Phylogenomics invokes the clade housing Cryptista, Archaeplastida, and *Microheliella maris*.

Euki Yazaki^{1, †, *}, Akinori Yabuki^{2, †, *}, Ayaka Imaizumi³, Keitaro Kume⁴, Tetsuo Hashimoto^{5,6},
and Yuji Inagaki^{6,7}

1: RIKEN iTHEMS, Wako, Saitama 351-0198, Japan

2: Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa 236-0001,
Japan

3: College of Biological Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8572, Japan.

4: Faculty of Medicine, University of Tsukuba, Tsukuba, Ibaraki, 305-8575, Japan

5: Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-
8572, Japan

6: Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba,
Ibaraki, 305-8572, Japan

7: Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki, 305-8572,
Japan

[†]EY and AY equally contributed to this work.

*Correspondence addressed to Euki Yazaki: euki87@gmail.com and Akinori Yabuki:
yabukia@jamstec.go.jp

Running title: The clade housing Cryptista, Archaeplastida, and *Microheliella maris*.

24 Abstract

25 As-yet-undescribed branches in the tree of eukaryotes are potentially represented by some of
 26 “orphan” protists (unicellular micro-eukaryotes), of which phylogenetic affiliations have not
 27 been clarified in previous studies. By clarifying the phylogenetic positions of orphan protists,
 28 we may fill the previous gaps in the diversity of eukaryotes and further uncover the novel
 29 affiliation between two (or more) major lineages in eukaryotes. *Microheliella maris* was
 30 originally described as a member of the phylum Heliozoa, but a pioneering large-scale
 31 phylogenetic analysis failed to place this organism within the previously described
 32 species/lineages with confidence. In this study, we analyzed a 319-gene alignment and
 33 demonstrated that *M. maris* represents a basal lineage of one of the major eukaryotic lineages,
 34 Cryptista. We here propose a new clade name “Pancryptista” for Cryptista plus *M. maris*. The
 35 319-gene analyses also indicated that *M. maris* is a key taxon to recover the monophyly of
 36 Archaeplastida and the sister relationship between Archaeplastida and Pancryptista, which is
 37 collectively called as “CAM clade” here. Significantly, Cryptophyceae tend to be attracted to
 38 Rhodophyta depending on the taxon sampling (ex., in the absence of *M. maris* and
 39 Rhodelphidia) and the particular phylogenetic “signal” most likely hindered the stable recovery
 40 of the monophyly of Archaeplastida in previous studies. We hypothesize that many
 41 cryptophycean genes (including those in the 319-gene alignment) recombined partially with the
 42 homologous genes transferred from the red algal endosymbiont during secondary
 43 endosymbiosis and bear a faint phylogenetic affinity to the rhodophytan genes.

44

45 **Keywords:** phylogenetic artifacts, endosymbiotic gene transfer, Cryptophyceae,
 46 Goniomonadea, Diaphoretickes.

47

1. Introduction

Our understanding of the evolutionary relationship among major eukaryotic groups has been progressed constantly. The foundation of the tree of eukaryotes was developed initially based on the combination of morphological characteristics (including those on the ultrastructural level) and molecular phylogenetic analyses of a single or few marker genes^{1–3}. In recent years, “phylogenomic” analyses—phylogenetic analyses of large-scale multigene alignments, particularly those comprising hundreds of genes—were often conducted to reconstruct deep splits in the tree of eukaryotes with high statistical support^{4–6}. For instance, recent phylogenomic analyses have constantly reconstructed the clade of stramenopiles, Alveolata, and Rhizaria (SAR clade)⁷, that of Opisthokonta, Amoebozoa, Breviatea, and Apusomonadida (Amorphea)⁸, and that of Collodictyonidae, Rigifilida, and *Mantamons* (CRuMs)⁹.

There are many unicellular micro-eukaryotes/lineages of which phylogenetic positions remain uncertain (“orphan” eukaryotes/lineages). Some of the current orphan eukaryotes/lineages most likely represent as-yet-unknown portions of the diversity of eukaryotes and hold clues to resolve the eukaryotic evolution. Prior to DNA sequencing experiments gaining in popularity in phylogenetic/taxonomic studies, diverse eukaryotes were isolated from the natural environments and examined by microscopes. If the morphological characteristics of the eukaryotes of interest showed no clear affinity to any other eukaryotes, their phylogenetic affiliations remained uncertain^{10–14}. The analyses of small subunit ribosomal DNA (SSU rDNA)—one of the most popular gene markers for organismal phylogeny—succeeded in finding the phylogenetic homes of many eukaryotes/lineages, of which morphological information was insufficient to resolve their phylogenetic affiliations^{15–18}. More recently, orphan eukaryotes/lineages, as well as newly found eukaryotes have been subjected to phylogenomic analyses^{7,8,19–26}.

Phylogenomic analyses are unlikely the silver bullet to all of the orphan eukaryotes/lineages recognized to date. For instance, the positions of Malawimonadida²⁷, Ancyromonadida⁹, Hemimastigophora²⁸, *Ancoracysta twista*²⁹, and *Microheliella maris*³⁰ could not be clarified even after phylogenomic analyses, implying that they are genuine deep branches that are critical to resolving the backbone of the tree of eukaryotes^{9,28–30}. In this study, we challenged to clarify the phylogenetic position of *M. maris* by analyzing a new phylogenomic alignment. *M. maris* was originally described as a member of the phylum Heliozoa based on the shared morphological similarities (e.g., the radiating axopodia with tiny granules and the centroplast)³¹. Cavalier-Smith et al. (2015)³⁰ then examined the phylogenetic position of *M. maris* by analyzing the alignment comprising 187 genes. Nevertheless, *M. maris* is still regarded as one of the orphan eukaryotes³², as the choice of the methods for tree reconstruction and taxon sampling affected largely the position of this eukaryote in the 187-gene phylogeny³⁰.

We here reassessed the phylogenetic position of *M. maris* by analyzing a new phylogenomic alignment comprising 319 genes (88,592 amino acid positions in total). The 319-gene phylogeny placed *M. maris* at the base of the Cryptista clade with high statistical support, suggesting that this eukaryote holds keys to understand the early evolution of Cryptista as well as Diaphoretickes. Indeed, we further demonstrated that *M. maris* and Rhodelphidia, which occupy the basal position of Cryptista and that of Rhodophyta, respectively, suppress the erroneous ‘signal’ attracting Cryptophyceae and Rhodophyta to each other and contribute to recovering (i) the monophyly of Archaeplastida and (ii) the sister relationship between Archaeplastida and the clade of Cryptista plus *M. maris*. Finally, we explored the biological ground for the phylogenetic artifact uniting Cryptophyceae and Rhodophyta together.

97 2. Methods

98 (a) Cell culturing and RNA-seq analysis

99 We generated the RNA-seq data from *M. maris* and *Hemiarma marina*, a species of
100 Goniomonadea, in this study. The culture of *M. maris* [studied in Yabuki et al. (2012)³¹] and
101 that of *H. marina* [established in Shiratori and Ishida (2016)³³] have been kept in the
102 laboratory and were utilized in this study. The harvested cells of both organisms were subjected
103 to RNA extraction using TRIzol (Life Technologies) by following the manufacturer's
104 instructions. We shipped the two RNA samples to a biotech company (Hokkaido System
105 Science) for cDNA library construction from the poly-A-tailed RNAs followed by sequencing
106 using the Illumina Hi-seq 2000 platform. For *M. maris*, 1.6×10^7 paired-end 100 bp reads (1.6
107 Gb in total) were obtained and then assembled into 30,305 unique contigs by TRINITY^{34,35}.
108 For *H. marina*, we obtained 1.9×10^7 paired-end 100 bp reads (1.9 Gb in total) and assembled
109 them into 41,539 unique contigs by TRINITY^{34,35}.

111 (b) Global eukaryotic phylogeny

112 To elucidate the phylogenetic position of *M. maris*, we prepared a phylogenomic alignment by
113 updating an existing dataset comprising 351 genes²⁸. For each of the 351 genes, we added the
114 homologous sequences retrieved from the transcriptomic data newly generated from *M. maris*
115 and *H. marina* in this study (see above), as well as other eukaryotes that were absent in the
116 original data²⁸, such as *Marophrys* sp. SRT127³⁶, two species of Rhodelphidia (i.e.,
117 *Rhodelphis limneticus* and *R. marinus*)²⁵, and *Ancoracysta twista*²⁹. Individual single-gene
118 alignments were aligned by MAFFT v. 7.205^{37,38} with the L-INS-i algorithm followed by
119 manual correction and exclusion of ambiguously aligned positions. For each of the single-gene
120 alignments, the maximum-likelihood (ML) phylogenetic tree was inferred by FASTTREE v.
121 2.1^{39,40} under the LG + Γ model with robustness assessed with a 100-replicate bootstrap
122 analysis. Individual single-gene trees were inspected to identify the alignments bearing

aberrant phylogenetic signal that disagreed strongly with any of a set of well-established monophyletic assemblages in the tree of eukaryotes, namely Opisthokonta, Amoebozoa, Alveolata, stramenopiles, Rhizaria, Rhodophyta, Chloroplastida, Glaucophyta, Haptophyta, Cryptista, Jakobida, Euglenozoa, Heterolobosea, Diplomonadida, Parabasalia, and Malawimonadida. 32 out of the 351 single-gene alignments were found to violate the above-mentioned criteria and were excluded from the phylogenomic analyses described below. The remaining 319 single-gene alignments (Table S1) were concatenated into a single phylogenomic alignment containing 82 taxa with 88,592 unambiguously aligned amino acid positions. The coverage for each single-gene alignment is summarized in Table S1.

We first subjected the final alignment comprising 319 genes from 82 taxa (GlobE alignment) to the ML method by IQ-TREE 1.6.12⁴¹ with the LG + Γ + F model, and the resultant ML tree was then used as a guide tree in a more thorough analysis with the LG + Γ + F + C60 + PMSF (posterior mean site frequencies) model⁴². The robustness of the ML phylogenetic tree was evaluated with a nonparametric ML bootstrap analysis with the LG + Γ + F + C20 + PMSF model (100 replicates). We also conducted Bayesian phylogenetic analysis with the CAT + GTR model using PHYLOBAYES-mpi v. 1.8a^{43,44}. In this analysis, two MCMC runs were run for 5,000 cycles with ‘burn-in’ of 1,250. The consensus tree with branch lengths and Bayesian posterior probabilities (BPPs) were calculated from the remaining trees.

We evaluated the contribution of fast-evolving positions in the GlobE alignment to the position of *M. maris*. Substitution rates of individual alignment positions were calculated over the ML tree by IQ-TREE 1.6.12⁴¹ and top 20%, 40%, 60%, and 80% fastest-evolving positions were then removed from the original alignment. The processed alignments were then subjected to the ML bootstrap analysis with the UFBOOT approximation⁴⁵ (1,000 replicates) by using IQ-TREE 1.6.12⁴¹ with the LG + Γ + F model. Henceforth, the alignment modification and the following ML bootstrap analyses are designated as “FPR (fast-evolving

position removal)” analysis.

We also examined the impact of the sampling of the genes in the GlobE alignment on the position of *M. maris* by “RGS (random gene sampling)” analyses described below⁴⁶. From the 319 genes in the GlobE alignment, 50 genes were randomly sampled and concatenated into a single alignment (“rs50g” alignment). The above procedure was repeated 50 times to obtain 50 of rs50g alignments. Likewise, we prepared (i) 50 of “rs100g” alignments comprising 100 randomly sampled genes, (ii) 10 of “rs150g” alignments comprising 150 randomly sampled genes, and (iii) 10 of “rs200g” alignments comprising 200 randomly sampled genes. The alignments comprising randomly sampled genes were subjected individually to the ML bootstrap analysis with the UFBOOT approximation (1,000 replicates) by using IQ-TREE 1.6.12⁴¹ with the LG + Γ + F model.

(c) Diaphoretickes phylogeny

To evaluate the impact of the inclusion of *M. maris* to the phylogenetic relationship among the species/lineages in Diaphoretickes, we excluded 22 taxa from the GlobE alignment to generate the second phylogenomic alignment, of which taxa were occupied mostly by members of Diaphoretickes. Note that the number of genes remained the same between the GlobE and the second, “Diaph” alignments. The Diaph alignment was subjected to both ML and Bayesian phylogenetic analysis under all the same conditions as described above, except that we used the LG + Γ + F + C60 + PMSF model for the ML bootstrap analysis. Both FPR and RGS analyses (see above) were applied to the Diaph alignment. We also conducted both FPR and RGS analyses after excluding Rhodelphidia and *M. maris* alternatively from the Diaph alignment.

The taxon sampling of the Diaph alignment was further modified by excluding (i) both Rhodelphidia and *M. maris*, (ii) Rhodelphidia, *M. maris*, and *P. bilix*, (iii) Rhodelphidia, *M. maris*, *P. bilix*, and Goniomonadea, and (iv) Rhodelphidia, *M. maris*, *P. bilix*, and

Cryptophyceae. We ran RGS analyses of all of the four alignments described above, and the last two were subjected to FPR analyses as well.

3. Results and Discussion

(a) *Microheliella maris* represents a lineage basal to Cryptista: Proposal of “Pancryptista”

We analyzed a transcriptome-based GlobE alignment consisting of 319 genes sampled from 82 eukaryotes, which represent the major taxonomic assemblages and several orphan taxa/lineages. The GlobE phylogeny recovered the major clades in eukaryotes, such as SAR, Amorphea, CRuMs, Discoba, and Cryptista with full statistical support in both ML and Bayesian methods (Fig. 1; see also Fig. S1). *M. maris* branched at the base of the Cryptista clade, which comprises *Palpitomonas bilix*, Goniomonadea including *Hemiarma marina*, and Cryptophyceae, with an MLBP of 100% and a BPP of 1.0. The GlobE alignment includes no data of Kathablepharidacea that is the other cryptistan subgroup, as their available data is extremely low site coverage. However, the lack of Kathablepharidacea most likely gave little impact on the phylogenetic position of *M. maris* relative to Cryptista, as far as the alignment includes *P. bilix* which is more basal than Kathablepharidacea in the Cryptista clade²⁰. The monophyly of Archaeplastida including Rhodelphidia (*Rhodelphis limneticus* and *R. marinus*) and Picozoa sp., both of which grouped with Rhodophyta, was recovered with an MLBP of 81% and a BPP of 1.0. The intimate affinity of Rhdelphidia and Picozoa to Rhodophyta in the GlobE phylogeny is consistent with the recent phylogenomic studies^{25,47}. Neither of the two recently proposed major clades in eukaryotes, T-SAR (Telonemia plus SAR)²³ and Haptista (Centrohelea plus Haptophyta)²¹, was reconstructed. Either or both ML and Bayesian phylogenetic analyses failed to give full statistical support to the nodes connecting the lineages/species in Diaphoretickes, namely Archaeplastida, Centrohelea, Haptophyta,

Telonemia, SAR, and Cryptista plus *M. maris*. Thus, we conclude that the 82-taxon phylogeny is insufficient to retrace the early evolution of Diaphoretickes with confidence.

We here examined the phylogenetic position of *M. maris* inferred from the GlobE alignment by the progressive removal of fast-evolving positions (FPR analyses). The contribution of fast-evolving positions in the GlobE alignment to the union of *M. maris* and Cryptista is most likely negligible, as the ultrafast bootstrap support values (UFBPs) for the clade comprising *M. maris* and Cryptista stayed 100% until the top 80% fastest-evolving positions were removed (purple line in Fig. S2A). We detected two conflicting phylogenetic signals regarding the position of *M. maris* relative to the members of Cryptista included in the GlobE alignment, one placing *M. maris* at the base of the Cryptista clade and the other uniting *M. maris* and *P. bilix* directly (red and yellow lines, respectively, in Fig. S2A). However, the former signal constantly dominated over the latter, regardless of the amount of fast-evolving positions in the alignment. Thus, we conclude that the basal position of *M. maris* to the Cryptista clade in the GlobE phylogeny (Fig. 1) is free from potential phylogenetic artifacts stemming from fast-evolving positions.

To evaluate the impact of gene sampling on the position of *M. maris* in the GlobE phylogeny, we randomly sampled 50 genes, 100 genes, 150 genes, and 200 genes from the 319 genes and concatenated them into “rs50g,” “rs100g,” “rs150g,” and “rs200g” alignments, respectively (note that the taxon sampling remained the same). The UFBPs for the clade comprising *M. maris* and Cryptista calculated from 50 of rs50g alignments and 50 of rs100g alignments distributed from 0 (or nearly 0) to 100% (Fig. S2B; see also Table S2 for the details), suggesting the phylogenetic signal in 100 or fewer genes are insufficient for stable recovery of the grouping of interest. Nevertheless, in the analyses of rs150g and rs200g alignments, the UFBP for the grouping of *M. maris* and Cryptista increased. Significantly, the support for the grouping of *M. maris* and Cryptista received UFBPs of 81.3-100% in the

analyses of rs200g alignments (the rightmost plot in Fig. S2B; see also Table S2). As observed in FPR analysis of GlobE alignment (Fig. S2A), rs50g and rs100g alignments appeared to contain two conflicting signals for the phylogenetic position of *M. maris*, one placing *M. maris* at the base of the Cryptista clade and the other uniting *M. maris* and *P. bilix* directly (Figs. S2C and D). In the analyses of rs150g alignments, the UFBP for the basal position of *M. maris* to the Cryptista clade appeared to be greater than that for the direct union of *M. maris* and *P. bilix* in 8 out of the 10 cases (Table S2). The same trend was observed in the analyses of rs200g alignments, albeit the signal for the clade grouping *M. maris* and *P. bilix* remained detectable (Fig. S2D). This series of the analyses demonstrated that the greater the number of genes included in the alignment, the greater the UFBP for the basal position of *M. maris* at the Cryptista clade. Based on the results described above, we conclude that the basal position of *M. maris* to the Cryptista clade is genuine, and henceforth designate the clade grouping *M. maris* and the previously known Cryptista as Pancryptista.

(b) The sister relationship between Archaeplastida and Pancryptista: Proposal of “CAM clade”

The Diaph alignment, which was generated by excluding 22 taxa from the GlobE alignment, was analyzed to explore the impact of *M. maris* on the phylogenetic relationship among the major lineages in Diaphoretickes. 21 out of the 22 taxa excluded from the GlobE alignment were not the member of Diaphoretickes—most of Opisthokonta, all discobids, *Paratrimastix pyriformis*, *Nutomonas longa*, and *Malawimonas jakobiformis*. We excluded a single member of Diaphoretickes, Picozoa sp., from the Diaph alignment due to its instability in the GlobE phylogeny, which likely stemmed from low site coverage in the GlobE alignment (Fig. 1). The phylogenetic relationship among the major Diaphoretickes lineages/species inferred from the Diaph alignment (e.g., the basal position of *M. maris* to the Cryptista clade) was essentially the

same as that inferred from the GlobE alignment (Figs. 1 and 2A). We here focus on the monophyly of Archaeplastida and the sister relationship between Archaeplastida and Pancryptista, both of which were fully supported in the ML and Bayesian analyses of the Diaph alignment (Fig. 2A).

Neither monophyly of Archaeplastida nor sister relationship between Archaeplastida and Cryptista has not been settled even by recently published phylogenomic studies. For instance, Gawryluk et al. (2019)²⁵ conducted the analyses of an alignment comprising 254 genes but could not settle the relationship among Chloroplastida, Glaucophyta, and Rhodophyta plus Rhodelphidia. The ML analysis of their 254-gene alignment put Cryptista within the three lineages of Archaeplastida, albeit Bayesian analysis of the same alignment recovered both monophyly of Archaeplastida and sister relationship between Archaeplastida and Cryptista. In another phylogenomic study, the monophyly of Archaeplastida was not reconstructed in either ML or Bayesian analysis of an alignment comprising 248 genes, as Cryptista was tied with Rhodophyta²³. Irisarri, Strassert, and Burki (2021)⁴⁸ recently analyzed a 311-genes alignment and demonstrated that taxon sampling, selection of alignment positions, and substitution models for phylogeny are critical to recovering the monophyly of Archaeplastida.

In contrast to the pioneering phylogenomic studies (see above), both monophyly of Archaeplastida and the sister relationship between Pancryptista and Archaeplastida were reconstructed from the Diaph alignment with full statistical support by both ML and Bayesian methods (Fig. 2A). Significantly, FPR analysis on the Diaph alignment appeared to have little impact on the UFBPs for the two nodes of interest (Fig. 2B). Both monophyly of Archaeplastida and sister relationship between Pancryptista and Archaeplastida received full or nearly full UFBPs until the top 60% fastest-evolving positions were removed (purple and green lines, respectively, in Fig. 2B). We also analyzed rs50g, rs100g, rs150g, and rs200g alignments

generated from the Diaph alignment (Figs. 2C, F, and I; see also Table S3 for the details). The results from RGS analyses clearly indicate that the UFBPs for the two nodes of interest (and that for the monophyly of Pancryptista) increased in proportion to the number of genes considered. We here conclude that Archaeplastida is a genuine clade as demonstrated by Irisarri, Strassert, and Burki (2021)⁴⁸, and Pancryptista is the closest relative of Archaeplastida and thus propose the sister relationship between Archaeplastida and Pancryptista as “CAM” clade. The proposed name is an acronym derived from the first letters of Cryptista, Archaeplastida, and *Microheliella*.

It is significant to note that the monophyly of Archaeplastida and the sister relationship between Archaeplastida and Cryptista was recovered by a 311-genes phylogeny prior to the *M. maris* data is available⁴⁸. Thus, we decided to evaluate systematically how the increment of *M. maris*, as well as that of Rhodelphidia, contributed to the recovery of the monophyly of Archaeplastida and the sister relationship between Archaeplastida and Pancryptista/Cryptista. We re-analyzed rs50g, rs100g, rs150g, and rs200g alignments generated from the Diaph alignment after excluding Rhodelphidia or *M. maris* (Figs. 2D, E, G, H, J, and K; see also Table S4 for the details). In the analyses of rs50g and rs100g alignments, the removal of Rhodelphidia/*M. maris* lowered the overall distributions of the UFBPs for the monophyly of Archaeplastida and sister relationship between Archaeplastida and Pancryptista/Cryptista (Figs. 2D, E, G, and H). However, most of the UFBPs for the two groupings of interest in the analyses of rs200g alignments were around of or greater than 90% (Figs. 2D, E, G, and H; Table S4). Likewise, the overall distribution of the UFBPs for the monophyly of Pancryptista was apparently lowered in the analyses of rs50g, rs100g, and rs150g alignments in the absence of Rhodelphidia (Fig. 2J). After *M. maris* was excluded, the monophyly of Cryptista was constantly recovered with full UFBPs, except a UFBP of 21.5% obtained in the analyses of a single rs50g alignment (Fig. 2K; Table S4). The removal of

Rhodelphidia or *M. maris* appeared to possess a moderate but apparent impact on the monophyly of Archaeplastida and sister relationship between Archaeplastida and Pancryptista/Cryptista in the alignments comprising 100 or fewer genes, albeit such impact can be overcome by an increment of the alignment size.

(c) On the artifactual grouping of Rhodophyta and Cryptophyceae

The phylogenetic analyses described above indicated that taxon sampling is a key to recover the monophyly of Archaeplastida and the sister relationship between Archaeplastida and Pancryptista (CAM clade) with confidence. Then, why did phylogenomic analyses, in which either or both of Rhodelphidia and *M. maris* were absent, often failed to recover the monophyly of Archaeplastida? For instance, a recent phylogenomic study²³, which considered neither Rhodelphidia nor *M. maris*, grouped Rhodophyta and Cryptista together instead of recovering the monophyly of Archaeplastida.

The absence of both Rhodelphidia and *M. maris* gave a greater impact on the UFBP for the monophyly of Archaeplastida and that for the sister relationship between Archaeplastida and Cryptista than the absence of either of the two lineages/species. Regardless of the number of randomly sampled genes, the distributions of the UFBPs for the two groupings of interest tend to be lower than the corresponding values calculated from the analyses excluding either Rhodelphidia or *M. maris* (compare Figs. 2D and E with Fig. 3A, and Fig. 2G and H with Fig. 3E; see also Table S3 for the details). Interestingly, the faint affinity between Rhodophyta and Cryptista became detectable in the absence of Rhodelphidia and *M. maris* (Fig. 3I). After *P. bilix* was additionally excluded (Cryptista was represented by Goniomonadea and Cryptophyceae), both UFBP for the monophyly of Archaeplastida and that for the sister relationship between Archaeplastida and Cryptista were further lowered (Fig. 3B and F). In stark contrast, the exclusion of *P. bilix* enhanced the affinity between Rhodophyta and

Cryptista (Fig. 3J). These results clearly indicated that, in the absence of Rhodelphidia and *M. maris*, *P. bilix* possesses a significant impact on the recovery of the monophyly of Archaeplastida by excluding Cryptista.

We further analyzed the alignments in which Cryptista was represented solely by Cryptophyceae or Goniomonadea. The most drastic results were obtained from the analyses of the alignments in which the member of Cryptophyceae were the sole representatives of Cryptista (Figs. 3C, G, and K; see also Table S4 for the details). After the exclusion of the non-photosynthetic lineages in CAM clade (i.e., Rhodelphidia, *M. maris*, *P. bilix*, and Goniomonadea), the union of Cryptista (i.e., Cryptophyceae) and Rhodophyta appeared to dominate over the monophyly of Archaeplastida, particularly in the analyses of larger-size alignments (Figs. 3G and K). The analyses of rs200g alignments recovered the union of Cryptista and Rhodophyta with UFBPs ranging from 43.2 to 97.6% (Table S4). Of note, the suppression of UFBPs appeared to be much severer to the sister relationship between Archaeplastida and Cryptista than the monophyly of Archaeplastida—the UFBPs for the former grouping were 0 or nearly 0, regardless of the alignment size (Fig. 3C; Table S4). Comparing to the analyses considering Cryptophyceae as the sole representative of Cryptista, we observed only the mild suppression of the monophyly of Archaeplastida and that of the sister relationship between Archaeplastida and Cryptista in the analyses of the alignments in which Cryptista was represented by Goniomonadea (Figs. 3D and H; Table S4). These observations coincide with the affinity between Goniomonadea and Rhodophyta being less intense (Fig. 3L) than that between Cryptophyceae and Rhodophyta (Fig. 3K).

We revealed that Rhodophyta was attracted to Goniomonadea and Cryptophyceae erroneously in the ML phylogenies inferred from the alignments lacking Rhodelphidia, *M. maris*, and *P. bilix* (see above). In the global eukaryotic phylogeny, Rhodelphidia interrupts the branch leading to the Rhodophyta clade. Likewise, *M. maris* and *P. bilix*, both of which are

basal branches of the Pancryptista clade, break the branch leading to the clade of Cryptophyceae and Goniomonadea (i.e., Cryptomonada). Thus, the grouping of Rhodophyta and Cryptomonada is most likely the phylogenetic artifact in which the two long branches, one exposed by the absence of Rhodelphidia and the other exposed by the absence of *M. maris* and *P. bilix*, attract to each other—long-branch attraction (LBA) artifact⁴⁹. Importantly, this phylogenetic artifact could not be overcome completely in the analyses of the alignments comprising at least 200 genes (Figs. 3J-L). If so, we anticipated that the putative phylogenetic artifact uniting Rhodophyta and Cryptomonada was enhanced further by the exclusion of Cryptophyceae (or Goniomonadea) (Figs. 3K and L), as this procedure extended the branch leading to the clade of Goniomonadea (or Cryptophyceae). The exclusion of Goniomonadea (i.e., Cryptophyceae were the sole representatives of Cryptomonada) appeared to enhance the putative phylogenetic artifact much greater degree than the exclusion of Cryptophyceae (i.e., Goniomonadea were the sole representatives of Cryptomonada) (compare Figs. 3C with D, Figs. 3G with H, and Figs. 3K and L). These results imply that the phylogenetic artifact uniting Rhodophyta and Cryptophyceae is substantially different from that uniting Rhodophyta and Goniomonadea. Altogether, we conclude that both size and taxon sampling, particularly the sampling of the members of CAM clade, in alignments heavily matter to reconstruct the monophyly of Archaeplastida and sister relationship between Archaeplastida and Pancryptista/Cryptista with confidence.

To pursue the reason why Rhodophyta is artifactually attracted to Cryptophyceae more severely than Goniomonadea (see above), we modified the Diaph alignment by excluding Rhodelphidia and all members of Pancryptista except Goniomonadea, and the resultant alignment was then subjected to FPR analysis (Fig. 3M). In the analysis of the alignment with full positions, the union of Goniomonadea and Rhodophyta received a UFBP of greater than 80%, while the monophyly of Archaeplastida was supported by a UFBP of smaller than 10%.

The analyses of the alignments after removing the top 20-60% fastest-evolving positions drastically increased the UFBP for the monophyly of Archaeplastida (89-100%; green line in Fig. 4M), while the UFBP for the grouping of Rhodophyta and Goniomonadea was reduced to less than 10% (red line in Fig. 3M). Thus, we conclude that the union of Rhodophyta and Goniomonadea is the typical LBA artifact stemming from fast-evolving positions. We repeated the same analysis described above but substituted Goniomonadea with Cryptophyceae (Fig. 3N). Unexpectedly, the union of Rhodophyta and Cryptophyceae received UFBP of 98%, 100%, 100%, 70%, and 92% in the analyses after removal of top 20%, 40%, 60%, and 80% fastest-evolving positions, respectively (red line in Fig. 3N). The UFBP for the monophyly of Archaeplastida was less than 10%, except the analysis after removal of the top 60% fastest-evolving positions gave a UFBP of 30% (green line in Fig. 3N). These results strongly suggest that, in terms of dependency on fast-evolving positions, the phylogenetic artifact uniting Rhodophyta and Cryptophyceae is distinct from the typical LBA artifact uniting Rhodophyta and Goniomonadea.

(d) The biological perspective on the signal uniting Rhodophyta and Cryptophyceae recovered in phylogenomic analyses

It is attractive to propose that the difference between the artifact uniting Rhodophyta and Cryptophyceae and that uniting Rhodophyta and Goniomonadea stems from the difference in lifestyle between the two closely related lineages in Cryptista. Goniomonadea is primarily heterotrophic and their nuclear genomes are free from endosymbiotic gene transfer (EGT)⁵⁰. Indeed, the series of the phylogenetic analyses described above demonstrated that the typical LBA was sufficient to explain the union of Rhodophyta and Goniomonadea (illustrated typically by Fig. 3M). In contrast, the extant member of Cryptophyceae possesses the plastids that were traced back to a red algal endosymbiont in the common ancestor of Cryptophyceae.

During the red algal endosymbiont being transformed into a host-governed plastid, a number of genes had been transferred from the endosymbiont nucleus to the host nucleus. If a phylogenomic alignment contains genes acquired from the red algal endosymbiont, such genes are the source of the phylogenetic “signal” uniting Cryptophyceae and Rhodophyta. However, we selected the 319 genes, each of which showed no apparent sign of EGT in the corresponding single-gene phylogenetic analysis, for the phylogenomic analyses in this study. Additionally, we calculated the log-likelihoods (lnLs) of two identical tree topologies except for the position of Cryptophyceae—one bearing the monophyly of Archaeplastida (Tree 1) and the other bearing the grouping of Rhodophyta and Cryptophyceae (Tree 2)—for each of the 319 single-gene alignments (note that Rhodelphidia, *M. maris*, *P. bilix*, and Goniomonadea were omitted from the alignments) (Figs. S5A and S5B). The 319 single-gene alignments were sorted by the lnL difference between the two test trees (normalized by the alignment lengths) and the top 10 alignments, which prefer Tree 2 over Tree 1 most, were subjected individually to the standard ML phylogenetic analyses (Fig. S6, see also Table S5 for the details). Nevertheless, we detected any strong phylogenetic affinity between Rhodophyta and Cryptophyceae in none of the 10 ML single-gene analyses (Fig. S6, see also Table S5). These results cannot be explained by a simple scenario assuming that a subset of the ‘cryptophycean genes’ in the phylogenomic alignment was in fact acquired endosymbiotically from the red algal endosymbiont as briefly mentioned in Cavalier-Smith et al., (2015)³⁰. Rather, we here propose that a potentially large number of the cryptophycean nuclear genes (including those composed of the phylogenomic alignment) are the chimeras of the sequence inherited vertically beyond the red algal endosymbiosis and that acquired from the red algal endosymbiont. In each chimeric gene, the phylogenetic signal from the red alga-derived gene portion is likely insufficient to unite Rhodophyta and Cryptophyceae together in the single-gene analysis. However, when multiple chimeric genes in the cryptophycean nuclear genomes

were included in a phylogenomic alignment, the phylogenetic signal from the red alga-derived gene portion becomes detectable as the union of Rhodophyta and Cryptista in the absence of Rhodelphidia and the basal branching taxa in Pancryptista, such as *M. maris* and *P. bilix*. Unfortunately, we additionally calculated the site-wise lnL differences between Trees 1 and 2, albeit no clear sign for the putative red algal gene fragments was detected (Fig. S7, see also Table S6).

4. Conclusion

In this work, we successfully deepen our understanding of the early evolution of eukaryotes. The phylogenomic analyses presented here demonstrated that *M. maris* is critical to understand the early evolution of Cryptista, as well as that of Archaeplastida. We also revealed that the deep branches of Archaeplastida and Cryptista—Rhodelphidia, *M. maris*, *P. bilix* (although not examined in this study, Picozoa most likely possesses the equivalent impact to the above-mentioned species/lineages, too)—are critical to suppress the cryptic and severe phylogenetic “signal” in cryptophycean genes.

Data accessibility

The transcriptome data and genome data of strain *M. maris* were deposited in DDBJ Sequence Archive under the accession nos. DRAXXX and DRAYYY. The assembled transcriptomes and genome sequences of *M. maris* and phylogenetic alignments analyzed in this study are available from the Dryad Digital Repository.

Acknowledgments

This study was supported by grants from the Japan Society for the Promotion of Science (18KK0203 & 19H03280) awarded to YI, (11J04684 & 17K19434) to AY, and was partially

supported by the Tree of Life Project of University of Tsukuba. The phylogenetic analyses conducted in this work have been carried out under the ‘Interdisciplinary Computational Science Program’ in the Center for Computational Sciences, University of Tsukuba.

Authors' contributions

EY and AY participated in the design of the study, carried out the molecular lab work, carried out data analyses including phylogenetic analyses and drafted the manuscript; AI prepared the phylogenomic dataset; KK and TH assembled the RNA-seq data; YI participated in the design of the study and revised the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Declaration of interests

We declare no competing interests.

Figure legends

Figure 1. Phylogenetic position of *Microheliella maris* inferred from the GlobE alignment.

The tree topology and branch lengths were inferred from the GlobE alignment (319 genes; 88,592 amino acid positions in total) by the maximum-likelihood (ML) and Bayesian methods. Bayesian analysis recovered principally an identical tree topology (Fig. S1). For each bipartition, the ML bootstrap support values (MLBPs) and Bayesian posterior probabilities (BPPs; if greater than 0.50) are shown. The bipartitions with dots indicate MLBPs of 100% and BPPs of 1.0. The bar graph for each taxon represents the percent coverage of the amino acid positions in the GlobE alignment.

Figure 2. Analyses assessing the impact of *Microheliella maris* and *Rhodolphidia* on the monophyly of Pancryptista, the monophyly of Archaeplastida, and CAM clade.

Note: We regard the sister relationship between Pancryptista and Archaeplastida as CAM clade. If Pancryptista (or Cryptista) was placed within Archaeplastida, we do not consider such assemblage as CAM clade. **(A)** The maximum-likelihood (ML) tree inferred from the Diaph alignment comprising 319 genes (88,592 amino acid positions in total). Clades of closely related taxa are collapsed as triangles. For the detailed ML tree, please refer to Fig. S2. Bayesian analysis recovered principally an identical tree topology (Fig. S3). ML bootstrap support values (MLBPs) and Bayesian posterior probabilities (BPPs; if greater than 0.50) are indicated on the bipartitions presented in the figure. **(B)** Analyses of Diaph alignment processed by fast-evolving position removal (FPR). We repeated ultrafast bootstrap analyses using IQ-TREE 1.6.12 on the Diaph alignment after excluding no position, the top 20, 40, 60, and 80% fastest-evolving positions. The plots in purple, blue, green, and red indicate the ultrafast bootstrap support values (UFBPs) for the monophyly of Pancryptista, the monophyly of Archaeplastida, CAM clade, and the union of Rhodophyta and Pancryptista, respectively. **(C-K)** Analyses of the alignments generated by random gene sampling (RGS). We sampled 50, 100, 150, and 200 genes randomly from the 319 genes in the Diaph alignment, concatenated into “rs50g,” “rs100g,” “rs150g,” and “rs200g” alignments, and subjected to ultrafast bootstrap analyses using IQ-TREE 1.6.12. We presented the UFBPs for CAM clade, the monophyly of Archaeplastida, and the monophyly of Pancryptista as box-and-whisker plots **(C)**, **(F)**, and **(I)**, respectively. The above-mentioned analyses were repeated after *Rhodolphis* spp. or *M. maris* were excluded from the alignments alternatively. The UFBPs from the analyses excluding *Rhodolphis* spp. and those from the analyses excluding *M. maris* are presented in **(D)**, **(G)**, and **(J)**, and **(E)**, **(H)**, and **(K)**, respectively. The UFBPs shown in the plots described above are summarized in Table S3.

Figure 3. Analyses assessing the phylogenetic affinity of Rhodophyta to Cryptophyceae and/or Goniomonadea.

(A-L) Analyses of the alignments generated by random gene sampling (RGS). We excluded both *Rhodolphis* spp. and *Microheliella maris* from the “rs50g,” “rs100g,” “rs150g,” and “rs200g” alignments, which were generated from the Diaph alignments (see Methods for the detail) and then subjected to the ultrafast bootstrap analyses using IQ-TREE 1.6.12. The ultrafast support values (UFBPs) for the sister relationship between Archaeplastida and Cryptista, the monophyly of Archaeplastida, and the union of Rhodophyta and Cryptista are presented as box-and-whisker plots **(A)**, **(E)**, and **(I)**, respectively. The ultrafast bootstrap analyses on the rs50g, rs100g, rs150g, and rs200g alignments were repeated after further exclusion of *Palpitomonas bilix* **(B, F, and J)**, *P. bilix* and Goniomonadea **(C, G, and K)**, and *P. bilix* and Cryptophyceae **(D, H, and L)**. The UFBPs shown in the plots described above are summarized in Table S4. **(M & N)** Analyses of the alignments processed by fast-evolving position removal (FPR). We modified the Diaph alignment in two ways, (i) the exclusion of *Rhodolphis* spp., *M. maris*, *P. bilix*, and Cryptophyceae and (ii) that of *Rhodolphis* spp., *M. maris*, *P. bilix*, and Goniomonadea. The two modified Diaph alignments were processed by FPR and further subjected to ultrafast bootstrap analyses. We plotted the UFBPs for the monophyly of the SAR clade (brown), those for the monophyly of Archaeplastida (green), and those for the grouping of Rhodophyta and Goniomonadea/Cryptophyceae (red).

References

1. Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vulot, D., Minge, M., Gall, F. Le, Romari, K., Thronsen, J., Botnen, A., Massana, R., et al. (2006). Telonemia, a new protist phylum with affinity to chromist lineages. Proc. R. Soc. B Biol. Sci. 273, 1833–1842.

2. Kim, E., Simpson, A.G.B., and Graham, L.E. (2006). Evolutionary Relationships of Apusomonads Inferred from Taxon-Rich Analyses of 6 Nuclear Encoded Genes. *Mol. Biol. Evol.* 23, 2455–2466.
3. Yabuki, A., Inagaki, Y., and Ishida, K. (2010). *Palpitomonas bilix* gen. et sp. nov.: A Novel Deep-branching Heterotroph Possibly Related to Archaeplastida or Hacrobia. *Protist* 161, 523–538.
4. Burki, F. (2014). The Eukaryotic Tree of Life from a Global Phylogenomic Perspective. *Cold Spring Harb. Perspect. Biol.* 6, a016147.
5. Burki, F., Roger, A.J., Brown, M.W., and Simpson, A.G.B. (2020). The New Tree of Eukaryotes. *Trends Ecol. Evol.* 35, 43–55.
6. Keeling, P.J., and Burki, F. (2019). Progress towards the Tree of Eukaryotes. *Curr. Biol.* 29, R808–R817.
7. Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, A., Nikolaev, S.I., Jakobsen, K.S., and Pawlowski, J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2, e790.
8. Brown, M.W., Sharpe, S.C., Silberman, J.D., Heiss, A. a, Lang, B.F., Simpson, A.G.B., and Roger, A.J. (2013). Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. *Proc. Biol. Sci.* 280, 20131755.
9. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.I., Hashimoto, T., Simpson, A.G.B., et al. (2018). Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. *Genome Biol. Evol.* 10, 427–433.
10. Patterson, D.J., and Simpson, A.G.B. (1996). Heterotrophic flagellates from coastal marine and hypersaline sediments in Western Australia. *Eur. J. Protistol.* 32, 423–448.
11. Lee, W.J. (2006). Some free-living heterotrophic flagellates from marine sediments of tropical Australia. *Ocean Sci. J.* 41, 75–95.
12. Hausmann, K., Weitere, M., Wolf, M., and Arndt, H. (2002). *Meteora sporadica* gen. nov. et sp. nov. (Protista incertae sedis) – an extraordinary free-living protist from the Mediterranean deep sea. *Eur. J. Protistol.* 38, 171–177.
13. Patterson, D.J., and Zolffel, M. (1991). Heterotrophic Flagellates of Uncertain Taxonomic Position. In *The Biology of Free-living Heterotrophic Flagellates*, D. J. Patterson and J. Larsen, eds. (Oxford University Press), pp. 427–475.
14. Woodcock, H.M. (1916). Observations on coprozoic flagellates: Together with a suggestion as to the significance of the kinetonucleus in the binucleata. *Philos. Trans. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character* 207, 375–412.
15. Hoppenrath, M., and Leander, B.S. (2006). Ebriid Phylogeny and the Expansion of the Cercozoa. *Protist* 157, 279–290.
16. Yubuki, N., and Leander, B.S. (2008). Ultrastructure and molecular phylogeny of *Stephanopogon minuta*: An enigmatic microeukaryote from marine interstitial environments. *Eur. J. Protistol.* 44, 241–253.
17. Yabuki, A., and Tame, A. (2015). Phylogeny and Reclassification of *Hemistasia phaeocysticola* (Scherffel) Elbrächter & Schnepf, 1996. *J. Eukaryot. Microbiol.* 62, 426–429.
18. Silberman, J.D., Collins, A.G., Gershwin, L.-A., Johnson, P.J., and Roger, A.J. (2004). Ellobiopsids of the Genus *Thalassomyces* are Alveolates. *J. Eukaryot. Microbiol.* 51, 246–252.
19. Yazaki, E., Kume, K., Shiratori, T., Eglit, Y., Tanifuji, G., Harada, R., Simpson, A.G.B., Ishida, K., Hashimoto, T., and Inagaki, Y. (2020). Barthelonids represent a deep-branching metamonad clade with mitochondrion-related organelles predicted to generate no ATP. *Proc. R. Soc. B Biol. Sci.* 287, 20201538.

- 573 20. Yabuki, A., Kamikawa, R., Ishikawa, S.A., Kolisko, M., Kim, E., Tanabe, A.S., Kume,
574 K., Ishida, K.I., and Inagaki, Y. (2014). Palpitomonas bilix represents a basal cryptist
575 lineage: Insight into the character evolution in Cryptista. *Sci. Rep.* 4.
- 576 21. Burki, F., Kaplan, M., Tikhonenkov, D. V., Zlatogursky, V., Minh, B.Q., Radaykina, L.
577 V., Smirnov, A., Mylnikov, A.P., and Keeling, P.J. (2016). Untangling the early
578 diversification of eukaryotes: A phylogenomic study of the evolutionary origins of
579 centrohelida, haptophyta and cryptista. *Proc. R. Soc. B Biol. Sci.* 283.
- 580 22. Zhao, S., Burki, F., Brte, J., Keeling, P.J., Klaveness, D., and Shalchian-Tabrizi, K.
581 (2012). Collodictyon-an ancient lineage in the tree of eukaryotes. *Mol. Biol. Evol.* 29,
582 1557–1568.
- 583 23. Strasser, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D. V., and Burki, F. (2019).
584 New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the
585 eukaryote tree of life. *Mol. Biol. Evol.* 36, 757–765.
- 586 24. Kawachi, M., Nakayama, T., Kayama, M., Nomura, M., Miyashita, H., Bojo, O.,
587 Rhodes, L., Sym, S., Pienaar, R.N., Probert, I., et al. (2021). Rappemonads are
588 haptophyte phytoplankton. *Curr. Biol.* 31, 2395–2403.e4.
- 589 25. Gawryluk, R.M.R., Tikhonenkov, D. V., Hehenberger, E., Husnik, F., Mylnikov, A.P.,
590 and Keeling, P.J. (2019). Non-photosynthetic predators are sister to red algae. *Nature*
591 572, 240–243.
- 592 26. Yang, J., Harding, T., Kamikawa, R., Simpson, A.G.B., and Roger, A.J. (2017).
593 Mitochondrial Genome Evolution and a Novel RNA Editing System in Deep-Branching
594 Heteroloboseids. *Genome Biol. Evol.* 9, 1161–1174.
- 595 27. Heiss, A.A., Kolisko, M., Ekelund, F., Brown, M.W., Roger, A.J., and Simpson, A.G.B.
596 (2018). Combined morphological and phylogenomic re-examination of malawimonads,
597 a critical taxon for inferring the evolutionary history of eukaryotes. *R. Soc. Open Sci.* 5.
- 598 28. Lax, G., Eglit, Y., Eme, L., Bertrand, E.M., Roger, A.J., and Simpson, A.G.B. (2018).
599 Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. *Nature* 564,
600 410–414.
- 601 29. Janoušková, J., Tikhonenkov, D. V., Burki, F., Howe, A.T., Rohwer, F.L., Mylnikov,
602 A.P., and Keeling, P.J. (2017). A new lineage of eukaryotes illuminates early
603 mitochondrial genome reduction. *Curr. Biol.* 27, 3717–3724.e5.
- 604 30. Cavalier-Smith, T., Chao, E.E., and Lewis, R. (2015). Multiple origins of Heliozoa from
605 flagellate ancestors: New cryptist subphylum Corbihelia, superclass Corbistoma, and
606 monophyly of Haptista, Cryptista, Hacrobia and Chromista. *Mol. Phylogenet. Evol.* 93,
607 331–362.
- 608 31. Yabuki, A., Chao, E.E., Ishida, K.-I., and Cavalier-Smith, T. (2012). Microheliella maris
609 (Microhelida ord. n.), an Ultrastructurally Highly Distinctive New Axopodial Protist
610 Species and Genus, and the Unity of Phylum Heliozoa. *Protist* 163, 356–388.
- 611 32. Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S.,
612 Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the Classification,
613 Nomenclature, and Diversity of Eukaryotes. *J. Eukaryot. Microbiol.* 66, 4–119.
- 614 33. Shiratori, T., and Ishida, K. (2016). A New Heterotrophic Cryptomonad: Hemiarma
615 marina n. g., n. sp. *J. Eukaryot. Microbiol.* 63, 804–812.
- 616 34. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I.,
617 Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length
618 transcriptome assembly from RNA-Seq data without a reference genome. *Nat.*
619 *Biotechnol.* 29, 644–652.
- 620 35. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J.,
621 Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence

reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.

36. Nishimura, Y., Shiratori, T., Ishida, K., Hashimoto, T., Ohkuma, M., and Inagaki, Y. (2019). Horizontally-acquired genetic elements in the mitochondrial genome of a centrohelid *Marophrys* sp. SRT127. *Sci. Reports* 2019 91 9, 1–10.

37. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066.

38. Katoh, K., and Standley, D.M. (2014). MAFFT: Iterative refinement and additional methods. *Methods Mol. Biol.* 1079, 131–146.

39. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* 26, 1641–50.

40. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5.

41. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.

42. Wang, H.C., Minh, B.Q., Susko, E., and Roger, A.J. (2018). Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* 67, 216–235.

43. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* 21, 1095–1109.

44. Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Syst. Biol.* 62, 611–615.

45. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522.

46. Inagaki, Y., Susko, E., Fast, N.M., and Roger, A.J. (2004). Covarion Shifts Cause a Long-Branch Attraction Artifact That Unites Microsporidia and Archaeobacteria in EF-1 α Phylogenies. *Mol. Biol. Evol.* 21, 1340–1349.

47. Schön, M.E., Zlatogursky, V. V., Singh, R.P., Poirier, C., Wilken, S., Mathur, V., Strassert, J.F.H., Pinhassi, J., Worden, A.Z., Keeling, P.J., et al. (2021). Picozoa are archaeplastids without plastid. *bioRxiv*, 2021.04.14.439778.

48. Irisarri, I., Strassert, J.F.H., and Burki, F. (2021). Phylogenomic Insights into the Origin of Primary Plastids. *Syst. Biol.* *in press*.

49. Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* 21, 163–193.

50. Cenci, U., Sibbald, S.J., Curtis, B.A., Kamikawa, R., Eme, L., Moog, D., Henrissat, B., Maréchal, E., Chabi, M., Djemiel, C., et al. (2018). Nuclear genome sequence of the plastid-lacking cryptomonad *Goniomonas avonlea* provides insights into the evolution of secondary plastids. *BMC Biol.* 16.