LETTER

https://doi.org/10.1038/s41586-019-1398-6

Non-photosynthetic predators are sister to red algae

Ryan M. R. Gawryluk[1,3,5]*, Denis V. Tikhonenkov[1,2,5]*, Elisabeth Hehenberger[1,4], Filip Husnik[1], Alexander P. Mylnikov[2] & Patrick J. Keeling[1]*

Rhodophyta (red algae) is one of three lineages of Archaeplastida[1], a supergroup that is united by the primary endosymbiotic origin of plastids in eukaryotes[2,3]. Red algae are a diverse and species-rich group, members of which are typically photoautotrophic, but are united by a number of highly derived characteristics: they have relatively small intron-poor genomes, reduced metabolism and lack cytoskeletal structures that are associated with motility, flagella and centrioles. This suggests that marked gene loss occurred around their origin[4]; however, this is difficult to reconstruct because they differ so much from the other archaeplastid lineages, and the relationships between these lineages are unclear. Here we describe the novel eukaryotic phylum Rhodelphidia and, using phylogenomics, demonstrate that it is a closely related sister to red algae. However, the characteristics of the two *Rhodelphis* species described here are nearly opposite to those that define red algae: they are non-photosynthetic, flagellate predators with gene-rich genomes, along with a relic genome-lacking primary plastid that probably participates in haem synthesis. Overall, these findings alter our views of the origins of Rhodophyta, and Archaeplastida evolution as a whole, as they indicate that mixotrophic feeding—that is, a combination of predation and phototrophy—persisted well into the evolution of the group.

Two previously undescribed eukaryovorous protists, *Rhodelphis limneticus* and *Rhodelphis marinus* (see Supplementary Information for taxonomic diagnosis), were isolated from a freshwater lake and marine coral sand, respectively. *Rhodelphis* are 10–13 μm, oval or tapered, slightly flattened cells with two subapical heterodynamic flagella (Fig. 1a–f and Extended Data Figs. 1a, 2a, b). Characteristic morphological features include umbrella-shaped glycostyles on the cell surface and flagellum (Fig. 1g, h, s and Extended Data Figs. 1d, 2c–e), perpendicularly oriented flagellar basal bodies (Fig. 1i, j and Extended Data Figs. 1e, k, 1, 2g, h) with outgoing striated structures and at least two fibrils (Fig. 1j–l and Extended Data Fig. 2e), one narrow and two wide microtubular bands (Fig. 1l–q and Extended Data Fig. 1h–l), a flagellar transition zone with a transverse plate at the cell surface and a proximal diaphragm through which a central pair of flagellar microtubules surrounded by a cylinder passes (Fig. 1q, r and Extended Data Fig. 1e–h), a sac-shaped double-layered smooth endoplasmic reticulum (Fig. 1 s–u and Extended Data Fig. 1o) and mitochondria with tubular cristae (Fig. 1 s–u and Extended Data Fig. 2g, h, j). Plastids were not observed.

To establish the evolutionary position of *Rhodelphis*, we sequenced transcriptomes from cultures and manually isolated cells, and generated a concatenated 153-taxon/253-protein supermatrix (153/253 dataset; 56,312 sites). Maximum-likelihood and Bayesian analyses recovered *Rhodelphis* as a well-supported sister to the red algae, with ultrafast bootstrap support of 97%, Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) support of 0.99 and a Bayesian posterior probability of 0.98 (Extended Data Fig. 3a, b). Analyses of a second supermatrix without the picozoan MS584-11[5] and *Telonema* (both of which had poor datasets; 151-taxon/253-protein supermatrix (151/253 dataset), 56,530 sites), recovered *Rhodelphis* and red algae with complete statistical support (Fig. 2a, b and Extended Data Fig. 4a, b), although approximately unbiased tests were unable to distinguish between Archaeplastida monophyly or paraphyly ($P = 0.6693$ and $P = 0.3397$, respectively). To examine the possibility that long-branch attraction affected the position of *Rhodelphis*, we carried out fast-site removal analyses, which showed that support for the sisterhood of *Rhodelphis* and red algae remained high for both 153/253 (Extended Data Fig. 3c) and 151/253 datasets (Fig. 2c). To test whether mixed gene ancestry (for example, from horizontal gene transfer) affected the phylogenetic placement of *Rhodelphis*, we calculated internode certainty values[6] from 253 bootstrapped single-gene trees and the 50 trees with the highest relative tree certainty. These analyses showed a degree of conflict that is similar to other ancient but well-established relationships (such as opisthokont monophyly) and that the internode certainty scores for the *Rhodelphis* and red algae bipartition increase in the 50 best-supported trees (Extended Data Fig. 5a, b), which also recover *Rhodelphis* as sisters to the red algae in a concatenated phylogeny (Extended Data Fig. 6). Coalescent species trees[7] estimated from the 253 and 50 single-gene tree datasets also recover *Rhodelphis* and red algae as sisters, with full support (Extended Data Fig. 7a, b).

Most Archaeplastida are photoautotrophic and phagotrophy is very rare. However, phagotrophy must have existed for the archaeplastid ancestor to take up the plastid, and must have persisted at least until the protoplastid became a reliable source of both energy and nutrients. This suggests a key mixotrophic intermediate stage; however, because the few known phagotrophic archaeplastids have been interpreted as secondarily derived, it has been widely assumed that phagotrophy was lost early in the evolution of archaeplastids[8]. We therefore characterized the *Rhodelphis* genome and transcriptomes further to investigate the ancestral states of phagotrophy and other characteristics that are seemingly absent from red algal and archaeplastid ancestors. Altogether, these analyses suggest that *Rhodelphis* possess larger and more gene-rich nuclear genomes than most red algae (Extended Data Table 1a–c) and genes with many more introns (nearly 40,000 spliceosomal introns were identified in genome–transcriptome comparisons in *R. limneticus*). But the most notable differences between *Rhodelphis* and red algae are in gene content: *Rhodelphis* possess genes that are associated with centrioles, autophagy and synthesis of glycosylphosphatidylinositol, all of which are absent from red algae[9]. Notably, genes that encode flagellar proteins are absent from red algae, whereas *Rhodelphis* encode homologues of 209 out of 361 high-confidence *Chlamydomonas* flagellar proteins[10], consistent with our microscopic observations.

Rhodelphis engulf whole bacteria and eukaryotic prey at the posterior end (Extended Data Fig. 1r–t and Supplementary Video 1), but a distinct feeding apparatus is not evident. Phagotrophy in *Rhodelphis* therefore differs from feeding in some prasinophytes (the only other known class in which phagotrophy occurs within Archaeplastida), which use a mouth-like opening, a tubular channel and a large permanent vacuole to engulf, transport and digest bacterial cells[11]. Specific genetic markers of phagotrophy are difficult to define; however, genome-level predictive models[12] indicate that the genetic repertoire of *Rhodelphis* is consistent with phagocytotic feeding (Extended Data Fig. 8).

[1]Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada. [2]Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences, Borok, Russia. [3]Present address: Department of Biology, University of Victoria, Victoria, British Columbia, Canada. [4]Present address: GEOMAR – Helmholtz Centre for Ocean Research Kiel, Division of Experimental Ecology, Kiel, Germany. [5]These authors contributed equally: Ryan M. R. Gawryluk, Denis V. Tikhonenkov. *e-mail: ryangawryluk@uvic.ca; tikho-denis@yandex.ru; pkeeling@mail.ubc.ca

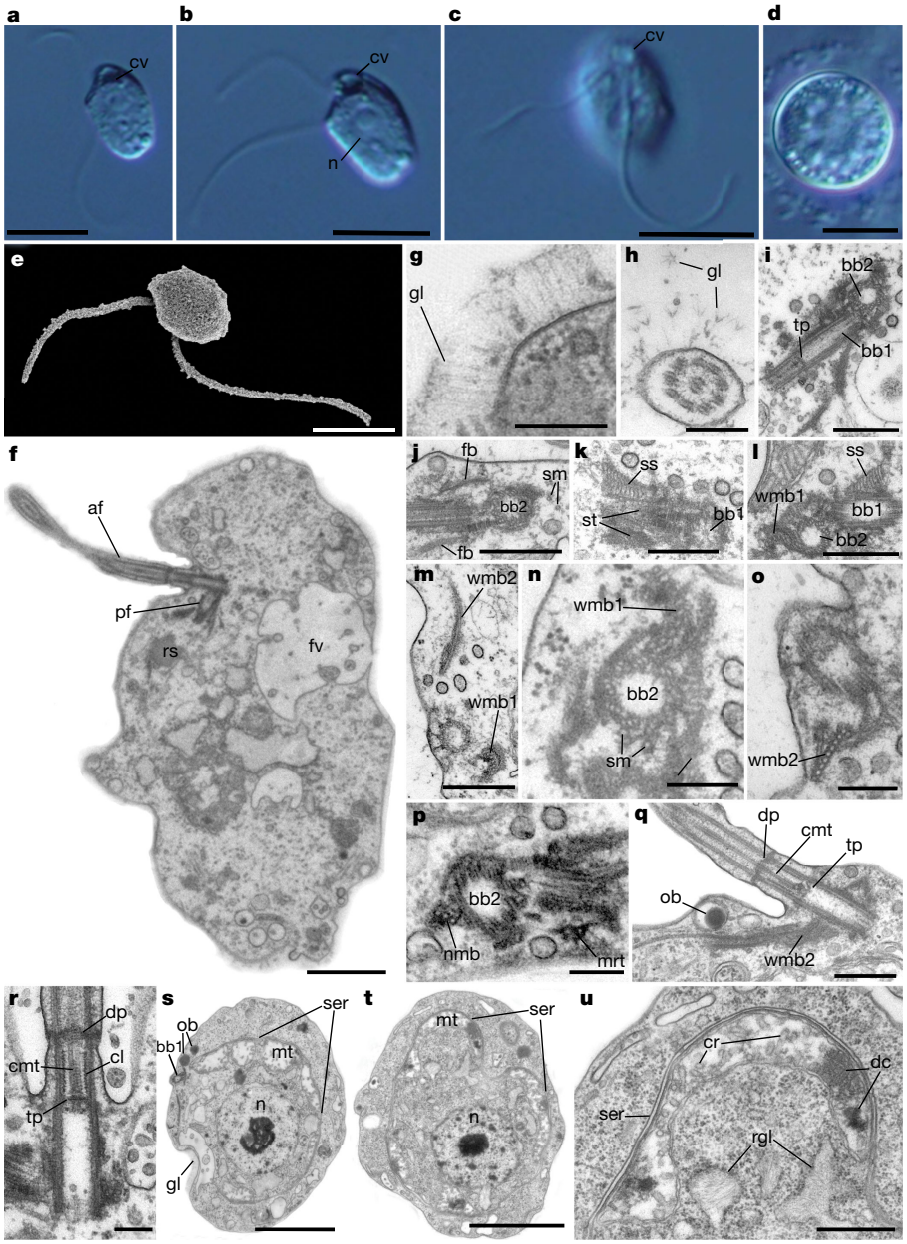


Fig. 1 | Cell morphology of *R. limneticus*. **a**–**c**, Living cells, visualized by light microscopy. **d**, Cyst, visualized by light microscopy. **e**, Scanning electron microscopy image highlighting the flagella. **f**–**u**, Cells, visualized by transmission electron microscopy. **f**, Longitudinal section. **g**, Glycostyles can be seen on the cell surface. **h**, Transverse section of the posterior flagellum covered with glycostyles. **i**–**l**, Arrangement of basal bodies, connecting structures and satellites. **m**–**p**, Arrangement of microtubular bands, microtubules and satellites. **q**, **r**, Structure of the flagellar transition zone. **s**, **t**, Transverse sections of the cell showing the sac-shaped smooth endoplasmic reticulum, nucleus and mitochondria. **u**, Area near the endoplasmic reticulum showing a single mitochondrion with dark condensations and vesicles with rudiments of glycostyles. af, anterior flagellum; cl, cylinder; cv, contractile vacuole; cr, cristae; cmt, central microtubules; dc, dark condensation; dp, diaphragm; fb, fibril; fv, food vacuole; gl, glycostyles; bb1, basal body of posterior flagellum; bb2, basal body of anterior flagellum; mt, mitochondrion; mrt, microtubule; n, nucleus; nmb, narrow microtubular band; ob, osmiophilic body; pf, posterior flagellum; rgl, rudiments of glycostyles; rs, reserve substance; ser, sac of smooth endoplasmic reticulum; sm, single microtubules; ss, striated structure; st, satellite of basal body; tp, transverse plate; wmb1, wide microtubular band 1; wmb2, wide microtubular band 2. Scale bars, 10 μm (**a**–**c**), 5 μm (**d**, **e**), 1 μm (**f**), 0.2 μm (**g**, **h**, **n**–**p**, **r**), 0.5 μm (**i**–**m**, **q**, **u**) and 2 μm (**s**, **t**). These experiments were repeated 50 (**a**–**d**), 3 (**e**) and 7 (**f**–**u**) times, with similar results.

Taken together, the cellular motility and phagotrophic feeding by *Rhodelphis* and the ancestral photosynthetic capacity of red algae indicates that their common ancestor was mixotrophic.

Although an ancestor of *Rhodelphis* must have been photosynthetic, no plastids were observed using microscopy, so we searched for genetic evidence of a relic plastid, as complete loss of the plastid is exceedingly rare[13,14]. In archaeplastids, nucleus-encoded proteins are targeted to plastids through N-terminal transit-peptide leaders that are recognized by TIC/TOC import complexes. We identified homologues of several plastid import proteins in *Rhodelphis*, including TIC20, TIC22, TIC32 and TOC75 (Extended Data Fig. 9b–e), as well as many proteins with a putative plastid function (Fig. 3a and Supplementary Table 1). Notably, plastid proteins of *Rhodelphis* encode leader sequences that are similar to archaeplastid transit peptides (Extended Data Fig. 9a); they do not have bipartite leader sequences or specific homologues of SELMA complex subunits[15] that are indicative of protein targeting to complex plastids by the endoplasmic reticulum. Taken together, the evidence from the analysis of the phylogenomics, plastid-targeting leaders and plastid import machinery is consistent with the conclusion that *Rhodelphis* have a primary plastid as with other members of Archaeplastida.

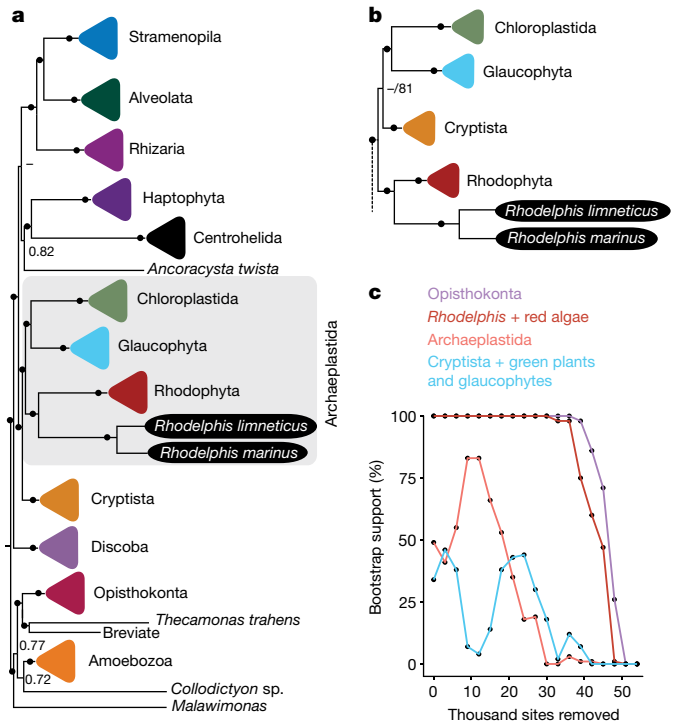


Fig. 2 | The evolutionary position of *Rhodelphis*. **a**, **b**, Bayesian (**a**; CAT + GTR)[20] and maximum-likelihood (**b**; 151/253 dataset; LG + C60 + F + G4)[21] analyses place *Rhodelphis* as a sister to red algae (Rhodophyta). Bayesian analyses also recover the monophyly of Archaeplastida, whereas maximum-likelihood analyses do not. Black dots denote full statistical support for the Bayesian tree (**a**; Bayesian posterior probability = 1.0) or maximum-likelihood tree (**b**; maximum-likelihood ultrafast bootstrap = 100%/SH-aLRT = 1.0); support values below 1.0 or 100% are shown in cases in which either value is below this value; support values <0.7 or 70% are not shown (indicated by a '–'). Complete Bayesian and maximum-likelihood trees for 151/253 and 153/253 datasets are shown in Extended Data Figs. 3 and 4, respectively. **c**, Bootstrap support for maximum-likelihood trees (PROT + CAT + LG + F) after progressive removal of the fastest-evolving amino acid sites (151/253 supermatrix) shows that the *Rhodelphis* and red algae relationship is robust, but that Archaeplastida paraphyly (Cryptista, and green plants and glaucophytes) may be the result of long branch attraction. Support for Opisthokonta monophyly serves as a control for the presence of sufficient information for phylogenomic inference. A parallel analysis for 153/253 is shown in Extended Data Fig. 3c.

Consistent with their lack of pigmentation, *Rhodelphis* encode almost no proteins that are involved in photosynthesis. The exceptions are ferredoxin and ferredoxin NADP$^+$ reductase, which probably participate in FeS cluster assembly through the plastidial SUF-type (sulfur assimilation) biosynthesis pathway (Fig. 3a). A mitochondrial-type iron sulfur cluster (ISC) pathway is also present in *Rhodelphis*, which suggests that the SUF system supports only plastid FeS proteins.

We found no evidence for the type II fatty acid biosynthesis pathway that is typical of plastids. Instead, *Rhodelphis* synthesize fatty acids in the cytosol, through a multidomain type I fatty acid synthase, which is found in numerous animals, fungi and protists, but not in archaeplastids (Fig. 3a). Furthermore, *Rhodelphis* also lack the plastid methylerythritol phosphate (MEP) isoprenoid biosynthetic pathway, and use the cytosolic mevalonate (MVA) pathway instead (Fig. 3a). The MVA pathway is patchily distributed among red algae, and has been independently lost numerous times[9], whereas the MEP pathway is found in most archaeplastids that have been examined to date. The distribution of both pathways highlights the potential for long-term functional redundancy to shape gene loss and metabolic reorganization.

By contrast, most *Rhodelphis* haem biosynthetic enzymes are probably targeted to plastids, with the exception of the first step that is catalysed by δ-aminolevulinic acid synthase (ALAS), which is probably a mitochondrial protein (Fig. 3a); this is probably the major reason that *Rhodelphis* plastids have not been lost. All of the putatively plastid-targeted haem proteins are of the typical plastid-type (phylogenetic trees are available through the Dryad Digital Repository), except ferrochelatase (HemH), which branches with homologues from some red algae and dinoflagellates that are in turn nested within a group

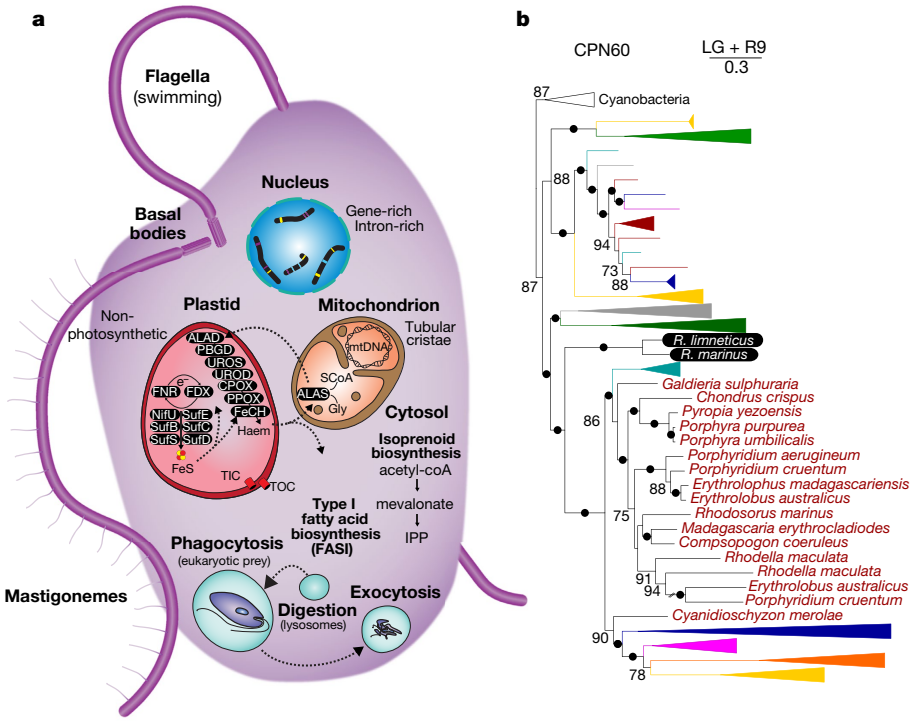


Fig. 3 | Genomic insights into *Rhodelphis* cell biology. **a**, *Rhodelphis* genome and transcriptome analyses reveal a relic plastid that generates FeS clusters and probably cooperates with mitochondria in haem biosynthesis, cytosolic isoprenoid and fatty acid synthesis, extensive nuclear mRNA splicing (coloured lines on chromosomes), phagocytosis (Extended Data Fig. 8 and Supplementary Video 1), basal bodies and flagella (a long flagellum was shortened to fit). Note that subcellular structures are not drawn to scale and the mitochondrial genome has not been verified to be circular. **b**, Maximum-likelihood phylogeny of Cpn60 with a predicted N-terminal transit peptide (Extended Data Fig. 9), indicating the presence of a primary plastid in *Rhodelphis*, although none were observed by microscopy. *Rhodelphis* Cpn60 groups basal to homologues from red algae and other eukaryotes with plastids derived from red algae (coloured wedges). Other putative plastid proteins are presented in Supplementary Table 1. IPP, isopentenyl pyrophosphate. Eukaryotic groups are indicated by coloured taxon names/collapsed clades: green, Viridiplantae; red, red algae; grey, glaucophytes; turquoise, cryptophytes; pink, stramenopiles; dark blue, haptophytes; yellow, dinoflagellates; orange, chrompodellids and apicomplexans; brown, opisthokonts, amoebozoans, apusozoans, ancyromonads and ciliates. Taxa or clades outlined in black are prokaryotic.

of bacterial homologues. Other red algae and lineages with red algal plastids—such as diatoms, cryptophytes and ochrophytes—have a typical cyanobacterial HemH[16]. This again suggests that there is redundancy of HemH types in the common ancestor of *Rhodelphis* and red algae.

Despite finding many nucleus-encoded putatively plastid-targeted proteins, we found no evidence that supports the existence of a plastid genome. No plastid DNA is present in the genomic datasets of *R. limneticus*, which were generated from over 300 million paired reads from both cultures and single cells. By contrast, mitochondrial DNA was readily identified; the genome could not be unambiguously assembled, but is >100 kilobases and individual contigs were connected in a single contig map. Similarly, no nucleus-encoded components of plastid genome replication, gene expression or translation systems were identified in any *Rhodelphis* dataset. Notably, many of the nucleus-encoded, putatively plastid-targeted *Rhodelphis* proteins that we did identify are encoded in plastid genomes of red algae. Taken together, we interpret these observations as strong evidence for the complete loss of plastid DNA in *Rhodelphis*. This has only been reported in a few non-photosynthetic plastids[17,18] and is in contrast to the gene-rich nature of the plastid genomes of red algae[19].

In conclusion, phylogenomic analyses strongly support the placement of *Rhodelphis* as a sister lineage to red algae. *Rhodelphis* are flagellate predators with primary, non-photosynthetic plastids that are involved in haem biosynthesis; all of which indicates that the ancestor of *Rhodelphis* and red algae was very different from previous models of the ancestors of red algae. This ancestor was probably a mixotrophic flagellate that obtained energy and nutrients from both photosynthetic plastids and phagotrophy, which suggests that phagotrophy persisted within Archaeplastida until well after the divergence of red algae from green plants and glaucophytes. The gene- and intron-rich genomes and complex pattern of biosynthetic pathway retention also provide insights into the potential for functional redundancy to persist over considerable periods of evolutionary time, followed by differential loss. Indeed, *Rhodelphis* reveals that the absence of phagotrophy and many other characteristics of the Archaeplastida as a whole are due to multiple convergent losses rather than an already established ancestral state.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; andstatements of data and code availability are available at at https://doi.org/10.1038/s41586-019-1398-6.

Received: 27 February 2019; Accepted: 13 June 2019;
Published online: 17 July 2019

1. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016147 (2014).
2. Archibald, J. M. The puzzle of plastid evolution. *Curr. Biol.* **19**, R81–R88 (2009).
3. Keeling, P. J. The number, speed, and impact of plastid endosymbioses in eukaryotic evolution. *Annu. Rev. Plant Biol.* **64**, 583–607 (2013).
4. Qiu, H., Price, D. C., Yang, E. C., Yoon, H. S. & Bhattacharya, D. Evidence of ancient genome reduction in red algae (Rhodophyta). *J. Phycol.* **51**, 624–636 (2015).
5. Yoon, H. S. et al. Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**, 714–717 (2011).
6. Salichos, L. & Rokas, A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331 (2013).
7. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
8. Spiegel, F. W. Contemplating the first Plantae. *Science* **335**, 809–810 (2012).
9. Qiu, H., Yoon, H. S. & Bhattacharya, D. Red algal phylogenomics provides a robust framework for inferring evolution of key metabolic pathways. *PLoS Curr.* **8**, https://doi.org/10.1371/currents.tol.7b037376e6d84a1be34af756a4d90846 (2016).
10. Pazour, G. J., Agrin, N., Leszyk, J. & Witman, G. B. Proteomic analysis of a eukaryotic cilium. *J. Cell Biol.* **170**, 103–113 (2005).
11. Maruyama, S. & Kim, E. A modern descendant of early green algal phagotrophs. *Curr. Biol.* **23**, 1081–1084 (2013).
12. Burns, J. A., Pittis, A. A. & Kim, E. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat. Ecol. Evol.* **2**, 697–704 (2018).
13. Gornik, S. G. et al. Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. *Proc. Natl Acad. Sci. USA* **112**, 5767–5772 (2015).
14. Xu, P. et al. The genome of *Cryptosporidium hominis*. *Nature* **431**, 1107–1112 (2004).
15. Gould, S. B., Maier, U.-G. & Martin, W. F. Protein import and the origin of red complex plastids. *Curr. Biol.* **25**, R515–R521 (2015).
16. Oborník, M. & Green, B. R. Mosaic origin of the heme biosynthesis pathway in photosynthetic eukaryotes. *Mol. Biol. Evol.* **22**, 2343–2353 (2005).
17. Smith, D. R. & Lee, R. W. A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiol.* **164**, 1812–1819 (2014).
18. Fernández Robledo, J. A. et al. The search for the missing link: a relic plastid in *Perkinsus*? *Int. J. Parasitol.* **41**, 1217–1229 (2011).
19. Muñoz-Gómez, S. A. et al. The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Curr. Biol.* **27**, 1677–1684 (2017).
20. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
21. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Cell isolation and culture establishment. *R. marinus* (clone Colp-29) was obtained from near-shore marine coral sand off the coast of a small island, Bay Canh, near Con Dao Island, South Vietnam (8° 39′ 58.6362″ N, 106° 40′ 49.7562″ E). The sample was collected from a depth of 40 cm on 3 May 2015.

R. limneticus (clone Colp-38) was obtained from a near-shore freshwater sample including organic debris, taken by Y. V. Dubrovsky (IEE NAS Ukraine) on 9 August 2016 from Lake Trubin (floodlands of Desna River, 51° 23′ 49.9986″ N, 32° 22′ 8.0004″ E), near Yaduty village, Chernigovskaya oblast, Ukraine.

The samples were examined on the third, sixth and ninth day of incubation in accordance with previously described methods[22]. Following isolation by glass micropipette, *R. marinus* and *R. limneticus* were propagated on the bodonids *Procryptobia sorokini* strain B-69 and *Parabodo caudatus* strain BAS-1, respectively, which were grown in marine Schmalz–Pratt's medium and spring water (Aqua Minerale, PepsiCo or PC Natural Spring Water, President's Choice) using the bacterium *Pseudomonas fluorescens* as food[23]. *R. limneticus* is currently being stored in a collection of live protozoan cultures at the Papanin Institute for Biology of Inland Waters, Russian Academy of Sciences; however, *R. marinus* perished after several months of cultivation.

Light and electron microscopy. Light microscopy observations of *R. limneticus* were made using a Zeiss AxioScope A.1 equipped with a differential interference contrast (DIC) water-immersion objective (63×) and an AVT HORN MC-1009/S analogue video camera. Observations of *R. marinus* were made using a Zeiss Axioplan 2 Imaging microscope equipped with a DIC objective (40×) and a Canon XL H1S video camera.

For scanning electron microscopy, cells from a culture in exponential growth phase were fixed with 2.5% glutaraldehyde (final concentration). The cells were mounted on a glass coverslip coated with poly-L-lysine for 30 min and subsequently rinsed three times with 0.1 M sodium cacodylate buffer (pH 7.34), which was diluted twice with spring water (PC Natural Spring Water, President's Choice). Next, cells were fixed in 1% osmium tetroxide for 1 h. The fixed cells were rinsed three times with distilled water, 10 min each time, and dehydrated with a graded ethanol series from 30% to absolute ethanol (10 min per step), followed by 100% hexamethyldisilazane (three times, 15 min each) and dried at 65 °C. Dry glass coverslips were mounted on aluminium stubs, coated with gold–palladium, and observed with a Hitachi S4700 scanning electron microscope (Hitachi High-Technologies Corporation).

For transmission electron microscopy, cells were centrifuged, fixed in a cocktail of 0.6% glutaraldehyde and 2% osmium tetroxide (final concentration) prepared using a 0.1 M cacodylate buffer (pH 7.2) for freshwater cells, or Schmaltz–Pratt medium for marine cells at 1 °C for 30–60 min and dehydrated in an alcohol and acetone series (30, 50, 70, 96 and 100%; 20 min per step). Finally, cells were embedded in a mixture of Araldite and Epon[24]. Ultrathin sections were obtained with an LKB ultramicrotome. Transmission electron microscopy observations were obtained using a JEM-1011 (JEOL) electron microscope.

Preparation of libraries and sequencing. *RNA and genomic DNA isolation.* Cells grown in clonal laboratory cultures were collected when the cultures had reached peak abundance and after the prey had been eaten (based on daily light microscopy observations). Cells were collected by centrifugation (1,000*g*, room temperature) onto a 0.8-μm membrane of a Vivaclear mini column (Sartorius Stedim Biotech, VK01P042); this was done separately for RNA and DNA extractions. Total RNA was then extracted using an RNAqueous-Micro Kit (Invitrogen, AM1931) and converted into cDNA using the Smart-Seq2 protocol[25]. Additionally, cDNA of *R. limneticus* was obtained from 20 single cells using the Smart-Seq2 protocol: cells were manually picked from the culture using a glass micropipette and transferred to a 0.2-ml thin-walled PCR tube containing 2 μl cell lysis buffer (0.2% Triton X-100 and RNase inhibitor (Invitrogen)). Total DNA was extracted from the filters using the MasterPure Complete DNA and RNA Purification Kit (Epicentre, MC85200).

The small subunit (SSU) rRNA genes of *R. marinus* and *R. limneticus* were amplified by polymerase chain reaction (PCR) using the general eukaryotic primers GGF (CTTCGGTCATAGATTAAGCCATGC) and GGR (CCTTGTTACGACTTCTCCTTCCTC) and 18SFU and 18SRU[26], respectively. PCR products were subsequently cloned (*R. marinus*) or sequenced directly (*R. limneticus*) using Sanger dideoxy sequencing.

R. limneticus transcriptome sequencing was performed on the Illumina MiSeq platform with read lengths of 300 bp using the NexteraXT protocol (Illumina, FC-131-1024) to construct paired-end libraries. *R. marinus* transcriptome sequencing was performed on the Illumina HiSeq platform (UCLA Clinical Microarray Core) with read lengths of 100 bp using the KAPA stranded RNA-seq kit (Roche) to construct paired-end libraries.

R. limneticus DNA was extracted from cultures (containing *R. limneticus*, prey and bacteria) using the MasterPure DNA Purification Kit (Epicentre) or obtained from three individually picked cells through whole-genome amplification using the TruePrime Single Cell WGA kit v.2.0 (Expedeon) according to the manufacturer's instructions. Libraries were generated at The Centre of Applied Genomics and 151-bp paired-end reads were sequenced on an Illumina HiSeq X. Whole-genome amplified DNA (TruePrime Single Cell WGA kit v.2.0 (Expedeon)) was sequenced on MinION, Oxford Nanopore Technologies using the Ligation Sequencing Kit 1D (SQK-LSK108, Oxford Nanopore Technologies). Covaris shearing was omitted to preserve long fragments. DNA was initially treated with T7 endonuclease to remove extremely branched DNA structures resulting from whole-genome amplification.

Sequencing dataset assembly and decontamination. Sequence quality and adaptor contamination of reads from transcriptomic datasets were assessed with FastQC[27]. Reads were trimmed with Trimmomatic-0.32 ILLUMINACLIP[28], with a maximum of two mismatches, a palindromeClipThreshold of 30 and a simpleClipThreshold of 10. Low-quality sequences were discarded, using a sliding window of 4 bp, a minimum quality score of 25 and a minimum trimmed length of 35 bp.

A strand-specific *R. marinus* HiSeq transcriptome was assembled with Trinity v.2.0.6[29], with the --SS_lib_type flag set to RF. MiSeq transcriptomes of *R. limneticus* from culture or 20-cell preparations were assembled in essentially the same manner, but without strand specificity. Transdecoder was used to infer the most likely open reading frame sequences, informed by blastp[30] and hmmscan queries of the Swissprot and Pfam databases, respectively (*E*-value cut-off = 1×10^{-5}). CD-HIT[31] was used to reduce the redundancy of the inferred protein dataset by clustering proteins with ≥95% identity. Extensive in silico decontamination was performed to remove sequences derived from the eukaryotic prey *P. sorokini* (*R. marinus*), *P. caudatus* (*R. limneticus*) and co-cultured bacteria. We used megablast to identify transcripts that were ≥95% identical to sequences of previously generated *P. sorokini* and *P. caudatus* MiSeq transcriptome datasets, along with previously generated HiSeq transcriptome datasets from protists that feed on *P. sorokini* or *P. caudatus* (that is, sequences that these datasets had in common were probably derived from *P. sorokini* or *P. caudatus*). Contaminating bacterial sequences were identified using megablast and blastp queries of the NCBI nt and nr databases. Nucleotide sequences that were ≥80% identical to bacterial entries were removed. For a protein sequence from the *Rhodelphis* datasets to be classified as bacterial, each of the top-15 blastp hits had to be most similar to a bacterial homologue and ≥70% identical to a bacterial protein. Assessment of transcriptome completeness was performed by searching 'eukaryote' protein datasets with BUSCO v.3.0.1[32], using default parameters.

The *R. limneticus* genome was assembled from HiSeq X reads (either culture or whole-genome-amplified DNA from single cells (WGA)) and MinION reads using SPAdes v.3.11.1[33] with kmer lengths of 21, 33, 55, 77, 99 and 121, and with the '–sc' flag activated for the single-cell assembly. For each of the WGA and culture genome assemblies, contigs from co-cultured contaminants (for example, kinetoplastid prey and bacteria) were identified using Autometa[34] and removed. As expected, the culture dataset was heavily contaminated, whereas WGA contigs were predominantly from *R. limneticus*. Assessments of genome assembly were performed using QUAST v.5.0.2[35].

A search for putative spliceosomal introns in *R. limneticus* was performed by aligning transcripts to the culture assembly, requiring a minimum 95% identity threshold, using GMAP v.2016-08-16[36]. Spliceosomal introns with GT/AG splice boundaries were extracted from the GMAP output with a custom Python script and visualized with WebLogo[37]. The total proportion of transcripts from *R. limneticus* mapping to the nuclear genome sequence was determined with isoblat v.3[38].

Searches for a plastid genome were carried out by querying *Rhodelphis* transcriptome and genome datasets with red algal plastid-encoded RNAs and proteins, and by searching for rRNAs of plastidial/cyanobacterial affinity with phyloFlash v.3.0[39]. All plastid-type proteins that were found are probably encoded by DNA in the nucleus and no plastidial/cyanobacterial rRNAs were found.

Phylogenomic dataset preparation and analysis. Construction of a phylogenomic supermatrix was performed essentially as previously described[40], using nearly the same dataset. In brief, blastp was used to identify *Rhodelphis* homologues, with an expect value threshold of $\leq 1 \times 10^{-30}$. Alignments were generated with MAFFT L-INS-I v.7.212[41] and trimmed automatically with BMGE v.1.12[42] based on the BLOSUM75 substitution matrix. Single-protein maximum-likelihood phylogenies—derived from 20 independent heuristic searches with RAxML v.8.1.6[43] using the PROTGAMMALGF model—were used to screen for paralogues and sequences that were probably derived from prey contamination. Individual trimmed alignments were concatenated with SCaFOs v.1.2.5[44], requiring a minimum of 15% coverage for inclusion. The final concatenated alignment included 153 taxa, 253 proteins and 56,312 amino acid sites (153/253 dataset); *R. marinus* and *R. limneticus* were well-represented (*R. marinus*, 98% of genes and 99% of sites; *R. limneticus*, 94% of genes and 93% of sites). Another alignment

was generated without the picozoan MS584-11 and *Telonema*, leaving 151 taxa, 253 proteins, and 56,530 sites (151/253 dataset); *R. marinus* and *R. limneticus* were again well represented (*R. marinus*, 98% of genes and 99% of sites; *R. limneticus*, 94% of genes and 93% of sites).

Maximum-likelihood phylogenomic tree reconstruction was performed using IQ-TREE v.1.5.5[21] with the LG matrix combined with the C60 protein mixture model and four gamma categories (that is, LG + C60 + F + G4). The results of 1,000 ultrafast bootstrap and SH-aLRT replicates are reported as a measure of statistical support for bipartitions. Bayesian analyses were carried out by running three independent Markov chain Monte Carlo chains with PhyloBayes MPI v.1.7[20], using an infinite mixture model and four discrete gamma categories (CAT + GTR + G4). Chains were run for more than 7,200 generations for both the 153/253 and 151/253 datasets, and the first 1,500 generations were discarded as burn-in. As is frequently seen in large-scale phylogenomic analyses, the chains failed to converge in each case; for the 153/253 dataset, maxdiff = 1 and meandiff = 0.0128946, and for the 151/253 dataset, maxdiff = 1 and meandiff = 0.0204329. Bayesian posterior probabilities are reported as a measure of statistical support for bipartitions.

The effect of fast-evolving sites on phylogenomic inference was examined by progressively removing the fastest-evolving sites in intervals of 3,000 and using RAxML (PROT + CAT + LG + F model) to generate 100 rapid-bootstrap replicates for each alignment. The fastest-evolving sites were identified using AgentSmith, based on the tree topologies generated in the IQ-TREE LG + C60 + F + G4 153- and 151-taxon trees. The support for the sister relationship of *Rhodelphis* and red algae was assessed for both trees; support for Archaeplastida monophyly versus Cryptista, glaucophytes and green plants (Cryptista and GG) was similarly tested for both datasets. Support for Picozoa, *Rhodelphis* and red algae was assessed only for the 153-taxon tree. In each case, the monophyly of Opisthokonta was tested as a positive control. Alternative tree topologies were generated from the maximum-likelihood trees using Mesquite v.3.5[45] and tested using the approximately unbiased test, as implemented in IQ-TREE v.1.5.5[21].

We carried out a number of analyses to test whether mixed gene ancestry could be affecting the phylogenetic analysis. RAxML[43] was used to compute internode certainty scores[6] by mapping bootstrapped single-gene trees (generated as above, except with the PROTCATLGF model) to the 151/253 maximum-likelihood phylogenomic tree (Extended Data Fig. 4b). We subsequently identified the 50 single-gene trees with the highest relative tree certainty (RTC) values with RAxML[43] (average RTC score for all 253 single-gene trees is 0.175; average for best 50 is 0.362), and computed internode certainty scores for this subset of the data, as above. To test whether the 50 highest RTC single-gene datasets recover the sister relationship of *Rhodelphis* and red algae, we concatenated the alignments with SCaFOs v.1.2.5[44] and performed phylogenetic analysis in IQ-TREE using the LG + PMSF + G model along with 1,000 ultrafast bootstrap and SH-aLRT replicates. ASTRAL-III[7] was used to calculate overall species trees from individual bootstrapped trees under a 'coalescence' framework, using default parameters and 100 multilocus bootstrap replicates.

Identification of putative plastid-targeted proteins. A screen for *Rhodelphis* proteins bearing putative N-terminal plastid transit peptides was performed with TargetP[46], with the 'Plant' flag activated. Proteins with a score corresponding to a 70% probability of plastid localization were retained for further manual inspection. Additionally, we queried the NCBI nr database with *Rhodelphis* protein sequences in search of proteins that were most similar to characterized plastidial or cyanobacterial homologues, and queried the *Rhodelphis* genome and transcriptome assemblies with proteins encoded by red algal plastid genomes. We considered proteins to potentially be plastidial if they fulfilled both of the above criteria, or if they were constituents of a metabolic unit of members which predominantly met the criteria.

Plastid protein phylogenies. Proteins of *R. marinus* and *R. limneticus* identified as being putatively plastid-targeted were either added to existing alignments (FeS cluster and haem biosynthesis pathway proteins) or used as queries in a blastp search (E-value threshold of 1×10^{-5}) against a comprehensive custom database, which contained representatives of plastid-bearing and non-plastidial groups (archaeplastids, cryptophytes, haptophytes, stramenopiles, dinoflagellates, chrompodellids, apicomplexans as well as opisthokonts, amoebozoans/apusozoans/ancyromonads and ciliates). For the TIC22 and TOC75 alignments, additional known homologues were used as queries, to extend the number of taxa obtained with *R. marinus* and *R. limneticus*. Results from blast searches were parsed for hits with a minimum query coverage of 50% and $E < 1 \times 10^{-25}$ or $E < 1 \times 10^{-5}$ (TIC/TOC and YCF proteins, respectively). The number of bacterial hits was constrained to 20 hits per phylum (for the Fibrobacteres–Chlorobi–Bacteroidetes group, most classes of Proteobacteria, the Planctomycetes–Verrucomicrobia–Chlamydiae group, Spirochaetes, Actinobacteria, Cyanobacteria (unranked) and Firmicutes) or 10 per phylum (remaining bacterial phyla) as defined by NCBI taxonomy. For initial tree reconstruction, corresponding sequences were aligned with MAFFT using the '--auto' option, ambiguously aligned positions were trimmed off with trimAL v.1.2[47] using a gap threshold of 20% and trees were calculated using FastTree[48] with default options. Resulting phylogenies and underlying alignments were manually inspected, and obvious contaminations and paralogues were removed. Cleaned sequence files as well as the pre-existing alignments with added *R. marinus*/*R. limneticus* sequences were filtered using prequel v.1.0.2[49] to remove stretches of non-homologous characters and then aligned using the G-INS-I algorithm of MAFFT in combination with VSM[50], using an α_{max} of 0.6, to reduce over-alignment. Alignments were trimmed as above and final trees were reconstructed using IQTREE[21], using ModelFinder[51] to identify the best model for each alignment based on the Bayesian information criterion. Node supports were calculated with 1,000 ultrafast bootstrap (UFBoot) replicates[52].

Comparison of *Rhodelphis* gene repertoire to red algae. We used the KEGG Automatic Annotation Server[53], using the bi-directional best hit, to functionally annotate *Rhodelphis* genes, along with genes encoded by the red algae *Cyanidioschyzon merolae*, *Galdieria sulphuraria*, *Chondrus crispus* and *Porphyra purpureum*. The resulting KEGG Orthology assignments were used to infer the overlap or difference in gene repertoire between *Rhodelphis* and red algae. Similarly, we used OrthoFinder v.2.0.0[54] to assess the overlap in orthologous gene sets between the same species (Extended Data Table 1c). Genome-based assessments of *Rhodelphis* trophic mode were done with *R. marinus* and *R. limneticus* Transdecoder proteins using PredictTrophicMode_Tool.R[12].

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw transcriptome and genome reads from *R. limneticus* and *R. marinus* are deposited in GenBank (PRJNA544719), along with full SSU rRNA gene sequences for *R. marinus* (MK966712) and *R. limneticus* (MK966713). Assembled transcriptomes and genomes, along with raw light and electron-microscopy images, individual gene alignments, concatenated and trimmed alignments, single-gene trees, and maximum-likelihood and Bayesian tree files for the 151-taxon and 153-taxon datasets have been deposited in Dryad (https://doi.org/10.5061/dryad.tr6d8q2). The family Rhodelphidae (urn:lsid:zoobank.org:act:80B5C004-2954-4A57-A411-482BCD29E85D), genus *Rhodelphis* (urn:lsid:zoobank.org:act:6D09D4D9-D9FC-4D0C-8FB2-55FD9DDEAD53) and species *Rhodelphis limneticus* (urn:lsid:zoobank.org:act:695ACD0B-8151-4609-97FC-A044A312BE22) and *Rhodelphis marinus* (urn:lsid:zoobank.org:act:84233191-4710-43D1-A2DA-914B8E7B7E01) have been registered with the Zoobank database (http://zoobank.org/).

Code availability

All unpublished code is available upon reasonable request from the corresponding authors.

22. Tikhonenkov, D. V., Mazeĭ, IuA. & Embulaeva, E. A. [Degradation succession of heterotrophic flagellate communities in microcosms]. *Zh. Obshch. Biol.* **69**, 57–64 (2008).
23. Tikhonenkov, D. V. et al. Description of *Colponema vietnamica* sp.n. and *Acavomonas peruviana* n. gen. n. sp., two new alveolate phyla (Colponemidia nom. nov. and Acavomonidia nom. nov.) and their contributions to reconstructing the ancestral state of alveolates and eukaryotes. *PLoS ONE* **9**, e95467 (2014).
24. Luft, J. H. Improvements in epoxy resin embedding methods. *J. Biophys. Biochem. Cytol.* **9**, 409–414 (1961).
25. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
26. Tikhonenkov, D. V., Janouškovec, J., Keeling, P. J. & Mylnikov, A. P. The morphology, ultrastructure and SSU rRNA gene sequence of a new freshwater flagellate, *Neobodo borokensis* n. sp. (Kinetoplastea, Excavata). *J. Eukaryot. Microbiol.* **63**, 220–232 (2016).
27. Andrews, S. FastQC: a quality control tool for high throughput sequence data. version 0.10.1 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2010).
28. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
29. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
30. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
31. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
32. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
33. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
34. Miller, I. J. et al. Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* **47**, e57 (2019).
35. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).

36. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
37. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
38. Ryan, J. F. Baa.pl: a tool to evaluate de novo genome assemblies with RNA transcripts. Preprint at https://arxiv.org/abs/1309.2087 (2013).
39. Gruber-Vodicka, H. R., Seah, B. K. B. & Pruesse, E. phyloFlash – rapid SSU rRNA profiling and targeted assembly from metagenomes. Preprint at https://www.biorxiv.org/content/10.1101/521922v1 (2019).
40. Burki, F. et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc. R. Soc. B* **283**, 20152802 (2016).
41. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
42. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
43. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
44. Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol. Biol.* **7** (Suppl. 1), S2 (2007).
45. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis. Version 3.5 https://www.mesquiteproject.org/ (2018).
46. Emanuelsson, O., Nielsen, H., Brunak, S. & von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **300**, 1005–1016 (2000).
47. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
48. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
49. Whelan, S., Irisarri, I. & Burki, F. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).
50. Katoh, K. & Standley, D. M. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* **32**, 1933–1942 (2016).
51. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
52. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
53. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
54. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).

Acknowledgements We thank L. Nguyen-Ngoc, H. Doan-Nhu, E. S. Gusev, Y. Dubrovsky and the staff of the Russian-Vietnam Tropical Centre, Coastal Branch for assistance with sample collection and trip management; S. A. Karpov for assistance with interpretation of transmission electron microscopy images; Compute/Calcul Canada for computational resources, especially the Orcinus (Westgrid) and Mammouth Parallèle II (Calcul Québec) clusters. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to P.J.K. (grant 227301). Field work in Vietnam is part of the project 'Ecolan 3.2' of the Russian-Vietnam Tropical Centre. R.M.R.G. was supported by a fellowship from the Canadian Institutes of Health Research and a grant from the Tula Foundation to the Centre for Microbial Diversity and Evolution. Cell isolation and culturing, generation of material for sequencing, light and electron microscopy and analysis were supported by the Russian Science Foundation to D.V.T. (grant 18-14-00239). F.H. is supported by an EMBO fellowship (ALTF 1260–2016).

Author contributions R.M.R.G., D.V.T. and P.J.K. designed the study. D.V.T. isolated and cultured cells. D.V.T. and F.H. generated material for sequencing and F.H. removed contaminant sequences from the genome assembly. D.V.T. and A.P.M. performed microscopy experiments. R.M.R.G. performed phylogenomic, genomic and transcriptomic analyses. E.H. performed phylogenetic analysis of plastid proteins. R.M.R.G., D.V.T. and P.J.K. wrote the manuscript with input from all authors.

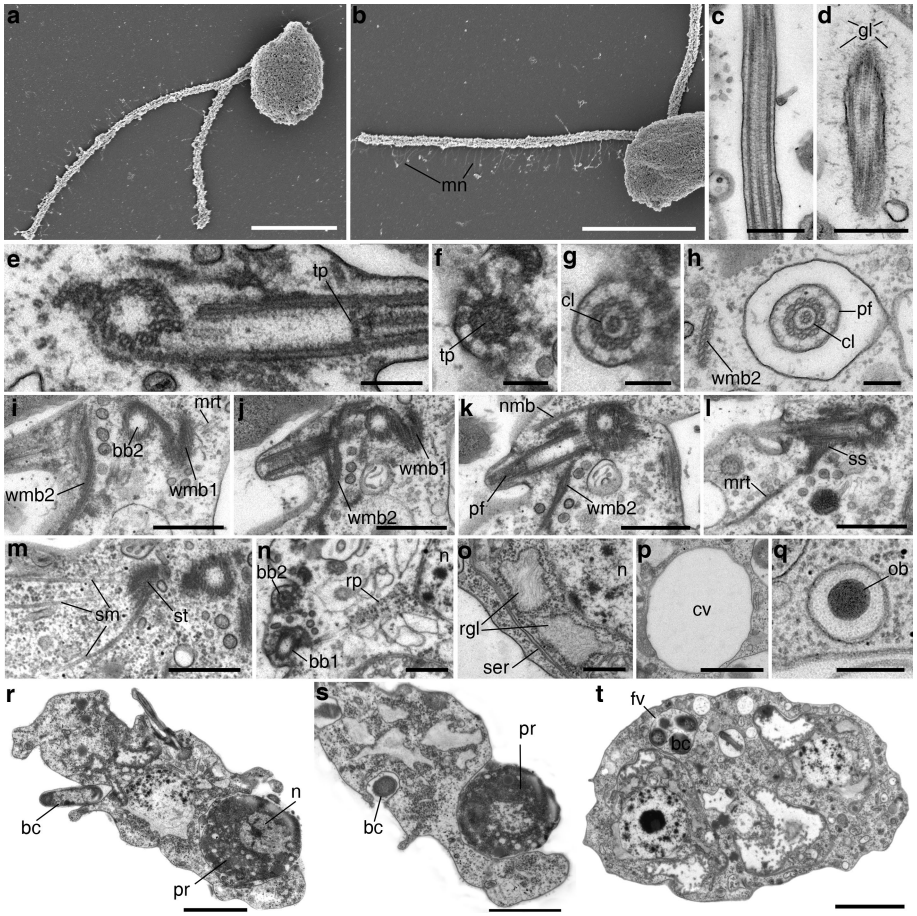
Competing interests The authors declare no competing interests.

Additional information

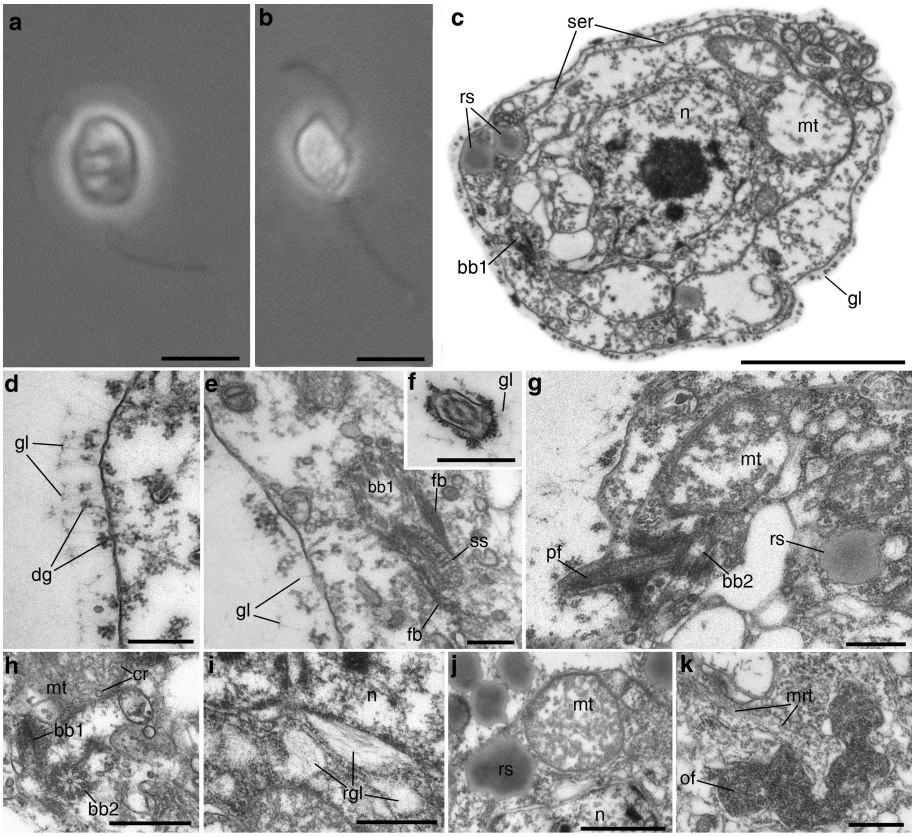
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1398-6.

Correspondence and requests for materials should be addressed to R.M.R.G., D.V.T. or P.J.K.

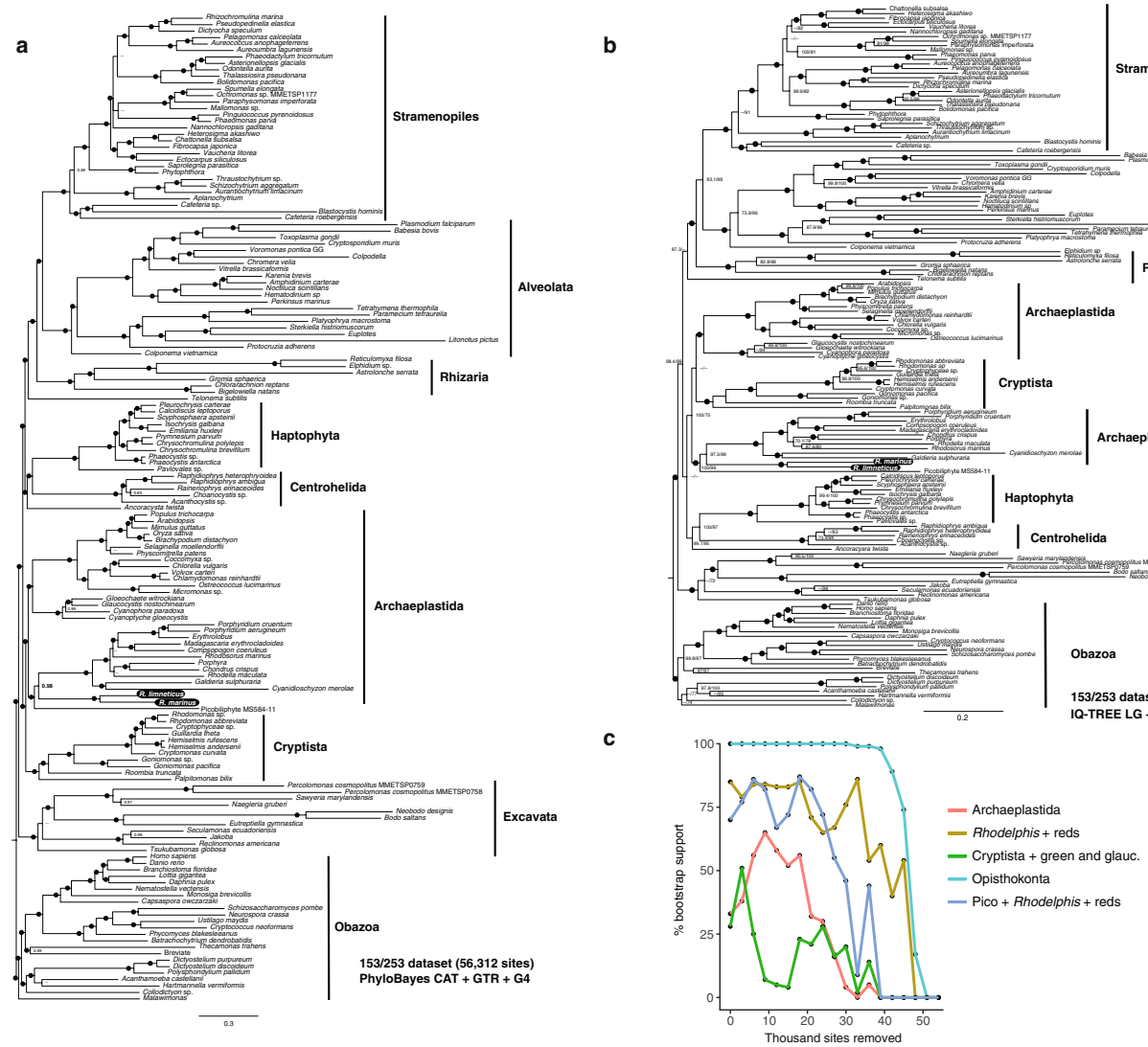
Peer review information *Nature* thanks Geoffrey McFadden and the other anonymous reviewer(s) for their contribution to the peer review of this work.



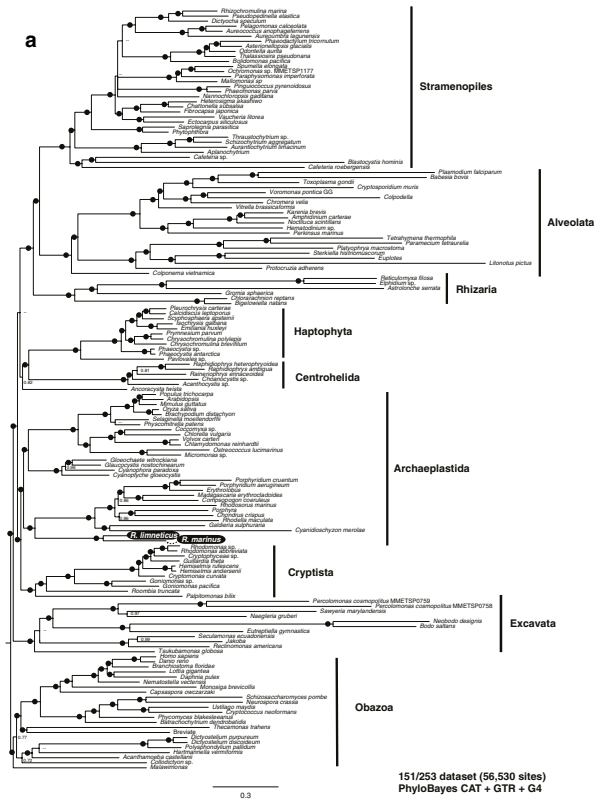
Extended Data Fig. 1 | Cell structure of *R. limneticus*. Related to Fig. 1. **a**, **b**, Scanning electron microscopy images showing the flagella and mastigonemes on the posterior flagellum. **c**, Section of an anterior flagellum. **d**, Section of a posterior flagellum. **e**–**g**, Arrangement of the transitional zone of a flagellum with transverse plate and cylinder. **h**, Wide microtubular band 2 accompanies the posterior flagellum. **i**–**l**, Cell sections from anterior to posterior. **m**, Single microtubules inside the cytoplasm. **n**, A rhizoplast connects the basal body of the posterior flagellum to the nucleus. **o**, Area of cell with nucleus, rudiments of glycostyles inside the vesicles and smooth endoplasmic reticulum. **p**, Contractile vacuole. **q**, Osmiophilic body. **r**, **s**, Phagocytosis of eukaryotic prey and bacteria. **t**, Cell section showing food vacuole with several engulfed bacterial cells. bc, bacterium; cl, cylinder; cv, contractile vacuole; fv, food vacuole; gl, glycostyles; bb1, basal body of posterior flagellum; bb2, basal body of anterior flagellum; mn, mastigonemes; mrt, microtubule; n, nucleus; nmb, narrow microtubular band; ob, osmiophilic body; pf, posterior flagellum; pr, eukaryotic prey; rgl, rudiments of glycostyles; rp, rhizoplast; ser, sac of smooth endoplasmic reticulum; sm, single microtubule; ss, striated structure; st, satellite of basal body; tp, transverse plate; wmb1, wide microtubular band 1; wmb2, wide microtubular band 2. Scale bars, 5 μm (**a**, **b**), 0.5 μm (**c**, **d**, **i**–**o**), 0.2 μm (**e**–**h**, **q**), 1 μm (**p**) and 2 μm (**r**–**t**). These experiments were repeated three (**a**, **b**) and seven (**c**–**t**) times, with similar results.

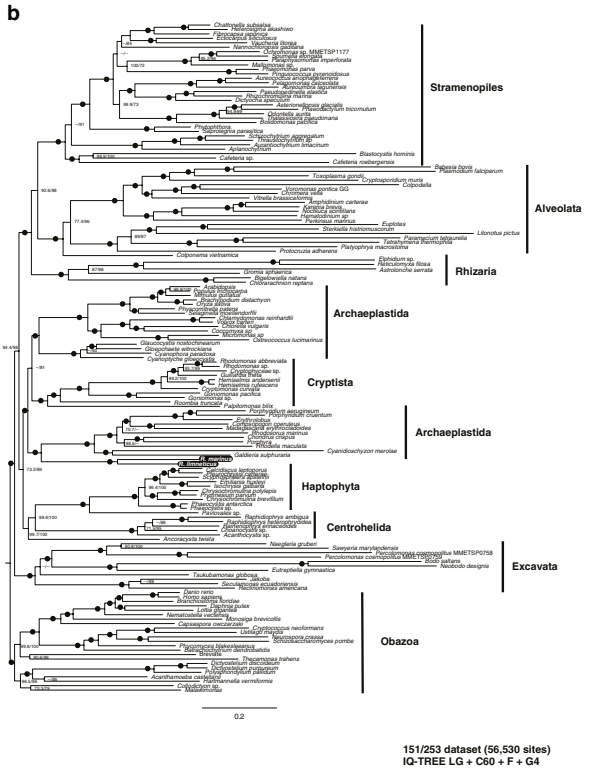


Extended Data Fig. 2 | Cell structure of *R. marinus*. **a**, **b**, Living cells, obtained by light microscopy. **c**, Longitudinal section of the cell, **d**, Region of the cell surface with glycostyles. **e**, Basal body of posterior flagellum with outgoing fibrils and striated structure. **f**, Section of the flagellum covered with glycostyles and dark granules. **g**, Emergence of a posterior flagellum. **h**, Basal body of the posterior flagellum and mitochondrion with tubular cristae. **i**, Formation of rudiments of glycostyles in perinuclear space. **j**, Nucleus, mitochondrion and reserve substance. **k**, Osmiophilic formation and microtubules. cr, cristae; dg, dark granules; fb, fibril; gl, glycostyles; bb1, basal body of posterior flagellum; bb2, basal body of anterior flagellum; mt, mitochondrion; mrt, microtubules; n, nucleus, of, osmiophilic (dark) formation; pf, posterior flagellum; rgl, rudiments of glycostyles; rs, reserve substance; ser, sac of smooth endoplasmic reticulum; ss, striated structure. Scale bars, 10 μm (**a**, **b**), 2 μm (**c**), 0.2 μm (**d**, **e**), 0.5 μm (**f**–**i**, **k**) and 1 μm (**j**). These experiments were repeated ten (**a**, **b**) and three (**c**–**k**) times, with similar results.

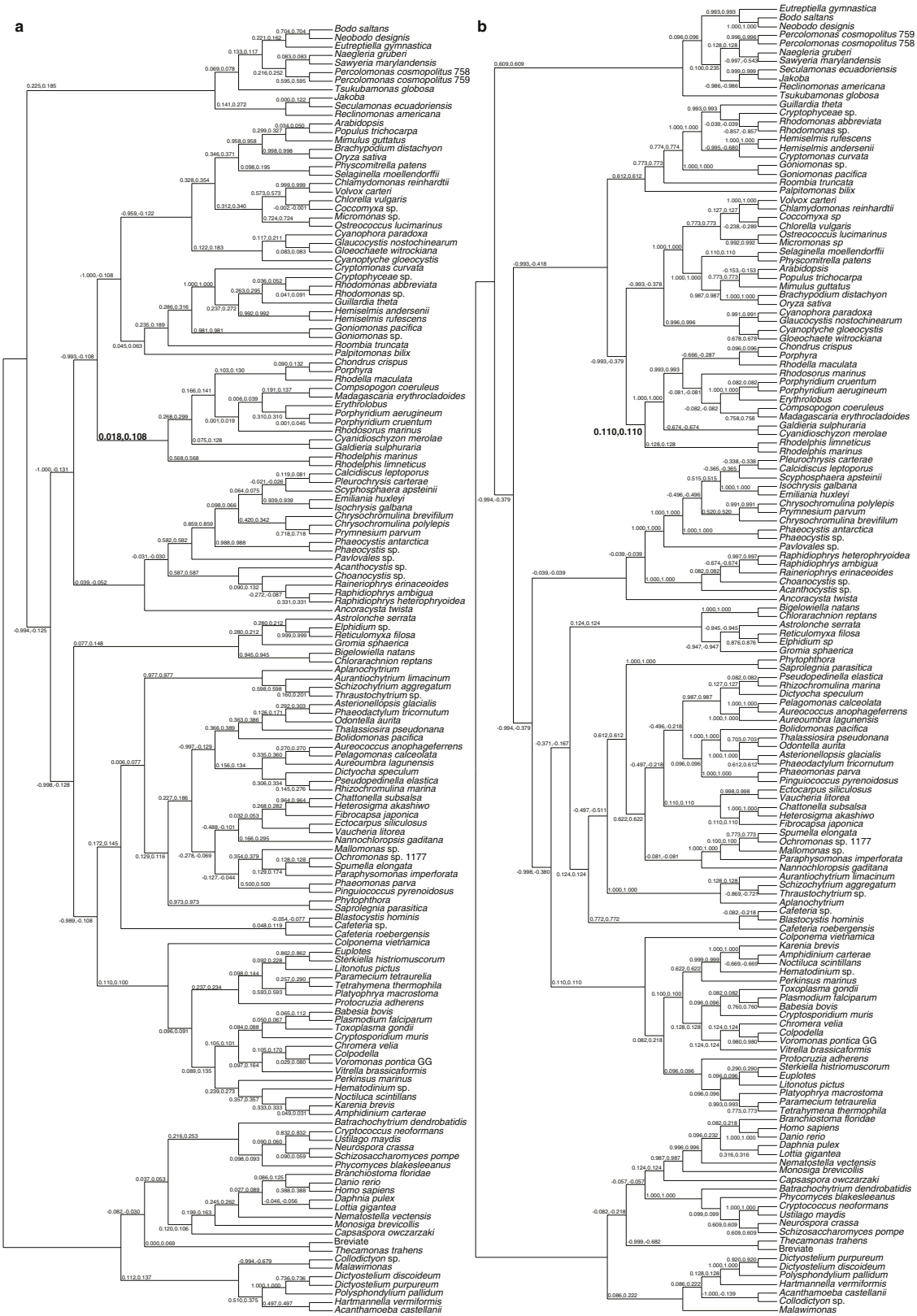


Extended Data Fig. 3 | Phylogenomic analysis of the concatenated 153/253 dataset. a, Bayesian tree using the CAT + GTR evolutionary model as implemented in PhyloBayes. **b**, Maximum-likelihood tree using the LG + C60 + F + G4 model as implemented in IQ-TREE. Black dots denote full statistical support (Bayesian posterior probability = 1.0, maximum-likelihood ultrafast bootstrap and SH-aLRT = 100%); support values <0.7/70% are not shown (indicated by '–'). **c**, Bootstrap support for maximum-likelihood trees (PROT + CAT + LG + F) after progressive removal of the fastest evolving amino acid sites shows that both the *Rhodelphis* and red algae and the picozoa, *Rhodelphis* and red algae relationships are relatively robust to data removal. Similar to the 151/253 dataset, support for Archaeplastida paraphyly (Cryptista, green plants and glaucophytes (green and glauc.)) decreases with data removal, whereas Archaeplastida monophyly support increases. Support for Opisthokonta monophyly serves as a control for the presence of sufficient information for phylogenomic inference.

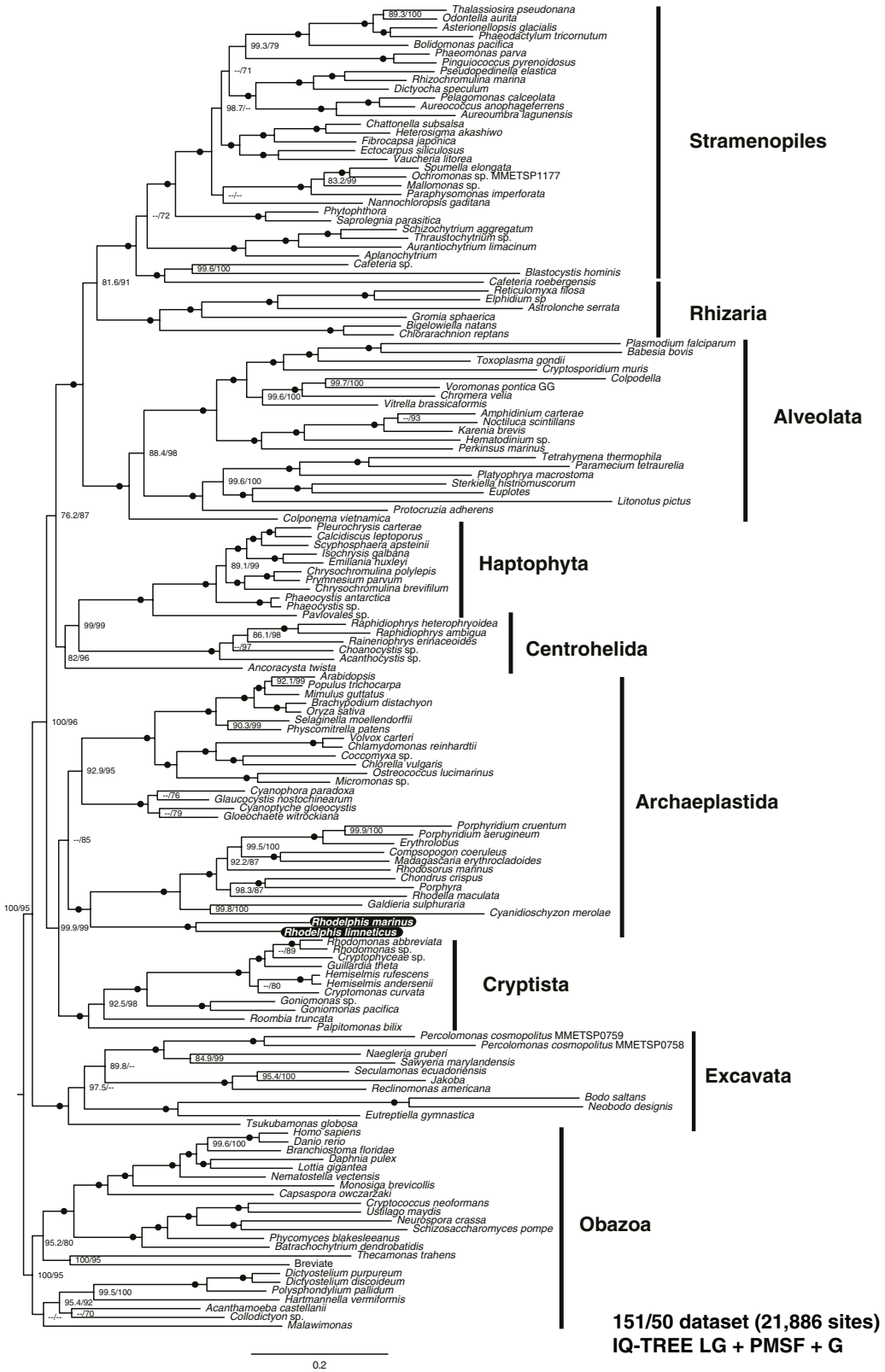




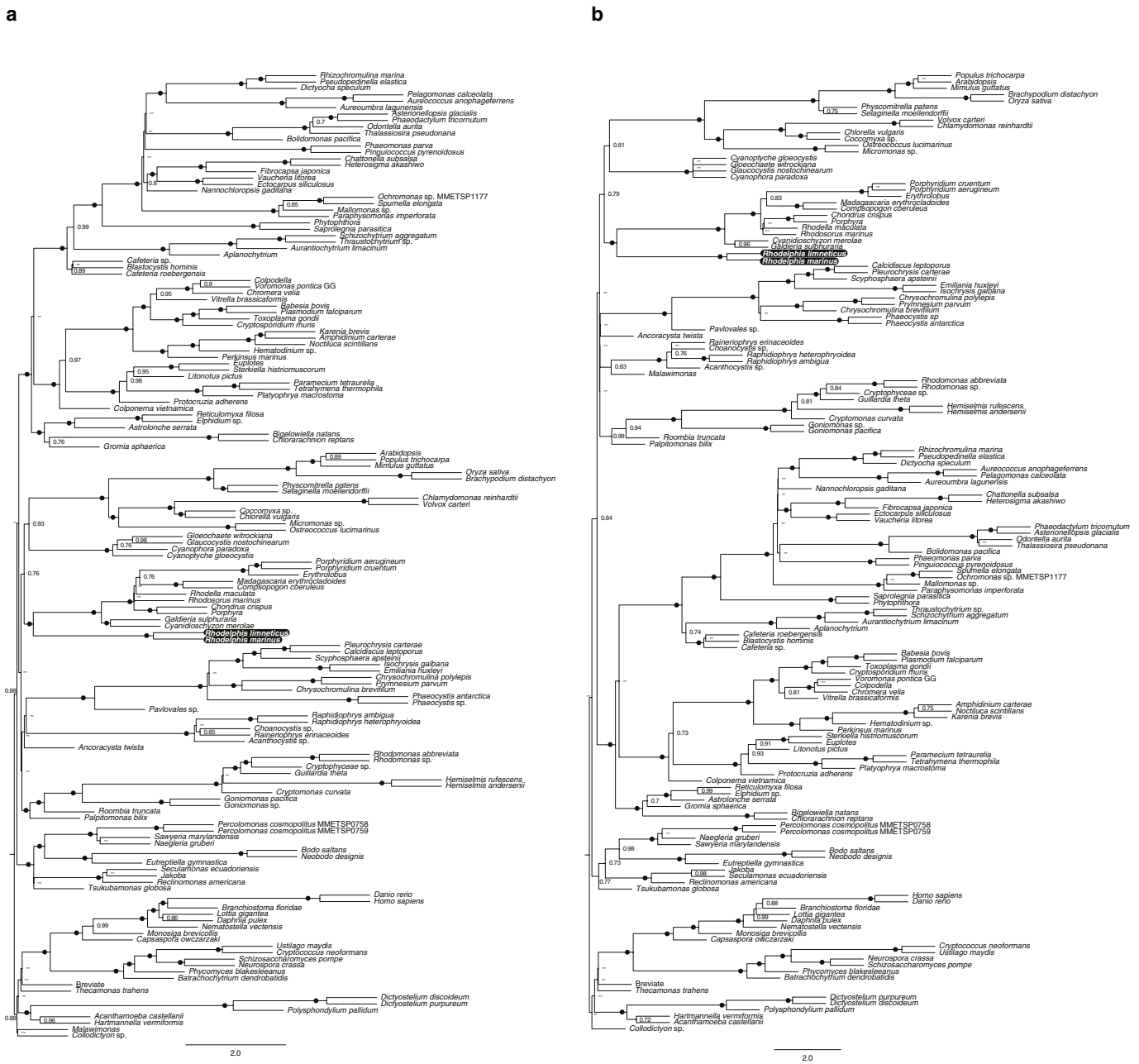
Extended Data Fig. 4 | Phylogenomic analysis of the concatenated 151/253 dataset. a, Bayesian tree using the CAT + GTR evolutionary model as implemented in PhyloBayes. **b**, Maximum-likelihood tree using the LG + C60 + F + G4 model as implemented in IQ-TREE. Black dots denote full statistical support (Bayesian posterior probability = 1.0, maximum-likelihood ultrafast bootstrap and SH-aLRT = 100%); support values <0.7/70% are not shown.



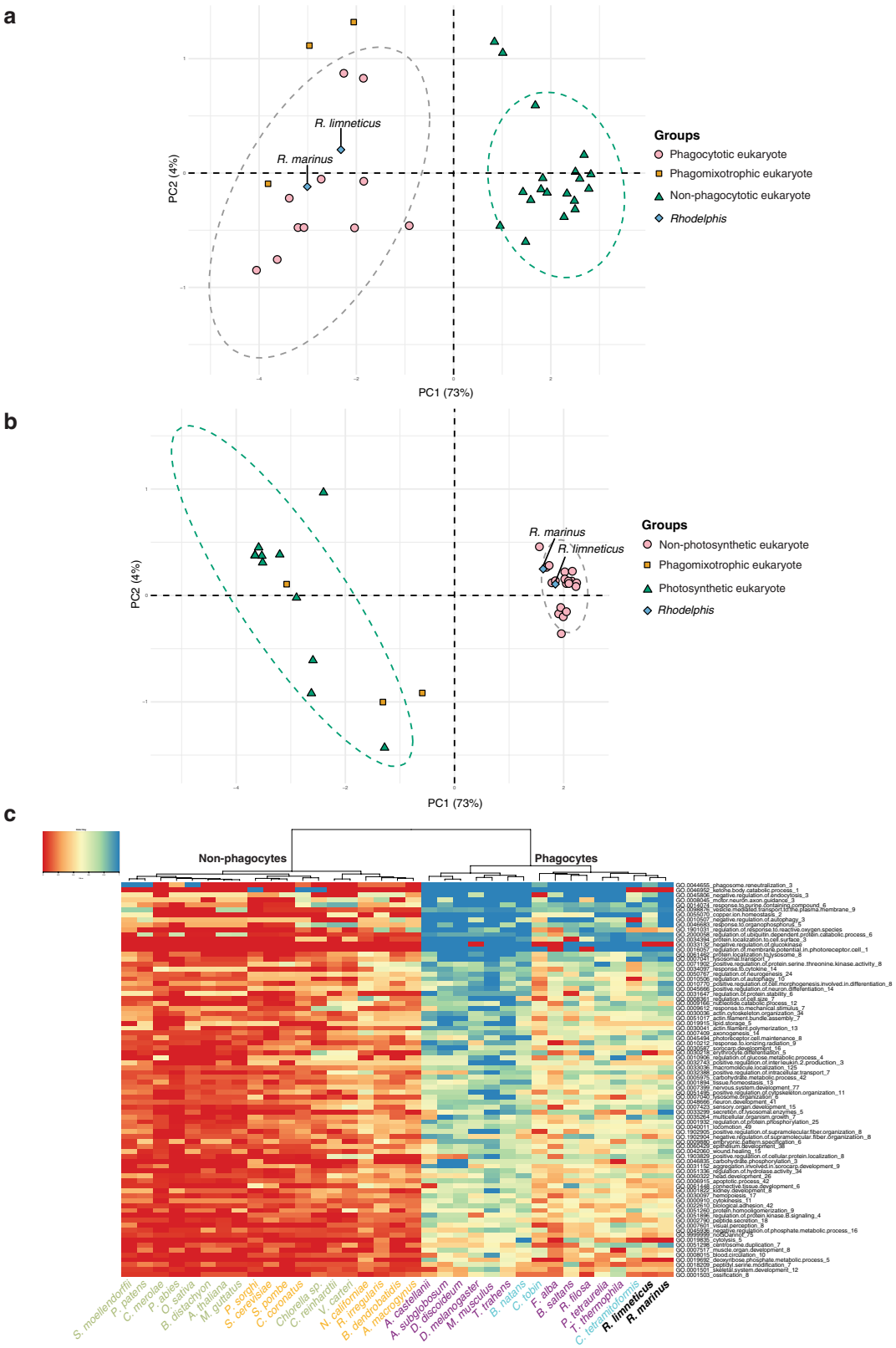
Extended Data Fig. 5 | Internode certainty analyses for the 151/253 and 151/50 datasets. Internode certainty and internode certainty-all scores were calculated with RAxML v.8.1.6 and mapped onto the 151/253 maximum-likelihood tree topology presented in Extended Data Fig. 4b. **a**, Internode certainty and internode certainty-all scores for 253 individual bootstrapped maximum-likelihood trees used to generate the concatenated alignment for Extended Data Fig. 4b. **b**, Internode certainty and internode certainty-all scores for the 50 trees among the 253-tree dataset that have the highest RTC scores, which are expected to improve the robustness of phylogenomic analyses. Internode certainty scores for the sister relationship of *Rhodelphis* and red algae are higher for the 50 best-supported trees, indicating that they also support this relationship, and with fewer conflicting signals.



Extended Data Fig. 6 | Phylogenomic analysis based on concatenation of 50 single-gene datasets with highest RTC scores. Maximum-likelihood tree using the LG + PMSF + G model as implemented in IQ-TREE (151 taxa, 50 proteins, 21,886 sites). Black dots denote full statistical support (maximum-likelihood ultrafast bootstrap and SH-aLRT = 100%); support values <0.7/70% are not shown. The sister relationship of *Rhodelphis* and red algae still receives full statistical support with a highly reduced, phylogenetically well-supported dataset.



Extended Data Fig. 7 | A coalescence phylogenomic framework recovers *Rhodelphis* as sister to red algae based on individual gene trees. Individual bootstrapped gene trees were generated with RAxML v.8.1.6 and used to generate a species tree with ASTRAL-III under default parameters and 100 bootstrap replicates. Support values <0.7/70% are not shown. **a**, Species trees were made from all 253 single-gene trees from the 151/253 dataset. **b**, Species trees were made from the 50 trees with the highest relative tree certainty. The sister relationship of *Rhodelphis* and red algae is recovered with both datasets, and is consistent with concatenated phylogenomic analyses.



Extended Data Fig. 8 | Genomic support for *Rhodelphis* as non-photosynthetic phagotrophs. a, **b**, Principal component (PC) plots of gene ontology (GO) category score from free-living phagocytes (**a**) and photosynthetic organisms (**b**). **a**, *Rhodelphis* associate with phagocytes, but not with photosynthetic eukaryotes. Dashed ellipses represent 95% confidence intervals based on training datasets using a model defined by free-living phagocytes (**a**; $n = 86$ GO categories, 474 proteins) and photosynthesis model (**b**; $n = 37$ GO categories, 243 proteins). **c**, Heat map of phagocyte GO terms showing (as in **a**) that *Rhodelphis* gene repertoires are similar to phagocytes. Analyses were performed using PredictTrophicMode_Tool.R.

Extended Data Fig. 9 | *Rhodelphis* encode plastid-targeted proteins with N-terminal targeting sequences and homologues of the TIC/TOC plastid import system. a, An alignment of plastid-type chaperonin 60 (related to Fig. 3b) shows that *Rhodelphis* nuclear genomes encode plastid-targeted proteins that have clear N-terminal extensions (cTP) relative to plastid-encoded red algal homologues (names in red) and cyanobacterial homologues (names in cyan), but lack a signal peptide (SP) characteristic of proteins targeted to complex secondary or tertiary plastids, as found in *Plasmodium* (orange). **b**–**e**, *Rhodelphis* nuclear genomes encode bona fide homologues of plastid protein import subunits TIC20, TIC32, TIC22 and TOC75, which are specific genetic markers for plastid presence.

Extended Data Table 1 | Summary of *Rhodelphis* genome and transcriptome data

a

R. limneticus WGA assembly statistics						
	>0 bp	>1000 bp	>5000 bp	>10000 bp	>25000 bp	>50000 bp
# contigs	54043	20225	7826	3768	655	54
Length	140673149	128280609	98710693	69888389	22743207	3260312
Largest contig (bp)	105305					
Final len gth (>500 bp)	134325183					
GC (%)	44.29					
N50	10548					
N75	4708					
L50	3503					
L75	8246					
N's per 100 kbp	51.04					

b

Species	# proteins	BUSCO complete	BUSCO fragmented	BUSCO missing	Mapped to genome	# introns
R.marinus	14,585	90.7%	4.3%	5.0%	N/A	N/A
R.limneticus	13,346	82.5%	10.9%	6.6%	97%	~39.5k

c

Species	*R. marinus*	*R. limneticus*	*G. sulphuraria*	*C. merolae*	*C. crispus*	*P. purpureum*
# genes	14585	13346	7174	4803	9807	8355
# genes in OG	8800	7534	5329	3878	4699	5784
# unassigned genes	5785	5812	1845	925	5108	2571
% genes in OG	60.3	56.5	74.3	80.7	47.9	69.2
% unassigned genes	39.7	43.5	25.7	19.3	52.1	30.8
# OG with spec.	5487	5006	3717	3280	3367	4028
% OG with spec.	75.7	69	51.3	45.2	46.4	55.5
# spec-specific OG	16	10	7	7	10	6
# genes in spec-specific OG	62	69	83	37	60	30
% genes in spec-specific OG	0.4	0.5	1.2	0.8	0.6	0.4

a, Assembly statistics for the *R. limneticus* WGA dataset. HiSeq X 150-bp paired-end and MinION reads were assembled with SPAdes v.3.11.1, the assembly (scaffolds) was decontaminated using Autometa and assessed with QUAST v.5.0.2. N50 and N75 values refer to the shortest contig lengths that make up 50% and 75% of the genome assembly, respectively. L50 and L75 refer to the smallest number of contigs that account for 50% and 75% of the assembly length, respectively. **b**, Summary of BUSCO reports for both *Rhodelphis* species based on decontaminated transcriptome data. *R. limneticus* results are from a combination of single-cell and culture datasets. Percentage of transcripts mapping to the *R. limneticus* genome was determined with isoblat and introns were inferred from GMAP alignments. High-quality *R. marinus* genome data were not generated before the death of the culture. **c**, Orthologous group (OG) distribution across red algae and *Rhodelphis*. OrthoFinder v.2.0.0 was used to infer orthologous groups between proteins inferred from *R. marinus* and *R. limneticus* transcriptomes and red algal genomes. *Rhodelphis* genomes are gene-rich in comparison to published red algal genomes, and retain components of evolutionarily conserved structures (such as flagella and centrioles) that have been lost in red algae. Species is abbreviated as 'spec'.

nature research

Corresponding author(s): Ryan Gawryluk

Last updated by author(s): May 26, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequence quality: FastQC v0.10.1 Transcriptome assembly: Trimmomatic v0.32;Trinity v2.0.6;Transdecoder v5.0.2;CD-HIT v4.6 Genome assembly: SPAdes v3.11.1
Data analysis	Transcriptome and genome analysis: QUAST v5.0.2; GMAP v2016-08-16; Autometa, cloned Jan 29, 2019 (https://bitbucket.org/jason_c_kwan/autometa); phyloFlash v3.0;R v3.5.0;BUSCO v3.0.1;BLAST+ v2.2.30;WebLogo v3 (http://weblogo.threeplusone.com/create.cgi);isoblat v0.3 Comparative genomics: OrthoFinder v2.0.0;KEGG Automatic Annotation Server (KAAS) (https://www.genome.jp/kegg/kaas/);TargetP v1.1 Phylogenomic and phylogenetic: mesquite v3.5;trimAL v1.2;MAFFT v7.212;SCaFOs v1.2.5;BMGE v1.12;RAxML v8.1.6;FastTree v2.1.7 SSE3;IQ-TREE v1.5.5;PhyloBayes MPI v1.7;AgentSmith;ASTRAL-III v5.6.2;Prequal v1.0.2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw transcriptome and genome reads from R. limneticus and R. marinus have been deposited in the GenBank Sequence Read Archive (SRA). Full SSU rRNA gene sequences have been deposited in GenBank under the accessions MK966712 (R. marinus) and MK966713 (R. limneticus). Assembled transcriptomes and genomes

have been deposited in Dryad (ACCESSION), along with raw light and electron microscopy images, individual gene alignments, concatenated and trimmed alignments, and ML and Bayesian tree-files for the 151-taxon and 153-taxon datasets . Zoobank accessions are also provided for family (urn:lsid:zoobank.org:act:80B5C004-2954-4A57-A411-482BCD29E85D), genus (urn:lsid:zoobank.org:act:6D09D4D9-D9FC-4D0C-8FB2-55FD9DDEAD53), and species (R. limneticus, urn:lsid:zoobank.org:act:695ACD0B-8151-4609-97FC-A044A312BE22; R. marinus, urn:lsid:zoobank.org:act:84233191-4710-43D1-A2DA-914B8E7B7E01).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study, we describe two species from a novel phylum of predatory eukaryotic microbes, Rhodelphidia. We performed detailed ultrastructural, transcriptomic/genomic, and phylogenomic analyses, showing that Rhodelphis are the sister lineage to red algae, and that they likely retain a non-photosynthetic plastid involved in heme biosynthesis.
Research sample	This research describes two new species, Rhodelphis limneticus and R. marinus, from a new phylum of predatory eukaryotic microbes that is the sister lineage to red algae. The organisms were collected from freshwater lake sediments and seawater sediments, respectively.
Sampling strategy	Sample size is not relevant to the present study.
Data collection	Samples were collected from freshwater and marine sediments, and the new organisms were subsequently grown in the laboratory. Microscopic data were recorded by D Tikhonenkov. Sequencing data were generated by the UCLA Clinical Microarray Core (R. marinus), The Centre of Applied Genomics (Toronto, Canada; R. limneticus), and in-house using a minION (F Husnik). Transcriptome/genome data were assembled by R Gawryluk.
Timing and spatial scale	Sampling relevant to the present study was carried out only two times: once from marine coral sand off of a small island, Bay Canh, near Con Dao Island, South Vietnam on May 3, 2015, and once from a near-shore freshwater sample on August 9, 2016. We had no reason to expect to find the organisms that we did, so there is no specific rationale to sampling sites.
Data exclusions	Sequencing data from prey organisms were excluded from the analyses as best as possible. To do this for transcriptomes (R.marinus and R. limneticus), we subtracted transcripts derived from prey (kinetoplastids and co-cultured bacteria) from the total datasets. For genomic datasets, we used automated binning (autometa) along with limited manual curation in order to reduce prey contributions to the genome datasets. The raw data associated with this are still accessible in the raw read files deposited in the NCBI SRA database.
Reproducibility	Microscopic analyses were conducted several times. Phylogenomic analyses were carried out with a number of different approaches (maximum likelihood, Bayesian etc.) and all associated datasets have been made available.
Randomization	Randomization is not relevant to the present study because organisms were not allocated into groups.
Blinding	Blinding was not relevant to the present study.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Climatic conditions in the field were not recorded and are not relevant to the study.
Location	1) Bay Canh island near Con Dao Island, South Vietnam (8.666288 N, 106.680488 E). 2) Lake Trubin (flood-lands of Desna River, 51.397222 N, 32.368889 E), near Yaduty village, Chernigovskaya oblast, Ukraine
Access and import/export	Habitats were accessed via a motor boat (location 1) and a car (location 2). No permissions were required for sampling in the selected sampling sites.
Disturbance	No disturbances to the sites were caused; we sampled a small amount of water and sand/debris from a marine and lake habitat.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Two clonal cultures of protists were isolated from marine coral sand and a freshwater sample containing organic debris.
Authentication	Phase contrast light microscopy and 18S rRNA gene sequencing was used for authentication.
Mycoplasma contamination	This is not relevant to protist cell culture.
Commonly misidentified lines (See ICLAC register)	This is not relevant to protist cell culture.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve laboratory animals.
Wild animals	The study did not involve wild animals (or any animals).
Field-collected samples	Monoeukaryotic cultures of Rhodelphis marinus and R. limneticus were established by isolating cells with a glass micropipette. Cultures were maintained at room temperature. R. marinus and R. limneticus were propagateed using the kinetoplastid protists Procryptobia sorokini B-69 and Parabodo caudatus BAS-1 as prey, respectively. The kinetoplastids were grown in marine Schmalz-Pratt's medium and spring water and preyed upon Pseudomonas fluorescens.
Ethics oversight	No ethical approval was required. The organisms described here are novel eukaryotic microbes (protists) that feed on other protists and pose no risks.

Note that full information on the approval of the study protocol must also be provided in the manuscript.