LA-UR-21-25757
Accepted Manuscript

Phylogenetic analyses and reclassification of the oleaginous marine species Nannochloris sp. “desiccata" (Trebouxiophyceae, Chlorophyta), formerly Chlorella desiccata, supported by a high-quality genome assembly

Sanders, Claire Kathleen
Hanschen, Erik Richard
Biondi, Thomas Christopher
Hovde, Blake
Kunde, Yuliya A.
Eng, Wyatt Lee Kok Ming
Kwon, Taehyung
Dale, Taraka T.

To be published in: Journal of Phycology

DOI to publisher's version: 10.1111/jpy.13242

Permalink to record:
http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-21-25757

MRS. CLAIRE K SANDERS (Orcid ID : 0000-0001-7848-2642)
DR. TARAKA DALE (Orcid ID : 0000-0001-9036-1210)

Corresponding author mail id : tdale@lanl.gov

Phylogenetic analyses and reclassification of the oleaginous marine species *Nannochloris* sp. 'desiccata' (trebouxiophyceae, chlorophyta), formerly *Chlorella desiccata*, supported by a high-quality genome assembly[1]

Claire K. Sanders, Erik R. Hanschen, Thomas C. Biondi, Blake T. Hovde, Yuliya A. Kunde,
Wyatt L. Eng, Taehyung Kwon, Taraka Dale[2]
Los Alamos National Laboratory, Bioscience Division
PO Box 1663, MS M888, Los Alamos, NM 87545

[2]tdale@lanl.gov; 505-412-7691

Running title - *Nannochloris* sp. 'desiccata' genome

Editorial Responsibility: H. Verbruggen (Associate Editor)

ABSTRACT

Microalgae are diverse, with many gaps remaining in phylogenetic and physiological understanding. Thus, studying new microalgae species increases our broader comprehension of biological diversity, and evaluation of new candidates as algal production platforms can lead to improved productivity under a variety of cultivation conditions. *Chlorella* is a genus of fast-growing species often isolated from freshwater habitats and cultivated as a source of nutritional supplements. However, the use of freshwater increases competition with other freshwater needs. We identified *Chlorella desiccata* to be worthy of further investigation as a potential algae production strain, due to its isolation from a marine environment and its promising growth and biochemical composition properties. Long-read genomic sequencing was conducted for *C. desiccata* UTEX 2526, resulting in a high-quality, near chromosome level, diploid genome with an assembly length of 21.55 Mbp in only 18 contigs. We also report complete circular mitochondrial and chloroplast genomes. Phylogenomic and phylogenetic analyses using nuclear, chloroplast, 18S rRNA, and actin sequences revealed that this species clades within strains currently identified as *Nannochloris* (Trebouxiophyceae, Chlorophyta), leading to its reclassification as *Nannochloris* sp. 'desiccata' UTEX 2526. The mode of cell division for this species is autosporulation, differing from the type species *N. bacillaris*. As has occurred across multiple microalgae genera, there are repeated examples of *Nannochloris* species reclassification in the literature. This high-quality genome assembly and phylogenetic analysis of the potential algal production strain *Nannochloris* sp. 'desiccata' UTEX 2526 provides an important reference and useful tool for further studying this region of the phylogenetic tree.

Keywords: algae, *Chlorella,* Chlorophyta, cryptic species, microalgae, morphology, *Nannochloropsis*, phylogeny, salinity, taxonomy

Abbreviations: BUSCO, Benchmarking Universal Single-Copy Orthologs; OMA, Orthologous Matrix; UTEX, Culture Collection of Algae at the University of Texas at Austin

INTRODUCTION

Investigation of microalgae as a feedstock for biotechnological and nutraceutical applications has shown great promise. Several species of *Chlorella* are being cultivated for nutritional supplements and have been investigated for biofuel production due to their high biomass accumulation rates (Negi et al. 2016, Posten and Chen 2016, Andrade and Andrade 2017, Huesemann et al. 2017, 2018, Bito et al. 2020). However, most species of *Chlorella* have been isolated from freshwater sources and show a growth reduction in marine media (Herrera-Valencia et al. 2011, Rai et al. 2014, Neofotis et al. 2016). The use of primarily brackish or marine water for cultivation reduces resource competition with crops, animals, and humans for freshwater and reduces production costs (Singh et al. 2011, Venteris et al. 2013). This brings to question whether a marine *Chlorella* with similar production properties as its freshwater counterparts can be identified. *Chlorella desiccata* was isolated from a marine environment, is cultivated in marine media, and is commercially available (UTEX 2437 and UTEX 2526; Margulis et al. 1988). This species shows potential as an algal production strain; it has promising growth rates and accumulates 25% triacylglycerides per dry weight within 3 days of nitrogen depletion (Zalogin and Pick 2014, Avidan and Pick 2015). Genetic engineering of this species has been demonstrated by expressing a foreign gene to control mosquito larvae (Borovsky et al. 2016).

Historically, phylogeny has been based on phenotype and morphology; however, this classification system in organisms such as microalgae can be inadequate to classify species, due to the lack of distinct feature descriptions that allow researchers to accurately identify that species in the future (Krienitz and Bock 2012, De Clerck et al. 2013). Additionally, microalgae often have extensive phenotypic plasticity (Morales and Trainor 1997, Morales et al. 2002, Schaum et al. 2013, Sassenhagen et al. 2015); for example, environment can heavily influence cell morphology, leading to further classification challenges (Luo et al. 2006, Bock et al. 2011, Darienko et al. 2019). This cryptic diversity has led to the frequent re-examination of algae species, both by more-thorough phenotypic analysis and via sequence analysis (Bock et al. 2011, Darienko et al. 2019, Krivina and Temraleeva 2020). Many *Chlorella* species isolated from marine environments and deposited into culture collections were originally identified on the basis of morphology. However, more recent analyses, utilizing 18S rRNA and rRNA internal transcribed spacer (ITS) sequences, showed that many of these species actually belonged to other genera such as *Chloroidium* or *Droopiella* (Darienko et al. 2019). In another example, the genus *Mychonastes* underwent a major revision after ITS analysis revealed that the entire genus *Pseudodictyosphaerium*, which had previously been reclassified based on lack of pyrenoids from *Dictyosphaerium*, should be included in the *Mychonastes* genus (Krienitz et al. 2011). While not renamed in the culture collection, *C. desiccata* was previously reclassified as *Mychonastes desiccatus* (Chlorophyceae), based on phenotypic properties (Margulis et al. 1988).

Nannochloris is another genus that has undergone substantial revisions. It was originally described by the species *N. bacillaris* and *N. coccoides*, which are portrayed as small, 2.5-3.5 µm, oval or coccoid organisms that divide into two daughter cells by binary fission (Naumann 1921, Butcher 1952). However, subsequent genetic analysis revealed that these two species are distinct enough to belong to different genera, and *N. coccoides* was reclassified as *Marvania coccoides*, leaving *N. bacillaris* as the type species (Henley et al. 2004). A look into species originally classified as *Nannochloris* showed that cell division phenotype within this genus encompass a wide variety of modes, including autosporulation, binary fission, and budding (Yamamoto et al. 2001). Extensive 18S rRNA analysis compromising many of these studied species showed a clearly distinct cluster that grouped *Nannochloris atomus* CCAP 251/7 and *Nannochoris maculatus* CCAP 251/3 with *Picochlorum oklamomensis* (Henley et al. 2004). Additionally, these strains all divide by autosporulation, further defining what Henley et al. (2004) described as the *Picochlorum* genus, supporting their reclassification to *P. atomus* and *P. maculatus*, respectively. It has since been noted, however, that the mode of cell division does not necessarily predict genetic relatedness, with Yamamoto et al. (2007) showing that closely related species demonstrate varied cell division methods. Culture conditions also strongly affect growth, with varying light, temperature, salinity, and CO_2 concentrations greatly influencing the reproductive capacity of cells (Cho et al. 2007, Miyagishima et al. 2014, Cross and Umen 2015). Thus, growth conditions can affect the number of daughter cells produced from a single division event as well as changing the apparent morphology (Vítová and Zachleder 2005), further supporting the concept that phenotypic and molecular characterization must be used in concert for phylogenetic classification (Krienitz and Bock 2012).

Chlorella is an oleaginous and salt-tolerant genera that warrants study as a potential algal production strain; therefore, we conducted a sequence anlaysis of a PCR amplicon of the *C. desiccata* UTEX 2526 18S rRNA gene. These data, as well as examination of cell division mode and genome size estimation, indicated that UTEX 2526 is neither a *Chlorella* nor a *Mychonastes* but is instead more likely to be related to *Nannochloris* or *Picochlorum*. To fully determine the relationship between UTEX 2526 and other known algal species, we conducted long-read genome sequencing, assembly, and annotation of *C. desiccata* UTEX 2526. Pylogenomic and phylogenetic analyses utilizing nuclear, chloroplast, 18S rRNA, and actin sequences show that this species consistently and significantly clades with several *Nannochloris* species, including

Nannochloris bacillaris, leading us to classify this strain as *Nannochloris* sp. 'desiccata' UTEX 2526 (Trebouxiophyceae, Chlorophyta). As no other high-quality genomes exist for any species of *Nannochloris*, the availability of this first complete genome, with 18 contigs, including eight full chromosomes, for this dynamic genus will assist the field in future classification, gene expression analysis, and development of genetic engineering tools (Hanschen and Starkenburg 2020).

MATERIALS AND METHODS

Strain information

A culture of *Chlorella desiccata* UTEX 2526 was purchased from the UTEX culture collection. *C. desiccata* UTEX 2437 was originally isolated from a desiccated microbial mat in Laguna Figueroa, Baja California, Mexico and was redeposited as an axenic culture, *C. desiccata* UTEX 2526, in 1989.

Culture conditions

Nannochloris desiccata UTEX 2526 (previously *Chlorella desiccata*) culture was maintained on 17.5 ppt salinity f/2 plates under 40 µmoles photons $\cdot$ m^{-2} $\cdot$ s^{-1} at room temperature. Liquid cultures were grown in modified, silicate-free, 35 salinity f/2 media in a 1% CO_2 atmosphere, as previously described (Gonzalez-Esquer et al. 2019). Prior to DNA extraction for sequencing, cultures were illuminated with 300 µmoles photons $\cdot$ m^{-2} $\cdot$ s^{-1} with a 16-hour light/8-hour dark cycle. 250 mL of culture was harvested, cells were centrifuged at 3200 rpm for 5 min, pellets were flash frozen, and stored at -80°C. For microscopy, cells were grown under 300 µmoles photons $\cdot$ m^{-2} $\cdot$ s^{-1} in continuous light, such that the cell cycle of cultures were not synchronized, samples from a rapidly dividing culture were placed at 4°C until analysis. Flow cytometry was conducted on cells grown in 1 L spin flask photobioreactors with 800 µmoles photons $\cdot$ m^{-2} $\cdot$ s^{-1} and a 16:8 h light:dark dark cycle with CO_2 on demand, maintaining a pH of 7.0.

18S rRNA gene amplicon sequencing

Degenerate primers that amplify the region 18S small subunit of eukaryotic rDNA genes were used (F1391-EukBr pair). The first round of PCR amplified 18S small subunit from 10 ng DNA using KAPA HiFi HotStart Ready Mix. The first round of PCR used a denaturation temperature of 94°C for 3 min, 20 cycles of 94°C for 45 s, 57°C for 60 s and 72°C for 60 s and followed by an extension of 72°C for 10 min. The second round of PCR used Nextera XT v2 indexes (Illumina, San Diego, CA, USA) and used a denaturation temperature of 95°C for 3 min, 8 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 30 s and followed by an extension of 72°C for 5 min. The amplicons were cleaned up using AMPure XP beads (Beckman Coulter, Brea, CA, USA). The unique tags allowed for multiple amplicons to be pooled together. The concentration of the amplicons pool was obtained using the Qubit dsDNA HS Assay (ThermoFisher Scientific, Waltham, MA, USA). The average size of the library was determined by the Agilent High Sensitivity DNA Kit (Agilent Santa Clara, CA, USA). An accurate library quantification was determined using the Library Quantification Kit – Illumina/Universal Kit (KAPA Biosystems, Wilmington, MA, USA). The amplicon pool was sequenced on the Illumina MiSeq using the MiSeq Reagent Kit v3 (600-cycle; Illumina) to generate paired end 301 bp reads in expectation that each amplicon would produce approximately 100,000 reads.

Blastn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to query highly similar sequences to UTEX 2526 18S rRNA gene sequences using the standard nucleotide collection database. The top hits for percent identity and query coverage were collected. Additional Blast searches against species *Chlorella variabilis* and *Mychonastes timauensis* were run as they were suspected close relatives from the literature.

Microscopy

A *Nannochloris desiccata* culture, grown without light synchronization, was stained in media with 1 µg · mL^{-1} Hoechst 33342 (CalBioChem San Diego, CA, USA), a DNA-specific stain, for 30 minutes at room temperature prior to being placed on a glass slide. Images were taken on a Zeiss Observer D1 with an Axiocam MRm. Hoechst images were collected using Filter Set 49, optimized for DAPI, and chlorophyll fluorescence was collected with Filter Set 43 optimized for Cy3. Images were falsely colored. Overlays were assembled in ImageJ v1.53a (Schneider et al. 2012).

Genome size estimation

Cells were fixed in 70% ethanol/30% water for 30 min at -20°C, centrifuged at 11,000 rpm at room temperature for 2 min and resuspended in modified f/2 media. 50 µg · mL^{-1} RNase A and

12.5 μM of the DNA stain Syto 9 (ThermoFisher) was added and cultures were incubated at 37°C for 30 min then allowed to cool to room temperature before analysis. Flow cytometry was conducted on an Accuri C6+ (BD Biosciences) flow cytometer, collecting fluorescence with a 533/30 bandpass filter with 488nm excitation. Flow cytometry DNA content controls included the fully sequenced species: *Picochlorum soloecismus* (haploid – 15.2 Mbp; Gonzalez-Esquer et al. 2018), *Picochlorum celeri* (diploid – 13.5 Mbp; Becker et al. 2020), and *Chlorella sorokiniana* 1412 (haploid – 57.9 Mbp; Hovde et al. 2018). The major fluorescence peak for each culture was gated and the arithmetic mean of the fluorescence value graphed against known genome size. A predicted genome size for *N. desiccata* UTEX 2526 was calculated from a linear fit to these points.

Sequencing

Cells from the frozen algal pellet were thawed, washed with 3 mL of BG-11 media (UTEX 2009), centrifuged for 10 min at 3,500 rpm at 4°C, then the supernatant was decanted and cells were resuspended in a pH 7.2 buffer consisting of 200 mM NaCl, 100 mM EDTA pH 8, 10 mM, Tris pH 7.2. An equal volume of 1% low melting point agarose was added to the cell suspension and mixed gently by swirling the tube. The mix was pipetted into plug molds and incubated at 4°C to solidify. Agarose plugs were incubated in protoplasting solution, consisting of 1 M sorbitol, 24 mM EDTA pH 8.0, 10 mM beta-mercaptoethanol, 100 μM sodium citrate, 4% hemicellulase, and 2% driselase, overnight at 37°C with gentle mixing to remove the cell wall. Next, the plugs were incubated overnight in a lysis buffer containing 1% N-Lauroylsarcosine in 0.5M EDTA pH 9.5 with 5 mg/mL proteinase K. Lysed plugs were washed three times with 1x TE buffer and digested overnight with beta-Agarase I. Genomic DNA was purified according to the High salt:Phenol:Chloroform:IsoAmyl Alcohol protocol recommended by Pacific Biosciences (Pacific Biosciences of California 2014). Briefly, 400 μL of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to 500 μL of DNA suspended in Elution Buffer (Pacific Biosiences, Menlo Park, CA) with 1M NaCl and 2 mM EDTA. The solutions were mixed and centrifuged at 5,000 rpm at RT for 5 min, the aqueous layer was removed and saved in a new tube. 400 μL of the elution buffer with 1 M NaCl and 2 mM EDTA was added to the non-aqueous phase, mixed and centrifuged at 5,000 rpm at RT for 5 min. The aqueous layer was removed and added to the aqueous phase from the previous step. The combined aqueous solution was mixed with an equal volume of Chloroform:Isoamyl Alcohol (24:1). After centrifuging at 5,000 rpm at RT for 5 min a 0.3X volume of 99.9% ethanol was added to the aqueous layer and mixed. The sample was centrifuged at 10,000 rpm at RT for 15 min. The supernatant was collected and a 1.7X volume of 99.9% ethanol was added, mixed, and centrifuged at 7,500 rpm, RT for 15 min. The supernatant was removed, and precipitated gDNA was washed with 500 μL of 70% ethanol, centrifuged for 5 min at 10,000 rpm at RT, and the supernatant was removed. The 70% ethanol washing step was repeated. The ethanol was removed and the pellet allowed to air dry for 5 min before resuspension in TE buffer at 4°C overnight. Purified DNA was concentrated with AMPure PB beads and fragmented on Megaruptor 2 to a target size of 20 kbp. SMRTbell low input HiFi libraries were constructed according to the PacBio Express low input HiFi Template Prep protocol v6 (Pacific Biosciences of California 2020). The library was sequenced using two 1M SMRT cells on the Pacific Biosciences Sequel instrument with sequencing chemistry v3.0 and DNA Polymerase 3.0. 20-hour sequencing movies were recorded.

De-novo genome assembly

HiFi read sequences were extracted from the movies using PacBio's CCS v4.2.0 (https://github.com/PacificBiosciences/ccs) module. Initially, the HiFi reads were assembled using HiCanu v2.1 (Nurk et al. 2020) with the parameter "--pacbio-hifi". Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.1.4 (Seppey et al. 2019) analysis with the chlorophyta_odb10 dataset in "genome" mode showed that 80.1% of genes were duplicated, which was indicative of a diploid genome. Subsequently, the HiFi reads were assembled using a diploid-aware assembler, hifiasm v0.12-r304 (Cheng et al. 2021), with default parameters. The chloroplast and mitochondrial genomes were assembled by first assembling the entire genome with Flye v2.8.2-b1689 (Kolmogorov et al. 2019) and then identifying contigs with considerably higher coverage and lower GC content compared to the genome as a whole. Two contigs met this criteria, and were confirmed to be the non-nuclear genomes through positive identification of mitochondria and chloroplast-specific genes in these contigs (Tillich et al. 2017).

Genome annotation

Automated gene prediction was performed using BRAKER2 version 2.1.5 (Gremme 2012, Brůna et al. 2021). Splice alignment information was collected by mapping protein homology data that was created from combining protein sequences of a closely related species *Picochlorum renovo* and species-specific predicted genes using a BUSCO-trained gene model based on OrthoDB v10 chlorophyta genes. A predicted gene set using the BUSCO-trained gene model was generated through BUSCO pipeline (Seppey et al. 2019), which trains gene models with AUGUSTUS by utilizing a set of conserved OrthoDB genes from genome sequence (Stanke and Waack 2003, Waterhouse et al. 2018). We first obtained the preliminary gene models from BUSCO primary runs across four OrthoDB datasets (eukaryota_odb10, stramenopiles_odb10, alveolata_odb10, and chlorophyta_odb10). The BUSCO run with chlorophyta_odb10 was selected due to its highest percentage of BUSCO completeness. We trained a refined gene model through a BUSCO secondary run with the preliminary gene model, and performed AUGUSTUS gene prediction using the following parameters: --strand=both, --genemodel=complete, --protein=on, --introns=off, --start=off, --stop=off, --UTR=off, --codingseq=off, --alternatives-from-sampling=true, --minexonintronprob=0.1, --minmeanexonintronprob=0.4, --maxtracks=1, and --gff3=on. Protein sequences corresponding to the BUSCO-trained predicted genes were extracted using getAnnoFasta.pl of AUGUSTUS package.

A total of 221 genes were identified by GenomeThreader (Gremme 2012) as valid training genes for the final gene prediction model. Functional Gene motifs and domains were added with InterProScan version 5.26-65.0-intel-2017b (Blum et al. 2021) against the (list used interproscan databases) databases. Gene functions were then allocated using the best BLASTP (Altschul et al. 1990) match to the UniProt SwissProt database (Bateman et al. 2021).

The chloroplast and mitochondrial genomes were annotated using Prokka v.1.14.5 (Seemann 2014) via the EDGE platform (Li et al. 2017) with parameters --evalue 1e-09 and with either --kingdom Mitochondria (for the mitochondrion genome) or –kingdom Bacteria (for the chloroplast genome). The annotations produced by Prokka were then manually curated using a combination of blastx, blastn, blastp and QIAGEN CLC Genomics Workbench 20.0 (https://digitalinsights.qiagen.com/) to create a final set of annotations.

To compare whole genome statistics genome information for *Picochlorum soloecismus*, *P. celeri*, *P. renovo*, *Chlorella sorokiniana* 1412, *C. variabilis* NC64A, and *Coccomyxa* sp. LA000219 were taken from GenBank assembly information.

Phylogenetic analyses

Phylogenetic analysis was conducted with multiple loci to demonstrate confidence, using multiple independently recombining loci in our phylogenetic assignment and taxonomic classification. This assumes sexual reproduction of Trebouxiophyte green algae based on the presence of genes related to meiosis (Blanc et al. 2010, Fučíková et al. 2015, Hovde et al. 2018, 2019). However, if Trebouxiophyte lineages are exclusively asexually reproducing our analysis still applies, as these three loci then effectively belong to the same locus. A single supermatrix was not produced because the available chloroplast, 18S rRNA, and actin sequences for all taxa would result in a high proportion of empty cells and correspondingly low statistical support. All closely related Trebouxiophyte chloroplast sequences were included in the chloroplast tree estimation. For the 18S and actin trees, NCBI was searched for Trebiouxiophyte sequences ensuring a broad sampling of species across Trebouxiophyte algae and specifically including taxa closely related to *Picochlorum* and *Nannochloris.* These loci included the 18S small subunit of the ribosomal RNA (nucleotide sequences), actin (nucleotide sequences), and a concatenation of 26 chloroplast genes (protein sequences based on higher levels of divergence; Table S1 in the Supporting Information; Hovde et al. 2018).

For each analysis, we generated a phylogeny using Bayesian Markov chain Monte Carlo, implemented in MrBayes version 3.2.6 (Ronquist et al. 2012), and maximum-likelihood analyses, implemented in RAxML version 8.2.10 (Stamatakis 2014). For each analysis, genes were aligned using Muscle version 3.8.31 (Edgar 2004) before concatenation if applicable. The best-fitting combination of partitioning scheme and protein substitution models for the chloroplast phylogeny was determined using PartitionFinder version 2.1.1 (Lanfear et al. 2017) using AICc and a greedy search algorithm with branch lengths linked across partitions. For the chloroplast analysis, a total of 26 possible partitions were initially defined (one for each protein) and the best-fitting strategy included 17 data blocks (Table S1). For the 18S and actin analyses, PartitionFinder was used to determine the substitution models, identifying a General Time-Reversible model of nucleotide substitution under the gamma-distributed model of rate heterogeneity with an estimate of proportion of invariable sites (GTR+Γ+I) for both analyses. Four independent Bayesian runs of four chains each (three heated chains and one cold chain) were run for $5×10^6$ generations with a burn-in of $1×10^6$ generations. Trees were sampled every

100 generations. We considered the runs to have adequately sampled the solution space when the standard deviation of split frequencies was below 1×10^{-2}. Trees were independently constructed using maximum likelihood (ML) methods with the rapid bootstrap analysis and the same partition scheme. 450, 350, and 400 ML replicate trees were used to estimate bootstrap support for the chloroplast, 18S rRNA, and actin phylogenies respectively. There are insufficient sequences of other genes in available databases to perform additional meaningful phylogenetic analyses.

A phylogenomic tree based on nuclear sequences was constructed using one-to-one orthologs predicted from protein sequences for 10 species: *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Coccomyxa subellipsoidea*, *Botyrococcus braunii*, *Chlorella sorokiniana* UTEX 1230, *Chlorella variabilis* NC64A, *Auxenochlorella protothecoides* UTEX 25, *Nannochloris* sp. desiccata UTEX 2526, *Picochlorum renovo*, and *Picochlorum* sp. soloecismus DOE101. We used Orthologous Matrix (OMA; Train et al. 2017, Altenhoff et al. 2018) to predict one-to-one orthologs from proteins. OMA predicted 227 one-to-one orthologs which were independently aligned using Muscle version 3.8.1 (Edgar 2004) and concatenated, resulting in 109,223 amino acid sites in the final alignment. As above, trees were independently constructed using RAxML version 8.2.10 (Stamatakis 2014) and MrBayes version 3.2.6 (Ronquist et al. 2012). The best-scoring protein substitution model was automatically estimated by RAxML to be LG (Le and Gascuel 2008). An automatically selected 50 ML replicate trees were used to estimate bootstrap support.

RESULTS

18S rRNA gene amplicon sequencing

UTEX 2526 is a marine alga that displays promising characteristics for biotechnology applications (Zalogin and Pick 2014, Avidan and Pick 2015, Borovsky et al. 2016). This strain was originally identified as a *Chlorella* sp. (Brown et al. 1985), but it was later changed to *Mychonastes desiccatus* and deposited into the UTEX culture collection as *Chlorella desiccata* (Margulis et al. 1988). This reclassification was based on morphology; to our knowledge, molecular techniques have not yet been used to classify this alga. In order to confirm the strain identity, we performed nucleotide sequencing of a PCR amplicon of the 18S rRNA gene sequence. Surprisingly, we found this 18S rRNA gene amplicon sequence to be much more similar to *Nannochloris bacillaris* (AB080300.1, 99.24% identity) than either *Chlorella* (*Chlorella variabilis* EdL_Cl2_3NB; (KF887351.1; 95.45% identity) or *Mychonastes* (*Mychonastes timauensis* CCAP 205/2; GQ477055; 87.41% identity). The only *Chlorella* partial sequence with >98% sequence identity was from *Chlorella* sp. Yanaqocha RA1, which has been shown to clade with *Nannochloris bacillaris* (Henley et al. 2004, Somogyi et al. 2019), rather than other *Chlorella*. These data suggested that this organism is neither *Mychonastes* nor *Chlorella* and may be more related to *Nannochloris* (Table S2 in the Supporting Information). Therefore, we sought to better characterize and understand the phylogenetic placement of UTEX 2526.

Microscopy

Algae in the *Nannochloris* genus have been proposed to be some of the smallest euaryotic organisms, with simple cellular structures and small genomes (Arai et al. 1998, Yamamoto et al. 2001). One of the defining characteristics of the *Nannochloris* type species, *N. bacillaris,* is the mode of cell division being binary fission (Naumann 1921, Ogawa et al. 1995, Yamamoto et al. 2007). In fact, a number of algae initially classified as belonging to *Nannochloris*, but found mainly to divide by autosporulation, have since been reclassified as the genus *Picochlorum* (Henley et al. 2004, Yamamoto et al. 2007). To determine the mode of cell division for UTEX 2526, we examined the cells by microscopy (Fig. 1). Brightfield (Fig. 1, panel i) and fluorescence images (Fig. 1, panels ii and iii) of DNA-stained cells (Hoescht 33342 fluorescent stain) revealed that UTEX 2526 cells grown in continuous light show a variety of stages of growth and division (Fig. 1, A and B). Cells were observed to be producing 2, 4, or larger numbers of daughter cells, with nuclei visible (Fig. 1). The mode of division is autosporulation, with a parental cell wall clearly visible surrounding the daughter cells (Fig. 1A). These results are consistent with those described by Margulis et al. (Margulis et al. 1988) and indicate that, while UTEX 2526 is most-related to *Nannochloris* by 18S rRNA gene amplicon analysis, it has a cell division mode more similar to *Picochlorum*.

Genome size estimation

Although cell morphology by microscopy can be similar in these picoplankton, the genome sizes of *Nannochloris* and *Picochlorum* are known to be different. The genomes of several

Picochlorum species have been sequenced, with genome sizes generally ranging from 13 to 15 Mbp (Foflonker et al. 2018, Gonzalez-Esquer et al. 2018, Dahlin et al. 2019, Becker et al. 2020). Although no high quality genome sequences are availale for *Nannochloris*, Yamamoto et al. (Yamamoto et al. 2001) and Arai et al. (Arai et al. 1998) estimated the genome size of *N. bacillaris* to be 20.3 Mbp.

To estimate the genome size of UTEX 2526 as a function of DNA content per cell, we stained the genomic DNA with a DNA-specific fluorescent stain (Syto 9) and analyzed the sample by flow cytometry. *Picochlorum soloecismus* (Gonzalez-Esquer et al. 2018), *Picochlorum celeri* (Becker et al. 2020), and *Chlorella sorokiniana* 1412 (Hovde et al. 2018) were selected as control strains with known genome sizes. The correlation between the Syto 9 staining intensity and known nuclear DNA content (genome size for the haploid species, and double the genome size for diploid *P. celeri*) produced a linear correlation of DNA content (Fig. 2). Using this correlation, the fluorescence value of UTEX 2526 resulted in an estimated genome size of 52.98 Mbp or, if diploid like *P. celeri* (Becker et al. 2020), an estimated haploid genome size of 26.49 Mbp (Dolezel and Bartos 2005). Due to ploidy uncertainty for UTEX 2526, these results alone were inclusive for supporting UTEX 2526 as a *Nannochloris*; however, these data did show that the UTEX 2526 genome size is significantly larger than known *Picochlorum* genomes. This experiment also ultimately informed genome sequencing parameters (below).

Genome assembly, annotation, and comparison

To more conclusively identify near-relatives of UTEX 2526, and to be able to more fully leverage UTEX 2526 as a biotechnology platform in the future, full genome sequencing and annotation was conducted. PacBio sequencing resulted in 87,681 HiFi reads with an N50 read length of 9,052 bp and a total read length of 673,508,441 bp. Initial assembly with HiCanu gave a genome size of 39,022,218 bp, however it was noted that 80.1% of BUSCOs were duplicated, indicating that this species is likely diploid. Assembly by the diploid-aware software hifiasm resulted in a primary nuclear assembly of UTEX 2526 that is 21,550,525 bp in 18 contigs (Table 1, Fig. 3). The maximum contig length is 2,785,254 bp with an N50 of 1,641,916 bp, and the GC content is 45.0%. The BUSCO score is 97.5%; BUSCO is one method to determine approximate genome completeness from universal single-copy orthologs and measures a near complete assembly of the UTEX 2526 genome (Table 1). In the assembly, there are 22 telomeric motifs (5′

TTTAGGG 3′), suggesting that this species has 11 nuclear chromosomes. There are eight contigs that contain telomeric motifs at both ends, indicating eight complete chromosomes (Fig. 2). This, in addition to the high assembly N50 and low contig number, indicates a near chromosome-level genome assembly. Both the chloroplast and mitochondrial genomes were successfully extracted from the assembly as single circular contigs (Figs. S1 and S2 in the Supporting Information). The chloroplast genome is 93,001 bp with a GC content of 33.4%, and the mitochondrial genome is 40,231 bp with a GC content of 32.2%. Genome annotation resulted in 8,705 genes with an average length of 1927 bp, resulting in approximately 77.5% of the genome being annotated as genes (Table 2). The genome assembly and annotations were deposited at DDBJ/ENA/GenBank under the accession JAGTXX000000000, reads are available under SRA SRR14363043. The version described in this paper is version JAGTXX010000000 within bioproject PRJNA704951.

Based on our assessment that UTEX 2526 is diploid, the estimated haploid genome size of 26.49 Mbp by flow cytometry is similar to the 21.55 Mbp haploid genome size determined by direct sequencing. Consequently, the total DNA content per cell is predicted to be 43.1 Mbp (twice the haploid genome size), which is smaller than (but similar to) the value of 52.98 Mbp calculated for UTEX 2526 by the flow cytometry experiment (Fig. 1). These results support the genome size results estimated by flow cytometry and also show that the UTEX 2526 genome size of 21.55 Mbp is very similar to the 20.3 Mbp genome of *Nannochloris bacillaris*.

A more global examination of whole genome statistics was conducted between UTEX 2526 and example genomes from *Picochlorum*, *Chlorella*, and *Coccomxya* (Table 3). Consistent with the flow cytometry measurements, the UTEX 2526 haploid genome size of 21.55 Mbp is larger than the 13.7-15.3 Mbp seen in the examined *Picochlorum* species and is much smaller than the 46-59 Mbp genomes of *Chlorella* and *Coccomyxa.* The low GC content of *N. desiccata* at 45.0% is very similar to the *Picochlorums* with ~46% and very different from the *Chlorellas* at 64-67%. The number of annotated genes is also more similar to the *Picochlorums* than the *Chlorellas*. The mitochondrial genome is closer in size to the *Picochlorums*, but like the nuclear genome, the size of the chloroplast genome falls between *Picochlorum* and *Chlorella*. Interestingly, the mitochondrial GC content is low, more similar to *Chlorella*, and there is little variation in chloroplast GC content between species. Unfortunately, information on GC content, organellar genomes, and number of genes are unavailable for *N. bacillaris* or other species currently assigned to the *Nannochloris* genus; nonetheless, these results show the closer relatedness of UTEX 2526 to *Picochlorum* than to *Chlorella* and other species.

Phylogenetic analysis

In order to phylogenetically place UTEX 2526 within a genus, we performed several independent phylogenetic analyses, including nuclear genome, concatenated chloroplast, 18S rRNA, and actin sequences. In the phylogenomic nuclear phylogeny, including over 200 nuclear orthologs, *Nannochloris desiccata* is sister group to several members of *Picochlorum* (Fig. 4A). While the position of *Auxenochlorella protothecoides* is poorly supported, UTEX 2526 is strongly supported as being more closely related to *Picochlorum* than *Chlorella*. Analysis of chloroplast sequences (Fig. 4B and Table S1) also support UTEX 2526 as closely related to *Picochlorum*, as well as *Marvania geminata*, which has been shown to be closely related to *N. bacillaris* (Yamamoto et al. 2007).

Because there are no published *Nannochloris* chloroplast sequences within the clade of *Marvania geminata* and *Picochlorum*, despite numerous taxa in this clade (Henley et al. 2004), analyses based on nuclear 18S rRNA and actin sequences were used to provide further phylogenetic resolution. Phylogenetic analysis of the full 18S rRNA sequences (versus the amplicon sequences discussed above) within Trebouxiophyceae show clear placement of UTEX 2526 within Chlorellales but outside Chlorellaceae, where most *Chlorellas* are found (Fig. 4C and Fig. S3 in the Supporting Information). Because *Mychonastes* sits within Chlorophyceae (L. Krienitz et al. 2011), it is in a different class from Trebouxiophyceae, substantiating our conclusion that UTEX 2526 is decidedly not a *Mychonastes*. UTEX 2526 is also placed outside of the *Picochlorum* clade and, rather, is closely related to *N. bacillaris* and *Chlorella* sp. Yanaqocha RA1 (Henley et al. 2004), as well as a number of other Nannochloris strains (Figs. 4C and S3).

Actin phylogeny has been previously performed as a means to describe relationships within the *Nannochloris* genus (Arai et al. 1998, Yamamoto et al. 2001, 2003, 2007). Actin genes can be an appropriate gene sequence for phylogenetic analyses, because they evolve slowly and can be aligned unambiguously (Doolittle 1992). Also, the number and variation of actin genes is of interest with respect to cellular complexity (Yamamoto et al. 2001). Therefore, we completed a phylogenetic analysis of the actin gene sequence (Fig. 4D), as well as an amino acid sequence alignment between the *N. bacillaris* and UTEX 2526 sequences (Fig. S4 in the Supporting Information). In the actin phylogeny, *N. desiccata* is closely related to *Chlorella desiccata* UTEX 2437 (to be renamed *Nannochloris desiccata* UTEX 2437, manuscript accepted) and *N. bacillaris* and notably less related to *Chlorella* or *Picochlorum* (Fig. 4D). Examination of the protein sequence alignment showed only a single amino acid difference between UTEX 2526 and *N. bacillaris* (Fig. S4), whereas a comparison of actin protein sequences between *N. bacillaris* and species ultimately classified as *Picochlorum* (Henley et al. 2004) showed wider variation (Yamamoto et al. 2001). As expected for a diploid assembly, UTEX 2526 has two copies of the actin gene, and their nucleotide sequences are identical to each other.

These analyses, utilizing numerous loci, further demonstrate that UTEX 2526 is neither a *Chlorella* nor a *Picochlorum*, but instead consistently show that UTEX 2526 has a close relationship to *Nannochloris*, including *N. bacillaris*, the closest relative with both 18S rRNA and actin sequences available (Arai et al. 1998, Yamamoto et al. 2003).

DISCUSSION

The species described in this paper, UTEX 2526, was originally classified as a *Chlorella* (Trebouxiophyceae); then, based on a subsequent in-depth phenotypic study, it was reclassified to the genus *Mychonastes* (Chlorophyceae). Due to its potential for biotechnological development, we sought to better understand the relationship between UTEX 2526 and other identified algal species. We began with a BLAST analysis of an 18S rRNA amplicon sequence, but the results indicated that UTEX 2526 was not a *Chlorella* as designated (Table S2). We followed this with microscopy and flow cytometry-based genome size estimation (Figs. 1 and 2). These data together pointed away from *Chlorella* and *Mychonastes* and indicated closer relationship to *Nannochloris* and *Picochlorum*. Rather than amplify additional single gene(s) for sequencing, we next opted to sequence and assemble the full genome. We anticipated that not only would a full genome provide further insight into the placement of UTEX 2526, but it would also provide a broader reference for this region of the phylogenetic tree and an important tool for evaluating future transcriptomic datasets and developing new genetic engineering approaches.

After genome sequencing and assembly, we performed whole genome analyses, which showed that UTEX 2526 is distinct from, but more characteristic of, *Picochlorum* than *Chlorella*, and further, that UTEX 2526 has a haploid genome size very similar to *Nannochloris bacillaris*. We then used sequence data to conduct multiple phylogenetic analyses, which consistently placed UTEX 2526 within Trebouxiophyceae, but outside of the *Chlorella* genus.

To further place UTEX 2526, a phylogenetic estimation using 18S rRNA sequences for 135 strains and species in the class Trebouxiophyceae showed placement in Chlorelles, with clear clustering outside of *Picochlorum* and within a grouping of *Nannochloris* (Figs. S3, 4C). Analysis of the actin phylogeny demonstrated a close relationship of UTEX 2526 with a previously deposited sequence from UTEX 2437 (the parental culture of UTEX 2526), but also with *N. bacillaris* (Fig. 4D). Lastly, phylogeny of 28 chloroplast proteins placed UTEX 2526 near *Marvania* and *Picochlorum* (Fig. 4B). This placement is consistent with the 18S rRNA and actin placements, given that no other *Nannochloris* chloroplast sequences are available.

Both 18S rRNA and actin analyses confirm the placement of UTEX 2526 within *Nannochloris*, which, despite morphological similarities, diverged from Chlorophyceae approximately 735 million years ago (Parfrey et al. 2011). Although clearly not a marine *Chlorella*, which formed our initial attraction to this strain, the relatively close relatedness of UTEX 2526, and closely-related *Nannochloris* strains and species, to the *Picochlorum* genus in all four phylogenies is of interest, as several species in that genus, including *P. soloecismus*, *P. celeri*, and *P. renovo*, are top performers for outdoor growth and biomass production for renewable fuels (Dahlin et al. 2019, Gonzalez-Esquer et al. 2019, Krishnan et al. 2021).

Based on the above work, the most appropriate phylogenetic placement of UTEX 2526 is within the genus *Nannochloris,* supporting the reclassification of this oleaginous marine species as *Nannochloris* sp. 'desiccata' UTEX 2526. This classification also leaves open questions about the different morphological features between current *Nannochloris* species. By 18S rRNA full sequence analysis (our most rich dataset), there are five species most closely related to *N. bacillaris*. The nearest species is *Nannochloris* sp. LH10HG6095, with *Chlorella* sp. Yanaqocha RA1 grouping nearby, while the three others clade slightly separately: *Nannochloris* sp. AICB 424, *Nannochloris* sp. TT-3-1-F, , and *N. desiccata* UTEX 2526 (Figs. 4C, S3). The original species of *N. bacillaris* isolated by Naumann and the strain used for *N. bacillaris* sequencing were isolated from freshwater sources, a river (Naumann 1921) and lake (Ogawa et al. 1995) respectively, and are considered to be true *N. bacillaris* species per Naumann's description. *Nannochloris* sp. LH10HG6095 clades with *N. bacillaris*, and because it was isolated from a grasslands, is presumably a freshwater strain. *Chlorella* sp. Yanaqocha RA1 also appears to have been isolated from a freshwater source (Aroz et al. 1998). Interestingly, *N. desiccata* UTEX 2526 and *Nannochloris* sp. TT-3-1-F were both isolated from saline environments, indicating that their evolutionary split from *N. bacillaris* may be related to habitat (we could not find the isolation location of *Nannochloris* sp. AICB 424).

Cell division modes, however, do not seem to follow this split in phylogeny. As noted earlier, *Nannochloris bacillaris* divides by binary fission. However, in the limited published data *Nannochloris* sp. LH10HG6095 (Hodač 2015) shows likely tetrads, indicating that it may divide by autosporulation. In similar confusion, *N. desiccata* UTEX 2526 divides by autosporulation, while *Nannochloris* sp. TT-3-1-F (Sommer et al. 2020) is more difficult to discern but may divide by binary fission, similar to descriptions by Ogawa (Ogawa et al. 1995) and Henley et al (Henley et al. 2004). Therefore, in spite of the reclassification of many autosporulating species from *Nannochloris* to *Picochlorum*, cell division modes may still be muddled across this part of the phylogenetic tree.

Although *Nannochloris bacillaris* has regularly been observed to divide by binary fission, some studies in other algae species have noted different modes of cell division occurring within a culture or between culture conditions (Bišová and Zachleder 2014, Yamagishi et al. 2017), including alternating between different types of fission. For example, cells generally observed to only divide by binary fission have been shown to be able to divide by multiple fission under more optimal growth conditions (Bišová and Zachleder 2014, Bensalem et al. 2018). These findings suggest that culture conditions may affect observed growth and morphology, such that direct comparisons within and between species cannot be made when conditions vary across experiments. Thus, in future work, it will be important to examine these microalgae under a range of conditions to evaluate if they can adopt varying modes of cell division. If it continues to hold true that *N. bacillaris* only divides by binary fission, and this remains a defining feature of *Nannochloris* (Krienitz et al. 1996), then it may be prudent to create a new genus name for the strains that divide by other cell division modes. Until then, the current species of the *Nannochloris* genus are the nearest relatives to UTEX 2526, justifying its reclassification to *Nannochloris*.

Decreasing costs and improving quality of *de novo* genome sequencing and assembly is opening new avenues of phylogenetic characterization and taxonomic classification that may more accurately reflect the relatedness of species. At the global genome level, looking at ploidy, genome size, and GC content gives indicators of relatedness, in this case showing that *Nannochloris* sp. 'desiccata' UTEX 2526 is closer to *Nannochloris* and *Picochlorum* than *Chlorella*. The placement of UTEX 2526 as *Nannochloris desiccata* is a clear demonstration of the need for increased classification based on in-depth sequence analysis. Genome sequencing allowed us to use several different full gene sequences to support our placement of *N. desiccata*, which was necessary in this phylogenetic region where significant reclassification has previously occurred. Genome sequencing not only supports more accurate classification, but also provides a valuable resource to both basic and applied biology studies. For example, high-quality genome assembly supports transcriptome analyses, providing reference genes in order to measure cellular responses to a given environment. Another critical component of genome assembly is the ability to identify gene targets and loci for applications in genetic engineering.

In summary, *Nannochloris* is a genus that has been studied for over a century, with the classification of species placed in this genus being fluid as more sequence data has become available. In spite of the long history of this genus, here we provide the first high-quality genome assembly of a *Nannochloris* strain. The *N. desiccata* UTEX 2526 genome was assembled into only 18 contigs, of which eight appear to be complete chromosomes, with telomeres at both ends. This high-quality genome will support further *Nannochloris* classification studies, which is necessary future work, as well as the investigation, including gene expression analysis, into the algal production potential of *N. desiccata*, specifically, and of *Nannochloris*, generally.

ACKNOWLEDGMENTS

We would like to thank Cheryl D. Gleasner for sequencing support.

This work was funded by the U.S. Department of Energy Bioenergy Technologies Office Annual Operating Plan project NL0025841. The authors declare no competing interests. LANL Release # LA-UR-21-25757.

REFERENCES

Altenhoff, A.M., Glover, N.M., Train, C.-M., Kaleb, K., Vesztrocy, A.W., Dylus, D., de Farias, T.M., Zile, K., Stevenson, C., Long, J., Redestig, H., Gonnet, G.H. & Dessimoz, C. 2018.

The OMA orthology database in 2018: Retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* 46:D477–85.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–10.

Andrade, C.J. de & Andrade, L.M. de 2017. An overview on the application of genus *Chlorella* in biotechnological processes. *J. Adv. Res. Biotechnol.* 2:1–9.

Arai, S., Takahashi, H., Takano, H., Sakai, A. & Kawano, S. 1998. Isolation, characterization, and chromosome mapping of an actin gene from the primitive green alga, *Nannochloris bacillaris* (Chlorophyceae). *J. Phycol.* 34:477–85.

Aroz, R., Hader, D. P. & Huss, V.A.R. 1998. DNA sequence and secondary structure of the small subunit rRNA (Accession No. Y14950) from the green alga *Chlorella* spec. Yanaqocha RA1 (Tre- bouxiophyceae, Chlorophyta). *Plant Physiol.* 116:1603–5.

Avidan, O. & Pick, U. 2015. Acetyl-CoA synthetase is activated as part of the PDH-bypass in the oleaginous green alga *Chlorella desiccata*. *J. Exp. Bot.* 66:7287–98.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. Da, Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S. & Zhang, J. 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49:D480–9.

Becker, S.A., Spreafico, R., Kit, J.L., Brown, R., Likhogrud, M., Fang, W., Posewitz, M.C., Weissman, J.C. & Radakovits, R. 2020. Phased diploid genome sequence for the fast-growing microalga *Picochlorum celeri*. *Microbiol Resour Announc*. 9:e00087-20.

Bensalem, S., Lopes, F., Bodénès, P., Pareau, D., Français, O. & Le Pioufl, B. 2018. Structural changes of *Chlamydomonas reinhardtii* cells during lipid enrichment and after solvent exposure. *Data Br.* 17:1283–7.

Bišová, K. & Zachleder, V. 2014. Cell-cycle regulation in green algae dividing by multiple fission. *J. Exp. Bot.* 65:2585–602.

Bito, T., Okumura, E., Fujishima, M. & Watanabe, F. 2020. Potential of *Chlorella* as a dietary supplement to promote human health. *Nutrients* 12:2524.

Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I. V., Claverie, J. M. & Van Etten, J.L. 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* 22:2943–55.

Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A. & Finn, R.D. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49:D344–54.

Bock, C., Krienitz, L. & Pröschold, T. 2011. Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea* 11:293–312.

Borovsky, D., Sterner, A. & Powell, C.A. 2016. Cloning and expressing trypsin modulating oostatic factor in *Chlorella desiccata* to control mosquito larvae. *Arch. Insect Biochem.*

Physiol. 91:17–36.

Brown, S., Margulis, L., Ibarra, S. & Siqueiros, D. 1985. Desiccation resistance and contamination as mechanisms of gaia. *BioSystems* 17:337–60.

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M. & Borodovsky, M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinform.* 3:1–11.

Butcher, R.W. 1952. Contributions to our knowledge of the smaller marine algae. *J. Mar. Biol. Assoc. U.K.* 31:175–91.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18:170–5.

Cho, S.H., Ji, S.C., Hur, S.B., Bae, J., Park, I. S. & Song, Y. C. 2007. Optimum temperature and salinity conditions for growth of green algae *Chlorella ellipsoidea* and *Nannochloris oculata*. *Fish. Sci.* 73:1050–6.

Cross, F.R. & Umen, J.G. 2015. The *Chlamydomonas* cell cycle. *Plant J.* 82:370–92.

Dahlin, L.R., Gerritsen, A.T., Henard, C.A., Kunde, Y., Hovde, B.T., Starkenburg, S.R., Wychen, S. Van, Linger, J.G., Posewitz, M.C. & Guarnieri, M.T. 2019. Development of a high-productivity, halophilic, thermotolerant microalga Picochlorum renovo. *Commun. Biol.* 2:1–9.

Darienko, T., Rad-Menéndez, C., Campbell, C. & Pröschold, T. 2019. Are there any true marine *Chlorella* species? Molecular phylogenetic assessment and ecology of marine *Chlorella*-like organisms, including a description of *Droopiella* gen. nov. *Syst. Biodivers.* 17:811–29.

De Clerck, O., Guiry, M.D., Leliaert, F., Samyn, Y. & Verbruggen, H. 2013. Algal Taxonomy: A Road to Nowhere? *J. Phycol.* 49:215–25.

Dolezel, J. & Bartos, J. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Ann. Bot.* 95:99–110.

Doolittle, R.F. 1992. Reconstructing history with amino acid sequences. *Protein Sci.* 1:191–200.

Edgar, R.C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–7.

Foflonker, F., Mollegard, D., Ong, M., Yoon, H.S. & Bhattacharya, D. 2018. Genomic analysis of *Picochlorum* species reveals how microalgae may adapt to variable environments. *Mol. Biol. Evol.* 35:2702–11.

Fučíková, K., Pažoutová, M. & Rindi, F. 2015. Meiotic genes and sexual reproduction in the green algal class Trebouxiophyceae (Chlorophyta). *J. Phycol.* 51:419–30.

Gonzalez-Esquer, C.R., Twary, S.N., Hovde, B.T. & Starkenburg, S.R. 2018. Nuclear, chloroplast, and mitochondrial genome sequences of the prospective microalgal biofuel strain *Picochlorum soloecismus*. *Genome Announc.* 6:1–2.

Gonzalez-Esquer, C.R., Wright, K.T., Sudasinghe, N., Carr, C.K., Sanders, C.K., Turmo, A., Kerfeld, C.A., Twary, S.N. & Dale, T. 2019. Demonstration of the potential of *Picochlorum soloecismus* as a microalgal platform for the production of renewable fuels. *Algal Res.* 43:101658.

Gremme, G. 2013. Computational Gene Structure Prediction. Ph.D. dissertation, University of Hamburg, Hamburg, Germany, 289 pp.

Hanschen, E.R. & Starkenburg, S.R. 2020. The state of algal genome quality and diversity. *Algal Res.* 50:101968.

Henley, W.J., Hironaka, J.L., Guillou, L., Buchheim, M.A., Buchheim, J.A., Fawley, M.W. & Fawley, K.P. 2004. Phylogenetic analysis of the "*Nannochloris*-like" algae and diagnoses of *Picochlorum oklahomensis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). *Phycologia* 43:641–52.

Herrera-Valencia, V.A., Contreras-Pool, P.Y., López-Adrián, S.J., Peraza-Echeverría, S. & Barahona-Pérez, L.F. 2011. The green microalga *Chlorella saccharophila* as a suitable source of oil for biodiesel production. *Curr. Microbiol.* 63:151–7.

Hodač, L. 2015. Green algae in soil: assessing their biodiversity and biogeography with molecular-phylogenetic methods based on cultures. Ph.D. dissertation, Georg August University of Göttingen School of Science, 185 pp.

Hovde, B.T., Deodato, C.R., Andersen, R.A., Starkenburg, S.R., Barlow, S.B. & Ann, R. 2019. *Chrysochromulina*: Genomic assessment and taxonomic diagnosis of the type species for an oleaginous algal clade. *Algal Res.* 37:307–19.

Hovde, B.T., Hanschen, E.R., Steadman Tyler, C.R., Lo, C., Kunde, Y., Davenport, K., Daligault, H., Msanne, J., Canny, S., Eyun, S., Riethoven, J.J.M., Polle, J. & Starkenburg, S.R. 2018. Genomic characterization reveals significant divergence within *Chlorella sorokiniana* (Chlorellales, Trebouxiophyceae). *Algal Res.* 35:449–61.

Huesemann, M., Chavis, A., Edmundson, S., Rye, D., Hobbs, S., Sun, N. & Wigmosta, M. 2018.

Climate-simulated raceway pond culturing: quantifying the maximum achievable annual biomass productivity of *Chlorella sorokiniana* in the contiguous USA. *J. Appl. Phycol.* 30:287–98.

Huesemann, M., Dale, T., Chavis, A., Crowe, B., Twary, S., Barry, A., Valentine, D., Yoshida, R., Wigmosta, M. & Cullinan, V. 2017. Simulation of outdoor pond cultures using indoor LED-lighted and temperature-controlled raceway ponds and phenometrics photobioreactors. *Algal Res.* 21:178–90.

Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P.A. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37:540–6.

Krienitz, L. & Bock, C. 2012. Present state of the systematics of planktonic coccoid green algae of inland waters. *Hydrobiologia* 698:295–326.

Krienitz, L, Bock, C., Dadheech, P.K. & Pröschold, T. 2011. Taxonomic reassessment of the genus *Mychonastes* (Chlorophyceae, Chlorophyta) including the description of eight new species. *Phycologia* 50:89–106.

Krienitz, L., Huss, V.A.R. & Hummer, C. 1996. Picoplanktonic Choricystis species (Chlorococcales, Chlorophyta) and problems surrounding the morphologically similar "*Nannochloris*-like algae." *Phycologia* 35:332–41.

Krishnan, A., Likhogrud, M., Cano, M., Edmundson, S., Melanson, J.B., Huesemann, M., McGowen, J., Weissman, J.C. & Posewitz, M.C. 2021. *Picochlorum celeri* as a model system for robust outdoor algal growth in seawater. *Sci. Rep.* 11:1–13.

Krivina, E.S. & Temraleeva, A.D. 2020. Identification problems and cryptic diversity of *Chlorella*-clade microalgae (Chlorophyta). *Microbiology* 89:720–32.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. & Calcott, B. 2017. PartitionFinder 2 new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34:772–3.

Le, S.Q. & Gascuel, O. 2008. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25:1307–20.

Li, P.E., Lo, C.C., Anderson, J.J., Davenport, K.W., Bishop-Lilly, K.A., Xu, Y., Ahmed, S., Feng, S., Mokashi, V.P. & Chain, P.S.G. 2017. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Res.* 45:67–80.

Luo, W., Pflugmacher, S., Pröschold, T., Walz, N. & Krienitz, L. 2006. Genotype versus phenotype variability in *Chlorella* and *Micractinium* (Chlorophyta, Trebouxiophyceae). *Protist* 157:315–33.

Margulis, L., Hinkle, G., Mckhann, H., Moynihan, B. & Brown, W. 1988. *Mychonastes desiccatus* BROWN sp. nova (Chlorococcales, Chlorophyta) - an intertidal alga forming achlorophyllous desiccation-resistant cysts. *Arch. Hydrobiol. Suppl.* 78:425–46.

Miyagishima, S., Fujiwara, T., Sumiya, N., Hirooka, S., Nakano, A., Kabeya, Y. & Nakamura, M. 2014. Translation-independent circadian control of the cell cycle in a unicellular photosynthetic eukaryote. *Nat. Commun.* 5:3807.

Morales, E.A. & Trainor, F.R. 1997. Algal phenotypic plasticity: Its importance in developing new concepts the case for Scenedesmus. *Algae* 12:147–57.

Morales, E.A., Trainor, F.R. & Schlichting, C.D. 2002. Evolutionary and ecological implications of plastic responses of algae. *Constancea*. 83.

Naumann, E. 1921. Notizen zur Systematik der Süsswasseralgen. V. Uber *Nannochloris*, eine neue Chlorophyceengattung. *Ark. FÖR Bot.* 16:1–19.

Negi, S., Barry, A.N., Friedland, N., Sudasinghe, N., Subramanian, S., Pieris, S., Holguin, F.O., Dungan, B., Schaub, T. & Sayre, R. 2016. Impact of nitrogen limitation on biomass, photosynthesis, and lipid accumulation in *Chlorella sorokiniana*. *J. Appl. Phycol.* 28:803–12.

Neofotis, P., Huang, A., Sury, K., Chang, W., Joseph, F., Gabr, A., Twary, S., Qiu, W., Holguin, O. & Polle, J.E.W. 2016. Characterization and classification of highly productive microalgae strains discovered for biofuel and bioproduct generation. *Algal Res.* 15:164–78.

Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M. & Koren, S. 2020. HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* 30:1291–305.

Ogawa, S., Ueda, K. & Noguchi, T. 1995. Division apparatus of the chloroplast in *Nannochloris bacillaris* (Chlorophyta). *J. Phycol.* 31:132–7.

Pacific Biosciences of California, I. 2014. https://www.pacb.com/wp-content/uploads/2015/09/Shared-Protocol-Guidelines-for-Using-a-Salt-Chloroform-Wash-to-Clean-Up-gDNA.pdf.

Pacific Biosciences of California, I. 2020. https://www.pacb.com/wp-content/uploads/Procedure-

Checklist-Preparing-HiFi-Libraries-from-Low-DNA-Input-Using-SMRTbell-Express-Template-Prep-Kit-2.0.pdf.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. & Katz, L.A. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl. Acad. Sci. USA* 108:13624–9.

Posten, C. & Chen, S.F. 2016. *Microalgae Biotechnology*. Springer International Publishing. 1–35 pp.

Rai, U., Deshar, G., Rai, B., Bhattarai, K., Dhakal, R. & Rai, S. 2014. Isolation and culture condition optimization of *Chlorella vulgaris*. *Nepal J. Sci. Technol.* 14:43–8.

Rodríguez, F., Feist, S.W., Guillou, L., Harkestad, L.S., Bateman, K., Renault, T. & Mortensen, S. 2008. Phylogenetic and morphological characterisation of the green algae infesting blue mussel *Mytilus edulis* in the North and South Atlantic oceans. *Dis. Aquat. Organ.* 81:231–40.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–42.

Sassenhagen, I., Wilken, S., Godhe, A. & Rengefors, K. 2015. Phenotypic plasticity and differentiation in an invasive freshwater microalga. *Harmful Algae* 41:38–45.

Schaum, E., Rost, B., Millar, A.J. & Collins, S. 2013. Variation in plastic responses of a globally distributed picoplankton species to ocean acidification. *Nat. Clim. Chang.* 3:298–302.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671–5.

Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–9.

Seppey, M., Manni, M. & Zdobnov, E.M. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. *In* Kollmar, M. [Ed.] *Methods in Molecular Biology: Gene Prediction*. Humana Press Inc., pp. 227–45.

Singh, A., Nigam, P.S. & Murphy, J.D. 2011. Renewable fuels from algae: An answer to debatable land based fuels. *Bioresour. Technol.* 102:10–6.

Sommer, V., Mikhailyuk, T., Glaser, K. & Karsten, U. 2020. Uncovering unique green algae and cyanobacteria isolated from biocrusts in highly saline potash tailing pile habitats, using an integrative approach. *Microorganisms* 8:1–37.

Somogyi, B., Felföldi, T., Solymosi, K., Makk, J., Homonnay, Z.G., Horváth, G., Turcsi, E., Böddi, B., Márialigeti, K. & Vörös, L. 2019. *Chloroparva pannonica* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta)-a new picoplanktonic green alga from a turbid, shallow soda pan. *Phycologia* 50:1–10.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–3.

Stanke, M. & Waack, S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215–25.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R. & Greiner, S. 2017. GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45:W6–11.

Train, C. M., Glover, N.M., Gonnet, G.H., Altenhoff, A.M. & Dessimoz, C. 2017. Orthologous Matrix (OMA) algorithm 2.0: More robust to asymmetric evolutionary rates and more scalable hierarchical orthologous group inference. *Bioinformatics* 33:i75–82.

UTEX, C.C. 2009. http://web.biosci.utexas.edu/utex/Media%20PDF/bg-11-medium.pdf.

Venteris, E.R., Skaggs, R.L., Coleman, A.M. & Wigmosta, M.S. 2013. A GIS cost model to assess the availability of freshwater, seawater, and saline groundwater for algal biofuel production in the United States. *Environ. Sci. Technol.* 47:4840–9.

Vítová, M. & Zachleder, V. 2005. Points of commitment to reproductive events as a tool for analysis of the cell cycle in synchronous cultures of algae. *Folia Microbiol. (Praha).* 50:141–9.

Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V. & Zdobnov, E.M. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35:543–8.

Yamagishi, T., Yamaguchi, H., Suzuki, S., Horie, Y. & Tatarazako, N. 2017. Cell reproductive patterns in the green alga *Pseudokirchneriella subcapitata* (=*Selenastrum capricornutum*) and their variations under exposure to the typical toxicants potassium dichromate and 3, 5-DCP. *PLoS ONE* 12:e0171259.

Yamamoto, M., Nishikawa, T., Kajitani, H. & Kawano, S. 2007. Patterns of asexual reproduction in *Nannochloris bacillaris* and *Marvania geminata* (Chlorophyta, Trebouxiophyceae). *Planta*. 226:917–27.

Yamamoto, M., Nozaki, H. & Kawano, S. 2001. Evolutionary relationships among multiple modes of cell division in the genus *Nannochloris* (Chlorophyta) revealed by genome size, actin gene multiplicity, and phylogeny. *J. Phycol.* 37:106–20.

Yamamoto, M., Nozaki, H., Miyazawa, Y., Koide, T. & Kawano, S. 2003. Relationship between presence of a mother cell wall and speciation in the unicellular microalga *Nannochloris* (Chlorophyta). *J. Phycol.* 39:172–84.

Zalogin, T.R. & Pick, U. 2014. Azide improves triglyceride yield in microalgae. *Algal Res.* 3:8–16.

	HiCanu assembly	Hifiasm assembly
Number of contigs	244	18
Total contig size	39,022,218 bp	21,550,525 bp
Contig N50	497,999 bp	1,641,916 bp
GC %	44.8%	45.0%
Contig N90	70,409 bp	653,437 bp
Max contig length	1,791,715 bp	2,785,254 bp
Complete BUSCOs	97.9%	97.5%
Complete and Single-Copy BUSCOs	17.8%	93.6%
Complete and Duplicated	80.1%	3.9%

Table 1: Continuity statistics and BUSCO scores of the nuclear genome assembly generated by HiCanu and hifiasm. The HiCanu assembly had a large amount of duplicated BUSCOs, suggesting that *Nannochloris desiccata* UTEX 2526 is diploid. As a result, we utilized the diploid-aware assembler, hifiasm, to generate a haploid reference.

BUSCOs		
Fragmented BUSCOs	0.9%	0.9%
Missing BUSCOs	1.2%	1.6%
Total BUSCO groups	1519	1519

Table 2: *Nannochloris desiccata* UTEX 2526 genome annotation statistics.

Number of genes	8705
Average gene length	1927
Average CDS length	1695

Average exon length	673
Average exons per gene	2.67
Average intron length	173
% of genome assembly as genes	77.5
% of genome as CDS	72.3

Table 3: Global genome statistics of *Nannochloris desiccata* UTEX 2526 compared to example *Picochlorum*, *Chlorella*, and *Coccomyxa* species. *N. bacillaris* has a genome size of 20.3 Mbp by chromosome number and size estimation (Arai et al. 1998); GC content and other genome information is unknown.

	***N. desiccata* UTEX 2526**	***P. soloecismus***	***P. celeri***	***P. renovo***	***C. sorokiniana* 1412**	***C. variabilis* NC64A**	***Coccomyxa* sp LA000219**
Genome size (haploid)	21.55Mbp	15.25Mbp	13.72Mbp	14.39Mbp	57.88Mbp	46.16Mbp	48.55Mbp
GC content	45.0%	46.0%	46.2%	46.2%	64.1%	67.0%	54.70%
Number of contigs in assembly	18 contigs	56 contigs	15 scaffolds	29 contigs	65 contigs	30 scaffolds	291 contigs/106 scaffolds
Ploidy	diploid	haploid	diploid	--	haploid	--	--
Number of genes	8705	7844	--	8902	12611	9791	--
GenBank accession	JAGTXX000000000.1	PJAJ00000000.1	JAACMV000000000.1	WJEA00000000.1	PKFD00000000.1	ADIC00000000.1	JUGA00000000.1
mitochondrial genome	40.2Kbp	38.7Kbp	--	36.3Kbs	61.8Kbs	78.5Kbs	--
Mitochondrial GC content	32.2%	40.9%	--	42.4%	30.1%	28.2%	--
Mitochondrial accession	OK569792	PJAJ01000013.1	--	Unpublished*	PKFD01000021.1	KP271968	JUGA01000053.1
Chloroplast genome	93.0Kbp	72.7Kbp	--	74.4Kbs	131.1Kbs	124.8Kbs	--
Chloroplast GC content	33.4%	32.0%	--	32.2%	33.9%	33.9%	--

Chloroplast accession	OK569791	PJAJ01000016.1	--	Unpublished*	PKFD01000065.1	KP271969	KN714624.1

* Sequencing and assembly completed at Los Alamos National Laboratory by Dalhin L.R., Hovde B.T., Starkenburg S.R., Posewitz C.M. and Guarnieri M.T.

Figure 1. *Nannochloris desiccata* UTEX 2526 cells in an unsynchronized culture. Two different fields of view (A and B) show various stages of cell division with different numbers of daughter cells being produced. Within each field of view, the cells were visualized (counterclockwise from top left) by (i) Brightfield, (ii) Hoechst 33342 fluorescence, (iii) Chlorophyll autofluorence and (iv) Overlay. The orange arrow in A1 indicates the presence of a parental cell wall.

Figure 2. Fluorescence values for cells stained with Syto 9 nucleic acid stain plotted against known nuclear DNA content. The cellular DNA content of *Nannochloris desiccata* UTEX 2526 was measured and genome size calculated (star) from the linear fit of the known genome size values of *Chlorella sorokiniana* 1412 (square), *Picochlorum celeri* (triangle), and *Picochlorum soloecismus* (circle). The size of the diploid genome size based on sequencing analysis is also shown for comparison (diamond).

Figure 3. Visual representation of the contigs in the nuclear genome assembly of *N. desiccata* UTEX 2526. The presence of a telomeric motif, 5′ TTTAGGG 3′, is indicated by shaded contig ends.

Figure 4. Phylogenetic placement of UTEX 2526 as *Nannochloris desiccata* (bold). (A) Phylogeny based on 227 nuclear proteins. Genera *Picochlorum* and *Nannochloris* are highlighted. (B) Phylogenetic analysis based on 26 concatenated chloroplast proteins. (C) A subset of the phylogenetic analysis based on 18S ribosomal RNA (also see Fig. S3 for complete analysis). (D) Phylogenetic analysis based on actin genes. Genera *Picochlorum* and *Nannochloris* are highlighted, as defined by (Henley et al. 2004), demonstrating the close

relationship of *Nannochloris desiccata* UTEX 2526 to *N. bacillaris* and the highly productive *Picochlorum* genus. All panels use both maximum likelihood and Bayesian approaches, with numbers at internal nodes indicating maximum likelihood bootstrap values and Bayesian posterior probabilities when above 70 and 0.95 respectively. Tree topology is from maximum likelihood analyses. Scale bars representing substitutions per site are included. All panels are rooted using *Chlamydomonas reinhardtii* and *Dunaliella salina* (Chlorophyceae) as outgroups.

Table S1. The best-fitting partition and protein substitution scheme for the chloroplast phylogeny as determined by PartitionFinder version 2.1.1.

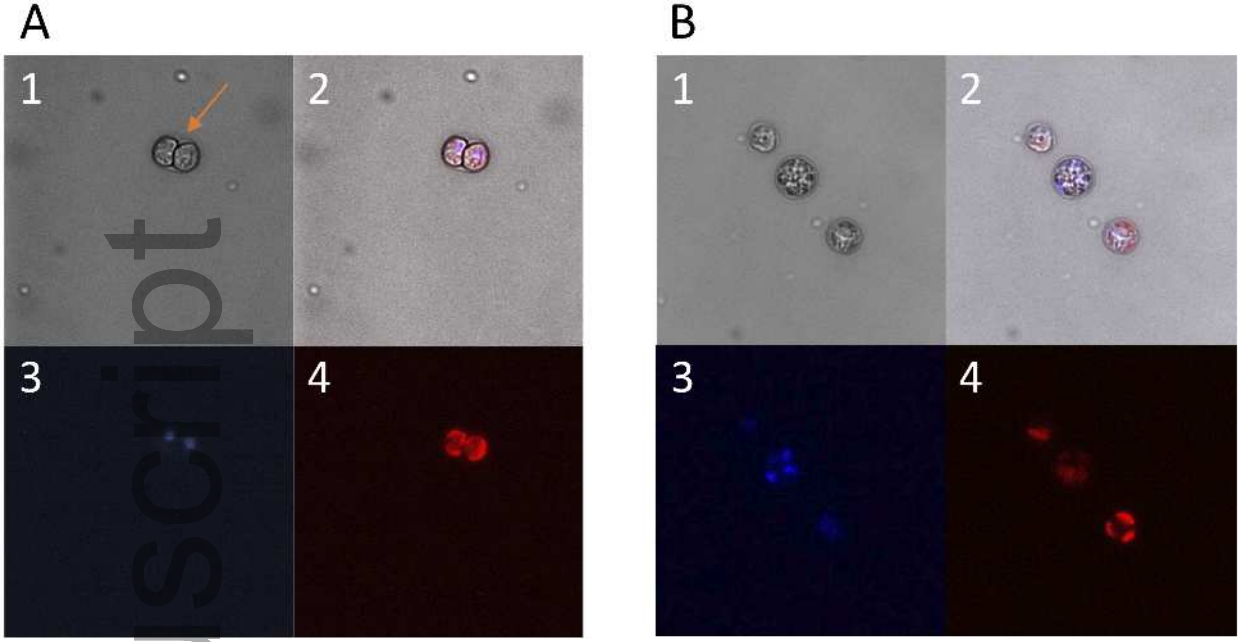
Table S2: Summary of 18S rRNA partial sequence sequences alignments, relative to a *Nannochloris desiccata* UTEX 2526 18S rRNA partial sequence. A BLASTn analysis of this partial 18S rDNA sequence, *Mychonastes* was not found in the top hits, and *Chlorella* was only found once. Instead, there were a range of organisms whose 18S rDNA aligned with our amplicon. First, the best alignment corresponded to *Coccomyxa parasitica*; however, the investigators of this sequence have indicated that this sequence was probably a contaminant and not actually *C. parasitica* (Rodríguez et al. 2008). The third and fourth alignments belong to organisms with insufficient information to assign an identification. Interestingly, however, the second-best alignment, as well as alignment ranks 5-7 and 9-10 (all with 98-100% sequence identity) were 18S rDNA partial sequences from organisms identified as belonging to the *Nannochloris* genus. The only *Chlorella* partial sequence with >98% sequence identity was from *Chlorella* sp. Yanaqocha RA1, which has been shown to clade with *Nannochloris bacillaris* (Henley et al. 2004, Somogyi et al. 2019), rather than other *Chlorella*. These data suggested that this organism is neither *Mychonastes* nor *Chlorella* and may be more related to *Nannochloris*. For species with available information, the type of isolation location and cell division mode is also shown. F = Freshwater, B = Brackish, S = Saline, BF = Binary Fission, A = Autosporulation, U = Unknown, (T) = Tentative, N/A = Not applicable/available.

Figure S1. Assembly and annotation of the *Nannochloris desiccata* UTEX 2526 chloroplast genome. Genes inside the outer circle are transcribed clockwise, and genes outside the outer circle are transcribed counter clockwise. The inner circle shows the GC content (dark grey) throughout the chloroplast genome, and the line in the middle of the inner circle is the 50% GC content threshold.

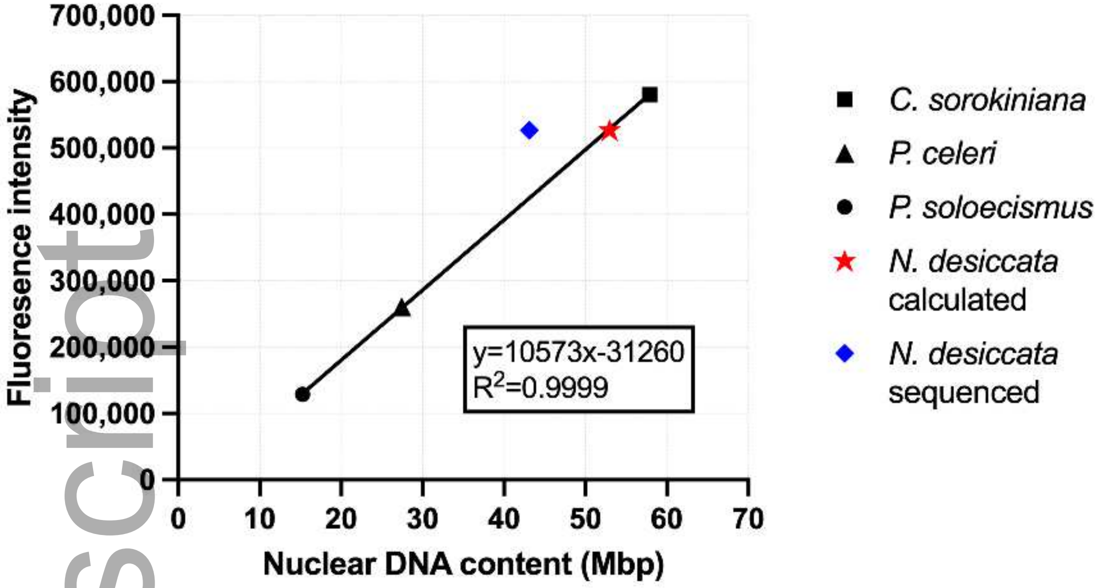
Figure S2. Assembly and annotation of the *Nannochloris desiccata* UTEX 2526 mitochondrial genome. Genes inside the outer circle are transcribed clockwise, and genes outside the outer circle are transcribed counter clockwise. The inner circle shows the GC content (dark grey) throughout the mitochondrial genome, and the line in the middle of the inner circle is the 50% GC content threshold.

Figure S3. Phylogenetic analysis of Trebouxiophyceae 18S rRNA sequences using maximum likelihood and Bayesian approaches. Numbers at internal nodes indicate maximum likelihood bootstrap values and Bayesian posterior probabilities when above 70 and 0.95 respectively. A scale bar representing substitutions per site is included.

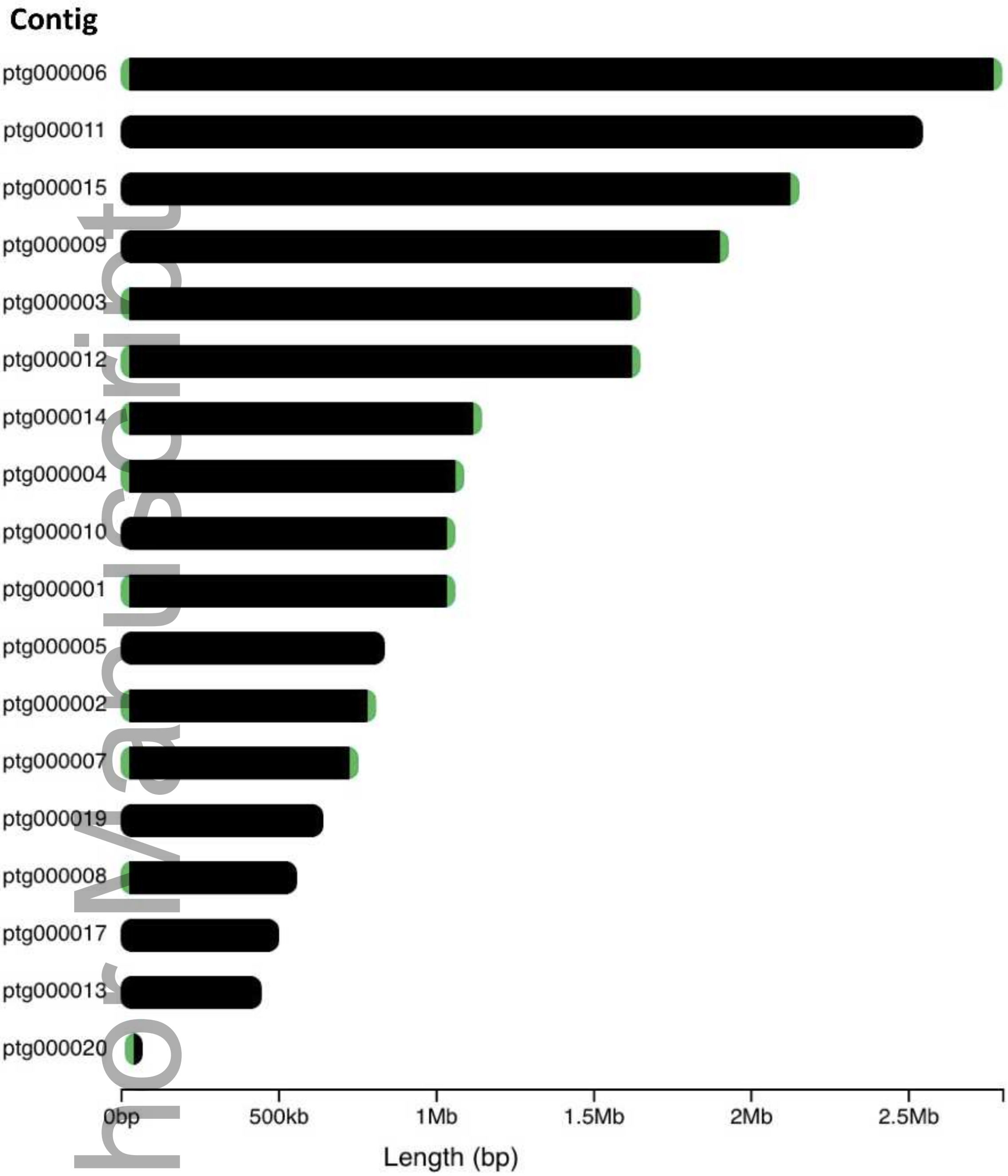
Figure S4. Alignment of the actin protein sequence between *Nannochloris desiccata* UTEX 2526 and *N. bacillaris*. The single amino acid difference (F63Y) is highlighted in yellow.



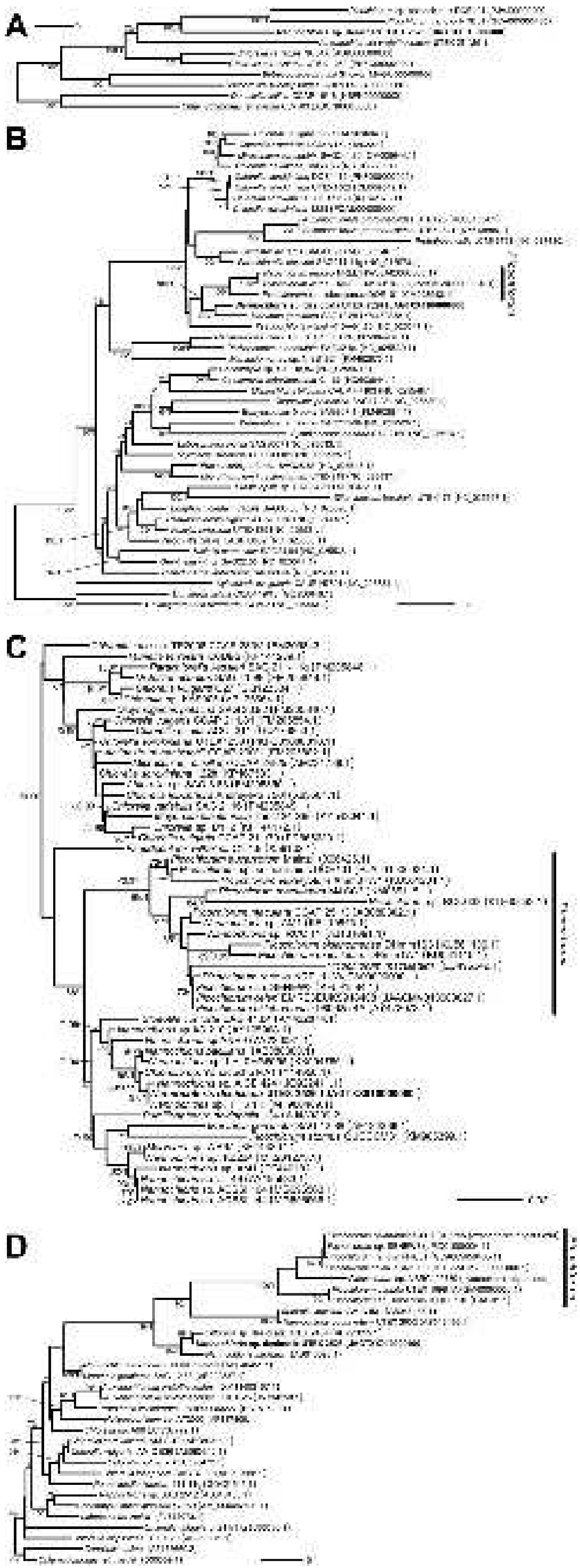
jpy_13242_f1.tiff



jpy_13242_f2.tiff



jpy_13242_f3.tiff



jpy_13242_f4.tiff