Received: 15 December 2021 | Accepted: 1 April 2022

DOI: 10.1002/ajb2.16015

RESEARCH ARTICLE

American Journal of Botany

Dynamic plastid and mitochondrial genomes in Chaetopeltidales (Chlorophyceae) and characterization of a new chlorophyte taxon

Tanner Robison[1,2] | Jessica M. Nelson[2] | Duncan A. Hauser[2] | Louise A. Lewis[3] | Fay-Wei Li[1,2]

[1]Plant Biology Section, Cornell University, Ithaca, NY, USA

[2]Boyce Thompson Institute, Ithaca, NY, USA

[3]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA

Correspondence

Fay-Wei Li, Plant Biology Section, Cornell University, Ithaca, NY, USA; Boyce Thompson Institute, Ithaca, NY 14853 USA.
Email: fl329@cornell.edu

Abstract

Premise: Chaetopeltidales is a poorly characterized order in the Chlorophyceae, with only two plastid and no mitochondrial genomes published. Here we describe a new taxon in Chaetopeltidales, *Gormaniella terricola* gen. et sp. nov. and characterize both of its organellar genomes.

Methods: *Gormaniella terricola* was inadvertently isolated from a surface-sterilized hornwort thallus. Light microscopy was used to characterize its vegetative morphology. Organellar genomes were assembled, annotated, and analyzed using a variety of software packages.

Results: The mitochondrial genome (66,927 bp) represents the first complete mitochondrial genome published for Chaetopeltidales. The chloroplast genome, measuring 428,981 bp, is one of the largest plastid genomes published to date and shares this large size and an incredible number of short, dispersed repeats with the other sequenced chloroplast genomes in Chaetopeltidales. Despite these shared features, the chloroplast genomes of Chaetopeltidales appear to be highly rearranged when compared to one another, with numerous inversions, translocations, and duplications, suggesting a particularly dynamic chloroplast genome. Both the chloroplast and mitochondrial genomes of *G. terricola* contain a number of mobile group I and group II introns, which appear to have invaded separately. Three of the introns within the mitochondrial genome encode homing endonucleases that are phylogenetically nested within those found in fungi, rather than algae, suggesting a possible case of horizontal gene transfer.

Conclusions: These results help to shed light on a poorly understood group of algae and their unusual organellar genomes, raising additional questions about the unique patterns of genome evolution within Chaetopeltidales.

KEYWORDS

algae, chloroplast genome, Gormaniella, intron, mitogenome, repeats

Our understanding of the evolution and diversity of Viridiplantae (i.e., land plants and green algae) has been greatly advanced by the sequencing of organellar genomes (Leliaert et al., 2012; Ruhfel et al., 2014; Gitzendanner et al., 2018; Bell et al., 2020; Sousa et al., 2020). The use of organellar genomes to resolve relationships among green algae has been particularly useful because some groups are difficult to separate based solely on morphological features (Leliaert et al., 2012; Fučíková et al., 2014; McManus et al., 2018). In addition to helping resolve phylogenetic relationships, the increased availability of chloroplast genomes has revealed striking patterns in the structural evolution across Viridiplantae. In land plants, the overall chloroplast genome structure (quadripartite organization, presence of an inverted repeat, resistance to new genes) has been largely conserved, and there is a high degree of conservation in gene order and content across land plants (Palmer, 1985; Mower and Vickrey, 2018). In contrast, while the chloroplast genomes of green algae have a similar structural organization, others more closely resemble the structure of mitochondrial genomes, which are more flexible and highly rearranged (Turmel and Lemieux, 2018).

Despite recent advancements in characterizing green algal diversity, studies exploring previously unsampled habitats have revealed that our current understanding remains incomplete (Lewis and Lewis, 2005; De Wever et al., 2009; Schmidt et al., 2011; Caisová, 2016; Watanabe

et al., 2016). Chaetopeltidales is a member of the strongly supported OCC clade (Oedogoniales, Chaetophorales, and Chaetopeltdidales), which comprises three of the five orders in Chlorophyceae (Brouard et al., 2010). Chaetopeltidales is an ecologically diverse order, with members found in terrestrial and freshwater habitats, or living epiphytically on plants and other algae (Sanchez-Puerta et al., 2006; Caisová, 2016; Watanabe et al., 2016). Among Chlorophyceae, Chaetopeltidales is one of the most obscure, understudied orders, with only two published chloroplast genomes and no published mitochondrial genomes (Brouard et al., 2010; Watanabe et al., 2016). These two chloroplast genomes are remarkable in their structural divergence. At 521 kb, *Floydiella terrestris* boasts the second largest sequenced chlorophycean chloroplast genome, while the chloroplast genome of *Koshicola spirodelphila* is fragmented into three smaller "minicircles" totaling 385k bp. The large size of these chloroplast genomes is made even more impressive by the fact that neither of them have inverted repeats.

In addition to their remarkable size and structure, the organellar genomes of Chaetopeltidales and the rest of the OCC clade are characterized by an abundance of introns; group II introns are the most prevalent (Turmel and Lemieux, 2018). Chloroplast genomes of members of the OCC clade also share a distinguishing feature in which three genes, *petD*, *psaC*, and *rbcL*, contain *trans*-splicing group II introns (Brouard et al., 2010). In some algal lineages, group I and group II introns have been shown to diversify and colonize new intragenomic locations through a combination of retrohoming and mutations to the exon binding sites (Brouard et al., 2016). Group I and group II introns are often highly variable in their distribution due to their mobility, although group I introns appear to be more stable at a given site (Turmel and Lemieux, 2018). Both groups also frequently contain sequences coding for putative homing endonucleases, which aid in the mobility of these introns (Lambowitz and Belfort, 2015).

Here we describe a novel green alga in Chaetopeltidales (Chlorophyceae), *Gormaniella terricola* gen. et sp. nov., isolated from a surface-sterilized hornwort thallus. We examined the morphology and conducted a phylogenetic analysis to support the phylogenetic placement of this new species. We also studied the structure and contents of both the chloroplast and mitochondrial genomes of this alga to determine whether they share some of the extreme features of related organellar genomes. These investigations revealed a notably massive chloroplast genome and distinctions from the other sequenced organelles of Chaetopeltidales. We also investigated the content and origins of the introns of this alga and the proliferation of repetitive sequences in intergenic regions.

MATERIALS AND METHODS

Isolation and culture

As part of our ongoing effort to isolate symbiotic cyanobacteria from hornworts (Nelson et al., 2019, 2021), we inadvertently obtained an unknown chlorophyte species. The hornwort, *Notothylas orbicularis*, was collected from a roadside drainage trench at Potato Hill State Forest (Tompkins County, New York; 42°20′37.8″N 76°16′09.1″W, 488 m a.s.l.; Appendix S1). Surface sterilization and initial culturing were done as described by Nelson et al. (2019).

Microscopy

Cultures were grown in Bold's Basal Medium broth or agar and examined using an Olympus BX60 light microscope (Olympus Corp, Tokyo, Japan) and bright field or differential interference contrast (DIC) optics. In some cases, India ink was added to the preparation to examine the extent of the mucilage layers.

DNA extraction and sequencing

DNA was extracted using the cetyl trimethylammonium bromide (CTAB) protocol as described previously (Nelson et al., 2019). Illumina library was prepared using SparQ DNA Frag & Library Prep kit and Adapter Barcode Set A (Quantabio, Beverly, MA, USA) following the manufacturer's protocol. The resulting library was pooled with 11 other samples and sequenced on an Illumina NexSeq500 mid-output flow cell (150-bp paired-end) at the Cornell Institute of Biotechnology.

Assembly and annotation

Both the mitochondrial and chloroplast genomes were assembled de novo using NOVOPlasty v2.7.2 (Dierckxsens et al., 2017). The entire mitochondrial genome of *Bracteacoccus aerius* was used as the seed sequence for mitochondrial genome assembly since there were no available genomes of closely related algae. For the chloroplast genome, *rbcL* from *Floydiella terrestris* was used as the seed sequences. In both assemblies, a *k*-mer of 33 was used and organellar genomes assembled into single, circular mapping contigs. Read mapping using bwa version 0.7.17 (Li and Durbin, 2009) was then performed, showing an average read depth of 2843. Genes were annotated in the chloroplast genome using the Transfer annotations tool in Geneious Prime (Biomatters, Auckland, New Zealand), where genes were predicted based on amino acid sequence similarity to *Floydiella terrestris*. Novel open reading frames (ORFs) were identified in Geneious Prime using the Find ORFs tool. Translation of ORFs >800 bp were queried using blastp. RNAweasel (Lang et al., 2007) was used to annotate rRNA and tRNA genes in both the chloroplast and mitochondrial genomes. Mitochondrial genes were annotated in a similar manner, but due to significant divergence in mitochondrial sequences, mitochondrial genomes from *Bulbochaete rectangularis* and *Stigeoclonium helveticum* (GenBank accessions MN810331 and MN810332, respectively) were used to

identify coding sequences. In some cases, especially in the mitochondrial genome, the annotated gene boundaries needed to be further refined manually. Finally, introns were annotated and classified using MFannot and RNAweasle (Lang et al., 2007). Organellar genome maps were generated using OGDRAW version 1.3.1 (Greiner et al., 2019).

Comparative genomics

Synteny analysis was performed manually, but MCScan (Tang et al., 2008) was used as a plotting tool, where each annotated canonical chloroplast gene in *F. terrestris*, *K. spirodelphila*, and *Gormaniella terricola* was treated as an ortholog. Syntenic gene blocks were then colored in Inkscape v0.92 to more easily visualize synteny across the three genomes. Repeat analysis was performed using Vmatch (http://www.vmatch.de/), searching for repeats greater than 30 bp long. To avoid overreporting of repeats, the supermax function was used to report the largest repeat not contained within another repeat. Then to infer the similarities among repeats in the genome, the sequences were clustered if they had greater than 80% sequence similarity as determined by the matchcluster function. These repeats were then visualized using Circos v0.69-8 (Krzywinski et al., 2009), mapping each repeat as a "link". Additionally, consensus sequences for the 20 most-abundant repeat clusters were generated using the Bioconductor package msa (Bodenhofer et al., 2015).

Alignment and phylogenetics

The nuclear 18S rDNA gene (1715 bp) was sequenced, and the alignment was largely based on the data matrix from Watanabe et al. (2016) in addition to a selection of appropriate sequences that were published later. These sequences were aligned with the additional sequence from *G. terricola* using PASTA 1.8.5 (Mirarab et al., 2015; Balaban et al., 2019), and phylogenetic trees were inferred using the default settings and the model selection function in IQ-TREE v2.0.3 (Minh et al., 2020) with 1000 bootstrap replicates. The selection of chloroplast genes for phylogenetic analysis (*rbcL*, *atpB*, and *psaB*) matched with those analyzed by Watanabe et al. (2016). The optimal partitioning schemes and nucleotide substitution models were found using PartitionFinder, partitioning for each codon position in each gene (Lanfear et al., 2017). Phylogenetic analysis was performed using IQ-TREE with 1000 non-parametric bootstrap replicates.

Retrieval and analysis of LAGLIDADG homing endonucleases

To retrieve sequences related to homing endonucleases found in the organellar genomes of *G. terricola*, we made a custom script in R in which the blastx hits were filtered to only retrieve the highest quality hits from each of the LAGLIDADG homing endonucleases. These sequences were then aligned using PASTA 1.8.5 (Mirarab et al., 2015; Balaban et al., 2019), and phylogenetic trees were inferred using IQ-TREE v2.0.3 with Ultrafast bootstrapping (UFBoot) for 1000 runs and 1000 SH-aLRT tests (Minh et al., 2020). Trees were then visualized using GGTREE v 3.0.2 (Yu et al., 2017).

RESULTS

Morphology of *Gormaniella terricola*

The algal cells are solitary or form packets of 1–8 cells, without producing filaments (Figure 1). Individual cells are 8–14 μm in diameter, depending on whether they are newly divided or are older. Each cell is uninucleate and contains a parietal chloroplast with one massive pyrenoid 2–2.5 μm in diameter and surrounded by a starch sheath; the pyrenoids are traversed. Cells that are in the process of mitosis contain two pyrenoids. Cell walls of individual cells are relatively thin. Cells in packets of two or four cells have their own wall and common walls shared by most recently divided cells. In older cultures, the packets can have thicker accumulations of wall material, but they never are as extensive, or branched, as is seen in older cultures of *Hormotilopsis*. Motile cells with four flagella were observed.

Phylogenetic analysis

Our 18S rDNA phylogenetic analysis placed *G. terricola* firmly in Chaetopeltidales and specifically grouped it with the 18S genes of *F. terrestris* (two accessions) and *Hormotilopsis gelatinosa*, although the relative arrangement of these four sequences was unresolved (Appendix S2) due to the low rate of nucleotide substitutions in 18S. A second analysis based on three chloroplast genes, *rbcL*, *atpB*, and *psaB*, corroborated the placement of *G. terricola* in Chaetopeltidales (Figure 2). These data also resolved *G. terricola* to be clearly distinct from but sister to *F. terrestris* and *H. gelatinosa*, providing support for designating it as a new genus.

Chloroplast genome

The chloroplast genome of *G. terricola* was assembled into one circular mapping chromosome with a length of 428,930 bp, making it one of the largest chloroplast genomes sequenced to date (Appendix S3). The chloroplast genome of *G. terricola* has a GC content of 34.0% and includes 106 annotated genes, 74 of which are protein coding, 28 code for tRNA, and three for rRNA (Appendix S4). *Gormaniella terricola* has nearly the same tRNA content to *F. terrestris* with the exception of an extra copy of *trnW* in *G. terricola*. The GC content of *G. terricola* (34%) is also very similar to *F. terrestris* (34.5%). Intergenic regions make up the vast majority of the total length of the genome

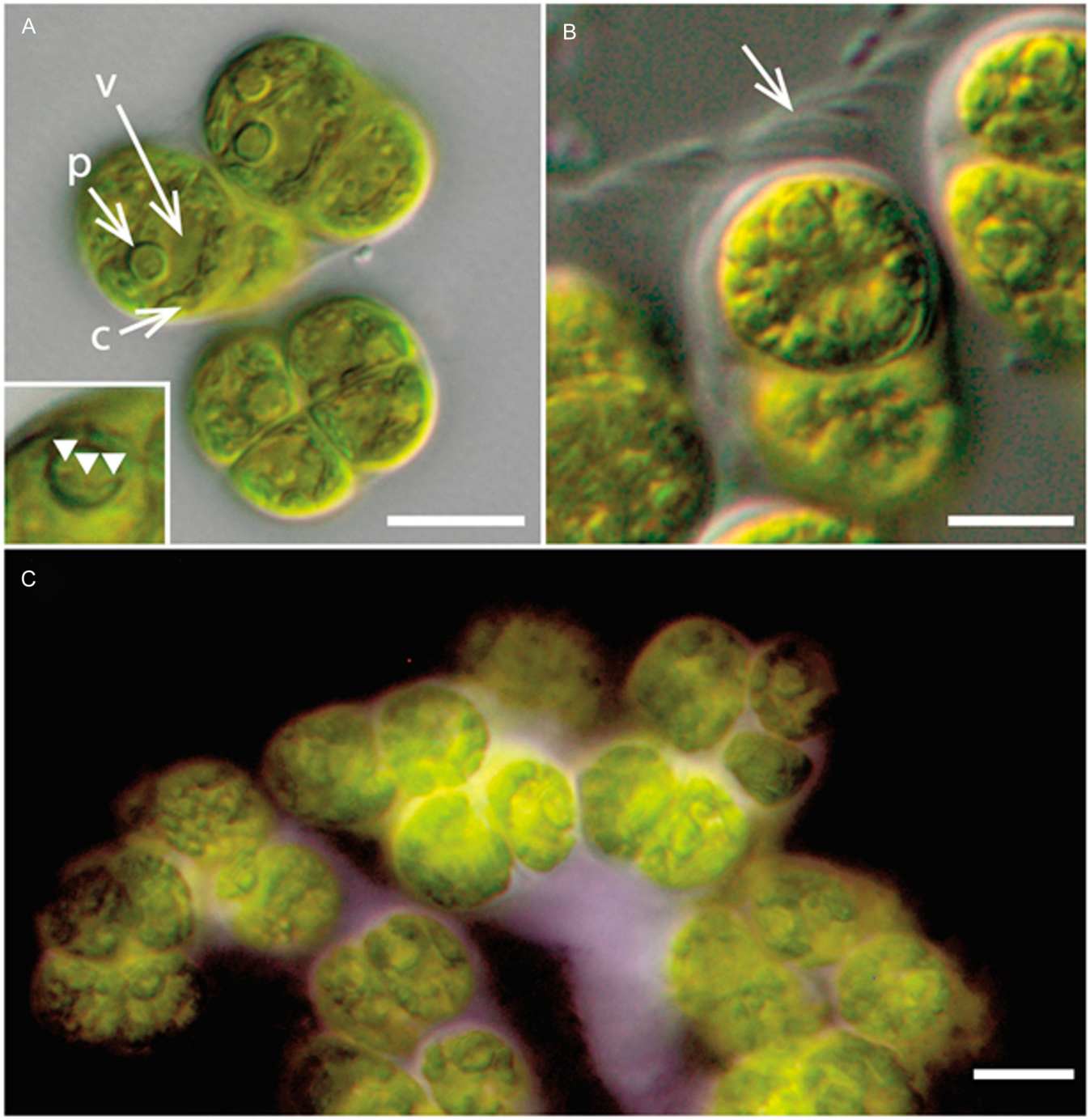

FIGURE 1 Light micrographs of vegetative morphology of *Gormaniella terricola* using differential interference contrast (DIC) optics. (A) Typical 4-celled packets seen in actively growing cultures, with each cell having a parietal chloroplast (c), prominent pyrenoid (p), and large central vacuole (v). Inset shows a pyrenoid of a different cell that is traversed by a channel (arrowheads). (B) In slightly older cultures, the cell packets accumulate a layered cell matrix (arrow). (C) Older cultures stained with India ink, revealing an abundant matrix (clear) around the mass of packets. Scale bar = 10 μm.

(302,333 bp), while protein-coding sequences and introns make up the bulk of the remaining genome (94,090 and 28,551 bp, respectively). The coding gene content of *G. terricola* was also similar to *F. terrestris*, except for a few annotated ORFs. Two ORFs were absent in *G. terricola* (*orf150* and *orf431*), two appear to have been pseudogenized (orf183 [552 bp] and *orf120* [376 bp]), and three large ORFs (>1300 bp) were newly identified in a ~30-kb intergenic region between *rps9* and *trnE*. Blastp searches of these new ORFs against the NCBI database did not reveal any significant similarities to known genes.

When compared to the chloroplast genome of *F. terrestris*, the genome of *G. terricola* has many blocks of conserved gene order (Figure 3, colored ribbons), but most of the similarities break down when compared to *K. spirodelophila*. The one

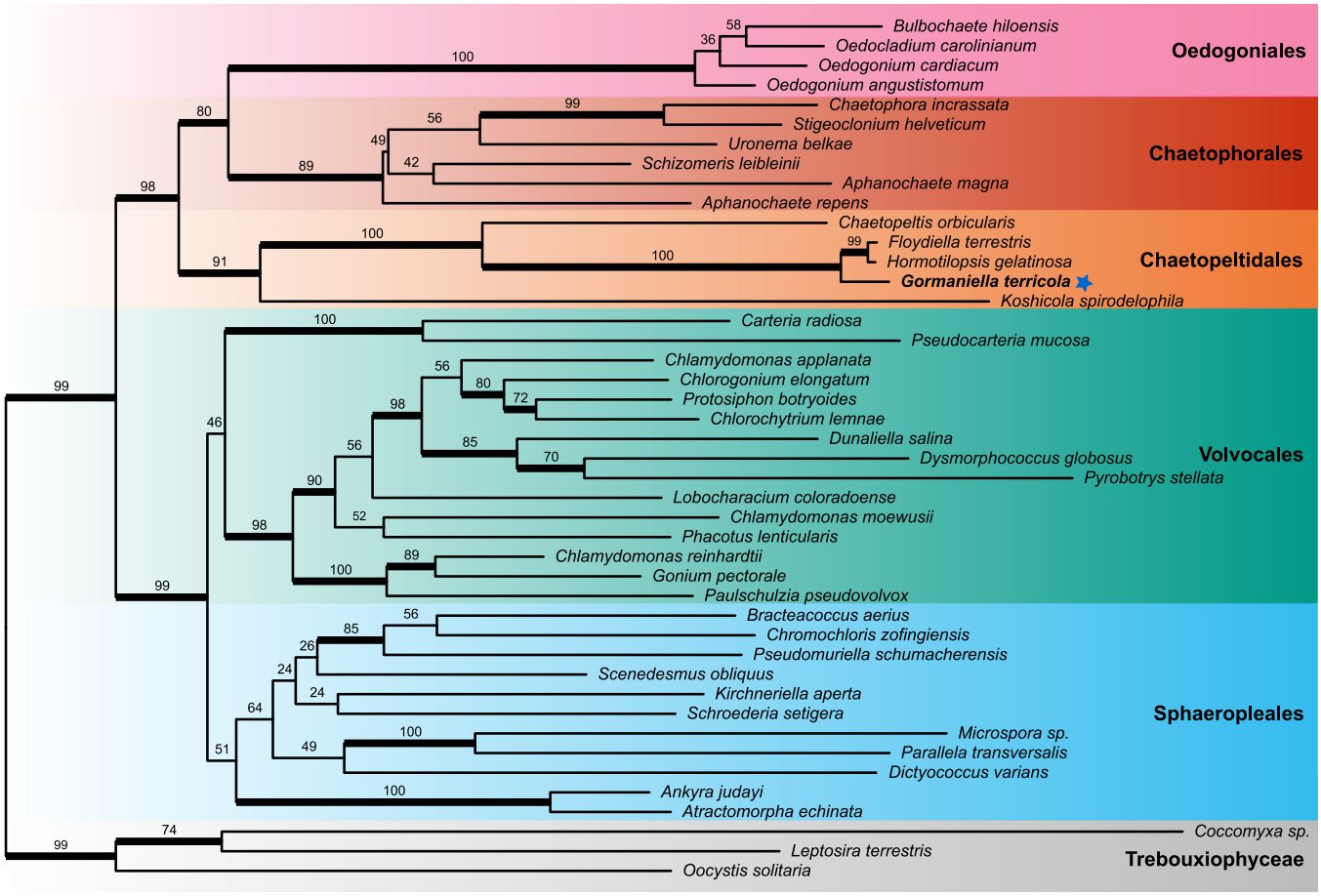

FIGURE 2 Phylogenetic position of *Gormaniella terricola*. The phylogeny was inferred based on three chloroplast loci (*rbcL*, *atpB*, and *psaB*) using maximum likelihood. The blue star marks the position of *G. terricola*. The numbers associated with branches are bootstrap support values and thickened branch thickenings indicate branches with ≥70 nonparametric bootstrap support. The tree was rooted with Trebouxiophyceae.

exception is the *psbB*, *psbT*, *psbH* block, which appears to be conserved across all three members of Chaetopeltidales and is similar to the conserved *psbB*, *psbT*, *psbN*, *psbH* gene block found in land plants. Like the other members of Chaetopeltidales, the chloroplast genome of *G. terricola* also lacks inverted repeats, although it does contain a suite of related short repeats (>30 bp) that are distributed in the intergenic regions (Figure 4A). The supermaximal repeats found in the chloroplast genome grouped into clusters based on sequence similarity, and the 20 largest clusters are reported in Table 1. These repeats are diverse and numerous, with 134 unique clusters (repeats with a frequency of two or more) in the chloroplast genome having greater than 80% sequence similarity and being longer than 30 bp, summing to 65,269 bp. These repeats often contain small palindromic sequences that, based on secondary structure prediction, may form short hairpin/cruciform structures.

Mitochondrial genome

The mitochondrial genome of *G. terricola* was also a circular mapping chromosome with a length of 64,781 bp and a GC content of 38% (Appendix S4). The total gene content of the mitochondrial genome is highly reduced compared to other chlorophyte algae, containing only 13 annotated protein-coding genes, all of which are essential to respiration. The mitogenome is less gene dense than the other members of the OCC clade, and has expanded intergenic regions (Turmel et al., 2020). The ribosomal RNA genes are highly fragmented, similar to what has been reported in sphaeroplealean algae (Fučíková et al., 2014), making it difficult to fully characterize these sequences. The mitochondrial genome also contains short repeats, but they are not nearly so numerous as in the chloroplast genome, nor can they be clustered into large groups based on similarity (Figure 4B).

Intron content

The chloroplast genome of *G. terricola* contained 29 introns, 17 of which are group I, 12 are group II. While this number is similar to the 26 introns found in *F. terrestris*, the location and origin of these introns is often different. Of the 29 introns, 19 share high sequence similarity and an insertion site with *F. terrestris*, with the other 10 appearing to be novel. Of these 10 novel introns, six are located in the *psbA* coding sequence, one is located in the *rrL* gene, and the remaining three are found in

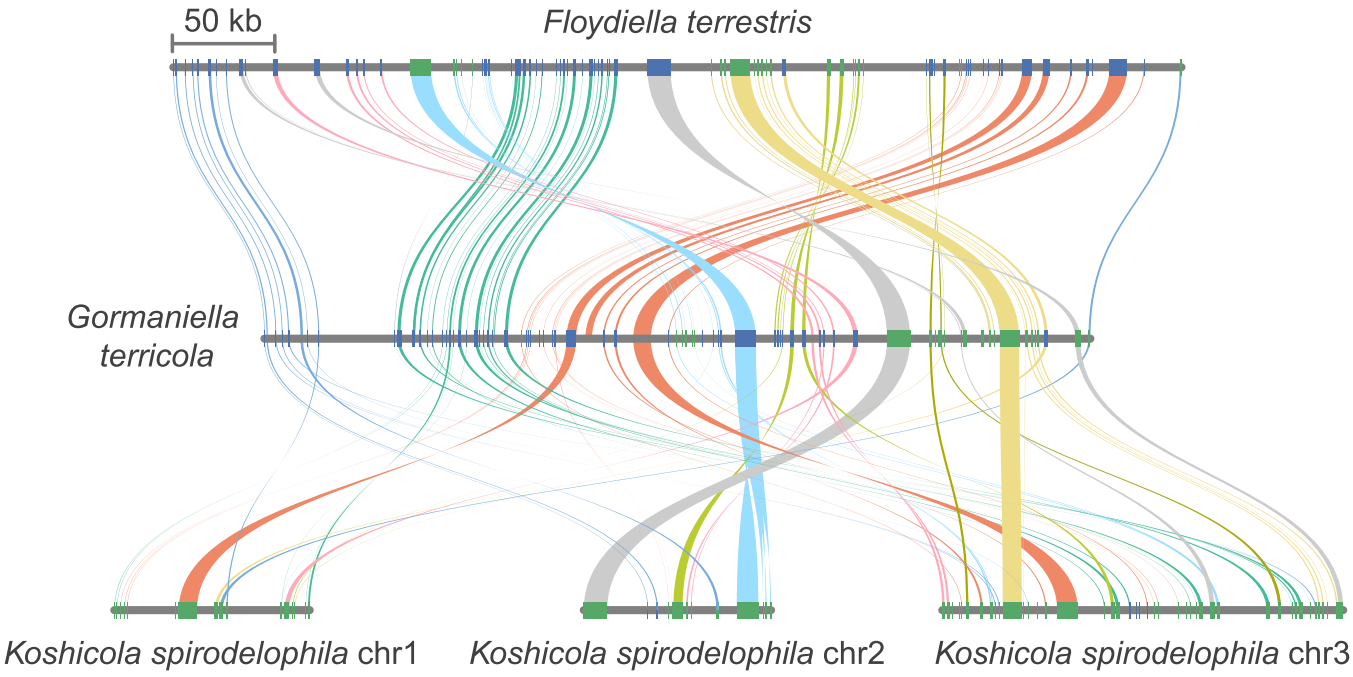

FIGURE 3 Comparison of gene order in Chaetopeltiales chloroplast genomes. Each ribbon on the diagram tracks the relative position of a single gene within a given chloroplast genome. Regions where gene order is conserved between *Floydiella terrestris* and *Gormaniella terricola* are colored to highlight synteny. Color of gene indicates direction of the gene (blue = forward, green = reverse).

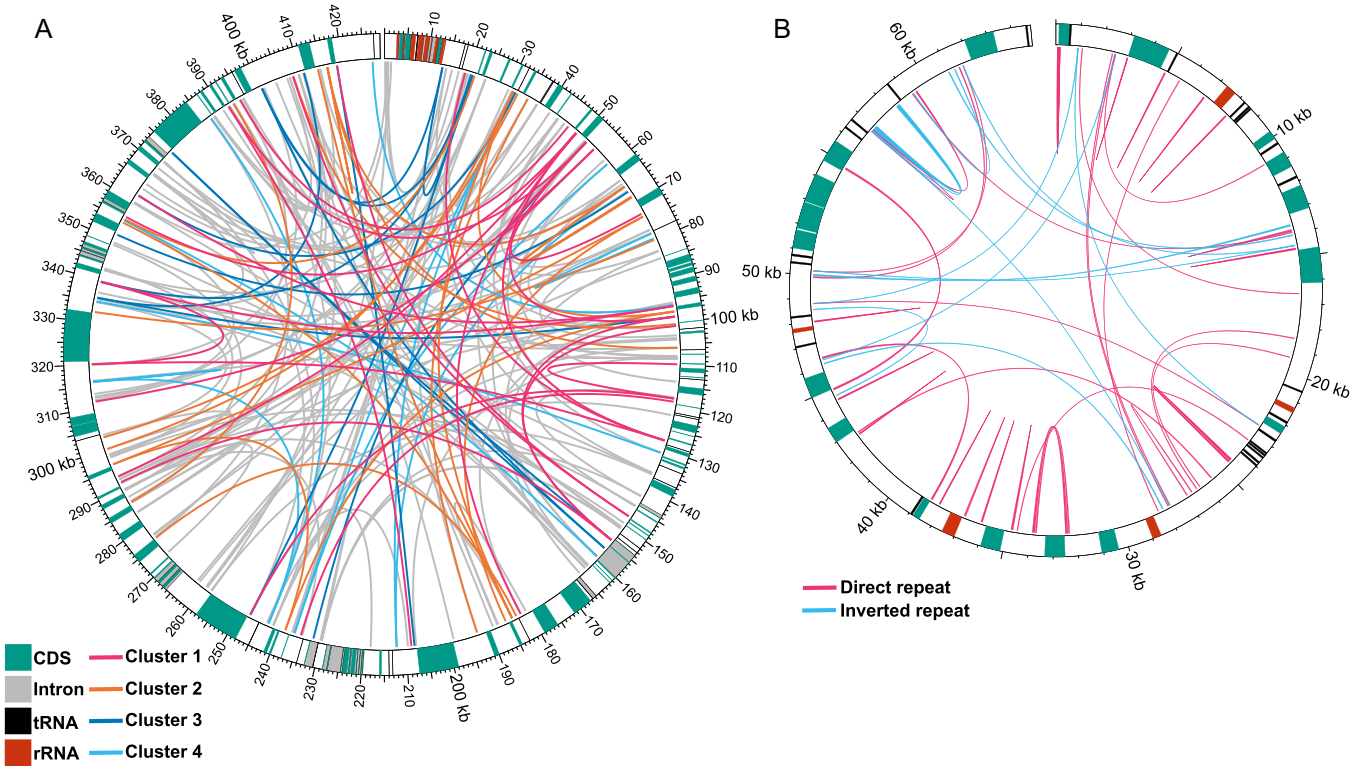

FIGURE 4 Graphical representation of repeat distributions in the (A) chloroplast and (B) mitochondria genomes of *Gormaniella terricola*. (A) Links indicate the positions of related repeat clusters, with color indicating the four most-abundant repeats clusters and the remaining all colored in gray. Colored blocks show the positions of genes. (B) Colored links indicate the position of either inverted (cyan) or direct repeats (pink).

TABLE 1 Summary of common repeat clusters found in the chloroplast genome. Uppercase letters have at least 80% consensus; lowercase letters indicate consensus of greater than 50%, but less than 80%. Dots (.) indicate less than 50%. Underlined sequences are possible short hairpin structures.

Copy no.	Consensus sequence	GC (%)	Length (bp)
24	agAAAAGGGCTTTTGCCTGAGCg.aGCTCAGGTGCAACCCGAACC	58	45
21	aaaccgAACGGGGTTTTTAAAAAGGCTGCGccct	50	36
16	ggGTGcAACCCGAACCTGAACGGgGTTCAGGCC	67	33
13	taGCCTGAACCCCGTTCAGGTTCGGGTTgCACCc	62	34
11	AAGGGCTAAAATT.gCTTCTAGAGCtCTTTTTT	36	33
10	GgGCGCAGCCGGAACCCTGTTCCGGTTTtt	61	31
10	ggttcGGGTTgCACCTGAGCtccGCTCAGGCAAAAGCCCTTTTCTttt	54	48
9	aaaagAAAAGGGCTTTTGCCtGAGCTCCGCTTTTTAAAaAGG.TGCAACCCgaa	43	54
9	tttt..aaagGGTTGCACCCTCTTTAAAAAGCgGAGCTCGGGCAAAAGCcc	49	51
8	CCCg.CCAAAAAAga.AAAGGGCTTTTGCC	48	32
8	atTTTTATGGGCTcTTTTcttATTaAAAGgaggg	32	34
8	ctcTAGAAG..AA.TTtAGCCCTT.CTGAGGGTTTTTtt	40	41
7	aAAAAAAAGGATTTTTATGGGCTcTTTTcTt	26	31
7	AACCCGAACgaaGTTCGGGCAGGGGCCCTC	67	31
7	GCCCTTatTAAAAAGCTCCGCCCAAAAAga	43	31
6	aAAAAAAGAAGAAGGCGGAGCTTTTAa.AAGg	32	35
6	AAAACCCgCTTCTTAAAAAAAA.TGGAAGg	33	34
6	t.AAAACCGGAACaGGGTTCCGGCTGCGCc	62	31
6	t.AAAGGAGGGTA.aAACCGGAACGGGGTTtt	47	34
6	aagGCCCGAACGGGGTTTTTTCAAAGGGTTGCACC	54	37

the coding sequences of *rbcL*, *psaA*, and *psbC*. Five of the introns in this chloroplast genome have ORFs encoding putative homing endonucleases, four of which are group I introns and have either HNH or LAGLIDADG homing endonucleases, while the group II intron within the *psaA* gene has an internal ORF which codes for an HNH homing endonuclease.

The mitochondrial genome contains seven introns, all of which are classified as group I. The locations of these introns are conserved when compared to the other members of the OCC clade, although the intron sequences bear little sequence similarity. Three of the seven introns have internal ORFs which encode for LAGLIDADG homing endonucleases.

Horizontal gene transfers of homing endonucleases

In addition to the aforementioned ORFs located within introns, there are additional freestanding ORFs in these organellar genomes. These ORFs code for putative LAGLIDADG homing endonucleases in the mitochondrial genome, and a freestanding ORF in the chloroplast genome has a high sequence similarity to GIY-YIG homing endonucleases. Among all the putative LAGLIDADG homing endonucleases, two (intron 3 of *cox1* and one of the freestanding ORFs) have the greatest sequence similarity to homing endonucleases found in fungi, rather than algae. Given the strong fungal signal observed in the blastp results of the homing endonucleases in the mitochondrial genome (Appendix S5), a phylogenetic analysis of the top 20 blastp hits from every LAGLIDADG homing endonuclease encoded in both the mitochondrial and chloroplast genomes was conducted. This analysis revealed not only that some of the mitochondrial endonucleases are indeed of fungal origin, but that the chloroplast encoded endonuclease is likely the result of horizontal gene transfer from bacteria (Figure 5). These results also confirm that such interspecies and interkingdom transfer of mobile introns has occurred multiple times throughout the algal lineage.

Taxonomic treatment: *Gormaniella* Lewis, Robison, Fay-Wei Li gen. nov.

Diagnosis

Green microalgae from terrestrial habitat. Vegetative cells are 8–14 μm in diameter, uninucleate and walled. Cells contain a

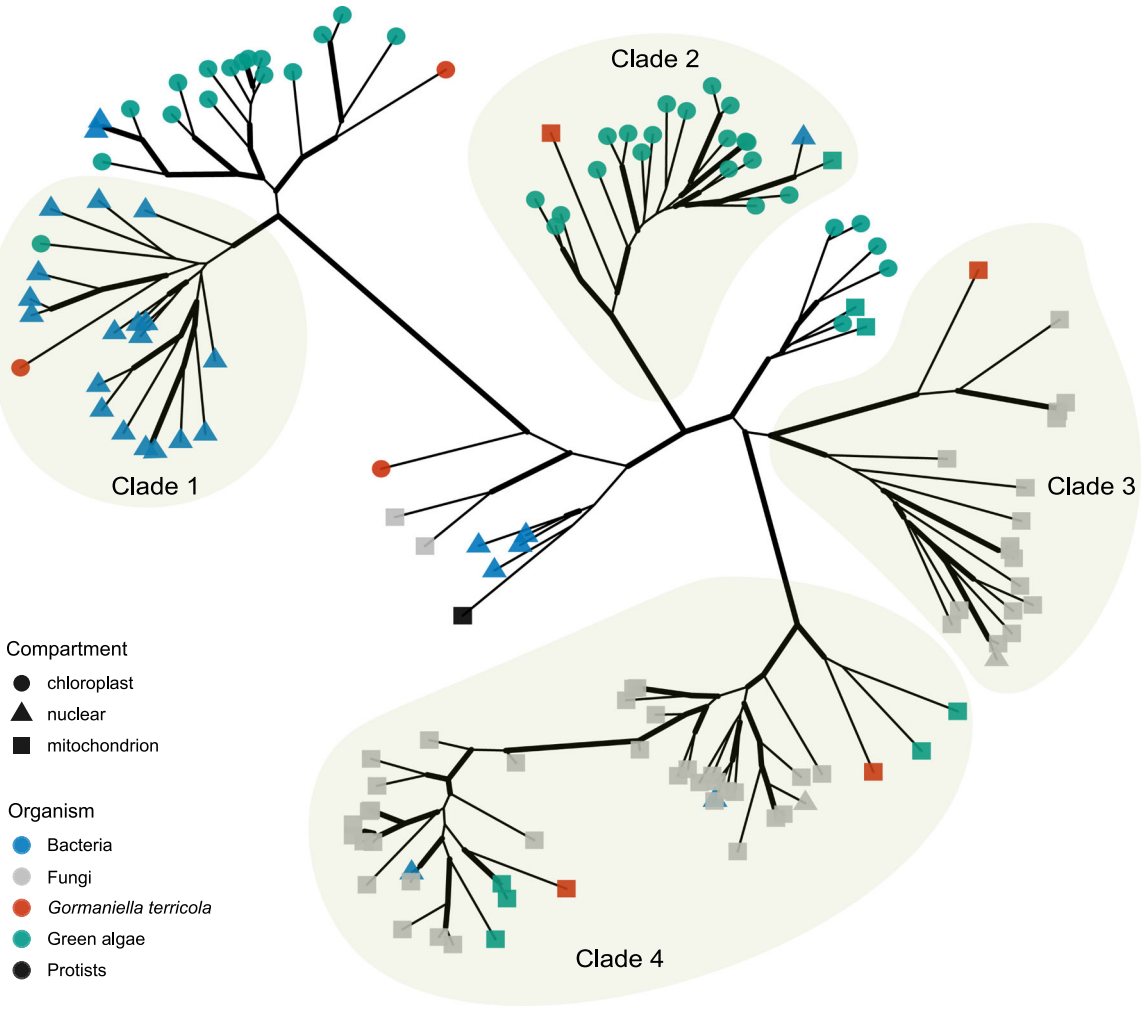

FIGURE 5 Phylogenetic relationships of blast hits to LAGLIDAD homing endonucleases encoded in the organellar genomes of *Gormaniella terricola*. Branch thickenings indicate branches with ≥95 UFBoot support and ≥70 SH-aLRT support. Tips are colored based on the broad phylogenetic grouping from which the sequences originated. Tip shapes indicate the genomic compartment from which they originate. Highlighted clades are those which have putative HGT events.

parietal chloroplast with 1 large pyrenoid with a traversing channel. Cells can be solitary or form packets of 2 or 4 cells due to successive bipartitioning, and the packets may accumulate into three-dimensional masses. Cell walls are thin in young cells, but thicken as cells age. Packets of cells in older cultures also accumulate an extensive, laminated, extracellular matrix. Distinct from *Floydiella* in forming an extensive extracellular matrix. Similar to the early life stages of *Hormotilopsis*, *Koshicola*, *Oncosaccus*, *Chaetopeltis*, and *Pseuduvella*, but lacking the extensive extracellular matrix stalks, the weak filaments, or blade- or disc-shaped thalli of those genera.

Etymology

The generic name *Gormaniella* is the first declension feminine Latin noun, to honor Amanda Gorman, 2017 USA National Youth Poet Laureate, and author of the poem *The Hill We Climb*.

Type species: *Gormaniella terricola* Lewis, Robison, Fay-Wei Li sp. nov.

Here designated as *Gormaniella terricola*. With characteristics of the genus, isolated from the thallus of the hornwort *Notothylas orbicularis*.

Etymology of the epithet

Nominative feminine singular of terricolus, earth-dwelling (terrestrial).

Type locality

Potato Hill State Forest, Tompkins County, New York, USA (42°20′37.8″N 76°16′09.1″W, elevation 488 m a.s.l.), 28 August 2018.

Holotype

Cells from the authentic strain fixed, glycerin mounted on a permanent slide, are deposited in the George Safford Torrey Herbarium, University of Connecticut, Storrs, CT, USA, CONN00228027.

DISCUSSION

A new taxon in Chaetopeltidales

The green algal order Chaetopeltidales is relatively small, containing seven genera and nine species (algaebase.org, 30 November 2021), two of which were described in the last decade (Caisová, 2016; Watanabe et al., 2016). Most species are known from freshwater habitats or are found in soil or associated with plants. Here we describe a new species, in a new genus, that was inadvertently isolated from the thallus of a hornwort and is characterized using morphology and molecular phylogenetics. Morphologically, *Gormaniella terricola* resembles members of *Hormotilopsis* and *Floydiella*, by formation of unicells or packets of 2–4 cells. However, *Floydiella* lacks the well-developed extracellular matrix seen in *Gormaniella*, and *Hormotilopsis* forms mucilaginous, often branched, cylindrical stalks, a trait not observed in *Gormaniella*. The vegetative morphology of *Gormaniella* also resembles an early stage in the development of *Chaetopeltis*, *Pseudulvella*, and *Oncosaccus*, which ultimately produce multicellular, flattened disc-like or blade-like thalli. The vegetative morphology of *Gormaniella* also is similar to the small packets of *Koshicola,* but *Gormaniella* does not form the weak filaments shown for this genus. *Gormaniella* does not share the morphology characteristic of *Dichranochaete*, which is unicellular and forms a distinct spiny cap and long setae (Caisová, 2016). Our findings echo those of Watanabe et al. (2016), that the relatively simple vegetative morphology characteristic of *Hormotilopsis*, especially the 2–4-celled packets, is repeatedly seen among phylogenetically distinct taxa in Chaetopeltidales.

Analysis of 18S rDNA is very useful for placing green algae into major lineages, and even into less inclusive groups such as orders and families, but it can have less utility for species level identification as the 18S rDNA sequence can be identical or nearly so among closely related species (e.g., Lewis and Flechtner, 2002; Hall et al., 2010; Yang et al., 2021). The results of the 18S analysis presented here (Appendix S2) supports *Gormaniella* as a member of the Chaetopeltidales, with close affinity to *Hormotilopsis gelatinosa* (the type species) and *Floydiella* (a monotypic genus), but distinct from *H. tetravacuolaris*, *Chaetopeltis*, *Pseudulvella*, *Koshicola*, and *Dichranochaete*. The phylogenetic analysis of three chloroplast genes (Figure 2) supports separation of *Gormaniella* from *Floydiella* and *Hormotilopsis gelatinosa* as well as from *H. tetravacuolaris*. Clearly, a taxonomic re-evaluation of the generic limits of *Hormotilopsis* is needed (as also demonstrated by Sanchez-Puerta et al., 2006, Watanabe et al., 2016), but such work is outside the scope of this study.

Evolution of chloroplast genome gigantism in Chaetopeltidales

While not as large as the chloroplast genome of *Floydiella terrestris* nor so unconventional as the three minicircle chromosomes found in *Koshicola spirodelphila*, the chloroplast genome of *Gormaniella terricola* nonetheless adds to the narrative of structural dynamism found in chloroplast genomes of Chatopelidales. In fact, the three members of Chaetopeltidales are among the top 10 largest chloroplast genomes published to date (Appendix S3), suggesting that large chloroplast genomes could be characteristic of the clade. While the size of the chloroplast genome of *G. terricola* cannot be totally explained by a single factor, two features appear to have the greatest effect.

The first and most significant is the expansion of intergenic regions, which makes up ~70% of the total length of the chloroplast genome, a figure comparable to the other members of Chaetopeltidales (*F. terrestris*, *K. spirodelphila*; 76% and 63%, respectively). In stark contrast, *Schizomeris leibleinii* and *Oedogonium cardiacum* (representatives from Chaetophorales and Oeodogoniales of the OCC clade) have chloroplast genomes that are much more gene dense, with intergenic regions comprising only 28% and 15% of their genomes, respectively (Figure 6). This intergenic expansion is likely the result of a proliferation of short, dispersed repeats, which appear to be a common theme across Chaetopeltidales, but not seen in the other orders in the OCC clade (Figure 6). These short, dispersed repeats also suggest a mechanism explaining the high levels of gene shuffling seen in Chaetopeltidales and suggest a possible mechanism by which the chloroplast genome of *K. spirodelphila* may have split into three "minicircles". In the chloroplast genome of *G. terricola*, detectable copies of these repeats account for 20% of the total size of the intergenic regions, although this number is, admittedly, conservative. When Vmatch searches allowing for shorter repeats (>20 bp) were conducted, as much as 75% of the intergenic region can be accounted for by repeats. These repeats often contain small palindromic sequences, similar to what is reported in chlamydomonadalean algae and *F. terrestris* (Smith, 2020a; Brouard et al., 2010) and may be folded into small stem loop structures like those hypothesized by Kelchner and Wendel (1996). Such hairpins are thought to drive small inversions in the hairpin loop and may be responsible for larger genome arrangements as well (e.g., Kelchner, 2000). In chlamydomonadalean algae, these repeats appear to be mutational hotspots and are nearly ubiquitous in the largest chloroplast genomes, but entirely absent in closely related, smaller chloroplast genomes. If such palindromic repeats are indeed the main drivers of the massive chloroplast genomes in Chaetopeltidales, then these repeats suggest a striking example of convergent genome

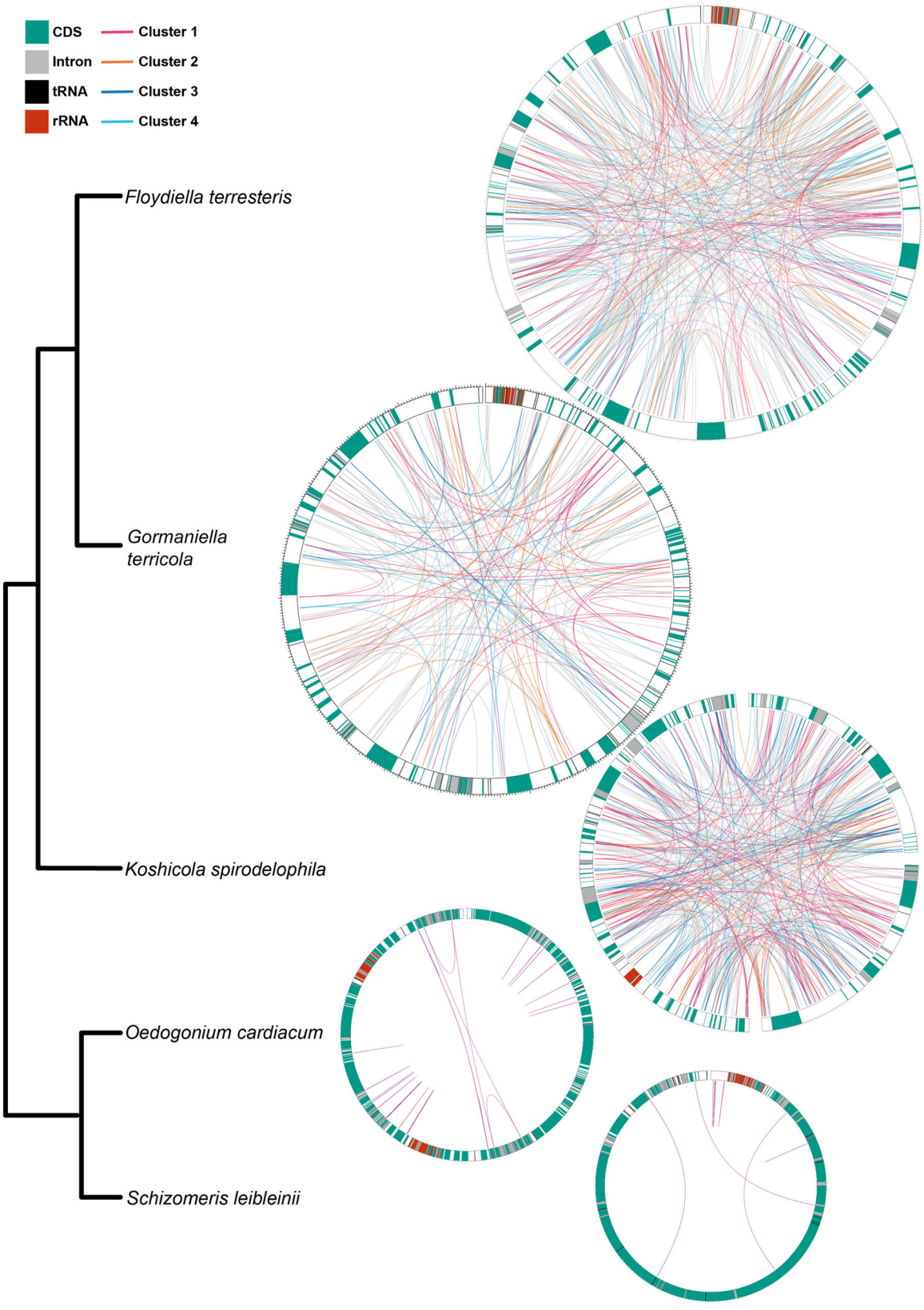

FIGURE 6 High repeat content in the chloroplast genomes of Chaetopeltidales, compared to other OCC orders, Oedogoniales (*Oedogonium*) and Chaetophorales (*Schizomeris*). Links indicate related repeat sequences found within a chloroplast genome, grouped into clusters based on similarity. The four most abundant clusters are displayed in distinct colors, while the remaining are all displayed as grey links. Circular genome plots are scaled roughly to relative genome size. Colored blocks indicate the position of an annotated gene. Clusters are groups of related short repeats.

evolution between Chaetopeltidales and Chlamydomonadales, although the repeats in Chaetopeltidales appear to be far more diverse than those reported in other green algae (Smith, 2020a). However, it remains unclear how these repetitive sequences initially proliferated. Like *F. terrestris*, *G. terricola* lacks a chloroplast-encoded reverse transcriptase (RT) gene, which would be required by the hypothesis that postulates an RT-mediated model of repeat proliferation (Pombert et al., 2005; Smith and Lee, 2009).

The second factor influencing the large chloroplast genome size *G. terricola* is the invasion of mobile introns and other foreign DNA. The total intron content of *G. terricola* is 28,551 bp, which is higher than that in *F. terrestris*, but not remarkably high relative to other members of the OCC clade (21,858 bp and 24,568 bp for *O. cardiacum* and *S. leibleinii*, respectively). In addition, the chloroplast genome of *G. terricola* is marked by a 30-kb "gene desert" containing four widely dispersed, large ORFs (879–1707 bp) that have no blastx hits on NCBI. Together, the intron content and the gene desert account for an additional 14% of the total genome size.

There are also differences in intron content and relative position between the chloroplast genomes of *F. terrestris* and *G. terricola*. The two organisms share a similar number of introns (26 and 29, respectively), but the location and content of these introns can vary significantly. In the rRNA operon, there are a combined 11 introns in *F. terrestris*, while there are six in *G. terricola*, one of which appears to be totally novel relative to *F. terrestris*. Not only are there different numbers of introns, but they have strikingly different content. Of the six introns in the rRNA operon of *G. terricola*, three have ORFs encoding homing endonucleases, none of which are shared with the single endonuclease of *F. terrestris*. Similarly in the gene *psbA*, *F. terrestris* has five introns, while *G. terricola* has 10, of which only three are shared with *F. terrestris*. Intron 4 of *G. terricola psbA* is a tandem duplication of intron 2 in *F. terrestris*.

The first look at the mitochondrial genome in OCC clade

While the chloroplast genomes of green algae have been characterized to some extent, their mitochondrial genomes have been comparatively undersequenced. Given this, it is unsurprising that the mitochondrial genome of *G. terricola* is the first published mitochondrial genome in Chaetopeltidales. In contrast to its massive chloroplast genome, the mitochondrial genome of *G. terricola* is fairly unremarkable in size (66,927 bp), especially when compared to some massive mitochondrial genomes in Chlorophyta, such as that of *Ostreobium queketti* (Repetti et al., 2020). The mitochondrial genome of *G. terricola* shares common features with many green algal lineages, such as the fragmentation and scattering of the small and large ribosomal subunits across the genome (Fučíková et al., 2014), the presence of *cox2a* but not *cox2b* (Rodríguez-Salinas et al., 2012), and the propensity for the invasion of mobile introns. While repeats are present in the mitochondrial genome, they are not as abundant in the chloroplast genome, nor do they appear to be playing such a key role in shaping the mitochondrial genome architecture. The two organellar genomes also do not share repeat elements as has been observed in Chlamydomonadales (Smith, 2020b).

Endonucleases and horizontal gene transfer

Both the mitochondrial and chloroplast genomes of *G. terricola* display a diversity of mobile introns, which often contain encoded homing endonuclease genes. In the chloroplast genome, the LAGLIDADG, GIY-YIG, and H-N-H families were all represented, whereas in the mitochondrial genome only LAGLIDADG endonucleases were encoded. We carried out a phylogenetic reconstruction of the LAGLIDADG endonucleases, given that they represent the majority of the endonucleases in both organellar genomes and given the strong fungal blastp hits that some returned. The resulting phylogeny (Figure 5) highlights the mobility of these intron elements and supports several distinct putative transfer events. The first is what appears to be at least one horizontal gene transfer event from fungi to green algae (Figure 5), where in clade 4 there are algal sequences nested within a largely fungal clade. There is also a third possible fungus to algae HGT in clade 3, but given this sequence is not nested very deeply within the clade, we are less confident in its relationship. The second is a pair of transfers of an endonuclease typically encoded in the chloroplast genomes of green algae to the mitochondrial genome (Figure 5, clade 2). Third is a transfer to the chloroplast genome of *G. terricola*, which appears to be derived from bacterial endonucleases (Figure 5, clade 1). While it is difficult to definitively determine the directionality of such HGT events, the phylogenetic placement of these endonucleases within clades represented by other organelles or distantly related organisms is strongly suggestive. If these inferred events are indeed examples of HGT, then it suggests that horizontal transfer of these mobile introns likely happens at a high frequency. Such transfer events are not unprecedented and extensive mitochondrial intron HGT was recently documented in brown algae (Starko et al., 2021). The extent to which these events occur may become clearer as greater algal organellar genome diversity is sampled in the coming years.

CONCLUSIONS

Here we describe a new taxon of green algae and report the detailed characterization of its organellar genomes. The chloroplast genome reveals a common pattern of gigantism in the order Chaetopeltidales, which is tied to the proliferation of a group of short dispersed repeats. The mitochondrial genome contains multiple LAGLIDADG homing endonucleases, some of which appear to be the result of HGT from fungi. Furthermore, this study indicates there is likely greater species

diversity within the OCC clade than is currently thought and that sampling from unconventional locations can help to uncover such diversity.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (DEB-1831428) to F.-W.L., and (DEB-1541539) to L.A.L. We thank the members of the Li lab, Associate Editor Sean Graham, and two anonymous reviewers for comments and suggestions.

DATA AVAILABILITY STATEMENT

The Illumina sequencing reads were deposited at NCBI SRA (PRJNA788523). The 18S sequence and the organellar genome assemblies were deposited at NCBI GenBank (OL839336 and OL839337). Custom script is available at https://github.com/TARobison/Robison_etal_2022/blob/main/EndonucleaseScript.

ORCID

Tanner Robison http://orcid.org/0000-0002-1629-9786
Jessica M. Nelson http://orcid.org/0000-0002-6062-6051
Louise A. Lewis http://orcid.org/0000-0003-0695-5134
Fay-Wei Li http://orcid.org/0000-0002-0076-0152

REFERENCES

Balaban, M., N. Moshiri, U. Mai, and S. Mirarab. 2019. TreeCluster: clustering biological sequences using phylogenetic trees. *PLoS One* 14: e0221068.

Bell, D., Q. Lin, W. K. Gerelle, S. Joya, Y. Chang, Z. N. Taylor, C. J. Rothfels, et al. 2020. Organellomic data sets confirm a cryptic consensus on (unrooted) land–plant relationships and provide new insights into bryophyte molecular evolution. *American Journal of Botany* 107: 91–115.

Bodenhofer, U., E. Bonatesta, C. Horejš-Kainrath, and S. Hochreiter. 2015. msa: an R package for multiple sequence alignment. *Bioinformatics* 31: 3997–3999.

Brouard, J.-S., C. Otis, C. Lemieux, and M. Turmel. 2010. The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the Chlorophyceae. *Genome Biology and Evolution* 2: 240–256.

Brouard, J.-S., M. Turmel, C. Otis, and C. Lemieux. 2016. Proliferation of group II introns in the chloroplast genome of the green alga *Oedocladium carolinianum* (Chlorophyceae). *PeerJ* 4: e2627.

Caisová, L. 2016. *Dicranochaete*—an enigmatic green alga with surprising adaptive capabilities. *Phycologia* 55: 219–229.

De Wever, A., F. Leliaert, E. Verleyen, P. Vanormelingen, K. Van der Gucht, D. A. Hodgson, K. Sabbe, and W. Vyverman. 2009. Hidden levels of phylodiversity in Antarctic green algae: further evidence for the existence of glacial refugia. *Proceedings of the Royal Society, B, Biological Sciences* 276: 3591–3599.

Dierckxsens, N., P. Mardulyn, and G. Smits. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.

Fučíková, K., P. O. Lewis, D. González-Halphen, and L. A. Lewis. 2014. Gene arrangement convergence, diverse intron content, and genetic code modifications in mitochondrial genomes of Sphaeropleales (chlorophyta). *Genome Biology and Evolution* 6: 2170–2180.

Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany* 215: 403–411.

Greiner, S., P. Lehwark, and R. Bock. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47: W59–W64.

Hall, J. D., K. Fučíková, C. Lo, L. A. Lewis, and K. G. Karol. 2010. An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie Algologie* 31: 529.

Kelchner, S. A., and J. F. Wendel. 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30: 259–262.

Kelchner, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

Krzywinski, M. I., J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.

Lambowitz, A. M., and M. Belfort. 2015. Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiology Spectrum 3: MDNA3*–0050–2014.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

Lang, B. F., M.-J. Laforest, and G. Burger. 2007. Mitochondrial introns: a critical view. *Trends in Genetics* 23: 119–125.

Leliaert, F., D. R. Smith, H. Moreau, M. D. Herron, H. Verbruggen, C. F. Delwiche, and O. De Clerck. 2012. Phylogeny and molecular evolution of the green algae. *Critical Reviews in Plant Sciences* 31: 1–46.

Lewis, L. A., and V. R. Flechtner. 2002. Green algae (Chlorophyta) of desert microbiotic crusts: diversity of North American taxa. *Taxon* 51: 443–451.

Lewis, L. A., and P. O. Lewis. 2005. Unearthing the molecular phylodiversity of desert soil green algae (Chlorophyta). *Systematic Biology* 54: 936–947.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

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

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

Mirarab, S., N. Nguyen, S. Guo, L.-S. Wang, J. Kim, and T. Warnow. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. Journal of Computational Biology 22: 377–386.

Mower, J. P., and T. L. Vickrey. 2018. Structural diversity among plastid genomes of land plants. *Advances in Botanical Research* 85: 263–292.

Nelson, J. M., D. A. Hauser, J. A. Gudiño, Y. A. Guadalupe, J. C. Meeks, N. Salazar Allen, J. C. Villarreal, and F.-W. Li. 2019. Complete genomes of symbiotic cyanobacteria clarify the evolution of vanadium-nitrogenase. *Genome Biology and Evolution* 11: 1959–1964.

Nelson, J. M., D. A. Hauser, and F.-W. Li. 2021. The diversity and community structure of symbiotic cyanobacteria in hornworts inferred from long-read amplicon sequencing. *American Journal of Botany* 108: 1731–1744.

Palmer, J. D. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* 19: 325–354.

Pombert, J.-F., C. Otis, C. Lemieux, and M. Turmel. 2005. The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights

into the branching order of chlorophyte lineages. *Molecular Biology and Evolution* 22: 1903–1918.

Repetti, S. I., C. J. Jackson, L. M. Judd, R. R. Wick, K. E. Holt, and H. Verbruggen. 2020. The inflated mitochondrial genomes of siphonous green algae reflect processes driving expansion of noncoding DNA and proliferation of introns. *PeerJ* 8: e8273.

Rodríguez-Salinas, E., H. Riveros-Rosas, Z. Li, K. Fucíková, J. J. Brand, L. A. Lewis, and D. González-Halphen. 2012. Lineage-specific fragmentation and nuclear relocation of the mitochondrial *cox2* gene in chlorophycean green algae (Chlorophyta). *Molecular Phylogenetics and Evolution* 64: 166–176.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Sanchez-Puerta, M. V., P. I. Leonardi, C. J. O'Kelly, and E. J. Cáceres. 2006. *Pseudulvella americana* belongs to the order Chaetopeltidales (class Chlorophyceae), evidence from ultrastructure and SSU rDNA sequence data 1. *Journal of Phycology* 42: 943–950.

Schmidt, S. K., R. C. Lynch, A. J. King, D. Karki, M. S. Robeson, L. Nagy, M. W. Williams, et al. 2011. Phylogeography of microbial phototrophs in the dry valleys of the high Himalayas and Antarctica. *Proceedings of the Royal Society, B, Biological Sciences* 278: 702–708.

Smith, D. R. 2020a. Can green algal plastid genome size be explained by DNA repair mechanisms? *Genome Biology and Evolution* 12: 3797–3802.

Smith, D. R. 2020b. Common repeat elements in the mitochondrial and plastid genomes of green algae. *Frontiers in Genetics* 11: 465.

Smith, D. R., and R. W. Lee. 2009. The mitochondrial and plastid genomes of *Volvox carteri*: bloated molecules rich in repetitive DNA. *BMC Genomics* 10: 132.

Sousa, F., P. Civáň, P. G. Foster, and C. J. Cox. 2020. The chloroplast land plant phylogeny: analyses employing better-fitting tree- and site-heterogeneous composition models. *Frontiers in Plant Science* 11: 1062.

Starko, S., T. T. Bringloe, M. S. Gomez, H. Darby, S. W. Graham, and P. T. Martone. 2021. Genomic rearrangements and sequence evolution across brown algal organelles. *Genome Biology and Evolution* 13: evab124.

Tang, H., J. E. Bowers, X. Wang, R. Ming, M. Alam, and A. H. Paterson. 2008. Synteny and collinearity in plant genomes. *Science* 320: 486–488.

Turmel, M., A.-S. Bélanger, C. Otis, and C. Lemieux. 2020. Complete mitogenomes of the chlorophycean green algae *Bulbochaete rectangularis* var. *hiloensis* (Oedogoniales) and *Stigeoclonium helveticum* (Chaetophorales) provide insight into the sequence of events that led to the acquisition of a reduced-derived pattern of evolution in the Chlamydomonadales and Sphaeropleales. *Mitochondrial DNA Part B* 5: 611–613.

Turmel, M., and C. Lemieux. 2018. Evolution of the plastid genome in green algae. *Advances in Botanical Research* 85: 157–193.

Watanabe, S., K. Fučíková, L. A. Lewis, and P. O. Lewis. 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: 865–875.

Yang, H.-P., M. Wenzel, D. A. Hauser, J. M. Nelson, X. Xu, M. Eliáš, F.-W. Li. 2021 *Monodopsis* and *Vischeria* genomes shed new light on the biology of eustigmatophyte algae. *Genome Biology and Evolution* 13: evab233.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8: 28–36.

SUPPORTING INFORMATION

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

Appendix S1. Collection site of *Gormaniella terricola*.

Appendix S2. Chaetopeltidales 18S maximum likelihood tree.

Appendix S3. Summary of largest chloroplast genomes available on NCBI.

Appendix S4. Maps of chloroplast and mitochondrial genomes.

Appendix S5. Endonucleases blastp results.

How to cite this article: Robison, T., J. M. Nelson, D. A. Hauser, L. A. Lewis, and F.-W. Li. 2022. Dynamic plastid and mitochondrial genomes in Chaetopeltidales (Chlorophyceae) and characterization of a new chlorophyte taxon. *American Journal of Botany* 109(6): 939–951. https://doi.org/10.1002/ajb2.16015