Received: 12 August 2025 | Accepted: 18 October 2025

DOI: 10.1111/jpy.70109

GENOME RESOURCE

Nuclear genome assembly and annotation for the phagotrophic green alga *Nephroselmis pyriformis*

Sophie Charvet[1,2] | **Eunsoo Kim**[1,3]

[1]Division of Invertebrate Zoology, American Museum of Natural History, New York City, New York, USA

[2]Department of Biology, School of Natural and Social Sciences, Susquehanna University, Selinsgrove, Pennsylvania, USA

[3]Division of EcoScience, Ewha Womans University, Seoul, South Korea

Correspondence
Sophie Charvet (Department of Biology, School of Natural and Social Sciences, Susquehanna University, Selinsgrove, Pennsylvania 17870, USA) and Eunsoo Kim (Division of EcoScience, Ewha Womans University, Seoul 03760, South Korea).
Email: charvet@susqu.edu, eunsookim@ewha.ac.kr or ekim1@amnh.org

Funding information
National Research Foundation of Korea, Grant/Award Number: RS-2025-00562426; Directorate for Biological Sciences, Grant/Award Number: NSF CAREER-1453639; Simons Foundation, Grant/Award Number: SF-876199

Editor: B. Palenik

Abstract

Prasinophytes are a globally distributed, diverse polyphyletic group of photosynthetic green algae, displaying a combination of presumably ancestral traits for the Chloroplastida. The Nephroselmidophyceae is a prasinophyte clade comprising small biflagellate algae that includes the marine mixoplankton *Nephroselmis pyriformis*, which was previously shown to feed on bacteria, particularly under low concentrations of dissolved inorganic nutrients. Given its mixotrophic proclivities and phylogenetic position, *N. pyriformis* has the potential to provide insights into early green algal evolution and ecophysiology. In this study, we have presented a highly contiguous nuclear genome assembly for *N. pyriformis*. We acquired both Illumina and PacBio long-read data to assemble a 70.8-Mbp nuclear genome annotated with a total of 19,330 protein-coding genes. The genome is inferred to be haploid based on the base frequency distribution at variable sites, together with prior cell biological information from a related species. When compared with four other green algal genomes, *N. pyriformis* displayed a relatively large proportion of ortholog genes shared with the Chlorophyta. The nuclear genomic data presented here will be valuable for a range of studies, including green algal phylogenomics, genome evolution, and phago-mixotrophy.

KEYWORDS
bioinformatics, Chlorophyta, microalgae, molecular, phytoplankton

INTRODUCTION

Prasinophytes are a polyphyletic assemblage of earlier-diverging green algae that constitute the Chlorophyta together with the core chlorophytes (Leliaert, 2019; Leliaert et al., 2012). Ubiquitous in the oceans, prasinophytes usually represent the predominant green algae and can quantitatively dominate total phytoplankton communities in certain regions such as the Arctic Ocean (Joli et al., 2017; Lovejoy et al., 2007), contributing to primary production despite their small size. Their numerical contribution is expected to increase as global changes affect the oceans (Li et al., 2009), which is of significance given their well-recognized roles in primary production and ocean carbon sinks (Bachy et al., 2022). In addition, a growing number of studies have revealed the presence of bacterivory among prasinophytes (Anderson et al., 2018; Bell & Laybourn-Parry, 2003; Bock et al., 2021; Maruyama & Kim, 2013; Mckie-Krisberg et al., 2015), indicating a role as consumers, with an influence on bacterial communities (Hartmann et al., 2012; Zubkov & Tarran, 2008) and marine carbon cycling (Ghyoot et al., 2017; Glibert & Mitra, 2022; Ward & Follows, 2016).

Abbreviations: BLAST, basic local alignment search tool; BUSCO, Benchmarking Universal Single-Copy Orthologs; HMW, high molecular weight; KEGG, Kyoto encyclopedia of genes and genomes; ML, maximum likelihood; SNP, single-nucleotide polymorphism.

Bacterivory has, so far, only been experimentally confirmed for members of the prasinophyte classes Mamiellophyceae, Nephroselmidophyceae, and Pyramimonadophyceae (Anderson et al., 2018; Bell & Laybourn-Parry, 2003; Bock et al., 2021; Mckie-Krisberg et al., 2015). This capacity for bacterivory likely originated from colorless ancestors that existed >1 billion years ago, prior to the acquisition of plastids (Archibald, 2009; Maruyama & Kim, 2020). Bacterivorous prasinophytes, therefore, might offer insights into the origin and early evolution of the Chloroplastida (aka Viridiplantae; Maruyama & Kim, 2013, 2020). Furthermore, the phylogenetic relationships of prasinophyte lineages with each other and with the core chlorophyta, which have yet to be fully resolved, would undoubtedly provide insights for better determining the common ancestor of the Chlorophyta (Leliaert et al., 2012; Turmel et al., 2009).

The recent surge in genomic sequencing has greatly improved our understanding of broad- and fine-scale relationships among green algae and within the chlorophytes (Brouard et al., 2010; Fang et al., 2018; Fučíková et al., 2014; Lemieux et al., 2007) and has allowed the preliminary characterization of prasinophyte classes (Crépeault et al., 2024; Leliaert et al., 2012; Lemieux et al., 2014; Li et al., 2020). Plastid and mitochondrion genomes have provided insights into the origin and spread of organelles (Leliaert et al., 2012; Sibbald & Archibald, 2020; Turmel & Lemieux, 2018), sometimes revealing novel relationships that have contradicted morphology-based classifications (Crépeault et al., 2024) or single-gene phylogenies (McManus et al., 2018). Whole nuclear genome sequencing can provide additional information beyond gene sequence data, such as gene inversions and organization, noncoding region lengths and location, and duplications, as well as other structural genomic features. Together, these can strengthen or refine our understanding of evolutionary relationships and processes (Li et al., 2020).

Despite an increase in the number of sequenced nuclear genomes over the last 2 decades, there are still relatively few prasinophyte genomes available in public databases. Currently, the most represented lineages among sequenced genomes are within the Mamiellophyceae (Derelle et al., 2006; Worden et al., 2009; Moreau et al., 2012; Dennu et al., 2023; Rey Redondo et al., 2024) and other picoplanktonic (<2 μm cell diameter) clades, such as the Chloropicophyceae (Lemieux et al., 2019) and Prasinodermophyceae (Li et al., 2020). Rarer are the genome assemblies for nano-sized prasinophytes possessing larger genomes, such as *Cymbomonas tetramitiformis*, the only member of the Pyramimonadophyceae to have been sequenced to date (Burns et al., 2015; Gyaltshen et al., 2023). Genomic resources for the Nephroselmidophyceae have so far been limited to organelle genomes, including chloroplast genomes of *Nephroselmis astigmatica*

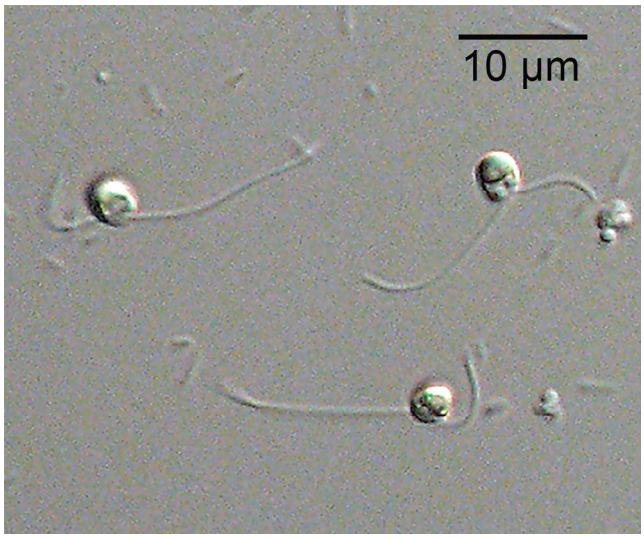

FIGURE 1 Light microscopic image of *Nephroselmis pyriformis* RCC618. Algal cells were observed using a 40× objective lens on an Axiovert 100 M (Carl Zeiss) equipped with a DP73 camera (Olympus). Cells are approximately 3.5 μm in length and bear two unequal flagella.

NIES-252 (Lemieux et al., 2014), *N. pyriformis* MED1 (Gastineau et al., 2021), and *N. olivacea* NIES-484 (Turmel, Otis, et al., 1999b) as well as the mitochondrion genome of *N. olivacea* NIES-484 (Turmel, Lemieux, et al., 1999a). Here, we present the nuclear genome assembly and annotation of *N. pyriformis* RCC618 (Figure 1), which has been experimentally shown to feed on bacteria (Bock et al., 2021).

MATERIALS AND METHODS

Algal culture, nucleic acid extraction, and whole genome sequencing

The prasinophyte *Nephroselmis pyriformis* RCC618 (=CCMP717) was obtained from the Roscoff Culture Collection (France). This strain was isolated from the Galveston Channel in the Gulf of Mexico (United States) on February 1, 1981, by L. Provasoli. The culture was maintained in f/2-Si medium (Guillard, 1975) at 18°C with a light cycle of 12 h at ~80 μmol photons · cm^{-2} · s^{-1}. For DNA extraction, the alga was grown for ~10 days to the highest density in a total of 1 L. High molecular weight (HMW) DNA was extracted using the MagAttract HMW DNA Kit (Qiagen). Briefly, the 1 L of culture was gently filtered onto 0.8-μm pore-sized polycarbonate filters (Millipore). Filters were placed in a mix of ATL Buffer (880 μL) and proteinase K (80 μL) and incubated at 50°C for ~3 h for enzymatic lysis of the prasinophyte cells. The lysate was then transferred in 200-μL aliquots to 2-mL microcentrifuge tubes with MagAttract beads and further processed according to

the manufacturer's instructions. The resulting HMW DNA was used for long-read sequencing at GeneWiz (Azenta Life Sciences). A 20-kb PacBio SMRTbell library was prepared with BluePippin size selection per the manufacturer's protocol and sequenced on the PacBio Sequel platform. An additional DNA extraction was performed using the PureLink Genomic DNA Mini Kit (Invitrogen) according to the manufacturer's instructions for short-read sequencing on the Illumina HiSeq 4000 platform (2 x 150 bp configuration). A total of 934,311 PacBio reads and 351,609,960 Illumina paired-end reads were generated, corresponding to 4.4 Gbp and 105 Gbp, respectively (Table S1).

Total RNA was collected from cultures exposed to two different nutrient conditions (unpublished data by Charvet and Kim). Briefly, a mixture of culture and an equal volume of RNA*later* (Sigma Aldrich) was gently filtered onto 0.8-μm polycarbonate filters (Sterlitech Corporation). The filters were then incubated with 1 mL of TRIzol (Invitrogen) for 5–10 min, after which the volume of TRIzol was used to flush each filter several times and transferred to a 1.7-mL tube. This was followed by the addition of 0.2 mL chloroform and incubation for 2–3 min. After centrifugation at 4°C for 15 min at 12,000× *g*, the clear aqueous phase was carefully transferred to a fresh tube and mixed with an equal volume of 70% ethanol before loading onto a PureLink RNA mini kit (Invitrogen) column for washing and on-column RNA elution per the manufacturer's instructions. Purified RNA samples were sent to GeneWiz (Azenta) for library preparation using NEBNext Ultra RNA Library Prep Kit for Illumina with the poly-A selection and sequencing on Illumina HiSeq (2 × 150 bp configuration). Data from two libraries, each comprising 35 and 37 million paired-end reads (~11 Gbp each), were used for gene prediction.

De novo genome assembly

The initial draft assembly was obtained using the hybrid assembler software MaSuRCA (v.3.2.6.2; Zimin et al., 2013) with the combination of long reads and paired-end short reads. Following the developer's instructions, the input reads were not pre-processed. MaSuRCA was run with the default options except for the following: USE_LINKING_MATES = 1, CA_PARAMETERS = cgwErrorRate = 0.15, KMER_COUNT_THRESHOLD = 2, JF_SIZE = 1,000,000,000. The algal nuclear scaffolds were identified by using a combination of genomic signatures and blastn analyses, a method that has proven effective in other protist genomic studies (e.g., Blaz et al., 2023; Gyaltshen et al., 2023). All the scaffolds were subjected to a blastn search (v2.8.1; Sayers et al., 2025) against the nucleotide (nt) database (v5), with which some clear cases of prokaryotic scaffolds could be identified (e.g., >95% identity over >90% the length of scaffold). Using the binning software MyCC (Lin & Liao, 2016) with the default option of 4-mer frequencies, we identified a cluster comprising 371 scaffolds that were enriched with algal nuclear sequences. Manual inspection of these scaffolds using AliView (v1.28; Larsson, 2014) revealed that 27 were entirely composed of low complexity segments and were therefore discarded. Contaminating scaffolds—including five of human and two of bacterial origin—were also excluded, resulting in a final assembly of 337 scaffolds. Outlier scaffolds with elevated coverage depth, identified through visual inspection of scatter plots (Figures S1, S2), were further examined for phylogenetic classification based on eggNOG-mapper annotations. All surveyed scaffolds contained genes affiliated with Chloroplastida and were retained in the final assembly. The resulting RCC618 nuclear genome assembly was evaluated two ways: (1) by estimating completeness, using BUSCO (v6.0; Tegenfeldt et al., 2025) with the eukaryota_odb12 and the vridiplantae_odb12 datasets and (2) by determining the coverage and sequencing depth. The latter was conducted with Bowtie2 (Langmead & Salzberg, 2013) and minimap2 (Lie et al., 2018) to map the Illumina and PacBio reads to the full assembly, respectively, followed by the output conversion into BAM files and depth calculation at each position or scaffold using SAMtools (v.1.20; Danecek et al., 2021) and an in-house script (kindly provided by Dr. Narechania).

Phylogenetic assessment of the sequenced strain RCC618

To confirm the taxonomic identity of the sequenced strain, an 18S rRNA gene sequence was identified from the genome assembly by blastn searches using the 18S rRNA gene sequence of *Nephroselmis pyriformis* strain MBIC11099 as a query. The RCC618 18S rRNA gene sequence was manually aligned with those sequences from *Nephroselmis* and other green algal species available in GenBank using Mesquite (v3.81; Maddison & Maddison, 2023). The final alignment (Data S1) included 1775 nucleotide sites and 54 sequences. The maximum likelihood (ML) tree was constructed using IQ-TREE (v2.1.2; Minh et al., 2020) with the best-fit model identified by ModelFinder (Kalyaanamoorthy et al., 2017). Ultrafast bootstrap analyses were based on 1000 re-samplings (Hoang et al., 2018).

Genome size and ploidy estimation

To estimate the genome size, all the acquired Illumina reads were subjected to identification and counting of 21-mers using jellyfish (v2.3.1; Marçais & Kingsford, 2011) with the --bf-size option

(to reduce the computational time). The output file was fed into GenomeScope 2.0 (Ranallo-Benavidez et al., 2020) to estimate genome size and repeat content. To estimate genome ploidy, we used nQuire (Weib et al., 2018), a statistical approach that infers ploidy levels from NGS data. For this, base frequency distributions at variable sites were calculated by aligning a subset of the Illumina reads (60,000,000 read pairs) to the final nuclear genome assembly using Bowtie2 (v2.3.5.1; Langmead & Salzberg, 2013). The output alignment in SAM format was converted into the BAM format, followed by sorting, using SAMtools (v1.20; Danecek et al., 2021). Finally, nQuire (Weib et al., 2018), with the denoising option, was used to calculate and plot the biallelic SNP ratios.

Gene structural annotation

The coding regions in the final *Nephroselmis pyriformis* nuclear genome assembly were annotated using the BRAKER3 pipeline (Galaxy v3.0.8+galaxy2; Gabriel et al., 2024) available on the Galaxy platform (The Galaxy Community, 2024). The genome was annotated in several approaches, including with or without soft-masking of repeat regions and with RNA-seq and/or protein evidence. Repeat masking was done using RepeatModeler (v2.0.3; Flynn et al., 2020) and RepeatMasker (v4.1.2; Tarailo-Graovac & Chen, 2009). RNA-seq evidence was prepared by mapping the reads that were sequentially treated for adapter trimming using Cutadapt (Martin, 2011), error correction using Rcorrector (Song & Florea, 2015), and removal of prokaryotic reads with Kraken (Wood & Salzberg, 2014) using the MiniKraken 2020 database. RNA-seq read mapping was done using STAR (v2.5.3a; Dobin et al., 2013). Protein evidence was generated by selecting the Viridiplantae subset from OrthoDB v11 proteins (Kuznetsov et al., 2023). BRAKER3 utilizes GeneMark-ETP (Brůna et al., 2024) and AUGUSTUS (v3.2.3; Stanke et al., 2008) prediction tools. Protein sequences were extracted from the resulting annotations using GffRead (Pertea & Pertea, 2020) and were subjected to Benchmarking Universal Single-Copy Orthologs (BUSCO) analyses (Tegenfeldt et al., 2025). To prevent inflation of duplicated gene counts in the BUSCO analysis, multiple isoforms were counted as a single gene. Functional annotations were performed using eggNOG-mapper (v2; Cantalapiedra et al., 2021) with eggNOG v5.0 database (Huerta-Cepas et al., 2019) as well as GhostKOALA (Kanehisa et al., 2016) plus the KEGGREST package in R (Tenenbaum & Volkening, 2021). The tRNA and rRNA genes were predicted with tRNAscan-SE (Chan & Lowe, 2019) and barrnap (v1.2; Seemann, 2013), respectively.

TABLE 1 *Nephroselmis pyriformis* RCC618 nuclear genome assembly statistics. BUSCO results were based on the use of the eukaryota_odb12 (129 genes) and viridiplantae_odb12 (822 genes) datasets.

# scaffolds	337
Total length	70.8 Mbp
N50	633 Kbp
GC content	64.2%
# protein-coding genes	19,330
# tRNA genes	80
# KEGG-annotated genes	8458 (43.8%) – eggNOG 8029 (41.5%) – GhostKOALA
BUSCO eukaryota: genome	C:90.7% [S:88.4%; D:2.3%]; F:3.9%; M:5.4%
BUSCO eukaryota: proteins	C:95.3% [S:93.0%; D:2.3%]; F:3.9%; M:0.8%
BUSCO viridiplantae: genome	C:87.3% [S:85.8%; D:1.5%]; F:5.8%; M:6.8%
BUSCO viridipantae: proteins	C:93.2% [S:91.2%; D:1.9%]; F:4.4%; M:2.4%

Abbreviations: C, complete BUSCOs; S, complete and single-copy BUSCOs; D, complete and duplicated BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs.

Orthologs and genome comparisons

OrthoFinder (v2.5.4; Emms & Kelly, 2019) was used to identify and investigate the orthologs present in the RCC618 genome in comparison with the core chlorophyte *Chlamydomonas reinhardtii* (Merchant et al., 2007) and prasinophytes *Cymbomonas tetramitiformis* (Gyaltshen et al., 2023), *Chloropicon primus* (Lemieux et al., 2019), and *Micromonas pusilla* (Worden et al., 2009). Venn diagrams to visualize the overlap in orthogroups shared by the various genomes were obtained from https://bioinformatics.psb.ugent.be/webtools/Venn/.

RESULTS AND DISCUSSION

Phylogenetic analyses of 18S rRNA gene from RCC618 (=CCMP717) confirmed its taxonomic placement as *Nephroselmis pyriformis* (Figure S3). The obtained nuclear genome assembly of *N. pyriformis* RCC618 had a total length of 70.8 Mbp comprising 337 scaffolds (derived from 495 contigs) with an average scaffold length of 0.2 Mbp and an N50 of 633 kbp (Table 1). Our assembly is close to the genome size estimate of ~60 Mbp based on k-mer analyses (Figure S4). The Illumina read alignment rate onto the *N. pyriformis* assembly was 78.29%, suggesting that the majority of raw reads came from the alga. Further, read alignment results showed that the nuclear genome was sequenced to 1115x depth with the Illumina reads, with 99.74% of all nucleotide positions mapped (Figure S1; Data S2). With the

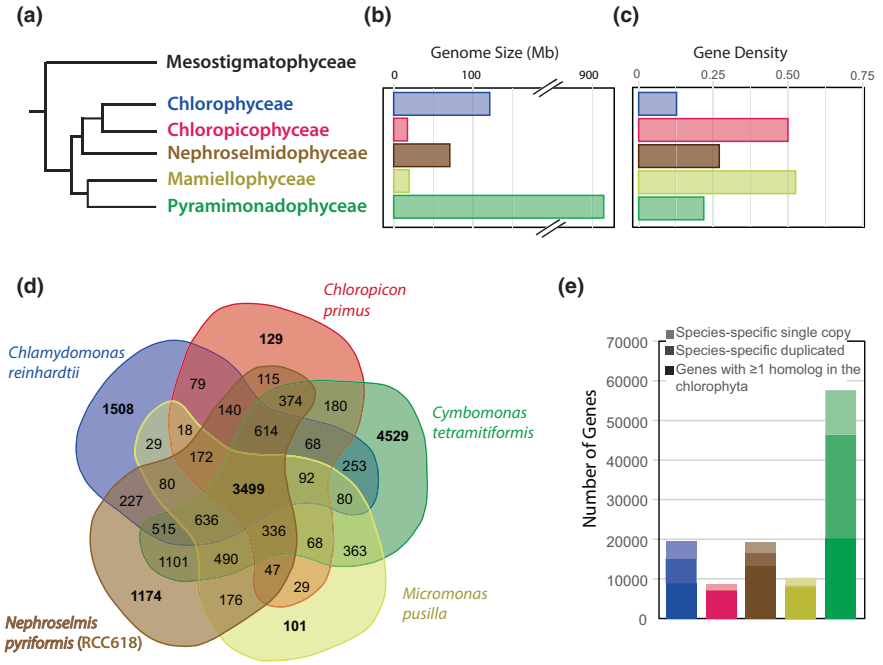

FIGURE 2 Genome characteristics of *Nephroselmis pyriformis* RCC618. The nuclear assembly was compared to assemblies from a core chlorophyte *Chlamydomonas reinhardtii*, and three other prasinophytes, *Cymbomonas tetramitiformis, Chloropicon primus, Micromonas pusilla* (a) to observe the differences in genome size (b) and gene density (c). Shared vs. unique genes are shown in the Venn diagram (d) and ortholog groups are compared (e).

PacBio long reads, the mapped coverage was 99.83%, and the sequencing depth was 39× (Figure S2; Data S3). Our newly generated *N. pyriformis* data thus surpassed the recommended sequencing depth for de novo assembly of a eukaryote genome (i.e., ≥50× Illumina data and ≥30× long-read data for a hybrid assembly; Chakraborty et al., 2016).

The RCC618 assembly showed high completeness, with BUSCO completeness scores of 90.7% and 87.3% when analyzed using the eukaryota_odb12 and viridiplantae_odb12 datasets, respectively (Table 1). The BUSCO completeness values attained for the RCC618 genome assembly were higher than those for four other algal genomes, except for *Chlamydomonas reinhardtii*, which exhibited a higher score in the viridiplantae_odb12 analysis (Table S2). Based on the eukaryota_odb12 database, the RCC618 assembly was the most complete and had the fewest missing BUSCOs (5.4%) among the five green algal genomes examined. These metrics suggest that the assembly reported here likely represents a near-complete reconstruction of the genome.

In annotating the *Nephroselmis pyriformis* genome with BRAKER3, higher-quality gene predictions (as evaluated by BUSCO scores) were obtained when only RNA-seq hints were used, rather than both RNA-seq and protein hints (Table S3). A comparable result was also achieved when the assembly was left unmasked for repeat regions. Therefore, we have presented the annotations generated from the soft-masked assembly using RNA-seq evidence only, which comprise 19,330 predicted protein-coding genes, of which 43.6% and 41.5% were functionally annotated with KEGG identifiers via eggNOG-mapper and GhostKOALA, respectively (Table 1). Compared to the genome assembly, the predicted proteins showed a higher proportion of complete BUSCOs and fewer missing BUSCOs (Table 1), indicating that the gene annotation was of high quality. The genome was additionally annotated with a total of 80 tRNA genes and seven rRNA gene clusters.

Suda et al. (1989, 2004) presented evidence for a zygotic sexual reproductive cycle in *Nephroselmis olivacea*. The non-motile zygote represents the only diploid stage of *N. olivacea* and undergoes meiosis to produce a total of four haploid vegetative biflagellate cells. Assuming *N. pyriformis* follows a similar life cycle, the genome of the vegetative cells from which our assembly was derived is haploid. This haploid state of the genome was supported by our SNP analyses: The bi-allelic SNP frequency histogram (Figure S5) does not display one or more peaks between 0.25 and 0.75 ratios, which would be expected for diploid, triploid, or tetraploid genomes (Weib et al., 2018).

In comparison with nuclear genomes from the chlorophyta, the *Nephroselmis pyriformis* genome is larger than those of the picoplanktonic prasinophytes *Chloropicon primus* and *Micromonas pusilla* but smaller than those of *Cymbomonas tetramitiformis* and *Chlamydomonas reinhardtii* (Figure 2a,b). With a GC content of 64.2%, the RCC618 nuclear genome is comparable to that of *M. pusilla* and *Chlamydomonas reinhardtii* (Table S2). The RCC618 nuclear genome was characterized by 0.27 genes per kb, revealing a gene density lower than that of two smaller genomes (*Chloropicon primus* and *M. pusilla*), but higher than that of larger genomes (*Chlamydomonas reinhardtii* and *Cymbomonas tetramitiformis*; Figure 2c). The RCC618 nuclear genome contained 1174 unique orthogroups and 3449 orthogroups shared with the genomes of *Chlamydomonas reinhardtii*, *Cymbomonas*

tetramitiformis, *Chloropicon primus*, and *M. pusilla* (Figure 2d). Compared to *Cymbomonas tetramitiformis*, whose species-specific duplicated and single genes represented two thirds of the total protein-coding genes, the RCC618 genome was dominated by protein-coding genes that had at least one homolog in other chlorophytes, representing more than half of its gene repertoire (Figure 2e).

Chloroplast genomes are available for *Nephroselmis olivacea* NIES-484 (Turmel, Lemieux, et al., 1999a; Turmel, Otis, et al., 1999b), *N. astigmatica* NIES-252 (Lemieux et al., 2014), and *N. pyriformis* MED1 (Gastineau et al., 2021). These studies on organellar genomes and nuclear SSU rDNA gene analyses have shown that *N. pyriformis* is sister to the rest of the *Nephroselmis* species (Yamaguchi et al., 2011), although earlier phylogenies indicated that *N. pyriformis* might group with the *N olivacea*/*N. viridis* clade, while the *N. astigmatica* and *N. anterostigmatica* group branche with other *Nephroselmis* species (Nakayama et al., 2007). Our draft assembly could aid in resolving the phylogenetic relationships within *Nephroselmis* when additional large-scale genomic data become available from different *Nephroselmis* species.

Constitutive mixoplankton, protists with permanent plastids that have the genetic capacity for both photosynthesis and phagotrophy, such as some prasinophytes, are versatile, highly adaptable organisms that tend to alternate between each trophic mode (Mitra et al., 2016). Their environmentally driven fluctuations between photosynthesis and predation, to limit the energetic cost of constantly using both simultaneously, likely have significant impacts on the food webs and biogeochemical cycles that rely on primary production (Glibert & Mitra, 2022; Mitra & Flynn, 2023; Ward & Follows, 2016). To better elucidate their behavior in the oceans, we need a stronger understanding of how constitutive mixoplankton adjust their trophic mode in response to surrounding environmental conditions (Anderson et al., 2018; Bock et al., 2021).

Comparative transcriptomics, essential in the study of gene expression, has already proved useful in investigating the mixotrophic nature and feeding behaviors of small constitutive mixoplankton among the chrysophytes (Lie et al., 2017, 2018), haptophytes (Liu et al., 2015, 2016), and prasinophytes (Charvet et al., 2024). However, assembly and annotation of transcriptomes can be limited by the lack of available nuclear genomes (Kwon et al., 2023). Hence, we expect that this *Nephroselmis pyriformis* nuclear genome will facilitate gene expression studies of the species, which is a known mixotroph with the ability to ingest bacteria in conditions of low nutrient availability (Bock et al., 2021). Interestingly, different species of *Nephroselmis* seem to respond differently to the availability of macronutrients, micronutrients, and vitamins (Anderson et al., 2018), highlighting the need for further, in-depth, investigations into feeding behaviors at the individual species level.

Our *Nephroselmis pyriformis* RCC618 nuclear genome represents a high-quality and nearly complete reconstruction. This assembly not only will enhance our understanding of *N. pyriformis* biology and ecology but also lays a critical foundation for future transcriptomic studies, particularly in the context of mixoplankton trophic flexibility and environmental adaptability. As more genomic data from related species become available, this resource will support a range of studies, including green algal phylogenomics, genome evolution, and species-specific ecological strategies.

AUTHOR CONTRIBUTIONS

Sophie Charvet: Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (lead); methodology (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Eunsoo Kim:** Conceptualization (equal); data curation (supporting); formal analysis (equal); funding acquisition (lead); investigation (supporting); methodology (supporting); project administration (equal); visualization (supporting); writing – original draft (supporting); writing – review and editing (equal).

ACKNOWLEDGMENTS

This work was supported by the Simons Foundation (SF-876199), the National Science Foundation (NSF CAREER-1453639), the Basic Science Research Program through the National Research Foundation of Korea (RS-2025-00562426), and funds generously provided by an anonymous donor (to EK).

DATA AVAILABILITY STATEMENT

The raw sequence data are available in the NCBI short read archive database (SRA) under BioProject PRJNA1303626 and BioSample SAMN50518886, with the following accession numbers: SRR34921878 for the Illumina reads and SRR34921877 for the PacBio reads. The de novo genome assembly is available through NCBI's Genome resources (*under submission*: SUB15687033). Additional annotation data, including predicted proteins, tRNA, and rRNA gene sequences, are available from the Figshare repository (https://doi.org/10.6084/m9.figshare.30281515).

ORCID

Eunsoo Kim https://orcid.org/0000-0001-6691-8461

REFERENCES

Anderson, R., Charvet, S., & Hansen, P. (2018). Mixotrophy in chlorophytes and haptophytes – effect of irradiance, macronutrient, micronutrient and vitamin limitation. *Frontiers in Microbiology*, *9*, 1704. https://doi.org/10.3389/fmicb.2018.01704

Archibald, J. M. (2009). The puzzle of plastid evolution. *Current Biology*, *19*, R81–R88. https://doi.org/10.1016/j.cub.2008.11.067

Bachy, C., Sudek, L., Choi, C. J., Eckmann, C. A., Nöthig, E. M., Metfies, K., & Worden, A. Z. (2022). Phytoplankton surveys in the arctic Fram Strait demonstrate the tiny eukaryotic alga *Micromonas* and other Picoprasinophytes contribute to deep sea export. *Microorganisms*, *10*, 961. https://doi.org/10.3390/microorganisms10050961

Bell, E., & Laybourn-Parry, J. (2003). Mixotrophy in the Antarctic phytoflagellate, *Pyramimonas gelidicola* (Chlorophyte: Prasinophyceae). *Journal of Phycology*, *39*, 644–649. https://doi.org/10.1046/j.1529-8817.2003.02152.x

Blaz, J., Galindo, L. J., Heiss, A. A., Kaur, H., Torruella, G., Yang, A., Alexa Thompson, L., Filbert, A., Warring, S., Narechania, A., Shiratori, T., Ishida, K., Dacks, J. B., López-García, P., Moreira, D., Kim, E., & Eme, L. (2023). One high quality genome and two transcriptome datasets for new species of *Mantamonas*, a deep-branching eukaryote clade. *Scientific Data*, *10*, 603. https://doi.org/10.1038/s41597-023-02488-2

Bock, N., Charvet, S., Burns, J., Gyaltshen, Y., Rozenberg, A., Kim, E., & Duhamel, S. (2021). Experimental identification and *in silico* prediction of bacterivory in green algae. *The ISME Journal*, *15*, 1987–2000. https://doi.org/10.1038/s41396-021-00899-w

Brouard, J. S., Otis, C., Lemieux, C., & Turmel, M. (2010). The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. *Genome Biology and Evolution*, *2*, 240–256. https://doi.org/10.1093/gbe/evq014

Brůna, T., Lomsadze, A., & Borodovsky, M. (2024). GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Research*, *34*, 757–768. https://doi.org/10.1101/gr.278373.123

Burns, J., Paasch, A., Narechania, A., & Kim, E. (2015). Comparative genomics of a bacterivorous green alga reveals evolutionary causalities and consequences of phago-mixotrophic mode of nutrition. *Genome Biology and Evolution*, *7*, 3047–3061. https://doi.org/10.1093/gbe/evv144

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, Orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research*, *44*, e147. https://doi.org/10.1093/nar/gkw654

Chan, P. P., & Lowe, T. M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods in Molecular Bioloy*, *1962*, 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Charvet, S., Bock, N. A., Kim, E., & Duhamel, S. (2024). Transcriptomics reveal a unique phago-mixotrophic response to low nutrient concentrations in the prasinophyte *Pterosperma cristatum*. *ISME Communications*, *4*, ycae083. https://doi.org/10.1093/ismeco/ycae083

Crépeault, O., Otis, C., Pombert, J., Turmel, M., & Lemieux, C. (2024). Comparative plastome and mitogenome analyses indicate that the marine prasinophyte green algae *Pycnococcus provasolii* and *Pseudoscourfieldia marina* (Pseudoscourfieldiophyceae class nov., Chlorophyta) represent morphotypes of the same species. *Journal of Phycology*, *60*, 1021–1027. https://doi.org/10.1111/jpy.13482

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*, giab008. https://doi.org/10.1093/gigascience/giab008

Dennu, L., Devic, M., Rigonato, J., Falciatore, A., Lozano, J.-C., Vergé, V., Mariac, C., Jaillon, O., The Dark Edge Genomics Sampling Team, Sabot, F., & Bouget, F.-Y. (2023). Biological and genomic resources for the cosmopolitan phytoplankton *Bathycoccus*: Insights into genetic diversity and major structural variations. *BioRxiv*, 1–38. https://doi.org/10.1101/2023.10.16.562038

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piégu, B., Ball, S. G., Ral, J. P., Bouget, F. Y., … Moreau, H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 1164711652. https://doi.org/10.1073/pnas.0604795103

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 238. https://doi.org/10.1186/s13059-019-1832-y

Fang, L., Leliaert, F., Novis, P. M., Zhang, Z., Zhu, H., Liu, G., Penny, D., & Zhong, B. (2018). Improving phylogenetic inference of core Chlorophyta using chloroplast sequences with strong phylogenetic signals and heterogeneous models. *Molecular Phylogenetics and Evolution*, *127*, 248–255. https://doi.org/10.1016/j.ympev.2018.06.006

Flynn, J., Hubley, R., Goubert, C., Rosen, J., Clark, A., Feschotte, C., & Smit, A. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Fučíková, K., Leliaert, F., Cooper, E. D., Škaloud, P., D'Hondt, S., De Clerck, O., Gurgel, C. F. D., Lewis, L. A., Lewis, P. O., Lopez-Bautista, J. M., Delwiche, C. F., & Verbruggen, H. (2014). New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Frontiers in Ecology and Evolution*, *2*, 1–12. https://doi.org/10.3389/fevo.2014.00063

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Research*, *34*, 769–777. https://doi.org/10.1101/gr.278090.123

Galaxy Community. (2024). The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, *52*, W83–W94. https://doi.org/10.1093/nar/gkae410

Gastineau, R., Konucu, M., Tekdal, D., Lemieux, C., Turmel, M., Witkowski, A., & Eker-Develi, E. (2021). A gene-rich and compact chloroplast genome of the green alga *Nephroselmis pyriformis* (N.Carter) Ettl 1982 from the shores of Mersin (Eastern Mediterranean Sea). *Mitochondrial DNA Part B: Resources*, *6*, 308–310. https://doi.org/10.1080/23802359.2020.1866461

Ghyoot, C., Lancelot, C., Flynn, K. J., Mitra, A., & Gypens, N. (2017). Introducing mixotrophy into a biogeochemical model describing a eutrophied coastal ecosystem: The southern North Sea. *Progress in Oceanography*, *157*, 1–11. https://doi.org/10.1016/j.pocean.2017.08.002

Glibert, P. M., & Mitra, A. (2022). From webs, loops, shunts, and pumps to microbial multitasking: Evolving concepts of marine microbial ecology, the mixoplankton paradigm, and implications for a future ocean. *Limnology and Oceanography*, *67*, 585–597. https://doi.org/10.1002/lno.1201

Guillard, R. (1975). Culture of phytoplankton for feeding marine invertebrates. In W. Smith & M. Chanley (Eds.), *Culture of marine invertebrate animals:* Proceedings – 1st Conference on Culture of Marine Invertebrate Animals Greenport (pp. 29–60). Springer Nature.

Gyaltshen, Y., Rozenberg, A., Paasch, A., Burns, J. A., Warring, S., Larson, R., Maurer-Alcalá, X. X., Dacks, J., Narechania, A., &

Kim, E. (2023). Long-read-based genome assembly reveals numerous endogenous viral elements in the green algal bacterivore *Cymbomonas tetramitiformis*. *Genome Biology and Evolution*, *15*, evad194. https://doi.org/10.1093/gbe/evad194

Hartmann, M., Grob, C., Tarran, G. A., Martin, A. P., Burkill, P. H., & Scanlan, D. J. (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 5756–5760. https://doi.org/10.1073/pnas.1118179109

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*, 518–522. https://doi.org/10.1093/molbev/msx281

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*, D309–D314. https://doi.org/10.1093/nar/gky1085

Joli, N., Monier, A., Logares, R., & Lovejoy, C. (2017). Seasonal patterns in Arctic prasinophytes and inferred ecology of *Bathycoccus* unveiled in an Arctic winter metagenome. *The ISME Journal*, *11*, 1372–1385. https://doi.org/10.1038/ismej.2017.7

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589.

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of Molecular Biology*, *428*, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., & Zdobnov, E. M. (2023). OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, *51*, D445–D451. https://doi.org/10.1093/nar/gkac998

Kwon, T., Hanschen, E. R., & Hovde, B. T. (2023). Addressing the pervasive scarcity of structural annotation in eukaryotic algae. *Scientific Reports*, *13*, 1687. https://doi.org/10.1038/s41598-023-27881-0

Langmead, B., & Salzberg, S. (2013). Fast gapped-read alignment with Bowtie2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923.Fast

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*, 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Leliaert, F. (2019). Green algae: Chlorophyta and Streptophyta. In T. M. Schmidt (Ed.), *Encyclopedia of microbiology* (4th ed., Vol. 2). Elsevier Inc. https://doi.org/10.1016/b978-0-12-809633-8.20890-x

Leliaert, F., Smith, D. R., Herron, M. D., Verbruggen, H., Delwiche, C. F., & Clerck, O. (2012). Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences*, *31*, 1–46. https://doi.org/10.1080/07352689.2011.615705

Lemieux, C., Otis, C., & Turmel, M. (2007). A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology*, *5*, 2. https://doi.org/10.1186/1741-7007-5-2

Lemieux, C., Otis, C., & Turmel, M. (2014). Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics*, *15*, 1–20.

Lemieux, C., Turmel, M., Otis, C., & Pombert, J. F. (2019). A streamlined and predominantly diploid genome in the tiny marine green alga *Chloropicon primus*. *Nature Communications*, *10*, 4061. https://doi.org/10.1038/s41467-019-12014-x

Li, L., Wang, S., Wang, H., Sahu, S. K., Marin, B., Li, H., Xu, Y., Liang, H., Li, Z., Cheng, S., Reder, T., Çebi, Z., Wittek, S., Petersen, M., Melkonian, B., Du, H., Yang, H., Wang, J., Wong, G. K., … Liu, H. (2020). The genome of *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nature Ecology & Evolution*, *4*, 1220–1231. https://doi.org/10.1038/s41559-020-1221-7

Li, W., Mclaughlin, F., Lovejoy, C., & Carmack, E. (2009). Smallest algae thrive as the Arctic Ocean freshens. *Science*, *326*, 539. https://doi.org/10.1126/science.1179798

Lie, A. A. Y., Liu, Z., Terrado, R., Tatters, A. O., Heidelberg, K. B., & Caron, A. (2018). A tale of two mixotrophic chrysophytes: Insights into the metabolisms of two *Ochromonas* species (Chrysophyceae) through a comparison of gene expression. *PLoS ONE*, *13*, 1–20. https://doi.org/10.1371/journal.pone.0192439

Lie, A. A. Y., Liu, Z., Terrado, R., Tatters, A. O., Heidelberg, K. B., & Caron, D. A. (2017). Effect of light and prey availability on gene expression of the mixotrophic chrysophyte, *Ochromonas* sp. *BMC Genomics*, *18*, 163. https://doi.org/10.1186/s12864-017-3549-1

Lin, H. H., & Liao, Y. C. (2016). Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Scientific Reports*, *6*, 12–19. https://doi.org/10.1038/srep24175

Liu, Z., Campbell, V., Heidelberg, K. B., & Caron, D. A. (2016). Gene expression characterizes different nutritional strategies among three mixotrophic protists. *FEMS Microbiology Ecology*, *92*, 1–11. https://doi.org/10.1093/femsec/fiw106

Liu, Z., Jones, A., Campbell, V., Hambright, K., Heidelberg, K., & Caron, D. (2015). Gene expression in the mixotrophic prymnesiophyte, *Prymnesium parvum*, responds to prey availability. *Frontiers in Microbiology*, *6*, 1–12. https://doi.org/10.3389/fmicb.2015.00319

Lovejoy, C., Vincent, W. F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R., Potvin, M., Massana, R., & Pedrós-Alió, C. (2007). Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic seas. *Journal of Phycology*, *43*, 78–89. https://doi.org/10.1111/j.1529-8817.2006.00310.x

Maddison, W. P., & Maddison, D. R. (2023). Mesquite: A modular system for evolutionary analysis. Version 4.02.

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*, 10–12. https://doi.org/10.14806/ej.17.1.200

Maruyama, S., & Kim, E. (2013). A modern descendant of early green algal phagotrophs. *Current Biology*, *23*, 1081–1084. https://doi.org/10.1016/j.cub.2013.04.063

Maruyama, S., & Kim, E. (2020). Evolution of photosynthetic eukaryotes; current opinion, perplexity, and a new perspective. *Symbiosis: Cellular, Molecular, Medical and Evolutionary Aspects*, *69*, 337–351. https://doi.org/10.1007/978-3-030-51849-3_12

Mckie-Krisberg, Z., Gast, R., & Sanders, R. (2015). Physiological responses of three species of Antarctic mixotrophic phytoflagellates to changes in light and dissolved nutrients. *Microbial Ecology*, *70*, 21–29. https://doi.org/10.1007/s00248-014-0543-x

McManus, H. A., Fučíková, K., Lewis, P. O., Lewis, L. A., & Karol, K. G. (2018). Organellar phylogenomics inform systematics in the green algal family Hydrodictyaceae (Chlorophyceae) and provide clues to the complex evolutionary history of plastid genomes in the green algal tree of life. *American Journal of Botany*, *105*(3), 315–329. https://doi.org/10.1002/ajb2.1066

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., … Grossman, A. R. (2007). The *Chlamydomonas genome* reveals the evolution of key animal and plant functions. *Science*, *318*, 245–250. https://doi.org/10.1126/science.1143609

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*, 1530–1534.

Mitra, A., & Flynn, K. J. (2023). Low rates of bacterivory enhances phototrophy and competitive advantage for mixoplankton growing in oligotrophic waters. *Scientific Reports*, *13*, 6900. https://doi.org/10.1038/s41598-023-33962-x

Mitra, A., Flynn, K., Tillmann, U., Raven, J., Caron, D., Stoecker, D., Not, F., Hansen, P., Hallegraeff, G., Sanders, R., Wilken, S., McManus, G., Johnson, M., Pitta, P., Våge, S., Berge, T., Calbet, A., Thingstad, F., Jeong, H., & Lundgren, V. (2016). Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: Incorporation of diverse mixotrophic strategies. *Protist*, *167*, 106–120. https://doi.org/10.1016/j.protis.2016.01.003

Moreau, H., Verhelst, B., Couloux, A., Derelle, E., Rombauts, S., Grimsley, N., Van Bel, M., Poulain, J., Katinka, M., Hohmann-Marriott, M. F., Piganeau, G., Rouzé, P., Da Silva, C., Wincker, P., Van de Peer, Y., & Vandepoele, K. (2012). Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology*, *13*(8), R74. https://doi.org/10.1186/gb-2012-13-8-r74

Nakayama, T., Suda, S., Kawachi, M., & Inouye, I. (2007). Phylogeny and ultrastructure of *Nephroselmis* and *Pseudoscourfieldia* (Chlorophyta), including the description of *Nephroselmis anterostigmatica* sp. nov. and a proposal for the Nephroselmidales ord. nov. *Phycologia*, *46*, 680–697.

Pertea, G., & Pertea, M. (2020). GFF utilities: GffRead and GffCompare. *F1000Research*, *9*, J-304. https://doi.org/10.12688/f1000research.23297.2

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3

Rey Redondo, E., Xu, Y., & Yung, C. C. M. (2024). Genomic characterisation and ecological distribution of *Mantoniella tinhauana*: A novel Mamiellophycean green alga from the Western Pacific. *Frontiers in Microbiology*, *15*, 1358574. https://doi.org/10.3389/fmicb.2024.1358574

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Connor, R., Feldgarden, M., Fine, A. M., Funk, K., Hoffman, J., Kannan, S., Kelly, C., Klimke, W., Kim, S., Lathrop, S., Marchler-Bauer, A., Murphy, T. D., O'Sullivan, C., Schmieder, E., … Pruitt, K. D. (2025). Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Research*, *53*, D20–D29. https://doi.org/10.1093/nar/gkae979

Seemann, T. (2013). Barrnap 0.9: rapid ribosomal RNA prediction. https://github.com/tseemann/barrnap

Sibbald, S. J., & Archibald, J. M. (2020). Genomic insights into plastid evolution. *Genome Biology and Evolution*, *12*(7), 978–990. https://doi.org/10.1093/GBE/EVAA096

Song, L., & Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, *4*(1), 48. https://doi.org/10.1186/s13742-015-0089-y

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Suda, S., Watanabe, M. M., & Inouye, I. (1989). Evidence for sexual reproduction in the primitive green alga *Nephroselmis olivacea* (Prasinophyceae). *Journal of Phycology*, *25*, 596–600. https://doi.org/10.1111/j.1529-8817.1989.tb00266.x

Suda, S., Watanabe, M. M., & Inouye, I. (2004). Electron microscopy of sexual reproduction in *Nephroselmis olivacea* (Prasinophyceae, Chlorophyta). *Phycological Research*, *52*, 273–283. https://doi.org/10.1111/j.1440-183.2004.00346.x

Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, *4*, 10.1. https://doi.org/10.1002/0471250953.bi0410s25

Tegenfeldt, F., Kuznetsov, D., Manni, M., Berkeley, M., Zdobnov, E. M., & Kriventseva, E. V. (2025). OrthoDB and BUSCO update: Annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research*, *53*, D516–D522. https://doi.org/10.1093/nar/gkae987

Tenenbaum, D., & Volkening, J. (2021). KEGGREST: Client-side REST access to the Kyoto encyclopedia of genes and genomes (KEGG).

Turmel, M., & Lemieux, C. (2018). Evolution of the plastid genome in green algae. *Advances in Botanical Research*, *85*, 157–193. https://doi.org/10.1016/bs.abr.2017.11.010

Turmel, M., Gagnon, M. C., O'Kelly, C. J., Otis, C., & Lemieux, C. (2009). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Molecular Biology and Evolution*, *26*, 631–648. https://doi.org/10.1093/molbev/msn285

Turmel, M., Lemieux, C., Burger, G., Lang, B. F., Otis, C., Plante, I., & Gray, M. W. (1999a). The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*: Two radically different evolutionary patterns within green algae. *Plant Cell*, *11*, 1717–1730. https://doi.org/10.1105/tpc.11.9.1717

Turmel, M., Otis, C., & Lemieux, C. (1999b). The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 10248–10253. https://doi.org/10.1073/pnas.96.18.10248

Ward, B. A., & Follows, M. J. (2016). Marine mixotrophy increases trophic transfer efficiency, mean organism size, and vertical carbon flux. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 2958–2963. https://doi.org/10.1073/pnas.1517118113

Weib, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, *19*, 122. https://doi.org/10.1186/s12859-018-2128-z

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*, R46.

Worden, A., Lee, J.-H., Mock, T., Rouzé, P., Simmons, M., Aerts, A., Allen, A., Cuvelier, M., Derelle, E., Everett, M., Foulon, E., Grimwood, J., Gundlach, H., Henrissat, B., Napoli, C., McDonald, S., Parker, M., Rombauts, S., Salamov, A., & Grigoriev, I. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *micromonas*. *Science*, *324*, 268–272. https://doi.org/10.1126/science.1167222

Yamaguchi, H., Suda, S., Nakayama, T., Pienaar, R. N., Chihara, M., & Inouye, I. (2011). Taxonomy of *Nephroselmis viridis* sp. nov. (Nephroselmidophyceae, Chlorophyta), a sister marine species to freshwater *N. olivacea*. *Journal of Plant Research*, *124*, 49–62. https://doi.org/10.1007/s10265-010-0349-y

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler.

Bioinformatics, *29*, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Zubkov, M., & Tarran, G. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature*, *455*, 224–226. https://doi.org/10.1038/nature07236

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1. 18S rRNA gene sequence alignment used to make the phylogenetic tree presented in Figure S1.

Data S2. Illumina sequencing coverage for each scaffold.

Data S3. PacBio sequencing coverage for each scaffold.

Figure S1. A scatter plot showing the Illumina sequencing depth of scaffolds.

Figure S2. A scatter plot showing the PacBio sequencing depth of scaffolds.

Figure S3. Maximum likelihood phylogenetic tree using 18S rDNA sequences, showing the placement of the strain RCC618 (in bold). A total of 1775 nucleotide sites were included and the best-fit model of TIM2e + I + G4 was selected for inference. The scale bar refers to the nucleotide substitutions per site. Bootstrap support values (≥50%) are shown at the corresponding nodes.

Figure S4. GenomeScope profile output. GenomeScope analyses based on the output from jellyfish run with the k-mer size of 21 suggest the estimated genome size of 59.85 Mbp and the repeat content of 19.42% when the genome is assumed to be haploid ($p = 1$).

Figure S5. Biallelic SNP frequency distribution. The *x*-axis represents allele frequency, and the *y*-axis shows the count. The distribution pattern supports the haploid state of the genome.

Table S1. Whole genome sequencing read statistics for *Nephroselmis pyriformis* RCC618.

Table S2. Statistics of nuclear genome assemblies for *Nephroselmis pyriformis* and four select green algal species. BUSCO analyses were run on the genome assemblies using the eukaryota_odb12 (129 genes) and viridiplantae_odb12 (822 genes) datasets. Abbreviations: complete BUSCOs (C); complete and single-copy BUSCOs (S); complete and duplicated BUSCOs (D); fragmented BUSCOs (F); missing BUSCOs (M).

Table S3. Comparison of predicted protein sets for *Nephroselmis pyriformis* RCC618 obtained with three different approaches. Completeness of the protein sets was evaluated with BUSCO on the eukaryota_odb12 (129 genes) and viridiplantae_odb12 (822 genes) datasets.

How to cite this article: Charvet, S., & Kim, E. (2025). Nuclear genome assembly and annotation for the phagotrophic green alga *Nephroselmis pyriformis*. *Journal of Phycology*, *61*, 1566–1575. https://doi.org/10.1111/jpy.70109