Title: Phylotranscriptomics points to multiple independent origins of multicellularity and cellular differentiation in the volvocine algae

Charles Ross Lindsey[1], Frank Rosenzweig[1,2], and Matthew D Herron[1*]

[1]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332

[2]Parker H. Petit Institute for Bioengineering and Biosciences

***Corresponding author**: xprinceps@gmail.com

Running Title: Phylotranscriptomics of the volvocine algae

Authorship Contributions: Performed the experiments (CRL), carried out the analyses (CRL, MDH), wrote the manuscript (CRL, MDH, RFR)

Acknowledgements: The authors gratefully acknowledge the core facilities at the Parker H. Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology for the use of their shared equipment, services and expertise. We are also grateful for useful discussion with and comments on the manuscript provided by Andy Chea, Emily Cook, Sa Geng, Armin Hallmann, Joseph Lachance, Bradley Olson, and James Umen. This work used the Hive cluster, which is supported by the National Science Foundation under grant number OAC-1828187. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA. This project was funded by NASA Astrobiology Grant# NNA17BB05A to Rosenzweig (PI) Herron (co-I); Rosenzweig NASA Exobiology Grant # 80NSSC20K0621; Rosenzweig start-up account DE00000308; NSF DEB-1723293 to Herron (PI) Rosenzweig (co-PI). This material is based upon work while Herron was serving at the National Science Foundation.

Conflict of Interest Statement: The authors declare that they have no conflict of interest.

Keywords: phylotranscriptomics, volvocine algae, green algae, phylogeny, multicellularity

Abstract

The volvocine algae, which include the single-celled species *Chlamydomonas reinhardtii* and the colonial species *Volvox carteri*, serve as a model in which to study the evolution of multicellularity and cellular differentiation. Studies reconstructing the evolutionary history of this group have often relied on datasets of one to a few genes for phylogenetic inference and ancestral character state reconstruction. These studies suggest that multicellularity evolved only once in the volvocine algae, that each of its three colonial families is monophyletic, and that there have been at least three independent origins of cellular differentiation in the group. We performed RNA-Seq on 55 strains representing 47 volvocine algal species, and obtained similar data from curated databases on 13 additional strains. We generated a dataset consisting of transcripts for 40 single-copy, protein-coding, nuclear genes, then subjected the predicted amino acid sequences of these genes to maximum likelihood, Bayesian inference, and coalescent-based analyses. These analyses show that multicellularity independently evolved at least twice in the volvocine algae and that the colonial family Goniaceae is not monophyletic. Our data further indicate that cellular differentiation independently arose at least four and possibly as many as six times within the group. Altogether, these results show how multicellularity and cellular differentiation are evolutionarily labile in the volvocine algae, affirming their importance for the study of major transitions in the history of life.

Introduction

The evolution of multicellularity is widely considered a major transition in the history of life (Grosberg & Strathmann, 2007; Maynard Smith & Szathmáry, 1995; Szathmáry & Maynard Smith, 1995; West *et al.*, 2015). Multicellularity not only gave rise to most of the visible life forms on the planet, but also opened the door to cellular differentiation, including that between somatic and reproductive cells, a hallmark feature of sexual reproduction in eukaryotes that exhibit morphological complexity (Hallmann, 2011; Szathmáry & Maynard Smith, 1995; Wahl & Murray, 2016). Questions regarding the evolution of multicellularity and cellular differentiation have been approached using the fossil record (Chen *et al*., 2014; Han & Runnegar, 1992; Schirrmeister *et al*., 2013), laboratory evolution (Herron *et al.*, 2019; Quintero-Galvis *et al.*, 2018; Ratcliff *et al.*, 2012, 2013), and comparative approaches that include superimposing cell biology upon molecular phylogeny (Kirk, 2005; Koyanagi, 2015) (Herron & Michod, 2008; Kirk, 2005; Koyanagi, 2015). The last of these approaches is predicated on the assumption that the cell biology and molecular phylogeny are mutually informative, an assumption that requires the phylogeny itself to be accurate.

The volvocine green algae have proved especially useful for investigating the major transition leading to multicellularity. The group consists of ~50 extant species which exhibit a range of body plans, cell numbers, sizes, and forms of sexual reproduction. The smallest of these are single-celled (e.g., *Chlamydomonas reinhardtii*); the largest, at up to 3 mm in diameter and up to 50,000 cells, are spheroidal, swimming colonies in the genus *Volvox*. Since the initial "very pleasant sight" of swimming *Volvox* colonies described by Van Leeuwenhoek more than 300 years ago (van Leeuwenhoek, 1700), the volvocine algae have come to be accepted as a useful model system in which to address questions related to the origins of multicellularity and cellular

differentiation (Kirk, 2001; Schmitt, 2003). Multiple species have now had their genomes sequenced (Craig *et al.*, 2021; Featherston *et al.*, 2018; Hanschen *et al.*, 2016; Merchant *et al.*, 2007; Prochnik *et al.*, 2010), and those of unicellular *C. reinhardtii* and multicellular *V. carteri* forma *nagariensis* are well-annotated (Merchant *et al.*, 2007; Prochnik *et al.*, 2010). However, the volvocine algae encompass more than two organisms representing alternative forms of life in terms of size and development. Vegetative forms range in characteristic cell number from 1 to ~50,000, and exhibit intermediate degrees of complexity likely similar to extinct ancestors. Further, multicellularity and cellular differentiation arose within the volvocine algae much more recently than those traits arose in animals: ~220 million years ago (Herron *et al.*, 2009) versus 600 million years ago (Brunet & King, 2017), respectively.

Evolution of the volvocine algae has sometimes been viewed as a linear progression in size and complexity (Lang, 1963; Larson *et al.*, 1992). Unicellular taxa such as *Chlamydomonas* occupy one end of this continuum while fully differentiated, multicellular taxa such as *Volvox* occupy the other. This concept, the 'volvocine lineage hypothesis', used a streamlined phylogeny of the volvocine algae to help explain how a multicellular species with complete germ-soma differentiation such as *Volvox* might evolve from a unicellular, *Chlamydomonas*-like ancestor. However, morphological and molecular phylogenetic studies suggest that the history of the volvocine algae may be more complicated, as cellular differentiation, different modes of sexual reproduction, and varying body plans appear to have evolved multiple times within the group (Hanschen *et al.*, 2018b; Herron *et al.*, 2010).

Current understanding of the major evolutionary relationships within this group has often been based on the analysis of five chloroplast gene sequences (Hanschen *et al.*, 2018; Herron *et al.*, 2009; Herron & Michod, 2008; Nakada *et al.*, 2010; Nozaki *et al.*, 2000, 2014, 2015).

Chloroplast gene-based phylogenies have also been used to carry out ancestral-state reconstructions (Grochau-Wright *et al*., 2017; Hanschen *et al*., 2018; Herron *et al*., 2010; Herron & Michod, 2008), opening a window on how multicellularity and cellular differentiation evolved within the volvocine algae. Overall, the branching order of most chloroplast gene-based phylogenies is defined by two related groups: *(i)* a set of unicellular species (e.g., *Chlamydomonas reinhardtii*) that are paraphyletic with respect to *(ii)* a clade that encompasses the three major families of colonial volvocine algae: Tetrabaenaceae (*Tetrabaena* and *Basichlamys*), Goniaceae (*Gonium* and *Astrephomene*), and Volvocaceae (*Colemanosphaera, Eudorina, Pandorina, Platydorina, Pleodorina, Volvox, Volvulina,* and *Yamagishiella*) (**Figure 1A** and **1D**). In this scheme, the Tetrabaenaceae is a sister group to the clade formed by the Goniaceae and Volvocaceae. Although this framework only takes into account family-level relationships, several conclusions can be drawn. *First*, the colonial species form a clade. *Second,* each of the three families is monophyletic. *Third,* monophyly among the colonial species implies that multicellularity evolved only once within that group with no reversion to unicellularity.

Two recent studies have called into question the monophyly of the colonial volvocine algae (**Figure 1C**). Pröschold *et al.* (2018) based their inferences on two datasets: one consisting of SSU *r*DNA sequences plus internal transcribed spacer (ITS) sequences 1 and 2, the other consisted of ITS sequences alone. By contrast, Nakada *et al.* (2019) used a single-gene 18S *r*RNA dataset. The taxonomic status of the Goniaceae has also been called into question by studies (**Figure 1B**) that indicate the group is either not monophyletic (Coleman, 1999; Nozaki *et al*., 2000) or that there is low support for a sister relationship between *Astrephomene* and *Gonium* (Nakada *et al*., 2019; Nozaki *et al*., 2000; Pröschold *et al.*, 2018). Moreover, a number of recent volvocine algal phylogenies leave uncertainty as to how many times cellular

differentiation evolved within the group. Chloroplast sequence data suggest at least 3 independent origins of cellular differentiation: in *Astrephomene*, in *Volvox* section *Volvox* (sometimes referred to as *Euvolvox*), and in the *Eudorina, Volvox, Pleodorina* (EVP) clade (**Figure 1B-D**). And within the EVP clade it is unclear whether cellular differentiation in *Pleodorina thompsonii, Volvox gigas* and *V. powersii,* and *Pleodorina starrii* and *P. indica* arose independently from *V. carteri* (**Figure 1B-D**).

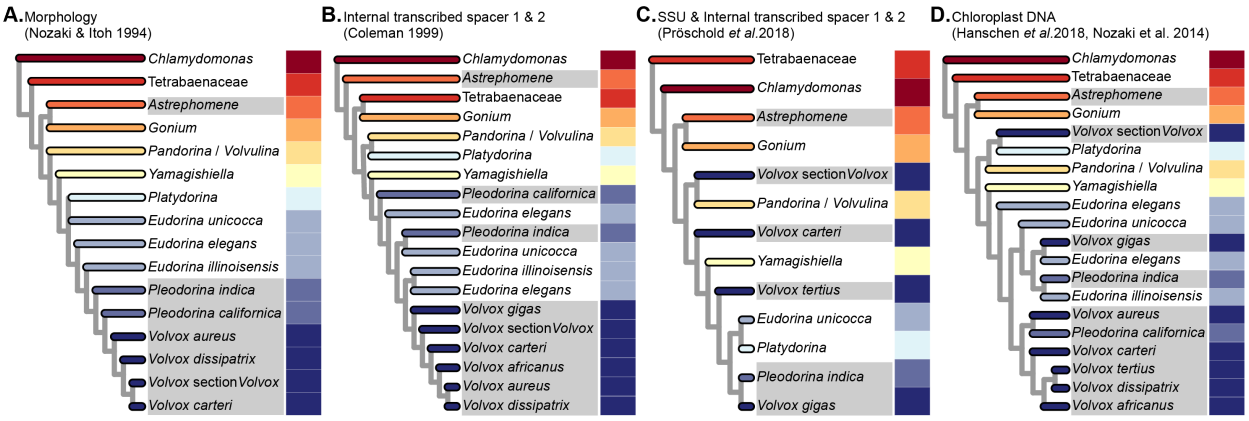

Figure 1. Phylogenies of the volvocine algae are not concordant: Four volvocine green algae phylogenies based on different types of data, displayed in chronological order of their appearance in the literature. Species highlighted in shades of gray exhibit somatic cell differentiation. The varying colors to the right of each phylogeny have been arbitrarily assigned to particular genera and are intended to be used as a visual aid to highlight differences among the phylogenies.

The foregoing uncertainties highlight the need for a new and more robust molecular phylogeny of the volvocine algae. These uncertainties may arise from incomplete taxonomic sampling, limited genetic sampling, or both. While five volvocine algal species have had their genomes sequenced, most taxonomically comprehensive phylogenetic inferences about this evolutionarily important group have been constructed using relatively small datasets. These consist of either the sequence of five chloroplast genes (Hanschen *et al*., 2018; Herron & Michod, 2008; Hu *et al*., 2019; Nozaki *et al*., 2014) representing an aggregate of ~6000

nucleotide positions, small (≤6) multi-gene datasets consisting of chloroplast gene(s) and ribosomal molecular markers, or both (Pröschold *et al*., 2018, Nakada *et al*., 2019). Moreover, the use of chloroplast genes in phylogenetic reconstruction can be problematic owing to the facts that they are not subject to sexual recombination, that they vary little among recently diverged species (Dong *et al.*, 2012), and that they can be subject to incomplete lineage sorting due to the retention of ancestral polymorphisms (Jakob & Blattner, 2006; Xu *et al.*, 2012).

Of special concern is the observation that volvocine phylogenies inferred using chloroplast genes **(Figure 1D)** conflict with those constructed using nuclear genes (**Figures 1B & 1C**) (Coleman, 1999; Nakada *et al*., 2019; Pröschold *et al*., 2018). While conflicts between chloroplast and nuclear phylogenies are not unusual (Rose *et al.*, 2020; Soltis & Kuzoff, 1995; Yu *et al.*, 2013), one recent study that used a multi-gene, nuclear dataset to infer a volvocine phylogeny is consistent with the general backbone relationships of the 5-gene chloroplast phylogenies (Zhang *et al*., 2019). However, this study based on nuclear genes is limited by relatively low taxonomic resolution. Specifically, the phylogeny of Zhang *et al.* (2019) reconstructed the same relationships inferred using chloroplast gene phylogenies **(Figure 1D)** but included only four colonial species.

Here, we seek to overcome a number of the limitations described above. We have adopted a phylotranscriptomic approach that uses a concatenated amino acid alignment of 40 nuclear protein-coding, single-copy genes. We sequenced whole transcriptomes of 55 strains encompassing 47 nominal species. Then, we increased taxonomic sampling by making use of previously published RNA-Seq data for 9 strains and shared amino acid sequences for 4 strains. Our goal was to derive a robust phylogeny of the volvocine algae that would enable inferences about the evolution of multicellularity, cellular differentiation, sexual dimorphism, and other

traits in this group. In this study, we present the most taxonomically comprehensive phylogeny yet produced of the volvocine algae using a nuclear dataset, including all described genera and multiple representatives of all genera that are not monotypic. Our results show that the colonial species do not form a clade, that the Goniaceae are not monophyletic, and that multicellularity has independently evolved at least twice within the volvocine algae, and cellular differentiation at least four times.

Materials and Methods

Strains and culture conditions Forty-two colonial strains of volvocine algae and two unicellular strains were obtained for this study from the National Institute for Environmental Studies (NIES, Japan). An additional eight colonial strains and two unicellular strains were obtained from the Culture Collection of Algae at the University of Göttingen (SAG, Germany). Finally, one strain was obtained from The Culture Collection of Algae at the University of Texas at Austin (UTEX, USA). Strain provenance and culture collection ID numbers are shown in **Table 1**. All cultures were grown at 20-26°C under cool-white LED lamps with an intensity of 2500-2700 lux under a 14-hour light/10-hour dark cycle. A detailed description of the medium used to culture each strain, as well each strain's morphology, degree of cellular differentiation, and gamete size are provided in **Supplementary Materials: Tables S1 and S2, respectively.**

Table 1. List of taxa used in this study + sequencing and assembly summary. Under Strain or Pubmed ID, "CC" refers to Chlamydomonas Culture Collection at the University of Minnesota (CC, USA), "F" refers to Culture Collection of Freshwater Algae at the Institute of Hydrobiology, Chinese Academy of Sciences (FACHB, China), "N" refers to National Institute for Environmental Studies (NIES, Japan), "S" refers to Culture Collection of Algae at the University of Göttingen (SAG, Germany), and "U" refers to The Culture Collection of Algae at the University of Texas at Austin (UTEX, USA). "QRPMK" and "TR" under RNA Extraction Method refer to QIAGEN RNeasy Plant Mini Kit and TRizol RNeasy, respectively. Strains assigned an asterisk represent data used from previously published studies.

Taxa	Strain or Pubmed ID	RNA Extraction Method	Total # of Raw Reads	Total # of Clean PE Reads	Total # of Contigs	N50	G/C %
Astrephomene gubernaculifera	N-418	QRPMK	61460551	44350080	78105	3894	0.61
A. perforata	N-564	QRPMK	63350620	46811342	72913	3613	0.61
Basichlamys sacculifera	N-566	QRPMK	66205749	48327288	134880	1945	0.69
Colemanosphaera angeleri	F 2363*		28119810	27000176	120763	3348	0.62
C. charkowiensis	F 2326*		27967726	26876865	103286	3382	0.62
	N-3383	TR	58777161	40550696	119848	2523	0.62
Chlamydomonas debaryana	S 11-55a	QRPMK	87455695	64552440	105096	2776	0.67
	S 70.81	QRPMK	50406749	42095108	103962	2858	0.64
C. globosa	S 81.72	TR	73820020	49419928	120062	2079	0.65
C. moewusii	S 11-16f	QRPMK	54728573	40436359	118833	2498	0.6
C. reinhardtii	CC-503*						
C. schloesseri	S 2486	QRPMK	70163393	49255021	64233	3094	0.66

***Chlorella variabilis* NC64a**	20852019*						
Eudorina cylindrica	F 2322*		26442110	25262148	103026	2846	0.64
	N-722	QRPMK	44479173	36297801	105029	2927	0.64
E. elegans	F 2321*						
	N-456	TR	84549961	56756776	122975	2108	0.63
	N-458	QRPMK	67451343	46725012	131335	2944	0.65
	N-568	QRPMK	83433797	69539684	120487	3614	0.64
	N-717	QRPMK	47625429	37616922	89118	2094	0.64
	N-719	QRPMK	57157324	43565312	98180	2976	0.63
	N-720	QRPMK	38134498	31388413	95662	3699	0.61
E. illinoisensis	N-460	QRPMK	62485922	49396583	97014	2615	0.64
E. minodii	N-856	QRPMK	83574258	66413244	128802	4127	0.61
E. peripheralis	N-725	QRPMK	46607143	36528930	100934	2322	0.62
E. unicocca	S 24-1c	TR	54482446	37679863	110391	2267	0.62
Gonium multicoccum	N-737	QRPMK	65984190	44301578	122279	3275	0.64
G. pectorale	N-2863*						
G. octonarium	N-851	TR	60029472	39826211	104857	2302	0.65
G. quadratum	N-653	QRPMK	62431743	48230702	100468	3789	0.64
G. viridistellatum	N-654	QRPMK	43286980	31852410	97991	2837	0.65
Pandorina colemaniae	F 2361*		25181003	23823370	80843	3292	0.62
	N-572	TR	49862534	37589048	84053	2237	0.62
P. morum	F 2362*		29837807	28442854	147990	2674	0.61
	N-890	QRPMK	39337532	24490235	133742	3123	0.61
Platydorina caudata	N-728	QRPMK	48702213	36896556	125206	3503	0.61
Pleodorina indica	N-736	QRPMK	57564627	44082516	192570	2006	0.63
P. japonica	U 2523	QRPMK	64897803	49495315	83662	3499	0.61
P. starrii	N-1362	TR	68790016	45935282	87641	2438	0.64
	N-1363	TR	59466876	44726861	128732	1695	0.63
P. thompsonii	N-4126	QRPMK	81077847	61986735	118940	4005	0.62
Tetrabaena socialis	N-571*		42302450	41232209	62014	1012	0.65
Vitreochlamys aulata	N-878	QRPMK	80710519	54651118	132115	2673	0.65
	S 80.81	QRPMK	41008694	32585611	84834	2100	0.65
V. nekrassovii	S 11-10	QRPMK	65658295	48341108	109360	3304	0.62
V. ordinata	N-882	TR	60631467	42981449	75434	2778	0.69
Volvox africanus	N-863	QRPMK	70739371	52378949	110066	4586	0.55
V. aureus	N-541	QRPMK	80821130	64315625	106245	4065	0.55
V. barberi	N-730	QRPMK	49805303	37197105	152910	2217	0.58
V. carteri* f. *kawasakiensis	N-732	TR	53994112	34239609	86004	1982	0.58
V. carteri* f. *nagariensis	20616280*						
	N-865	QRPMK	65624222	50641434	117314	3947	0.57
V. carteri* f. *weismannia	N-866	QRPMK	25913610	20987137	78230	2636	0.57
V. dissipatrix	N-4128	QRPMK	77137081	51045569	112190	5146	0.55
V. ferrisii	N-3986	QRPMK	49810357	37522413	91045	3281	0.57
V. gigas	N-867	QRPMK	50554912	37592146	151302	2499	0.63
V. globator	S 199.80	QRPMK	67413919	52861449	100861	4005	0.56
V. kirkiorum	N-543	QRPMK	44941633	34650729	89533	3081	0.56
V. obversus	N-868	QRPMK	67394604	50823546	99171	3881	0.58
V. ovalis	N-2569	QRPMK	52242003	35319700	107999	3643	0.53
V. powersii	N-4127	QRPMK	55232784	42410688	140237	2908	0.62
V. tertius	N-544	QRPMK	86222840	62563240	108551	5454	0.54
Volvulina boldii	N-893	QRPMK	25665262	20161297	158265	1593	0.62
V. compacta	F 2337*		27543035	26374210	90261	3550	0.61
	N-582	QRPMK	53045384	38991781	112583	3380	0.63
V. pringsheimii	N-895	TR	82233959	55849890	128304	3122	0.62
V. steinii	S 90-1	QRPMK	46133638	37699635	89515	3635	0.62
Yamagishiella unicocca	F 2364*		30922957	29574420	64415	3609	0.63

Previously published transcriptomic data Previously published RNA-seq data used for this study include those for *Tetrabaena socialis* NIES-571 (Featherston *et al*., 2018) as well as for 8 strains from the Culture Collection of Freshwater Algae at the Institute of Hydrobiology, Chinese Academy of Sciences (FACHB, China; Hu *et al*., 2020). Taxon names and culture collection or Pubmed IDs used in previously published studies are designated with an asterisk in **Table 1**. All previously published RNA-seq data used in this study were accessed via the National Center for Biotechnology Information (NCBI) Sequence Read Archive using the accession numbers shown in **Supplementary Materials: Table S3**.

RNA extraction procedures Two protocols were used to isolate total RNA: a modified version of the TRizol RNeasy method described by Matt & Umen (2018) and a slightly modified QIAGEN RNeasy Plant Mini Kit protocol. For a detailed description of each please see **Supplementary Materials: File lindseyetal_rna_extr_protocols.docx**. Information on which protocol was used for a given strain is provided in **Table 1.**

Library Preparation and Sequencing Before generating a sequencing library, RNA quality and quantity were assessed by Nanodrop and Qubit (Thermo Fisher Scientific, Waltham, MA 02451 USA). RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA 95051 USA. mRNA was isolated using poly T beads, whereafter Illumina libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit. Library concentrations were determined fluorometrically; sequencing was carried out on the Illumina

NovaSeq 6000 platform (Illumina, Inc., San Diego, CA 92122 USA) to generate 151 bp paired-end reads.

Quality Control of Reads Raw read quality was assessed through FastQC v.0.11.8 with an additional FastQC assessment post-trimming. Quality control of the raw reads was completed with Trimmomatic v.0.39 (Bolger *et al*., 2014) where the bases at the 5' and 3' end of each read are trimmed if found to be below a quality score of 3. A 4-base sliding window approach was used to trim the rest of the read once average quality fell below a score of 15; reads that were below a minimum length of 36 bases were discarded (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). If adapter content was detected by FastQC the additional ILLUMINACLIP step was used with the 'TruSeq3-PE-2.fa' file provided by the Trimmomatic developers. If performed, the following ILLUMINACLIP parameters were used: 2:30:10 at the beginning of each command line. This allows for 2 "seed" mismatches where the "seed" is a short segment of the adapter that is being aligned in every section of the read. If more than 2 mismatches occurred, no trimming of the read occurred. Additionally, there had to be at least 30 matched bases in the paired-end palindrome read alignment and at least 10 matched bases between an adapter sequence and read.

De Novo Assembly SOAPdenovo-Trans v1.0.4 (Xie *et al*., 2014) was used to assemble *de novo* transcriptomes from the quality filtered, paired-end reads using a k-mer size of 25 (SOAPdenovo-Trans-31mer all -s <config input file> -o <outfile> -K 25). GapCloser from the SOAPdenovo package was utilized to close gaps in each transcriptome using the same configuration file, which contains read-specific information and file paths, from the previous step

(-b <config file> -a <.scafSeq file output by SOAPdenovo-Trans> -o <outfile> -l <max read length, int value> -t <thread number>). Default parameters were used for CD-HIT v4.8.1 (Fu *et al.*, 2012) to reduce redundant transcripts from our *de novo* transcriptomes.

Orthologous Gene Identification for Phylotranscriptomic Analysis The evolutionary history of the volvocine algae dates back at least 200 million years (Herron *et al*., 2009). Over this timescale nucleotide sequences become saturated with substitutions, diminishing their phylogenetic utility (Hasegawa & Hashimoto, 1993; Yang & Roberts, 1995). Amino acid sequences were therefore chosen for our alignments, as they are known to be more reliable for ascertaining distant evolutionary relationships. De Clerck and colleagues identified 58 nuclear protein-coding, single copy genes that were members of highly conserved gene families across the green algae (*Chlorophyceae, Prasinophytes,* and *Trebouxiophyceae*) and land plants (*Streptophyta*) (De Clerck *et al*., 2018). Their amino acid alignment of the 58 nuclear protein-coding genes that includes *Chlamydomonas reinhardtii* CC-503*, Chlorella variabilis* NC64A*, Gonium pectorale* NIES-2863*, and Volvox carteri* HK10 was kindly shared with our research team. Out of the 58 genes shared, we used 40 for our gene alignments. In order to identify those specific genes in the *de novo* transcriptomes of our taxa a Basic Local Alignment Search Tool (BLAST) server was established in our lab, and a unique BLAST database for each taxon was created following the instructions in the BLAST manual. A BLASTP search using the *C. reinhardtii* CC-503*, G. pectorale* NIES-2863*,* and *V. carteri* HK10 genes from De Clerck *et al.* as our query sequences enabled us to identify the orthologous genes for each of our taxa.

Gene Sequence Alignments and Phylotranscriptomic Analysis The BLASTP results were used to identify the scaffold and open read frame where each gene was located in a strain's transcriptome. Using a proprietary Python script (**Supporting Materials**: **File translate_scaffold.py)** each scaffold was extracted from its transcriptome and translated in the appropriate reading frame, then the translated scaffold was added to an alignment file. For consistency, we generated *de novo* transcriptomes since we lacked a reference genome for most of our sequenced strains. At times, a gene was found to be incomplete for a given taxon due to assembler or sequencing error after manual examination. When this was determined to be the case, the gene was manually stitched together. This was done in a highly conservative manner: if we could not ascertain whether or not a gene was incomplete due to assembler or sequencing error then it was excluded from the alignment for the given species. Amino acid scaffolds were aligned using MUSCLE v3.8.31 (Edgar, 2004). Alignments were also subjected to manual alignment in Aliview v1.26 (Larsson, 2014); extraneous data were trimmed, leaving only the aligned genes. Ambiguously aligned regions were eliminated from each alignment leaving only conserved and reliably aligned regions for phylogenetic analysis using the following parameters in Gblocks v0.91b (Castresana, 2000): -t=p -b3=8 -b4=2 -b5=h -b6=y. Phyutility v2.7.1 (S. A. Smith & Dunn, 2008) was used to concatenate all gene alignment files.

Phylogenies were inferred under maximum-likelihood (ML), Bayesian inference (BI), and coalescent-based (CB) models. The ML and BI analyses used a multi-gene partitioned dataset where the best evolutionary model for each gene was predicted by ProtTest v3.4.2 under the Akaike Information Criterion (AIC). For information regarding each predicted evolutionary model please refer to **Supplementary Materials: Table S4**. The ML analysis was conducted using IQtree v1.6.12 (Nguyen *et al*., 2015) under partition models (Chernomor *et al*., 2016).

Support values reported for the IQtree ML analysis were estimated through the bootstrap technique where 1,000 ultrafast bootstrap replicates were generated (Hoang *et al*., 2018). The BI analysis was performed with MrBayes 3.2.7a (Ronquist *et al*., 2012) with 3 heated and 1 cold Markov chains, where trees were sampled every 1,000 generations for a total of 1,000,000 generations with 1,000 trees discarded at the beginning of each chain (ngen =100000000, samplefreq=1000, burnin=1000, nruns=4, nchains=4, starttree=random).

ASTRAL (C. Zhang *et al*., 2018) was used to perform the coalescent-based analysis where all 40 single-gene phylogenies (**Supplementary Materials: File lindseyetal_single_gene_phylogenies**) produced by RAxML v8.2.12 (Stamatakis, 2014) were used as the input after collapsing branches with low bootstrap support (<10) using Newick Utilites v1.6 (Junier & Zdobnov, 2010). Posterior probabilities were assessed for the Bayesian and coalescent-based analyses in MrBayes and ASTRAL, respectively. Lastly, Approximately Unbiased (AU) tests with 100,000 RELL re-samplings were conducted to test certain key topologies and hypotheses using IQtree (-zw 100000 -au) (**Supplementary Materials: Figure S1**).

Results

De novo transcriptome data makes possible 40 single-gene alignments. We sampled a total of 68 taxa, carrying out whole-transcriptome sequencing on 55 strains and using RNA-Seq data from 13 strains from previously published sources. Together, these 68 taxa encompassed all presumed major lineages of the colonial volvocine algae and 9 of their nearest unicellular relatives, plus an outgroup, *Chlorella vulgaris*, from the Trebouxiophyceae (**Table 1**). All described volvocine genera were included, with multiple species represented for every genus that is not monotypic. Comprehensive taxon sampling was not possible, since several described species, especially in the genus *Volvox*, are no longer available in culture collections. While our main focus was to resolve relationships within the colonial volvocine algae, our study included several closely related unicellular taxa from the genera *Chlamydomonas* and *Vitreochlamys* in order to provide better phylogenetic resolution of the *Reinhardtinia* clade.

The total number of raw reads generated from RNA-sequencing for each species ranged from 25,665,262 to 87,455,695 reads with an average of 60,194,849 reads generated per species. After quality trimming of the raw reads, in which low-quality reads and adapter sequences were discarded, the total number of clean paired-end reads ranged from 20,161,297 to 69,539,684 reads with an average of 44,416,935 reads per species (**Table 1**). From our RNA-seq data, we generated 55 *de novo* transcriptomes and mined each transcriptome using local BLAST databases for the single-copy, nuclear protein coding genes shared with us by De Clerck and colleagues (De Clerck *et al*., 2018). We treated the data from previously published studies in the same fashion as data generated in our lab by filtering the raw reads through quality trimming, then assembling *de novo* transcriptomes using the same programs and parameters (**Table 1**).

From the RNA-seq data, we assembled a total of 40 single-gene alignments that were later concatenated to a single alignment representing an aggregate of 12,650 amino acids, equivalent to 37,950 nucleotide positions, with a total of 5992 parsimony-informative sites. Numbers of informative positions in the single-gene alignments ranged from 40 to 446. For both the single-gene alignments and the concatenated alignment ambiguously aligned regions were trimmed out, leaving only the conserved and reliably aligned regions. Single-gene alignments were subjected to ML and BI analyses in order to infer single-gene phylogenies. Single-gene phylogenies were then further analyzed using a coalescent-based approach. The concatenated multi-gene alignment was partitioned so that the appropriate model of protein substitution was applied to each gene for the supermatrix phylogenetic approach under ML and BI.

Colonial volvocine algae are not monophyletic We sampled 67 taxa of core unicellular and multicellular organisms from the clade *Reinhardtinia* of the order Chlorophyceae plus 1 unicellular outgroup taxon from the Trebouxiophyceae (68 taxa altogether, **Table 1**). Trees inferred using maximum likelihood (ML), Bayesian inference (BI), and coalescence-based (CB) analyses were generally well-supported with some topological differences between the ML and BI analyses relative to the CB analysis, as described below.

Three general conclusions emerge from these analyses: *First*, the colonial volvocine algae are not monophyletic; s*econd,* at least two independent origins of multicellularity occurred within the volvocine algae; *third,* once multicellularity evolved no extant lineage reverted to the ancestral unicellular state. The three main families of the volvocine algae are represented by the Tetrabaenaceae, Goniaceae, and Volvocaceae; only the Tetrabaenaceae and Volvocaceae appear to be monophyletic in all three phylogenetic analyses (see **Figures 2** and **3**).

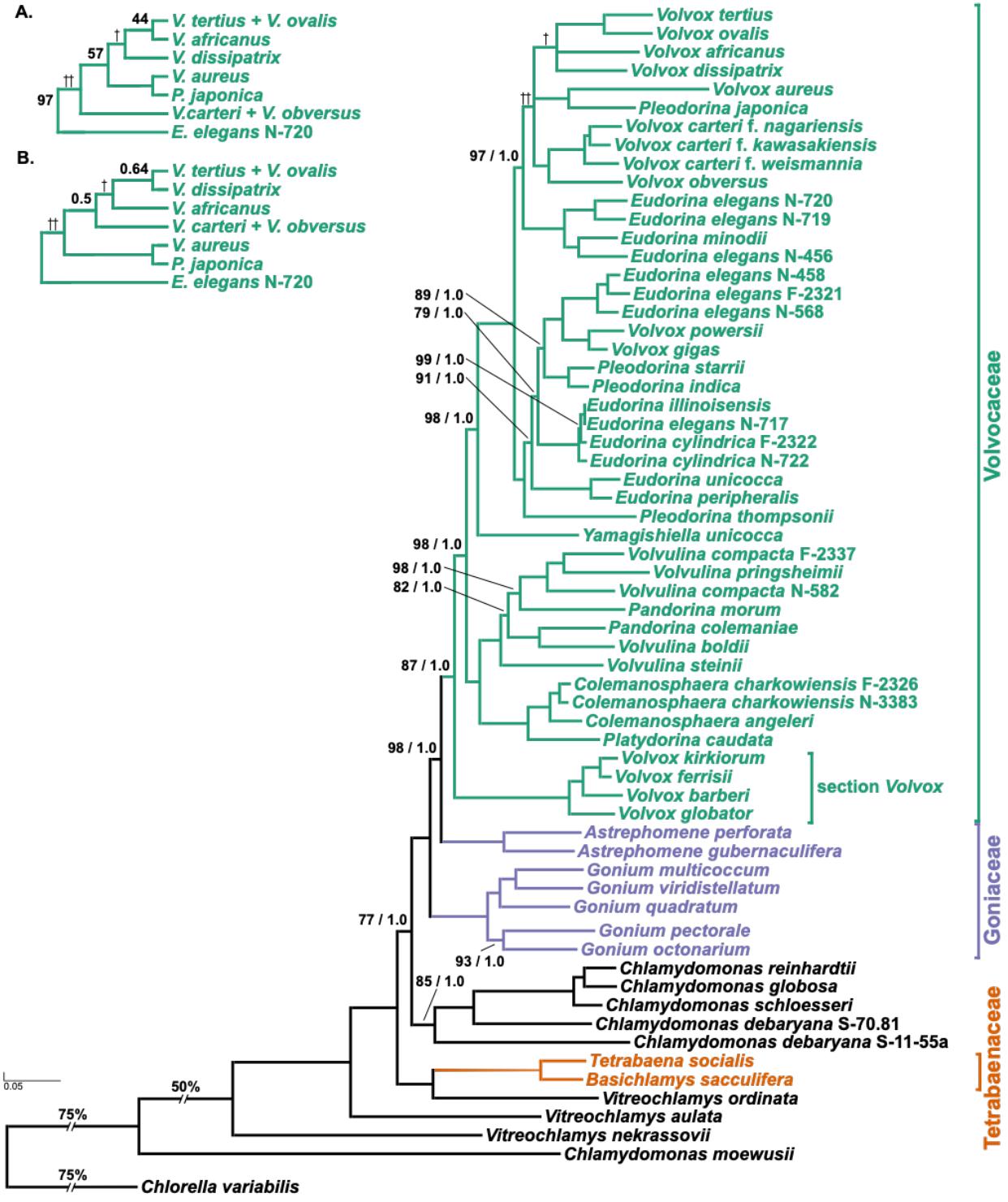

Figure 2. Molecular phylogeny of the colonial volvocine algae (Tetrabaenaceae, Goniaceae, and Volvocaceae) and closely related unicellular taxa represented by *Chlamydomonas* and *Vitreochlamys* with *Chlorella variabilis* as the outgroup. The phylogenetic tree shown is based

on a multi-gene dataset of single-copy, protein-coding nuclear genes (12,650 aligned amino acid positions of 68 taxa) inferred using the maximum likelihood method in IQtree, the branching order of which is nearly identical to that inferred in the Bayesian Inference using MrBayes. Numbers on branches represent bootstrap values and Bayesian posterior probabilities, respectively (all support values not shown are MLBS = 100, BPP = 1.0). Branch lengths correspond to genetic divergence, as indicated by the scale bar. Members of the Tetrabaenaceae, Goniaceae, and Volvocaceae are denoted in orange, purple, and green, respectively; unicellular species are denoted in black. Daggers correspond to branching order differences between the maximum likelihood and Bayesian inference analyses that are represented in greater detail in Figures 2A and 2B, respectively.

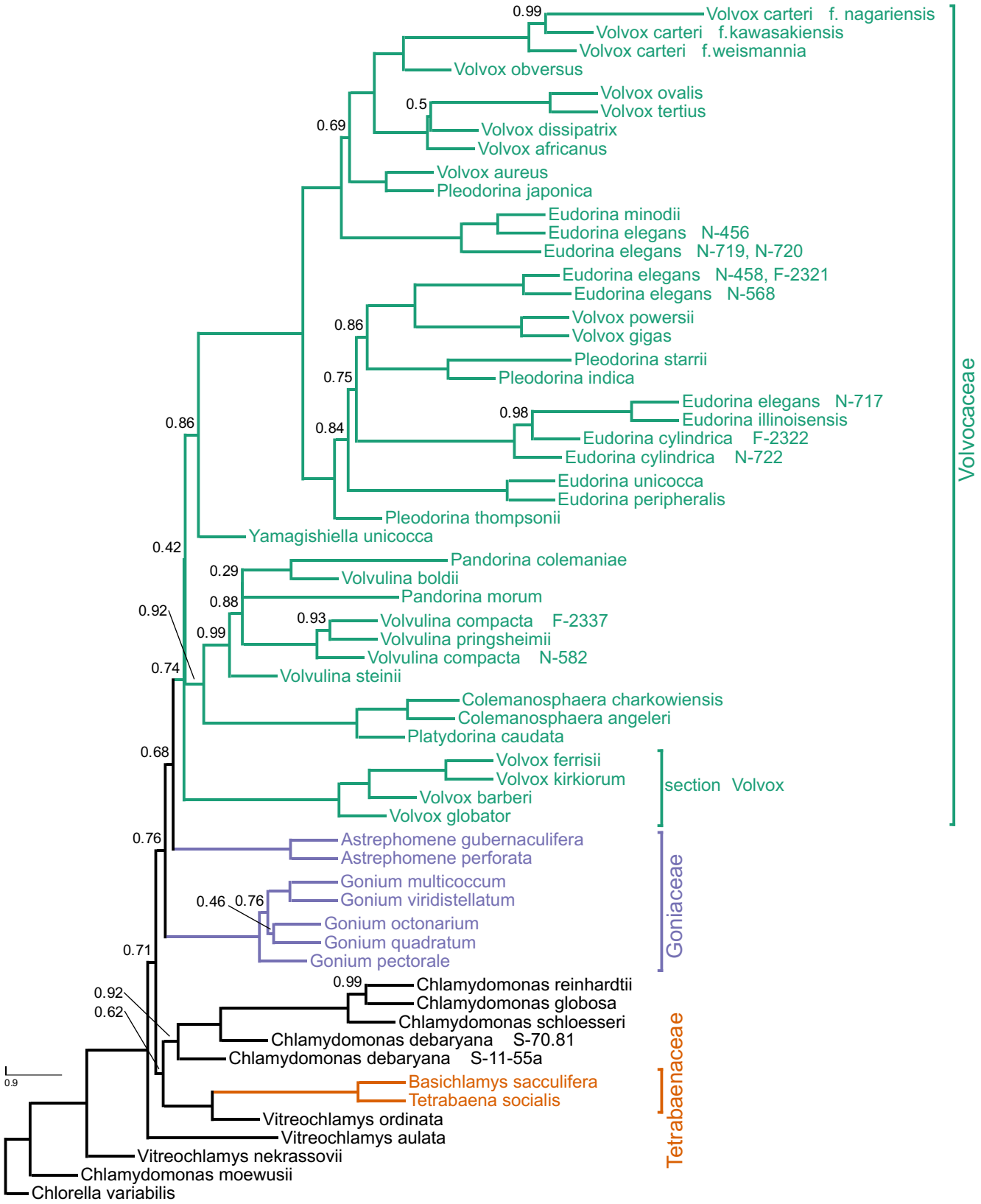
Volvocaceae
Volvox carteri f. nagariensis
Volvox carteri f.kawasakiensis
Volvox carteri f.weismannia
Volvox obversus
Volvox ovalis
Volvox tertius
Volvox dissipatrix
Volvox africanus
Volvox aureus
Pleodorina japonica
Eudorina minodii
Eudorina elegans N-456
Eudorina elegans N-719, N-720
Eudorina elegans N-458, F-2321
Eudorina elegans N-568
Volvox powersii
Volvox gigas
Pleodorina starrii
Pleodorina indica
Eudorina elegans N-717
Eudorina illinoisensis
Eudorina cylindrica F-2322
Eudorina cylindrica N-722
Eudorina unicocca
Eudorina peripheralis
Pleodorina thompsonii
Yamagishiella unicocca
Pandorina colemaniae
Volvulina boldii
Pandorina morum
Volvulina compacta F-2337
Volvulina pringsheimii
Volvulina compacta N-582
Volvulina steinii
Colemanosphaera charkowiensis
Colemanosphaera angeleri
Platydorina caudata
Volvox ferrisii
Volvox kirkiorum
Volvox barberi
Volvox globator
section Volvox
Goniaceae
Astrephomene gubernaculifera
Astrephomene perforata
Gonium multicoccum
Gonium viridistellatum
Gonium octonarium
Gonium quadratum
Gonium pectorale
Tetrabaenaceae
Chlamydomonas reinhardtii
Chlamydomonas globosa
Chlamydomonas schloesseri
Chlamydomonas debaryana S-70.81
Chlamydomonas debaryana S-11-55a
Basichlamys sacculifera
Tetrabaena socialis
Vitreochlamys ordinata
Vitreochlamys aulata
Vitreochlamys nekrassovii
Chlamydomonas moewusii
Chlorella variabilis
0.99
0.5
0.69
0.86
0.75
0.98
0.84
0.86
0.42
0.29
0.88
0.93
0.99
0.92
0.74
0.68
0.76
0.46
0.76
0.71
0.99
0.92
0.62
0.9

Figure 3. Phylogeny of the volvocine algae inferred using a coalescent-based approach through ASTRAL. This tree was inferred using our 40 single-gene phylogenies. Numbers on branches represent posterior probabilities (support values not shown are CPP = 1.0). Members of the Tetrabaenaceae, Goniaceae, and Volvocaceae are denoted in orange, purple, and green, respectively; unicellular species are denoted in black.

In each analytical framework, the Tetrabaenaceae was found to be sister to *Vitreochlamys ordinata* (Maximum Likelihood Bootstrap (MLBS) =100, Bayesian Posterior Probabilities (BPP)=1.0, Coalescent Posterior Probabilities (CPP)=1.0 rather than to the Goniaceae + Volvocaceae. A sister relationship between the Tetrabaenaceae and *V. ordinata* was observed in 39% of our single-gene phylogenies, and appears in 74% of our 4-taxa, unrooted, single-gene phylogenies (**Figure 4**). However, the maximum likelihood and Bayesian inference analyses differ from the coalescent-based analysis in the phylogenetic position of the Tetrabaenaceae. In the former, the Tetrabaenaceae are sister to *V. ordinata*, and Tetrabaenaceae + *V. ordinata* is sister to a clade containing several species of *Chlamydomonas* (including *C. reinhardtii*), Goniaceae, and Volvocaceae (MLBS=100, BPP=1.0) (**Figure 2**). In the coalescent-based analysis the Tetrabaenaceae + *V. ordinata* are sister to several species of *Chlamydomonas* (CPP=0.62) in a clade that is sister to the Goniaceae + Volvocaceae (CPP=0.71) (**Figure 3**). Although the branching order among *C. reinhardtii*, Tetrabaenaceae + *V. ordinata*, and Goniaceae + Volvocaceae differs among analyses, all three indicate that the colonial volvocine algae are not monophyletic. This implies independent origins of multicellularity in the Tetrabaenaceae and in the Goniaceae + Volvocaceae. It should be noted that when *Vitreochlamys ordinata* was included among the colonial taxa an Approximately Unbiased test failed to reject a constrained tree whose branching order showed the colonial volvocine algae to be monophyletic (p=0.107). However, when *V. ordinata* was excluded, a constrained tree showing monophyly

among colonial taxa was strongly rejected (p=5.46e-40) (**Supplementary Materials: Figure S1A**).

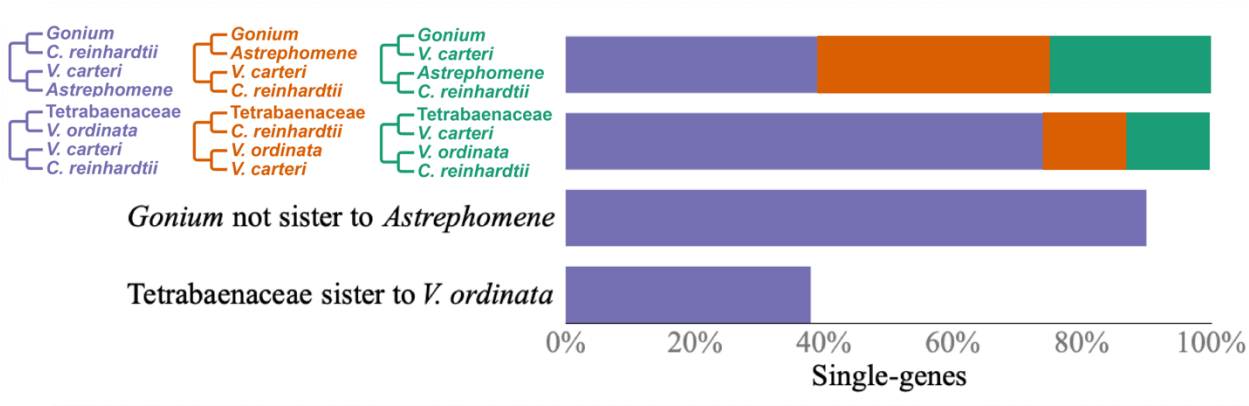

Figure 4. Phylogenetic relationships between *Gonium* + *Astrephomene* and the Tetrabaenaceae + *Vitreochlamys ordinata*. All single-gene phylogenies were inferred using RAxML under the appropriate evolutionary model as estimated by ProtTest. The percentage of single-gene phylogenies representing a specific four-taxon, unrooted tree is represented by the purple, orange, and green shading. For all relationships involving *V. ordinata*, 39 out of 40 single-gene phylogenies were used due to *V. ordinata* not appearing in one of the inferences.

Gonium and Astrephomene are not monophyletic. Our results indicate that, contrary to prior reports (Hanschen *et al*., 2018; Herron *et al*., 2009; Herron & Michod, 2008; Nakada *et al.*, 2019; Nozaki *et al.*, 1996; Nozaki, 1996; Nozaki *et al.*, 1997, 2000; Nozaki & Itoh, 1994; Pröschold *et al.*, 2018), *Astrephomene* is not sister to *Gonium* (**Figures 2** and **3**), and therefore the family Goniaceae is not monophyletic. This inference is strengthened by observations that 90% of our single-gene phylogenies show that *Gonium* and *Astrephomene* are not sister taxa, as do 64% of our four-taxon, unrooted phylogenies (**Figure 4**). All three of our analyses indicate that *Astrephomene* is monophyletic and sister to the Volvocaceae clade (MLBS=98, BPP=1.0, CPP=0.68), with *Gonium* sister to *Astrephomene* + Volvocaceae (MLBS=100, BPP=1.0, CPP=0.76). Furthermore, we performed an Approximately Unbiased test where the monophyly

of the Goniaceae was tested against our finding of paraphyly for the Goniaceae. The null hypothesis, monophyly of the Goniaceae, was rejected (p=0.0446) (**Table S5B**).

***Volvox* section *Volvox* is sister to the remaining Volvocaceae.** Our analyses indicate that *Volvox* section *Volvox* is not a subclade within either the *Pandorina* + *Volvulina* + *Colemanosphaera* (PVC) or *Eudorina* + *Volvox* +*Pleodorina* (EVP) subclades, as suggested by previous analyses (Coleman, 1999; Hanschen *et al*., 2018; Nozaki *et al*., 2000, 2014, 2015; Nozaki & Itoh, 1994; Pröschold *et al*., 2018). Rather, all of our analyses indicate that section *Volvox* is monophyletic and sister to the remaining Volvocaceae (MLBS=87, BPP=1.0, CPP= 0.74). An Approximately Unbiased test strongly rejected the null hypotheses of *Volvox* section *Volvox* being sister to *Colemanosphaera* (p-AU=4.64e-88) or sister to the PVC clade (p-AU=0.0332) (**Supplementary Materials: Figure S1C**).

***Platydorina caudata* is sister to *Colemanosphaera*, and *Pandorina* is paraphyletic with respect to *Volvulina*.** Within the PVC clade, *Colemanosphaera* appears to be monophyletic with high support (MLBS=100, BPP=1.0, CPP=1.0) and sister to *Platydorina* (MLBS=100, BPP=1.0, CPP=1.0) The clade containing *Colemanosphaera* and *Platydorina* is sister to *Pandorina* + *Volvulina* (MLBS=100, BPP=1.0, CPP=0.92). Additionally, *Pandorina* is paraphyletic with respect to *Volvulina* (**Figures 2** and **3**).

The genera Eudorina, Volvox and Pleodorina are polyphyletic. *Yamagishiella unicocca* is sister to the *Eudorina*+*Volvox*+*Pleodorina* (EVP) clade, which encompasses two large subclades (MLBS=98, BPP=1.0, CPP=0.86) (**Figures 2** and **3**). Moreover, the genus *Volvox* appears to be

polyphyletic, with members represented across the two EVP subclades and the section *Volvox* clade. Members of both the *Pleodorina* and *Eudorina* genera are suggested to be polyphyletic across the PVC and EVP subclades.

Unicellular taxa are nested within the clade containing the colonial volvocine algae + Chlamydomonas reinhardtii. Of the unicellular taxa, *Chlamydomonas debaryana, C. globosa*, *C. reinhardtii*, *C. schloesseri*, and *Vitreochlamys ordinata* are nested within the clade containing the colonial volvocine algae. The closest unicellular relative to the clade that contains the colonial algae + *C. reinhardtii* is suggested to be *V. aulata* (**Figures 2** and **3**). By contrast, the unicellular taxon *V. ordinata* is suggested to be sister to the Tetrabaenaceae (MLBS=100, BPP=1.0, CBB=1.0). As noted above, the maximum likelihood and Bayesian inference analyses show *V. ordinata* + Tetrabaenaceae to be sister to *C. reinhardtii* + Goniaceae + Volvocaceae (MLBS=100, BPP=1.0). In contrast, the coalescent-based analysis shows *V. ordinata* + Tetrabaenaceae + *C. reinhardtii* to be sister to Goniaceae + Volvocaceae (CPP=0.71). All analyses indicate that *C. debaryana* SAG 11-55a is sister to *C. debaryana* SAG 80.81 + *C. schloesseri*, *C. globosa*, and *C. reinhardtii* (MLBS=85, BPP=1.0, CBB=0.92). Consistent with a prior study, *C. moewusii* appears to be more distantly related to the colonial volvocines than is *Vitreochlamys nekrassovii* (Herron & Michod, 2008).

Discussion

Why revisit the volvocine algal phylogeny? Prior phylogenetic analyses of this evolutionarily important group have often been based on few genes and, in some cases, limited taxon sampling. Perhaps it is therefore not surprising that phylogenies of the volvocine algae are not concordant

(**Figure 1**). Here, we offer what is currently the most taxonomically comprehensive phylogeny of the volvocine algae inferred using a multi-gene nuclear dataset. To overcome past limitations, we utilized two phylogenetic inference approaches: a supermatrix analysis using a partitioned dataset and a coalescent-based analysis. Our supermatrix analysis consisted of a concatenated dataset of 40 single-copy nuclear, protein coding genes as determined by De Clerck and colleagues (De Clerck *et al*., 2018); this dataset was partitioned in order to apply the appropriate evolutionary model to each specific gene within the supermatrix. Partitioning allows for the use of multiple substitution models, averting the risk of inaccurately representing evolution of the "supergene." Additionally, a coalescent-based analysis was conducted using ASTRAL to further counteract another limitation of a supermatrix analysis: the possibility that single-gene trees may yield conflicting evolutionary inferences. Lastly, because the use of *Chlamydomonas reinhardtii* as an outgroup for the volvocine algae has recently been called into question (Nakada *et al.*, 2019; Pröschold *et al*., 2018), we used *Chlorella variabilis*. Because it is a member of the Trebouxiophyceae, the sister group to the Chlorophyceae, *C. variabilis* should not bias our phylogenetic inferences.

Our results conflict with prior volvocine algal phylogenies in four respects. Our analyses lead us to several conclusions that differ from prior phylogenies. *First*, we find that members of the Tetrabaenaceae are more closely related to some unicellular species than they are to the Goniaceae and Volvocaceae. *Second*, monophyly of the family Goniaceae is not supported. *Third*, section *Volvox* is suggested to be sister to the remaining Volvocaceae. *Fourth*, cellular differentiation independently arose at least four and perhaps as many as six times.

All three of our phylogenetic analyses indicate that the colonial volvocines are not monophyletic (**Figures 2** and **3**). This finding represents a major departure from earlier chloroplast gene-based volvocine phylogenies (Hanschen *et al*., 2018; Herron *et al*., 2009; Herron & Michod, 2008; Hu et al., 2019; Nakada *et al*., 2010; Nakazawa *et al.*, 2001; Nozaki *et al*., 2000, 2014), phylogenies based on morphological characters (Nozaki, 1996; Nozaki & Itoh, 1994), phylogenies inferred using ITS 1 and 2 sequences (Coleman, 1999), as well as less taxonomically comprehensive phylogenies inferred using nuclear data (Zhang *et al.*, 2019), all of which suggest that the colonial volvocine algae are monophyletic. Consistent with Pröschold *et al.* (2018) and the single-gene phylogeny inferred by Nakada *et al*. (2019), our results support the view that multicellularity independently evolved at least twice within this group.

The Tetrabaenaceae, which exhibits the simplest form of multicellularity, is more closely related to certain unicellular species than it is to the Goniaceae or Volvocaceae. Our results differ in key respects from a recent volvocine algae phylogeny offered by Zhang *et al*. (2019), which like ours is based on single-copy nuclear genes. Zhang *et al*. (2019) sought to understand the evolutionary relationships between two psychrophilic algae: *Chlamydomonas* sp. ICE-L and *Tetrabaena socialis* N-691. To do so they constructed a phylogeny consisting of ICE-L, N-691, three colonial *Volvox* strains, and eight unicellular species, including *C. reinhardtii*.

We further investigated this lack of concordance between our findings and those of Zhang *et al*. (2019). First, we confirmed that *T. socialis* N-691 and *T. socialis* N-571 are conspecific. Nozaki and Ohtani (1992) examined *T. socialis* N-691 and found that N-691 was morphologically identical to *T. socialis* N-571 except for forming slightly larger colonies. Next, we compared ITS-1, 5.8S rRNA, and ITS-2 sequences from Mai & Coleman (1997) between the

two *T. socialis* strains; we found that they were 92% identical. Then, we performed a BLAST search of the aforementioned sequences: the first ten BLAST hits all matched the same sequences from *T. socialis* strains with 100% query coverage. Using the BLAST software, neighbor joining and fast minimum evolution distance trees were inferred from the BLAST results; the trees produced show *T. socialis* strains N-571 and N-691 to be reciprocally monophyletic. Once we confirmed that N-691 and N-571 were conspecific, we were able to replicate the branching order produced by Zhang *et al*. (2019) using our concatenated 40-gene dataset (**Supplementary Materials: Figure S2A**). For our initial tree, we sampled our strains of *Chlamydomonas reinhardtii*, *C. moewusii*, *T. socialis*, *Volvox aureus*, *V. carteri* f. *nagariensis*, and *V. globator* to match taxa that were used in that study. For an outgroup species we sampled *Chlorella variabilis*. Multiple studies have shown that the accuracy of phylogenetic reconstruction can be improved by increasing the number of taxa sampled (Hedtke *et al.*, 2006; Pollock *et al.*, 2002; Zwickl *et al.*, 2002). When we added more taxa and performed Maximum Likelihood analysis on the new dataset, the three colonial volvocine families were no longer monophyletic. The Tetrabaenaceae formed a clade with *Vitreochlamys ordinata*, and this clade appeared sister to *C. reinhardtii* + Goniaceae + Volvocaceae. (**Supplementary Materials: Figure S2B**). These results indicate that the finding that *T. socialis* N-691 is sister to the Volvocaceae is an artefact of limited taxonomic sampling.

Approximately Unbiased (AU) tests were performed to test monophyly of the colonial volvocine algae. When *Vitreochlamys ordinata* was included among the colonial taxa in a constraint tree, the constrained tree (**Supplementary Materials: Figure S1A T2**) showing monophyly of the colonial taxa was not rejected (p=0.107). However, when *V. ordinata* was excluded among the colonial taxa in the constraint tree, (**Supplementary Materials: Figure**

S1A T3) monophyly of the colonial taxa was strongly rejected (p=5.46e-40). This contradiction makes it impossible to claim unequivocally that the multicellular taxa are polyphyletic.

The family Goniaceae is not monophyletic. Multiple volvocine phylogenies have concluded that the Goniaceae is monophyletic (Hanschen *et al.*, 2018a, 2018b; Herron *et al.*, 2009; Herron & Michod, 2008; Nakada *et al.*, 2019; Nozaki, 1996; Nozaki *et al.*, 1996, 1997, 2000; Nozaki & Itoh, 1994; Pröschold *et al.*, 2018). Our analyses suggest otherwise (**Figure 2** and **3**): we find that the Goniaceae is not monophyletic, and specifically that *Astrephomene* is not sister to *Gonium*. Instead, we find that *Astrephomene* is sister to the Volvocaceae (MLBS= 98, BPP= 1.0, CBPP= 0.68). We performed an AU test to test monophyly in this family. A constrained tree in which the Goniaceae are monophyletic (**Supplementary Materials: Figure S1B T2**) was rejected (p=0.0446) when tested against our unconstrained phylogeny showing the Goniaceae to be non-monophyletic (p=0.955) (**Supplementary Materials: Figure S1B T1**). This finding is further supported by the observation that in 36 of 40 single-gene trees *Astrephomene* and *Gonium* are not sister to one another (**Figure 4**).

We should note that prior attempts to ascertain whether or not the Goniaceae are monophyletic have produced mixed results, sometimes with low support values for the relevant relationships. Nozaki and colleagues (1995) published four phylogenies inferred using a single chloroplast gene and different inference methods; all four trees either showed low support for monophyly of the Goniaceae or suggested a topology where *Astrephomene* is sister to Goniaceae + Volvocaceae. Coleman (1999) inferred a volvocine phylogeny based on ITS-1 and ITS-2 sequences that showed *Astrephomene* sister to Tetrabaenaceae + Goniaceae + Volvocaceae;

however, the bootstrap support for this suggested relationship was listed as "above the 50% level" in a range of 50-75%, indicating weak support for the branching order.

Other phylogenies suggesting monophyly in the Goniaceae do so with weak or contradictory support (Nakada *et al*., 2019; Nozaki *et al*., 2000; Pröschold *et al.*, 2018). Nozaki *et al*. (2000) inferred a volvocine phylogeny using 5 chloroplast genes; the neighbor joining tree that sampled *Astrephomene* and *Gonium* reports two bootstrap support values from majority-rule and ML analyses: 73 and 74, respectively. Neither qualifies as strong support for monophyly of the Goniaceae. Pröschold *et al.* (2018) inferred phylogenies from three datasets: (*i*) SSU rDNA + ITS-1 and ITS-2 sequences, (*ii*) ITS-1 and ITS-2 sequences, and (*iii*) the sequence of *rbcL,* the large subunit of RuBisCo. The phylogeny inferred using the first reported monophyly of the Goniaceae to be "moderately supported" by bootstrap values of 81, 54, and 63 from maximum likelihood analyses produced, respectively, by PAUP, RAxML and a PAUP neighbor-joining analysis. The tree inferred using the second dataset showed that monophyly of the Goniaceae to be "weakly supported in bootstrap and Bayesian analyses." The *rbcL*-based phylogeny suggested that *Astrephomene* and *Gonium* were not monophyletic. Instead, *Astrephomene* was sister to *Volvulina steinii*, and together these two were sister to the other Volvocacean taxa + Tetrabaenaceae + Goniaceae + a *Chlamydomonas debaryana* strain (referred to as *Edaphochlamys debaryana* in Pröschold *et al.* (2018)). Finally, using a dataset consisting of 5 core chloroplast genes and 18S rRNA, Nakada *et al.* (2019) reported strong support (BPP=1.0) for monophyly of the Goniaceae by Bayesian analysis, but weak support for monophyly of *Astrephomene* and *Gonium* by ML analysis (MLBS=65).

Our data indicate that at least one of three main colonial volvocine families is not monophyletic. Moreover, this finding is consistent with some – but not all – of the analyses

recently reported by Pröschold *et al.* (2018) and Nakada *et al*. (2019). That said, past morphological and ultrastructural studies suggesting a close relationship between *Astrephomene* and *Gonium* (Nozaki, 1990; Nozaki & Itoh, 1994; Nozaki & Kuroiwa, 1992) should not be disregarded. *Astrephomene* and *Gonium*, along with the Tetrabaenaceae, are unique among the colonial volvocines in that each cell, rather than the entire colony, is surrounded by a tripartite boundary, the most conserved morphological feature of the Volvocaceae (Kirk *et al.*, 1986). This feature distinguishes their mode of colony formation from other colonial algae within the Volvocaceae.

Section Volvox is suggested to be sister to the remaining Volvocaceae. A third major difference between our results and prior volvocine phylogenies concerns the position of the section *Volvox* clade. Older studies based on the *rbcL* chloroplast gene (Nozaki *et al*., 1996), ITS-1 and ITS-2 sequences (Coleman, 1999), and morphology (Nozaki and Itoh, 1994) suggest that section *Volvox* belongs to a clade that encompasses *Eudorina*, *Pleodorina*, and other *Volvox* species. More recent studies of the volvocine algae based on 5 chloroplast genes or based on multiple datasets that include 1 chloroplast gene (Pröschold *et al*. (2018) suggest that the section *Volvox* belongs to a clade that includes *Pandorina*, *Volvulina*, *Platydorina* (Hanschen *et al.*, 2018a; Herron & Michod, 2008), and (in the studies where it was also studied) *Colemanosphaera* (Hu *et al.*, 2019; Nozaki *et al.*, 2014). Our results, however, suggest with moderate support (MLBS=86, BPP=1.0, CPP=0.74) that section *Volvox* is sister to the remaining Volvocaceae (**Figures 2** and **3**). AU tests rejected both section *Volvox* being sister to *Colemanosphaera* + *Platydorina* (p=4.64e-88) (**Supplementary Materials: Figure S1C T2**) and section *Volvox* being sister to the PVC clade (p=0.0332) (**Supplementary Materials: Figure S1C T3**), when tested against

our unconstrained phylogeny (p=0.967) (**Supplementary Materials: Figure S1C T1**). These results, along with moderate support from our phylogenetic support values (MLBS=86, BPP=1.0, CPP=0.74) bolster our finding that section *Volvox* is sister to the remaining Volvocaceae (**Figures 2** and **3**).

Cellular differentiation independently arose at least four times in the volvocine algae. The last major difference between our results and earlier phylogenies concerns the number of independent origins of cellular differentiation. Prior literature suggests that cellular differentiation independently evolved at least three times: once in *Astrephomene*, once in section *Volvox,* and at least once in the EVP clade (Grochau-Wright *et al.*, 2017; Herron & Michod, 2008). By contrast, our results show a *minimum* of four independent origins of cellular differentiation: once in *Astrephomene*, once in section *Volvox*, and at least twice in the EVP clade (**Figure 5A**). And we cannot exclude the possibility of two additional independent origins in the branches leading to *Pleodorina starrii* and *Volvox gigas* (**Figure 5A**). In *Astrephomene*, section *Volvox, Pleodorina*, and *Volvox dissipatrix*, differentiated cells carry out the function of motility, whereas undifferentiated cells participate in both motility and reproduction (Herron *et al.*, 2009, 2010). The remaining *Volvox* species within the EVP clade have all evolved specialized germ cells for reproduction and somatic cells for motility (Herron *et al*., 2009, 2010).

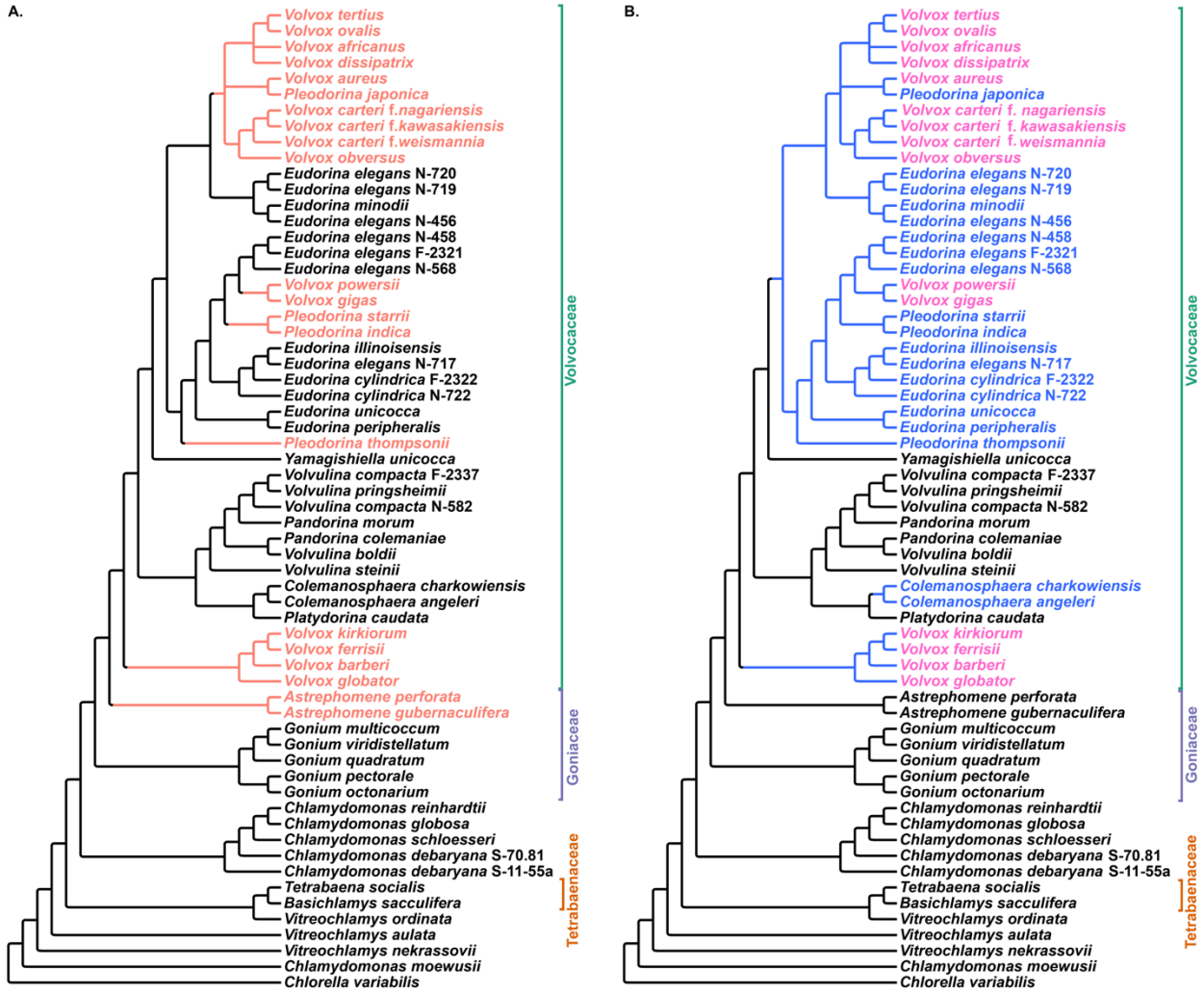

Figure 5. (A) Phylogeny of the volvocine algae highlighting the lineages in which soma differentiation has evolved (peach). This tree indicates a minimum of four and maximum of six independent origins of cellular differentiation **(B)** Phylogeny of the volvocine algae highlighting the lineages that are isogamous (black), anisogamous (blue), and oogamous (names in pink only). Both phylogenies were inferred from a dataset of single-copy, protein-coding nuclear genes (12,650 aligned amino acid positions of 68 taxa) inferred using the maximum likelihood method in IQtree.

Our results reinforce several prior phylogenetic inferences about the volvocine algae.

ISOGAMY IS THIS GROUP'S ANCESTRAL STATE. Consistent with past studies, our results suggest that isogamy, the production of similar sized, motile gametes, is the ancestral mode of sexual reproduction among the volvocine algae (**Figure 5B** and **Supplementary Materials: Table S2**).

Isogamy is present in the unicellular genera *Chlamydomonas* and *Vitreochlamys*, and is retained within the multicellular genera *Astrephomene*, *Basichlamys*, *Gonium*, *Pandorina*, *Platydorina*, *Tetrabaena*, *Volvulina*, and *Yamagishiella*. By contrast, *Colemanosphaera*, *Eudorina*, *Pleodorina*, and *Volvox* have all evolved either anisogamy or oogamy (Herron, 2016; Nozaki *et al*., 2014; da Silva, 2018; Umen, 2011). Anisogamy appears to have independently evolved at least three times from an isogamous ancestor in section *Volvox* and in both *Colemanosphaera* and EVP. Conventional anisogamy, which consists of two motile gamete types of unequal size, appears in *Colemanosphaera*, *Eudorina*, and *Pleodorina.* Oogamy, a specialized form of anisogamy where the female gamete is immotile and significantly larger than the motile, male gamete, is suggested to have independently evolved at least three times in lineages leading to section *Volvox*, *V. gigas* + *V. powersii*, and in the clade containing *V. africanus*, *V. aureus*, *V. carteri*, *V. dissipatrix*, *V. obversus*, *V. ovalis*, and *V. tertius* (Herron, 2016; Nozaki *et al*., 2014; da Silva, 2018). This latter finding differs from Hanschen *et al.* (2018b) who reported at least two independent origins of oogamy among the volvocine algae. Our results clearly suggest an independent origin of oogamy in section *Volvox*. Due to the branching order of the Hanschen *et al*. (2018b) phylogeny, however, we cannot be sure that the most recent common ancestor of *V. gigas* and *V. powersii* and *V. carteri* and other *Volvox* species within that clade was not oogamous.

THE VOLVOCACEAE, WITH THE EXCEPTION OF SECTION VOLVOX, IS MONOPHYLETIC. Results from our nuclear multi-protein dataset are congruent with previous studies (Coleman, 1999; Grochau-Wright *et al*., 2017; Hanschen *et al*., 2018a; Herron *et al.*, 2009; Herron & Michod, 2008; Hu *et al*., 2019, 2020; Nakada *et al*., 2010, 2019; Nozaki *et al*., 2000, 2014; Nozaki & Itoh, 1994; Pröschold *et al*., 2018) in indicating monophyly of the family Volvocaceae

as well as its division, with the exception of the section *Volvox*, into two principal clades: PVC and EVP. Within the PVC clade, our results add further support to the view that the genera *Pandorina* and *Volvulina* are not reciprocally monophyletic (Coleman, 1999, 2001; Hanschen *et al*., 2018b; Herron *et al*., 2009; Herron & Michod, 2008; Nozaki *et al*., 2000, 2014). Also, consistent with other multi-gene analyses *Colemanosphaera* appears sister to *Platydorina* (Hanschen *et al*., 2018a; Nozaki *et al*., 2014, 2015).

A NUMBER OF VOLVOCINE GENERA ARE NOT MONOPHYLETIC. Our results also support prior work suggesting that the genera *Volvox*, *Eudorina* and *Pleodorina* are not monophyletic (Coleman, 1999; Grochau-Wright *et al.*, 2017; Hanschen *et al.*, 2018a, 2018b; Herron *et al.*, 2009, 2010; Herron & Michod, 2008; Nozaki, 2003; Nozaki *et al.*, 2000, 2006, 2014, 2015). Historically, the genus *Volvox* has been divided into 4 sections – *Copelandosphaera*, *Janetosphaera*, *Merrillosphaera*, and *Volvox* – based on morphological (Smith, 1944) and molecular data (Nozaki, 2003). A recent section-level revision of the genus *Volvox* (Nozaki *et al*., 2015) resulted in the creation and deletion of sections *Besseyosphaera* and *Copelandosphaera*, respectively. Hereafter, we will only refer to the revised taxonomic sections proposed by Nozaki *et al*. (2015), with which our Bayesian inference and coalescent-based results are in agreement (**Supplementary Materials: Figure S3**). Our Bayesian inference and coalescent-based analyses further suggest that each of the four sections is monophyletic, and that none encompass novel taxa not listed by Nozaki *et al*. (2015). The branching order of our ML analysis, however, suggests that section *Merrillosphaera* is not monophyletic (**Supplementary Materials: Figure S3A**), although the support values for the relevant relationships are low. Specifically, our ML analysis suggests that *V. africanus*, *V. dissipatrix*, *V. ovalis*, and *V. tertius* form a clade with *V. aureus* and *P. japonica* that is separate from the other *Merrillosphaera* taxa

(MLBS=59). In contrast, our BI and CB analyses suggest with moderate support (BPP=0.50, CPP=0.69) that the *Merrillosphaera* species are monophyletic (**Supplementary Materials: Figure S3B**). Heeding our support values rather than only the branching order, we propose that the taxonomic system of the genus *Volvox* as outlined by Nozaki and colleagues (2014) be retained.

Regarding unicellular species, our results confirm prior studies showing the genus *Vitreochlamys* to be polyphyletic (Nakada *et al*., 2019; Nakazawa *et al*., 2001). Both our aggregate phylogenies and 15/39 of our single-gene phylogenies indicate that *V. ordinata* is sister to the Tetrabaenaceae (**Figure 4**), consistent with Nakazawa *et al.* (2001). This last point, along with our findings that *V. aulata* is the closest unicellular relative to the clade containing the colonial volvocine algae + multiple *Chlamydomonas* species (including *C. reinhardtii*), suggests that at least some members of *Vitreochlamys* are very closely related to the colonial volvocine algae. This relationship had been previously suggested by other studies (Nakada *et al.*, 2019; Nozaki *et al.*, 1999) including Nakazawa *et al.* (2001), whose ultrastructural studies uncovered striking similarities in how these taxa formed pyrenoids and eyespot apparati (stigma), and established their tripartite cell walls.

Chlamydomonas is a polyphyletic genus (Craig *et al.*, 2021; Nakada *et al.*, 2008, 2019; Pröschold *et al.*, 2001) composed of at least 500 species (Pröschold *et al*., 2001). Although we sampled only a handful of *Chlamydomonas* species, our data support this view and broadly agree with the *Chlamydomonas* relationships inferred by Pröschold and colleagues (2018), who used a combination of molecular phylogenetic analyses, sporangium wall lysis tests, and ultrastructural analyses. Our data strongly support *C. schloesseri* being sister to *C. reinhardtii* + *C. globosa* (MLBS=100, BPP=1.0, CPP=1.0) and designating *C. schloesseri* as a "true" *Chlamydomonas*

species, as suggested by Pröschold *et al*. (2018). Our study is also in agreement with a recent study by Craig *et al.* (2021) that shows *C. schloesseri* being sister to *C. reinhardtii* + *C. globosa*. Also, like Pröschold *et al*. (2018), our analyses indicate that *C. debaryana* SAG 70.81 is sister to *Chlamydomonas schloesseri* and its relatives (MLBS=100, BPP= 1.0, CPP=1.0). However, unlike the Pröschold *et al.* (2018) study which proposed that strain *C. debaryana*/*Edaphochlamys debaryana* (SAG 11-55a) is sister to the Tetrabaenaceae, our analyses support the view that *C. debaryana*/*Edaphochlamys debaryana* is more closely related to *C. reinhardtii* (MLBS=84, BPP=1.0, CPP=0.92) than to the colonial algae. Our finding is further supported by Craig *et al.* (2021) who inferred that *C. debaryana*/*Edaphochlamys debaryana* + *Chlamydomonas sphaeroides* is sister to the clade containing *C. schloesseri* + *C. reinhardtii* + *C. globosa*. Our placement of *C. debaryana* (SAG 11-55a) could be a result of limited (N=6) sampling within the *Chlamydomonas* genus, which was more extensively sampled by Pröschold *et al*. (2018) (N>30).

Unresolved issues and future directions. Our phylogeny of the *Chlamydomonas* clade calls into question the position of the Tetrabaenaceae, the multicellular volvocine algae that have the smallest cell number (4 cells per colony). Despite the high support from all phylogenetic analyses for the Tetrabaenaceae + *V. ordinata* being sister to the clade containing *C. reinhardtii* + the two other colonial families (MLBS=100, BPP=1.0), we recognize that moderate support for *C. reinhardtii* being sister to the other two colonial clades (MLBS=77, BPP=1.0) limits our ability to fully resolve the position of the Tetrabaenaceae. To clarify the position of this colonial family we suggest two approaches. *First*, we recommend using a larger orthologous nuclear gene dataset to better resolve their position. *Second*, additional unicellular taxa should be sampled from VLE groups 1 and 2 (Pröschold *et al*., 2018). The latter recommendation would not only

allow for better resolution of those *Chlamydomonas* species, it would also serve to resolve whether or not the Tetrabaenaceae is sister to *C. debaryana*, as suggested by Pröschold *et al.* (2018). We have demonstrated here that increased taxon sampling can affect branching order in a volvocine phylogeny (**Supplementary Materials: Figure S2**). By sampling even more unicellular taxa, specifically within the genus *Chlamydomonas*, better phylogenetic resolution could likely be achieved.

Future research directions that can be taken using our transcriptomic dataset include ancestral-state reconstruction of traits involving cellularity, cellular differentiation, and gamete size as well as discerning the evolutionary history of gene families across the volvocine algae as a whole and within its major clades.

Conclusions

The volvocine algae, which include the single-celled species *Chlamydomonas reinhardtii* and the colonial species *Volvox carteri*, serve as a model for studying the evolution of multicellularity and cellular differentiation. Speculations as to how these traits evolved in this group have often been based on phylogenetic inferences limited to one or a few genes. Using a 40-protein dataset, we have shown that the Tetrabaenaceae and the Goniaceae + Volvocaceae represent two independent origins of multicellularity, and that cellular differentiation has independently evolved at least four, and possibly six, times within this group. Our results suggest that both multicellularity and cellular differentiation are evolutionary labile traits within the volvocine algae. We have established a robust phylogeny of this group, which we hope will assist future efforts aimed at re-evaluating ancestral character states and understanding the origins of multicellularity and cellular differentiation in the volvocine green algae.

Data Archiving

All raw data generated and used for this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA701495. Accession numbers for our raw RNA-Seq reads range from SAMN17884613 to SAMN17884667. For detailed information regarding accession number assignment to a specific taxon please refer to **Table S5** in **Supplementary Materials.**

Literature cited

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brunet, T., & King, N. (2017). The Origin of Animal Multicellularity and Cell Differentiation. *Developmental Cell*, *43*(2), 124–140. https://doi.org/10.1016/j.devcel.2017.09.016

Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, *17*(4), 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chen, L., Xiao, S., Pang, K., Zhou, C., & Yuan, X. (2014). Cell differentiation and germ–soma separation in Ediacaran animal embryo-like fossils. *Nature*, *516*(7530), 238–241. https://doi.org/10.1038/nature13766

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, *65*(6), 997–1008. https://doi.org/10.1093/sysbio/syw037

Coleman, A. W. (1999). Phylogenetic analysis of "Volvocacae" for comparative genetic studies. *Proceedings of the National Academy of Sciences*, *96*(24), 13892–13897. https://doi.org/10.1073/pnas.96.24.13892

Coleman, A. W. (2001). Biogeography and speciation in the *Pandorina*/*Volvulina* (chlorophyta) superclade. *Journal of Phycology*, *37*(5), 836–851. https://doi.org/10.1046/j.1529-8817.2001.01043.x

Craig, R. J., Hasan, A. R., Ness, R. W., & Keightley, P. D. (2021). Comparative genomics of *Chlamydomonas*. *The Plant Cell*. https://doi.org/10.1093/plcell/koab026

De Clerck, O., Kao, S.-M., Bogaert, K. A., Blomme, J., Foflonker, F., Kwantes, M., Vancaester, E., Vanderstraeten, L., Aydogdu, E., Boesger, J., Califano, G., Charrier, B., Clewes, R., Del Cortona, A., D'Hondt, S., Fernandez-Pozo, N., Gachon, C. M., Hanikenne, M., Lattermann, L., … Bothwell, J. H. (2018). Insights into the Evolution of Multicellularity from the Sea Lettuce Genome. *Current Biology*, *28*(18), 2921-2933.e5. https://doi.org/10.1016/j.cub.2018.08.015

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Featherston, J., Arakaki, Y., Hanschen, E. R., Ferris, P. J., Michod, R. E., Olson, B. J. S. C., Nozaki, H., & Durand, P. M. (2018). The 4-Celled *Tetrabaena socialis* Nuclear Genome Reveals the Essential Components for Genetic Control of Cell Number at the Origin of Multicellularity in the Volvocine Lineage. *Molecular Biology and Evolution*, *35*(4), 855–870. https://doi.org/10.1093/molbev/msx332

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Grochau-Wright, Z. I., Hanschen, E. R., Ferris, P. J., Hamaji, T., Nozaki, H., Olson, B. J. S. C., & Michod, R. E. (2017). Genetic basis for soma is present in undifferentiated volvocine green algae. *Journal of Evolutionary Biology*, *30*(6), 1205–1218. https://doi.org/10.1111/jeb.13100

Grosberg, R. K., & Strathmann, R. R. (2007). The Evolution of Multicellularity: A Minor Major Transition? *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 621–654. https://doi.org/10.1146/annurev.ecolsys.36.102403.114735

Hallmann, A. (2011). Evolution of reproductive development in the volvocine algae. *Sexual Plant Reproduction*, *24*(2), 97–112. https://doi.org/10.1007/s00497-010-0158-4

Han, T., & Runnegar, B. (1992). Megascopic eukaryotic algae from the 2.1-billion-year-old negaunee iron-formation, Michigan. *Science*, *257*(5067), 232–235. https://doi.org/10.1126/science.1631544

Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., & Michod, R. E. (2018a). Repeated evolution and reversibility of self-fertilization in the volvocine green algae*. *Evolution*, *72*(2), 386–398. https://doi.org/10.1111/evo.13394

Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., & Michod, R. E. (2018b). Multicellularity Drives the Evolution of Sexual Traits. *The American Naturalist*, *192*(3), E93–E105. https://doi.org/10.1086/698301

Hanschen, E. R., Marriage, T. N., Ferris, P. J., Hamaji, T., Toyoda, A., Fujiyama, A., Neme, R., Noguchi, H., Minakuchi, Y., Suzuki, M., Kawai-Toyooka, H., Smith, D. R., Sparks, H., Anderson, J., Bakarić, R., Luria, V., Karger, A., Kirschner, M. W., Durand, P. M., … Olson, B. J. S. C. (2016). The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nature Communications*, *7*(1), 1–10. https://doi.org/10.1038/ncomms11370

Hasegawa, M., & Hashimoto, T. (1993). Ribosomal RNA trees misleading? *Nature*, *361*(6407), 23–23. https://doi.org/10.1038/361023b0

Herron, M. D., Borin, J. M., Boswell, J. C., Walker, J., Chen, I.-C. K., Knox, C. A., Boyd, M., Rosenzweig, F., & Ratcliff, W. C. (2019). De novo origins of multicellularity in response to predation. *Scientific Reports*, *9*(1), 2328. https://doi.org/10.1038/s41598-019-39558-8

Herron, M. D., Desnitskiy, A. G., & Michod, R. E. (2010). Evolution of developmental programs in *Volvox* (Chlorophyta). *Journal of Phycology*, *46*(2), 316–324. https://doi.org/10.1111/j.1529-8817.2009.00803.x

Herron, M. D., Hackett, J. D., Aylward, F. O., & Michod, R. E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proceedings of the National Academy of Sciences*, *106*(9), 3254–3258. https://doi.org/10.1073/pnas.0811205106

Herron, M. D., & Michod, R. E. (2008). Evolution of complexity in the volvocine algae: Transitions in individuality through Darwin's eye. *Evolution*, *62*(2), 436–451. https://doi.org/10.1111/j.1558-5646.2007.00304.x

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hu, Y., Xing, W., Song, H., Hu, Z., & Liu, G. (2020). Comparison of colonial volvocine algae based on phylotranscriptomic analysis of gene family evolution and natural selection. *European Journal of Phycology*, *55*(1), 100–112. https://doi.org/10.1080/09670262.2019.1663269

Hu, Y., Xing, W., Song, H., Zhu, H., Liu, G., & Hu, Z. (2019). Evolutionary Analysis of Unicellular Species in Chlamydomonadales Through Chloroplast Genome Comparison With the Colonial Volvocine Algae. *Frontiers in Microbiology*, *10*, 1351. https://doi.org/10.3389/fmicb.2019.01351

Jakob, S. S., & Blattner, F. R. (2006). A Chloroplast Genealogy of Hordeum (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequences for Phylogenetic Inference. *Molecular Biology and Evolution*, *23*(8), 1602–1612. https://doi.org/10.1093/molbev/msl018

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, *26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Kirk, D. L. (2001). Germ–Soma Differentiation in *Volvox*. *Developmental Biology*, *238*(2), 213–223. https://doi.org/10.1006/dbio.2001.0402

Kirk, D. L. (2005). A twelve-step program for evolving multicellularity and a division of labor. *BioEssays*, *27*(3), 299–310. https://doi.org/10.1002/bies.20197

Kirk, D. L., Birchem, R., & King, N. (1986). The extracellular matrix of *Volvox*: A comparative study and proposed system of nomenclature. *J. Cell Sci.*, *80*, 207–231.

Koyanagi, K. O. (2015). Inferring Cell Differentiation Processes Based on Phylogenetic Analysis of Genome-Wide Epigenetic Information: Hematopoiesis as a Model Case. *Genome Biology and Evolution*, *7*(3), 699–705. https://doi.org/10.1093/gbe/evv024

Lang, N. J. (1963). Electron Microscopy of the Volvocaceae and Astrephomenaceae. *American Journal of Botany*, *50*(3), 280–300.

Larson, A., Kirk, M. M., & Kirk, D. L. (1992). Molecular Phylogeny of the Volvocine Flagellates. *Molecular Biology and Evolution*, *9*(1), 85–105.

Larsson, A. (2014). AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, *30*(22), 3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Mai, J. C., & Coleman, A. W. (1997). The Internal Transcribed Spacer 2 Exhibits a Common Secondary Structure in Green Algae and Flowering Plants. *Journal of Molecular Evolution*, *44*(3), 258–271. https://doi.org/10.1007/PL00006143

Matt, G. Y., & Umen, J. G. (2018). Cell-Type Transcriptomes of the Multicellular Green Alga *Volvox carteri* Yield Insights into the Evolutionary Origins of Germ and Somatic Differentiation Programs. *G3: Genes|Genomes|Genetics*, *8*(2), 531–550. https://doi.org/10.1534/g3.117.300253

Maynard Smith, J., & Szathmáry, E. (1995). The major transitions in evolution. Oxford University Press.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Marechal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., … Grossman, A. R. (2007). The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science*, *318*(5848), 245–250. https://doi.org/10.1126/science.1143609

Nakada, T., Misawa, K., & Nozaki, H. (2008). Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. *Molecular Phylogenetics and Evolution*, *48*(1), 281–291. https://doi.org/10.1016/j.ympev.2008.03.016

Nakada, T., Tomita, M., & Nozaki, H. (2010). *Volvulina compacta* (Volvocaceae, Chlorophyceae), new to Japan, and its phylogenetic position. *J. Jpn. Bot.*, *85*, 364–369.

Nakada, T., Tsuchida, Y., & Tomita, M. (2019). Improved taxon sampling and multigene phylogeny of unicellular chlamydomonads closely related to the colonial volvocalean

lineage Tetrabaenaceae-Goniaceae-Volvocaceae (Volvocales, Chlorophyceae). *Molecular Phylogenetics and Evolution*, *130*, 1–8. https://doi.org/10.1016/j.ympev.2018.09.013

Nakazawa, A., Krienitz, L., & Nozaki, H. (2001). Taxonomy of the unicellular green algal genus *Vitreochlamys* (Volvocales), based on comparative morphology of cultured material. *European Journal of Phycology*, *36*(2), 113–128. https://doi.org/10.1080/09670260110001735268

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nozaki, H. (1990). Ultrastructure of the extracellular matrix of *Gonium* (Volvocales, Chlorophyta). *Phycologia*, *29*(1), 1–8. https://doi.org/10.2216/i0031-8884-29-1-1.1

Nozaki, H. (1996). Morphology and evolution of sexual reproduction in the Volvocaceae (Chlorophyta). *Journal of Plant Research*, *109*(3), 353–361. https://doi.org/10.1007/BF02344484

Nozaki, H. (2003). Origin and evolution if the genera *Pleodorina* and *Volvox* (Volvocales). *Biologia*, *58*(4), 425–431.

Nozaki, H., Ito, M., Uchida, H., Watanabe, M. M., Takahashi, H., & Kuroiwa, T. (1997). Phylogenetic analysis of *Yamagishlella* and *Platydorina* (volvocaceae, chlorophyta) based on rbcl gene sequences. *Journal of Phycology*, *33*, 272–278.

Nozaki, H., & Itoh, M. (1994). Phylogenetic relationships within the colonial volvocales (chlorophyta) inferred from cladistic analysis based on morphological data. *Journal of Phycology*, *30*(2), 353–365.

Nozaki, H., Itoh, M., Watanabe, M. M., & Kuroiwa, T. (1996). Ultrastructure of the vegetative colonies and systematic position of *Basichlamys* (Volvocales, Chlorophyta). *European Journal of Phycology*, *31*(1), 67–72. https://doi.org/10.1080/09670269600651211a

Nozaki, H., & Kuroiwa, T. (1992). Ultrastructure of the extracellular matrix and taxonomy of *Eudorina*, *Pleodorina* and *Yamagishiella* gen. Nov.(Volvocaceae, Chlorophyta). *Phycologia*, *31*(6), 529–541.

Nozaki, H., Matsuzaki, R., Yamamoto, K., Kawachi, M., & Takahashi, F. (2015). Delineating a New Heterothallic Species of *Volvox* (Volvocaceae, Chlorophyceae) Using New Strains of "*Volvox africanus*." *PLOS ONE*, *10*(11), e0142632. https://doi.org/10.1371/journal.pone.0142632

Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S., & Watanabe, M. M. (2000). Origin and Evolution of the Colonial Volvocales (Chlorophyceae) as Inferred from Multiple, Chloroplast Gene Sequences. *Molecular Phylogenetics and Evolution*, *17*(2), 256–268. https://doi.org/10.1006/mpev.2000.0831

Nozaki, H., Ohta, N., Takano, H., & Watanabe, M. M. (1999). Reexamination of phylogenetic relationships within the colonial Volvocales (Chlorophyta): An analysis of *atp*B and *rbc*L gene sequences. *Journal of Phycology*, *35*(1), 104–112. https://doi.org/10.1046/j.1529-8817.1999.3510104.x

Nozaki, H., Ott, F. D., & Coleman, A. W. (2006). Morphology, molecular phylogeny and taxonomy of two new species of *Pleodorina* (Volvoceae, Chlorophyceae). *Journal of Phycology*, *42*(5), 1072–1080. https://doi.org/10.1111/j.1529-8817.2006.00255.x

Nozaki, H., Yamada, T. K., Takahashi, F., Matsuzaki, R., & Nakada, T. (2014). New "missing link" genus of the colonial volvocine green algae gives insights into the evolution of oogamy. *BMC Evolutionary Biology*, *14*(1), 1–11. https://doi.org/10.1186/1471-2148-14-37

Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L. K., Hellsten, U., Chapman, J., Simakov, O., Rensing, S. A., Terry, A., Pangilinan, J., Kapitonov, V., Jurka, J., Salamov, A., … Rokhsar, D. S. (2010). Genomic Analysis of Organismal Complexity in the Multicellular Green Alga *Volvox carteri*. *Science*, *329*(5988), 223–226. https://doi.org/10.1126/science.1188800

Pröschold, T., Darienko, T., Krienitz, L., & Coleman, A. W. (2018). *Chlamydomonas schloesseri* sp. Nov. (Chlamydophyceae, Chlorophyta) revealed by morphology, autolysin cross experiments, and multiple gene analyses. *Phytotaxa*, *362*(1), 21–38. https://doi.org/10.11646/phytotaxa.362.1.2

Pröschold, T., Marin, B., Schlösser, U. G., & Melkonian, M. (2001). Molecular Phylogeny and Taxonomic Revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas* Ehrenberg and *Chloromonas* Gobi, and Description of *Oogamochlamys* gen. Nov. And *Lobochlamys* gen. Nov. *Protist*, *152*(265–300), 36.

Quintero-Galvis, J. F., Paleo-López, R., Solano-Iguaran, J. J., Poupin, M. J., Ledger, T., Gaitan-Espitia, J. D., Antoł, A., Travisano, M., & Nespolo, R. F. (2018). Exploring the evolution of multicellularity in *Saccharomyces cerevisiae* under bacteria environment: An

experimental phylogenetics approach. *Ecology and Evolution*, *8*(9), 4619–4630. https://doi.org/10.1002/ece3.3979

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, *109*(5), 1595–1600. https://doi.org/10.1073/pnas.1115323109

Ratcliff, W. C., Herron, M. D., Howell, K., Pentz, J. T., Rosenzweig, F., & Travisano, M. (2013). Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nature Communications*, *4*(1), 1–7. https://doi.org/10.1038/ncomms3742

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rose, J., Toledo, C. A., Lemmon, E. M., Lemmon, A. R., & Sytsma, K. (2020). Out of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in the Flowering Plant Genus *Polemonium* (Polemoniaceae) Suggests Widespread Historical Gene Flow Despite Limited Nuclear Signal. *Systematic Biology*, *70*(1), 162–180.

Schirrmeister, B. E., de Vos, J. M., Antonelli, A., & Bagheri, H. C. (2013). Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences*, *110*(5), 1791–1796. https://doi.org/10.1073/pnas.1209927110

Schmitt, R. (2003). Differentiation of germinal and somatic cells in *Volvox carteri*. *Current Opinion in Microbiology*, *6*(6), 608–613. https://doi.org/10.1016/j.mib.2003.10.007

Smith, G. M. (1944). A Comparative Study of the Species of *Volvox*. *Transactions of the American Microscopical Society*, *63*(4), 265–310. https://doi.org/10.2307/3223302

Smith, S. A., & Dunn, C. W. (2008). Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, *24*(5), 715–716. https://doi.org/10.1093/bioinformatics/btm619

Soltis, D. E., & Kuzoff, R. K. (1995). Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). *Evolution*, *49*(4), 727–742. https://doi.org/10.1111/j.1558-5646.1995.tb02309.x

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Szathmáry, E., & Maynard Smith, J. (1995). The major evolutionary transitions. *Nature*, *374*(6519), 227–232. https://doi.org/10.1038/374227a0

van Leeuwenhoek, A. (1700). Part of a letter from Mr Antony van Leeuwenhoek, concerning the worms in sheeps livers, gnats, and animalcula in the excrements of frogs. *Philosophical Transactions of the Royal Society of London*, *22*, 509–518.

Wahl, M. E., & Murray, A. W. (2016). Multicellularity makes somatic differentiation evolutionarily stable. *Proceedings of the National Academy of Sciences*, *113*(30), 8362–8367. https://doi.org/10.1073/pnas.1608278113

West, S. A., Fisher, R. M., Gardner, A., & Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, *112*(33), 10112–10119. https://doi.org/10.1073/pnas.1421402112

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.-W., Li, Y., Xu, X., Wong, G. K.-S., & Wang, J. (2014). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*(12), 1660–1666. https://doi.org/10.1093/bioinformatics/btu077

Xu, B., Wu, N., Gao, X.-F., & Zhang, L.-B. (2012). Analysis of DNA sequences of six chloroplast and nuclear genes suggests incongruence, introgression, and incomplete lineage sorting in the evolution of *Lespedeza* (Fabaceae). *Molecular Phylogenetics and Evolution*, *62*(1), 346–358. https://doi.org/10.1016/j.ympev.2011.10.007

Yang, Z., & Roberts, D. (1995). On the use of nucleic acid sequences to infer early branchings in the tree of life. *Molecular Biology and Evolution*, *12*(3), 451–458. https://doi.org/10.1093/oxfordjournals.molbev.a040220

Yu, W.-B., Huang, P.-H., Li, D.-Z., & Wang, H. (2013). Incongruence between Nuclear and Chloroplast DNA Phylogenies in *Pedicularis* Section *Cyathophora* (Orobanchaceae). *PLoS ONE*, *8*(9), e74828. https://doi.org/10.1371/journal.pone.0074828

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(S6), 153. https://doi.org/10.1186/s12859-018-2129-y

Zhang, Z., Qu, C., Yao, R., Nie, Y., Xu, C., Miao, J., & Zhong, B. (2019). The Parallel Molecular Adaptations to the Antarctic Cold Environment in Two Psychrophilic Green Algae. *Genome Biology and Evolution*, *11*(7), 1897–1908. https://doi.org/10.1093/gbe/evz104