Plant Ecology and Evolution 156 (3): 311–325, 2023
https://doi.org/10.5091/plecevo.105762

RESEARCH ARTICLE

Johansenicoccus eremophilus gen. et sp. nov., a novel evolutionary lineage in Chlorophyceae with unusual genomic features

Karolina Fučíková[1], Melissa Taylor[1,2], Louise A. Lewis[3], Brian K. Niece[1], Aleeza S. Isaac[1,4], Nicole Pietrasiak[5]

1 Department of Biological and Physical Sciences, Assumption University, Worcester, MA, U.S.A.
2 School of Nursing, Massachusetts College of Pharmacy and Health Sciences, Worcester, MA, U.S.A.
3 Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, U.S.A.
4 College of Osteopathic Medicine, University of New England, Biddeford, ME, U.S.A.
5 School of Life Sciences, University of Nevada - Las Vegas, Las Vegas, NV, U.S.A.

Corresponding author: Karolina Fučíková (k.fucikova@assumption.edu)

Academic editor: Brecht Verstraete ♦ **Received** 1 May 2023 ♦ **Accepted** 14 July 2023 ♦ **Published** 6 September 2023

Abstract

Background – Green algae are a diverse group of photosynthetic eukaryotes, yet are still vastly understudied compared to land plants. For many years, green algae were characterized based on their morphology and life cycles. More recently, phylogenetic and genomic analyses have been added to the phycological toolkit for a better understanding of algal biodiversity and evolutionary history.
Material and methods – A desert strain of green algae was isolated from Joshua Tree National Park (JTNP) in southern California as part of a larger biodiversity survey. The alga's nuclear rRNA genes as well as the chloroplast genome were sequenced, annotated, and analysed in addition to a morphological assessment.
Results – Morphologically this strain is especially similar to *Pseudomuriella* and *Rotundella*, and its lipid profile resembles that of other soil algae, but phylogenomic analyses demonstrate that it is a distinct evolutionary lineage in Chlorophyceae. The alga exhibits several unusual genomic features, the most remarkable being its highly derived yet apparently functional nuclear rRNA genes, 18S and 28S. Both genes are GC-rich and bear many compensatory base changes to maintain a similar secondary structure to that of other green algae. The chloroplast genome has a distinct gene order and repeat arrangement from other published green algal plastomes, but contains the expected genes and also provides phylogenetically informative data.
Conclusion – We conclude that the strain be placed into a new species and genus in the class Chlorophyceae, and propose the name *Johansenicoccus eremophilus* for this new taxon. *Johansenicoccus eremophilus* exemplifies science's insufficient understanding of the range of genomic variations among inconspicuous soil algae.

Keywords

biological soil crust, chlorophyte, coccoid, direct repeat, inverted repeat, microalga

INTRODUCTION

Green algae span a great diversity of body plans, from simple solitary unicells to complex filaments and thalli with differentiated cell types, and they inhabit a broad range of environments in nearly every corner of Earth (Leliaert et al. 2012 and references within). Their occurrence in the soils of desert regions has been long known and numerous new species have been described from such environments (e.g. Bischoff and Bold 1963; Fučíková et al. 2014a; Saber

Plant Ecology and Evolution is published by Meise Botanic Garden and Royal Botanical Society of Belgium.

et al. 2018). This continued stream of discoveries indicates that additional, yet unknown, species likely dwell in desert soils. Many green algal taxa are morphologically similar to each other, which complicates their identification, adequate assessment of their geographic ranges, and the understanding of their ecological roles in the desert environment. Continued field collections, careful microscopic and genetic assessments of cultured material, as well as high-throughput sequencing of environmental samples can give us a better understanding of the complex soil communities that are otherwise hidden to the eye.

Studies of green algae from soils have shown remarkable adaptations. Desert soil algae have distinct mechanisms for coping with environmental stresses (Gray et al. 2007), and even temperate soil algae can remain viable after much-prolonged desiccation (Lewis and Trainor 2012). Genomic and transcriptomic studies have provided new insights into the biology of green soil-dwellers, focusing on the mechanisms protecting them against high UV radiation (Karsten and Holzinger 2014) and desiccation (Peredo and Cardon 2020). Additionally, membrane lipid composition and lipid production have been shown as important in handling temperature stress in plants (Zheng et al. 2011), and algal lipid production has been of great interest to biologists because of its potential in biofuel and other industries (Aratboni et al. 2019). However, our knowledge of the role of lipids in the physiology of desert algae is still very limited. Similarly, little is known about whether the organellar genomes of soil-dwelling, and especially desert-dwelling green algae contain adaptive features aiding their survival in hot and arid environments.

Within the green algal phylum Chlorophyta, the class Trebouxiophyceae is especially known to harbour numerous soil-dwelling species and includes many desert-dwelling lineages among them (Fučíková et al. 2014b). The class Chlorophyceae may be better known for its aquatic colony-forming representatives with charismatic morphologies such as *Volvox* Linnaeus, but for some time there has been mounting evidence of desert habitat affiliation of numerous chlorophyceans as well (Lewis and Lewis 2005; Flechtner et al. 2013; Fučíková et al. 2014a). Because desert green algae (and soil algae in general) often have simple coccoid morphologies with few distinct characteristics, confident identification of chlorophytes and trebouxiophytes from desert soils almost always requires the use of isolation, culturing, and molecular phylogenetic tools.

The class Chlorophyceae recently received detailed systematic attention and the taxonomic structure within the class was examined using chloroplast genome data (Fučíková et al. 2019). Consistent with previous literature, the study found strong support for the OCC clade, containing the orders Oedogoniales, Chaetopeltidales, and Chaetophorales. The OCC group is sister to the SV clade, which contains the order Volvocales (sometimes also referred to as Chlamydomonadales in literature) and the order Sphaeropleales, as well as several incertae sedis lineages. In fact, the monophyly of the order Sphaeropleales was not strongly supported, and to aid future communication about the systematics of the group, the clade names Scenedesminia and Treubarinia may be used alongside the family names Microsporaceae, Sphaeropleaceae, and several lineages that have yet to be given higher-taxon or clade names. Adding newly discovered lineages of incertae sedis to the current data set could help resolve the backbone of the chlorophycean phylogeny, possibly by breaking up the long solitary branches that are common in that part of the tree.

In the present study, we focus on a single enigmatic isolate from a biological soil crust originally collected from the Joshua Tree National Park (JTNP) by Flechtner et al. (2013). We present multiple lines of evidence for its taxonomic distinctness despite its simple, common morphology. We explore the relationship between the environment of origin and GC content in nuclear ribosomal genes in green algae and demonstrate the highly unusual nature of the nuclear rRNA loci in the JTNP alga. We compare the alga's chloroplast genome to those of other green algae and note the unusual structure of the plastome. Finally, we contribute data on the alga's lipid profile to provide a more complete characterization of the new taxon.

MATERIAL AND METHODS

Sample collection, strain isolation, and culture conditions

The material was collected as part of a larger study focused on the biodiversity of JTNP biological soil crusts (Pietrasiak et al. 2011; Flechtner et al. 2013). The sample WJT24 was collected at coordinates 33°45'47"N, 115°47'46"W on 13 Jun. 2006. The site was located north of Cottonwood, CA, U.S.A. and described as having sandy, gravelly soil derived from a granitic outcrop west of Eagle Mountains. The site was vegetated with desert plants typical for the Colorado Desert and also exhibited well developed biological soil crusts. A composite sample of the topsoil (0–1 cm soil depth) was collected systematically every 5 m along a 50 m transect (Pietrasiak et al. 2011).

A 1 g subsample of the composite soil was rehydrated in 100 ml Bold's Basal Medium (BBM) (Bischoff and Bold 1963) and dilution-plated on an agar plate, and algal colonies were picked as described in Flechtner et al. (1998, 2013). The culture, labelled WJT24VFNP31, was maintained on agarized BBM under 16:8 light:dark cycle at 18°C. Prior to lipid extraction, the cultures were placed in a growth rig with 40 W fluorescent lights at 15°C and 2000 to 4000 lux, 12:12 light:dark cycle, for three weeks or until reaching sufficient biomass for lipid extraction.

Microscopic examination

The morphology of the strain WJT24VFNP31 was first assessed using light microscopy to survey and measure cell size, shape, and internal structures visible under a light microscope. The Olympus BH-2 photomicroscope with Nomarski DIC optics was used for observations and photographs were obtained using an Olympus DP25 camera (Olympus, Center Valley, CA, USA). Cells from 2–4-week-old cultures were observed. To determine the number of nuclei in young and mature vegetative cells, the cells were first left overnight in a permeation buffer containing 0.1% Triton X and 0.05% Tween 20. Cells were washed three times with phosphate buffer saline (PBS), then fixated using 4% formalin for at least 2 hours. Samples were washed and combined with DAPI (diluted from stock 1:1000) for 7–10 minutes. Samples were washed again five times with PBS in order to remove any residual DAPI that would cause background staining. Cells were mounted onto a microscope slide in distilled H_2O and were visualized using a Nikon Eclipse 90i microscope (Nikon Instruments Inc., Melville, NY, USA) with an Andor camera (Andor Technology Ltd., Belfast, Northern Ireland).

DNA extraction, PCR, and sequencing

DNA was extracted from WJT24VFNP31 using MO BIO PowerPlant Pro DNA Isolation Kit (Qiagen, Germantown, MD, USA) following the manufacturer's protocol with the exception of using distilled water instead of Buffer 7. The total genomic DNA was sequenced using Illumina MiSeq (Illumina Inc., San Diego, CA, USA) technology (150 bp paired reads). Two million reads (2,078,788) were obtained in total from two separate MiSeq runs.

The sequences of the 18S and 28S genes were initially obtained using PCR and Sanger sequencing, following the protocols described in Saber et al. (2018) and Shoup and Lewis (2003). The Sanger results were consistent with the Illumina data and are not discussed in detail in this article. However, the fact that they were obtained from a different DNA extraction and matched the high-throughput sequencing data helps rule out potential sequencing artifacts and contamination. A low-coverage region in the Illumina assembly of the rRNA genes was well covered by five Sanger reads, which further adds confidence to our results.

Genome assembly and annotation

The Illumina reads were paired, trimmed and assembled de-novo using Geneious v.6 (Biomatters Inc., Boston, MA, USA; details in Fučíková et al. 2014c, 2016) and inspected by eye. Chloroplast tRNA and rRNA genes and intron conserved elements were identified using the RNAweasel tool (Lang et al. 2007, http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl), and their boundaries were edited based on alignment with other Chlorophycean sequences. Conserved protein-coding genes were identified by reference assembly of published plastid genes (*Ankyra judayi* NC_029735.1) to the WJT24VFNP31 plastid contig. Highly variable genes (e.g. *ycf1*, *ftsH*) were initially identified by the ORF prediction tool in Geneious and subsequent BLAST searches. All gene boundaries were verified based on alignment with the data set of Fučíková et al. (2019). The genome map (Supplementary material 1) was drawn using OGDRAW v.1.3.1 (Greiner et al. 2019).

After the full contigs for rDNA and the chloroplast genome were obtained, a reference assembly was performed in Geneious to visually inspect coverage and its distribution along the contigs. Stringent conditions were applied in the reference assembly, allowing no more than 1% mismatches and 1% gaps, with a word size of 20 nucleotides. Areas of low coverage were inspected by eye to rule out mis-assemblies.

rRNA secondary structure estimation

The NCBI nucleotide database was searched for closest matches for the rRNA gene sequences of WJT24VFNP31 using BLAST (Altschul et al. 1990). The results, in combination with existing 18S and 28S alignments, were used to determine the boundaries of the genes and internal and external transcribed spacers in the rDNA.

The secondary structure of the 18S gene was visualized using the *Chlamydomonas reinhardtii* P.A.Dangeard model from https://bioinformatics.psb.ugent.be/webtools/rRNA/secmodel/Crei_SSU.html (Van de Peer et al. 2000). The WJT24VFNP31 18S sequence was manually mapped onto this structure and from the mapping, compensatory base changes (CBCs) in paired regions were determined, as well as other substitutions in the nucleotide sequence and changes in the predicted secondary structure. Structure was verified using Mfold (Zuker 2003) in stems and loops with many nucleotide substitutions.

The secondary structure of the internal transcribed spacer 2 (ITS2) was predicted using the ITS2 Database (Koetschan et al. 2010, http://its2.bioapps.biozentrum.uni-wuerzburg.de), implementing the Identity matrix (because the ITS2 PAM 50 model yielded no results) with otherwise default settings for template-based modelling. *Spermatozopsis exsultans* Korshikov and *Ankyra judayi* (G.M.Smith) Fott were selected as templates after the initial search for templates yielded no results. Alternative structures of the individual helices were explored using Mfold (Zuker 2003).

Phylogenetic analyses

Nucleotide sequences of 59 chloroplast genes were aligned with the existing data from Fučíková et al. (2019), concatenated (69 taxa, 33,483 nucleotide positions) and analysed using MrBayes v.3.2 (Ronquist et al. 2012), applying the GTR+I+Γ model over 6,250,000 generations

and two MCMC chains in each of two parallel runs. Sites were partitioned by codon position across the entire data set. The first 20% of each run was discarded as burn-in. An analysis of the translated amino acid data set was also conducted, applying a mixed AA model (averaging over ten amino acid models) in MrBayes. The OCC clade was used to root the tree, consistent with previously published literature (Fučíková et al. 2019 and references within).

The evolution of the 18S GC content was reconstructed using fastAnc in the R package phytools v.1.2 (Revell 2012). The trait (18S GC content) was mapped onto the chloroplast nucleotide phylogeny. Additionally, the GC content of the coding chloroplast regions (including exons of protein-coding genes, rRNA and tRNA genes but not intronic or other hypothetical or unidentified ORFs) was mapped onto the chloroplast phylogeny.

To examine whether GC content in the 18S gene is correlated with the temperature of the locality of origin, we first conducted a Pearson correlation test in base R on the uncorrected temperature and 18S GC data, which were available for all but two of the Chlorophycean species included in our phylogenetic data set. We then regressed the 18S GC and average temperature of the warmest month at the locality of origin and corrected for the effect of phylogenetic relatedness using the phylogenetically independent contrasts (PIC) method (Felsenstein 1985) in phytools. The complete annotated R code and the underlying data set is available at https://github.com/KarolinaFucikova/GC_mapping, with a permanent DOI https://doi.org/10.5281/zenodo.8011298.

Lipid extraction and analysis

Three replicate cultures were grown for three weeks or longer until sufficient biomass was produced for extraction. A minimum of 15 mg of wet algal biomass was harvested and suspended in gas-chromatography vials with methanol. The open vials were placed in a glass desiccator overnight with Drierite desiccant. After weighing the vials for their dry mass the next day (optimum dry mass being above 5 mg), transesterification was conducted using 200 µl of 2:1 $CHCl_3$:methanol, 300 µl of 0.6 M HCl:methanol, and 25 µl of C13:0 methyl ester internal standard. The vials were tightly sealed with parafilm to prevent evaporation. After heating the samples at 85°C for an hour in a water bath, the vials were cooled and 1 ml of hexane was added to each vial. The vials were then vortexed, allowed to rest for 1–4 hours, then 500 µl of the top layer was removed and placed in a new vial (Van Wychen et al. 2015).

All lipids were quantitated by a Varian CP-3800 Gas Chromatograph using a Saturn 2000 Mass Spectrometry detector (Varian, Palo Alto, CA, USA). The gas chromatography analysis was conducted using a Phenomenex ZB-WAX capillary column (Phenomenex, Torrance, CA, USA) with helium as carrier gas with the following parameters: helium flow 1 mL/min, temperature program: 100°C for 1 min, 25°C/min heating up to 200°C, hold for 1 min, 5°C/min heating up to 250°C, hold for 7 min. For mass spectrometry, a two-minute solvent delay was followed by electron impact ionization and detection of masses from 45–400 m/z. Quantitation was accomplished with the Total Ion Chromatogram for all lipids except for C22:5 (quant ions 79 + 91 m/z) and C24:0 (quant ion 382 m/z) because their peaks overlapped.

The lipid content, composition, and saturation indexes were calculated in MS Excel. The percent total lipid content of the samples was calculated using the initial dry mass of the algae after they were desiccated. Based on total lipid content, estimated percent composition of the various fatty acids were determined.

RESULTS

Morphological description

Cells were solitary, spherical or nearly so (Fig. 1A–E). Young cells in two-week and one-month-old cultures were 4–6 µm in diameter. Mature cells were up to ca 12 µm in diameter. Older cells had visible cytoplasmic granules/deposits that were dark reddish-brown and occurred outside of the plastids (Fig. 1E, highlighted with arrowheads). Young cells contained at first a single cup-shaped chloroplast, which then became lobed and broke up into multiple parietal plastids as the cells aged. No pyrenoid was observed. Cell wall did not appear to thicken appreciably with age. Reproduction via four autospores was observed (Fig. 1C), but additional sporangia-like cells were seen, which were large and irregular in shape (Fig. 1F, G). Senescent cultures of 6+ months were deep orange in colour. DAPI staining demonstrated that young cells have a single nucleus (Fig. 2A, B) and maturing cells can have two or more (Fig. 2C, D). Chloroplast DNA stains with DAPI as well, which makes the precise determination of the number of nuclei difficult.

Phylogenetic position of the strain WJT24VFNP31

The phylogenetic relationships among Chlorophyceae, derived from the sequences of 59 chloroplast genes, were largely consistent with the phylogenies of Fučíková et al. (2019), and similar discrepancies were found between the nucleotide and amino acid topologies. In both analyses (nucleotide and amino acid), the position of the strain WJT24VFNP31 was sister to *Spermatozopsis similis* H.R.Preisig & M.Melkonian with absolute support, and the two taxa formed a clade sister to the family Sphaeropleaceae (*Ankyra* + *Atractomorpha*), also with absolute support of BPP = 1.0 (Fig. 3). The position of this four-taxon clade with respect to the main SV lineages (Volvocales, Scenedesminia, Treubarinia, Microsporaceae, and several incertae sedis taxa) differs between analyses, as it did in Fučíková et al. (2019). Specifically, in the amino acid analysis the Sphaeropleaceae

+ *Spermatozopsis* + WJT24VFNP31 appeared sister to Volvocales, consistently with Fučíková et al. (2019).

Chloroplast genome size, structure, and content

Under the stringent reference assembly conditions, which gave a conservative (low) estimate, the chloroplast genome coverage ranged from 12 to 139 (mean 75.9). The chloroplast genome was assembled in two equally probable structures: either in a single circular contig with two direct (not inverted) repeats containing ribosomal genes, or in two circular contigs containing one of the repeats each. Without long-read sequencing data, it is not possible to determine the true structure of the genome, and it is also possible that the genome exists in both forms. Nevertheless, all expected genes were detected in the contig(s) and upon visual inspection there were no areas of suspected mis-assembly. Therefore, we concluded that the chloroplast genome sequence is likely complete, despite the uncertainty about its structure.

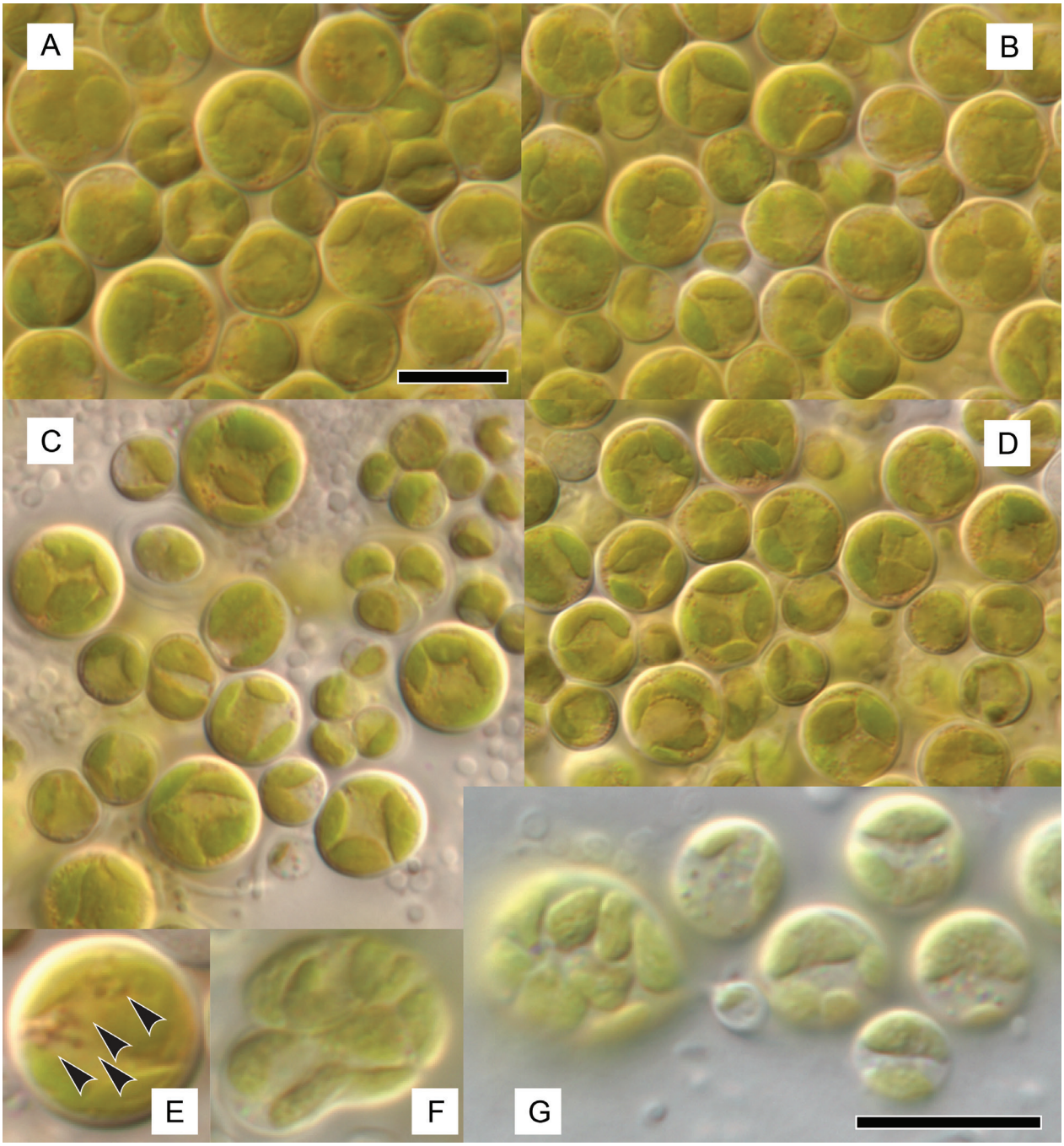

Figure 1. Light micrographs capturing the morphology of *Johansenicoccus eremophilus* gen. et sp. nov. **A**, **B**. Young and mature vegetative cells. **C**. Vegetative cells and autospore formation. **D**. Mature vegetative cells with clear multiple chloroplasts and extra-plastidic inclusions. **E**. Detail of vegetative cell; arrows point to unidentified extra-plastidic inclusions. **F**, **G**. Irregularly shaped cells, possibly in the process of spore production. Scale bars represent 10 µm, second scale bar pertains to E–G.

The chloroplast genome of WJT24VFNP31 (GenBank OQ849777) is 235,854 bp in size. Of the two possible structural configurations we present the single circular molecule with two direct repeats. The repeats were confirmed by their twofold read coverage compared to the rest of the plastome, and contain ribosomal RNA genes, two tRNA genes (trnA-UGC and trnI-GAU), and part of the *petA* gene. No introns were found in the repeats (Supplementary material 1).

All genes expected in Sphaeropleales and Chlorophyceae incertae sedis were detected in the WJT24VFNP31 plastome. In addition, the uncommon (in Chlorophyceae) trnR-UCG gene was present, as was a trnT-AGU gene. The genome also contained a trans-spliced *psaA* gene, split into two exons at position 269, as is common in the SV clade, but contiguous at position 89, a condition only reported from a handful of SV taxa. A full account of gene and intron content across analysed chlorophyceans is presented in Supplementary material 2.

Nuclear rDNA GC content and secondary structure

The full nuclear ribosomal gene set is deposited in GenBank under accession number OQ849776. The top BLAST match for the 18S gene was 84.1% similar (*Heterochlamydomonas inaequalis* Ed.R.Cox & T.R.Deason), and others in the 83–84% range included other Volvocales, Sphaeropleales (e.g., *Mychonastes* P.D.Simpson & S.D.Van Valkenburg), and even Trebouxiophyceae (*Neocystis* Hindák), further underscoring the uniqueness and uncertain phylogenetic position of WJT24VFNP31 and its rDNA sequence.

The GC content in the 18S gene was 54.9%, which is 3% higher than the next highest GC in the data set, which is in *Jenufa perforata* Němcová, M.Eliáš, Škaloud & Neustupa (51.9%). The third highest was *Borodinellopsis texensis* Dykstra (51.1%), illustrating that high GC content is not exactly clade-specific, although high 18S GC seems to be more common among incertae sedis taxa (lineages outside of OCC, Volvocales, and Scenedesminia, Fig. 4).

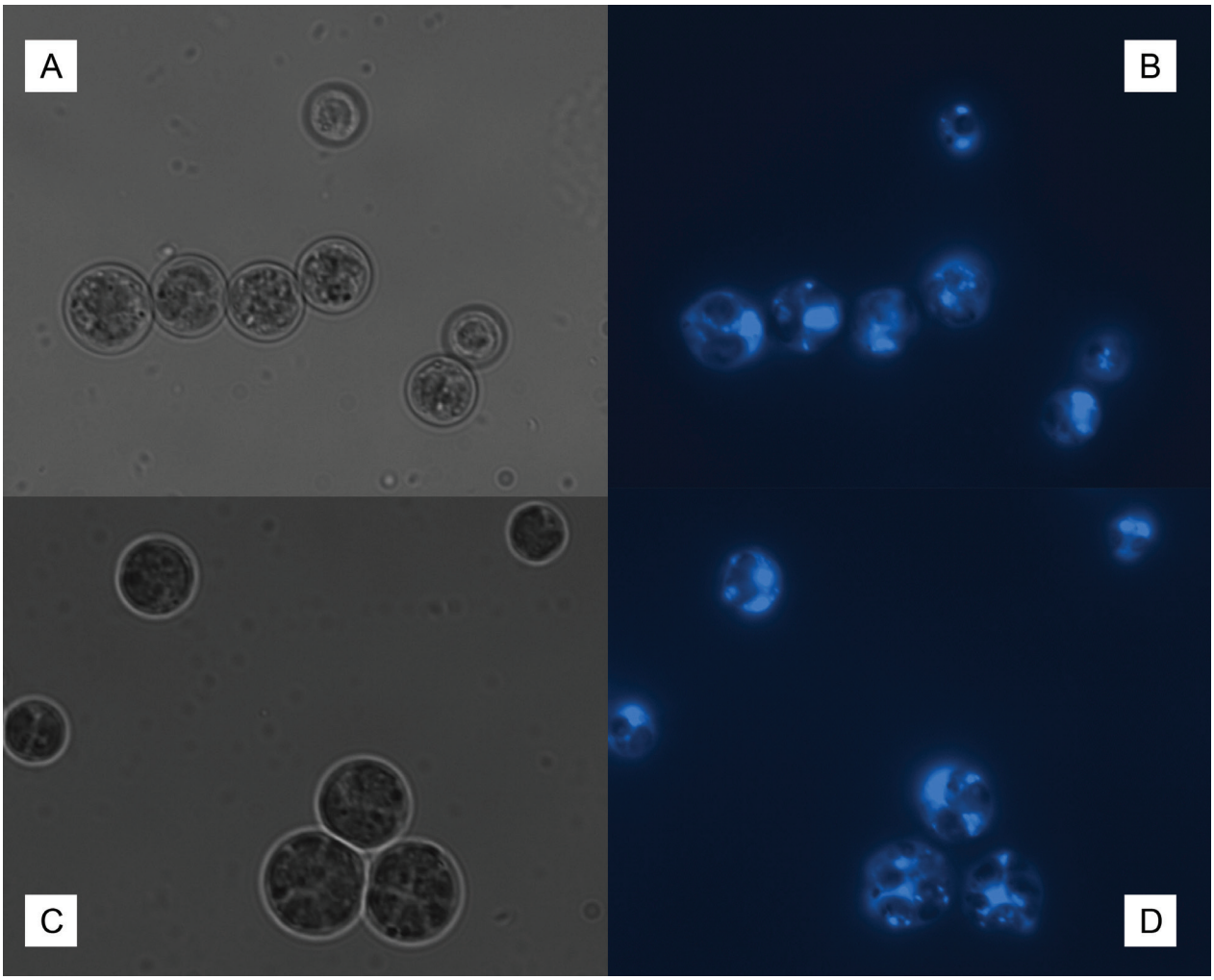

Figure 2. Micrographs of DAPI-stained cells of *Johansenicoccus eremophilus* gen. et sp. nov. **A**, **B**. Young vegetative cells. **C**, **D**. Mature vegetative cells. A, C. Bright field. B, D. Fluorescence micrographs corresponding to A and C, respectively.

The full compilation of GC content in the rRNA genes and chloroplast coding regions of Chlorophyceae is available in our GitHub repository (https://doi.org/10.5281/zenodo.8011298). The 18S secondary structure, compared to that of *Chlamydomonas reinhardtii*, seemed largely maintained despite the numerous substitutions and elevated GC content (Supplementary materials 3, 4). Many of the substitutions and some structural changes were unique to WJT24VFNP31.

The ITS2 structure was significantly different from those currently deposited in the ITS2 Database. The Database tools readily detected the proximal stem and the borders of the spacer, however, the initial search for suitable templates did not yield any results. Two of the putatively closest relatives of the strain WJT24VFNP31 were therefore selected as templates for structural folding: *Spermatozopsis* (represented in the database only by *S. exsultans*), and *Ankyra judayi*. No high-quality model

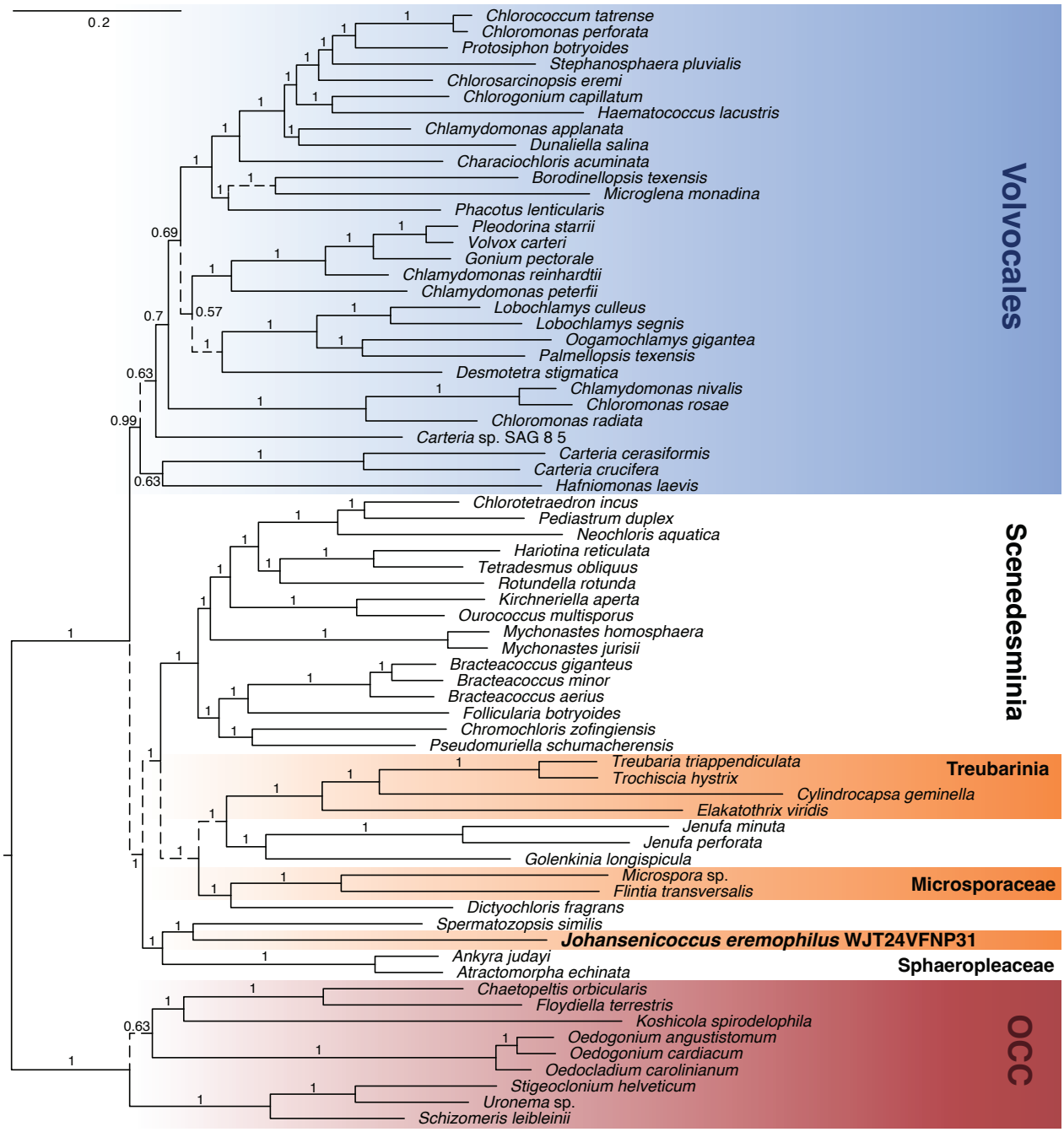

Figure 3. Bayesian consensus phylogeny based on an analysis of 59 concatenated protein-coding chloroplast genes, with the nucleotide data partitioned by codon position. Numbers on branches represent Bayesian Posterior Probability. Solid branches represent relationships that were also recovered in the corresponding Bayesian analysis of the amino acid data. Dashed branches indicate relationships that were not recovered in the amino acid analysis. Scale bar represents the estimated number of nucleotide substitutions per site.

could be derived from the *Spermatozopsis* template. Percentages of helix transfer were 64, 35, 50, and 22 for the four helices, respectively. Mapping was also performed onto the *Ankyra* template and yielded helix transfer percentages of 66, 37, 60, and 50. *Ankyra judayi* was thus a better fit as a template for WJT24VFNP31. The predicted ITS2 structure modelled onto the *Ankyra* template contained many extra bulges (unpaired regions), and therefore alternative structures of individual helices were explored. Mfold (Zuker 2003) yielded some structures visually more similar to the helices of the templates and other algae (Supplementary material 5), but the full secondary structure of WJT24VFNP31's ITS2, as well as its functionality remains unresolved.

Evolution of GC content in Chlorophyceae

Figure 4 shows how high the 18S GC content is compared to other Chlorophyceae, and that this feature evolved along the long branch subtending WJT24VFNP31 from an ancestor with low to moderate GC content. The sister taxon to WJT24VFNP31 is *Spermatozopsis similis*, which has a low GC content in its 18S gene (47.8%). The figure also shows the uniquely high GC content in *Mychonastes* among the otherwise low-GC Scenedesminia. This could explain the previous studies' difficulty with phylogenetically placing this genus. The 28S gene generally follows the trend of the 18S gene across Chlorophyceae – taxa with high 18S GC also have a high 28S GC and vice versa. WJT24VFNP31 has extremely high 28S GC (58.1%) compared to all other taxa examined.

In contrast to the nuclear ribosomal GC content, the GC content of the chloroplast coding regions in WJT24VFNP31 is 38.5%, which is not remarkably high compared to other Chlorophyceae. Both species of *Mychonastes* have over 40% GC, and the volvocalean *Stephanosphaera pluvialis* Cohn has the highest chloroplast coding GC of 44%. Supplementary material 6 shows both mappings in colour and our GitHub repository contains the full data set for further exploration (https://doi.org/10.5281/zenodo.8011298).

The Pearson correlation test yielded a significant positive correlation between the 18S GC content and maximum temperature at the locality of origin ($p = 0.0014$), and the corresponding plot showed the strain WJT24VFNP31 as an outlier with extremely high 18S GC, even given the warm, desert climate at its locality of origin (Fig. 5). When WJT24VFNP31 was excluded from the data set, the Pearson correlation was still significantly positive ($p = 0.0054$, $r = 0.3607$). After correction for the

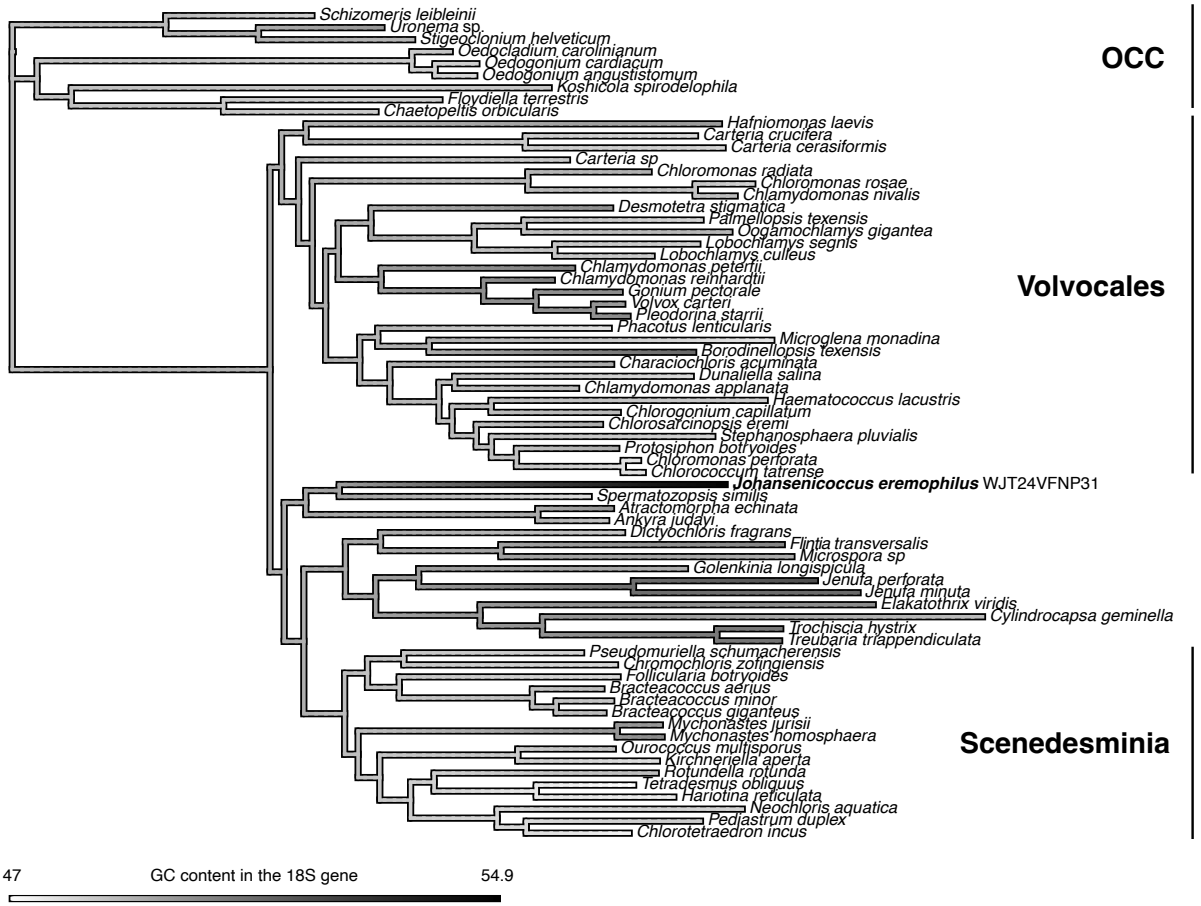

Figure 4. Evolution of GC content in the 18S nuclear ribosomal gene mapped onto an independently derived phylogeny (based on data from 59 chloroplast genes). Darker colour signifies higher 18S GC content.

effect of phylogeny (WJT24VFNP31 included), the 18S GC content still correlated positively and significantly with temperature ($p = 0.0004$, adjusted $R^2 = 0.1838$, slope of 0.0842, Supplementary material 7). No significant correlation with temperature was detected in the GC of the chloroplast coding regions (https://doi.org/10.5281/zenodo.8011298).

The chloroplast ribosomal genes were not investigated in depth, as data for several taxa are not available due to the partial nature of the published genome sequences. In most algae in our data set, the GC content in chloroplast ribosomal genes was 10–20% higher than the general coding GC in the chloroplast. The strain WJT24VFNP31 and *Bracteacoccus giganteus* H.W.Bischoff & Bold had the highest GC content in the rrs gene: 53.7%. All three included species of *Bracteacoccus* had rrs GC over 53%, while most other taxa had rrs GC < 52%. There was a weak positive correlation between temperature and rrs GC: $r = 0.2733$ and $p = 0.0499$. The data set is available in our GitHub repository for further exploration (https://doi.org/10.5281/zenodo.8011298).

Fatty acid profile

The most abundant fatty acid in the strain WJT24VFNP31 was alpha-linolenic acid (C 18:3 n-3, 56.0% on average), followed by palmitic (C16:0, 24.7%) and vaccenic (C18:1 trans-11, 9.7%) acids (Supplementary material 8). Oleic (C18:1 cis-9), linoleic (C18:2), and gamma-linolenic (C18:3 n-6) acids were also present in detectable amounts. On average, 73% of the fatty acids detected by the GC-MS method were unsaturated.

TAXONOMIC TREATMENT

Johansenicoccus Fučíková & Pietrasiak, **gen. nov.**
Fig. 1

Type species. *Johansenicoccus eremophilus* Fučíková & Pietrasiak.

Diagnosis. Spherical, broadly oval or irregular cells with multiple parietal chloroplasts in maturity, without pyrenoids. Reproduction via autospores. Resembles several coccoid genera, especially *Bracteacoccus* Tereg, *Chromochloris* Kol & F.Chodat, *Pseudomuriella* N.Hanagata, and *Rotundella* Fučíková, P.O.Lewis & L.A.Lewis, but is distinct phylogenetically as evidenced by analyses of 18S rDNA and *rbcL* data (GenBank accessions OQ849776 and OQ849777, respectively).

Etymology. Named in honour of Dr Jeffrey R. Johansen, the Brontosaurus of desert soil algae.

Registration. http://phycobank.org/103939

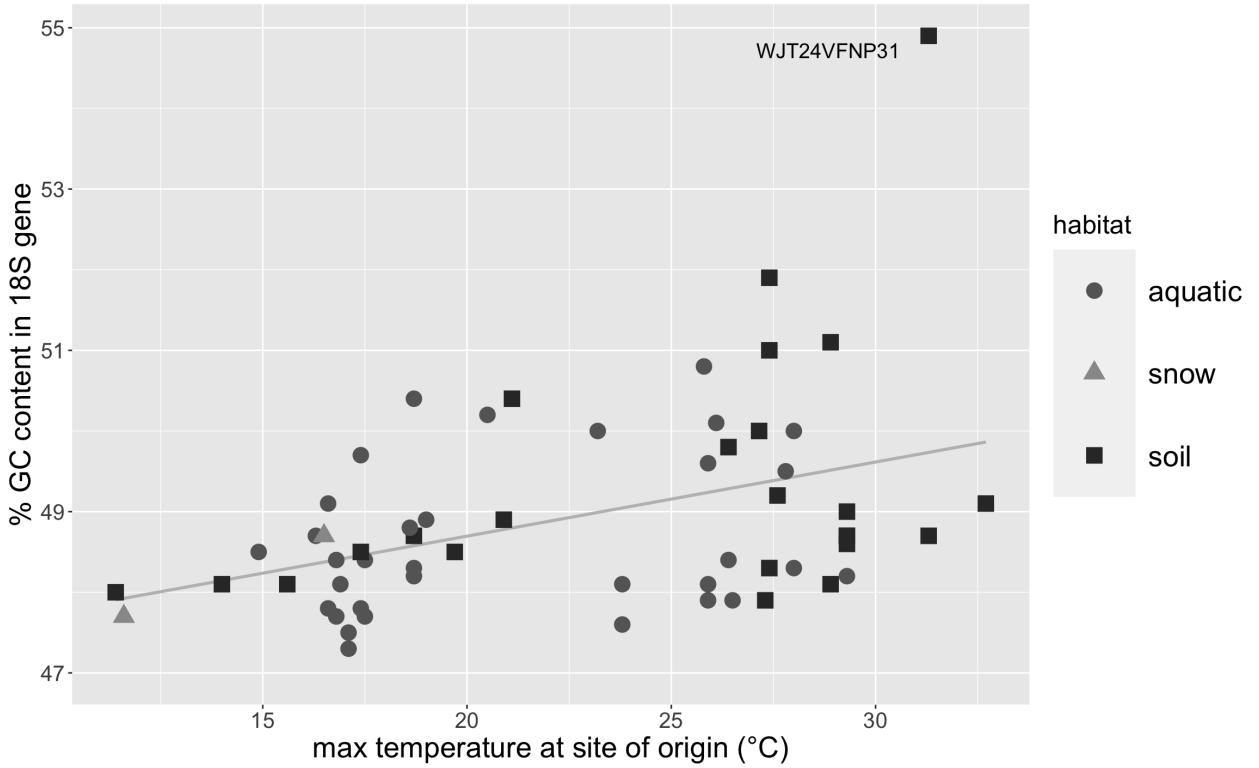

Figure 5. Regression analysis of the relationship between temperature at the site of origin and GC content in the 18S gene across 59 species of green algae. Temperature refers to the average temperature in the warmest month at or near the site of the species/strain's collection. Algae from different habitats (aquatic, snow, soil) are represented with differently shaped symbols. The strain WJT24VFNP31 (*Johansenicoccus eremophilus* gen. et sp. nov.) is labelled. Grey line represents the fitted regression line.

Johansenicoccus eremophilus Fučíková & Pietrasiak, **sp. nov.**
Fig. 1

Type. USA – **California** • Joshua Tree National Park; 33°45'47"N, 115°47'46"W; 13 Jun. 2006; holotype: fixed algae on microscope slide, CONN [CONN00234349]; authentic culture: UTEX B 3223.
Description. Cells solitary, spherical or rarely oval or irregular, 4–12 µm in diameter. Chloroplasts parietal without pyrenoids. In young cells, chloroplast single and cup-shaped or lobed, nucleus single. Multiple chloroplasts and nuclei in mature cells. Older cells contain darkly reddish-brown cytoplasmic granules outside the plastids. Cell wall thin and smooth, not thickening appreciably with age. Reproduction via four or more autospores; zoospore production uncertain. In culture, the cells form elevated colonies with lighter, well-defined margins, which grow and merge into a firm, glossy lawn. Senescent cultures accumulate orange pigments, turning from green to olive and eventually to deep orange in colour.
Distribution. So far only known from the type locality, Joshua Tree National Park, USA.
Habitat and ecology. Biological soil crust, desert.
Etymology. The name reflects the desert-dwelling nature of the species, as erêmos means desert in ancient Greek and phílos means loving.
Registration. http://phycobank.org/103940

DISCUSSION

The morphology of *J. eremophilus* is a fairly common one among Chlorophyceae. Spherical cells with multiple parietal, pyrenoid-free chloroplasts are also found in *Bracteacoccus*, *Bracteamorpha* Fučíková, P.O.Lewis & L.A.Lewis, *Chromochloris*, *Pseudomuriella*, *Rotundella*, and *Tumidella* Fučíková, P.O.Lewis & L.A.Lewis. These genera are all made up of soil-dwelling species and include numerous desert-dwellers. Additionally, the aquatic genus *Dictyococcus* Gerneck can be considered morphologically similar but is readily distinguished by the inflexed edges of its chloroplasts (Fučíková et al. 2011, 2014a). The phylogenetic distribution of the *Bracteacoccus*-like morphology suggests convergent evolution and possible association with soil habitats. To date, *J. eremophilus* is the only taxon with this morphology placed conclusively outside of the clade Scenedesminia (Fig. 3).

The genus *Tumidella* was not represented in our phylogenetic data set but was shown as related to *Bracteacoccus* and other Scenedesminia by Fučíková et al. (2014a) with high statistical support. The cells of *Tumidella* grow quite large and have numerous tiny chloroplasts in maturity. *Bracteamorpha* and some species of *Bracteacoccus* also tend to have more numerous, small chloroplasts at maturity, as well as larger vegetative cells and in some cases a thick secondary cell wall – all of which distinguish these genera from *Johansenicoccus*. Visually, *J. eremophilus* is therefore most easily mistaken for *Chromochloris*, *Pseudomuriella*, *Rotundella*, and smaller *Bracteacoccus* species. *Rotundella rotunda* Fučíková, P.O.Lewis & L.A.Lewis (interestingly also found in JTNP) cells also appear to have cytoplasmic granules (seen in Fučíková et al. 2014a: fig. 1) like *J. eremophilus,* though their pigmentation in *R. rotunda* is not as obvious as in our *Johansenicoccus* material. These granules could possibly be lipid droplets, which could be confirmed in future studies using visualization techniques that were beyond the scope of this study, or the cells could be subjected to nitrogen starvation to observe whether the droplets will grow larger and more numerous, as has been documented in other green algae under such conditions (e.g. Jaeger et al. 2016; Liu et al. 2022).

A confident identification of any of the aforementioned genera should involve the sequencing of at least one molecular marker. We recommend the chloroplast gene *rbcL*, as we cannot be sure if the anomalous nuclear rDNA is unique to only the strain WJT24VFNP31 or if similar sequence would be found in any newly discovered *Johansenicoccus* isolates in the future.

Spermatozopsis similis appears to be the closest relative of *J. eremophilus*. Their relationship received absolute statistical support in our analyses, although the two taxa are separated by considerable genetic distance (Fig. 3). The two species share very little in terms of morphology, aside from both being single-celled organisms. *Spermatozopsis* is a freshwater-dwelling, naked flagellate with two currently recognized species in the genus. The quadriflagellate *S. exsultans* is the type species, in addition to the biflagellate *S. similis* (Preisig and Melkonian 1984). Both *Spermatozopsis* species share a similar, spirally-twisted cell shape.

The sister clade to *S. similis* and *J. eremophilus* is the family Sphaeropleaceae, represented in our data by *Ankyra judayi* and *Atractomorpha echinata* L.R.Hoffmann. The position of *S. similis* in the phylogenetic proximity of Sphaeropleaceae is consistent with Marin (2012) and Fučíková et al. (2019) and was not changed by the addition of *J. eremophilus* to the data set. In this context of related taxa, *J. eremophilus* stands out as the only spherical-celled coccoid and the only soil dweller, further expanding the range of morphologies in the clade that already contains flagellates, filaments, and large coenocytic forms.

Unique nuclear ribosomal loci

The nuclear ribosomal region of *Johansenicoccus eremophilus* contains the 18S gene, 5.8S gene, 28S gene, and two internal transcribed spacers in the expected order, but the rRNA genes are extremely GC-rich compared to other Chlorophyceae (Figs 4, 5). The maintenance of 18S secondary structure (Supplementary materials 3, 4) suggests that the unusual 18S sequence codes for a functional ribosomal subunit. This is consistent with the fact that in our genomic data the 'abnormal' ribosomal region is abundant and therefore unlikely

to be a deteriorating pseudogene. Additional reference assemblies have hinted at the presence of another low-copy rDNA region (also very divergent from other GenBank sequences), but it was represented only by a few reads and we were unable to assemble it fully. We have confirmed the unusual 18S (main copy) with PCR and Sanger sequencing to rule out chimeric assemblies in the high-throughput data. The PCR primers did not amplify any other 18S copy that could be a putative paralog. Therefore, while it is possible that other, paralogous copies of the gene and spacer set exist in the genome of *J. eremophilus*, all of our evidence currently suggests that the unusual rDNA genes discussed here code for the components of functional ribosomes in the species.

One previously explored hypothesis is that high GC content may be an adaptation to hot climates because of the higher thermal stability of the G-C pairing compared to the A-T/A-U pairing. While there does not seem to be a correlation between GC content and temperature for whole genomes or the (more or less) freely evolving third codon positions, the pattern has been consistently recovered for structural RNA in prokaryotes (Hurst and Merchant 2001), where species living in higher temperatures tend to have higher GC content in their ribosomal genes. This could be due to the high selective pressure to maintain the ability to translate proteins, which is contingent of a functional structure of ribosomes and therefore the secondary structure of rRNA after transcription. Such an issue would not necessarily arise in chromatin (DNA). Somewhat contrastingly, however, Šmarda et al. (2014) reported higher genomic GC content as indicative of adaptation to cold and drought stress in monocots. The drivers of GC content are therefore likely quite complex.

In our study we uncovered a significant positive correlation between 18S GC content and warmest-month temperature at the locality of origin (Fig. 5), consistent with Hurst and Merchant (2001). However, temperature or desert habitat alone do not explain the unusual and GC-rich DNA sequence of the rRNA genes in *J. eremophilus*. Other desert algae, such as *Rotundella rotunda* (also from JTNP), do not have nearly such extreme GC levels in their ribosomal genes. Therefore, our current data only provide tentative support for the thermal adaptation hypothesis, and more research is needed to properly explain the unusual nature of the *J. eremophilus* nuclear ribosomal loci.

This uniqueness in ribosomal genes all but precludes phylogenetic analysis using 18S or 28S, which are otherwise very commonly used in algal systematics. Using secondary structure to guide the alignment and analysis, as recommended by e.g. Wolf et al. (2008) and Czech and Wolf (2020), does not remedy the extremely long branch subtending *J. eremophilus* in any attempted 18S-based phylogenies (not shown) and the placement of the species in such phylogenies does not receive high support. ITS2, another commonly used marker, faces similar problems (being very divergent from other green algal taxa, Supplementary material 5), though this marker is more appropriate for genus- and species-level resolution and we currently only have one strain and species of *Johansenicoccus*. If additional representatives of the genus are found in the future, ITS2 may still be of use for species-level taxonomy.

Unusual chloroplast genome

The chloroplast genome of *Johansenicoccus eremophilus* is nearly 236 kbp, which makes it the largest among its closest relatives, *Spermatozopsis* (135 kbp, Fučíková et al. 2019) and Sphaeropleaceae (157 kbp in *Ankyra*, Fučíková et al. 2016). The gene content in the plastome of *J. eremophilus* is nearly perfectly consistent with its phylogenetic position in the Chlorophyceae (details in Supplementary materials 1, 2). Like its closest relatives, *J. eremophilus* has very few introns in its plastome. Like other non-OCC chlorophyceans, *J. eremophilus* lacks the gene *psaM* and has a trans-spliced *psaA*. The two-exon configuration of *psaA* (as opposed to a three-exon configuration) is not common, and is known from four divergent non-OCC chlorophyceans, *J. eremophilus*, *Golenkinia longispicula* Hegewald & Schnepf (incertae sedis), *Chloromonas radiata* (T.R.Deason & Bold) T.Pröschold, B.Marin, U.W.Schlösser & M.Melkonian (Volvocales), and *Tetradesmus obliquus* (Turpin) M.J.Wynne (Scenedesminia), suggesting that the two-exon state may be convergently evolved.

Unlike the unsurprising gene content, the structure of the *J. eremophilus* plastome is noteworthy, being either a single circular molecule with two large direct repeats containing blocks of rRNA genes (Supplementary material 1), or two smaller circular molecules with one rRNA block present in each circle. While our data cannot confidently resolve which of the alternative configurations is the true one, both are unusual in Chlorophyceae. Multiple chloroplast chromosomes have been reported in the OCC taxon *Koshicola spirodelophila* Shin Watanabe, Fučíková & L.A.Lewis (Chaetopeltidales, Watanabe et al. 2016) and outside of green algae it is known in dinoflagellates (Zhang et al. 1999). A similar situation is known for some mitochondrial genomes as well (Smith et al. 2010). In contrast, the single-chromosome configuration would be unique among green algae because of the orientation of the rRNA gene-containing repeats. Most plants and green algae contain two copies of the ribosomal RNA gene set too, but it is positioned on an inverted repeat (IR) segment in their chloroplast genome. In the Chlorophyceae, species in Chaetophorales and Chaetopeltidales lack an IR and only possess a single copy of each of their rRNA genes. A genome with "un-inverted" repeats, however, is not known from any green algal lineage, though this configuration was recently reported in two species of the vascular plant genus *Selaginella* P.Beauv. (Xu et al. 2018; Mower et al. 2019), and outside of green plants also in the heterokont alga *Rhizochromulina marina* D.J.Hibberd & Chrétiennot-Dinet (Han et al. 2019). A similar orientation (tandem, or

direct repeats positioned immediately next to each other on the genome) was also found in several euglenophytes (Wiegert et al. 2012; Maciszewski et al. 2022). Further, Wang and Lanfear (2019) reported that in most plants (though not *Selaginella tamariscina* (P.Beauv.) Spring) the plastome exists in two equally frequent versions, which have the short single copy region oriented in the opposite directions. Given these new discoveries, the unusual plastome structure of *J. eremophilus* is perhaps less surprising. More in-depth investigations could include long-read sequencing data and the statistical approach of Wang and Lanfear (2019).

Lipid composition

Microalgal lipid production is subject to a growing body of literature, as microalgal cultivation promises advances in biofuels, feedstock, and nutritional supplements. Green algal biomass tends to be rich in 16- and 18-carbon fatty acids, and often contains significant amounts of unsaturated lipids. Our results are consistent with this body of literature. For example, *Tetradesmus obliquus*, a commonly used algal model, was found to contain similar fatty acids as found in our study (Choi et al. 1987), though in a different order of abundance: linoleic (only found in small amounts in our study), palmitic (second most-abundant and the only significantly represented saturated fatty acid in both studies) and alpha-linolenic (most abundant in our study). Unsaturated fatty acid composition was higher in *Tetradesmus* G.M.Smith than in *J. eremophilus* (80% vs 73%). Saber et al. (2018) analysed the lipid composition of two Sphaeroplealean desert algae and cited fatty-acid composition roughly in agreement with Hong et al. (2012), with the abundant alpha-linolenic and palmitic acids being consistent with our study.

The evolutionary significance of lipid composition has been well studied in vascular plants. Such studies have shown that plant cells can desaturate their membrane lipids in response to lower temperature and vice versa. The mechanisms of the maintenance of membrane fluidity are complex especially in environments where temperature fluctuates rapidly on a daily basis, as can be true in alpine environments and deserts (Zheng et al. 2011 and references within). Therefore, a comparative study would be needed to determine whether the lipid content and composition in *J. eremophilus* reflects any adaptations to its desert lifestyle. Such a follow-up study would also need to explore different culturing conditions. Breuer et al. (2013), for example, demonstrated that nitrogen deprivation can alter the fatty acid composition in *Tetradesmus* significantly, and Nzayisenga et al. (2020) showed that heterotrophic growth with varying availability of organic carbon also affects the lipid production in several green algal species. Perhaps most importantly, the study would have to grow the algae at different temperatures, and ideally directly compare *J. eremophilus* to other desert and non-desert strains.

CONCLUSION

In summary, we have used multiple lines of data to characterize a novel lineage of green algae, *Johansenicoccus eremophilus*. The alga has several unusual genomic properties, which contrast with its simple and common morphology. Some of the unique genomic features may be adaptive to the desert lifestyle of the alga. So far, ours is the only find of this taxon worldwide with no known close relatives, but additional surveys of soil algae may well discover new species of *Johansenicoccus* in the future. Such discoveries would allow for better understanding of the gene and genome evolution in this lineage.

ACKNOWLEDGEMENTS

We thank the Joshua Tree National Park for permission to survey biological soil crusts in 2006 and characterize the park's soil algal flora. This strain was obtained under national park permit #JOTR-2006-SCI-0018. Sampling efforts in Joshua Tree National Park were supported by the California Desert Research Fund at The Community Foundation, Robert Lee Graduate Student Research Grant, and the Phycological Society Grant in Aid of Research awarded to NP during her Ph.D. studies. We thank Dr Valerie R. Flechtner for spearheading the collection and study of WJT green algae. Genomic sequencing was supported by NSF grant DEB-1036448. Lipid analysis materials and equipment, as well as the summer support for MT were provided by the Assumption University Department of Biological and Physical Sciences. Summer support for AI was provided by the Assumption University Honors Program. Analyses were carried out at the Computational Biology Core Facility of the University of Connecticut.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aratboni HA, Rafiei N, Garcia-Granados R, Alemzadeh A, Morones-Ramírez JR (2019) Biomass and lipid induction strategies in microalgae for biofuel production and other applications. Microbial Cell Factories 18: 178. https://doi.org/10.1186/s12934-019-1228-4

Bischoff HW, Bold HC (1963) Some soil algae from enchanted rock and related algae species. Phycological Studies 44(1): 1–95.

Breuer G, Evers WAC, de Vree JH, Kleinegris DMM, Martens DE, Wijffels RH, Lamers PP (2013) Analysis of fatty acid content and composition in microalgae. Journal of Visual Experiments 80: 50628. https://doi.org/10.3791/50628

Choi KJ, Nakhost Z, Barzana E, Karel M (1987) Lipid content and fatty acid composition of green algae *Scenedesmus obliquus* grown in a constant cell density

apparatus. Food Biotechnology 1(1): 117–128. https://doi.org/10.1080/08905438709549660

Czech V, Wolf M (2020) RNA consensus structures for inferring green algal phylogeny: a three–taxon analysis for *Golenkinia*/*Jenufa*, Sphaeropleales and Volvocales (Chlorophyta, Chlorophyceae). Fottea 20(1): 68–74. https://doi.org/10.5507/fot.2019.016

Felsenstein J (1985) Phylogenies and the comparative method. The American Naturalist 125(1): 1–15. https://doi.org/10.1086/284325

Flechtner VR, Johansen JR, Clark WH (1998) Algal composition of microbiotic crusts from the Central Desert of Baja California, Mexico. Great Basin Naturalist 58(4): 295–311. https://doi.org/10.2307/41713069

Flechtner VR, Pietrasiak N, Lewis LA (2013) Newly revealed diversity of green microalgae from wilderness areas of Joshua Tree National Park (JTNP). Monographs of the Western North American Naturalist 6(1): 43–63. https://doi.org/10.3398/042.006.0103

Fučíková K, Rada JC, Lewis LA (2011) The tangled taxonomic history of *Dictyococcus*, *Bracteacoccus* and *Pseudomuriella* (Chlorophyceae, Chlorophyta) and their distinction based on a phylogenetic perspective. Phycologia 50(4): 422–429. https://doi.org/10.2216/10-69.1

Fučíková K, Lewis PO, Lewis LA (2014a) Putting *incertae sedis* taxa in their place: a proposal for ten new families and three new genera in Sphaeropleales (Chlorophyceae, Chlorophyta). Journal of Phycology 50(1): 14–25. https://doi.org/10.1111/jpy.12118

Fučíková K, Lewis PO, Lewis LA (2014b) Widespread desert affiliation of trebouxiophycean algae (Trebouxiophyceae, Chlorophyta) including discovery of three new desert genera. Phycological Research 62(4): 294–305. https://doi.org/10.1111/pre.12062

Fučíková K, Lewis PO, González-Halphen D, Lewis LA (2014c) Gene arrangement convergence, diverse intron content, and genetic code modifications in mitochondrial genomes of Sphaeropleales (Chlorophyta). Genome Biology and Evolution 6(8): 2170–2180. http://doi.org/10.1093/gbe/evu172

Fučíková K, Lewis PO, Lewis LA (2016) Chloroplast phylogenomic data from the green algal order Sphaeropleales (Chlorophyceae, Chlorophyta) reveal complex patterns of sequence evolution. Molecular Phylogenetics and Evolution 98: 176–183. https://doi.org/10.1016/j.ympev.2016.01.022

Fučíková K, Lewis PO, Neupane S, Karol KG, Lewis LA (2019) Order, please! Uncertainty in the ordinal-level classification of Chlorophyceae. PeerJ 7: e6899. https://doi.org/10.7717/peerj.6899

Gray DW, Lewis LA, Cardon ZG (2007) Photosynthetic recovery following desiccation of desert green algae (Chlorophyta) and their aquatic relatives. Plant, Cell & Environment 30(10): 1240–1255. https://doi.org/10.1111/j.1365-3040.2007.01704.x

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47(W1): W59–W64. https://doi.org/10.1093/nar/gkz238

Han KY, Maciszewski K, Graf L, Yang JH, Andersen RA, Karnkowska A, Yoon HS (2019) Dictyochophyceae plastid genomes reveal unusual variability in their organization. Journal of Phycology 55(5): 1166–1180. https://doi.org/10.1111/jpy.12904

Hong JW, Kim SA, Chang JW, Yi J, Jeong JE, Kim S, Kim SH, Yoon HS (2012) Isolation and description of a Korean microalga, *Asterarcys quadricellulare* KNUA020, and analysis of its biotechnological potential. ALGAE 27(3): 197–203. https://doi.org/10.4490/algae.2012.27.3.197

Hurst LD, Merchant AR (2001) High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proceedings of the Royal Society of London Series B: Biological Sciences 268(1466): 493–497. https://doi.org/10.1098/rspb.2000.1397

Jaeger D, Pilger C, Hachmeister H, Oberländer E, Wördenweber R, Wichmann J, Mussgnug JH, Huser T, Kruse O (2016) Label-free *in vivo* analysis of intracellular lipid droplets in the oleaginous microalga *Monoraphidium neglectum* by coherent Raman scattering microscopy. Scientific Reports 6: 35340. https://doi.org/10.1038/srep35340

Karsten U, Holzinger A (2014) Green algae in alpine biological soil crust communities: acclimation strategies against ultraviolet radiation and dehydration. Biodiversity and Conservation 23(7): 1845–1858. https://doi.org/10.1007/s10531-014-0653-2

Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J (2010) The ITS2 Database III - sequences and structures for phylogeny. Nucleic Acids Research 38: D275–D279. https://doi.org/10.1093/nar/gkp966

Lang BF, Laforest M-J, Burger G (2007) Mitochondrial introns: a critical view. Trends in Genetics 23: 119–125. https://doi.org/10.1016/j.tig.2007.01.006

Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, De Clerck O (2012) Phylogeny and molecular evolution of the green algae. Critical Reviews in Plant Sciences 31(1): 1–46. https://doi.org/10.1080/07352689.2011.615705

Lewis LA, Lewis PO (2005) Unearthing the molecular phylodiversity of desert soil green algae (Chlorophyta). Systematic Biology 54(6): 936–947. https://doi.org/10.1080/10635150500354852

Lewis LA, Trainor FR (2012) Survival of *Protosiphon botryoides* (Chlorophyceae, Chlorophyta) from a Connecticut soil dried for 43 years. Phycologia 51(6): 662–665. https://doi.org/10.2216/11-108.1

Liu T, Chen Z, Xiao Y, Yuan M, Zhou C, Liu G, Fang J, Yang B (2022) Biochemical and morphological changes triggered by nitrogen stress in the oleaginous microalga *Chlorella vulgaris*. Microorganisms 10(3): 566. https://doi.org/10.3390/microorganisms10030566

Maciszewski K, Fells A, Karnkowska A (2022) Challenging the importance of plastid genome structure conservation: new insights from euglenophytes. Molecular Biology and Evolution 39(12): msac255. https://doi.org/10.1093/molbev/msac255

Marin B (2012) Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. Protist 163(5): 778–805. https://doi.org/10.1016/j.protis.2011.11.004

Mower JP, Ma PF, Grewe F, Taylor A, Michael TP, VanBuren R, Qiu YL (2019) Lycophyte plastid genomics: extreme variation in GC, gene and intron content and multiple inversions between a direct and inverted orientation of the rRNA repeat. New Phytologist 222(2): 1061–1075. https://doi.org/10.1111/nph.15650

Nzayisenga JC, Niemi C, Ferro L, Gorzsas A, Gentili FG, Funk C, Sellstedt A (2020) Screening suitability of northern hemisphere algal strains for heterotrophic cultivation and fatty acid methyl ester production. Molecules 25(9): 2107. https://doi.org/10.3390/molecules25092107

Peredo EL, Cardon ZG (2020) Shared up-regulation and contrasting down-regulation of gene expression distinguish desiccation-tolerant from intolerant green algae. Proceedings of the National Academy of Science 117(29): 17438–17445. https://doi.org/10.1073/pnas.1906904117

Pietrasiak N, Johansen JR, Drenovsky RE (2011) Geologic composition influences distribution of microbiotic crusts in the Mojave and Colorado Deserts at the regional scale. Soil Biology and Biochemistry 43: 967–974. https://doi.org/10.1016/j.soilbio.2011.01.012

Preisig HR, Melkonian M (1984) A light and electron microscopical study of the green flagellate *Spermatozopsis similis* spec. nova. Plant Systematics and Evolution 146: 57–74. https://doi.org/10.1007/BF00984054

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3(2): 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3): 539–542. https://doi.org/10.1093/sysbio/sys029

Saber AA, Fučíková K, McManus HA, Guella G, Cantonati M (2018) Novel green algal isolates from the Egyptian hyper-arid desert oases: a polyphasic approach with a description of *Pharao desertorum* gen. et sp. nov. (Chlorophyceae, Chlorophyta). Journal of Phycology 54(3): 342–357. https://doi.org/10.1111/jpy.12645

Shoup S, Lewis LA (2003) Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta). Journal of Phycology 39(4): 789–796. https://doi.org/10.1046/j.1529-8817.2003.03009.x

Šmarda P, Bureš P, Horová L, Leitsch IJ, Mucina L, Pacini E, Tichý L, Grulich V, Rotreklová O (2014) Ecological and evolutionary significance of genomic GC content in diversity in monocots. Proceedings of the National Academy of Science 111(39): E4096–E4102. https://doi.org/10.1073/pnas.1321152111

Smith DR, Hua J, Lee RW (2010) Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. Current Genetics 56(5): 427–438. https://doi.org/10.1007/s00294-010-0311-5

Van de Peer Y, De Rijk P, Wuyts J, Winkelmans T, De Wachter R (2000) The European small subunit ribosomal RNA database. Nucleic Acids Research 28(1): 175–176. https://doi.org/10.1093/nar/28.1.175

Van Wychen S, Ramirez K, Laurens LM (2015) Determination of total lipids as fatty acid methyl esters (FAME) by in situ transesterification, Laboratory Analytical Procedure (LAP). National Renewable Energy Laboratory (NREL), Golden, CO, USA. https://www.nrel.gov/docs/fy16osti/60958.pdf [accessed 16.08.2023]

Wang W, Lanfear R (2019) Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. Genome Biology and Evolution 11(12): 3372–3381. https://doi.org/10.1093/gbe/evz256

Watanabe S, Fučíková K, Lewis LA, Lewis PO (2016) Hiding in plain sight: *Koshicola spirodelophila* gen. et sp. nov. (Chaetopeltidales, Chlorophyceae), a novel green alga associated with the aquatic angiosperm *Spirodela polyrhiza*. American Journal of Botany 103(5): 865–875. https://doi.org/10.3732/ajb.1500481

Wiegert KE, Bennett MS, Triemer RE (2012) Evolution of the chloroplast genome in photosynthetic euglenoids: a comparison of *Eutreptia viridis* and *Euglena gracilis* (Euglenophyta). Protist 163: 832–843. https://doi.org/10.1016/j.protis.2012.01.002

Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T (2008) ProfDistS: (profile-) distance based phylogeny on sequence—structure alignments. Bioinformatics 24(20): 2401–2402. https://doi.org/10.1093/bioinformatics/btn453

Xu Z, Xin T, Bartels D, Li Y, Gu W, Yao H, Liu S, Yu H, Pu X, Zhou J, Xu J, Xi C, Lei H, Song J, Chen S (2018) Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. Molecular Plant 11(7): 983–994. https://doi.org/10.1016/j.molp.2018.05.003

Zhang Z, Green BR, Cavalier-Smith T (1999) Single gene circles in dinoflagellate chloroplast genomes. Nature 400: 155–159. https://doi.org/10.1038/22099

Zheng G, Tian B, Zhang F, Tao F, Li W (2011) Plant adaptation to frequent alterations between high and low temperatures: remodelling of membrane lipids and maintenance of unsaturation levels. Plant, Cell & Environment 34(9): 1431–1442. https://doi.org/10.1111/j.1365-3040.2011.02341.x

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31(13): 3406–3415. https://doi.org/10.1093/nar/gkg595

SUPPLEMENTARY MATERIALS

Supplementary material 1
Putative structure of the chloroplast genome of *Johansenicoccus eremophilus* gen. et sp. nov. The chloroplast genome may be present in the cells in one of two configurations, indistinguishable by our short-read data. The structure shown in this figure is a single circular molecule with direct repeats containing the block of rRNA genes. Two smaller circular molecules, each with one ribosomal block, are an equally plausible structure based on our data.
Link: https://doi.org/10.5091/plecevo.105762.suppl1

Supplementary material 2
Gene content and intron positions across chlorophycean plastomes. Asterisks (*) indicate that the gene is located in an inverted repeat, or in the case of WJT24VFNP31 in a direct repeat. Numbers refer to insertion positions of introns, if present. TR = trans-spliced intron, C = cis-spliced introns (if both cis- and trans-spliced are present in the same gene, otherwise cis is assumed). Blue colour of cells indicates the complete sequence of the gene is available; light blue indicates a partial sequence where uncertainty may exist about intron presence/absence. Grey colour indicates an incomplete genome assembly where the gene in question is missing but may be expected to be present based on its presence in related taxa. White cells indicate gene is absent from the completely assembled plastome.
Link: https://doi.org/10.5091/plecevo.105762.suppl2

Supplementary material 3
18S rRNA secondary structure of *Johansenicoccus eremophilus* gen. et sp. nov. (GenBank accession number OQ849776) with CBCs, HCBCs, and other base-pairing changes highlighted in colour, and substitutions and structural differences unique to *Johansenicoccus* highlighted with arrows, compared to *Chlamydomonas reinhardtii*.
Link: https://doi.org/10.5091/plecevo.105762.suppl3

Supplementary material 4
18S secondary structure of *Chlamydomonas reinhardtii* (GenBank accession number M32703) with highlighted differences (with colour and arrows) from *Johansenicoccus eremophilus* gen. et sp. nov.
Link: https://doi.org/10.5091/plecevo.105762.suppl4

Supplementary material 5
Results of the nuclear internal transcribed spacer 2 (ITS2) secondary structure modeling in Mfold and ITS2 Database. a: secondary structures of the ITS2 helices of *Johansenicoccus eremophilus* gen. et sp. nov. predicted by Mfold. b: secondary structure of *J. eremophilus* predicted by ITS2 Database using *Ankyra judayi* as template. c: *Ankyra judayi* ITS2 secondary structure. d: *Spermatozopsis exsultans* ITS2 secondary structure.
Link: https://doi.org/10.5091/plecevo.105762.suppl5

Supplementary material 6
Evolution of GC content in the 18S nuclear ribosomal gene (left) and chloroplast coding regions (right) mapped onto the chloroplast-derived phylogeny (based on data from 59 chloroplast genes), colour representation.
Link: https://doi.org/10.5091/plecevo.105762.suppl6

Supplementary material 7
Regression analysis of the relationship between temperature at the site of origin and GC content in the 18S gene across 59 species of green algae, corrected for phylogenetic relatedness using phylogenetically independent contrasts (PICs). Temperature refers to the average temperature in the warmest month at or near the site of the species/strain's collection. Dashed line represents the fitted regression line.
Link: https://doi.org/10.5091/plecevo.105762.suppl7

Supplementary material 8
Lipid composition in *Johansenicoccus eremophilus* gen. et sp. nov. Percentages were determined from three replicate extractions and analyses. Saturated and unsaturated fatty acids are marked; the "other" category was a mix of both. Bar graph on the bottom shows the same data but adds error bars (standard error of mean) to show variation among the three replicate extractions.
Link: https://doi.org/10.5091/plecevo.105762.suppl8