as the original one with 99% support for NMs as sister group to Cyanidiophytina (Fig. 2g).

In addition, tree topology comparisons using the AU test (test 7, Fig. 2h) rejected all topologies except the original ML tree

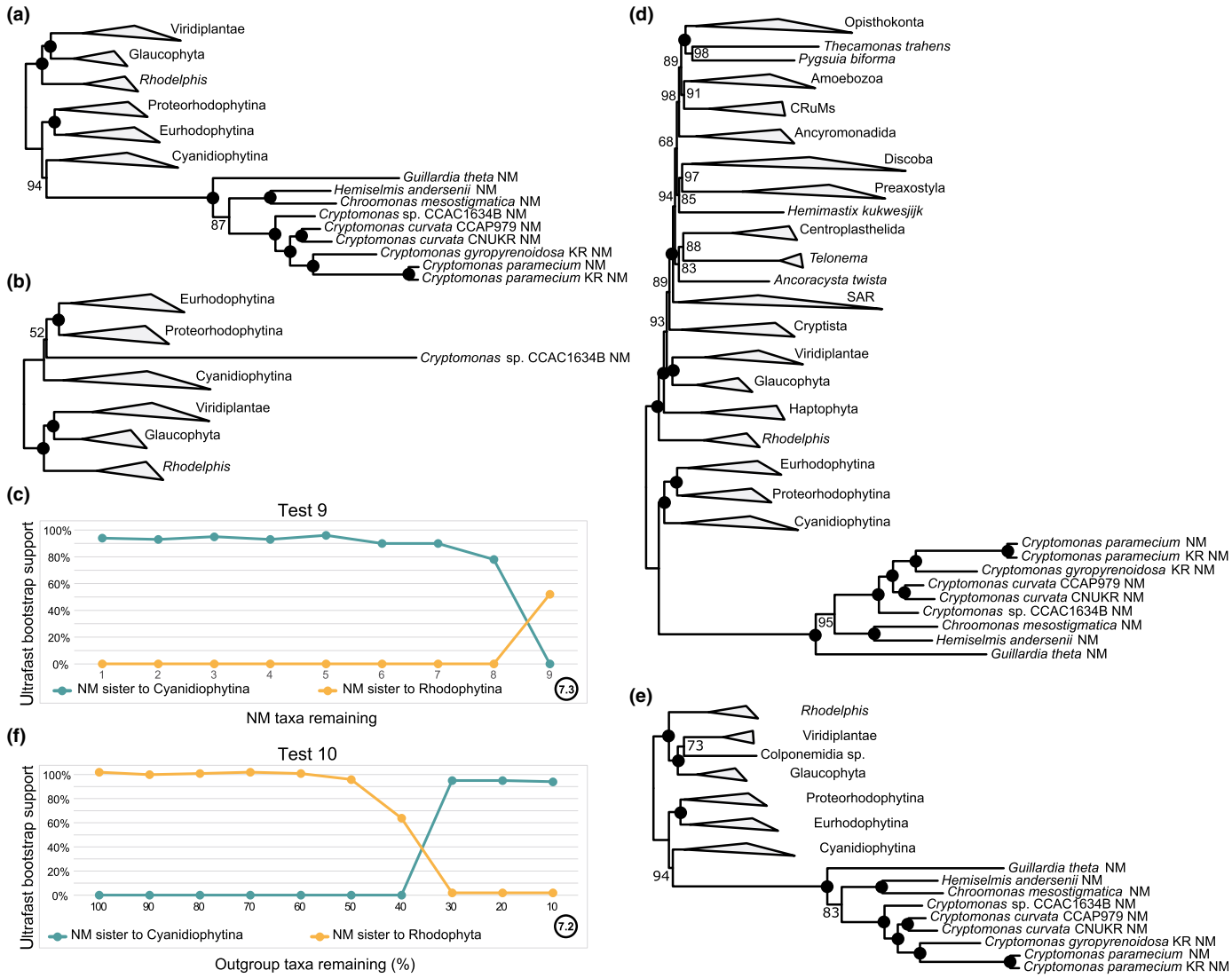


Fig. 3 Results of the topology robustness tests 9 and 10 after inclusion of additional taxa. (a) Maximum likelihood (LG + C60 + G + F, 1000× ultrafast bootstrap, 100% values are indicated with black dots) tree including all 9 available NM genomic datasets (maximum number of NMs in test 9). (b) Maximum likelihood (LG + C60 + G + F, 1000× ultrafast bootstrap, 100% values are indicated with black dots) tree including only the single NM genomic dataset with the shortest branch (minimum number of NMs in test 9). (c) Changes in ultrafast bootstrap support for the sister relationship between cryptophyte NMs and Cyanidiophytina or alternative topologies as calculated based on the progressive removal of cryptophyte NMs (test 9). (d) Maximum likelihood (LG + C60 + G + F, 1000× ultrafast bootstrap, 100% values are indicated with black dots) tree including all nine available NM genomic datasets and 68 additional outgroup taxa (maximum number of taxa in test 10). (e) Maximum likelihood (LG + C60 + G + F, 1000× ultrafast bootstrap, 100% values are indicated with black dots) tree including all 9 available NM genomic datasets and only 10% of the outgroup taxa with the shortest branches (minimum number of taxa in test 10). (f) Changes in ultrafast bootstrap support for the sister relationship between cryptophyte NMs and Cyanidiophytina or alternative topologies as calculated based on increasingly reduced numbers of taxa (test 10).

(Fig. 2g: topology 7.1). The topology consistent with previous plastid-encoded proteins analyses, that is, NMs sister to Rhodophytina, could not be rejected, although the *P*-value (0.0924) was close to the significance threshold (Fig. 2g: topology 7.3). Notably, the topology briefly supported in test 2 (Fig. 2g: topology 7.2) was rejected by the AU test (*P*-value = 0.0154). The ML tree reconstructed based on a modified supermatrix with 7% of the most functionally divergent sites removed (test 8) again recovered the original topology with 99% support for the monophyly of NMs and Cyanidiophytina (Fig. 2g).

We tested the effect of increasing the taxon sampling of NMs by including five new NM assemblies from the genus *Cryptomonas* (Kim *et al.*, 2022; George *et al.*, 2023). The ultrafast bootstrap support for cryptophyte NMs being sister to Cyanidiophytina dropped only slightly to 94% (Fig. 3a) in this analysis and remained above 90% throughout test 9, where we sequentially removed NMs from the longest–branching one to the shortest-branching one. Only when two NMs remained in the dataset did the support drop to 78%, and with only a single NM present, an alternative topology (NMs sister to

Rhodophytina; 7.3 in Fig. 2g) was slightly favored with an ultrafast bootstrap value of 52% (Fig. 3c).

Finally, we tested the effect of increasing the taxon sampling of outgroups sequences by including 68 additional outgroup taxa from a broad diversity of eukaryotes. The cryptophyte NMs branched sister to all Rhodophyta (7.2 in Fig. 2g), and the tree was not fully consistent with the expected eukaryotic phylogeny (e.g. Haptophyta branched within Archaeplastida). When 50% of the outgroup taxa with the longest branches were removed in test 10, the support for the alternative topology started to diminish, and at 30%, the originally recovered topology of NMs sister to Cyanidiophytina received support of 95% and remained over 90% in trees with further reduced outgroups (Fig. 3e,f).

Discussion

Our analyses show that the hypothesis of cryptophyte NMs branching sister to Cyanidiophytina is the best supported one by the phylogenetic analyses of conserved genes still preserved in the cryptophyte nucleomorph genomes. NM genomes have most likely experienced evolutionary pressures that differ from those acting on plastid and nuclear genomes, which have led to increased evolutionary rates and compositional biases in NMs (Patron *et al.*, 2006; Baurain *et al.*, 2010; Grisdale *et al.*, 2019). However, NM genomes appear to keep evolutionary signal, as shown by the overall congruence of the internal phylogeny of cryptophytes based on NM, nuclear (e.g. Greenwold *et al.*, 2023), and plastid (e.g. Kim *et al.*, 2017; Suzuki *et al.*, 2022) data, and we did not find any strong indication that our results were influenced by long-branch attraction or sequence compositional bias. In fact, reanalyzing the supermatrix with fewer divergent data did not show any trend toward alternative topologies. Interestingly, this position was also recovered in a previous pan-eukaryotic phylogenomic analysis, which included NM data, although with low support (Strassert *et al.*, 2021). The alternative position of NMs as sister to Rhodophytina is not strongly rejected by the NM data and, given its recovery in previous plastid genome phylogenetic analyses (Iida *et al.*, 2007; Janouškovec *et al.*, 2010; Ševčíková *et al.*, 2015; Kim *et al.*, 2017), we cannot at this stage favor one over the other. The evolutionary history of secondary red plastids clearly remains an open question.

It is difficult to imagine a mechanism by which a red algal plastid and a red algal nucleus of different origins would end up together as components of the current complex plastid of cryptophytes. Therefore, we should assume that they indeed derive from the same red algal lineage, and their different phylogenetic signals are not reflecting different evolutionary histories. The discrepancy between the two topologies: (1) NMs sister to Rhodophytina, well supported by plastidial data and (2) NMs sister to Cyanidiophytina, well supported by NM data, can be explained by an undiscovered phylogenetic artifact, by an unknown gene transfer event that resulted in different phylogenetic signals in the two sources of genomic data, or by an undiscovered deep rhodophyte lineage at the origin of CASH plastids.

The plastidial and NM genomes presumably undergo different evolutionary pressures, and it is conceivable that this resulted in a so far undiscovered systematic compositional bias in one or both sets of genomes, leading to different phylogenetic signals. Indeed, conflicting phylogenetic signals between plastids and other compartments have been reported, for example, in coralline red algae (Lee *et al.*, 2018) and Apicomplexa (Muñoz-Gómez *et al.*, 2019a). Furthermore, the case of the recently discovered 'leptophytes' shows that plastidial genomes may not provide congruent phylogenetic signal for fully resolving ancient, closely spaced speciation events, even with dense taxon sampling, large datasets, and sophisticated phylogenetic methods (Jamy *et al.*, 2026; Shrestha *et al.*, 2025).

Alternatively, either the plastidial or NM genomes might be enriched in genes of a different origin as a result of ancient horizontal gene transfer (HGT). For example, if the entire complex plastid was derived from a Rhodophytina-related ancestor, as suggested by the plastidial data, the Cyanidiophytina-like phylogenetic signal could be a result of predation on Cyanidiophytina prey by the cryptophyte host and frequent HGT replacement of genes in the NM genome. An opposite scenario could also be conceived, in which plastidial genes were replaced by HGT from Rhodophytina prey. However, this may be less likely as plastidial membranes present a less permeable barrier than the nuclear envelope, perforated by nuclear pores. These scenarios would be somewhat analogous to the 'red carpet' hypothesis explaining the presence of rhodophyte proteins in proteomes of species with secondary plastids of green origin (Ponce-Toledo *et al.*, 2019). A period of retention of two distinct endosymbionts, as exemplified by certain dinotoms (Hehenberger *et al.*, 2014), or even more complicated endosymbiotic architectures, like the karyoklepty described in ciliates (Johnson *et al.*, 2007), could also have happened in an ancestor of cryptophytes, resulting in HGT into one of the symbiotic genomes. All the presented scenarios are highly speculative and very difficult to test, given the close relationship between the two alternative source lineages and the challenges of precisely reconstructing single-gene phylogenetic trees.

Most importantly, our analyses neither show a close relationship between cryptophyte NMs and any of the six classes of Rhodophytina, nor their origin from within the known extant diversity of Cyanidiophytina. On the contrary, they support a deep and ancient origin of the cryptophyte NMs and, therefore, presumably also of all CASH plastids, within red algae. They likely emerged as sister to either Cyanidiophytina or Rhodophytina, two alternative positions placed just a single node away from each other.

Our results suggest that the secondary red endosymbiosis occurred either before the diversification of extant rhodophyte classes, or later, but from a so-far-unknown deep-branching lineage of red algae. This currently unknown lineage has either gone extinct (except for the complex plastids derived from it) or still exists, waiting to be discovered. Indeed, the discovery of a novel deep-branching lineage in the phylogenetic vicinity of Rhodophyta would not be unprecedented (Gawryluk *et al.*, 2019). Because smaller genetic divergence allows for better phylogenetic resolution, the identification of a novel microbial lineage closely related to an organelle of interest may be crucial

for resolving a disagreement between evolutionary scenarios, as demonstrated by the case of *Gloeomargaritales*, a recently described group of cyanobacteria (Couradeau *et al.*, 2012). The close relationship of *Gloeomargaritales* to primary plastids has enabled resolving the controversy over whether these organelles derived from early- or late-branching cyanobacteria (Ponce-Toledo *et al.*, 2017).

Furthermore, if the sister position of cryptophyte NMs to Cyanidiophytina recovered in our analyses gets confirmed, it would support a freshwater and/or thermoacidophilic ancestry of CASH plastids, consistent with the notion that most of the early evolution of photosynthetic eukaryotes occurred in freshwater environments (Lewis, 2017; Ponce-Toledo *et al.*, 2017; Sánchez-Baracaldo *et al.*, 2017).

Acknowledgements

This work was funded by the European Research Council Advanced Grants Protist World (No. 322669, PL-G) and PlastEvol (No. 787904, DM), the ERC Starting grant MacroEpik (No. 803151, LE), and a NASA Exobiology Program grant (80NSSC24K1875 to SAM-G). We thank the UNICELL single-cell genomics platform (deemteam.fr/en/unicell) for help in DNA preparation and Nanopore sequencing. Open access publication funding provided by COUPERIN CY26.

Competing interests

None declared.

Author contributions

LVFN, PL-G and DM designed and supervised the study. FB and MC performed culturing, DNA extraction, and sequencing. LVFN, SAM-G and LE performed data collection and phylogenetic analyses. LVFN, SAM-G, LE, PL-G and DM wrote and edited the manuscript. All authors read and approved the final manuscript.

ORCID

Fabian van Beveren https://orcid.org/0000-0003-2883-3090
Maria Ciobanu https://orcid.org/0000-0002-7504-2358
Laura Eme https://orcid.org/0000-0002-0510-8868
Purificación López-García https://orcid.org/0000-0002-0927-0651
David Moreira https://orcid.org/0000-0002-2064-5354
Sergio A. Muñoz-Gómez https://orcid.org/0000-0002-6200-474X
Lukáš V. F. Novák https://orcid.org/0000-0001-7773-9119

Data availability

All new sequences used in the phylogenomic analysis were deposited in NCBI GenBank under accession nos. –PP114032" > PP113013–PP114032.

References

Ali RH, Bogusz M, Whelan S. 2019. Identifying clusters of high confidence homologies in multiple sequence alignments. *Molecular Biology and Evolution* **36**: 2340–2351.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Baurain D, Brinkmann H, Petersen J, Rodríguez-Ezpeleta N, Stechmann A, Demoulin V, Roger AJ, Burger G, Lang BF, Philippe H. 2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular Biology and Evolution* **27**: 1698–1709.

van Beveren F, Eme L, López-García P, Ciobanu M, Moreira D. 2022. Independent size expansions and intron proliferation in red algal plastid and mitochondrial genomes. *Genome Biology and Evolution* **14**: evac037.

Bianchini G, Sánchez-Baracaldo P. 2024. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecology and Evolution* **14**: e10873.

Bodył A. 2018. Did some red alga-derived plastids evolve via kleptoplastidy? A hypothesis. *Biological Reviews* **93**: 201–222.

Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of Eukaryotes. *Trends in Ecology & Evolution* **35**: 43–55.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *Journal of Eukaryotic Microbiology* **46**: 347–366.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit – interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics* **10**: 1361–1374.

Couradeau E, Benzerara K, Gérard E, Moreira D, Bernard S, Brown GE, López-García P. 2012. An early-branching microbialite cyanobacterium forms intracellular carbonates. *Science* **336**: 459–462.

Criscuolo A, Gribaldo S. 2010. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology* **10**: 210.

Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng L-T, Wu X, Reith M, Cavalier-Smith T, Maier U-G. 2001. The highly reduced genome of an enslaved algal nucleus. *Nature* **410**: 1091–1096.

Douglas SE, Murphy CA, Spencer DF, Gray MW. 1991. Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. *Nature* **350**: 148–151.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Computational Biology* **7**: e1002195.

Füssy Z, Oborník M. 2024. Complex endosymbioses I: from primary to complex plastids, serial endosymbiotic events. In: Maréchal E, ed. *Plastids: methods and protocols*. New York, NY, USA: Springer US, 21–41.

Gaston D, Susko E, Roger AJ. 2011. A phylogenetic mixture model for the identification of functionally divergent protein residues. *Bioinformatics* **27**: 2655–2663.

Gawryluk RMR, Tikhonenkov DV, Hehenberger E, Husnik F, Mylnikov AP, Keeling PJ. 2019. Non-photosynthetic predators are sister to red algae. *Nature* **572**: 240–243.

George EE, Barcytė D, Lax G, Livingston S, Tashyreva D, Husnik F, Lukeš J, Eliáš M, Keeling PJ. 2023. A single cryptomonad cell harbors a complex community of organelles, bacteria, a phage, and selfish elements. *Current Biology* **33**: 1982–1996.

Greenwold MJ, Merritt K, Richardson TL, Dudycha JL. 2023. A three-genome ultraconserved element phylogeny of cryptophytes. *Protist* **174**: 125994.

Grisdale CJ, Smith DR, Archibald JM. 2019. Relative mutation rates in nucleomorph-bearing algae. *Genome Biology and Evolution* **11**: 1045–1053.

Hehenberger E, Imanian B, Burki F, Keeling PJ. 2014. Evidence for the retention of two evolutionary distinct plastids in dinoflagellates with diatom endosymbionts. *Genome Biology and Evolution* **6**: 2321–2334.

Hoef-Emden K, Marin B, Melkonian M. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. *Journal of Molecular Evolution* **55**: 161–179.

Hoff KJ, Stanke M. 2013. WebAUGUSTUS – a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**: W123–W128.

Iida K, Takishita K, Ohshima K, Inagaki Y. 2007. Assessing the monophyly of chlorophyll-c containing plastids by multi-gene phylogenies under the unlinked model conditions. *Molecular Phylogenetics and Evolution* **45**: 227–238.

Inagaki Y, Blouin C, Susko E, Roger AJ. 2003. Assessing functional divergence in EF-1α and its paralogs in eukaryotes and archaebacteria. *Nucleic Acids Research* **31**: 4227–4237.

Irisarri I, Strassert JFH, Burki F. 2022. Phylogenomic insights into the origin of primary plastids. *Systematic Biology* **71**: 105–120.

Jamy M, Huber T, Antoine T, Ruscheweyh H-J, Paoli L, Pelletier E, Delmont TO, Burki F. 2026. Identification of a deep-branching lineage of algae using environmental plastid genomes. *Nature Communications* **17**: 662.

Janouškovec J, Horák A, Oborník M, Lukeš J, Keeling PJ. 2010. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences, USA* **107**: 10949–10954.

Johnson MD, Oldach D, Delwiche CF, Stoecker DK. 2007. Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* **445**: 426–428.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kim E, Archibald JM. 2013. Ultrastructure and molecular phylogeny of the cryptomonad *Goniomonas avonlea* sp. nov. *Protist* **164**: 160–182.

Kim JI, Moore CE, Archibald JM, Bhattacharya D, Yi G, Yoon HS, Shin W. 2017. Evolutionary dynamics of cryptophyte plastid genomes. *Genome Biology and Evolution* **9**: 1859–1872.

Kim JI, Tanifuji G, Jeong M, Shin W, Archibald JM. 2022. Gene loss, pseudogenization, and independent genome reduction in non-photosynthetic species of *Cryptomonas* (Cryptophyceae) revealed by comparative nucleomorph genomics. *BMC Biology* **20**: 227.

Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**: 1095–1109.

Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology* **62**: 611–615.

Lee JM, Song HJ, Park SI, Lee YM, Jeong SY, Cho TO, Kim JH, Choi H-G, Choi CG, Nelson WA *et al.* 2018. Mitochondrial and plastid genomes from coralline red algae provide insights into the incongruent evolutionary histories of organelles. *Genome Biology and Evolution* **10**: 2961–2972.

Lewis LA. 2017. Hold the salt: freshwater origin of primary plastids. *Proceedings of the National Academy of Sciences, USA* **114**: 9759–9760.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*: 1303.3997.

Ludwig M, Gibbs SP. 1987. Are the nucleomorphs of cryptomonads and *Chlorarachnion* the vestigial nuclei of eukaryotic endosymbionts? *Annals of the New York Academy of Sciences* **503**: 198–211.

Ludwig M, Gibbs SP. 1989. Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: morphology, division and DNA-DAPI fluorescence. *Journal of Phycology* **25**: 385–394.

Maier U-G, Hofmann CJB, Eschbach S, Wolters J, Igloi GL. 1991. Demonstration of nucleomorph-encoded eukaryotic small subunit ribosomal RNA in cryptomonads. *Molecular and General Genetics MGG* **230**: 155–160.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**: 4647–4654.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Moore CE, Curtis B, Mills T, Tanifuji G, Archibald JM. 2012. Nucleomorph genome sequence of the cryptophyte alga *Chroomonas mesostigmatica* CCMP1168 reveals lineage-specific gene loss and genome complexity. *Genome Biology and Evolution* **4**: 1162–1175.

Moreira D, Le Guyader H, Philippe H. 1999. Unusually high evolutionary rate of the elongation factor 1 alpha genes from the Ciliophora and its impact on the phylogeny of eukaryotes. *Molecular Biology and Evolution* **16**: 234–245.

Muñoz-Gómez SA, Durnin K, Eme L, Paight C, Lane CE, Saffo MB, Slamovits CH. 2019a. *Nephromyces* represents a diverse and novel lineage of the Apicomplexa that has retained apicoplasts. *Genome Biology and Evolution* **11**: 2727–2740.

Muñoz-Gómez SA, Hess S, Burger G, Lang BF, Susko E, Slamovits CH, Roger AJ. 2019b. An updated phylogeny of the Alphaproteobacteria reveals that the parasitic Rickettsiales and Holosporales have independent origins. *eLife* **8**: e42535.

Muñoz-Gómez SA, Mejía-Franco FG, Durnin K, Colp M, Grisdale CJ, Archibald JM, Slamovits CH. 2017. The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Current Biology* **27**: 1677–1684.

Muñoz-Gómez SA, Susko E, Williamson K, Eme L, Slamovits CH, Moreira D, López-García P, Roger AJ. 2022. Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nature Ecology & Evolution* **6**: 253–262.

Novák Vanclová AMG, Dorrell RG. 2024. Complex plastids across the eukaryotes: an overview of inherited and convergently evolved characters. In: Schwartzbach SD, Kroth PG, Oborník M, eds. *Endosymbiotic organelle acquisition: solutions to the problem of protein localization and membrane passage*. Cham: Springer International, 39–88.

Park SI, Cho CH, Ciniglia C, Huang T-Y, Liu S-L, Bustamante DE, Calderon MS, Mansilla A, McDermott T, Andersen RA *et al.* 2023. Revised classification of the Cyanidiophyceae based on plastid genome data with descriptions of the Cavernulicolales ord. nov. and Galdieriales ord. nov. (Rhodophyta). *Journal of Phycology* **59**: 444–466.

Patron NJ, Rogers MB, Keeling PJ. 2006. Comparative rates of evolution in endosymbiotic nuclear genomes. *BMC Evolutionary Biology* **6**: 46.

Petersen J, Ludewig A-K, Michael V, Bunk B, Jarek M, Baurain D, Brinkmann H. 2014. *Chromera velia*, endosymbioses and the Rhodoplex hypothesis – plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH lineages). *Genome Biology and Evolution* **6**: 666–684.

Pietluch F, Mackiewicz P, Ludwig K, Gagat P. 2024. A new model and dating for the evolution of complex plastids of red alga origin. *Genome Biology and Evolution* **16**: evae192.

Ponce-Toledo RI, Deschamps P, López-García P, Zivanovic Y, Benzerara K, Moreira D. 2017. An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology* **27**: 386–391.

Ponce-Toledo RI, López-García P, Moreira D. 2019. Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytologist* **224**: 618–624.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**: e9490.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. *Current Protocols in Bioinformatics* **70**: e102.

Pronk LJU, Medema MH. 2022. Whokaryote: distinguishing eukaryotic and prokaryotic contigs in metagenomes based on gene structure. *Microbial Genomics* **8**: 000823.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**: 901–904.

Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG, Burki F, de Vargas C. 2022. EukProt: a database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal* **2**: e56.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**: 24–26.

Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. 2017. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proceedings of the National Academy of Sciences, USA* **114**: E7737–E7745.

Sanchez-Puerta MV, Delwiche CF. 2008. A hypothesis for plastid evolution in chromalveolates. *Journal of Phycology* **44**: 1097–1107.

Sarai C, Tanifuji G, Nakayama T, Kamikawa R, Takahashi K, Yazaki E, Matsuo E, Miyashita H, Ishida K, Iwataki M *et al.* 2020. Dinoflagellates with relic endosymbiont nuclei as models for elucidating organellogenesis. *Proceedings of the National Academy of Sciences, USA* **117**: 5364–5375.

Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S *et al.* 2021. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **50**: D20–D26.

Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, Jenkins J, Schmutz J, Přibyl P, Fousek J *et al.* 2015. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Scientific Reports* **5**: 10134.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**: 492–508.

Shrestha B, Romero MF, Villada JC, Consortium PM, Crysten B-H, Schulz F. 2025. Global metagenomics reveals plastid diversity and unexplored algal lineages. *bioRxiv*: 2025.03.28.644651.

Si Quang L, Gascuel O, Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**: 2317–2323.

Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24**: 715–716.

Steenwyk JL, Buida TJ III, Labella AL, Li Y, Shen X-X, Rokas A. 2021. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* **37**: 2325–2331.

Stiller JW, Schreiber J, Yue J, Guo H, Ding Q, Huang J. 2014. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nature Communications* **5**: 5764.

Strassert JFH, Irisarri I, Williams TA, Burki F. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nature Communications* **12**: 1879.

Susko E, Roger AJ. 2007. On reduced amino acid alphabets for phylogenetic inference. *Molecular Biology and Evolution* **24**: 2139–2150.

Suzuki S, Matsuzaki R, Yamaguchi H, Kawachi M. 2022. What happened before losses of photosynthesis in cryptophyte algae? *Molecular Biology and Evolution* **39**: msac001.

Tanifuji G, Onodera NT, Wheeler TJ, Dlutek M, Donaher N, Archibald JM. 2010. Complete nucleomorph genome sequence of the nonphotosynthetic alga *Cryptomonas paramecium* reveals a core nucleomorph gene set. *Genome Biology and Evolution* **3**: 44–54.

Whelan S, Irisarri I, Burki F. 2018. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**: 3929–3930.

Wong DK, Grisdale CJ, Slat VA, Rader SD, Fast NM. 2023. The evolution of pre-mRNA splicing and its machinery revealed by reduced extremophilic red algae. *Journal of Eukaryotic Microbiology* **70**: e12927.

Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. *Proceedings of the National Academy of Sciences, USA* **99**: 15507–15512.

Yoon HS, Nelson W, Lindstrom SC, Boo SM, Pueschel C, Qiu H, Bhattacharya D. 2017. Rhodophyta. In: Archibald JM, Simpson AGB, Slamovits CH, eds. *Handbook of the protists*. Cham: Springer International, 89–133.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Dataset S1 Preliminary phylogenetic trees of individual markers reconstructed using FastTree.

Dataset S2 Detailed information about the composition of the phylogenomic supermatrix.

Dataset S3 Phylogenomic supermatrix constructed from all markers used in the analysis.

Dataset S4 Single-gene trees of all the markers incorporated in the supermatrix.

Dataset S5 Site-specific evolutionary rates for each position in the supermatrix.

Dataset S6 Preliminary phylogenetic trees of individual markers reconstructed using IQ-TREE for tests 9 and 10.

Dataset S7 Alternative phylogenetic trees reconstructed in tests 1–6 and 8–10.

Fig. S1 Flowchart of the bioinformatic methods.

Table S1 List of taxa used in the analyses including sources of genetic data.

Table S2 Single-gene alignments of the markers.

Table S3 Details about marker removal based on branch lengths (tests 1–3).

Table S4 Details about the calculation of amino acid compositional bias.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.