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

Jonathan Featherston,*[1,2] Yoko Arakaki,[3] Erik R. Hanschen,[4] Patrick J. Ferris,[4] Richard E. Michod,[4] Bradley J.S.C. Olson,[5] Hisayoshi Nozaki,[3] and Pierre M. Durand[1]

[1]Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

[2]Agricultural Research Council, Biotechnology Platform, Pretoria, South Africa

[3]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Bunkyo-ku, Tokyo, Hongo, Japan

[4]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ

[5]Division of Biology, Kansas State University, Manhattan, KS

***Corresponding author**: E-mail: featherstonj@arc.agric.za.

Associate editor: Nicole King

Abstract

Multicellularity is the premier example of a major evolutionary transition in individuality and was a foundational event in the evolution of macroscopic biodiversity. The volvocine chlorophyte lineage is well suited for studying this process. Extant members span unicellular, simple colonial, and obligate multicellular taxa with germ-soma differentiation. Here, we report the nuclear genome sequence of one of the most morphologically simple organisms in this lineage—the 4-celled colonial *Tetrabaena socialis* and compare this to the three other complete volvocine nuclear genomes. Using conservative estimates of gene family expansions a minimal set of expanded gene families was identified that associate with the origin of multicellularity. These families are rich in genes related to developmental processes. A subset of these families is lineage specific, which suggests that at a genomic level the evolution of multicellularity also includes lineage-specific molecular developments. Multiple points of evidence associate modifications to the ubiquitin proteasomal pathway (UPP) with the beginning of coloniality. Genes undergoing positive or accelerating selection in the multicellular volvocines were found to be enriched in components of the UPP and gene families gained at the origin of multicellularity include components of the UPP. A defining feature of colonial/multicellular life cycles is the genetic control of cell number. The genomic data presented here, which includes diversification of cell cycle genes and modifications to the UPP, align the genetic components with the evolution of this trait.

Key words: **Volvocines, multicellularity, *Tetrabaena socialis*, evolutionary genomics.**

Introduction

The evolution of multicellular organisms from unicellular ancestors is an example of a transition in the unit of selection and evolution, in this case, from the level of the cell to the level of the multicellular group. Explaining such a major evolutionary transition (*sensu* Maynard Smith and Szathmary 1995) on evolutionary principles is a challenge, because explanatory concepts such as fitness and individuality are themselves changing during the transition. Critical in this endeavor to explain the origin of multicellular life is understanding the initial steps involved in the transition from a unicellular life cycle to a life cycle involving groups of cells (Maliet et al. 2015; Rashidi et al. 2015). The volvocine green algae provide a compelling model system to deconstruct these early steps (Starr 1968; Kirk and Harper 1986; Kirk 1999, 2005; Herron and Michod 2008; Herron 2009). The value of the volvocines as a model system is due to 1) the relatively recent evolution of multicellularity in the lineage (∼234 Mya) (Herron et al. 2009) when compared to lineages such as land plants and metazoa (600–1000 MYA) (Sharpe et al. 2015), 2) the availability of extant taxa with varying levels of organismal complexity, and 3) with phylogenetic relationships sufficiently resolved for meaningful comparative analyses (Larson et al. 1992; Coleman 1999; Nozaki et al. 2000; Nakada et al. 2008; Herron et al. 2009; Nozaki et al. 2014, 2016).

The volvocine algae form a monophyletic lineage within the *reinhardtinia*-clade of the order Chlamydomonadales (Nakada et al. 2008). Members include single-celled species (e.g., *Chlamydomonas reinhardtii*), simple colonial forms where multicellularity can be facultative (e.g., the 4-celled *Tetrabaena socialis* and the 16-celled *Gonium pectorale*) to complex multicellular taxa with germ-soma differentiation (e.g., *Volvox carteri*) (Nishii and Miller 2010; Coleman 2012). Multicellular members of the lineage are divided into three families: the Tetrabaenaceae (e.g., *T. socialis*), the Goniaceae (e.g., *G. pectorale*) and the Volvocaceae (e.g., *Pleodorina starrii* and *V. carteri*) (Coleman 1999; Nozaki et al. 2000). Comparative genomic analyses of *C. reinhardtii*

Mol. Biol. Evol. doi:10.1093/molbev/msx332 Advance Access publication December 26, 2017

(Merchant et al. 2007), *G. pectorale* (Hanschen et al. 2016), and *V. carteri* f. *nagariensis* (hereafter *V. carteri*) (Prochnik et al. 2010) have revealed protein-coding potential to be similar across the lineage. Consequently, research into volvocine developmental complexity has largely focused on gene exaptation/cooption rather than protein-coding novelty (Miller and Kirk 1999; Kirk 2001; Nishii et al. 2003; Duncan et al. 2006; Hanschen et al. 2014, 2016; Olson and Nedelcu 2016; Grochau-Wright et al. 2017). Genome comparisons have furthered our understanding of some of the important developmental traits. These include: the association between expanded extracellular matrix (ECM) genes and morphological expansion of the ECM in *V. carteri*, (Prochnik et al. 2010; Hanschen et al. 2016) and the master regulator of germ-soma differentiation in *V. carteri* regA (Kirk et al. 1999) and closely related *regA*-cluster genes (Duncan et al. 2006; Hanschen et al. 2014) originated near the split between the Goniaceae and Volvocaceae (Hanschen et al. 2014, 2016; Grochau-Wright et al. 2017).

In accordance with the expectation that the transition to multicellular life involves construction of a life cycle at the level of the cell group (Maliet et al. 2015), previous work, including genomic analyses, in the volvocine algae has shown that cell cycle genes are among the earliest diversifying gene families (Prochnik et al. 2010; Hanschen et al. 2016). These include the expansion of *CYCD1* genes relative to *C. reinhardtii* in *G. pectorale* and *V. carteri* (Prochnik et al. 2010; Hanschen et al. 2016) and a role for the retinoblastoma pathway in cell cycle modifications associated with multicellularity (Hanschen et al. 2016). The palintomic multiple-fission life cycle of the unicellular *C. reinhardtii* includes a brief phase as a group where division has occurred but separation is incomplete. This brief phase is rapidly followed by separation during which there is no growth. In contrast, cell growth in *T. socialis* occurs in the group context. The order of cell cycle events has changed in *T. socialis* such that the separation stage occurs after cell growth and before division in the group life cycle instead of before growth in the unicellular cycle (Maliet et al. 2015). Another major difference between the multicellular program of division and the unicellular program is that the maximum number of divisions (n) is genetically determined in the multicellular taxa and not in *C. reinhardtii* (Kirk 2005).

To date, sequenced colonial and multicellular volvocines are from the Goniaceae and Volvocaceae families. Members of the Tetrabaenaceae family, which is sister to the Goniaceae and Volvocaceae families (fig. 1), possess morphological traits that can be used to make inferences about the colonial forms ancestral to the Goniaceae and Volvocaceae families. The most well studied member of the Tetrabaenaceae family is the 4-celled *T. socialis* (supplementary fig. S1, Supplementary Material online), which is amongst the "simplest" of known colonial multicellular eukaryotes (Nozaki 1986; Arakaki et al. 2013). The nuclear genome of *T. socialis* is analyzed here (organelle genomes have been described elsewhere [Featherston et al. 2016]). Our analyses confirm the limited differences in protein coding potential between the unicellular *C. reinhardtii* and the colonial/multicellular taxa. A small set of orthologous gene families were identified that arose at the evolution of colonial living and an even smaller set of families showed expansion in colonial/multicellular taxa. Although few families were gained or expanded, these include a number of genes with putative functions related to DNA repair, surveillance of DNA damage or are homologous to metazoan oncogenes. We show here that this diversification of cell cycle genes began at the earliest possible stage with the evolution of a 4-celled Tetrabaena colony and that the genetic determination of colony size was likely one of the earliest group properties to evolve. Methods to detect positive and accelerating selection, as well as other findings, identified numerous genes involved at most points in the ubiquitin proteasomal pathway (UPP) and we discuss a role for the UPP in the genetic control of cell number.

Results and Discussion

Genome Assembly

A final *Tetrabaena socialis* (NIES-571) genome assembly was generated after combining Allpaths-LG (Gnerre et al. 2011) and SPAdes (Bankevich et al. 2012) assemblies from coverage in excess of 400×. The combined meta-assembly produced 20,363 contigs, 5,856 scaffolds and totalled 135.7 MB in length. The N50 values for contigs and scaffolds were 7.4 and 145.9 KB, respectively (table 1). The *T. socialis* genome is highly GC-skewed with a GC content of 66%. Kmer analyses with Jellyfish (v2.0.0) (Marcais and Kingsford 2011) and GenomeScope (Vurture et al. 2017) (e.g., supplementary fig. S2, Supplementary Material online), Allpaths-LG and BBtools (http://jgi.doe.gov/data-and-tools/bbtools/, last accessed January 3, 2018) did not converge on the same estimated genome size or repeat content (supplementary table S1, Supplementary Material online). The estimated genome is likely to be at least ~120 MB with repeat content in excess of 20%, which is similar to *C. reinhardtii* (19%) and *V. carteri* (20%) (Prochnik et al. 2010).

Transcriptome Assembly and Annotation

A de novo transcriptome was generated using Bridger De Novo (r2014-12-01) (Chang et al. 2015) from 75 million paired RNAseq reads (v4 SBS 2 × 125 bp chemistry). TransRate (1.0) (Smith-Unna et al. 2016) was used to evaluate assembly quality and to filter contigs ($n = 23{,}559$). The TransRate Optimal Assembly Score was 0.362. By comparison, 50% of 155 publically available de novo transcriptomes have been shown to have a TransRate Optimal Assembly Score of $\leq$0.35 (Smith-Unna et al. 2016). A total of 21,823 (92%) assembled transcripts mapped to the genome assembly (90% mapped with >95% of contig length). Computationally derived annotations were generated using the MAKER2 pipeline (v2.31.3) (Holt and Yandell 2011) with evidence for computational predictions derived from de novo assembled transcripts, protein sequences from publically available volvocine genomes (from *C. reinhardtii*, *G. pectorale*, and *V. carteri*, supplementary table S2, Supplementary Material online) as well as the SWISS-PROT database (Boutet et al. 2016) of curated proteins. Automated annotations produced 15,513 protein-coding loci, which is a value similar to other volvocines (table 1). A total of 65% of proteins modeled by MAKER2

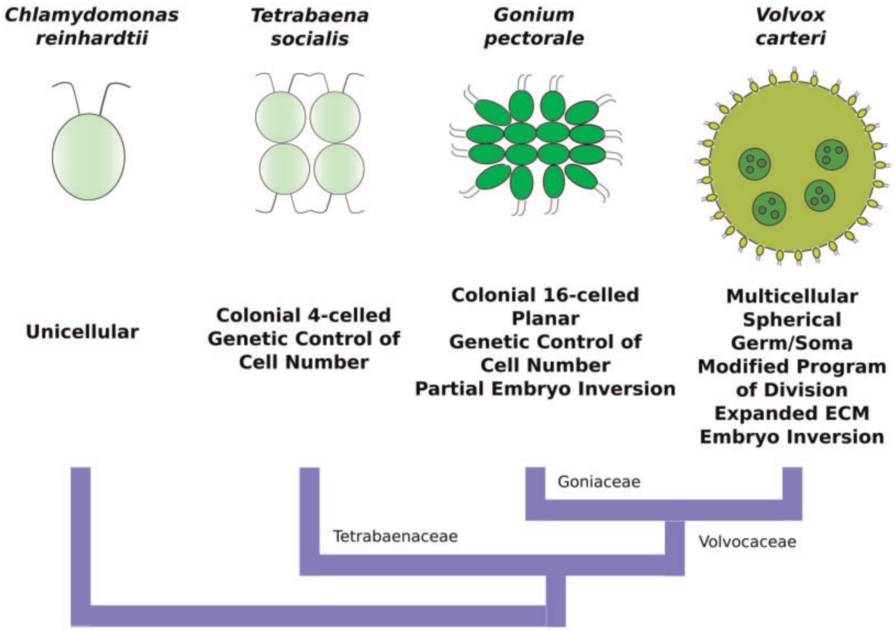

Fig. 1. Traits relevant to the current investigation and phylogenetic relationships of genome-sequenced volvocines.

Table 1. Summary Statistics of Volvocine Genomes Used in This Analysis.

	Chlamydomonas reinhardtii V5.5	*Tetrabaena socialis*	*Gonium pectorale*	*Volvox carteri V1*[a]
Assembly size	111.1 MB	135.7 MB	148.8 MB	137.8 MB
Contig N50	90.6 KB	7.4 KB	16.2 KB	43,981
Scaffold N50	6,617 KB	146 KB	1,276 KB	1,491 KB
Number of scaffolds	88	5,858	2,373	1,265
GC content	64.1	66.0	64.5	56.0
Number of coding genes	17,737	15,513	17,984	14,520
Average intron length	279.17	635	349.83	496.67
Genes with Pfam-A domains	10,889[b]	9,803	10,029	8,456

[a] For this analysis, *Volvox carteri* version 1 data were used (see supplementary text, Supplementary Material online).

[b] Including alternative splice variants.

were in possession of at least one Pfam-A domain (Finn et al. 2014) and 93.8% of Eukaryotic Clusters of Orthologous Genes (KOG) domains (Parra et al. 2007) could be identified in the predicted proteome. Nonredundant homologs for at least 91% of *C. reinhardtii* KOG proteins ($n = 406$) (Parra et al. 2007) were identified in the *T. socialis* predicted proteome, indicating that at least 90% of genes were modeled. Manual annotation of gene families (e.g., pherophorins or supplementary table S24, Supplementary Material online orthologues) showed similar levels of model coverage. The average intron length in *T. socialis* is greater than those of other volvocines (table 1). This is likely due to a mixture of the highly repetitive nature of the genome and due to assembly artifacts that are the result of inflated average mate-pair insert distances spanning exons. Using OrthoMCL (v2.09) and OrthoVenn orthologous proteins shared by all volvocines numbered 5199 (supplementary fig. S3, Supplementary Material online). BLASTP searches to volvocine proteins as well as UNIPROT databases identified significant matches ($E = 10^{-7}$) for 90% of *T. socialis* predicted proteins (supplementary table S3, Supplementary Material online) and the percentage of reciprocal BLAST matches to proteins from model organism *C. reinhardtii* was comparable for *T. socialis* when compared to *G. pectorale* and *V. carteri* (supplementary table S4, Supplementary Material online).

Domain Content

Comparing domains in *C. reinhardtii* with those of the multicellular taxa (*T. socialis*, *G. pectorale*, and *V. carteri*) in a 2-way comparison identified only 61 domains as significantly enriched ($P = 0.05$). With false discover rate (FDR) correction only 12 of these domains were significant ($q = 0.05$). Of the 61 domains at an alpha of 0.05 only 18 were enriched in the multicellular taxa (the other 43 were enriched in *C. reinhardtii*). A further 30 domains were identified as present in *T. socialis*, *G. pectorale*, and *V. carteri* but absent from the proteome of *C. reinhardtii* (supplementary table S5, Supplementary Material online). Domains that were significantly enriched ($n = 61$) and domains unique to the multicellular taxa ($n = 30$) were combined (total $n = 91$) and examined for their presence in selected eukaryotic model species (supplementary table S2 and fig. S4, Supplementary Material online). The majority of these domains were not enriched in volvocine or indeed chlorophyte specific domains (supplementary fig. S5, Supplementary Material online) but rather the majority ($n = 59$) were of a more universal eukaryotic origin (supplementary fig. S5, Supplementary Material online). These findings are consistent with previous volvocine genomes, which suggested that extensive domain novelty was not a major feature of the evolution of multicellularity in the volvocines (Prochnik et al. 2010; Hanschen et al. 2016). With the exception of the gametolysin domain (Hanschen et al. 2016) enriched domains between *C. reinhardtii* and *V. carteri* (Prochnik et al. 2010) mostly show loss in *V. carteri* (supplementary table S6, Supplementary Material online) and this is particularly evident for ankyrin domains (supplementary table S7, Supplementary Material online).

Phylogenetic Reconstruction of Gene Family Expansion and Contraction

OrthoMCL (2.09) (Li et al. 2003) was used to generate gene family clusters from the proteomes of all available genome sequenced chlorophyta ($n = 13$) (supplementary table S2, Supplementary Material online). Following Hanschen et al (2016), parsimony reconstruction of gene family expansion and contraction was performed using the Count software package (v10.04) (Csuros 2010) using Dollo and Wagner's parsimony (symmetric and asymmetric). We focused on the results of Wagner's symmetrical parsimony as this produced the most conservative set of gene families expanded at the origin of multicellularity (supplementary table S8, Supplementary Material online). The largest expansion of gene families ($n = 2657$) in the phylogeny was for the Chlorophyceae (at the split between the chlorophyta orders Chlorophyceae and Trebouxiophyceae), which points towards substantial protein-coding diversification and increasing complexity in this order since separating from other lineages (fig. 2) (Hanschen et al. 2016). However, it remains unknown if this increased proteome complexity was important for the evolution of multicellularity.

Using results from Wagner's symmetric parsimony a set of 131 gene families was gained/originated at the origin of volvocine coloniality/multicellularity (fig. 2). For simplicity these genes were labeled as the "origin of multicellularity" (ORIMC) gene set (supplementary table S9, Supplementary Material online). Argot2 (http://www.medcomp.medicina.unipd.it/Argot2/, last accessed January 3, 2018) (Falda et al. 2012), InterProScan (Quevillon et al. 2005) as well as BLASTP (Altschul et al. 1997) homology to *C. reinhardtii* proteins were used to annotate the ORIMC gene set. BLASTP queries to the RefSeq nonredundant (nr) database (O'Leary et al. 2016) identified homology matches to proteins from a diverse range of organisms including both unicellular and multicellular taxa (supplementary fig. S6, Supplementary Material online). The majority of matches ($n = 59$) are to proteins present in both unicellular and multicellular taxa. Of the 131 families, homologues (outside of the volvocine lineage) were not identified for 52 families ($E = 1 \times 10^{-10}$). A substantial proportion of ORIMC families (40%) are therefore lineage specific without an ascribed function, which highlights that the evolution of multicellularity is associated with lineage specific mechanisms at the molecular level. Domains present in these lineage-specific families include a number of domains with a role in nucleic acid binding suggesting that some of these proteins have a role in transcription or translation regulation (table 2). Only three families had matches to proteins from multicellular taxa with no matches to proteins from unicellular taxa, which reveals that the evolution of multicellularity in the volvocine lineage did not arise due to the acquisition or retention of genes that are unique to multicellular organisms or functions. Instead, the ORIMC families are abundant in lineage specific families and genes that are present in both unicellular and multicellular organisms.

A number of families that originate at the ORIMC are candidates for further exploration many of which have putative functions related to developmental or regulatory processes. Of particular interest may be those homologous to DNA repair, DNA damage surveillance and cancer genes, including: breast cancer susceptibility1-like (*BRCA1-like*) gene, colon cancer-associated protein Mic1-like gene, RECQ-helicases, structural maintenance of chromosome 3 proteins (with a recF/recN domain), MutL C-terminal dimerization domain containing proteins and a family of proliferating cell nuclear antigen domain containing proteins (Strzalka and Ziemienowicz 2011). In relation to the Ubiquitin Proteasomal Pathway (UPP), a family of AAA-type ATPases with a possible role in proteosomal degradation was gained (Bar-Nun and Glickman 2012). Of the 131 ORIMC families, homology ($E = 10^{-10}$) to *C. reinhardtii* proteins was not identified for proteins from 30 families. Using TBLASTN to search the *C. reinhardtii* genome with proteins from these 30 families it was found that 16 gene families had no homology at $E = 10^{-5}$ while a further five showed no homology at $E = 10^{-10}$. There were two instances where alignments indicated that an orthologue is present in *C. reinhardtii* while all other BLAST alignments ($E < 10^{-10}$) were partial or due to repetitive motifs. Argot2 only assigned eight gene ontology (GO) terms to the 30 families. BLASTP searches against the RefSeq nonredundant (O'Leary et al. 2016) and UNIPROT-TREMBL databases (Schneider et al. 2009) collectively identified only 18 families where matches were not to volvocine proteins, suggesting that certain of the 30 families may have been gained by horizontal transfer events. Families with no homology to *C. reinhardtii* proteins were likely either lost in *C. reinhardtii*, gained by horizontal transfer or arose de novo near the origin of multicellularity. Homology to *C. reinhardtii* proteins could be identified for proteins from the remaining ORIMC families ($n = 101$), however, it was not determined for how many a *C. reinhardtii* orthologue might be present. The exclusion of *C. reinhardtii* proteins from these 101 OrthoMCL Markov family clusters is most likely accounted for by protein diversification, including paralogue divergence from gene duplication events but is not suggestive of ab initio gene formation.

Gene families expanded in the multicellular taxa relative to *C. reinhardtii* numbered only 27 (supplementary table S10, Supplementary Material online). Amongst expanded families was a family of MutL-related genes and, furthermore, enrichment analysis of expanded families using *C. reinhardtii* v5 annotations (Lopez et al. 2011; Blaby et al. 2014) identified the KEGG pathway "Mismatch repair."

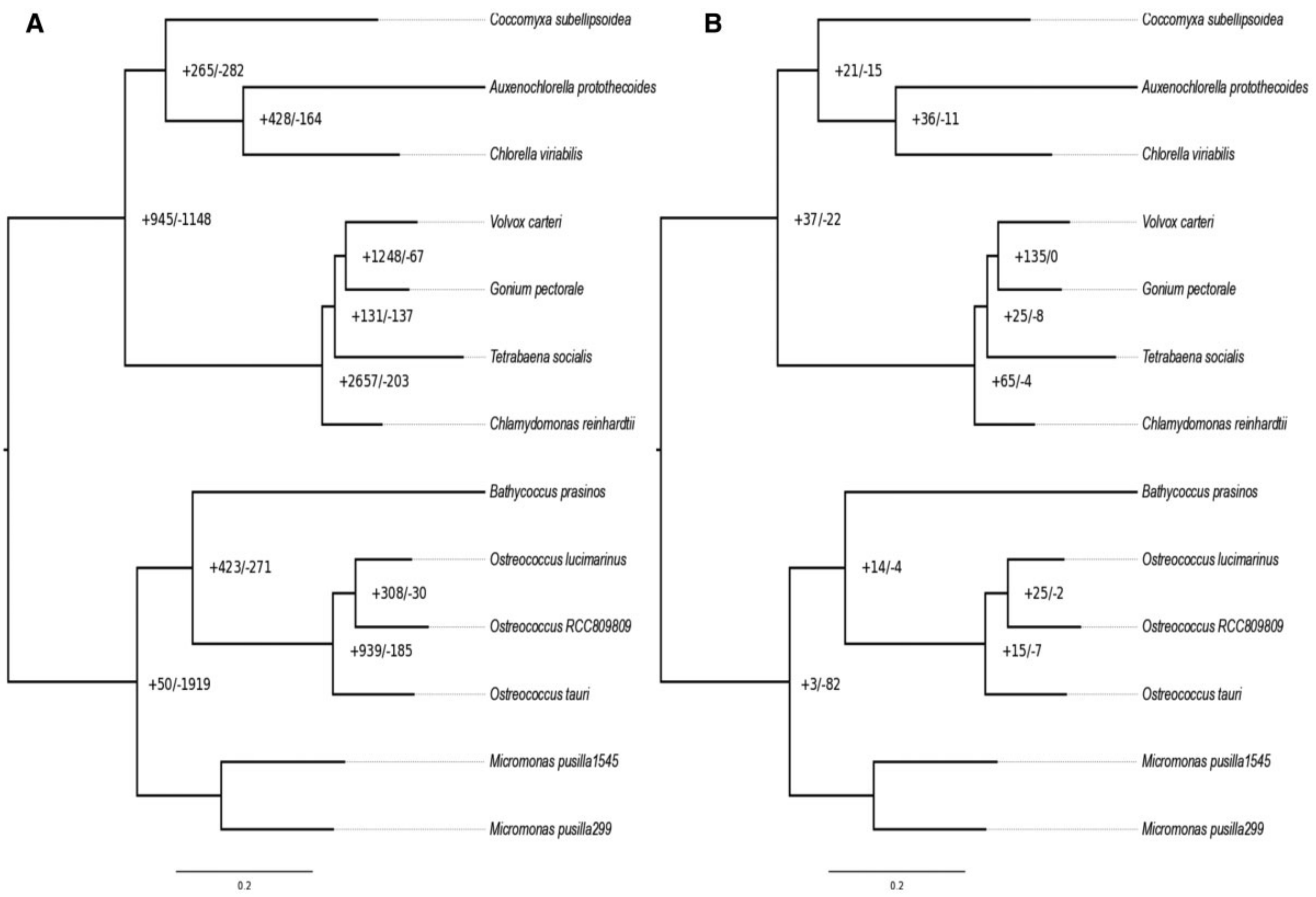

Fig. 2. Wagner's symmetrical parsimony reconstruction of gene family history for whole genome sequenced chlorophytes. (*A*) Gene families gained and lost (supplementary table S9, Supplementary Material online). (*B*) Gene familes expanded and contracted (supplementary table S10, Supplementary Material online). Phylogeny based on 927 single orthologue protein families (100% bootstrap support for all branches).

Table 2. Domains of Proteins that Arose at the Origin of Multicellularity for Proteins without Homology to the Proteins in the RefSeq Database.

Domain	PFAM Annotation
PAN domain	Protein–protein protein–carbohydrate interactions
BED zinc finger domain	DNA binding
Protein kinase domain	Phosphorylation frequently regulatory
Cyclic nucleic acid domain	Binding of cAMP and GMP
Up-frameshift suppressor 2 domain	Nonsense mediated mRNA decay
DUF1762	Polymerase II transcription
MAPEG domain	Membrane bound metabolic functions
PPPDE domain	Role in the Ubiquitin proteasomal pathway
CAAD domain of cynobacterial aminoacyl-tRNA synthetase	Coupling of tRNAs to amino acids
DUF4436	Unknown
RAP domain	RNA binding
BTZ domain	Post transcriptional regulation

In terms of gene function, families of protein kinases and DNA-repair related genes were identified both amongst families gained and expanded at the ORIMC. As molecules with diverse functions it is unsurprising to identify diversification of protein kinases (Francino 2005) and protein kinase expansions have previously been associated with the evolution of multicellularity or increasing organismal complexity in other lineages (Rensing et al. 2008; Stajich et al. 2010; de Mendoza and Ruiz-Trillo 2011; Miller 2012; Schultheiss et al. 2012; Glockner et al. 2016). The expansion or diversification of DNA repair genes (Rad51-type) has been identified in the multicellular brown alga *Ectocarpus siliculosus* (Cock et al. 2010), the importance of which is not known and the significance of gained/expanded families of DNA repair-related genes in the multicellular volvocines requires further investigation. There is currently no evidence to suggest that replication fidelity is improved in multicellular volvocines and the general trend across life suggests that mutation rates rise with increased organismal complexity (Lynch et al. 2016). This does not, however, preclude specific exceptions to this general trend. Based on the premise that increased organismal complexity potentially results in increased risk of harm from mutations and mutator genotypes (Lynch 2008) then a possibility for gains in DNA repair-related genes in the multicellular taxa might be related to increased redundancy in repair mechanisms. Both protein kinase signaling and DNA repair pathways may play a role in regulating the cell cycle (Haring et al. 1995; Branzei and Foiani 2008).

Protein Orthologues with Evidence of Codon-Level Positive Selection

Codeml (Yang 2007) site models of positive selection were applied within the Potion pipeline (v1.1.2) (Hongo et al. 2015) to identify evidence for positive selection amongst orthologue families. Analysis was restricted to OrthoMCL families without paralogues with a single protein from each of the 4-volvocine taxa and 73 family clusters with evidence of positive selection were identified ($q = 0.05$ and m7/8 models, supplementary table S11, Supplementary Material online). It is expected that substantially more families with evidence of positive selection would be detected if paralogues were included. GO enrichment analysis applied to *C. reinhardtii* v5.5 annotations (Blaby et al. 2014) from the 73 families identified 82 enriched terms. Most enriched terms were related to metabolic processes (supplementary table S12, Supplementary Material online). Selection acting upon metabolic processes might be evidence of metabolic adaptations to group living since the initial step of group formation is stressful to organisms due to the challenges of dealing with dying cells and nutrient and waste exchange (Durand et al. 2016; Sathe and Durand 2016). Alternatively, selection acting upon metabolic pathways may be indicative of a shift in nutrient acquisition from a mixotrophic lifestyle as per *C. reinhardtii* to obligate autotrophy, which is a characteristic of the multicellular volvocines. In addition, terms associated with developmental processes were amongst the enriched terms. A number of terms related to the UPP were enriched that ranged from targeting for proteolysis by ubiquitination to proteasomal components. In addition, the MapMan annotation term (Klie and Nikoloski 2012) "protein degradation ubiquitin" and the KEGG pathway "Proteosome" pathway were enriched. Genes from the 73 families with putative functions of interest include genes related to mRNA processing/splicing and polymerization, protein translation, protein shuttling, Ca^{2+} sensing and signaling, UPP-related genes, chromatin-remodeling, and FIZZY-related kinase activity.

Genome-Level Scans for Conservation and Accelerating Selection

A whole-genome multiple sequence alignment of *C. reinhardtii*, *T. socialis*, *G. pectorale*, and *V. carteri* was generated, referenced to *C. reinhardtii* and examined for signature of selection using the PHAST suite of algorithms (v1.1.) (Hubisz et al. 2011). Highly conserved noncoding elements identified with PhastCons in the volvocine genomes were remarkably sparse—only 101 elements of average length 58.9 bp were identified (5,951 bp) (supplementary tables S13 and S14, Supplementary Material online). This was almost entirely due to the minimal alignment stemming from noncoding sequences, which amounted to only 57,898 bp. The paucity of alignments stemming from noncoding regions can be explained by a high-degree of noncoding genome diversification, simple repeats masking (including removal of duplicate alignments) and the gap content of volvocine genome assemblies, which are highest in the *T. socialis* (27%) and *G. pectorale* (21%) assemblies.

PhyloP (Pollard et al. 2010) was used to identify features that, relative to *C. reinhardtii*, showed evidence of accelerating selection in the multicellular taxa. Features were divided according to *C. reinhardtii* v5.5 annotations into genes (whole genes including untranslated regions), CDS, 3′UTR, 5′UTR and intergenic regions. With false discovery rate correction ($q = 0.05$) 129 genes were identified as accelerating in the multicellular taxa relative to *C. reinhardtii* (supplementary table S15, Supplementary Material online) of which 24 were accelerating faster than the neutral rate (the remaining 105 genes were accelerating in the multicellular taxa relative to *C. reinhardtii* but not faster than the neutral rate). Amongst the 24 more rapidly evolving genes were the retinoblastoma (*MAT3/RB*) gene and an *SPC97/SPC98* member of the spindle pole body component (also known as Gamma tubulin interacting protein 2 [*GCP2*]). *GCP* genes in plants are involved in microtubule cytoskeleton organization and spindle integrity (Janski et al. 2012). In addition, the tubulin beta chain 1 gene was accelerating relative to *C. reinhardtii*. Both genes may be of interest for further investigation because microtubule interactions are important for volvocine developmental traits such as embryo inversion in *V. carteri* (Nishii et al. 2003). GO enrichment analysis of the 129 genes identified enriched annotations such as "generative cell differentiation", "regulation of cell cycle" and "actin filament-based movement." Enriched molecular functions included "cyclin binding" and "ubiquitin activating enzyme activity." Enriched MapMan terms included "RNA transcription", "cell organization", "protein degradation ubiquitin E1" and "cell division" (supplementary table S16, Supplementary Material online).

The term "accelerating selection" implies that loci were evolving more rapidly than the neutral rate or, for branch tests, acceleration indicates that loci are evolving more rapidly on a particular phylogenetic branch than for the rest of the tree. Acceleration might indicate positive selection but may also indicate relaxation of purifying selection and/or even ultimately pseudogenization. Therefore, in order to identify if genes accelerating in the multicellular taxa showed evidence of positive selection Potion analysis was repeated specifically for families containing accelerating genes, but with slightly relaxed filters (see supplementary text, Supplementary Material online) employed and with the inclusion of families containing paralogues. The 129 genes were associated with 124 orthoMCL families, of which 67 were valid within the Potion pipeline and 25 (37%) displayed evidence of positive selection (q-value $= 0.05$). We considered this to be a substantial proportion because accelerating genes without evidence for positive selection might be undergoing pseudogenization or acceleration may be occurring in noncoding parts of gene sequences rather than in CDS.

A total of 42 CDS features were identified (from 38 genes—supplementary table S17, Supplementary Material online) as accelerating in the multicellular taxa relative to *C. reinhardtii* with 16 accelerating faster than the neutral rate and 6 that were also identified as accelerating in the whole-gene analysis. Accelerating CDS features of interest include a putative chromatin remodeling gene and a cytoplasmic dynein heavy chain. Corroborating whole gene analysis and indicating that the coding regions of genes are under acceleration are the ubiquitin activating enzyme 2, ubiquitin ligase 1, GCP2, tubulin beta chain 2, a putative transcription regulator and a putative histone protein. GO enrichment analysis identified the cellular compartment terms "microtubule organizing center", "SWI/SNF complex," and "SWI/SNF-type complex" and the molecular function term "ubiquitin activating enzyme activity" as enriched. As per the gene-level analysis, accelerating CDS regions were examined for positive selection using Potion for 34 OrthoMCL families, which were filtered to 14 and of which 11 were significant for positive selection (79%). This indicates a good correlation, albeit from a limited number of families, between phyloP results and protein-level Codeml analysis, suggesting that the majority of accelerating CDS are under positive selection rather than under-going pseudogenization. Perhaps due to ~234 Mya of separation and low-levels of alignment stemming from noncoding regions no 3′UTRs, 5′UTRs or intergenic regions were identified as accelerating significantly faster than the neutral rate of noncoding DNA evolution. This is similar to the noncoding regions of volvocine organelle genomes, which are replete with extensive tracts of noncoding regions intractable to alignment (Smith and Lee 2009; Smith et al. 2013). Diversification of noncoding regions also suggests extensive divergence of regulatory elements and gene regulatory mechanisms (Kianianmomeni 2015), which in other lineages have been associated with the evolution of multicellularity (Sebe-Pedros et al. 2016). As per findings from proteome-wide scans of positive selection amongst orthologues, accelerating genes and CDS with evidence of positive selection include genes with functions related to the UPP (e.g., ubiquitin-activating enzyme 2 and ubiquitin ligase 1), Ca^{2+} sensing and signaling (e.g., calmodulin 4) and protein shuttling (e.g., heat shock protein 70), which is further evidence of selection acting on these processes. Genes with these functions are likely to be important for basic cellular functions and regulatory processes and have been associated with cell cycle regulatory functions (Goldknopf et al. 1980; Ciechanover et al. 1984; Kao et al. 1985; Goebl et al. 1988; von Kampen and Wettern 1991; Cheng et al. 2005; Choi and Husain 2006).

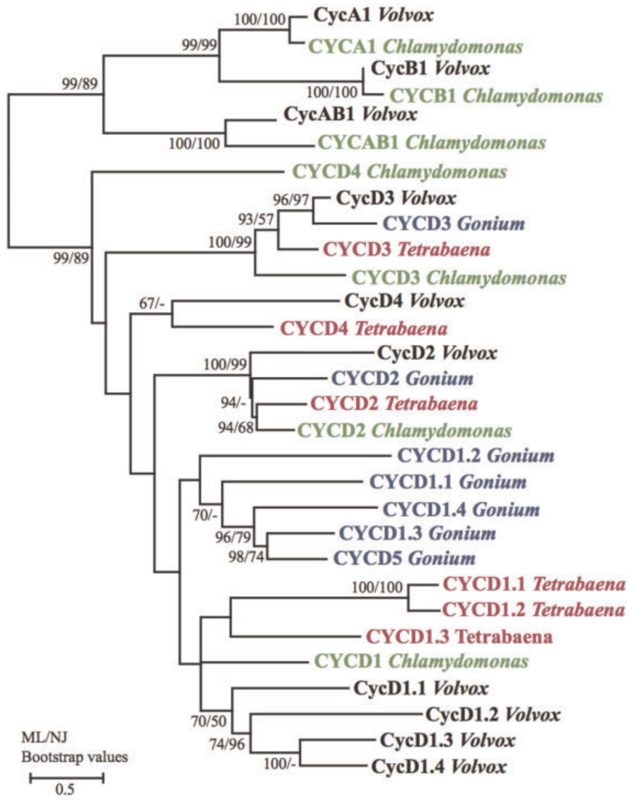

FIG. 3. Maximum likelihood and Neighbour-joining phylogeny of cyclins from *Chlamydomonas reinhardtii* (green), *Tetrabaena socialis* (magenta), *Gonium pectorale* (blue), and *Volvox carteri* (black). The phylogeny shows the expansion of *CYCD1* from a single copy in *Chlamydomonas reinhardtii* to three copies in *Tetrabaena socialis* and further expansion to four copies in *Gonium pectorale* and *Volvox carteri*.

Cyclins and the Retinoblastoma Gene

Theoretical models (Maliet et al. 2015; Rashidi et al. 2015) and observations of traits in extant taxa such as *T. socialis* (Arakaki et al. 2013) suggest that alterations to the cell cycle was one of the primary and contingent steps to emerge at the origin of group living (Kirk 2005; Herron and Michod 2008). The underlying genetic components that control the cell cycle are therefore of special interest. The primary genetic components of the cell cycle regulation pathway are similar throughout the volvocine lineage (Bisova et al. 2005; Merchant et al. 2007; Olson et al. 2010; Prochnik et al. 2010; Cross and Umen 2015) and are generally similar to those found in most eukaryotes. However, the cyclin D1 (*CYCD1*) gene has expanded through tandem duplication from a single copy in *C. reinhardtii* to four copies in *V. carteri* and *G. pectorale* (Prochnik et al. 2010; Hanschen et al. 2016). The expansion of *CYCD1* was identified in the *T. socialis* genome, three copies of *CYCD1* were identified and confirmed using rapid amplification of cDNA (RACE) and Sanger sequencing (fig. 3, supplementary tables S18 and S19, Supplementary Material online). Two of the *CYCD1* copies are located in close proximity on the same scaffold while the third copy is found on a separate scaffold (supplementary fig. S7, Supplementary Material online). The expansion of *CYCD1* genes in *T. socialis* suggests that *CYCD1* expansion traces back to the formation of colonial multicellularity and that an additional duplication event (resulting in four copies) most probably occurred later, after the *G pectorale* and *V carteri* lineage split from the *Tetrabaena* lineage. *CYCD1* expansion in the simple colonial *T. socialis* supports the hypothesis that the expansion was integral for the development of colonial multicellularity. As per other volvocines (Merchant et al. 2007; Prochnik et al. 2010; Hanschen et al. 2016) the repertoire of cyclin-dependent kinases is similar in *T. socialis* (supplementary fig. S8, Supplementary Material online).

In most eukaryotic lineages including the volvocines D-type cyclins interact with the retinoblastoma (*RB*) gene (a notable exception being yeast) (Muller et al. 1994; Umen and Goodenough 2001; Olson et al. 2010; Desvoyes et al. 2014; Narasimha et al. 2014; Li et al. 2016). The volvocine

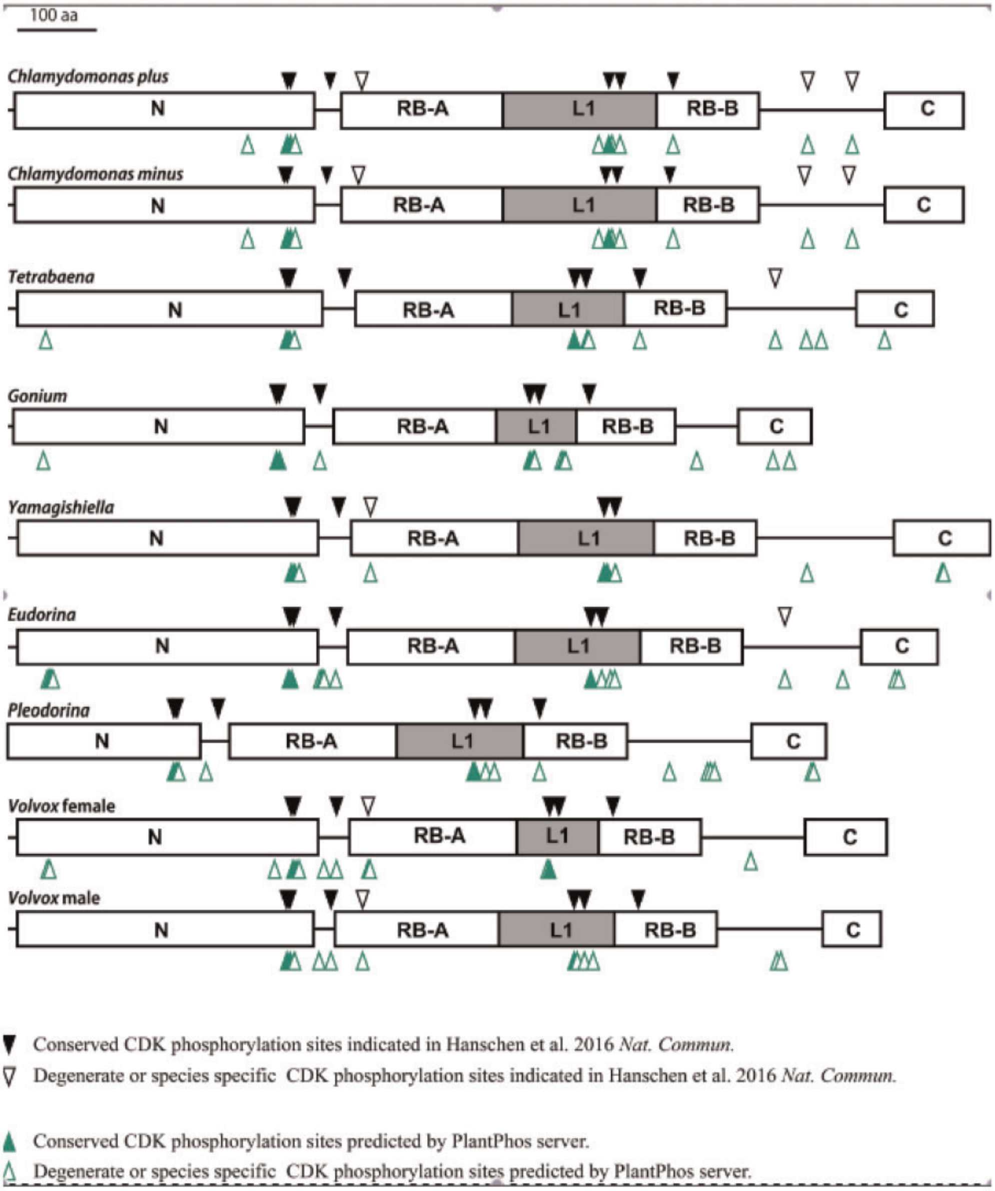
100 aa
Chlamydomonas plus
N
RB-A
L1
RB-B
C
Chlamydomonas minus
Tetrabaena
Gonium
Yamagishiella
Eudorina
Pleodorina
Volvox female
Volvox male
▼ Conserved CDK phosphorylation sites indicated in Hanschen et al. 2016 Nat. Commun.
▽ Degenerate or species specific CDK phosphorylation sites indicated in Hanschen et al. 2016 Nat. Commun.
▲ Conserved CDK phosphorylation sites predicted by PlantPhos server.
△ Degenerate or species specific CDK phosphorylation sites predicted by PlantPhos server.

Fig. 4. Putative phosphorylation targets in volvocine retinoblastoma protein sequences. Putative phosphorylation targets were manually identified and identified using the PlantPhos server.

RB homolog (also known as *MAT3*) is important for the formation of colonial multicellularity (Hanschen et al. 2016) and may be involved in oogamous sexual reproduction in *V. carteri* (Kianianmomeni et al. 2008; Ferris et al. 2010; Hiraide et al. 2013). Transformation of a *MAT3/RB* negative strain of *C. reinhardtii* with the *G. pectorale RB* gene was found to cause colony formation in the unicellular *C. reinhardtii* (Hanschen et al. 2016). The transformative effect of the *G. pectorale RB* gene was proposed to be due to specific differences of the RB protein structure observed between *C. reinhardtii* RB and the *G. pectorale* and *V. carteri* RB proteins. Including *T. socialis*, *G. pectorale*, *V. carteri*, and other published RB protein sequences from *Yamagishiella*, *Eudorina*, and *Pleodorina* (Hiraide et al. 2013) the colonial/multicellular taxa were confirmed to possess a shortened linker region between the RB-A and RB-B domains (supplementary table S20 and fig. S9, Supplementary Material online). Phylogenetic analysis of MAT3/RB proteins was in accordance with Hiraide et al. (2013) (supplementary fig. S10, Supplementary Material online). Following Hanschen et al. (2016) cyclin-dependant kinase (CDK) phosphorylation targets within the linker region between RB-B and the carboxy (C)-terminal domain were identified in *T. socialis*, *Eudorina spp.*, and *Yamagishiella spp.* (fig. 4). In addition, the PlantPhos server was used to computationally identify additional targets within this region in all examined taxa. Findings show that phosphorylation targets within this region are not an exclusive feature of the *C. reinhardtii* MAT3/RB protein and do not correlate with the evolution of multicellularity. Results from the PlantPhos server, in particular, show extensive variation in putative phosphorylation targets amongst MAT3/RB protein sequences, and therefore, altered MAT3/RB phosphorylation may be of significance. However, the absence of phosphorylation targets between the RB-B and RB-C domains cannot be associated with the emergence of multicellularity. To further investigate the *RB* gene, the exon/intron boundaries of *RB* genes from *C. reinhardtii*, *T. socialis*, *G. pectorale*, and *V. carteri* were examined (supplementary fig. S11, Supplementary Material online). Aligned gene structures identified three introns present in *T. socialis*, *G. pectorale*, and *V. carteri* but are absent in *C. reinhardtii*. The exact intron positions and sequence composition in these regions are not identical and, thus, gene structures are not identical. However, the intron gains are at least indicative of a predisposition for intron gain at these specific locations and, of more interest, introns gained in these three regions are limited to the multicellular taxa. It is unknown if these introns are important for transcript regulation or if the intron gains have no influence on function. However, if the former is true, it might contribute towards the transformative ability of *G. pectorale MAT3/RB* to induce colonies in a strain of *C. reinhardtii* lacking *MAT3/RB*.

From analysis of cell cycle genes in *T. socialis* we can conclude that the expansion of D1-type cyclins and alterations to the *MAT3/RB* gene (shortened RB-A to RB-B linker region) trace near to the origin of colonial of group living, which further supports a role for these genetic alterations for a multicellular program of cell division.

The Ubiquitin Proteasomal Pathway

Although other classes of genes identified in the current investigation are of interest in relation to the genetic control of division number (e.g., protein kinases) the UPP is of particular interest because our various analyses identified genes related to most steps in this pathway (supplementary table S21, Supplementary Material online). Targeted degradation of proteins via the UPP is integral to various cellular processes including but not limited to the removal of misfolded proteins, deflagellation and the regulation of cell cycle proteins (Teixeira and Reed 2013). The basic components of the pathway include a series of enzymatic steps to bind ubiquitins (or ubiquitin-like molecules) to proteins whereupon they are shuttled to the proteasome and degraded, releasing free ubiquitin molecules and degraded protein products. Specialized ligases (E3 ligases) that target specific proteins take the form of RING-U-box complexes, which vary but may consist of cullins, transducins, E2 conjugators, and various other molecules (e.g., SKP1 in the Skp1-Cullin-F-box complex) (Bai et al. 1996; Hershko 2005). Complexes such as the Skp1-Cullin-F-Box complex form the basis of key cell cycle checkpoints. Migration to the proteasome may be facilitated and regulated by chaperones such as heat-shock proteins (Noga et al. 1997; Park et al. 2007; Shiber et al. 2013). Aspects of the UPP identified in our analyses (supplementary table S21, Supplementary Material online; fig. 5) include: ubiquitin activation, conjugation and ligation enzymes (E1, E2, and E3), shuttling and regulation of ubiquitinated targets to the proteasome (*HSP70*), proteasomal subunits/ATPases as well as components of the cullin-RING ubiquitin ligase-type complexes. The scope and varying nature (e.g., positive selection vs. gained families) of UPP-related modifications identified in our analyses render targeted functional characterization challenging. Together, however, they point to systemic modifications of this pathway in the multicellular taxa. Regulation via the UPP is important for many aspects of cell biology and modifications of this pathway will have far-reaching and unpredictable consequences and hence it cannot be stated with certainty that UPP-related genes identified in our analyses have specifically contributed towards increased developmental complexity in the lineage. Based on of the well-known role in many organisms of the UPP in cell cycle regulation where it has been shown to be an integral means of regulating cell cycle progression (Goldknopf et al. 1980; Ciechanover et al. 1984; Pagano 1997; Nakayama and Nakayama 2005, 2006a, 2006b; Tu et al. 2012; Vlachostergios et al. 2012; Teixeira and Reed 2013) the modifications of the UPP pathway will impact cell cycle regulation in the volvocines. The UPP is not well characterized in *C. reinhardtii* (von Kampen et al. 1995; Vallentine et al. 2014) although a relationship between the UPP and the *C. reinhardtii* cell cycle has been demonstrated (von Kampen and Wettern 1991; von Kampen et al. 1995). Specifically, the expression of ubiquitin encoding genes is associated with stress responses in *C. reinhardtii* and their expression has furthermore been associated with progression through the cell cycle (von Kampen and Wettern 1991; von Kampen et al. 1995). In *C. reinhardtii*, the number of divisions is highly correlated with the size of the mother cell entering the division phase of the life cycle (Craigie and Cavalier-Smith 1982; Cross and Umen 2015). A key step that evolved early in the evolution of volvocine multicellularity was the modification of the genetic program to control the number of rounds of cell division (Kirk 2005; Herron and Michod 2008). Specifically, the maximum number of divisions is genetically controlled in all colonial volvocines and this feature is even present in the comparatively simple 4-celled *T. socialis* (Arakaki et al. 2013). An intriguing feature of the multicellular volvocines is that the maximum number of cell divisions is a characteristic feature of each species, for example, 8-celled *T. socialis* colonies are rare (requiring three divisions). This suggests that the genetic control program is differentially regulated. Such regulation may occur through various mechanisms, not the least being alternative regulation of key cell cycle regulatory molecules such as RB, which has been shown to be a key regulatory molecule at specific points in the cell cycle (Umen and Goodenough 2001; Olson et al. 2010; Hanschen et al. 2016). Another molecule of interest is CDKG1, because CDKG1 concentration has been correlated with the number of cell divisions in *C. reinhardtii*, and after the final round of palintomic division in *C. reinhardtii* CDKG1 is undetectable (Li et al. 2016). Evidence suggests that CDKG1 undergoes degradation between rounds of palintomic divisions (Li et al. 2016) and hence serves as example of how targeted degradation may control cell division in the volvocines. Modification to the UPP identified in the current investigation associate with even the deeply rooted *T. socialis*, therefore, although we do not present specific evidence these modifications are associated with a genetic program to control the number of divisions in the volvocines, we propose that because of the known role of the UPP in the cell cycle

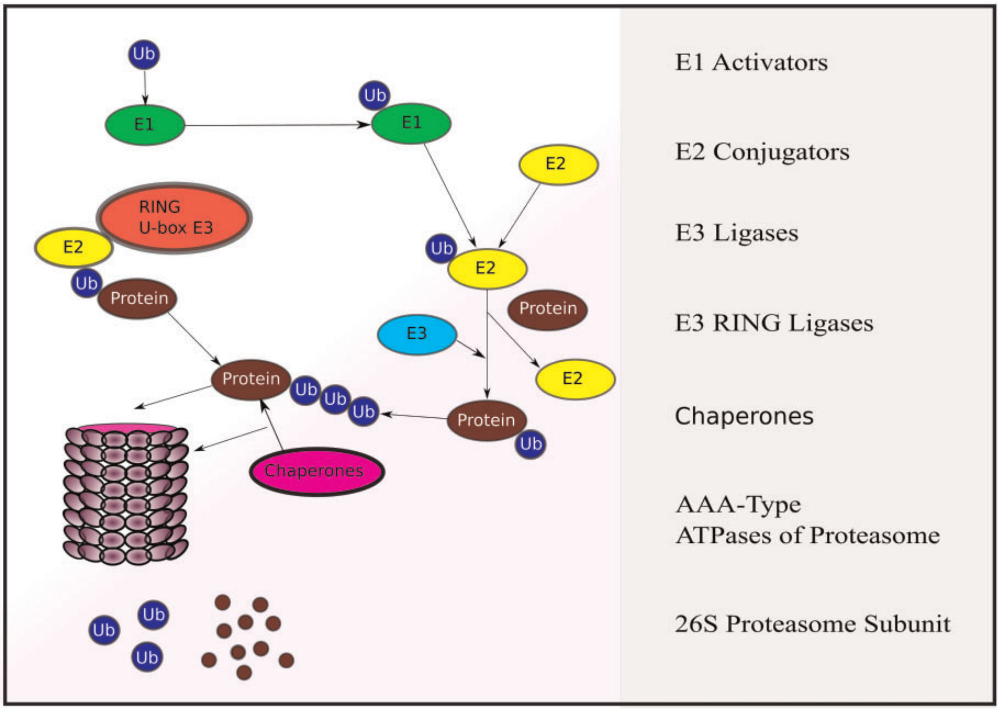

Fig. 5. The core aspects of ubiquitin proteasomal pathway (UPP) and components highlighted in the current study. E1 activators, E2 conjugators, E3 ligases, and the chaperone *HSP70*, a 26S proteasomal regulatory subunit, three AAA-ATPase putative components of the proteasome, and two transducins similar to Cullin-RING E3 ligase complex transducins showed evidence of positive/accelerating selection. Furthermore, a family of RING/U-box E3 type ligases, a family of RPM1 interacting proteins, and an FTSH AAA-type protease were gained at the origin of multicellularity. Together these data point towards modification to multiple aspects of the UPP in the multicellular volvocines relative *Chlamydomonas reinhardtii*.

regulation, including in *C. reinhardtii*, that the modifications are likely to be important for this key evolutionary development. The role of the UPP might conceivably be through regulated degradation of other determinants of the number of divisions such as CDKG1 or MAT3/RB or other cell cycle genes such as the *CYCD1* genes.

Additional Analysis of Volvocine Families Associated with Developmental Complexity

Extracellular Matrix Proteins

The two major protein families found in the extracellular matrix (ECM) of volvocines are the pherophorins and matrix metalloproteases (Godl et al. 1995, 1997; Sumper and Hallmann 1998; Hallmann 1999). A total of 36 pherophorin and 34 matrix metalloproteases (MMPs) genes were identified in *T. socialis*. Separate phylogenies of manually curated pherophorins and MMPs that included proteins from *C. reinhardtii* (Merchant et al. 2007; Blaby et al. 2014), *T. socialis*, *G. pectorale* (Hanschen et al. 2016), and *V. carteri* (Prochnik et al. 2010; Hanschen et al. 2016) highlight orthologues shared throughout the lineage as well as species-specific expansions (supplementary figs. S12 and S13, Supplementary Material online). As per previous findings (Hanschen et al. 2016), widespread expansion of pherophorins and MMP genes are not a feature of the colonial taxa such as *T. socialis* or *G. pectorale* but rather are limited to *V. carteri* and are, therefore, best associated with the expansion of the extracellular matrix rather than initial transformation of the cell wall into an extracellular matrix. The *C. reinhardtii* cell wall hydroxyproline-rich glycoprotein 1 (*GP1*) gene (Adair et al. 1987) is absent from the genomes of *T. socialis*, *G. pectorale*, and *V. carteri* (supplementary table S11, Supplementary Material online). The significance of this is unknown, however, it might be important for understanding the initial transformation of cell wall components into an ECM, which is a feature that all multicellular volvocines share.

RegA-Related Sequences

The evolution of germ-soma division of labor is a key development in complex multicellular volvocines such as *Pleodorina* and *Volvox spp*. The somatic *regenerator* (*regA*) is a putative transcription factor that is expressed in *Volvox carteri* somatic cells where it is thought to suppress cell growth via suppression of chloroplast biogenesis. Homologues of *regA* (regA-like sequences [*rls*] genes) are found throughout the lineage (Harper et al. 1987; Kirk 1994; Choi et al. 1996; Kirk 1997; Meissner et al. 1999; Duncan et al. 2006, 2007; Hanschen et al. 2014, 2016; Grochau-Wright et al. 2017). The genes *rlsA*, *rlsB*, *rlsC*, and *rlsN*/*O* have been identified as syntenically linked to *regA* in a number of Volvocaceae and show a greater phylogenetic relationship to *regA* than other *rls* genes (Hanschen et al. 2014; Grochau-Wright et al. 2017). Together with *regA* these genes have been labeled as *regA*-cluster genes and are absent in *C. reinhardtii* and *G. pectorale***.** A total of eight *RLS* genes were identified in *T. socialis*, which is the same number as in *G. pectorale* but less than *C. reinhardtii* ($n = 12$) and *V. carteri* ($n = 14$). Phylogenetic relationships (supplementary fig. S14, Supplementary Material online) of RLS proteins from *C. reinhardtii*, *T. socialis*, *G. pectorale*, and *V. carteri* f. *nagariensis* show that *regA*-cluster genes are absent in *T. socialis* and confirm the close relationship of RLS1/RlsD proteins to RegA-cluster proteins (Duncan et al. 2006, 2007; Hanschen et al. 2014) (supplementary fig. S8, Supplementary Material online), consistent with the hypothesis that the *regA*-cluster genes evolved from an ancestral *RLS1*/*rlsD* gene (Hanschen et al. 2014, 2016).

Fasciclin-Domain Containing Proteins

A *V. carteri* fasciclin-domain containing protein known as the algal-cell adhesion molecule (algal-CAM) has previously been shown to be required for *V. carteri* embryogenesis and embryo inversion (Huber and Sumper 1994). Protein phylogenetic relationships of fasciclin-domain containing proteins from *C. reinhardtii* ($N = 21$), *T. socialis* ($n = 18$), *G. pectorale* ($n = 17$), and *V. carteri* ($n = 17$) show that 2 *V. carteri* proteins (Jgi|Volca1|127393 and Jgi|Volca1|108097) cluster separately with the annotated *V. carteri* algal-CAM protein (Jgi|Volca1|127389) (fig. 6). Orthologues of algal-CAM were not found in *C. reinhardtii*, *T. socialis*, or *G. pectorale*. While *G. pectorale* undergoes a form of partial inversion *V. carteri* undergoes complete embryo inversion. The absence of algal-CAM orthologues in *C. reinhardtii*, *T. socialis*, and *G. pectorale* supports that algal-CAMs are required for complete embryo inversion in *V. carteri* (Huber and Sumper 1994). As yet it is unknown how CAMs are involved in inversion. A possible role might be that CAMs are required for reconstitution of cellular connectivity after cellular inversion around cytoplasmic bridges and are candidates for further exploration as cell adhesion molecules, which may play an important role in volvocine developmental biology.

Conclusions

We present findings from the nuclear genome of one of the simplest eukaryotic colonial multicellular organisms—the 4-celled volvocine *T. socialis*—as well as comparative genomic analyses with the three other available volvocine genomes. In keeping with previous findings, extensive increases in proteome complexity are not associated with the origin of coloniality in this lineage (Prochnik et al. 2010; Hanschen et al. 2016). However, gene families gained near the origin of coloniality were enriched in genes related to developmental processes such as DNA repair, protein kinase activity, cell adhesion and extracellular functions. Furthermore, gene families gained at the origin of coloniality are enriched in lineage-specific genes (40%) indicating that at the molecular level the transition to multicellularity involves a strong component of lineage specificity. In addition, over 200 genes with evidence of either positive or accelerating selection trace back to *T. socialis*, which suggests that the evolution of volvocine multicellularity has been shaped by numerous genetic loci. Diversification of cell cycle genes as well as modification to components of the UPP were identified that emerged at the beginning of colonial living and associate with the evolution of a genetic program for the control of cell number.

Materials and Methods

Culturing, Nucleic Acid Extraction, and Sequencing

An axenic preparation of *T. socialis* NIES-571 was cultured in tris-acetate-phosphate medium (Gorman and Levine 1965) as per Arakaki et al (2013). Total DNA was extracted using standard *Volvox* DNA isolation method (Miller et al. 1993) and RNA was extracted 2 h into the dark phase of the life cycle with Thermo Fisher Tri reagent (Waltham, MA, USA). Paired-end sequencing libraries for DNA were prepared using the Illumina TruSeq DNA library preparation kit (San Diego, CA, USA) and libraries with two insert sizes were generated (∼180 and ∼450 bp). A 6 kb insert mate-pair library was generated using the Illumina Nextera Mate-Pair Library Kit (San Diego, CA, USA) and automated pulse-field size selection on the Sage BioScience Blue-Pippen (0.75% agarose cassette, Beverley, MA, USA). An RNA-seq library was prepared with the TruSeq stranded-mRNA kit (cat # RS-122-2101). Sequencing was performed on an Illumina HiScanSQ (2 × 100 bp), HiSeq 2500 (2 × 100bp and 2 × 125 bp), and MiSeq (2 × 300 bp).

Genome Assembly

An Allpaths-LG assembly (v48057) (Gnerre et al. 2011) was generated using mate-pair data and paired-end data from short reads (2 × 100 bp) as well as longer MiSeq reads that were trimmed to 150 bp. A SPAdes (v3.1.1) assembly (Bankevich et al. 2012) was generated using only MiSeq data (2 × 300bp), using the multi-Kmer approach (maximum kmer of 127) and subsequent scaffolding with mate-pair data was performed with Opera (v1) (Gao et al. 2011). The Allpaths-LG and SPAdes assemblies were combined using MetaAssembler (Wences and Schatz 2015) with the Allpaths-LG assembly selected as the "master" assembly. Further improvements to the assembly in terms of gap content and accuracy were made using Gapfiller (Boetzer and Pirovano 2012) and Pilon (Walker et al. 2014). Multi-kmer analysis was performed with Jellyfish (v2.0.0) (Vurture et al. 2017), Allpaths-LG and BBtools (http://jgi.doe.gov/data-and-tools/bbtools/, last accessed January 3, 2018).

Transcriptome Assembly and De Novo Gene Prediction

Transcriptome assembly was performed from ∼150 million reads using Bridger *de novo* (r2014-12-01) (Chang et al. 2015) ($k = 25$) and was examined for quality using TransRate (v1.0) (Smith-Unna et al. 2016). Automated gene predictions were performed using MAKER2 (v2.31.3) (Holt and Yandell 2011) using protein homology data and de novo transcripts generated by Bridger assembly. Consensus models were generated by MAKER2 from genes predicted by Augustus (v3.02) (Stanke et al. 2004), Genemark (v4.10) (Borodovsky et al. 2003), and SNAP (v2013-11-29) (Korf 2004). Functional annotations of proteins were derived using HMMER3 (v3.1 b) (Zhang and Wood 2003), InterProScan v5 (Quevillon et al.

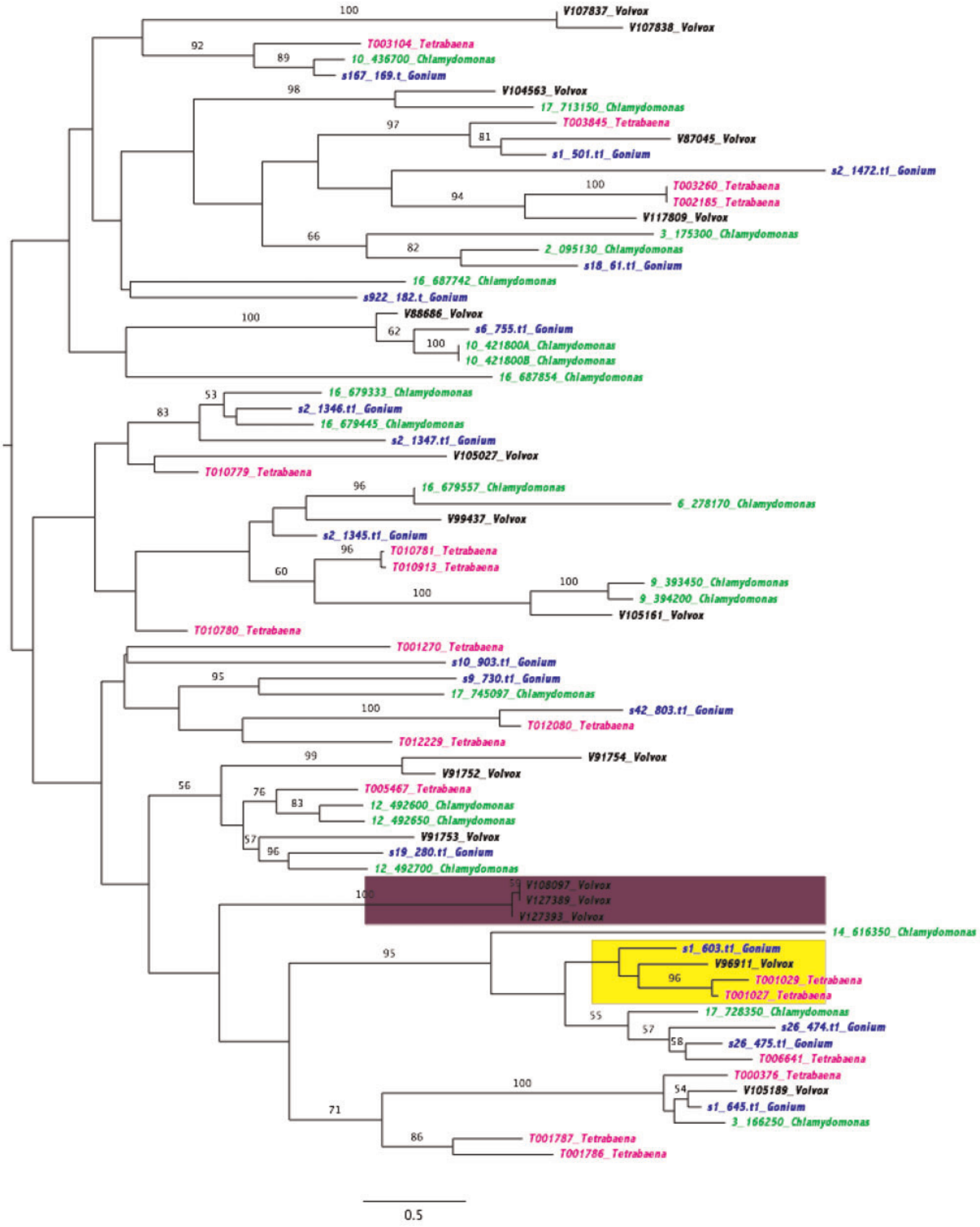

Fig. 6. Phylogeny of fasciclin-domain containing volvocine proteins. Included in the phylogeny are proteins from *Chlamydomonas reinhardtii* (green), *Tetrabaena socialis* (magenta), *Gonium pectorale* (blue), and *Volvox carteri* (black). Highlighted in yellow is a branch identified by Count to have been gained at the origin of multicellularity. Highlighted in purple are *Volvox* algal-CAM proteins.

2005), BLASTP (Altschul et al. 1997) to the UniProt TREMBL database (Schneider et al. 2009) as well as by identifying homology to proteins from sequenced volvocines.

Global Coding Content

Proteomes were collected for all available genome sequenced chlorophyta (supplementary table S2, Supplementary Material online), additional transcriptome-derived proteomes from chlorophyta in the Chlamydomonadales order, yeast and selected model metazoans and plants. Domains were identified using hmmsearch (HMMER v3.1) against Pfam-A. Domain content was compared between *C. reinhardtii* and multicellular taxa with a χ^2-test. Benjamini–Hochberg false discovery rate (FDR) correction (Hochberg and Benjamini 1990) was applied ($q = 0.05$), however, for further analysis of phylogenetic origin domains significant at an alpha of 0.05 were considered. Domains with significant difference in abundance ($P = 0.05$) as well as domains shared by multicellular volvocines are absent in *C. reinhardtii* were identified for presence in opistokonta (yeast and selected metazoa), chlorophyta, and plantae. Orthologue protein families for all available genome-sequenced chlorophyta were generated with OrthoMCL (v2.09) (Li et al. 2003). An RAxML (v8.2.0) phylogeny (Stamatakis 2006) was first generated from 1:1 orthologue families (singletons, $n = 927$). The expansion and contraction of Markov families were examined within the Count environment (v10.04) (Csurös 2010) using Dollo and Wagner's parsimony (symmetric and asymmetric—gain cost of 2). Families gained in the multicellular volvocines and families with expanded protein numbers in the multicellular volvocines were further annotated with manual curation of outputs from BLAST, InterProScan, and Argot2 (Falda et al. 2012) results.

Positive Selection in Orthologue Protein Families

OrthoMCL family clustering was repeated except only volvocine proteomes were included. Identification of families with evidence of positive selection was performed using the Potion pipeline (v1.1.2) (Codeml with site-models) (Yang 2007; Hongo et al. 2015). Only singleton families were examined. Identified families with evidence for positive selection ($q = 0.05$) were examined for Gene Ontology, MapMan and KEGG pathway enrichment via the Algal Functional Annotation Tool (http://pathways.mcdb.ucla.edu/algal/index.html, last accessed January, 2018) (Lopez et al. 2011).

Genomic Elements with Evidence of Conservation or Acceleration

Whole genome alignments for volvocine genomes were performed using LAST (Kiełbasa et al. 2011) and TBA-Multiz (Blanchette et al. 2004) and multiple sequence alignments (MSA) were referenced to the *C. reinhardtii* genome thereby removing alignments that did not include *C. reinhardtii* sequences. The PHAST suite of software (v1.1.) was used to analyze elements within the MSA for evidence of conservation and acceleration. After the PhastCons (Pollard et al. 2010) parameters Gamma and Omega were optimized (supplementary table S22, Supplementary Material online) PhastCons was used to identify discreet highly conserved regions while gene (whole genes including untranslated regions), coding sequences, intergenic regions, and untranslated gene regions ($3'$ and $5'$ ends of genes) were examined in separate analyses for evidence of acceleration in the multicellular taxa using phyloP with Benjamini–Hochberg FDR correction applied (Hochberg and Benjamini 1990). Enrichment analyses were performed as per families with evidence of positive selection (Lopez et al. 2011). Potion was used to examine accelerating genes (identified by either gene-level analysis or CDS-level analysis) for evidence of selection, this time with families containing paralogues included.

Specific Gene Families

Tetrabaena socialis orthologues of cell cycle genes, matrix metalloproteases, pherophorins, and *regA*-related genes were identified. D-type cyclins and the *MAT3/RB* gene were sequenced using capillary sequencing with primers designed from the genome sequence (supplementary table S23, Supplementary Material online). Furthermore, fasciclin-domain containing proteins from all volvocines were identified. Protein phylogenies of these families were generated that included *T. socialis* proteins and where possible curated sets of proteins from *C. reinhardtii*, *G. pectorale*, and *V. carteri* from Hanschen et al (2016) were used. Orthologues of the "Prochnik Table" genes (Prochnik et al. 2010; Hanschen et al. 2016) were identified in *T. socialis* (supplementary table S24, Supplementary Material online).

Data Availability

Raw data used in this study have been deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra; last accessed January 3, 2018) with the accession number SRX3367147 (paired-end genomics DNA sequences, 2 × 100 bp), SRX336146 (Miseq paired-end sequences, 2 × 300 bp), SRX3367144 (RNA sequences), and SRX3367145 (mate-pair sequences). The annotated genome assembly has been deposited in the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/; last accessed January 3, 2018) with the accession number of PGGS00000000. The version described in this paper is version PGGS01000000.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by a National Aeronautics and Space Administration grant to P.M.D, P.J.F. and R.E.M (NNX13AH41G). B.J.S.C.O and R.E.M were supported by a National Science Foundation grant (MCB-1412395). P.M.D is funded by the University of Witwatersrand Department of Science and Technology National Research Foundation Centre of Excellence in Palaeosciences. Sequencing was partly supported by a University of Witwatersrand, Faculty of Health Science grant awarded to J.F.

References

Adair WS, Steinmetz SA, Mattson DM, Goodenough UW, Heuser JE. 1987. Nucleated assembly of Chlamydomonas and Volvox cell walls. *J Cell Biol.* 105(5): 2373–2382.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17): 3389–3402.

Arakaki Y, Kawai-Toyooka H, Hamamura Y, Higashiyama T, Noga A, Hirono M, Olson BJ, Nozaki H. 2013. The simplest integrated multicellular organism unveiled. *PLoS One* 8(12): e81641.

Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ. 1996. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* 86(2): 263–274.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5): 455–477.

Bar-Nun S, Glickman MH. 2012. Proteasomal AAA-ATPases: structure and function. *Biochim Biophys Acta* 1823(1): 67–82.

Bisova K, Krylov DM, Umen JG. 2005. Genome-wide annotation and expression profiling of cell cycle regulatory genes in *Chlamydomonas reinhardtii*. *Plant Physiol.* 137(2): 475–491.

Blaby IK, Blaby-Haas CE, Tourasse N, Hom EF, Lopez D, Aksoy M, Grossman A, Umen J, Dutcher S, Porter M. 2014. The Chlamydomonas genome project: a decade on. *Trends Plant Sci.* 19(10): 672–680.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* 14(4): 708–715.

Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13(6): R56.

Borodovsky M, Lomsadze A, Ivanov N, Mills R. 2003. Eukaryotic gene prediction using GeneMark.hmm. *Curr Protoc Bioinformatics*. 1:4.6:4.6.1-4.6.1.

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the uniprot knowledgebase: how to use the entry view. *Methods Mol Biol.* 1374:23–54.

Branzei D, Foiani M. 2008. Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol.* 9(4): 297–308.

Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X. 2015. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* 16(1): 30.

Cheng Q, Pappas V, Hallmann A, Miller SM. 2005. Hsp70A and GlsA interact as partner chaperones to regulate asymmetric division in Volvox. *Dev Biol.* 286(2): 537–548.

Choi G, Przybylska M, Straus D. 1996. Three abundant germ line-specific transcripts in *Volvox carteri* encode photosynthetic proteins. *Curr Genet.* 30(4): 347–355.

Choi J, Husain M. 2006. Calmodulin-mediated cell cycle regulation: new mechanisms for old observations. *Cell Cycle* 5(19): 2183–2186.

Ciechanover A, Finley D, Varshavsky A. 1984. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* 37(1): 57–66.

Cock JM, Sterck L, Rouze P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, et al. 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* 465(7298): 617–621.

Coleman AW. 2012. A comparative analysis of the Volvocaceae (Chlorophyta)(1). *J Phycol.* 48(3): 491–513.

Coleman AW. 1999. Phylogenetic analysis of "Volvocacae" for comparative genetic studies. *Proc Natl Acad Sci U S A.* 96(24): 13892–13897.

Craigie RA, Cavalier-Smith T. 1982. Cell volume and the control of the Chlamydomonas cell cycle. *J Cell Sci.* 54:173–191.

Cross FR, Umen JG. 2015. The Chlamydomonas cell cycle. *Plant J.* 82(3): 370–392.

Csurös M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26(15): 1910–1912.

de Mendoza A, Ruiz-Trillo I. 2011. The mysterious evolutionary origin for the *GNE* gene and the root of bilateria. *Mol Biol Evol.* 28(11): 2987–2991.

Desvoyes B, de Mendoza A, Ruiz-Trillo I, Gutierrez C. 2014. Novel roles of plant RETINOBLASTOMA-RELATED (RBR) protein in cell proliferation and asymmetric cell division. *J Exp Bot.* 65(10): 2657–2666.

Duncan L, Nishii I, Harryman A, Buckley S, Howard A, Friedman NR, Miller SM. 2007. The VARL gene family and the evolutionary origins of the master cell-type regulatory gene, regA, in *Volvox carteri*. *J Mol Evol.* 65(1): 1–11.

Duncan L, Nishii I, Howard A, Kirk D, Miller SM. 2006. Orthologs and paralogs of *regA*, a master cell-type regulatory gene in *Volvox carteri*. *Curr Genet.* 50(1): 61–72.

Durand PM, Sym S, Michod RE. 2016. Programmed cell death and complexity in microbial systems. *Curr Biol.* 26(13): R587–R593.

Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, Cilia E, Velasco R, Fontana P. 2012. Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* 13(Suppl 4): S14.

Featherston J, Arakaki Y, Nozaki H, Durand PM, Smith DR. 2016. Inflated organelle genomes and circular-mapping mtDNA probably existed at the origin of coloniality in volvocine green algae. *Eur J Phycol.* 51:369–377.

Ferris P, Olson BJSC, De Hoff PL, Douglass S, Casero D, Prochnik S, Geng S, Rai R, Grimwood J, Schmutz J, et al. 2010. Evolution of an expanded sex-determining locus in Volvox. *Science* 328(5976): 351–354.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, et al. 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42(Database issue): D222–D230.

Francino MP. 2005. An adaptive radiation model for the origin of new gene functions. *Nat Genet.* 37(6): 573–577.

Gao S, Sung WK, Nagarajan N. 2011. Opera: reconstructing optimal genomic scaffolds with high-throughput paired-end sequences. *J Comput Biol.* 18(11): 1681–1691.

Glockner G, Lawal HM, Felder M, Singh R, Singer G, Weijer CJ, Schaap P. 2016. The multicellularity genes of dictyostelid social amoebas. *Nat Commun.* 7:12085.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* 108(4): 1513–1518.

Godl K, Hallmann A, Rappel A, Sumper M. 1995. Pherophorins: a family of extracellular matrix glycoproteins from Volvox structurally related to the sex-inducing pheromone. *Planta* 196(4): 781–787.

Godl K, Hallmann A, Wenzl S, Sumper M. 1997. Differential targeting of closely related ECM glycoproteins: the pherophorin family from Volvox. *Embo J.* 16(1): 25–34.

Goebl MG, Yochem J, Jentsch S, McGrath JP, Varshavsky A, Byers B. 1988. The yeast cell cycle gene *CDC34* encodes a ubiquitin-conjugating enzyme. *Science* 241(4871): 1331–1335.

Goldknopf IL, Sudhakar S, Rosenbaum F, Busch H. 1980. Timing of ubiquitin synthesis and conjugation into protein A24 during the HeLa cell cycle. *Biochem Biophys Res Commun.* 95(3): 1253–1260.

Gorman DS, Levine RP. 1965. Cytochrome *f* and plastocyanin: their sequence in the photosynthetic eletron transport chain of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci U S A.* 54(6): 1665–1669.

Grochau-Wright ZI, Hanschen ER, Ferris PJ, Hamaji T, Nozaki H, Olson BJSC, Michod RE. 2017. Genetic basis for soma is present in undifferentiated volvocine green algae. *J Evol Biol.* 30(6): 1205–1218.

Hallmann A. 1999. Enzymes in the extracellular matrix of Volvox: an inducible, calcium-dependent phosphatase with a modular composition. *J Biol Chem.* 274(3): 1691–1697.

Hanschen ER, Ferris PJ, Michod RE. 2014. Early evolution of the genetic basis for soma in the volvocaceae. *Evolution* 68(7): 2014–2025.

Hanschen ER, Marriage TN, Ferris PJ, Hamaji T, Toyoda A, Fujiyama A, Neme R, Noguchi H, Minakuchi Y, Suzuki M, et al. 2016. The Gonium pectorale genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat Commun.* 7:11370.

Haring MA, Siderius M, Jonak C, Hirt H, Walton KM, Musgrave A. 1995. Tyrosine phosphatase signalling in a lower plant: cell-cycle and oxidative stress-regulated expression of the Chlamydomonas eugametos *VH-PTP13* gene. *Plant J.* 7(6): 981–988.

Harper JF, Huson KS, Kirk DL. 1987. Use of repetitive sequences to identify DNA polymorphisms linked to *regA*, a developmentally important locus in Volvox. *Genes Dev.* 1(6): 573–584.

Herron MD. 2009. Many from one: lessons from the volvocine algae on the evolution of multicellularity. *Commun Integr Biol.* 2(4): 368–370.

Herron MD, Hackett JD, Aylward FO, Michod RE. 2009. Triassic origin and early radiation of multicellular volvocine algae. *Proc Natl Acad Sci U S A.* 106(9): 3254–3258.

Herron MD, Michod RE. 2008. Evolution of complexity in the volvocine algae: transitions in individuality through Darwin's eye. *Evolution* 62:436–451.

Hershko A. 2005. The ubiquitin system for protein degradation and some of its roles in the control of the cell-division cycle (Nobel lecture). *Angew Chem Int Ed Engl.* 44(37): 5932–5943.

Hiraide R, Kawai-Toyooka H, Hamaji T, Matsuzaki R, Kawafune K, Abe J, Sekimoto H, Umen J, Nozaki H. 2013. The evolution of male-female sexual dimorphism predates the gender-based divergence of the mating locus gene *MAT3/RB*. *Mol Biol Evol.* 30(5): 1038–1040.

Hochberg Y, Benjamini Y. 1990. More powerful procedures for multiple significance testing. *Stat Med.* 9(7): 811–818.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.

Hongo JA, de Castro GM, Cintra LC, Zerlotini A, Lobo FP. 2015. POTION: an end-to-end pipeline for positive Darwinian selection detection in genome-scale data through phylogenetic comparison of protein-coding genes. *BMC Genomics* 16:567.

Huber O, Sumper M. 1994. Algal-CAMs: isoforms of a cell adhesion molecule in embryos of the alga Volvox with homology to Drosophila fasciclin I. *EMBO J.* 13(18): 4212–4222.

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform.* 12(1): 41–51.

Janski N, Masoud K, Batzenschlager M, Herzog E, Evrard JL, Houlne G, Bourge M, Chaboute ME, Schmit AC. 2012. The GCP3-interacting proteins GIP1 and GIP2 are required for gamma-tubulin complex protein localization, spindle integrity, and chromosomal stability. *Plant Cell* 24(3): 1171–1187.

Kao HT, Capasso O, Heintz N, Nevins JR. 1985. Cell cycle control of the human *HSP70* gene: implications for the role of a cellular E1A-like function. *Mol Cell Biol.* 5(4): 628–633.

Kianianmomeni A. 2015. Potential impact of gene regulatory mechanisms on the evolution of multicellularity in the volvocine algae. *Commun Integr Biol.* 8(2): e1017175.

Kianianmomeni A, Nematollahi G, Hallmann A. 2008. A gender-specific retinoblastoma-related protein in *Volvox carteri* implies a role for the retinoblastoma protein family in sexual development. *Plant Cell* 20(9): 2399–2419.

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res.* 21(3): 487–493.

Kirk DL. 1999. Evolution of multicellularity in the volvocine algae. *Curr Opin Plant Biol.* 2(6): 496–501.

Kirk DL. 1997. The genetic program for germ-soma differentiation in Volvox. *Annu Rev Genet.* 31:359–380.

Kirk DL. 1994. Germ cell specification in *Volvox carteri*. *Ciba Found Symp.* 182:2–15. discussion 15-30.

Kirk DL. 2001. Germ-soma differentiation in volvox. *Dev Biol.* 238(2): 213–223.

Kirk DL. 2005. A twelve-step program for evolving multicellularity and a division of labor. *Bioessays* 27(3): 299–310.

Kirk DL, Harper JF. 1986. Genetic, biochemical, and molecular approaches to Volvox development and evolution. *Int Rev Cytol.* 99:217–293.

Kirk MM, Stark K, Miller SM, Muller W, Taillon BE, Gruber H, Schmitt R, Kirk DL. 1999. *regA*, a Volvox gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. *Development* 126:639–647.

Klie S, Nikoloski Z. 2012. The choice between MapMan and gene ontology for automated gene function prediction in plant science. *Front Genet.* 3:115.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.

Larson A, Kirk MM, Kirk DL. 1992. Molecular phylogeny of the volvocine flagellates. *Mol Biol Evol.* 9(1): 85–105.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13(9): 2178–2189.

Li Y, Liu D, Lopez-Paz C, Olson BJSC, Umen JG. 2016. A new class of cyclin dependent kinase in Chlamydomonas is required for coupling cell size to cell division. *Elife* 5:e10767.

Lopez D, Casero D, Cokus SJ, Merchant SS, Pellegrini M. 2011. Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics* 12:282.

Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180(2): 933–943.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 17(11): 704–714.

Maliet O, Shelton DE, Michod RE. 2015. A model for the origin of group reproduction during the evolutionary transition to multicellularity. *Biol Lett.* 11(6): 20150157.

Marcais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurences of k-mers. *Bioinformatics* 27(6): 764–770.

Maynard Smith J, Szathmáry E. 1995. The major transitions in evolution. Oxford; New York: W.H. Freeman Spektrum.

Meissner M, Stark K, Cresnar B, Kirk DL, Schmitt R. 1999. Volvox germline-specific genes that are putative targets of RegA repression encode chloroplast proteins. *Curr Genet.* 36(6): 363–370.

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L, et al. 2007. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318(5848): 245–250.

Miller SM, Kirk DL. 1999. *glsA*, a Volvox gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. *Development* 126(4): 649–658.

Miller SM, Schmitt R, Kirk DL. 1993. *Jordan*, an active Volvox transposable element similar to higher plant transposons. *Plant Cell* 5(9): 1125–1138.

Miller WT. 2012. Tyrosine kinase signaling and the emergence of multicellularity. *Biochim Biophys Acta* 1823(6): 1053–1057.

Muller H, Lukas J, Schneider A, Warthoe P, Bartek J, Eilers M, Strauss M. 1994. Cyclin D1 expression is regulated by the retinoblastoma protein. *Proc Natl Acad Sci U S A.* 91(8): 2945–2949.

Nakada T, Misawa K, Nozaki H. 2008. Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. *Mol Phylogenet Evol.* 48(1): 281–291.

Nakayama KI, Nakayama K. 2005. Regulation of the cell cycle by SCF-type ubiquitin ligases. *Semin Cell Dev Biol.* 16(3): 323–333.

Nakayama KI, Nakayama K. 2006. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* 6(5): 369–381.

Nakayama KI, Nakayama K. 2006. Ubiquitin system regulating G1 and S phases of cell cycle. *Tanpakushitsu Kakusan Koso* 51(10 Suppl): 1362–1369.

Narasimha AM, Kaulich M, Shapiro GS, Choi YJ, Sicinski P, Dowdy SF. 2014. Cyclin D activates the Rb tumor suppressor by mono-phosphorylation. *Elife* 3: e02872.

Nishii I, Miller SM. 2010. Volvox: simple steps to developmental complexity?. *Curr Opin Plant Biol.* 13(6): 646–653.

Nishii I, Ogihara S, Kirk DL. 2003. A kinesin, *invA*, plays an essential role in volvox morphogenesis. *Cell* 113(6): 743–753.

Noga M, Hayashi T, Tanaka J. 1997. Gene expressions of ubiquitin and *hsp70* following focal ischaemia in rat brain. *Neuroreport* 8(5): 1239–1241.

Nozaki H. 1986. Sexual reproduction in *Gonium sociale* (Chlorophta, Volvocales). *Phycologia* 25(1): 29–35.

Nozaki H, Misawa K, Kajita T, Kato M, Nohara S, Watanabe MM. 2000. Origin and evolution of the colonial volvocales (Chlorophyceae) as inferred from multiple, chloroplast gene sequences. *Mol Phylogenet Evol.* 17(2): 256–268.

Nozaki H, Ueki N, Isaka N, Saigo T, Yamamoto K, Matsuzaki R, Takahashi F, Wakabayashi KI, Kawachi M. 2016. A new morphological type of Volvox from Japanese large lakes and recent divergence of this type and *V. ferrisii* in two different freshwater habitats. *PLoS One* 11(11): e0167148.

Nozaki H, Yamada TK, 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 Evol Biol.* 14(1): 37.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44:D733–D745.

Olson BJ, Nedelcu AM. 2016. Co-option during the evolution of multicellular and developmental complexity in the volvocine green algae. *Curr Opin Genet Dev.* 39:107–115.

Olson BJ, Oberholzer M, Li Y, Zones JM, Kohli HS, Bisova K, Fang SC, Meisenhelder J, Hunter T, Umen JG. 2010. Regulation of the Chlamydomonas cell cycle by a stable, chromatin-associated retinoblastoma tumor suppressor complex. *Plant Cell* 22(10): 3331–3347.

Pagano M. 1997. Cell cycle regulation by the ubiquitin pathway. *Faseb J* 11(13): 1067–1075.

Park SH, Bolender N, Eisele F, Kostova Z, Takeuchi J, Coffino P, Wolf DH. 2007. The cytoplasmic Hsp70 chaperone machinery subjects misfolded and endoplasmic reticulum import-incompetent proteins to degradation via the ubiquitin-proteasome system. *Mol Biol Cell.* 18(1): 153–165.

Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9): 1061–1067.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20(1): 110–121.

Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, et al. 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329(5988): 223–226.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33(Web Server issue): W116–W120.

Rashidi A, Shelton DE, Michod RE. 2015. A Darwinian approach to the origin of life cycles with group properties. *Theor Popul Biol.* 102:76–84.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319(5859): 64–69.

Sathe S, Durand PM. 2016. Cellular aggregation in *Chlamydomonas* (Chlorophyceae) is chimaeric and depends on traits like cell size and motility. *Eur J Phycol.* 51(2): 129–138.

Schneider M, Lane L, Boutet E, Lieberherr D, Tognolli M, Bougueleret L, Bairoch A. 2009. The UniProtKB/Swiss-Prot knowledgebase and its plant proteome annotation program. *J Proteomics* 72(3): 567–573.

Schultheiss KP, Suga H, Ruiz-Trillo I, Miller WT. 2012. Lack of Csk-mediated negative regulation in a unicellular SRC kinase. *Biochemistry* 51(41): 8267–8277.

Sebe-Pedros A, Ballare C, Parra-Acero H, Chiva C, Tena JJ, Sabido E, Gomez-Skarmeta JL, Di Croce L, Ruiz-Trillo I. 2016. The dynamic regulatory genome of Capsaspora and the origin of animal multicellularity. *Cell* 165(5): 1224–1237.

Sharpe SC, Eme L, Brown MW, Roger AJ. 2015. Timing the origins of multicellular eukaryotes through phylogenomics and relaxed molecular clock analysis. In: Trillo IR, Nedelcu AM, editors. Evolutionary transitions to multicellular life. Netherlands: Springer, pp. 3–29.

Shiber A, Breuer W, Brandeis M, Ravid T. 2013. Ubiquitin conjugation triggers misfolded protein sequestration into quality control foci when Hsp70 chaperone levels are limiting. *Mol Biol Cell.* 24(13): 2076–2087.

Smith DR, Hamaji T, Olson BJ, Durand PM, Ferris P, Michod RE, Featherston J, Nozaki H, Keeling PJ. 2013. Organelle genome complexity scales positively with organism size in volvocine green algae. *Mol Biol Evol.* 30(4): 793–797.

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

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26(8): 1134–1144.

Stajich JE, Wilke SK, Ahren D, Au CH, Birren BW, Borodovsky M, Burns C, Canback B, Casselton LA, Cheng CK, et al. 2010. Insights into evolution of multicellular fungi from the assembled chromosomes of the mushroom *Coprinopsis cinerea* (*Coprinus cinereus*). *Proc Natl Acad Sci U S A.* 107(26): 11889–11894.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21): 2688–2690.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32(Web Server issue): W309–W312.

Starr RC. 1968. Cellular differentiation in Volvox. *Proc Natl Acad Sci U S A.* 59(4): 1082–1088.

Strzalka W, Ziemienowicz A. 2011. Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. *Ann Bot.* 107(7): 1127–1140.

Sumper M, Hallmann A. 1998. Biochemistry of the extracellular matrix of Volvox. *Int Rev Cytol.* 180:51–85.

Teixeira LK, Reed SI. 2013. Ubiquitin ligases and cell cycle control. *Annu Rev Biochem.* 82:387–414.

Tu Y, Chen C, Pan J, Xu J, Zhou ZG, Wang CY. 2012. The Ubiquitin Proteasome Pathway (UPP) in the regulation of cell cycle control and DNA damage repair and its implication in tumorigenesis. *Int J Clin Exp Pathol.* 5(8): 726–738.

Umen JG, Goodenough UW. 2001. Control of cell division by a retinoblastoma protein homolog in Chlamydomonas. *Genes Dev.* 15(13): 1652–1661.

Vallentine P, Hung C, Xie J, van Hoewyk D. 2014. The ubiquitin-proteasome pathway protects *Chlamydomonas reinhardtii* against selenite toxicity, but is impaired as reactive oxygen species accumulate. *AoB Plants* 6. doi: 10.1093/aobpla/plu062.

Vlachostergios PJ, Voutsadakis IA, Papandreou CN. 2012. The ubiquitin-proteasome system in glioma cell cycle control. *Cell Div.* 7(1): 18.

von Kampen J, Nielander U, Wettern M. 1995. Expression of ubiquitin genes in *Chlamydomonas reinhardtii*: involvement in stress response and cell cycle. *Planta* 197(3): 528–534.

von Kampen J, Wettern M. 1991. Ubiquitin-encoding mRNA and mRNA recognized by genes encoding ubiquitin-conjugating enzymes are differentially expressed in division-synchronized cultures of *Chlamydomonas reinhardtii*. *Eur J Cell Biochem.* 55:312–317.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202–2204.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11): e112963.

Wences AH, Schatz MC. 2015. Metassembler: merging and optimizing de novo genome assemblies. *Genome Biol.* 16:207.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8): 1586–1591.

Zhang Z, Wood WI. 2003. A profile hidden Markov model for signal peptides generated by HMMER. *Bioinformatics* 19(2): 307–308.