

Gene loss, co-option and the evolution of developmental complexity in the volvocine algae

by

Leticia Berenice Jiménez Marín

B.S., Universidad Nacional Autónoma de México, 2012

M.S., Universidad Nacional Autónoma de México, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Abstract

The evolution of multicellularity is a Major Evolutionary Transition that has occurred in each domain of life and has evolved independently numerous times in diverse eukaryotic lineages. Multicellularity is associated with changes to developmental complexity; this is thought to be a consequence of multicellular species collectively organizing into higher level individuals. Hence, extant descendants of early multicellular lifeforms exhibit a wealth of morphological and developmental innovations. Despite the recurring nature and biological importance of multicellular evolution, determining the molecular mechanisms underlying it presents several challenges. Most notably, in most multicellular lineages there are long divergence times between multicellular species and their unicellular relatives that obscures the genetic signature of multicellular evolution. Moreover, most systems used to study the evolution of multicellularity lack important experimental toolkits to dissect the molecular basis of multicellular evolution. However, the volvocine algae have several advantages to the study the molecular bases of the transition to multicellularity and developmental complexity. Extant volvocine species exhibit a stepwise increase in multicellular and developmental complexity; member species have phenotypes ranging from unicellular, to undifferentiated multicellular, and to differentiated multicellular. This group evolved multicellularity relatively recently (~250 MYA) resulting in a high degree of genomic similarity and conserved synteny that suggests few genetic changes are required for lineages to undergo dramatic shifts in multicellular and developmental complexity. Multicellularity in the volvocines is thought to have evolved by co-option of key genes, such as cell cycle and developmental regulators. However, these co-opted genes appear to be an exception rather than a rule underlying developmental features observed in multicellular volvocine species (Chapter 1). In this work, comparative genomics (Chapters 2 and 3),

mathematical modeling (Chapter 2), molecular biology (Chapters 2 and 5), comparative transcriptomics (Chapter 4), and reverse genetics (Chapter 5) approaches are used to provide an integrative view on how gene loss and co-option have impacted the volvocine transition to multicellularity and increased developmental complexity, with implications for other systems. Results show that genetic co-option of *fsII* impacted the evolution of cell-cell adhesion – a key aspect of multicellular life, in undifferentiated multicellular *Gonium pectorale*. Furthermore, a putative environmental response gene, *rlsD*, has overlapping targets with its paralog, a cell differentiation regulator in *Volvox carteri*. Importantly, further results demonstrate that significant loss of conserved genes occurred in the volvocine lineage. These bursts of gene loss likely occurred after the divergence between the lineage of unicellular *Chlamydomonas* and the ancestor of multicellular volvocines, potentially setting the stage for multicellularity to evolve. Moreover, gene losses are shared across two volvocine families that independently evolved cell differentiation programs, further suggesting that gene loss might have facilitated the evolution of complex phenotypes across the volvocine lineage. Gene loss events likely impact the evolutionary paths and constraints of other genes, potentially resulting in molecular network rewiring and stabilization of the multicellular state. As means to expand these findings to other systems, a mathematical model is developed that suggests the evolution of many multicellular traits might be the product of gene-regulatory and functional innovations that do not necessarily require expansion of genetic repertoires.

Gene loss, co-option and the evolution of developmental complexity in the volvocine algae

by

Leticia Berenice Jiménez Marín

B.S., Universidad Nacional Autónoma de México, 2012

M.S., Universidad Nacional Autónoma de México, 2017

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:
Major Professor
Bradley J.S.C. Olson, Ph.D.

Copyright

© Leticia Berenice Jiménez Marín 2023.

Abstract

The evolution of multicellularity is a Major Evolutionary Transition that has occurred in each domain of life and has evolved independently numerous times in diverse eukaryotic lineages. Multicellularity is associated with changes to developmental complexity; this is thought to be a consequence of multicellular species collectively organizing into higher level individuals. Hence, extant descendants of early multicellular lifeforms exhibit a wealth of morphological and developmental innovations. Despite the recurring nature and biological importance of multicellular evolution, determining the molecular mechanisms underlying it presents several challenges. Most notably, in most multicellular lineages there are long divergence times between multicellular species and their unicellular relatives that obscures the genetic signature of multicellular evolution. Moreover, most systems used to study the evolution of multicellularity lack important experimental toolkits to dissect the molecular basis of multicellular evolution. However, the volvocine algae have several advantages to the study the molecular bases of the transition to multicellularity and developmental complexity. Extant volvocine species exhibit a stepwise increase in multicellular and developmental complexity; member species have phenotypes ranging from unicellular, to undifferentiated multicellular, and to differentiated multicellular. This group evolved multicellularity relatively recently (~250 MYA) resulting in a high degree of genomic similarity and conserved synteny that suggests few genetic changes are required for lineages to undergo dramatic shifts in multicellular and developmental complexity. Multicellularity in the volvocines is thought to have evolved by co-option of key genes, such as cell cycle and developmental regulators. However, these co-opted genes appear to be an exception rather than a rule underlying developmental features observed in multicellular volvocine species (Chapter 1). In this work, comparative genomics (Chapters 2 and 3),

mathematical modeling (Chapter 2), molecular biology (Chapters 2 and 5), comparative transcriptomics (Chapter 4), and reverse genetics (Chapter 5) approaches are used to provide an integrative view on how gene loss and co-option have impacted the volvocine transition to multicellularity and increased developmental complexity, with implications for other systems. Results show that genetic co-option of *fsII* impacted the evolution of cell-cell adhesion – a key aspect of multicellular life, in undifferentiated multicellular *Gonium pectorale*. Furthermore, a putative environmental response gene, *rlsD*, has overlapping targets with its paralog, a cell differentiation regulator in *Volvox carteri*. Importantly, further results demonstrate that significant loss of conserved genes occurred in the volvocine lineage. These bursts of gene loss likely occurred after the divergence between the lineage of unicellular *Chlamydomonas* and the ancestor of multicellular volvocines, potentially setting the stage for multicellularity to evolve. Moreover, gene losses are shared across two volvocine families that independently evolved cell differentiation programs, further suggesting that gene loss might have facilitated the evolution of complex phenotypes across the volvocine lineage. Gene loss events likely impact the evolutionary paths and constraints of other genes, potentially resulting in molecular network rewiring and stabilization of the multicellular state. As means to expand these findings to other systems, a mathematical model is developed that suggests the evolution of many multicellular traits might be the product of gene-regulatory and functional innovations that do not necessarily require expansion of genetic repertoires.

Table of Contents

List of Figures.....	xi
List of Tables.....	xiii
Acknowledgements.....	xiv
Chapter 1 - The curious case of multicellularity in the volvocine algae	1
Abstract	2
Introduction	2
Variations on a theme – evolution of development in the volvocine algae	4
How can developmental complexity arise from a widely shared genomic repertoire?	10
Combinatorial co-option and creative destruction.....	13
Figures and tables	15
Chapter 2 - Gene loss at a transition to multicellularity	18
Abstract	19
Introduction	19
Results	22
Contraction outpaces expansion of conserved genetic and functional elements as complexity increases.....	23
Neither co-option nor gene age account for the all of volvocine developmental complexity	28
Gene loss is a driver for regulatory innovation.....	30
Discussion	31
Methods	35
Strains and genomes.....	35
Algae growth conditions and library preparation	36
Genome assembly	37
Evidence-based gene prediction.....	37
Identification of orthologous groups.....	38
Identification of Pfam domains.....	38
Analysis of other genomic features.....	38
Identification of gene count trends.....	39
Identification of positively selected genes and whole-genome-based phylogeny	40
Characterization of gene loss patterns.....	40
Identification and analysis of known co-opted genes	41
Prediction of lineage specific genes.....	42
Identification of protein disorder and protein binding site disorder	42
2D gel electrophoresis and silver stain.....	42
Western blot and antibodies.....	43
Phylogenetic analysis	44
Regulatory network model	44
Figures and tables	46
Chapter 3 - Gene loss during independent gains of cellular differentiation in the volvocine algae	84
Abstract	85
Introduction	85

Results	88
The <i>Astrephomene</i> genome is similar to that of other volvocines	88
Contraction of conserved genetic elements is shared across the volvocine lineage	89
Gene loss events likely occurred at the ancestor of Goniaceae+Volvocaceae	91
Expanded Cyclin D repertoire is conserved across Goniaceae and Volvocaceae	92
Discussion	93
Methods	95
Strains and genomes	95
Identification of orthologous groups and Pfam domains	96
Analysis of other genomic features	96
Identification of gene count trends	97
Identification of positively selected genes and generation of whole-genome phylogenetic tree	97
Characterization of gene loss patterns	97
Identification and analysis of known co-opted genes and prediction of lineage specific genes	98
Figures and tables	99
Chapter 4 - Transcriptional regulation of cellular differentiation by <i>rlsD</i> in <i>Volvox carteri</i>	117
Introduction	118
Summary of previous work (Ortega Escalante, 2018)	120
Results	121
<i>rlsD</i> acts as both as a gene activator and repressor	121
Genes involved in ribosomal biogenesis, photosynthesis and ECM modification are <i>rlsD</i> targets	122
<i>rlsD</i> overexpression results in biased regulation of cell-type related genes	123
Growth, reproduction and stress related genes are regulated by <i>rlsD</i>	125
Discussion	127
Methods	130
Library preparation, RNA-sequencing and read assembly	130
Gene expression analysis and cell-type related expression	130
PFAM enrichment	131
Assembly of candidate gene sets for differential expression analysis	131
Figures and tables	134
Chapter 5 - Cell adhesion gene <i>fsII</i> is important for multicellularity in colonial alga <i>Gonium pectorale</i>	158
Introduction	159
Results	161
<i>fsII</i> is an ortholog of a cell adhesion molecule in <i>Volvox</i> with differential regulation in <i>Chlamydomonas</i>	161
Antisense RNA of Gp <i>FsII</i> results in loss of multicellularity in <i>Gonium</i>	162
Heterologous expression of Gp <i>FsII</i> produces multicellular <i>Chlamydomonas</i>	162
Cytokinesis is disrupted in gain-of-function cell-wall compromised <i>Chlamydomonas</i> strains	163
Discussion	164
Methods	166
Strains and culture conditions	166

Cell cycle synchrony assay	166
Cloning	167
Nuclear transformation of <i>Gonium</i> and <i>Chlamydomonas</i>	168
RNA interference and gain-of-function experiments.....	169
RNA methods and qPCR.....	170
Microscopy analysis.....	170
Figures and tables	172
Chapter 6 - Conclusions.....	187
References	190

List of Figures

Figure 1.1. The evolution of volvocine developmental complexity is marked by gains and losses of traits.....	15
Figure 1.2. Gene loss contributes to molecular network rewiring for volvocine multicellularity in concert with limited co-option and expansion of functional units.....	16
Figure 2.1. Volvocine algal genomes are similar.	46
Figure 2.2. Gene loss outpaces gain in volvocine algae genomes.....	47
Figure 2.3. Number of Pfams per gene does not change in the five volvocine algae species.....	49
Figure 2.4. Gene copy numbers of nine internal control genes in the six volvocine species.	50
Figure 2.5. A) Transcription factor and B) protein kinase family loss and gain in the six volvocine genomes.....	52
Figure 2.6. Gene loss in the volvocine algae occurs primarily by gradual decay.	53
Figure 2.7. GO terms for biological process and molecular functions of Chlamydomonas orthologs of retained (55.8% of total retained genes), decayed (24% of total decayed genes), and deleted (25.3% of total deleted genes) genes.	54
Figure 2.8. Number of genes per phylostratum (PS) for six volvocine species.	55
Figure 2.9. Number of Chlamydomonas orthologs of genes that underwent decay or deletion in multicellular volvocines per phylostratum (PS).	56
Figure 2.10. Phylogenetic analysis of cell cycle genes.....	58
Figure 2.11. Phylogenetic analysis of matrix metalloprotease (MMP) genes.	60
Figure 2.12. Phylogenetic analysis of pherophorin genes.....	62
Figure 2.13. Phylogenetic analysis of inv genes.	63
Figure 2.14. Phylogenetic analysis of glsA genes.	64
Figure 2.15. Percentage frequency distribution of disordered proteins in the volvocine algae. ...	65
Figure 2.16. Protein-protein interactions differ among volvocine species.	66
Figure 2.17. Raw images for A) Silver stained gels; B) α -tubulin and C) β -actin blots for Chlamydomonas, Gonium and Eudorina.	67
Figure 2.18. A) β -actin and B) α -tubulin phylogenetic trees using Coccomyxa subellipsoidea (Chlorophyta) as an outgroup.	68
Figure 2.19. Gene loss in a fully connected network (c=1) of genes (N=10) yields a higher proportion of novel stable network states than gene gain.	69
Figure 2.20. Wagner model simulates loss (orange) and duplication (blue) of k genes out of a network of N=10 genes over different network connectivities.....	70
Figure 2.21. Number of orthoMCL clusters for a range of inflation values.	71
Figure 2.22. Genome-wide phylogeny of red algae (magenta) compared to a chlorophycean outgroup (green) inferred by PosiGene.....	72
Figure 3.1. Volvocine genomes are similar.....	99
Figure 3.2. Gene loss outpaces gain in volvocine genomes.....	100
Figure 3.3. A) Transcription factor and B) protein kinase family loss and gain in seven volvocine genomes.....	103
Figure 3.4. Number of Pfams per gene does not change in seven volvocine algae species.	104
Figure 3.5. Gene loss in the volvocine algae occurs primarily by gradual decay.	105
Figure 3.6. Expansion of Cyclin D repertoire is conserved across multicellular volvocines.	107
Figure 4.1. rlsD overexpression changes transcriptional landscape for a wide array of genes...	134

Figure 4.2. rlsD targets genes encoding ribosomal and photosynthetic Pfams for downregulation, as well as carbon metabolism and ECM Pfam encoding genes for both up-and downregulation, among others.....	135
Figure 4.3. rlsD targets are biased towards downregulation of germ related genes and upregulation of soma related genes.....	136
Figure 4.4. Expression patterns of candidate targets of rlsD (set 1-12).....	138
Figure 4.5. Expression patterns of candidate targets of rlsD (set 13-24).....	140
Figure 4.6. Expression patterns of candidate targets of rlsD (set 25-27).....	141
Figure 4.7. rlsD and regA share multiple targets, but phenotypic output differs.....	142
Figure 5.1. Expression patterns of GpFsl1 and CrFSL1 over the life cycle of Chlamydomonas and Gonium.....	172
Figure 5.2. Amino acid conservation of orthologs of GpFsl1 in Volvox and Chlamydomonas.....	173
Figure 5.3. GpFsl1 knock-down Gonium are unicellular.....	175
Figure 5.4. Percentage unicellularity in WT Gonium (K3) and GpFsl1 knock-down Gonium lines.....	176
Figure 5.5. GpFsl1 knock-down lines have reduced amounts of transcript relative to WT Gonium (K3).....	177
Figure 5.6. GpFsl gain-of-function Chlamydomonas 21gr can be multicellular.....	178
Figure 5.7. Percentage multicellularity in WT Chlamydomonas (21gr) (orange) and GpFsl1 gain-of-function 21gr lines (blue).....	179
Figure 5.8. GpFsl gain-of-function Chlamydomonas 5325 and Crfsl1 strains fail cytokinesis... 180	
Figure 5.9. Percentage multicellularity in WT <i>Chlamydomonas</i> (5325) and <i>GpFsl1</i> gain-of-function 5325 and <i>Crfs11</i> lines (blue).....	181
Figure 5.10. Construction of A) GpFsl1-hi-anti B) GpFSL1-C-3xHA.....	182

List of Tables

Table 2.1. BUSCO completeness values of seven volvocine algae annotated genomes.....	73
Table 2.2. Genomic features of six volvocine algae species.....	74
Table 2.3. Shared orthologous group and pfam count for six volvocine algae species.....	76
Table 2.4. Conserved gene counts for six volvocine species.....	77
Table 2.5. Histone H1 gene counts for volvocine algae.....	78
Table 2.6. Histone family copy number and tail variant number of six volvocine algae.....	79
Table 2.7. Histone gene counts for Rhodophytes.....	80
Table 2.8. Distribution and type of gene losses in multicellular volvocine algae.....	81
Table 2.9. Matrix metalloprotease (MMP) and pherophorin gene counts for six volvocine species.....	83
Table 3.1. Counts for shared orthologous groups and Pfams across seven volvocine species. ..	111
Table 3.2. Conserved gene counts for seven volvocine species.....	112
Table 3.3. Histone H1 gene counts for seven volvocine species.....	113
Table 3.4. Histone family copy number and tail variant number for seven volvocine species...	114
Table 3.5. Distribution and type of gene losses in multicellular volvocine algae.....	116
Table 4.1. Pfam enrichment of differentially expressed genes across induction conditions pf <i>rlsD</i> overexpression.	157
Table 5.1. Fasciclin-encoding gene counts for five volvocine species.....	183
Table 5.2. Primers used for this work.....	185
Table 5.3. Settings for <i>Gonium</i> electroporation.	186

Acknowledgements

In theory, it takes a village to raise a child. If that is the case, it takes about two floors worth of Chalmers Hall and a copious amount of junk food to raise a PhD candidate to write their thesis. I wish to thank the people who have taken even small parts in this experience (if I do not explicitly name you, thank you!). Chris Toomajian, Thomas Mueller, Anna Zynovyeva, Kathrin Shrick, Michi Tobler, Kristin Michel, Ray ‘radiation man’ Boller, Gabe Figueroa, Tari Phillips, Sarah Hacker, Lydia Waner, Nirupama Kotian, Fabiha Rahman, MCDB attendees, Evolution journal club peeps, K-SACNAS peeps, Arrow peeps, Dee Yoesting, Val Olson, Jennifer Regier, Rita Goss, Tom Merchant, Allan Schrag, Leah, Ryan, Zoe, and Alex Oswald, for great conversation, insightful comments, dealing with my nonsense and overall kindness.

Next, I must thank my friends far and near. Far-friends, you have actively participated of helping me arrive at where I am today. Your patience, love, and trust are priceless to me, and I cannot wait for adventures to come. My MHK friends – Brian Jacobs, Jacob Parker, Ruben Lerma, Emily Burghardt, Daniel Quiring, Amanda Somers, Courtney Hill, Gabi Shipman, John Tumberger, David ‘David Thomas’ Hayes, Ashmita Pandey, Bianca Morejón, Luke Messer, and Patil Tavitian also have my undying love and gratitude, because it was you guys who made this adventure feel achievable.

I also wish to thank my colleagues from the *Volvax* community and every colleague not mentioned yet from the Olson Lab, Platt Lab, and Miller Lab for their pivotal role in helping me become a better, kinder, scientist, peer, and mentor. I am lucky to have met such talented people. I am especially grateful to my undergrads: without you, I would not have amassed the knowledge that is compiled here; I would not have known the feeling of sharing the joy of discovery the way

I know it now, and I would certainly have no consistent source of professional satisfaction. It has been a privilege to mentor you.

I wish to thank my committee members, Erika Geisbrecht, Tom Platt, and Jocelyn McDonald, for always sharing your knowledge with sincerity and kindness. I also thank Stephen Miller for being a wonderful collaborator and being among the most patient and nice scientists I have met. Stephanie Shames and Jocelyn McDonald took the time to help me navigate key aspects of being a woman in science with such grace that I can only hope to be half as inspiring to others as you were to me. I also want to thank Tom Platt for his perennial sunny disposition and sincere advice on career moves. I have attempted to express my gratitude and appreciation in order of difficulty, so this is the hardest part for sure. I am at a loss of words to describe how grateful I am to my advisor, Brad Olson. Brad, you provided me with *too many* opportunities. Your trust and mentorship through thick and thin (in these years you became a double-dad and COVID happened!) have quite literally changed my life. Thank you for being a fantastic mentor.

Finally, I want to thank Lynn and my family. It took me a couple of weeks to wake up in the US and not feel like ‘any day now I ought to fly back home’. It took me a couple of years to wake up in the US and feel like my life here made sense. Now I am crying as I write this because somehow I have a home in Manhattan, KS. This would not have happened without Lynn. *Nunca podré agradecerle lo suficiente a mi familia el apoyo, amor y confianza que han invertido en mi. Mispapás, mi hermana, mis padrinos y mi abuelito me han inspirado a intentar ser mejor... principalmente porque quiero ser la persona que ustedes ven en mi. Los quiero.*

If you have made it this far and have any intention of continuing your read, I will leave you with the undying words of Burger King man the Invincible (Michael Rayner): ‘*LOOK AT IT! LOOK AT IT AND ENJOY IT!!!!*’

Chapter 1 - The curious case of multicellularity in the volvocine algae

Berenice Jiménez-Marín and Bradley J.S.C. Olson

Front. Genet., 15 February 2022
Sec. Evolutionary and Genomic Microbiology
<https://doi.org/10.3389/fgene.2022.787665>

Author contributions: BJM: Conceptualization, investigation, writing—original draft, visualization. BJSCO: Conceptualization, supervision, writing—review and editing, project administration, funding acquisition.

Abstract

The evolution of multicellularity is a major evolutionary transition that underlies the radiation of many species in all domains of life, especially in eukaryotes. The volvocine green algae are an unconventional model system that holds great promise in the field given its genetic tractability, late transition to multicellularity, and phenotypic diversity. Multiple efforts at linking multicellularity-related developmental landmarks to key molecular changes, especially at the genome level, have provided key insights into the molecular innovations or lack thereof that underlie multicellularity. Twelve developmental changes have been proposed to explain the evolution of complex differentiated multicellularity in the volvocine algae. Co-option of key genes, such as cell cycle and developmental regulators has been observed, but with few exceptions, known co-option events do not seem to coincide with most developmental features observed in multicellular volvocines. The apparent lack of “master multicellularity genes” combined with no apparent correlation between gene gains for developmental processes suggest the possibility that many multicellular traits might be the product of gene-regulatory and functional innovations; in other words, multicellularity can arise from shared genomic repertoires that undergo regulatory and functional overhauls.

Introduction

The evolution of multicellular organisms is a Major Transition in Evolution where unicells relinquish their individuality to collectively coordinate for the development of a complex, higher organizational level individual (John Maynard Smith, 1995; Grosberg and Strathmann, 2007; Knoll, 2011). Multicellularity appears to be a successful solution for adapting to the environment; it evolved independently multiple times in all domains of life (Bonner, 2000; Grosberg and Strathmann, 2007; Ruiz-Trillo and Nedelcu, 2015) having generated extant

descendants that showcase a wealth of morphological and developmental innovation. Unicellular organisms have been shown to readily evolve multicellular lifestyles in response to selection (Grosberg and Strathmann, 2007) suggesting that the genetic requirements for multicellularity are not a major hurdle for the transition itself to occur (Rokas, 2008; Arenas-Mena, 2017).

Indeed, experimental evolution studies demonstrate the rapid ability of unicells such as yeast (Ratcliff et al., 2013), and green algae (Martin E. Boraas, 1998; Sathe and Durand, 2015; Fisher et al., 2016; Herron et al., 2019) to transition to multicellularity. Although the recurring nature of the evolution of multicellularity provides a useful base for comparative studies, determining the mechanisms underlying the evolution of multicellularity remains enigmatic due to technical obstacles such as long divergence times between multicellular models and their unicellular relatives and limited toolkits for experimental work in multicellularity models. Tractable models for understanding the evolution of multicellularity are thus indispensable to define the molecular players underlying multicellular evolution. A curious clade of chlorophytes, the volvocine algae, are an important system for addressing the molecular basis of multicellular evolution.

The volvocines are a well-established model system for the study of multicellularity (Nedelcu and Michod, 2003; Kirk, 2005; Miller and Technau, 2010; Niklas, 2014; Grochau-Wright et al., 2017) that encompasses ~50 extant multicellular species and their closest unicellular relative, *Chlamydomonas reinhardtii* (Kirk, 1997; Coleman, 1999; Umen and Olson, 2012) (Figure 1.1A). Multicellular volvocines are arranged into families Tetrabaenaceae (*Basichlamys*, *Tetrabaena*), Goniaceae (*Gonium*, *Astrephomene*) and Volvocaceae (*Pandorina*, *Eudorina*, *Yamagishiella*, *Volvox*) (Hanschen et al., 2018a; Lindsey et al., 2021), and display morphologies ranging from bowl-shaped undifferentiated colonies to differentiated spheroids

(Kirk, 1997; Coleman, 2012; Yamashita et al., 2016) (Figure 1.1A). Despite their phenotypic differences, this diverse group of algae evolved from their *Chlamydomonas*-like unicellular ancestor relatively recently during the Triassic period (around 250 MYA) (Herron et al., 2009). The late evolution of multicellularity of the volvocines has made them attractive subjects for comparative genomic analyses, which have shown that their genomes are highly similar and share conserved synteny (Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2017; Jiménez-Marín et al., 2021). The remarkable genetic similarity between volvocine species suggests the genetic signals relating to multicellularity might still be traceable within the lineage and makes this algal system suitable for the study of multicellularity at the molecular level. The genetic tractability of the volvocines, coupled to the increasing availability of functional genomic tools in key species (Lerche and Hallmann, 2009; Umen and Olson, 2012; Lerche and Hallmann, 2013), their phenotypic diversity and recent divergence from a *Chlamydomonas*-like ancestor, provide a unique framework for determining the common principles and functional players behind multicellularity, which impacts the broader fields of evolution and development.

Variations on a theme – evolution of development in the volvocine algae

The seeming stepwise acquisition of phenotypes that the volvocines display (from unicellular to undifferentiated and differentiated multicellular) have been used as the basis for a hypothesis of sequential acquisition of developmental complexity through genetic co-option (Kirk, 2005; Olson and Nedelcu, 2016). David Kirk's seminal work on this subject suggested that twelve developmental changes could explain the evolution of differentiated multicellularity in *Volvox carteri* (Kirk, 2005) (Figure 1.1A, highlighted in blue). At least four of Kirk's developmental changes are required for undifferentiated multicellularity as exhibited by *Tetrabaena* and *Basichlamys* (Figure 1.1). The best described undifferentiated multicellular species, *Gonium*

pectorale, covers six of Kirk's twelve steps (Umen and Olson, 2012) (Figure 1.1), suggesting that within the volvocines, the evolution of undifferentiated multicellularity requires the most developmental changes.

The lineage hypothesis of co-option proposes that volvocine multicellularity resulted from modification of cellular structures and developmental cycles under an otherwise conserved phenotypic background (Kirk, 1997; Bonner, 2000). This appears to be true for volvocine morphology; except for species with differentiated germ cells, the cells that compose volvocine colonies strongly resemble their closest unicellular relative, *Chlamydomonas*. *Chlamydomonas* cells are ovoid shaped and asymmetrical (Holmes and Dutcher, 1989) (Figure 1.1A). At their anterior end, each cell has two flagella that serve chemosensory and motility functions. At their posterior end, cells have a distinctive large cup-shaped chloroplast that occupies the bulk of the intracellular space (Umen and Olson, 2012). Multicellular volvocine cells are enveloped in a proteinaceous extracellular matrix (ECM) that resembles the ECM of animal cells rather than cellulosic cell walls of plants (Sumper and Hallmann, 1998). Interestingly, the multicellular volvocine ECM is biochemically and genetically conserved with *Chlamydomonas*' cell wall (Adair et al., 1987; Prochnik et al., 2010; Hanschen et al., 2016), suggesting the ancestral genes were repurposed for new multicellularity functions. Different species secrete various amounts of ECM, with the most extreme example being *Volvox*. Cells in *Volvox* are enveloped in ECM layers such that it comprises roughly 99% of the spheroid volume (Kirk, 1997).

Likewise, the multicellular volvocine life cycle is very similar to that of *Chlamydomonas*. All volvocines are haploid and capable of both vegetative and sexual reproduction (Hoops et al., 2006). Vegetative reproduction occurs in individual cells that undergo a modified cell cycle called "multiple fission" (Cross and Umen, 2015). Except for germ cells in *Volvox*, every

volvocine cell is flagellated and hence provides motility to the whole organism until cell division occurs, at which point the cells retract their flagella, undergo division, daughter cells regrow their flagella and begin swimming again. Given that the volvocine cells are not symmetrical, in multicellular species the organism as a whole must retain defined cell orientations in the context of their bowl (e.g. *Gonium*) or spheroid (e.g. *Pandorina*, *Eudorina*, *Pleodorina* & *Volvox*) morphologies and the number of cells in the colony for the organism to successfully swim (Hoops, 1993).

At the molecular level, one of key genetic innovations for volvocine multicellularity is co-option of the cell cycle regulator and tumor suppressor ortholog Retinoblastoma (RB). RB (encoded by the *MAT3* gene) from multicellular *Gonium* causes unicellular *Chlamydomonas* to be multicellular (Hanschen et al., 2016). As in plants and animals, all volvocines have an RB ortholog that regulates their cell cycles through cyclin-dependent kinase (CDK) phosphorylation (Ferris et al., 2010; Hanschen et al., 2016). RB from multicellular volvocine species differ in their regulatory domains from their ortholog in unicellular *Chlamydomonas*. First, RB from multicellular volvocines have shorter linkers between the RB-A and RB-B domains compared to *Chlamydomonas* (Ferris et al., 2010; Hiraide et al., 2013; Hanschen et al., 2016; Featherston et al., 2017). Second, RB from multicellular volvocines lack conservation in C-terminal domain and CDK phosphorylation sites that are present in their unicellular relative (Hiraide et al., 2013; Hanschen et al., 2016; Featherston et al., 2017). Additionally, *Volvox carteri* has different sex-based isoforms of RB that correlate with oogamy (Ferris et al., 2010). While cases of gene family expansion in the volvocines are limited, several multicellular volvocines, including *Gonium*, show expansion of cyclin D1 gene family (Prochnik et al., 2010; Hanschen et al., 2016). Cyclin D genes regulate the cell cycle by dimerizing with CDKs and phosphorylating RB

proteins (Cross and Umen, 2015; Hanschen et al., 2016; Featherston et al., 2017; Jiménez-Marín et al., 2021); whether cyclins are involved in the evolution of volvocine multicellularity is yet to be determined. Regardless of the impact of cyclins, it would seem likely that cell cycle dependent regulation of gene expression by RB is different for *Chlamydomonas* compared to its multicellular volvocine relatives, possibly driving the evolution of multicellular development in the latter.

Aside from RB and cyclin D1s, other known co-option events do not seem to coincide with Kirk's six steps to undifferentiated multicellularity (Figure 1.1B, steps 1-6). For instance, 'ECM-related' genes, which provide structure and cohesion to volvocine colonies, were likely co-opted for multicellularity. However, *Gonium* and *Chlamydomonas* share similar numbers of key ECM-related genes despite the former being multicellular and the latter unicellular (Hanschen et al., 2016; Olson and Nedelcu, 2016; Jiménez-Marín et al., 2021). Instead, expansion and diversification of ECM-related genes (pherophorins, matrix metalloproteinases- MMPs) seems to relate more to changes in organismal size than to the evolution of colonial life (Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2017; Jiménez-Marín et al., 2021). Similarly, genes that became co-opted for key volvocine developmental processes originated much before the appearance of the phenotypes with which they are associated (Figure 1.1).

Differentiated multicellularity in the volvocines has been long thought to be a consequence of developmental innovations that originate from further modification of structures and co-option of genes present in undifferentiated multicellular ancestors, and likely in extant undifferentiated sister species. Spheroidal and bowl shaped volvocine embryos need to undergo a process akin to gastrulation in which the colony shape must change to ensure a final configuration that allows the organism to swim (Hallmann, 2006a). This process, called

inversion, is known to occur in the Goniaceae (all but *Astrephomene* (Yamashita et al., 2016)), and Volvocaceae (Hallmann, 2006a). The product of *invA*, a kinesin, is known to play a key role in inversion in *Volvox*. Orthologs of *invA* are present in several volvocine species, regardless of their morphology, and noteworthily, *Chlamydomonas* also has an ortholog for this gene, *iar1*. Surprisingly, the *Chlamydomonas iar1* can rescue *Volvox* inversion-defective *invA* mutants (Nishii et al., 2003). However, whether *Gonium*'s ortholog of *invA* has a role in inversion remains to be demonstrated. Likewise, co-chaperone *glsA*, which is involved in asymmetric cell division in *Volvox* – a process tied to cell differentiation in this and other volvocines, has orthologs in species that do not undergo cell differentiation (Jiménez-Marín et al., 2021). The *Chlamydomonas* ortholog for *glsA*, much like *iar1*, can rescue gonidialess *glsA* *Volvox* mutants (Cheng et al., 2003).

Perhaps the best-known example of co-option for a key volvocine developmental processes preceding the phenotype itself is *Volvox*'s cell differentiation master regulator *regA*. Whilst there are *regA* homologs in *Chlamydomonas* and *Gonium* (*RLS1*), there are no orthologs of it in either species (Hanschen et al., 2014; Hanschen et al., 2016). In the species that have it, *regA* is part of a gene cluster that likely evolved through tandem duplication shortly after Goniaceae split from Volvocaceae (Hanschen et al., 2014; Grochau-Wright et al., 2017). Indeed, the *Pandorina* (Volvocaceae) genome encodes for the full *reg* cluster, despite *Pandorina* being an undifferentiated species (Grochau-Wright et al., 2017). Ancestral state reconstruction of the *reg* cluster supports the hypothesis that the genetic basis for somatic cell differentiation in the Volvocaceae preceded cellular differentiation itself and was subject to co-option (Grochau-Wright et al., 2017). A recent transcriptome study of *Volvox* gonidia and somatic cells demonstrated the different cell types have markedly distinct transcriptional programs, and

suggest that co-option of certain gene expression programs for differentiation might have driven cell type specialization (Matt and Umen, 2018). A major open question in the field is whether *regA* acts as a key regulator of these transcriptional programs during development to establish the germ and somatic cell lineages (Olson and Nedelcu, 2016).

Traits related to sexual reproduction are oftentimes highly dynamic evolutionarily. In the case of the volvocines, there are members whose reproductive morphology range from isogamous (*Chlamydomonas*, *Gonium*, *Pandorina*) to anisogamous (*Eudorina*) and oogamous (*Volvox*). However, the evolution of anisogamy does not appear to be correlated to the evolution of cellular differentiation, though it does correlate with increased complexity (Hanschen et al., 2018a). The mating-type (MT) loci of different volvocines experience frequent turnover, and their distribution (as well as that of MT linked genes) does not seem to relate to the evolution of anisogamy within the clade. Thus, it is hypothesized that the evolution of volvocine males might be the consequence of altering the function of *MID*, a sex determining putative transcription factor, or of its target genes, as opposed to the acquisition of novel genes controlling gamete size (Hamaji et al., 2018). Empirical, phylogenetics-based evidence supports predictions that multicellularity is a driver of derived sexual traits, starting with anisogamy and then with sexual dimorphism (Hanschen et al., 2018a).

While the co-option of *RB*, *glsA*, *regA* and other examples here given demonstrates that shared genetic elements can be repurposed for a wide array of biological roles, surprisingly few ‘multicellularity genes’ are found in the volvocines (Figure 1.1B). Unlike the initial shift to undifferentiated multicellularity that was likely boosted by co-option of *RB*, it would seem like subsequent genetic changes related to ECM maintenance, inversion, and cell differentiation have roles in stabilizing colonial life and controlling subsequent developmental processes rather than

impacting multicellularity itself (Figure 1.1B). Indeed, Kirk's streamlined framework has undergone revisions in light of the discovery of a dynamic history of gains and losses of morphological and developmental traits (Nanjundiah et al., 2018; Lindsey et al., 2021) (Figure 1.1A). Hence, the volvocine phylogenetic tree is under constant revision (Nozaki et al., 2003; Hanschen et al., 2018b; Lindsey et al., 2021). Genera *Eudorina*, *Pleodorina* and *Volvox* are not monophyletic (Figure 1.1A), and family Goniaceae likely isn't either (Lindsey et al., 2021). Ancestral state reconstruction shows that there have been repeated instances of evolution of cell differentiation and loss of cell differentiation within the volvocines (Grochau-Wright et al., 2017; Lindsey et al., 2021) (Figure 1.1A). Other landmark traits, including the evolution of spheroidal morphologies, inversion, and expansion/contraction of the ECM, seem to have occurred independently more than once as well (Herron et al., 2009; Ruiz-Trillo and Nedelcu, 2015) (Figure 1.1A). The most extreme example of independent evolution is *Astrephomene*, a proposed (but debated) sister genus of *Gonium*, which evolved a spheroidal shape, has sterile somatic cells without the *reg* cluster, and rearranges itself without undergoing inversion (Yamashita et al., 2016). Upon inspecting gains and losses in morphology, it would seem that the transition to undifferentiated multicellularity is not as readily gained as lost, again supporting the idea that this step is more developmentally complex to evolve. The frequent shifts between simplification and complexification of traits along multicellular volvocine history paint a picture that does not support widespread genetic innovation as a driver of developmental complexity.

How can developmental complexity arise from a widely shared genomic repertoire?

One of the most striking features of multicellular evolution in the volvocine algae is how few new genes seem to be required to drive complex developmental changes. Moreover, many such

genes are largely conserved in 1:1 orthology across the entire lineage (e.g., *invA*, *glsA*, RB/MAT3) and relatively few examples of gene expansion (e.g., cyclin Ds, *regA*, MMPs). Surprisingly, comparative analyses of multicellular volvocine genomes to unicellular *Chlamydomonas* (Jiménez-Marín et al., 2021) combined with experimental approaches support the impression that few key genes were co-opted for volvocine developmental traits, and that other processes such as *de novo* gene evolution are much less important (Hanschen et al., 2016; Olson and Nedelcu, 2016; Jiménez-Marín et al., 2021) to multicellularity in this system. Thus, alternative mechanisms must be behind the dynamic developmental and evolutionary picture the volvocine algae have painted.

Inferences on the minimal role of genetic innovation made through phylogenetic approaches have been confirmed through comparative genomics. Orthologous groups, protein family (Pfam) domains, regulatory protein families, and even histones except for H1 are less numerous in *Volvox* than they are in less developmentally complex volvocines. Contraction of shared groups among the genomes of *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* relative to *Chlamydomonas* is more frequent than is expansion of other shared elements (Jiménez-Marín et al., 2021). Interestingly, the highly conserved synteny between volvocine genomes facilitates mechanistic analysis of gene loss (that is, analysis of what genes have been lost and the manner of their loss). This analysis shows that losses mainly occur through progressive decay (Jiménez-Marín et al., 2021). What is more, a significant proportion of analyzed gene losses are common to *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* but not to *Chlamydomonas*, strongly suggesting that bursts of gene loss at the last common ancestor (LCA) of the multicellular volvocines might have played a significant role in the transition to multicellularity of this lineage (Jiménez-Marín et al., 2021).

The high sequence similarity shared among volvocine genomes has allowed for detailed advances in our understanding of how small changes have ‘primed’ this algal lineage for multicellularity. However, the small-scale innovations observable in the genome cannot account for the variety of phenotypes that the clade itself possesses. For instance, there are similar abundances of transcription factors (TFs) between *Chlamydomonas* and *Gonium* (Hanschen et al., 2016), though there is evidence that certain TF families as well as other regulatory elements are contracting as developmental complexity increases in the multicellular volvocines (Jiménez-Marín et al., 2021). Making sense of these seemingly contradictory data is as challenging as it is promising, and some interesting insight is already becoming available.

The study of gene loss as a driver of evolutionary innovation is relatively recent (Go et al., 2005; Kvitek and Sherlock, 2013; Albalat and Canestro, 2016; Fernandez and Gabaldon, 2020; Guijarro-Clarke et al., 2020), but mathematical modeling suggests that gene loss might be related to changes to the molecular network of the impacted species in such a way that novel interactions might arise (Jiménez-Marín et al., 2021). This could translate to the ancestor of the colonial volvocines evolving multicellularity whilst setting up the resulting lineages for a wide array of developmental outcomes that might explain the discordance between the emergence of gene families, their co-option, and the emergence of more complex phenotypes among species.

Despite a general trend of reduction of histone copy numbers, *Gonium* histones H2B have a greater N-terminal variant diversity than other volvocines do (Jiménez-Marín et al., 2021). This might mean that differential combinations in histone tails between species could fuel differential post-translational change programs between species that could facilitate increases in developmental complexity even in the context of gene loss. Furthermore, protein-protein interaction (PPI) analysis of *Chlamydomonas*, *Gonium* and *Eudorina* strongly suggests that the

proteomic makeup of each species is vastly different even for conserved gene products (Jiménez-Marín et al., 2021). These novel findings hint at the possibility that post-genomic programs have a strong role in generating and stabilizing more developmentally complex morphologies (Figure 1.2). Perhaps that is why it has proven so difficult to find ‘multicellularity genes’: differential usage of shared functional repertoires is a likely path to varying phenotypes that cannot be readily identified through comparative genomics alone.

Combinatorial co-option and creative destruction

The elusive nature of the molecular players behind evolution of multicellularity and other Major Transitions in biological complexity might be due in part because they do not necessarily have specific genetic determinants. Rather, a few key master regulators of gene expression (e.g., *RB* and *regA*) undergo extensive co-option that results in a cascade of genome-wide developmental expression rewiring. In addition to this, it is feasible that epigenetic changes add to differential expression patterns between species; differences in histone diversity and abundance between species (Jiménez-Marín et al., 2021) provide some initial support to this possibility. Among the effects of global network rewiring is the formation of novel interactions at the protein level, which yields a combinatorial co-option of function that alters the organism’s functional repertoire without significantly shifting its genetic repertoire. In other words, rather than duplication and divergence to co-opt the function of genes, their combinations during development are altered for biological novelty. Functions that lose their adaptive value or become dispensable are gradually lost, as they are no longer needed.

The consequence of this framework is that, outside of a few key regulators, there may not be such a thing as “multicellularity genes”. Protein-protein interaction network rewiring should be capable of producing divergent phenotypes without the intervention of extensive genetic

novelty. While on its face this mode of evolutionary novelty does not fit preconceptions, it is biologically sound. Unicellular ancestors were jack-of-all trades and had to perform many different functions. As they evolved into undifferentiated and differentiated multicellular organisms, these functions were either used as a “parts bin” for novel functions or their functions were no longer needed and lost. While increased sampling of volvocine genomes might increase the resolution on the history of key multicellularity genes, this framework hints at the need for the exploitation of other ‘omic’ approaches (transcriptomics, proteomics) in conjunction with the usage and development of molecular biology tools (transformation, RNAi, CRISPR) that allow for comparisons between species and between WT and experimental strains under diverse conditions.

Upon transitioning to multicellularity, colonial organisms can switch from strictly temporal gene regulation to spatial or spatial temporal regulation patterns (Mikhailov et al., 2009). In the context that multicellularity has occurred numerous times under vastly different genetic backgrounds, it seems feasible that the many efforts to review the genetic underpinnings of this transition have come up short because multicellular evolution is not a process that requires a handful of genes to orchestrate organismal integration; rather, it requires a complete overhaul of how an organism uses whatever genes it has. Our future understanding of multicellularity will likely require a similar scientific overhaul, where the integration of information at different levels (DNA, RNA, phenotypic data) should lead to a more complete picture of the greater and lesser hurdles underlying this fascinating Major Transition.

Figures and tables

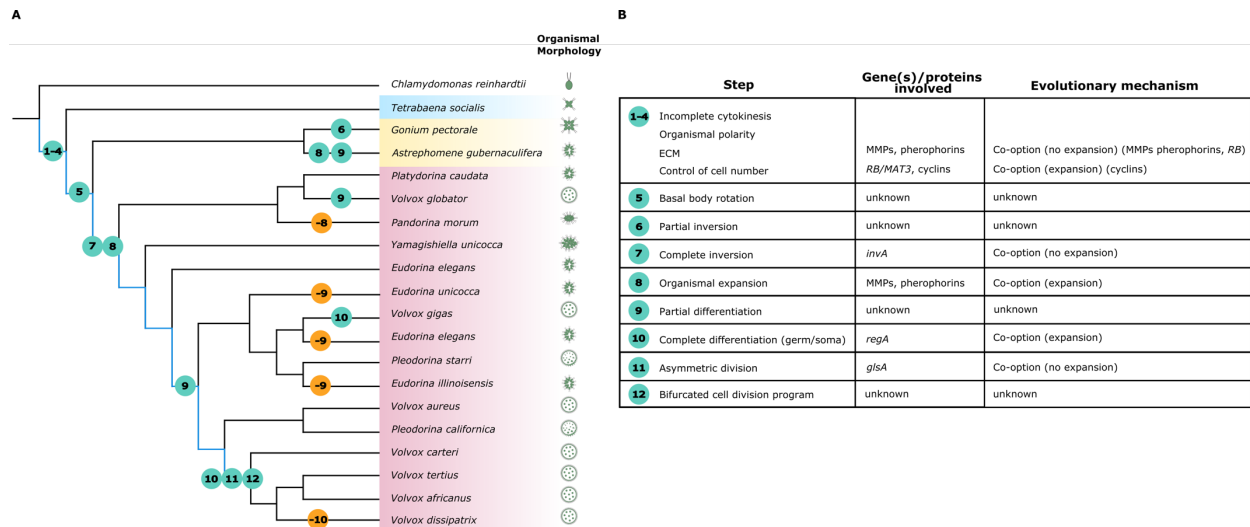


Figure 1.1. The evolution of volvocine developmental complexity is marked by gains and losses of traits.

Adapted from (Ruiz-Trillo and Nedelcu, 2015; Lindsey et al., 2021). A) Phylogeny of the volvocine algae. Multicellular volvocine families Tetrabaenaceae, Goniaceae, and Volvocaceae are demarcated in blue, yellow, and pink respectively. Kirk's lineage explanation for volvocine multicellularity is highlighted in blue. Traits associated to Kirk's twelve steps are marked by circles; numbers within circles detail what step is described. Teal circles represent evolution of traits and orange circles represent loss of traits. Organismal morphology for depicted species is not to scale. B) Kirk's twelve steps do not always have a known genetic origin, and when they do, it is not always co-opted via duplication and divergence.

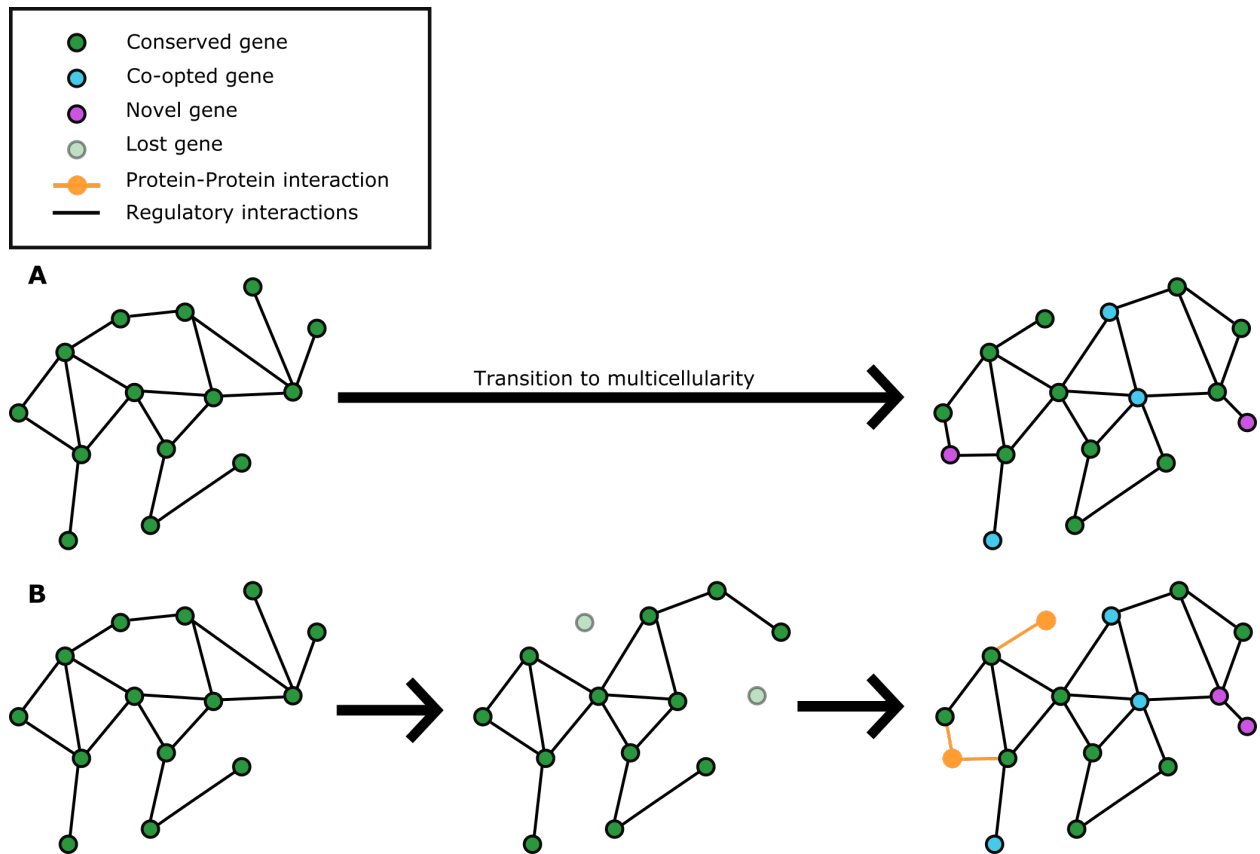


Figure 1.2. Gene loss contributes to molecular network rewiring for volvocine multicellularity in concert with limited co-option and expansion of functional units.

A) current understanding of how molecular networks in a unicellular, *Chlamydomonas*-like ancestor of the volvocines evolved upon the transition to multicellularity (Adami et al., 2000; Trigos et al., 2018). In this model, network elements are co-opted for multicellular function, and expanded to varying degrees for different multicellular lineages (magenta nodes) with minimal changes to the core ‘ancestral’ (unicellular) network. B) changes in molecular networks from a unicellular, *Chlamydomonas*-like ancestor of the volvocines by gene loss and limited co-option to evolve new functions for multicellularity. Some network elements were lost (light blue disconnected nodes), which caused a reconfiguration of the network to promote novel

interactions (dotted magenta edges), co-option events (cyan nodes), and expansion of units with novel multicellularity related functions (magenta nodes).

Chapter 2 - Gene loss at a transition to multicellularity

Berenice Jiménez-Marín, Jessica B. Rakijas, Antariksh Tyagi, Aakash Pandey, Erik R. Hanschen, Jaden Anderson, Matthew G. Heffel, Thomas G. Platt, Bradley J. S. C. Olson

Under revision at *Scientific Reports*
Available as preprint on bioRxiv (DOI:10.1101/2021.02.16.431445)

Author contributions: BJM: Conceptualization, software, formal analysis, investigation, writing – original draft, visualization. JBR: conceptualization, methodology, validation, investigation, visualization. AT: validation, formal analysis, investigation, data curation, visualization. AP: conceptualization, methodology, software, formal analysis, visualization. ERH: conceptualization, investigation, writing – review & editing. JA: conceptualization, investigation. MGH: validation, formal analysis. TGP: conceptualization, methodology, software, formal analysis supervision, writing – review and editing, visualization. BJSCO: conceptualization, methodology, software, formal analysis, writing – review & editing, supervision, project administration, funding acquisition.

Abstract

Multicellular evolution is a major transition associated with momentous diversification of multiple lineages and increased developmental complexity. The volvocine algae comprise a valuable system for the study of this transition, as they span from unicellular to undifferentiated and differentiated multicellular morphologies despite their genomes being similar, suggesting multicellular evolution requires few genetic changes to undergo dramatic shifts in developmental complexity. Here, the evolutionary dynamics of six volvocine genomes were examined, where a gradual loss of genes was observed in parallel to the co-option of a few key genes. Protein complexes in the six species exhibited novel interactions, suggesting that gene loss could play a role in evolutionary novelty. This finding was supported by gene network modeling, where gene loss outpaces gene gain in generating novel stable network states. These results suggest gene loss, in addition to gene gain and co-option, may be important for the evolution of developmental complexity.

Introduction

The evolution of multicellular organisms is a prerequisite for the emergence of complex organismal body plans (Grosberg and Strathmann, 2007; Knoll, 2011). However, the genomic basis for multicellular evolution is poorly understood (Olson and Nedelcu, 2016). The volvocine algae are a valuable model system for studying multicellular evolution because they have undergone a recent transition to multicellularity (~200 MYA) (Herron et al., 2009), and the ~100 member species span a wide range of developmental complexity within a size range from 10 μm to 3 mm (Hanschen et al., 2018a). This group includes unicellular *Chlamydomonas*, undifferentiated multicellular bowl-shaped *Gonium* and spheroidal *Pandorina* (8-16 cells),

multicellular isogamous *Yamagishiella* (32 cells), multicellular anisogamous *Eudorina* (32 cells), and multicellular differentiated *Volvox* (>500 cells) (Kirk, 2005; Coleman, 2012).

Based on morphology, differentiated multicellular volvocines were predicted to have evolved from their unicellular ancestors by stepwise acquisition of developmental processes, such as establishment of organismic polarity, genetic control of cell number, size expansion, and division of labor among cell lineages (Kirk, 2005) (Figure 2.1A, cartoons). However, phylogenetic approaches suggest that diversification among the volvocine algae was rapid (Herron et al., 2009) and many of the genetic changes allowing for the evolution of developmental complexity took place early in the evolution of the volvocines. This suggests that genetic gains for each step might not fully explain the organismal complexity of the volvocine algae (Hanschen et al., 2016).

It has been proposed that shared genetic toolkits and pathways that act as developmental patterning modules can lead to divergent phenotypic outputs in other multicellular lineages (Salazar-Ciudad, 2006; Newman and Bhat, 2009). Indeed, the genomes of *Chlamydomonas reinhardtii*, *Gonium pectorale*, and *Volvox carteri* (hereafter referred to by genus) are highly similar and syntenic (Merchant et al., 2007; Blaby et al., 2014; Hanschen et al., 2016). Moreover, a homolog of the cell cycle regulator and tumor suppressor *Retinoblastoma* (RB) is sufficient for undifferentiated multicellularity (Hanschen et al., 2016), suggesting co-option of a developmental regulator can result in complex developmental patterns required for multicellularity (Olson and Nedelcu, 2016). Thus, the evolution and diversification of the volvocine algae is thought to have heavily relied on the co-option of key functional elements.

^aHenceforth called *Chlamydomonas*, *Gonium*, *Pandorina*, *Eudorina* and *Volvox*

Co-option, however, is unlikely to be the only evolutionary process behind volvocine multicellularity. Gene loss is found in many lineages undergoing evolutionary change (Go et al., 2005; Kvitek and Sherlock, 2013; Albalat and Canestro, 2016; Fernandez and Gabaldon, 2020; Guijarro-Clarke et al., 2020), and it is likely a major force in eukaryotic evolution (Krylov et al., 2003) with the potential to influence evolutionary trajectories (Sharma et al., 2018), to the extent that gene loss has been proposed to be a dominant force in genome evolution overall (Wolf and Koonin, 2013). Nevertheless, its action in concert with other evolutionary drivers is not well understood (Bolotin and Hershberg, 2015; Bhattacharya et al., 2018; Bowles et al., 2019; Lopez-Escardo et al., 2019). The study of gene loss is challenging because loss events can be the result of trait evolution (e.g. through relaxed selection), but they can also cause phenotypic adaptation (Sharma et al., 2018). Moreover, gene losses can impact an organism's evolvability and speed of adaptation (Helsen et al., 2020), depending on their dispensability and position within the regulatory and protein-protein interaction network (Krylov et al., 2003; Helsen et al., 2020).

To investigate whether gene loss played a role in volvocine evolution, we report the assembly of *Pandorina morum* and annotation of genomes from *P. morum*, *Yamagishiella unicocca* and *Eudorina sp.*^a These newly analyzed genomes, in addition to those of *Chlamydomonas*, *Gonium*, and *Volvox* (Merchant et al., 2007; Prochnik et al., 2010; Hanschen et al., 2016), were subjected to comprehensive analysis of genic evolutionary trajectory. We demonstrate that significant loss of conserved genes has occurred in the volvocine lineage, and through a gene network model, we show that gene loss outperforms gene gain in generating stable novelty. Our results additionally suggest that only a few genes were co-opted for volvocine multicellularity. Furthermore, differential protein-protein interactions (PPI) are observed between species. These results suggest that gene loss at the transition to multicellularity

may impact molecular network structure such that novel functions associated to changes in developmental complexity can evolve.

Results

Volvocine genomes share genomic and functional toolkits

We validated the annotation of seven closely related volvocine algae exhibiting major changes in morphology (Figure 2.1A cartoons) using BUSCO (Kirk, 2005) (Seppey M., 2019). BUSCO inferences suggest that the *Chlamydomonas*, *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina*, and *Volvox* annotated genomes are of high quality (Table 2.1). However, BUSCO analysis of the *Tetrabaena* genome showed it had low completeness values (63.6%). Hence, it was not used for downstream analyses despite being an available volvocine genome. Genome comparisons (Table 2.2), orthologous group and protein family domain (Pfam) catalogs for the analyzed species and their chlorophyte outgroups were compiled, and a genome-wide reconstruction of volvocine phylogeny was generated (Figure 2.2A) based on orthologous groups, which is in close agreement with previously determined phylogenies based on smaller gene sets (Nozaki et al., 2003; Herron and Michod, 2008; Hanschen et al., 2016).

Overall, 15,155 orthologous groups were identified; of these, 6,297 (41.6%) are shared among the six genomes, and 2,417 are species specific (15.9%, Figure 2.1B, Table 2.3). Analysis of identified Pfams (El-Gebali et al., 2019) in the six species shows 1,743 Pfam domains. Of these, 1,184 (67.9%) are shared in all six species, and 96 (5.5%) are species specific. Notably, 8,922 (58.9%) orthologous groups, and 1,479 (84.9%) Pfams are shared by at least 5 species (Figure 2.1C). These results support previous findings of minimal species-specific genomic and functional toolkits between the volvocine genomes (Prochnik et al., 2010; Hanschen et al., 2016), despite their marked developmental and morphological differences.

Contraction outpaces expansion of conserved genetic and functional elements as complexity increases

Given the similarity between volvocine algae genomes (Figure 2.1B, C), we hypothesized that small differences between the *Chlamydomonas*, *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* genomes contribute to the evolution of developmental complexity. Rates of evolution of shared genes and protein domains (7,942 orthologous groups and 1,338 Pfam domains) for the six species were analyzed by normalization to the distance from their Last Common Ancestor (LCA) to account for potential signatures of gene or domain gain and loss. This analysis revealed that 402 (5.1%) orthologous groups have undergone significant changes, with a bias towards gene loss (176 expanding orthologous groups, 226 contracting, Figure 2.2A outliers, shaded grey area). An independent approach to evaluate contraction of orthologous groups using GLOOME (Cohen et al., 2010) confirmed that 141 of the 226 contracting orthologous groups (62.4%) underwent losses at different points of volvocine evolution. Of the remaining groups, 51 of the 226 (22.6%) cannot be subjected to GLOOME analysis (see Methods), which means that only 34 orthologous groups (15%) show discrepancies between both methods.

Surprisingly, while most Pfam domains were conserved, significantly expanding domains were less abundant than their contracting counterparts (51 and 127, respectively, Figure 2.2B outliers, shaded grey area). Pfam domain counts per gene did not show significant variation between the six species (Figure 2.3), and contraction trends were not observed in highly conserved genes actin, β Tubulin, α -Tubulin, chloroplast ferredoxin, mtATP-synthase α , mtATP-synthase β , and dynein heavy chain (Figure 2.4, Table 2.4).

Co-option of master regulators such as transcription factors (TFs) or protein kinases (PKs) is thought to be important for evolutionary innovation (Carroll, 2001; Hanschen et al.,

2016). Analysis of all identified TF and PK families in the volvocine algae (47 and 69 respectively) revealed that 5 families of TFs (10.6%) underwent significant contraction, while none expanded significantly (Figure 2.2C, Figure 2.5A). Similarly, 10 families of PKs (14.5%) underwent significant change, where 1 of the PK families expanded and 9 contracted (Figure 2.2D, Figure 2.5B). These data suggest that volvocine algae have lost genes with conserved functions, while gene gain occurred at a much lower rate and frequency (Figure 2.2A-F, Figure 2.5).

To gain further insight into gene loss as a mechanism of multicellular evolution, we analyzed histone genes for their evolutionary dynamics. Chromatin is an important regulator of gene expression, where post-translational modifications of histone N-termini are conserved and well understood (Waterborg et al., 1995). Thus, histone genes were analyzed for their evolutionary dynamics. All families of histones, except linker H1, were reduced in number as developmental complexity increases in the volvocine algae (Figure 2.2E, Tables 2.5 – 2.6). Analysis of histone N-termini methylation site variation for H2A, H2B, H3, and H4 shows that H2A, H2B and H4 underwent copy number reduction (Figure 2.2F, Table 2.6). However, Histone H3 exhibited greater N-terminal variant diversity in *Gonium* (Figure 2.2F, Table 2.6). These data suggest that post-translational changes in histones may facilitate increases in developmental complexity through altering gene expression in the context of gene loss.

Red algae (Rhodophyceae) also display various degrees of developmental complexity and exhibit signatures of gene loss (Bhattacharya et al., 2018). Hence, orthologous groups (Figure 2.2A), Pfam domains (Figure 2.2B), TFs (Figure 2.2C), PKs (Figure 2.2D) and histone (Table S2.7) dynamics were examined in parallel to their volvocine counterparts. This analysis confirmed the presence of contraction signatures in red algae, agreeing with earlier reports

(Bhattacharya et al., 2018), as well as the presence of expansion signatures (Figure 2.2D). These results validate the usefulness of a regression-based approach to identify candidate genetic and functional units with a history of expansion or contraction.

To assess the impact of positive selection on increases to developmental complexity, orthologous groups of genes with representatives in all six species were subjected to dN/dS analysis. Previous reports indicate that few genes in the *Gonium* and *Volvox* genomes show a signature of undergoing positive selection, as measured by the ratio of non-synonymous to synonymous mutations (dN/dS) (Hanschen et al., 2016). Using *Chlamydomonas* as the reference for ratio calculation 77, 51, 4, 19 and 3 genes exhibited positive selection in *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina*, and *Volvox*, respectively (Figure 2.2G). Using *Volvox* as the reference, there were 372, 82, 53, 4 and 17 genes with a significant dN/dS value in *Chlamydomonas*, *Gonium*, *Pandorina*, *Yamagishiella*, and *Eudorina*, respectively (Figure 2.2G). Among multicellular volvocines, *Gonium* has a larger number of genes under positive selection, with the multicellular species examined having comparable albeit lower numbers of genes under positive selection. These results suggest that positive selection is not the only force involved in volvocine evolution; rather, it seems like a combination of processes (co-option, gain, and loss of genes) might have set the stage for instances of positive selection, especially upon the transition to undifferentiated multicellularity.

Gene loss occurs mainly by progressive decay

As means to distinguish between evolutionary gene loss events (and the mechanism underlying them) and false positive losses from technical limitations, we examined conservation of genome structure at the sequence level for all volvocine genomes. This approach is difficult to do in other phyla, where gene loss is inferred indirectly through phylogenetic analysis (Fernandez and

Gabaldon, 2020; Guijarro-Clarke et al., 2020). The genomes of the volvocines, however, are highly syntenic (Prochnik et al., 2010; Hanschen et al., 2016). The remarkable conservation of syntenic regions across volvocine genomes allows for detailed examination of gene loss events, as syntenic blocks of conserved genes may have remnants of genes lost. Cross-syntenic block comparisons between species were made against the best annotated genome for the least developmentally complex organism available (*Chlamydomonas*). *Chlamydomonas* genes from contracting orthologous groups that are not assembled into chromosomes were filtered out. If syntenic regions could not be found for the selected *Chlamydomonas* genes and their orthologs in the other five species, those genes would too be eliminated from further analysis. Thus, only a sub-set of conserved loci where gene order was preserved in well-assembled, well understood and syntenic sequence blocks that belong to contracting orthologous groups was examined.

Examination of the mechanisms underlying gene loss was performed on genes belonging to the 226 contracting orthologous groups at the sequence level (Figure 2.2A). The contracting groups contain 1,230 genes assembled into *Chlamydomonas* chromosomes, of which 1,184 (91.2%) in 199 (88.1%) orthologous groups had loci that were adequately assembled and syntenous in all genomes for further analysis. This analysis demonstrated 43 of 1,184 genes in contracting orthologous groups are retained - that is, that despite belonging to contracting orthologous groups, these genes have an ortholog in all six volvocine species, while 1,141 loci exhibit gene loss (Figure 2.6A, B). The mechanisms of gene loss were then examined by tBLASTx to determine if remnants of the lost genes (gene decay) were present at the locus or if remnants of the gene were completely lost (gene deletion). Our results show gene decay is more abundant than deletion in all the colonial species; 72.7% of *Gonium*, 65.8% of *Pandorina*, 72.9% of *Yamagishiella*, 73.2% of *Eudorina* and 51.4% of *Volvox* gene losses show decay signatures

(Figure 2.6A). Observed signatures of decay confirm that our observations of gene loss are not likely to be caused by technical issues with genome comparisons, but rather are, for the most part, evolutionary losses of gene function. Thus, the volvocines exhibit a burst of gene loss primarily by decay during the transition to undifferentiated multicellularity that appears to have continued as developmental complexity increased.

Cross-species comparisons of retention, decay, and deletion events show that most candidate gene losses, 975 loci (82.3%), are due to decay (Figure 2.6B). Surprisingly, our analysis also suggests 1,017 (85.9%) loci are gene loss candidates in all five multicellular species. Of these, 880 (74.3%) occurred by gradual decay (Figure 2.6C, Table 2.8). The presence of decay signatures eliminates the possibility that these loci represent gene gains of *Chlamydomonas*. In the case of deletions, only 166 loci (14.0%) were completely absent in all multicellular species, which could either represent complete gene loss in the multicellular lineage or *Chlamydomonas*-specific gene gains. GO-term analysis of *Chlamydomonas* orthologs of retained, decayed, and deleted genes was performed. Since the GO annotation of volvocine genomes is far from complete, only 24 of 43 retained gene orthologs, 234 of 975 decayed gene orthologs, and 42 of 166 lost gene orthologs could be analyzed. Genes associated to conserved biological processes (e.g. metabolic processes, ion transport) and molecular functions (e.g. protein kinase activity, nucleic acid binding) have undergone loss by decay, deletion, or both (Figure 2.7). The predominance and progressive distribution of decay signatures suggest that a major gene loss event(s) occurred after the LCA of colonial volvocine algae diverged from the lineage of rather than that the *Chlamydomonas* lineage experienced extensive gene gains after diverging from multicellular volvocines.

Neither co-option nor gene age account for the volvocine developmental complexity

Previous analysis of the volvocine genomes indicated that co-option of genes has occurred in *Gonium* and *Volvox* (Hanschen et al., 2016). In this expanded phylostratigraphic analysis we confirmed the high degree of conservation of ancestral genes between volvocine genomes, as 51.1 to 69.8% of the genes per species fell under phylostratum 1 (PS-1, cellular organisms) and PS-2 (Eukaryotes) (Figure 2.8). Multicellular volvocines are traditionally divided into three families (Tetrabaenaceae, Goniaceae, and Volvocaceae); this means that genetic innovation relating to increased developmental complexity should be represented in family and genus strata. Contrary to that expectation, 0 to 4.3% of the genes per species were found in PS-7 (family), while 0 to 9.1% of the genes per species were found in PS-8 (genus). These results support co-option as a key mechanism in volvocine evolution, and do not coincide with a scenario of increases in developmental complexity resulting and being proportional to genetic innovation, as previously thought (Kirk, 2005).

Furthermore, genes undergoing decay and deletion are not exclusively younger, dynamic genes. *Chlamydomonas* orthologs of decaying genes are conserved; 72.2% fall under phylostrata 1 to 6 (Figure 2.9). Likewise, 73.7% of the *Chlamydomonas* orthologs for deleted genes are classified into PS1-6 (Figure 2.9). These results suggest that while roughly 27% of observed gene loss events reflect the evolutionary dynamics of young genes that might not be involved in increasing developmental complexity, most of the observed losses are likely linked to more consequential evolutionary events.

Previously, genes important for *Volvox* development were hypothesized to reflect key co-option events, including those needed for the evolution of multicellularity (Prochnik et al., 2010; Hanschen et al., 2016; Olson and Nedelcu, 2016). Hence, phylogenetic analyses were used to

gain deeper understanding of co-option in the volvocine algae. Cyclin D1 genes (Figure 2.10F), certain extracellular matrix (ECM) related genes (Table 2.9, Figure 2.11, 2.12), and the *Volvox* cell differentiation gene *regA* (Grochau-Wright et al., 2017) follow a duplication and divergence pattern that coincides with increases in complexity. However, other developmentally relevant genes such as other cyclins (Figure 2.10 A-E), *invA*, *B*, *C*, (Figure S2.13) and *glsA*, (Figure 2.14) do not. Genes that encode intrinsically disordered proteins (IDPs) were also examined, since these proteins might be co-opted for novel developmental functions in the volvocine algae (Dunker et al., 1998; Dunker et al., 2008). A relationship between complexity and protein disorder in TFs has been reported for a variety of organisms (Yruela et al., 2017). However, no significant change in the IDP content, frequency or distribution was observed between the volvocine algae genomes (Figure 2.15), even when testing exclusively for genes encoding protein-binding proteins (Figure 2.15).

Protein-protein interactions differ between volvocines even for conserved proteins

To understand how developmental complexity could evolve in the context of gene loss in the volvocine algae, the dynamics of their protein-protein interactions (PPIs) were examined. We reasoned that changes in PPIs could lead to impactful consequences to volvocine developmental complexity, even though few genes are undergoing positive selection or co-option (Figure 2.2G). Examination of silver-stained 2D-PAGE gels of total protein extracted from *Chlamydomonas*, *Gonium* and *Eudorina* shows extensive differences in the proteomic makeup of each species (Figure 2.16A, Figure S2.17A).

Changes in morphological and developmental complexity in the volvocine algae are thought to be related to the cytoskeleton (Kirk and Nishii, 2001). This is because cytoskeletal elements common to *Chlamydomonas* and *Volvox* are involved in cell division in both species,

and in the latter also in inversion (Kirk and Nishii, 2001). However, many cytoskeletal proteins are also highly conserved, as they are components of structures with essential roles in the cell (Li et al., 2004). Tubulin and actin proteins are highly conserved in the volvocine algae (Figure 2.18). Hence, the PPIs of actin and tubulin were examined to test for protein-level functional diversification in the context of changes to developmental complexity. As expected, complexes of similar size were found in *Chlamydomonas*, *Gonium*, and *Eudorina* 2D-PAGE immunoblots (Figure 2.16B, C, Figure 2.17), where minor differences in migration patterns for both proteins in the denaturing dimension are likely caused by post-translational modifications. However, novel α -tubulin complexes were shifted along the native axis, suggesting the presence of larger, species-specific complexes containing tubulin (Figure 2.16B, arrows). Similarly, 2D-PAGE immunoblots for β -actin showed a common streak in all three species, but also displayed species-specific actin-containing complexes that are larger than the shared complex (Figure 2.16C, arrows). Cultures used to extract protein were asynchronous; hence, the differential migration patterns between species are not likely to result from comparing species at differential life stages. Taken together, these results indicate that while the genomes of the volvocine algae are highly similar and experiencing gene loss, PPIs between conserved gene products are forming species-specific complexes, which could account for the observed differences in biological complexity between the volvocines.

Gene loss is a driver for regulatory innovation

To evaluate the potential of gene loss as a mechanism of evolutionary transitions, gene loss and gain were modeled using an adaptation of the Wagner gene regulatory network model (Wagner, 1994). Our model mirrors the effects of gene gain in Wagner's original model, which demonstrated that networks more often reach a new equilibrium state following an intermediate

number of gene duplications than when there are many or few duplications (Wagner, 1994). Our model predicts similar outcomes following gene loss with the proportion of novel stable gene network configurations being higher following gene deletions as compared to an equal number of duplications (Figure 2.19). Upon iterating the model over a broad range of network connectivity, similar results were obtained (Figure 2.20). Through network-wide loss of robustness, novel stable network states might result in regulatory rewiring that allows for major changes to developmental programs. In sum, our evidence indicates that the signature of gene loss in the volvocine genomes may represent an example of a broadly used mechanism for the evolution of developmental complexity.

Discussion

Our work further challenges the common assumption that developmental complexity and genomic repertoire are positively correlated (Adami et al., 2000; Trigos et al., 2018). If this were the case, the developmental gains in *Volvox* and species that exhibit seemingly intermediate steps of developmental complexity (typified by *Gonium*, *Pandorina*, *Yamagishiella* and *Eudorina*) might be the consequence of the acquisition of novel genes and/or the result of co-option events correlating to their developmental gains (Cheng et al., 2003; Nishii et al., 2003; Olson and Nedelcu, 2016; Grochau-Wright et al., 2017). However, we discovered that a very small set of expanding and co-opted genes are found in the five multicellular species (Table 2.9, Figure 2.10 – 2.14), some of which have an evolutionary signature of sequential co-option. With few exceptions (e.g., *reg*, extracellular matrix (ECM) genes, cyclin D), it was not possible to assign novel genes to specific developmental gains in the volvocine algae, despite the complexity of these novel traits.

Notwithstanding the similarities between volvocine genomes (Figure 2.1B, C), genetic losses in conserved genomic regions have occurred in volvocine history (Figure 2.6). Orthologous groups and Pfam domains both show more numerous significant gene losses than they do gains, and the rate of loss for some of these groups can be quite dramatic (Figure 2.2A, B), even in the absence of significant variation in the number of Pfam domain counts per gene between the six species (Figure 2.3). These results agree with previous work describing pervasive Pfam loss events across eukaryotes (Zmasek, 2011). Regulatory and histone genes, which are important for development (Carroll, 2001; Hanschen et al., 2016; Gaiti et al., 2017), are not exceptions to this trend (Figure 2.2C, D, E). The progressive contraction of histone genes as complexity increases for the volvocine lineage allows for the possibility of gene loss to be directly related to increased complexity in this lineage, either as a driver or as a consequence of shifts in selective pressures.

The non-uniform distribution of gene loss events across the five multicellular species (Figure 2.6A, C), as well as the signatures of decay found for many of these genes (Figure 2.6B), contend that the observed patterns are not a consequence of annotation disparities or analytical biases. Events of gene loss throughout evolution are not unique to the volvocines, suggesting that this is a widespread phenomenon (Wolf and Koonin, 2013; Bhattacharya et al., 2018; Bowles et al., 2019; Lopez-Escardo et al., 2019). However, this is the first time to our knowledge that the mode of gene loss has been traced within a lineage. The mechanisms underlying the observed rates of decay for genes lost remain enigmatic, but their presence allows for the inference that most gene loss events described here are *bona fide* and, not *Chlamydomonas* specific gene gains. Evidence for gene loss through gradual decay in the context minimal duplication and divergence differs from the model of differential loss of paralogs that has been proposed for metazoans

(Fernandez and Gabaldon, 2020). It also appears to differ from other lineages (Zmasek, 2011) in the sense that there is no clear functional bias related to gene loss (Figure 2.7); instead, it would seem that a wide array of functions are differentially preserved.

Previous reports posited that gene duplication is a costly process due to the expense of synthesizing the gene products and the impact of increased function/gene dosage alteration on the cell (Adler et al., 2014). Moreover, novel stable regulatory networks can be readily generated via gene loss (Figure 2.19, Figure 2.20). Hence, both theoretical and biological systems suggest that gene gains are neither the only nor the fastest route to biological innovation. Indeed, it has been hypothesized that the addition and subtraction of elements in regulatory networks across species (True, 2001), as well as co-option and differential spatiotemporal combinations of conserved genes/networks and processes (Benitez et al., 2018) participate of generating some key aspects of organismal body plans. Our data supports the idea that changes to developmental complexity, such as the transition to multicellularity, could be driven by re-purposing existing genes for new functions while eliminating others as opposed to by generating of extensive genetic novelty (Jimenez-Marin and Olson, 2022).

The patterns of gene loss and PPI diversification in the volvocine algae illustrate vigorous genomic and proteomic shifts that likely may underlie changes to evolutionary and developmental constraints between members of a single lineage. A significant proportion of the gene losses (~55%) identified in the volvocine algae were shared by the five colonial species, whereas other gene loss patterns were less numerous (Figure 2.6C). The ancestor of undifferentiated *Gonium* (and the other multicellular species studied) would then appear to be a candidate hotspot for initiation of gene loss and co-option. Analysis of genes with a signature for positive selection shows *Gonium* has the largest number of genes under positive selection among

the species we studied (Figure 2.2G), and it appears to have undergone diversification of histone H3 N-terminal tails (Figure 2.2F). Perhaps the developmental complexity portrayed by *Volvox* was made possible by the same mechanisms and tools visible in *Gonium*, such as co-option of RB (Olson and Nedelcu, 2016). Recent findings suggest volvocine undifferentiated multicellularity has evolved more than once (Lindsey et al., 2021). Likewise, differentiated multicellularity has arisen multiple times (Nozaki et al., 2000; Albalat and Canestro, 2016), as has the volvocine spheroidal colony morphology (Yamashita et al., 2016). For instance, *Astrephomene* is notably similar to *Eudorina* phenotypically despite not belonging to Volvocaceae (Lindsey et al., 2021), and its embryogenesis and development is markedly different from that of other spheroidal volvocines (Yamashita et al., 2016). Future work comparing independent instances of evolution of simple multicellularity will no doubt provide valuable insight into the mechanistic similarities and differences underlying these convergent phenotypes. For the time being, the comparisons between *Chlamydomonas*, *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* support the possibility that complex multicellularity is a relatively straightforward consequence of simple multicellularity. Should this be the case, a future challenge would be to test this hypothesis in other multicellularity models as means to assess if the true evolutionary transition lies in establishing the conditions that stabilize undifferentiated multicellularity through loss of gene functions that might be necessary to maintain the unicellular state. In sum, our results indicate that biological innovation coincides with gene loss events, which could allow for reconfiguration of gene networks, differential use of existing functional repertoires, and reduction of the impact of genetic gains.

Methods

Strains and genomes

Chlamydomonas reinhardtii CC-503 MT⁺ genome version 5.5 (Merchant et al., 2007) (GCA_000002595.3), *Tettrabaena socialis* NIES 571 (GCA_002891735.1) (Featherston et al., 2017), *Gonium pectorale* K3-F3-4 MT⁻ NIES-2863 (Hanschen et al., 2016) (GCA_001584585.1), and *Volvox carteri f. nagariensis*, Eve S1 (Prochnik et al., 2010) genome versions 1 (GCA_000143455.1) and 2.1 (available on Phytozome) (Prochnik et al., 2010; Goodstein et al., 2012) were used for all analyses. Genome assemblies for *Yamagishiella* and *Eudorina* were kindly provided by Dr. Hisayoshi Nozaki before publication (Hamaji et al., 2018). All analyses involving *Yamagishiella* were performed using *Y. unicocca* strains 2012-1026-YU-F2-6 MT⁺ NIES-3982, and 2012-1026-YU-F2-1 MT⁻ NIES-3983. For *Eudorina*, *Eudorina* sp. strains 2010-623-F1-E4 female NIES-3984 and 2010-623-F1-E2 male NIES-3985 were utilized. Analyzed red algae genomes have been published previously: *Porphyra umbilicalis* (Collen et al., 2013) (GCA_002049455.2), *Pyropia yezoensis* (Bhattacharya et al., 2013) (GCA_009829735.1), *Chondrus crispus* (Nozaki et al., 2007) (GCA_000350225.2), *Porphyridium purpureum* (Bhattacharya et al., 2013) (GCA_000397085.1) and *Cyanidioschyzon merolae* (Nozaki et al., 2007) (GCA_000091205.1). Chlorophyte outgroup species genomes have been published previously: *Chromochloris zofingiensis* (available on Phytozome) (Roth et al., 2017), *Chlorella variabilis* (Blanc et al., 2010) (ADIC000000000), *Micromonas pusilla* (Worden et al., 2009) (GCA_000151265.1), and *Ostreococcus lucimarinus* (Palenik et al., 2007) (GCA_000092065.1).

Algae growth conditions and library preparation

Cultures of *Pandorina morum* 5 (MT⁻), a gift from Dr. Hisayoshi Nozaki, were initially decontaminated by collecting colonies grown in Standard Volvox Media (SVM) by centrifugation at 1000 *g* for 5 min at 25°C and removal of the growth media. These colony pellets were overlaid with 50% Percoll® (Millipore-Sigma) in SVM plus Acetate (SVMA) followed by illumination with 100 uE of light at the top of the tube. The top 1 mL of colonies that were able to phototax through this were then plated for single colonies on SVMA agar plates supplemented with 50 ug/mL carbenicillin, 50 ug/mL cefotaxime, 10 ug/mL trimethoprim and 20 ug/mL chloramphenicol. Cultures from single isolates were established in SVM and grown with 0.5% CO₂ bubbling at 30°C, 350 $\mu\text{mol m}^{-2}\text{sec}^{-1}$ light intensity on a 14:10 day:night cycle.

Cultures were harvested by centrifugation at 600*g* for 10 min with 0.005% Tween-20. DNA was isolated using a magnetic bead protocol to obtain high molecular weight DNA (Mayjonade et al., 2016). Illumina TruSeq PCR-free libraries were prepared from DNA that was either unfractionated or fractionated on a Pippin Prep (Sage Science) in the size ranges of 1kb, 2kb, 5kb and 10kb. Following library preparation, they were sequenced on an Illumina HiSeq2500 in Rapid Run mode to an estimated coverage of 25X. High molecular weight DNA was also subjected to Pippin Prep fractionation for 15 kb molecules, followed by PacBio SMRTbell library preparation. Libraries were sequenced on a PacBio Sequel system to an estimated 10X coverage. All sequencing was performed by Genewiz (South Plainfield, New Jersey, USA).

Yamagishiella unicocca 2012-1026-YU-F2-1 (NIES-3983, MT⁻) and *Eudorina* sp. strain 2010-623-F1-E4 (NIES-3984, female) were axenically cultured for RNA-seq experiments. Algae were grown SVM with 0.5% CO₂ bubbling at 30°C, 350 $\mu\text{mol m}^{-2}\text{sec}^{-1}$ light intensity on a 14:10 day:night cycle. Triplicate cultures were harvested by centrifugation at 600*g*

for 10 min with 0.005% Tween-20. RNA was extracted using a Quick-RNA Plant Miniprep Kit (Zymo Research). RNA-Seq libraries were prepared using the TruSeq Stranded mRNA kit (Illumina, Inc. San Diego, California, 20020595) according to manufacturer's conditions. Axenic cultures for 2D PAGE were grown asynchronously as follows: *Chlamydomonas reinhardtii* strain 21gr (*Chlamydomonas* collection CC-4889) was grown in TAP media to a density of $\sim 10^6$ cells/mL in ambient conditions with 0.5% CO₂ bubbling. *Gonium pectorale* (strain Kaneko4, MT⁺ NIES-1711) was grown in SVM + acetate media to a density of $\sim 10^6$ colonies/mL in ambient conditions with 0.5% CO₂ bubbling. *Eudorina* sp. strain 2010-623-F1-E4 (NIES-3984, female) was grown in SVM to a density of $\sim 10^6$ colonies/mL at 30°C with 350 $\mu\text{mol m}^{-1} \text{sec}^{-1}$ light intensity on a 14:10 day: night cycle with 0.5% CO₂ bubbling.

Genome assembly

Illumina reads were assembled *de novo* with ABySS 2.1.5 (Simpson et al., 2009) using a range of k-mers from 21-89, in increments of 4. Unitigs from each of the assemblies were then filtered against a database of known bacterial genomes to remove presumed bacterial endophyte sequences. Remaining unitigs were assembled with ABySS under the same conditions as above. Resulting scaffolds were merged to PacBio reads and assembled using Canu 1.8 (Koren et al., 2017) (Bioproject PRJNA787305).

Evidence-based gene prediction

The gene models for *Yamagishiella* and *Eudorina* genomes were generated using RNA-seq data and the AUGUSTUS 3.2.3 software package (Stanke and Morgenstern, 2005), in a manner similar to gene predictions for *Volvox* and *Gonium* (Prochnik et al., 2010; Hanschen et al., 2016). The gene models for *Yamagishiella* and *Eudorina* genomes were generated using RNA-seq data (Bioproject PRJNA787306 and PRJNA787302) and the AUGUSTUS 3.2.3 software package

(Stanke and Morgenstern, 2005), in a manner similar to gene predictions for *Volvox* and *Gonium* (Prochnik et al., 2010; Hanschen et al., 2016; Hanschen et al., 2018b). AUGUSTUS predicted genes were merged with the genome mapped reads and de-duplicated to remove any splice variants.

Identification of orthologous groups

Orthologous groups of volvocine algae and red algae genomes were determined using OrthoMCL 2.0.9 (Li et al., 2003). The optimum inflation value was determined empirically by testing a variety of values ranging from 1.2 to 4.0 in increments of 0.1 (Figure 2.21). The inflation value of 1.9 was used for all analyses.

Identification of Pfam domains

The diversity and abundance of known Pfam domains in the volvocine algae and red algae genomes were identified using Pfam Scan (Finn et al., 2014) and database version 31. Pfam Scan and associated data collection Perl scripts available at: <ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/> (El-Gebali et al., 2019).

Analysis of other genomic features

Transcription factors and protein kinases in the annotated volvocine algae genomes were identified using the iTAK version 1.7 program and database (Zheng et al., 2016). The *Chlamydomonas* gene definitions on GenBank were used to identify histone H1 (Accession #XM_001696120) and quantified (Figure 2.2E second panel, Table 2.5). Histones H2A, H2B, H3, H4 (Accession # L41841) genes in the volvocine algae were identified using the reciprocal best BLAST hit method with a e-value cutoff of 1e-5. H2A, H2B, H3 and H4 genes were further filtered using e-values of 1e-10 and quantified (Figure 2.2E panels 3-6, Figure 2.2F first panel, Table 2.6). The first 30 amino acids from each histone were used for analysis of duplication and

presence of predicted post-translational modification sites (lysine methylation) using MethyIsight (Biggar et al., 2020). N-tails per histone per species that had different predicted sites with threshold ≥ 0.5 were considered tail variants and quantified (Figure 2.2F second panel, Table 2.6). Histone genes were also identified in the five red algae genomes using gene definitions from the *Chondrus* genome annotation (Table 2.7). Eight internal control genes, Actin, mitochondrial ATP-synthase α subunit (mATP-A), mitochondrial ATP-synthase β subunit mATP-B, α -tubulin, β -tubulin, β -2 tubulin, flagellar dynein heavy chain, chloroplast ferridoxin were identified using OrthoMCL (Table 2.4).

Identification of gene count trends

Rate of gene number change was calculated by applying a linear regression model on the relationship between gene counts per orthologous group per species and distance of said species from their Last Common Ancestor (LCA, Figure 2.1A). Only groups with non-identical gene copy numbers and comprising genes from at least 2 species were used for this analysis. Rate of change values per orthologous group were subjected to quartile analysis to identify outlier groups (Figure 2A). This same analysis was performed on Pfam domains (Figure 2.2B), transcription factors (Figure 2.2C), and protein kinases (Figure 2.2D) for volvocine species and for red algae species. Regression and quartile analyses were performed using custom Python scripts.

The orthologous groups used for the regression analysis were also subjected to gene gain and loss parsimony-based analysis using GLOOME (Cohen et al., 2010) web server. Since GLOOME requires patterns of presence/absence of gene families to assess gain and loss trends, orthologous groups where the number of member genes per species are all greater than zero were not included. The contracting orthologous groups from the rate of change analysis described

above were then searched against the GLOOME output and quantified. Contracting orthologous groups that belong to the non-included dataset were also quantified.

Identification of positively selected genes and whole-genome-based phylogeny

Positively selected genes were identified using the PosiGene pipeline (Sahm et al., 2017). Two independent PosiGene analyses were performed using *Chlamydomonas* and *Volvox* genome annotations as anchors. Chlorophytes *Chlorella variabilis*, *Micromonas pusilla*, *Ostreococcus lucimarinus* and *Chromochloris zofingiensis* were used as outgroups. Similarly, the red algae genes under positive selection were identified using *Porphyra umbilicalis* as anchor species as it has previously been reported to be closest to the LCA of red algae (Alexander et al., 2003; Yang et al., 2016). A gene was considered to be positively selected if the dN/dS value (HA foreground omega) was greater than 1 and the Bonferroni-corrected p-value was less than 0.05. As a part of the workflow, PosiGene generates a phylogenetic tree from the isoform assignments and calculates the distance to LCA (Figure 2.1A, Figure 2.22).

Characterization of gene loss patterns

To evaluate annotation completeness, BUSCO v4.0.2 (Seppey M., 2019) was run in protein mode using the Chlorophyta lineage dataset for each volvocine species (Table S2.2).

Orthologous groups showing significant contraction (Figure 2.2A) were selected for gene loss analysis based on whether they included *Chlamydomonas* genes. *Chlamydomonas* loci that were not assigned to chromosomes were not analyzed further. Orthologous groups that included *Chlamydomonas* loci with the aforementioned characteristics (target loci) served as anchors for obtention of genes up and downstream, called the ‘syntenic neighborhood’. Orthologs for loci in the syntenic neighborhood, as well as target loci for *Gonium*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* were retrieved from their respective orthologous groups. Instances of no

ortholog found for target loci were subjected to tBLASTx (Altschul et al., 1990) 2.2.18+ searches using tailored databases based on syntenic neighborhoods for each species. Non-informative (loss of syntenic neighborhoods or no BLAST results) genes were omitted from further analyses. tBLASTx hits were filtered for quality (e-value < 1e-5). Hits passing quality threshold were considered to be evidence of a gene remnant suggestive of loss via decay. Genes with no hits passing quality threshold and conserving synteny are candidate deleted genes. Decayed and deleted genes were assigned to bins corresponding to their pattern of absence (e.g. a gene lost by whichever mechanism only in *Gonium*, but present in the other clades; see Figure 2.6C and Table 2.8), or by type of event, where decayed gene clusters consist of orthologs with at least one signature of decay in multicellular species, and deleted gene clusters are those where all losses happen via deletion (Figure 2.6B). Percentages of decay and deletion cases were calculated, and a Chi square goodness-of-fit test was applied to determine statistical significance of observed patterns using a uniform distribution as expected probability distribution. All analyses except for tBLASTx search were performed using custom Python scripts. Treemaps for GO analysis of retained, decayed and deleted genes were generated using REVIGO (Supek et al., 2011) with default parameters and against *Chlamydomonas reinhardtii* database.

Identification and analysis of known co-opted genes

Cell cycle genes in *Pandorina*, *Yamagishiella* and *Eudorina* genomes were annotated as previously described (Prochnik et al., 2010; Hanschen et al., 2016) using tBLASTn. Transcriptome data and manual curation (Prochnik et al., 2010; Hanschen et al., 2016) was used to prepare gene models for *Yamagishiella* and *Eudorina*. Known pherophorins are composed of two pherophorin domains (Pfam DUF3707) that are connected by a variable length hydroxyproline-rich repeat region (Kirk et al., 1986). All genes containing two or more

pherophorin domains were identified as pherophorin genes. Matrix metalloproteases (MMPs) contain a single metalloprotease domain (Peptidase M11, PF05548). Genes containing this Pfam domain were selected for the existence of the [HQ]EXXHHXGXXH motif in the gene model (Hallmann, 2003). The *invA*, *invB*, *invC* and *glsA* genes were identified in *Pandorina*, *Yamagishiella* and *Eudorina* by the presence of kinesin (PF00225), TPT (PF03151), glycosyl-transferase for dystroclygan (PF13896), and DnaJ (PF00226) domains respectively (Cheng et al., 2003; Nishii et al., 2003).

Prediction of lineage specific genes

Phylostratigraphy (Domazet-Loso et al., 2007) was used to predict the lineage of genes in the volvocine algae genomes as described previously (Hanschen et al., 2016). The phylogenetic classes of volvocine algae were defined in the input data as per NCBI taxonomy. All protein sequences were searched against the NCBI non-redundant database from version 2.6+ (Altschul et al., 1990), with an e-value cutoff of 1e-3. Phylostratigraphic classification of each gene ranged from PS-1 to PS-9. Gene level phylostratigraphic classification was converted to Pfam domain level classification.

Identification of protein disorder and protein binding site disorder

Protein disorder was calculated using DISOPRED3 (Ward et al., 2004; Jones and Cozzetto, 2015), that yields percentage disorder in each protein sequence as well as that of protein binding sites. The frequency distribution of percentage protein disorder was calculated and plotted.

2D gel electrophoresis and silver stain

Approximately 10⁶ cells/colonies of *Chlamydomonas*, *Gonium*, and *Eudorina* were pelleted and snap frozen in liquid nitrogen. Cells were lysed by six rounds of liquid nitrogen freeze/thaw lysis in in low salt TBS (50 mM Tris, 50 mM NaCl, pH 7.5) and protease/phosphatase inhibitors

(Plant Protease Inhibitor Cocktail (Sigma, Burlington, Massachusetts), 1 mM Benzamidine, 1 mM PMSF, 1 mM Na₃VO₄, 5 mM NaF, 0.5 mM β -glycerophosphate, 0.1 mM EDTA) (Olson et al., 2010)). Insoluble material was pelleted by centrifugation at 20,000 g for 10 min at 4°C. Supernatant was removed and protein content was quantified by Bradford assay (BioRad, Hercules, California) (Olson and Markwell, 2007). For blue native PAGE, 20 μ g of protein per well was loaded into a 4-12% native PAGE gel (Fisher, Waltham, Massachusetts) with native running buffer (Fisher) and light blue buffer at the cathode (Fisher). Gel was run until the Coomassie dye front had exited the gel. Lanes were excised using a razor and incubated in 1x Tris-MOPS running buffer (Genscript, Piscataway, New Jersey) with 100 mM DTT and 5 mM sodium bisulfite at 55°C for 10 min. Gel slice was loaded into a NuPAGE 4-12% Bis-Tris gel (Fisher) and run with Tris-MOPS running buffer with 5 mM sodium bisulfite in the cathode buffer until the blue dye front had exited the gel. Silver staining was carried out using a Silver Staining Kit (Pierce) according to manufacturer's protocols.

Western blot and antibodies

Western transfer was performed in Tris-Glycine transfer buffer to a nitrocellulose membrane (0.45 μ m pore, Pall Scientific, Port Washington, New York) (Olson et al., 2010). Membranes were blocked in 1% milk (α -tubulin) or 5% BSA (β -Actin) for 1 hr. Primary antibodies (mouse anti- α tubulin, Sigma T6074 1:5000; mouse anti- β actin HRP, Santa Cruz sc-47778 HRP, 1:5000, Dallas, Texas) were incubated at 4°C overnight with gentle shaking. Membranes were washed with TBS 0.1% Tween-20 (TBST) and tubulin blots were probed with anti-mouse secondary (Pierce 31430, Rockford, Illinois) for 1 hr. Actin blots did not require a secondary antibody, as mouse anti- β actin is already HRP conjugated. Membranes were developed using

Premium Western Blotting Reagent (LI-COR, Lincoln, Nebraska) and scanned for chemiluminescence on a C-Digit Scanner (LI-COR).

Phylogenetic analysis

In order to determine phylogenetic relationships, the genes from all six volvocine genomes and *Coccomyxa subellipsoidea* (Chlorophyta, GCF_000258705.1) were aligned using MUSCLE v3.8.425 (Edgar, 2004) and a phylogenetic tree was produced using RaxML v8 (Stamatakis, 2014) with protein gamma model and automatic model selection. A rapid bootstrapping analysis with 1000 bootstrap replicates was performed and a majority ruled consensus tree was produced.

Regulatory network model

A Wagner regulatory network model (Wagner, 1994) was designed as follows: a gene network of N genes whose on or off expression state at a time t is given as:

$$\vec{S}(t) := (S_1(t), \dots, S_N(t))$$

Where $S_i(t)$ is the expression state of the i^{th} gene at time t , $S_i=0$ reflects an off state, and $S_i=1$ reflects an on state. The state $S_i(t)$ is determined by the network of interactions affecting gene i . The expression state of gene i can be changed due to regulatory interactions. These changes can be modelled as a set of differential equations:

$$S_i(t + \tau) = \sigma \left[\sum_{j=1}^N w_{ij} S_j(t) \right]$$

Here, $\sigma(x)$ is the sign function similar to Wagner's model and w_{ij} is the weight of interaction between gene i and gene j . The connectivity matrix is given by w . Overall connectivity of the network (c) is given by the average fraction of entries in w that are different from zeros (i.e. $c \in (0,1)$). The effect of gene duplication on a random network was assessed as in Wagner (Wagner, 1994).

The effect of gene deletion on a random network was assessed as follows. A network of size N was considered. The initial and final states of the gene network were randomly chosen where each individual gene has an equal probability of being on or off (i.e. $p(S_i(0) = 1) = p(S_i^{eq} = 1) = 0.5$). Entries in w were randomly and independently chosen from the Gaussian distribution ($\mu = 0, \sigma = 1$). Only weight matrices that lead to stable expression patterns ($\vec{S}^{eq}$) as $t \rightarrow \infty$ were retained. We iterated the differential equation 100 times to confirm if w results in $\vec{S}^{eq}$ when the initial state is $\vec{S}(0)$. If not, a new w was chosen. Once the network set $(\vec{S}(0), w, \vec{S}^{eq})$ was found, k genes ($k \in (1,9)$) were randomly deleted from the $\vec{S}(0)$ to give $\vec{S}^\Delta(0)$. This was done by randomly setting the expression state of the k genes to 0. The connectivity matrix w was updated such that the effect of remaining genes were set to 0. However, the effect of the deleted genes on the remaining genes were randomly selected from the Gaussian distribution. The updated connectivity matrix w was then used to simulate the new final expression state $\vec{S}^{\Delta eq}()$. Hamming distance (d_h) was calculated between $\vec{S}^{\Delta eq}$ and $\vec{S}^{eq}$. For each k , 1000 iterations of simulation were done. The effect of gene deletion was reported as the proportion of instances where $d_h > k$. Calculation of proportion of novel stable networks for gene gain and loss was performed using custom scripts.

Figures and tables

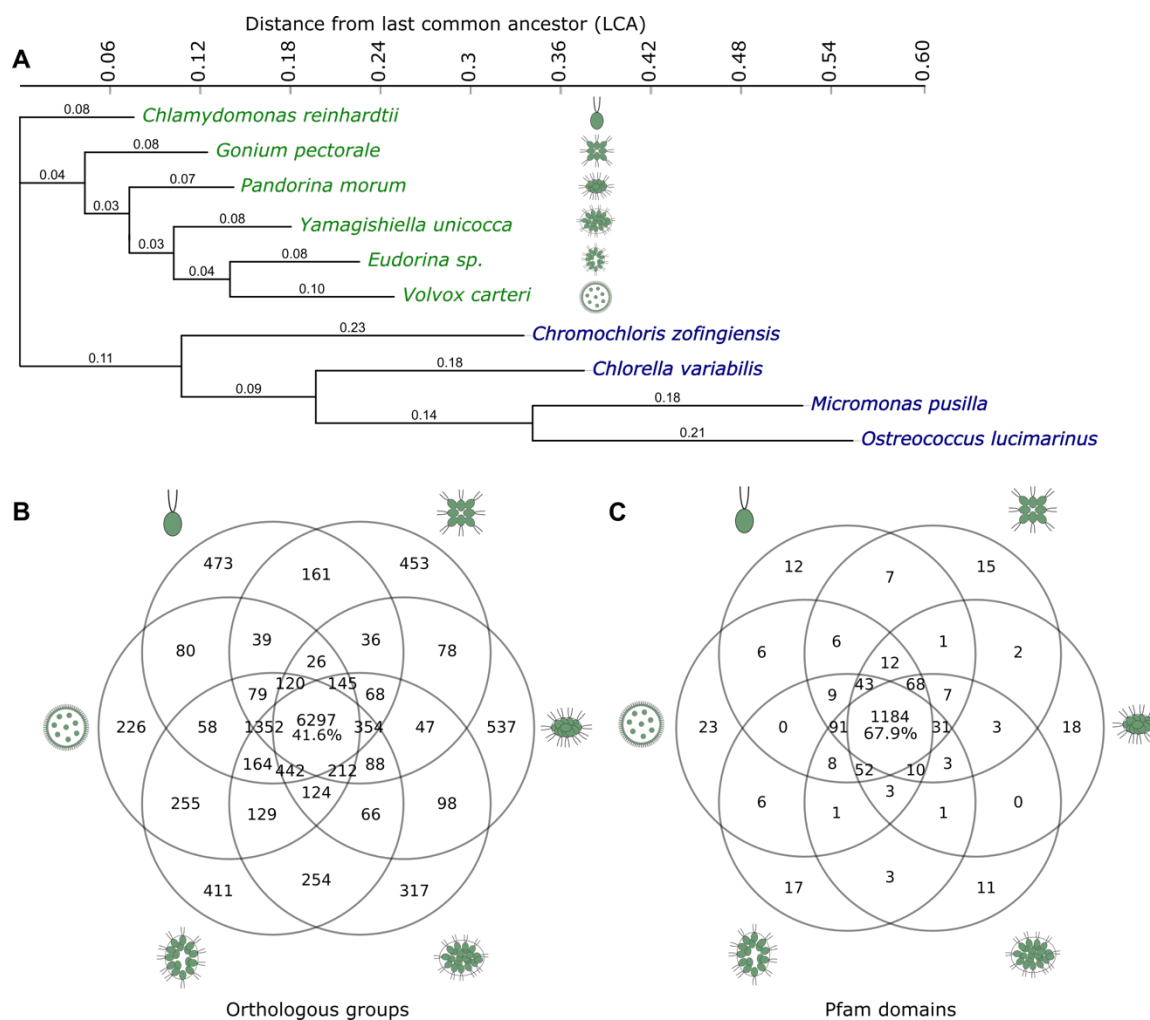


Figure 2.1. Volvocine algal genomes are similar.

A) Genome-wide phylogeny of the volvocine algae (green) compared to other chlorophyte outgroups (blue) inferred by PosiGene. Branch values indicate distance from the nearest node. Note *Chlamydomonas* is unicellular and closely related to colonial volvocines. B) Orthologous groups shared between the six volvocine species. C), Pfam domains shared between six volvocine species.

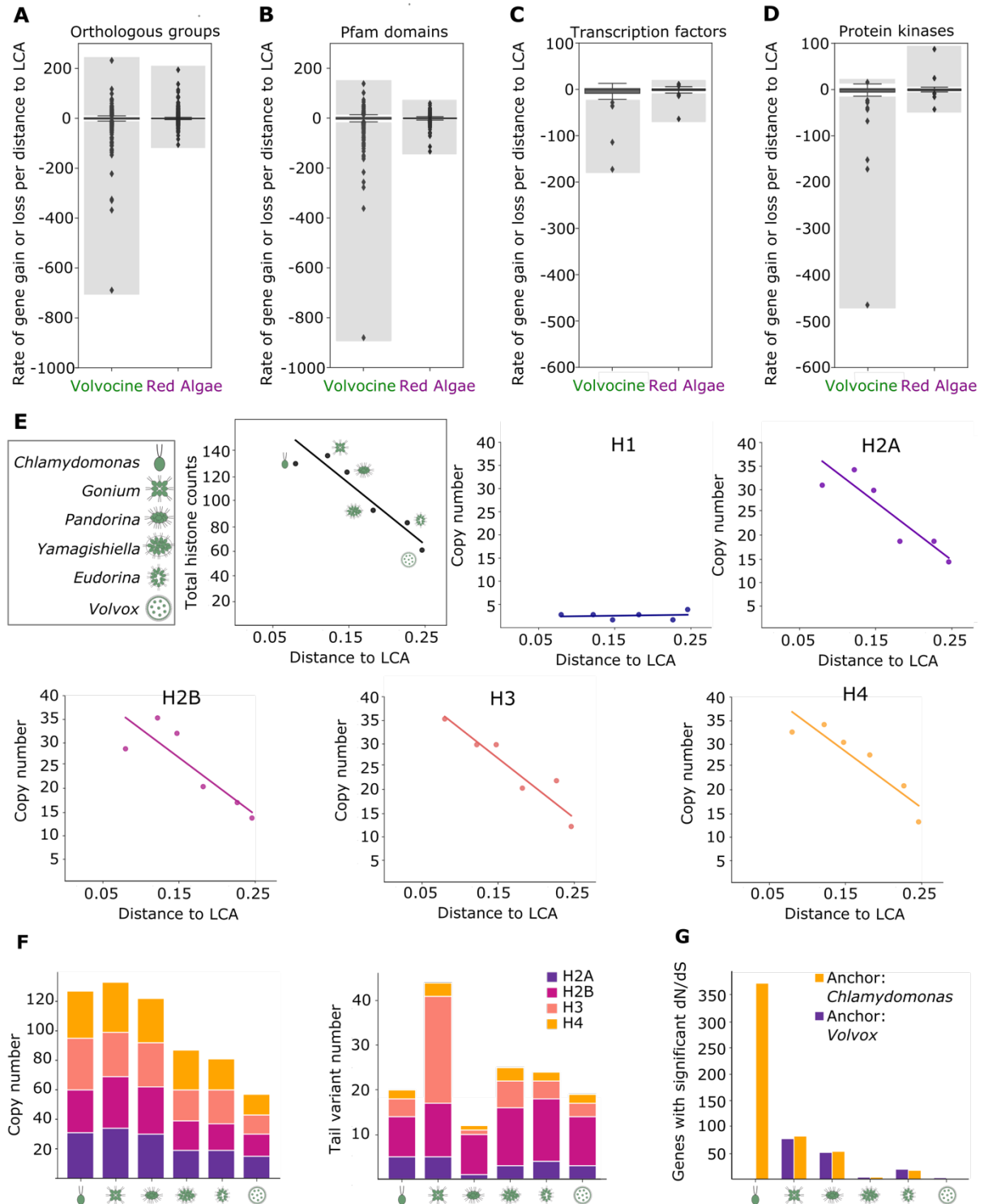


Figure 2.2. Gene loss outpaces gain in volvocine algae genomes.

Distribution of the gene loss and gain rates per distance to LCA in the volvocine genomes compared to Rhodophyceae (red algae) of A) orthologous groups, B) Pfam domains, C)

transcription factors, and D) protein kinases. Outlier groups, which represent significant loss (negative values) or gain (positive values) rates, are within the shaded grey area. E) Histone copy number relative to distance to LCA shows a decreasing trend except for H1 in the volvocine genomes. Top left panel describes total histone counts per species. F) Histone copy numbers (left) are reduced with minimal loss of tail variants (right). G) Quantification of positively selected genes per species using *Chlamydomonas* (yellow) and *Volvox* (purple) as reference genomes.

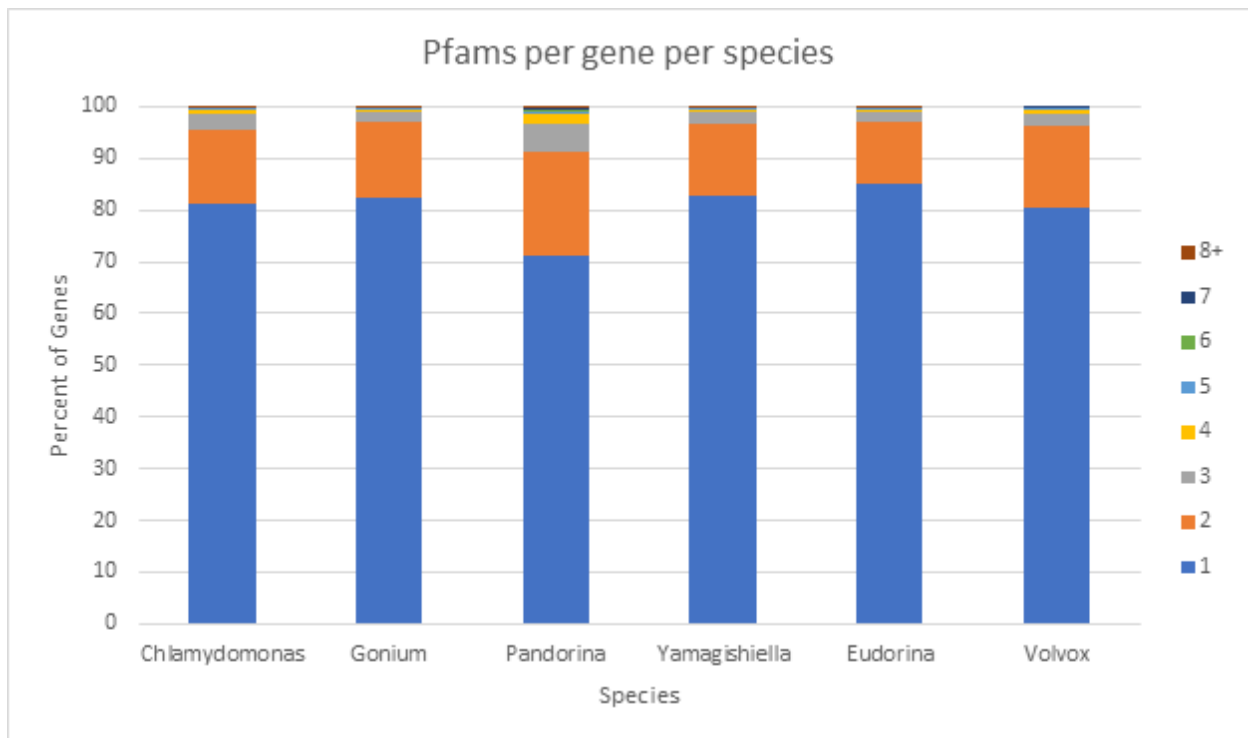


Figure 2.3. Number of Pfams per gene does not change in the five volvocine algae species.

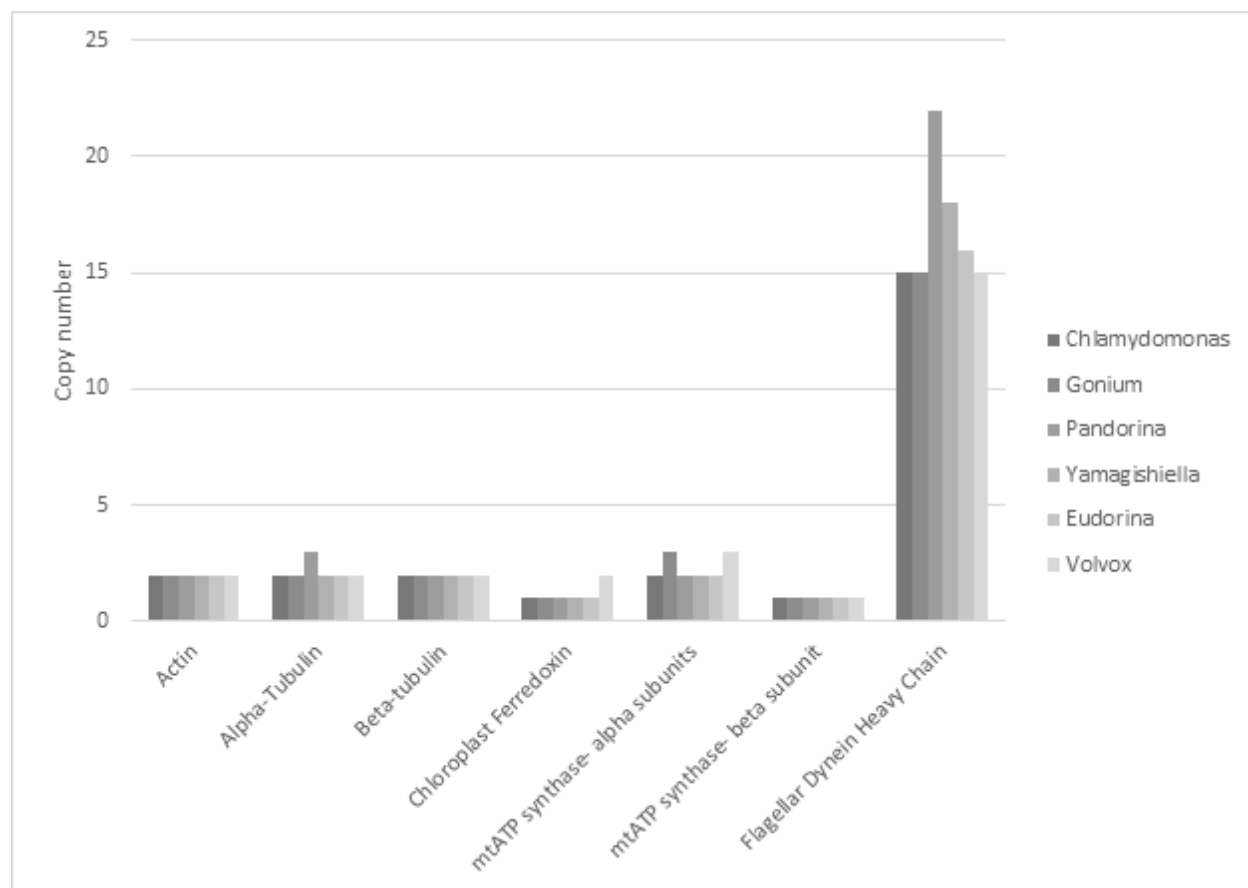
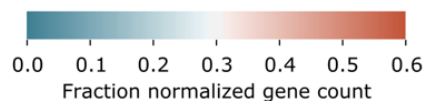
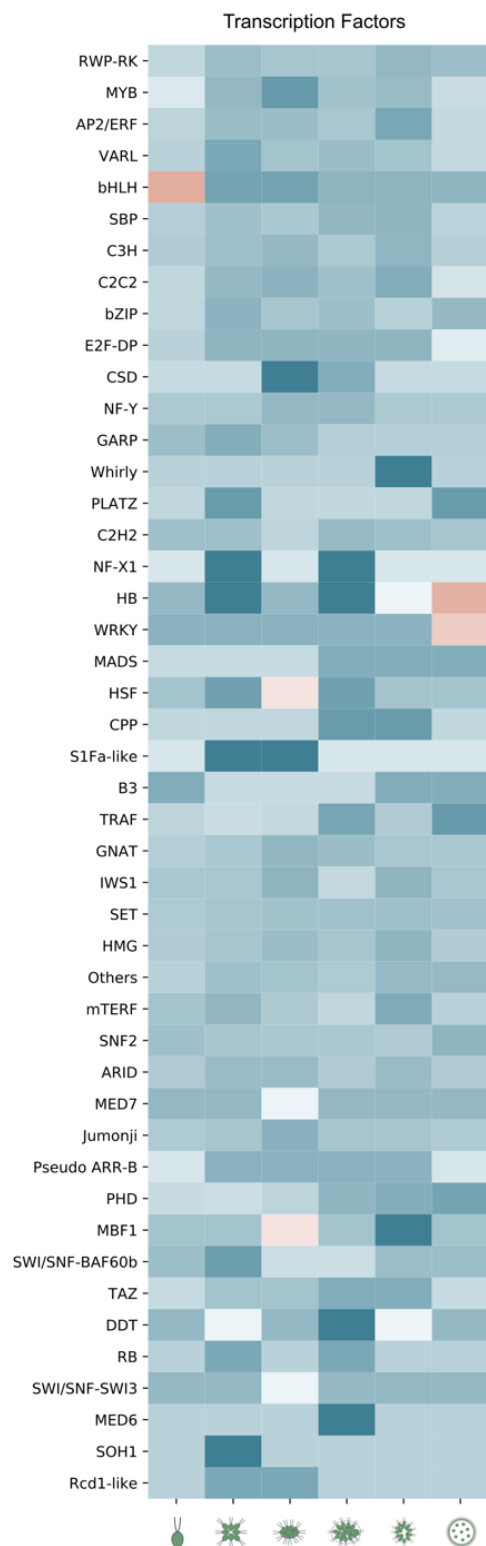


Figure 2.4. Gene copy numbers of nine internal control genes in the six volvocine species.



A



B

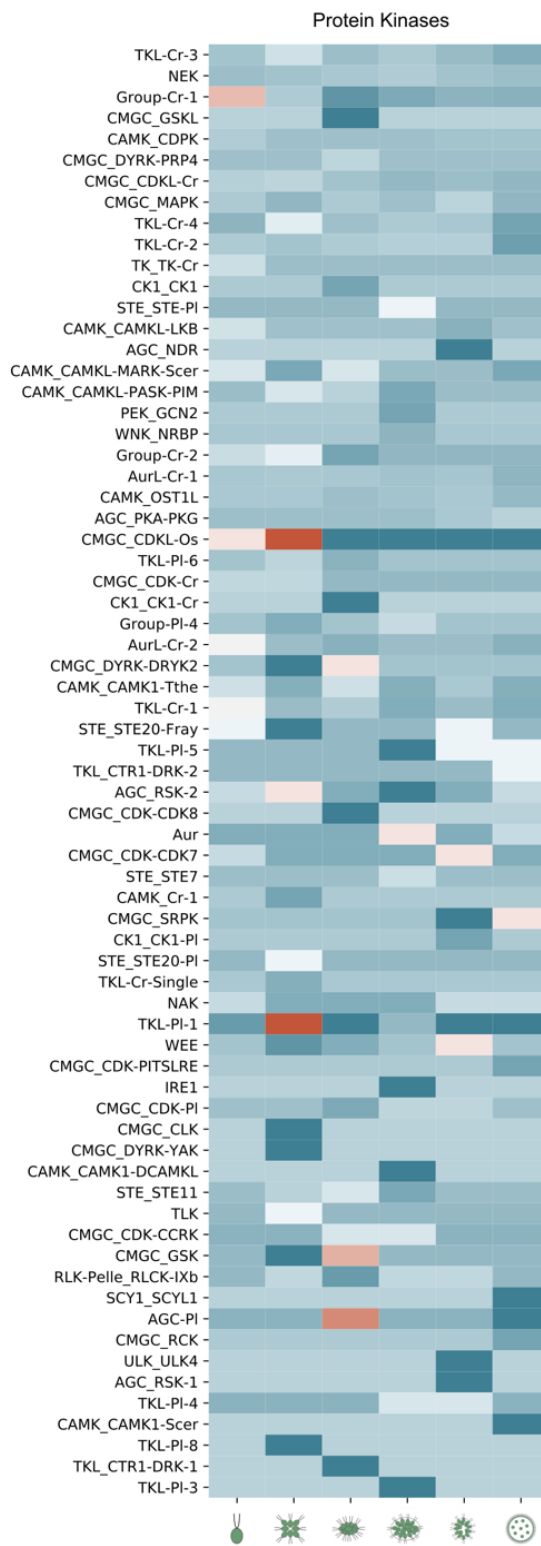


Figure 2.5. A) Transcription factor and B) protein kinase family loss and gain in the six volvocine genomes.

Relative gene counts represent the fraction normalized count relative to counts of the other species.

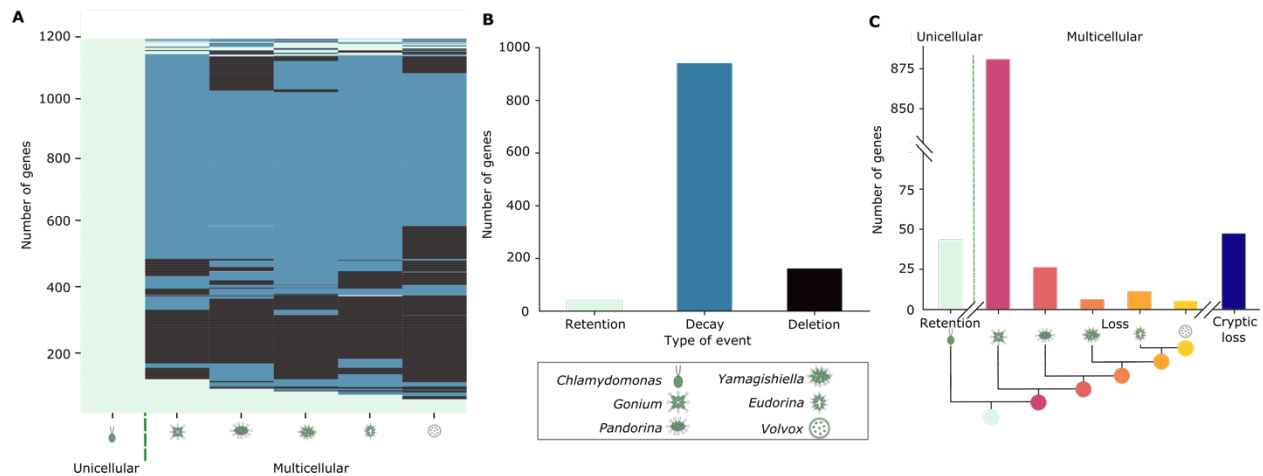


Figure 2.6. Gene loss in the volvocine algae occurs primarily by gradual decay.

A) Retention, decay, and deletion of 1,184 loci belonging to orthologous groups undergoing contraction in volvocine species. B) Cross species quantification of retention (light green), decay (blue), and deletion (black) of loci in contracting orthologous groups. C) Quantification of events of retention and loss by decay relative to evolutionary relationships between volvocine species. Losses that do not follow phylogenetic distributions are considered cryptic (dark blue).



Figure 2.7. GO terms for biological process and molecular functions of *Chlamydomonas* orthologs of retained (55.8% of total retained genes), decayed (24% of total decayed genes), and deleted (25.3% of total deleted genes) genes.

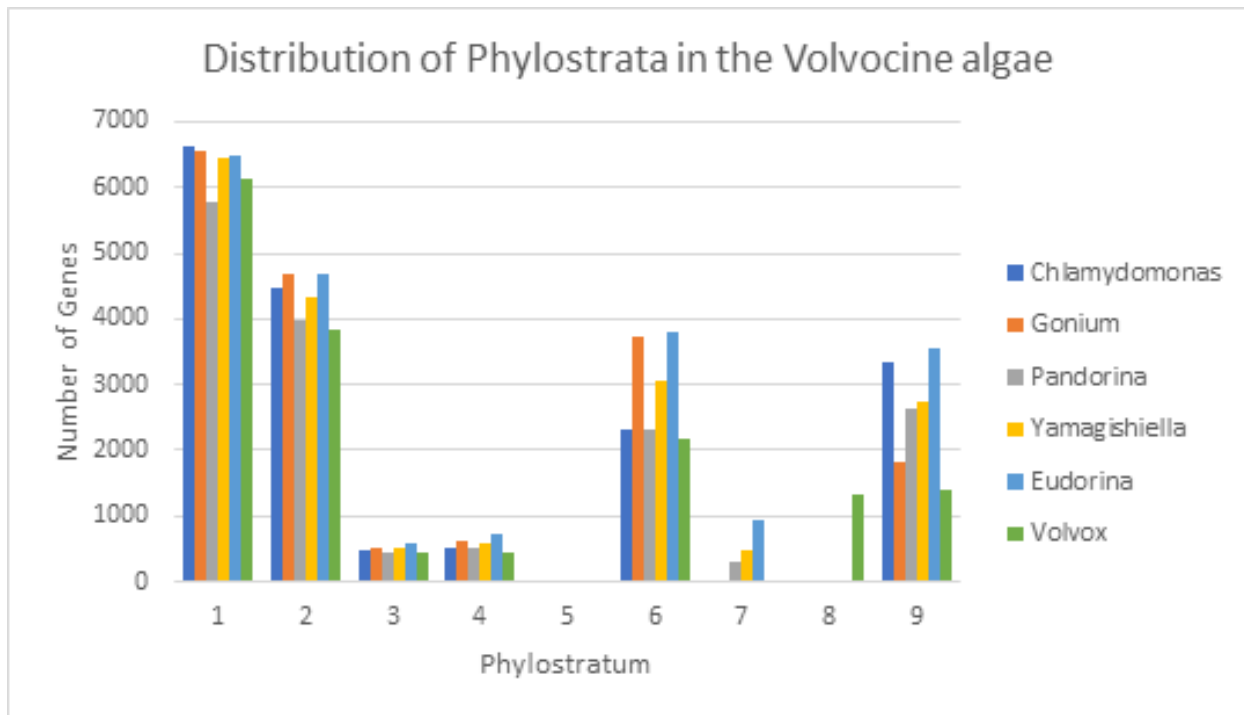


Figure 2.8. Number of genes per phylostratum (PS) for six volvocine species.

PS-1 corresponds to cellular organisms (least specific) and PS-9 corresponds to species (most specific).

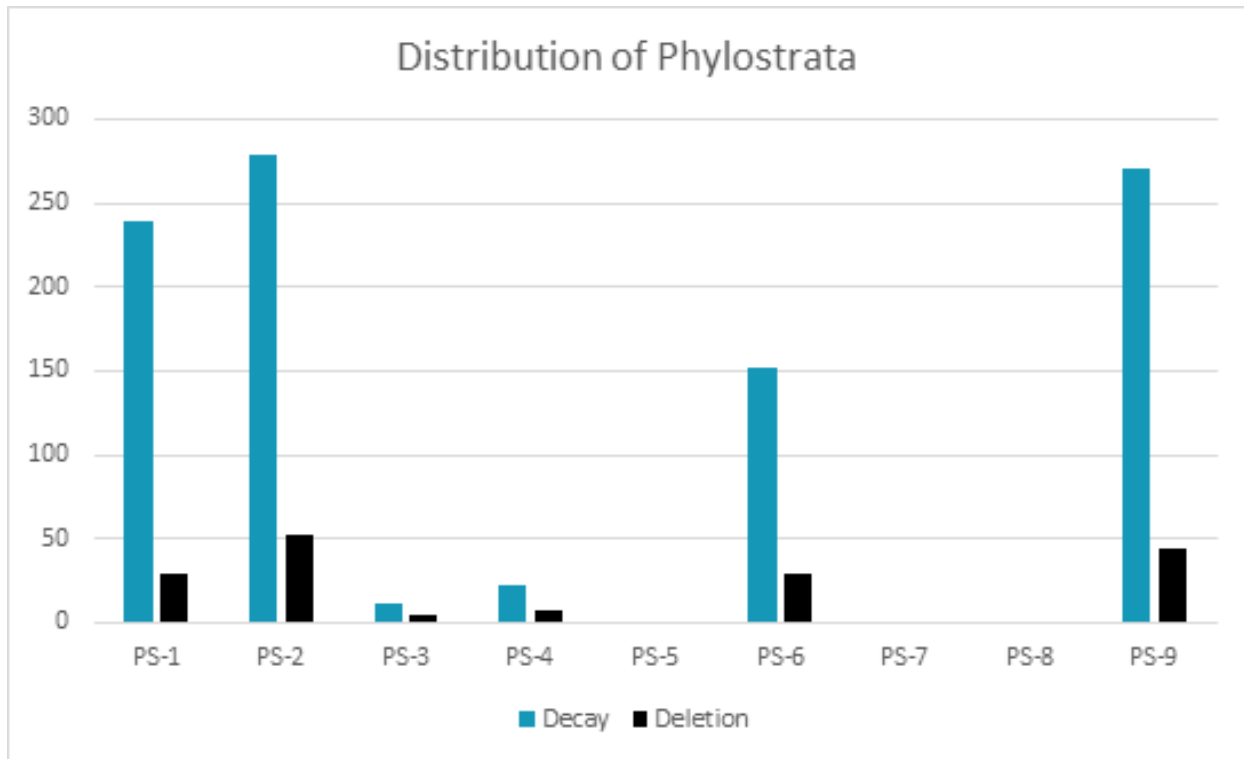


Figure 2.9. Number of *Chlamydomonas* orthologs of genes that underwent decay or deletion in multicellular volvocines per phylostratum (PS).

PS-1 corresponds to cellular organisms (least specific) and PS-9 corresponds to species (most specific).

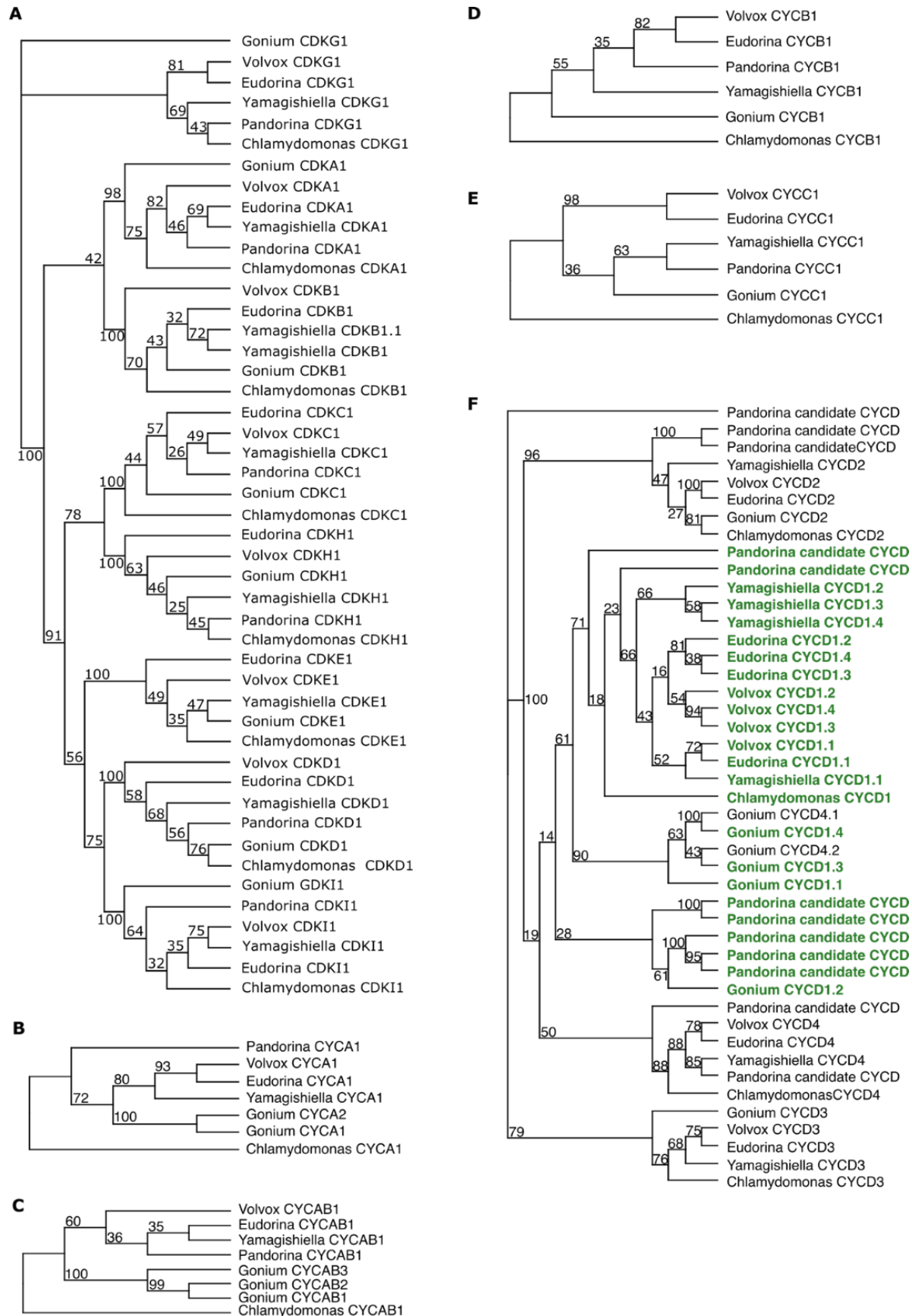


Figure 2.10. Phylogenetic analysis of cell cycle genes.

A) Phylogeny of cyclin dependent kinases (CDKs). B-F) Phylogeny of cyclins (cyc). CycD1s are in green.



Figure 2.11. Phylogenetic analysis of matrix metalloprotease (MMP) genes.

Chlamydomonas genes are in blue, *Gonium* in magenta, *Yamagishiella* in salmon, *Eudorina* in orange, and *Volvox* in yellow. Values on the nodes represent percentage bootstrap support.

Figure 2.12. Phylogenetic analysis of pherophorin genes.

Chlamydomonas genes are in blue, *Gonium* in magenta, *Yamagishiella* in salmon, *Eudorina* in orange, and *Volvox* in yellow. Values on the nodes represent percentage bootstrap support.

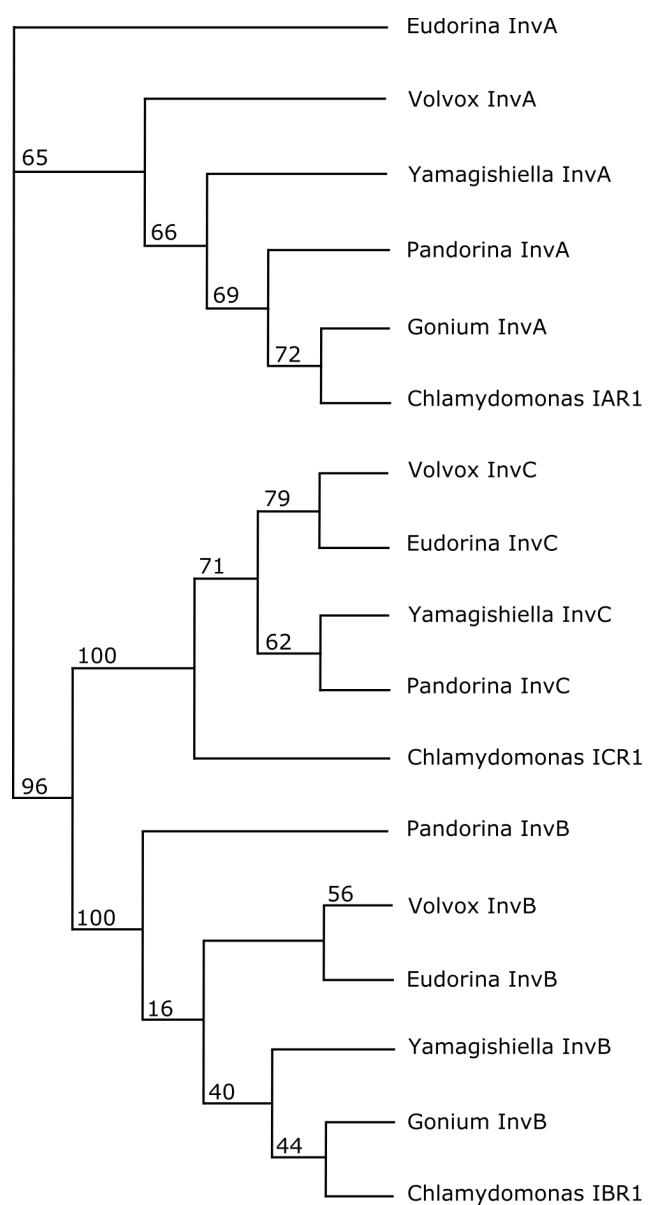


Figure 2.13. Phylogenetic analysis of *inv* genes.

Values on the nodes represent percentage bootstrap support.

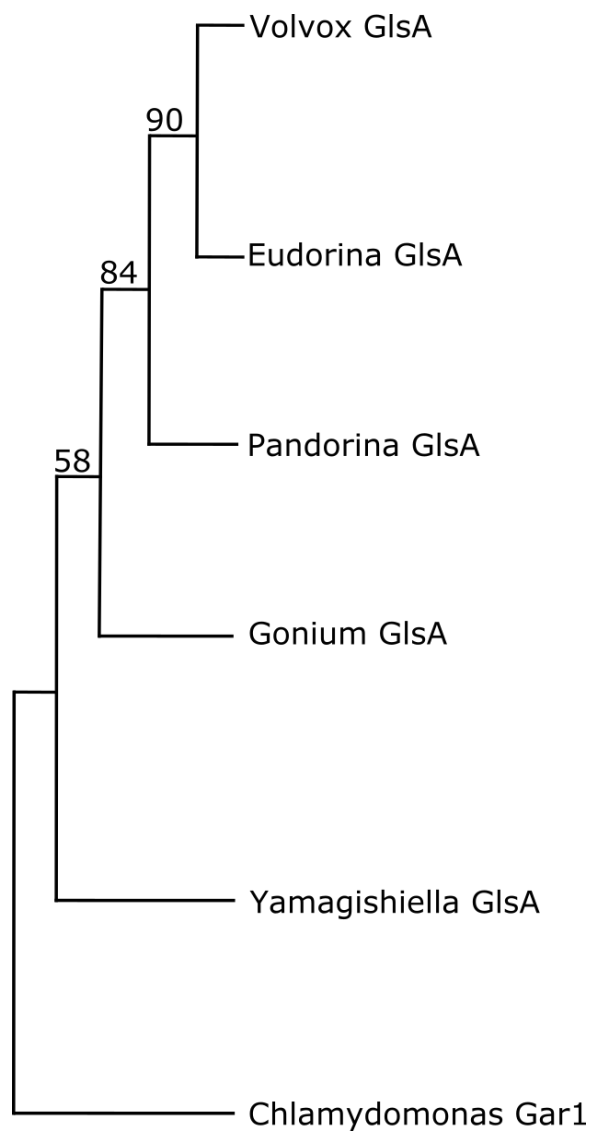


Figure 2.14. Phylogenetic analysis of *glisA* genes.

Values on the nodes represent percentage bootstrap support.

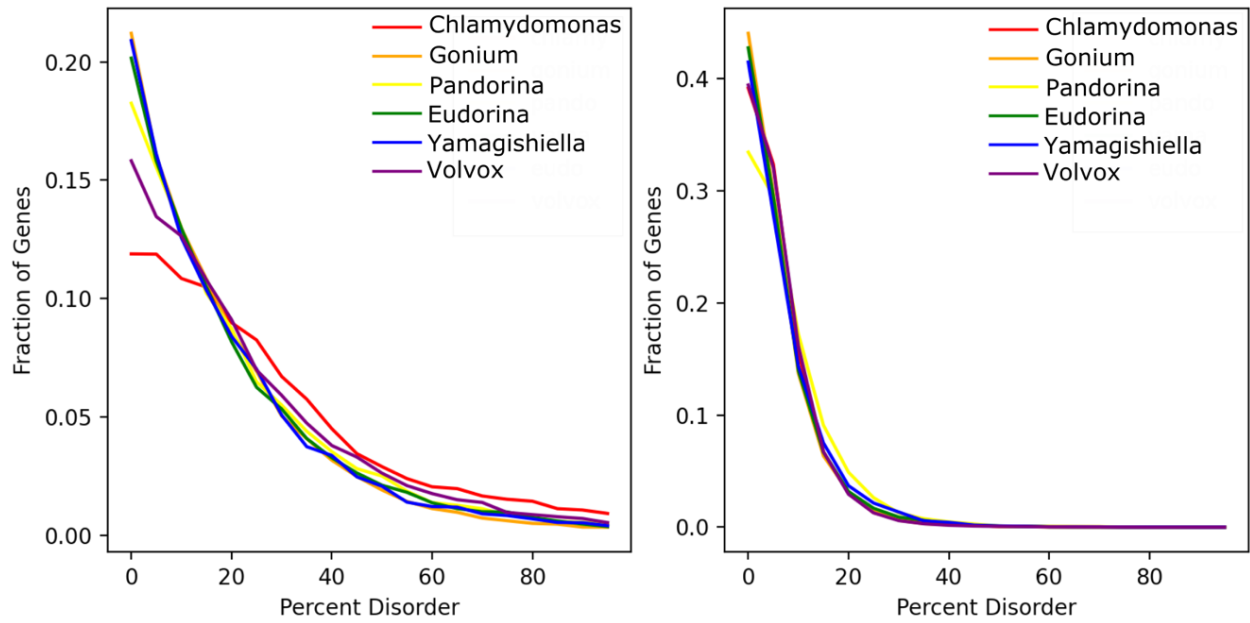


Figure 2.15. Percentage frequency distribution of disordered proteins in the volvocine algae.

A) Percentage distribution of genes encoding protein-binding proteins with an increasing amount of disorder. B) Percentage distribution of genes encoding proteins with an increasing amount of disorder.

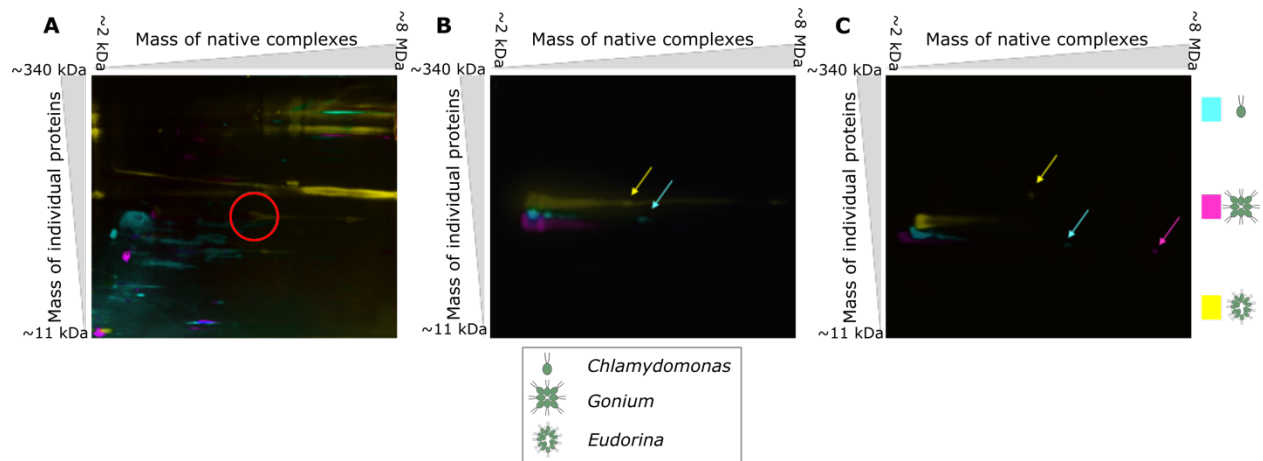


Figure 2.16. Protein-protein interactions differ among volvocine species.

Blue Native (BN) PAGE followed by SDS-PAGE of lysates from *Chlamydomonas* (pseudocolored cyan), *Gonium* (pseudocolored magenta), and *Eudorina* (pseudocolored yellow).

A) Overlay of silver stains. RuBisCO signal is circled in red. B) Overlay of western blot signals for α -tubulin. Tubulin signals for species-specific complexes are indicated by color coded arrows. C) Overlay of western blot signals for β -actin. Actin signals from unique complexes are indicated by color coded arrows.

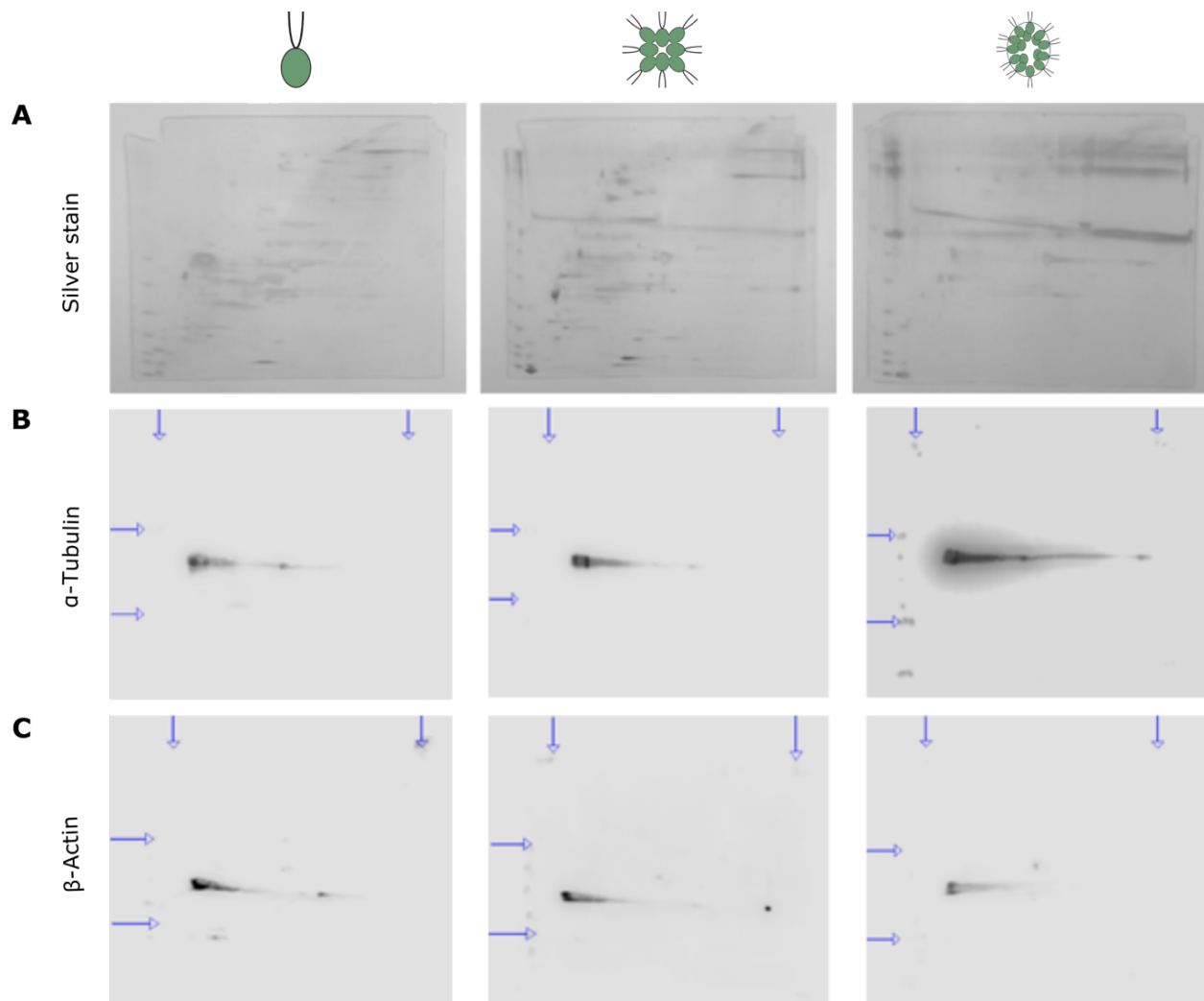


Figure 2.17. Raw images for A) Silver stained gels; B) α -tubulin and C) β -actin blots for *Chlamydomonas*, *Gonium* and *Eudorina*.

Vertical blue arrows indicate boundaries of the native gel slice. Horizontal blue arrows indicate 80 kDa (top) and 25 kDa (bottom) bands on the SDS-PAGE standards ladder. Silver stains are notched at the top corners to indicate the boundaries of the native gel slice.

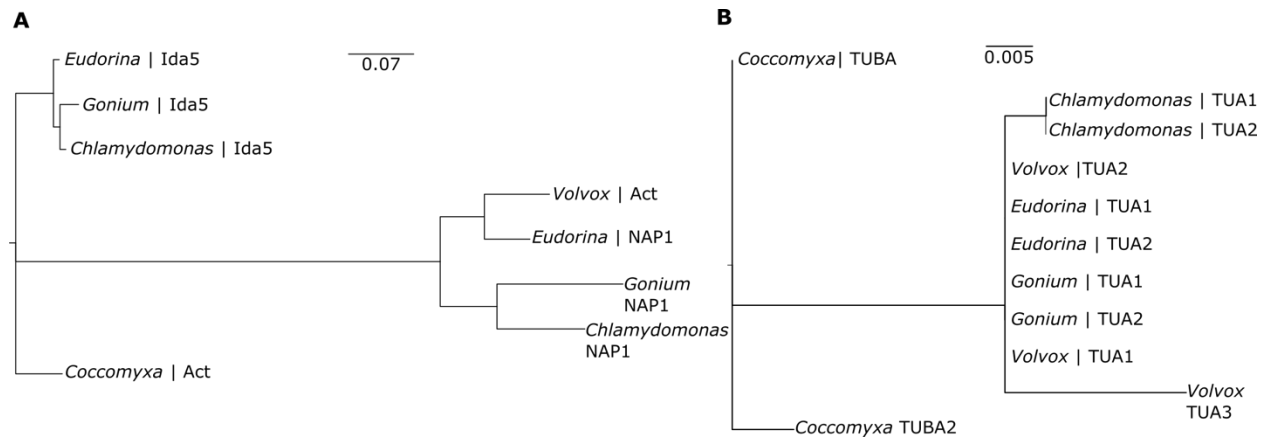


Figure 2.18. A) β -actin and B) α -tubulin phylogenetic trees using *Coccomyxa subellipsoidea* (Chlorophyta) as an outgroup.

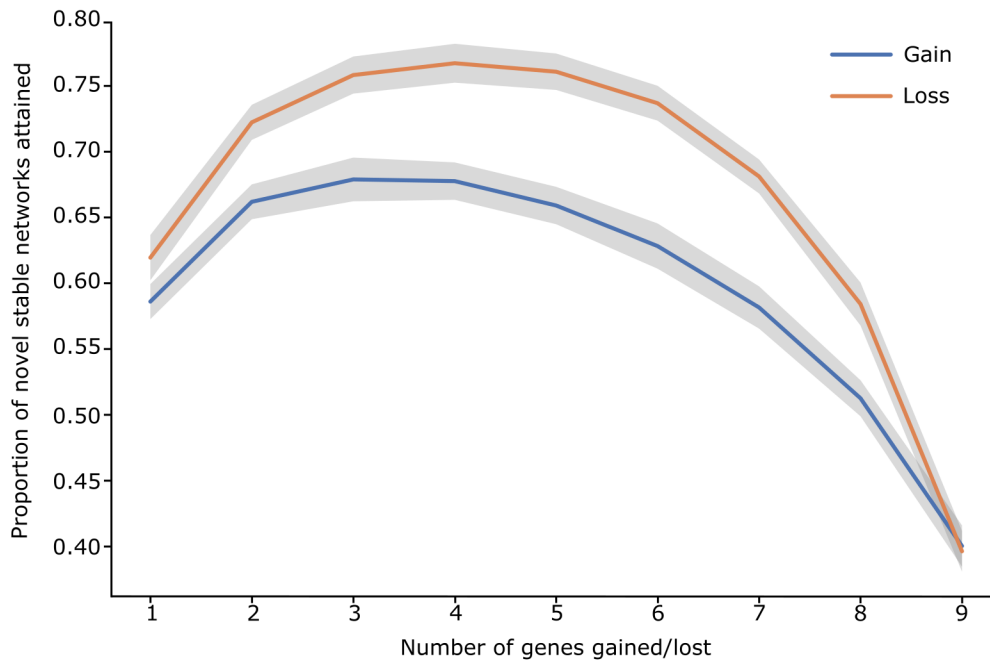


Figure 2.19. Gene loss in a fully connected network ($c=1$) of genes ($N=10$) yields a higher proportion of novel stable network states than gene gain.

Wagner model simulates loss (orange) and duplication (blue) of k genes. Grey shading represents variance.

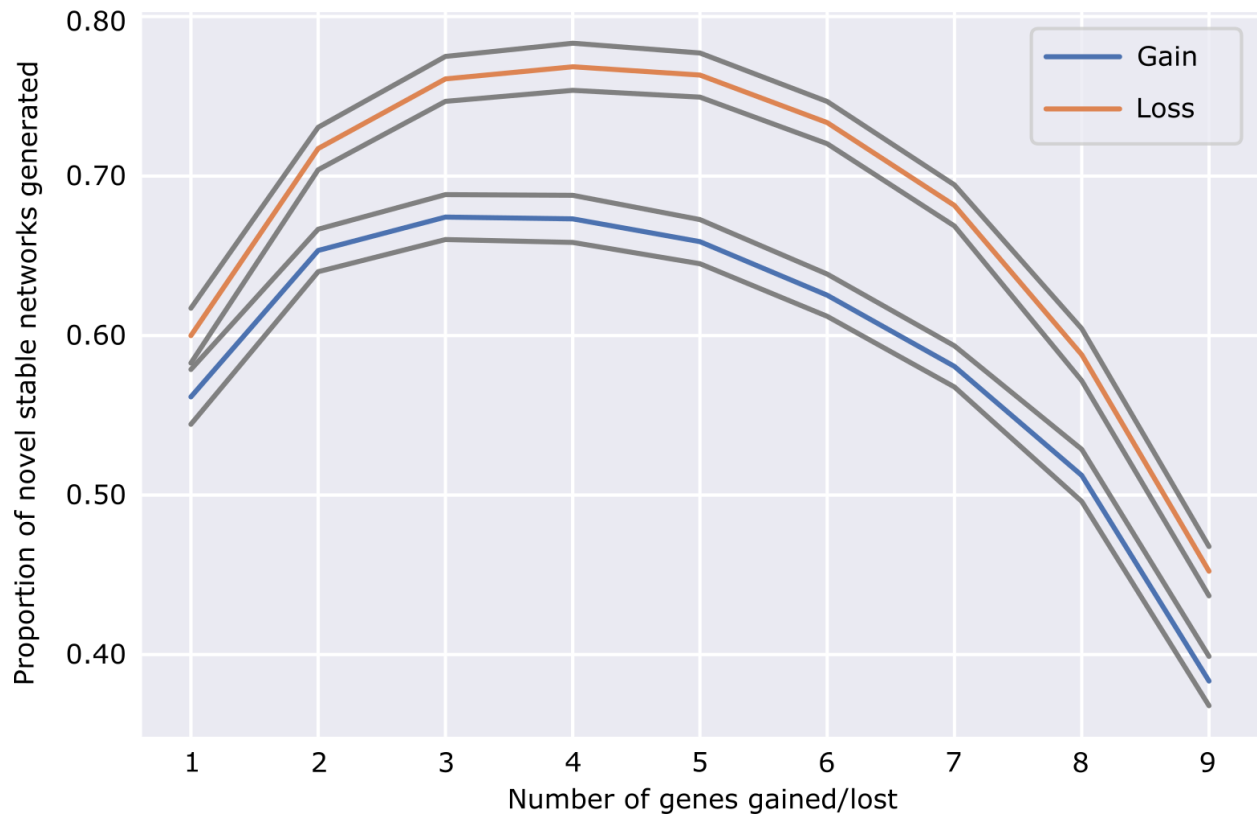


Figure 2.20. Wagner model simulates loss (orange) and duplication (blue) of k genes out of a network of $N=10$ genes over different network connectivities.

Gene loss yields a higher proportion of novel stable network states than gene gain regardless of changes to network connectivity. Grey lines represent variance.

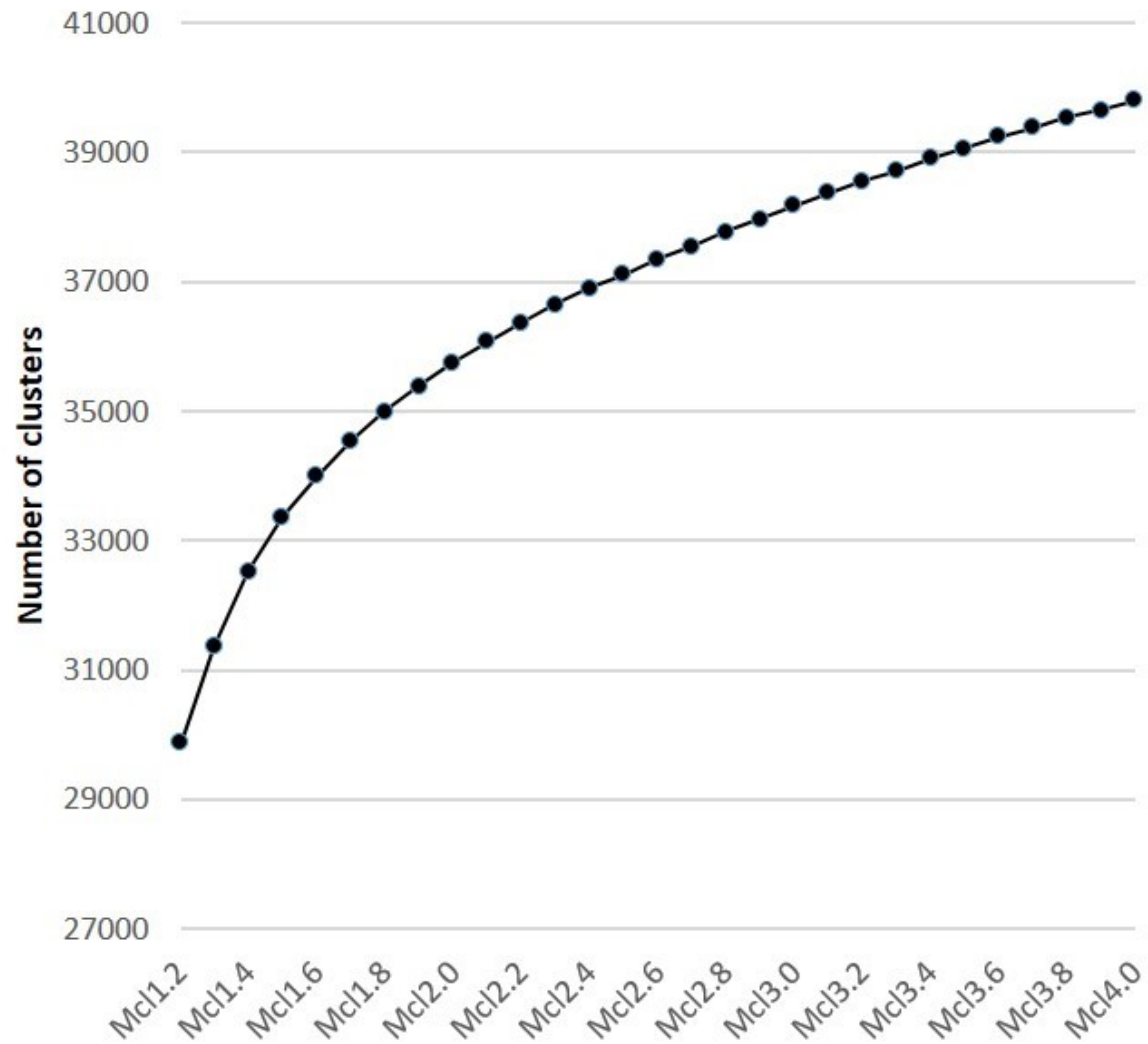


Figure 2.21. Number of orthoMCL clusters for a range of inflation values.

Singltons not included.

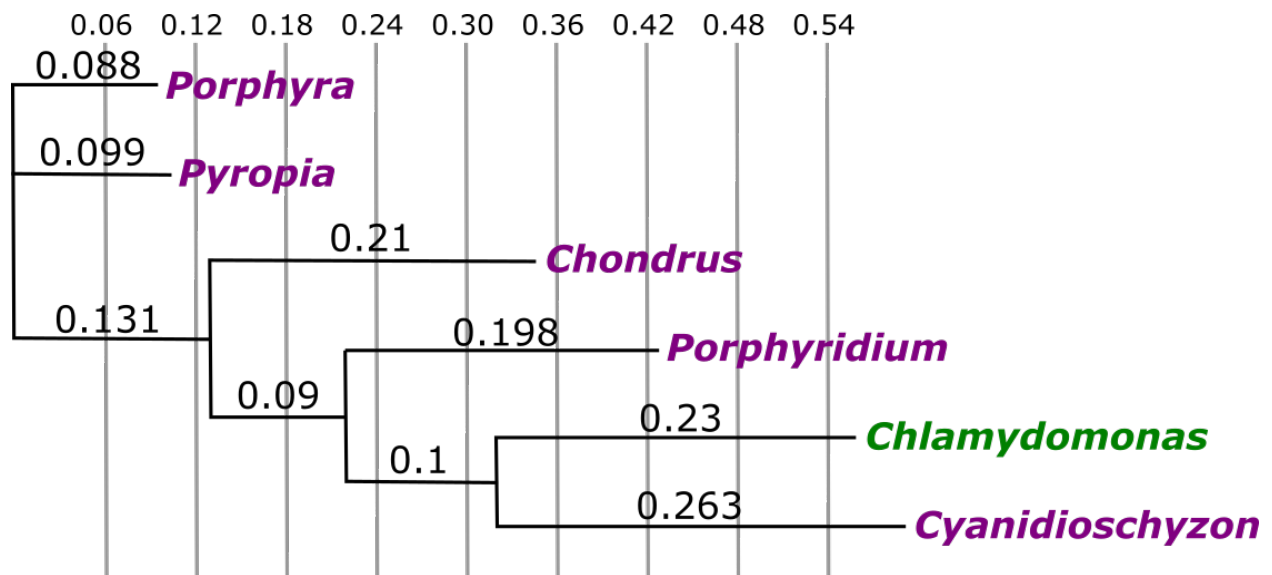


Figure 2.22. Genome-wide phylogeny of red algae (magenta) compared to a chlorophycean outgroup (green) inferred by PosiGene.

Branch values indicate distance from the nearest node.

Species	Complete (%)	Single (%)	Duplicated (%)	Fragmented (%)	Missing (%)	Total
<i>Chlamydomonas</i>	1503 (99)	1411 (92.9)	92 (6.1)	5 (0.3)	11 (0.7)	1519
<i>Tettrabaena</i>	967 (63.6)	936 (61.6)	31 (2.0)	277 (18.2)	275 (18.2)	1519
<i>Gonium</i>	1415 (93.2)	1400 (92.2)	15 (1)	69 (4.5)	35 (2.3)	1519
<i>Pandorina</i>	1193 (78.5)	1115 (73.4)	78 (5.1)	46 (3)	280 (18.5)	1519
<i>Yamagishiella</i>	1422 (93.6)	992 (65.3)	430 (28.3)	58 (3.8)	39 (2.6)	1519
<i>Eudorina</i>	1340 (88.2)	838 (55.2)	502 (33)	106 (7)	73 (4.8)	1519
<i>Volvox</i>	1464 (96.4)	1332 (87.7)	132 (8.7)	16 (1.1)	39 (2.5)	1519

Table 2.1. BUSCO completeness values of seven volvocine algae annotated genomes.

Completeness values describe how many BUSCOS are conserved between query genomes and the Chlorophyta database. BUSCOS are classified as Complete, Fragmented, and Missing.

Complete BUSCOS can be Single or Duplicated.

	<i>Chlamydomonas</i>	<i>Gonium</i>	<i>Pandorina</i>	<i>Yamagishiella</i> Minus	<i>Yamagishiella</i> Plus	<i>Eudorina</i> Plus	<i>Eudorina</i> Minus	<i>Volvox</i>
% G and C content	64.1	64.5	62.8	62.9	62.8	63.4	63.2	56
Protein coding Genes	17741	17948	15976	18180	18416	20744	22924	14247
Total transcripts	19526	17984	16542	30755	31705	32233	38492	16075
Ave Exons Per transcript	8.62	6.94	8.67	7.06	6.93	6.38	5.73	7.91
Average Exon length	260.99	211.04	215.8	205.93	205.37	207.78	211.58	254.39
Average Intron length	279.17	408.89	457.7	348.37	348.28	396.73	382.3	399.5
Average UTR5 length	203.5	232.57	187.64	235.97	237.61	235.23	233.72	231.69
Average UTR3 length	772.74	711.23	426.37	702.19	708.88	725.83	769.03	1105.09

Table 2.2. Genomic features of six volvocine algae species.

Present in	Number of orthologous groups	Number of Pfams
C	473	12
G	453	15
P	537	18
Y	317	11
E	411	17
V	226	23
CG	161	7
CP	50	4
CY	103	2
CE	91	1
CV	80	6
GP	78	2
GY	162	3
GE	120	5
GV	46	1
PY	98	0
PE	67	4
PV	33	0
YE	254	3
YV	96	1
EV	255	6
CGP	36	1
CGY	82	4
CGE	45	2
CGV	39	6
CPY	32	2
CPE	14	1
CPV	20	2
CYE	51	1
CYV	47	5
CEV	58	0
GPY	47	3
GPE	39	0
GPV	15	1
GYE	86	3
GYV	40	1
GEV	46	1
PYE	66	1

PYV	34	1
PEV	50	0
YEV	129	1
CGPY	68	7
CGPE	44	3
CGPV	26	12
CGYE	132	5
CGYV	72	12
CGEV	79	9
CPYE	84	1
CPYV	35	12
CPEV	38	6
CYEV	164	8
GPYE	88	3
GPYV	33	2
GPEV	37	2
GYEV	122	2
PYEV	124	3
CGPYE	354	31
CGPYV	145	68
CGPEV	120	43
CGYEV	1352	91
CPYEV	442	52
GPYEV	212	10
CGPYEV	6297	1184
Total	15155	1743

Table 2.3. Shared orthologous group and pfam count for six volvocine algae species.

Species names are encoded as follows: C=*Chlamydomonas*, G=*Gonium*, P=*Pandorina*,

Y=*Yamagishiella*, E=*Eudorina*, V=*Volvox*

Species	Actin	mtATP-A	mtATP-B	Flagellar Dynein heavy chain	Chloroplast Ferridoxin	α -Tubulin	β -Tubulin	Flagellar Dynein Heavy Chain
<i>Chlamydomonas</i>	2	2	1	15	1	2	2	15
<i>Gonium</i>	2	3	1	15	1	2	2	15
<i>Pandorina</i>	2	2	1	22	1	3	2	22
<i>Yamagishiella</i>	2	2	1	18	1	2	2	18
<i>Eudorina</i>	2	2	1	16	1	2	2	16
<i>Volvox</i>	2	3	1	15	2	2	2	15

Table 2.4. Conserved gene counts for six volvocine species.

Species	H1
<i>Chlamydomonas</i>	3
<i>Gonium</i>	3
<i>Pandorina</i>	2
<i>Yamagishiella</i>	3
<i>Eudorina</i>	2
<i>Volvox</i>	4

Table 2.5. Histone H1 gene counts for volvocine algae.

Histone	Species	Copy number	Tail var. number
H2A	<i>Chlamydomonas</i>	31	5
	<i>Gonium</i>	34	5
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	19	3
	<i>Eudorina</i>	19	4
	<i>Volvox</i>	15	3
H2B	<i>Chlamydomonas</i>	29	9
	<i>Gonium</i>	35	12
	<i>Pandorina</i>	32	9
	<i>Yamagishiella</i>	20	13
	<i>Eudorina</i>	18	14
	<i>Volvox</i>	15	12
H3	<i>Chlamydomonas</i>	35	4
	<i>Gonium</i>	30	24
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	21	6
	<i>Eudorina</i>	23	4
	<i>Volvox</i>	13	3
H4	<i>Chlamydomonas</i>	32	2
	<i>Gonium</i>	34	3
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	27	3
	<i>Eudorina</i>	21	2
	<i>Volvox</i>	14	2

Table 2.6. Histone family copy number and tail variant number of six volvocine algae.

Species	H1	H2A	H2B	H3	H4
<i>Porphyra</i>	1	6	4	74	2
<i>Pyropia</i>	1	3	1	22	1
<i>Chondrus</i>	1	3	2	21	3
<i>Porphyridium</i>	1	3	2	3	2
<i>Cyanidioschyzon</i>	1	3	2	2	2

Table 2.7. Histone gene counts for Rhodophytes.

Absence in	Total	Decay	Deletion
GPYEV	1017	880	137
GPYE	0	0	0
GPYV	8	6	2
GPEV	2	1	1
GYEV	5	5	0
PYEV	30	26	4
GPY	0	0	0
GPE	0	0	0
GPV	2	2	0
GYE	1	1	0
GYV	0	0	0
GEV	1	1	0
PYE	1	1	0
PYV	8	5	3
PEV	6	4	2
YEV	8	6	2
GP	1	1	0
GY	0	0	0
PY	0	0	0
PE	2	1	1
PV	6	6	0
GE	0	0	0
YE	0	0	0
GV	2	1	1
EV	11	11	0
YV	5	3	2
G	3	3	0
P	4	3	1
Y	3	2	1
E	1	1	0
V	14	5	9

Table 2.8. Distribution and type of gene losses in multicellular volvocine algae.

Species names are encoded as follows: G=*Gonium*, P=*Pandorina*, Y=*Yamagishiella*,
E=*Eudorina*, V=*Volvox*. All values have a P-value<0.01 and reject uniform distribution of decay
and deletion events.

Species	MMP genes	Pherophorin genes
<i>Chlamydomonas</i>	34	41
<i>Gonium</i>	22	21
<i>Pandorina</i>	42	38
<i>Yamagishiella</i>	38	32
<i>Eudorina</i>	63	69
<i>Volvox</i>	58	68

Table 2.9. Matrix metalloprotease (MMP) and pherophorin gene counts for six volvocine species.

Chapter 3 - Gene loss during independent gains of cellular differentiation in the volvocine algae

Berenice Jiménez-Marín and Bradley J.S.C. Olson

Submitted at *Front. Cell and Dev Biol.*

Author contributions: BJM: Conceptualization, methodology, software, formal analysis, investigation, writing – original draft, visualization. BJSCO: conceptualization, writing – review & editing, supervision, project administration, funding acquisition.

Abstract

The evolution of development is intimately tied to the evolution of multicellularity. The integration of each cell into an organized individual through cell specialization is thought to stabilize multicellularity; hence, germ-soma differentiation has evolved repeatedly from undifferentiated ancestors. The volvocine algae are an outstanding model system for addressing questions at the juncture of multicellularity, development, and complexity. In this work, the evolutionary dynamics of seven volvocine genomes, including independently-differentiated *Astrephomene gubernaculifera* and *Volvox carteri* were examined. Our results show that gene loss events likely occurred before the lineages of *Astrephomene* and *Volvox* diverged. Moreover, convergent losses of histone H2A, H2B, H3 and H4 copies in *Astrephomene* and *Volvox* are observed. Finally, an expanded repertoire of cyclin D genes that is conserved across other multicellular volvocines is also found in *Astrephomene*, suggesting that shared gene losses and co-option of key genes could have played a role in facilitating convergent increases to developmental complexity across multicellular species.

Introduction

The evolution of multicellularity is a major transition that is intimately linked to the evolution of development (Leigh et al., 1995; Grosberg and Strathmann, 2007; Knoll, 2011). Both multicellularity and germ-soma differentiation have repeatedly evolved from unicellular/undifferentiated ancestors across both prokaryotic and eukaryotic lineages (Grosberg and Strathmann, 2007; Herron et al., 2013; Ruiz-Trillo and Nedelcu, 2015). Hence, the convergence of disparate lineages onto multicellular states with cell differentiation has been proposed to be a consequence of either strategy's selective advantages (Michod et al., 2004; Ratcliff et al., 2013; Fisher et al., 2016; Herron et al., 2019), and germ-soma division of labor is

thought to play a paramount role, though not an indispensable one, in stabilizing multicellularity (Michod, 2007). Multicellular evolution and cell differentiation are associated with dramatic increases to developmental complexity of organisms, to the extent that the number of cell types in an organism has been proposed as a measure of complexity (McShea, 1996). Given that changes to developmental complexity are associated with impactful phenotypic innovation, it has been generally expected that expansions to the genetic repertoire underlie increases to developmental complexity. Our understanding the evolution of developmental complexity at the molecular level is, however, very limited due to long divergence times between multicellular models and their closest unicellular relatives, loss of extant species representing ‘intermediate’ states of complexity, and limited toolkits for experimental work (Yamashita et al., 2021; Jimenez-Marín and Olson, 2022).

The volvocine green algae have proven powerful for addressing questions on the evolution of multicellularity, cell differentiation, and overall developmental complexity. The volvocines comprise extant member species ranging from unicellular to multicellular undifferentiated and to differentiated multicellular (Kirk, 2005; Coleman, 2012; Yamashita et al., 2016; Hanschen et al., 2018a; Lindsey et al., 2021). Multicellular and differentiated-state traits within the volvocine lineage evolved recently (~200 MYA) (Herron et al., 2009), and several member species have been sequenced (Merchant et al., 2007; Prochnik et al., 2010; Hanschen et al., 2016; Hamaji et al., 2018; Jiménez-Marín et al., 2021), revealing strong similarities at the genomic level between species, as well as conservation of synteny between gene groups (Prochnik et al., 2010; Hanschen et al., 2016; Jiménez-Marín et al., 2021). Surprisingly, recent findings suggest that increased developmental complexity in the volvocines coincides with a trend for gene loss (Jiménez-Marín et al., 2021). Bursts of gene loss common to multicellular

species have been proposed to provide opportunities for novel protein-protein interactions to occur (Jiménez-Marín et al., 2021), suggesting developmental complexity need not emerge from substantial genetic innovation (Jimenez-Marín and Olson, 2022).

Several multicellular traits have evolved independently multiple times across the volvocine lineage (Herron et al., 2010; Hanschen et al., 2018a; Lindsey et al., 2021; Jimenez-Marín and Olson, 2022). Differentiated multicellularity has only occurred, however, in two of the three Families that encompass volvocine lineage, Volvocaceae and Goniaceae.

Differentiation appears to have evolved across the Volvocaceae more than once (Herron and Michod, 2008; Grochau-Wright et al., 2017; Lindsey et al., 2021). In *Volvox carteri* (Volvocaceae), a vegetative spheroid composed of thousands of flagellated somatic cells imbedded in extracellular matrix (ECM) that appears to maintain the integrity of the spheroid (Hallmann and Kirk, 2000) holds a small number (~16) of large, unflagellated gonidial cells. Conversely, in *Astrephomene gubernaculifera* (Goniaceae), a spheroid of up to 60 flagellated cells and up to 4 somatic cells at its posterior end (Yamashita et al., 2016). *Volvox* and *Astrephomene* are not only phenotypically distinct; their embryogenesis is markedly different (Yamashita et al., 2016), and it has been proposed that the genetic underpinnings of their developmental program are different as well (Yamashita et al., 2021).

The recent sequencing and assembly of the *Astrephomene* genome allows for addressing questions on convergent evolution of developmental complexity at the molecular level across 7 volvocine species. Since *Astrephomene* appears to have converged with *Volvox* at greater developmental complexity phenotypically but not necessarily at the genetic co-option level, we hypothesize that gene loss bursts might have shaped the evolution of *Astrephomene*'s increased developmental complexity as they did in other volvocine species. Given that gene loss is found

as developmental complexity increases for members of Volvocaceae, *Astrephomene* is an ideal independent test of whether gene loss represents a common trend, or if a burst of genetic innovation is instead observed.

Results

The *Astrephomene* genome is similar to that of other volvocines

The *Astrephomene* genome was recently sequenced, yielding valuable insights into how co-option of different genes, even in a strongly shared genomic context, may result in converging phenotypes and developmental programs (Yamashita et al., 2021). As means to complement results on the shared genomic context between *Astrephomene* and other volvocines, comparisons between the genomes of *Astrephomene* and another six volvocine algae spanning from unicellular to multicellular differentiated (*Chlamydomonas reinhardtii* - unicellular, *Gonium pectorale* – multicellular undifferentiated, Goniaceae, *Pandorina morum* – multicellular undifferentiated, spheroidal, Volvocaceae, *Yamagishiella unicocca*– multicellular undifferentiated, spheroidal, Volvocaceae, *Eudorina* sp. – multicellular undifferentiated, spheroidal, Volvocaceae, and *Volvox carteri* – multicellular differentiated, spheroidal, Volvocaceae) were made. Orthologous group and protein family domain (Pfam) catalogs for the analyzed species and their chlorophyte outgroups were compiled, and a genome-wide reconstruction of volvocine phylogeny was generated based on orthologous groups, which is in close agreement with recent results suggesting that *Astrephomene* forms a sister clade to Volvocaceae rather than falling within Goniaceae (Lindsey et al., 2021) (Figure 3.1A). Overall, 15,664 orthologous groups were identified; of these, 5,421 (34.6%%) are shared among the seven genomes, and 2,485 are species specific (15.7%, Figure 3.1B, Table 3.1). Analysis of the evolutionary trajectory of Pfams (El-Gebali et al., 2019) in the seven species shows 1,981 Pfam

domains. Of these, 1,296 (65.4%) are shared in all seven species, and 97 (4.9%) are species specific (Figure 3.1C, Table 3.1).

Contraction of conserved genetic elements is shared across the volvocine lineage

Rates of evolution of 9,329 orthologous groups and 1,569 Pfam domains for the seven species were analyzed by normalization to the distance from their Last Common Ancestor (LCA) as previously reported (Jiménez-Marín et al., 2021). This analysis shows that 536 (5.7%) orthologous groups have undergone significant changes, with a small bias towards gene loss (242 expanding orthologous groups, 294 contracting, Figure 3.2A outliers, shaded grey area). Likewise, while most Pfam domains were conserved, significantly expanding domains were less abundant than their contracting counterparts (78 and 121, respectively, Figure 3.2B outliers, shaded grey area).

Analysis of all identified transcription factor (TF) and protein kinase (PK) families in the seven volvocine algae (46 and 73 respectively) revealed that 5 families of TFs (10.9%) underwent significant contraction, while none expanded significantly (Figure 3.2C, Figure 3.3). Similarly, 10 families of PKs (13.7%) underwent significant change, where 1 of the PK families expanded and 9 contracted (Figure 3.2D, Figure 3.3). The trends of contraction of conserved elements across the volvocine gradient of complexity thus appears to be common to the clade itself. Notably, Pfam domain counts per gene did not show significant variation between the seven species (Figure 3.4), and contraction trends were not observed in highly conserved genes Actin, β -Tubulin, α -Tubulin, chloroplast ferredoxin, mtATP-synthase α , mtATP-synthase β , and dynein heavy chain (Table 3.2).

To address if there might be differences in regulation of gene expression mechanics for *Astrephomene*, histone genes were analyzed for their evolutionary dynamics. All families of histones, except linker H1, were reduced in number as developmental complexity increases in the volvocine algae (Figure 3.2E, Tables 3.3 and 3.4). Analysis of histone N-termini methylation site variation for H2A, H2B, H3, and H4 shows that copy number reduction has occurred to varying degrees in these four histones (Figure 3.2F, Table 3.4). Interestingly, *Astrephomene* has histone copy numbers that are closer to those of *Eudorina* and *Volvox* than to *Pandorina* or *Gonium* (Figure 3.2F). Histone N-terminal variant diversity in *Astrephomene* is not significantly different from that of other species, suggesting the possibility that convergent histone copy number loss relates to increases in developmental complexity through altering gene expression. Moreover, *Gonium*'s expanded H3 tail diversity might be a species-specific phenomenon.

To evaluate the impact of positive selection on *Astrephomene*'s lineage and contrast it to volvocacean increases to developmental complexity, orthologous groups of genes with representatives in all seven species were subjected to dN/dS analysis. Using *Chlamydomonas* as the reference, 77, 48, 51, 4, 19 and 3 genes exhibited positive selection in *Gonium*, *Astrephomene*, *Pandorina*, *Yamagishiella*, *Eudorina*, and *Volvox*, respectively (Figure 3.2G). Using *Volvox* as the reference, there were 372, 82, 55, 53, 4 and 17 genes with a significant dN/dS value in *Chlamydomonas*, *Gonium*, *Astrephomene*, *Pandorina*, *Yamagishiella*, and *Eudorina*, respectively (Figure 3.2G). These results suggest that positive selection has impacted roughly the same number of genes as *Pandorina*, and fall within the trends observed for other multicellular volvocines.

Gene loss events likely occurred at the ancestor of Goniaceae+Volvocaceae

The genomes of many volvocines are highly syntenic (Prochnik et al., 2010; Hanschen et al., 2016; Jiménez-Marín et al., 2021). Synteny of the genetic neighborhood of *rlsI* has been shown for *Astrephomene* (Yamashita et al., 2021), outlining the possibility that *Astrephomene*'s genome is syntenic with that of other volvocines. As means to test for the presence of syntenic blocks and evaluate whether signatures of gene loss observed in other volvocines are present on *Astrephomene*, 294 contracting orthologous groups were examined at the sequence level (Figure 3.2A).

The contracting groups contain 1,343 genes assembled into *Chlamydomonas* chromosomes, of which 1,293 (96.2%) in 255 (86.7%) orthologous groups had loci that were adequately assembled in all genomes for further analysis. This analysis demonstrated 43 of 1,293 genes in contracting orthologous groups are retained (preservation of orthologs across all species), while 1,250 loci exhibit gene loss (Figure 3.5A, B). The mechanisms of gene loss were then examined by tBLASTx to determine if remnants of the lost genes (gene decay) were present at the locus or if remnants of the gene were completely lost (gene deletion). Our results show gene decay is more abundant than deletion in all the colonial species; 72.7% of *Gonium*, 68% of *Astrephomene*, 66.8% of *Pandorina*, 74% of *Yamagishiella*, 74.6% of *Eudorina* and 53% of *Volvox* gene losses show decay signatures (Figure 3.5A).

Cross-species comparisons of retention, decay, and deletion events show that most candidate gene losses, 1,088 loci (84.1%), are due to decay (Figure 3.5B). Surprisingly, our analysis also suggests 1,082 (83.7%) loci are gene loss candidates in all six multicellular species. Of these, 957 (75.6%) occurred by gradual decay (Figure 3.5C, Table 3.5). In the case of deletions, only 162 loci (12.5%) were completely absent in all multicellular species, which could

either represent complete gene loss in the multicellular lineage or *Chlamydomonas*-specific gene gains. Thus, the volvocines exhibit bursts of gene loss, primarily by decay, during the transition to undifferentiated multicellularity at Goniaceae+Volvocaceae.

Expanded Cyclin D repertoire is conserved across Goniaceae and Volvocaceae

Several key genes underlying *Volvox* development have resulted from co-option events (Prochnik et al., 2010; Hanschen et al., 2016; Olson and Nedelcu, 2016; Grochau-Wright et al., 2017). In *Gonium* and sequenced Volvocaceae, Cyclin D1 genes (Hanschen et al., 2016; Jiménez-Marín et al., 2021), and the *Volvox* cell differentiation gene *regA* (Grochau-Wright et al., 2017) follow a duplication and divergence pattern that coincides with increases in complexity, while other developmentally relevant genes to *Volvox* such as other cyclins, *invA*, *B*, *C*, and *glsA*, appear to have been co-opted without undergoing expansion (Jiménez-Marín et al., 2021). However, it appears to be that different genes might have been co-opted for *Astrephomene*'s development (Yamashita et al., 2021), which likely resulted in the fundamental differences in embryogenesis between the two species. Given that the evolution of volvocine multicellularity is hypothesized to have occurred partly through expansion of cyclin D genes (Hanschen et al., 2016), phylogenetic analyses of Cyclins and Cyclin Dependent Kinases (CDKs) was used to address whether this co-option event could be common to Goniaceae+Volvocaceae. As observed for other multicellular volvocines (Hanschen et al., 2016; Jiménez-Marín et al., 2021), Cyclin D1 genes are expanded in *Astrephomene* (Figure 3.6F), whereas other cyclins and CDKs (Figure 3.6A-E) do not appear to have undergone conserved expansions. These results suggest that the expansion of cyclin Ds occurred before Goniaceae and Volvocaceae diverged, and perhaps played a role in the evolution of volvocine multicellularity.

Furthermore, these results highlight that neither *Astrephomene* nor *Volvox* required further expansion of key cell-cycle regulatory circuits to evolve cell differentiation.

Discussion

Despite the unique embryogenesis of *Astrephomene*, obvious similarities between volvocine genomes remain (Figure 3.1B, C). As per previous analysis (Jiménez-Marín et al., 2021), both orthologous groups and Pfam domains have more contracting than they do expanding members as within the volvocines (Figure 3.2A, B), even in the absence of significant variation in the number of Pfam domain counts per gene between the six species (Figure 3.4). Indeed, *Astrephomene* does not seem to have significant differences in either its orthologous group and Pfam repertoire when compared to the rest of the analyzed species (Figure 3.1B, C, Table 3.1).

While it is true that *Astrephomene* is absent from 4,184 (31.7%) of all orthologous groups shared by two or more species (Table 3.1), this number is comparable the number of orthologous groups that *Pandorina* is not in (4,207 groups, 31.9% of groups shared by two or more species) (Table 3.1). These results agree with previous work comparing *Astrephomene* to a smaller set of species (*Gonium* and *Volvox*) (Yamashita et al., 2021). The more complete survey of orthologous group diversity from comparing seven volvocine species in this work suggests that the ancestors of *Astrephomene* and *Pandorina* underwent gene loss to a similar extent. Interestingly, only 2,128 of the orthologous groups shared by at least two species are absent in both *Astrephomene* and *Pandorina* (Table 3.1). Whether this pattern is the result of a shared burst of loss at the ancestor of both lineages or convergence towards loss of certain families followed by independent losses remains to be elucidated.

Astrephomene appears to fall into the volvocine trend of losing genes mainly by decay as well. What is more, there are few *Astrephomene*-specific gene losses (Table 3.5); rather, losses in

conserved genomic regions are for the most part shared across all multicellular species (Figure 3.5C) or shared by *Astrephomene* and the Volvocaceae (Figure 3.5C). The commonality of gene loss across the multicellular volvocines in addition to the similar set of cell cycle genes (Figure 3.6) support the hypothesis that bursts of gene loss, key co-option events and minimal gene gains paved the evolutionary path to increased developmental complexity across the full lineage. The pronounced contraction of histone genes in *Astrephomene*, which is comparable to that of *Volvox* (Figure 3.2E, F), further supports this possibility. It will be interesting to evaluate if and how such copy number reductions impact the regulatory landscape throughout the development of both species, especially during cell differentiation.

Whether the same (cyclin Ds, Figure 3.6 F) or different (*regA*, MYB transcription factors (Yamashita et al., 2021)) genetic elements became co-opted during the instances of increased complexity typified by *Volvox* and *Astrephomene*, the underlying dynamics of loss that are shared between multicellular volvocines certainly appear to set the stage for a relatively straightforward path to increased developmental complexity. The interplay between dynamic processes at the molecular level and convergence at the metabolic and phenotypic level (Yamashita et al., 2016; Matt and Umen, 2018; Yamashita et al., 2021) is one of many layers, including that of gene regulation. The volvocine algae are now providing insight into how acquisition and loss of molecular players might alter the regulatory landscape in ways that promote the differential usage of conserved pathways and functional elements.

The patterns of cyclin co-option and gene loss across the multicellular volvocines analyzed in this work suggest that the ancestor of the Goniaceae+Volvocaceae lineage underwent genomic changes that set the stage for future developmental innovations – convergent or not. Under this scenario, changes to developmental complexity, such as the cell differentiation

programs in *Astrephomene* and *Volvox*, could be the product of ‘fine tuning’ of regulatory circuits through differential expression of genes belonging to key conserved pathways. Should this be the case, the regulatory networks of *Astrephomene* and *Volvox* should be topologically similar, even if the nodes (genes) forming convergent network structures are different. In line with this, the convergent losses of histone gene copies between *Astrephomene* and *Volvox* (Figure 2F) pose the exciting possibility that epigenetic regulation might play a pivotal role on the evolution of volvocine developmental programs. Altogether, our results indicate that biological innovation can and does coincide with gene loss events.

Methods

Strains and genomes

Chlamydomonas reinhardtii CC-503 MT⁺ genome version 5.5 (Merchant et al., 2007) (GCA_000002595.3), *Gonium pectorale* K3-F3-4 MT⁻ NIES-2863 (Hanschen et al., 2016) (GCA_001584585.1), and *Volvox carteri f. nagariensis*, Eve S1 (Prochnik et al., 2010) genome versions 1 (GCA_000143455.1) and 2.1 (available on Phytozome) (Prochnik et al., 2010; Goodstein et al., 2012), and *Astrephomene gubernaculifera* NIES 4017 (Yamashita et al., 2021) (BMAR01000001–BMAR01000206) were used for all analyses. Genome assemblies for *Yamagishiella* (*Y. unicocca* strains 2012-1026-YU-F2-6 MT⁺ NIES-3982, and 2012-1026-YU-F2-1 MT⁻ NIES-3983) and *Eudorina* (*Eudorina* sp. strains 2010-623-F1-E4 female NIES-3984 and 2010-623-F1-E2 male NIES-3985) were kindly provided by Dr. Hisayoshi Nozaki before publication (Hamaji et al., 2018). Chlorophyte outgroup species genomes have been published previously: *Chromochloris zongingiensis* (available on Phytozome) (Goodstein et al., 2012; Roth et al., 2017), *Chlorella variabilis* (Worden et al., 2009) (ADIC000000000), *Micromonas pusilla*

(Worden et al., 2009) (GCA_000151265.1), and *Ostreococcus lucimarinus* (Palenik et al., 2007) (GCA_000092065.1).

Identification of orthologous groups and Pfam domains

Orthologous groups of volvocine algae were determined using OrthoMCL 2.0.9 (Li et al., 2003). An inflation value of 1.9 was used for all analyses as determined previously (Jiménez-Marín et al., 2021). The diversity and abundance of known Pfam domains in the volvocine algae was identified using Pfam Scan (Finn et al., 2014) and database version 35. Pfam Scan and associated data collection Perl scripts available at: <http://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/> (El-Gebali et al., 2019).

Analysis of other genomic features

Transcription factors and protein kinases in *Astrephomene* were identified using the iTAK version 1.7 program and database (Zheng et al., 2016). The *Chlamydomonas* gene definitions on GenBank were used to identify histone H1 (Accession #XM_001696120) and quantified (Figure 2E second panel, Supplementary Table 3). Histones H2A, H2B, H3, H4 (Accession # L41841) genes in the *Astrephomene* were identified using the reciprocal best BLAST hit method with a e-value cutoff of 1e-5. H2A, H2B, H3 and H4 genes were further filtered using e-values of 1e-10 and quantified (Figure 2E panels 3-6, Figure 2F first panel, Supplementary Table 4). The first 30 amino acids from each histone were used for analysis of duplication and presence of predicted post-translational modification sites (lysine methylation) using Methysight (Biggar et al., 2020). N-tails per histone per species that had different predicted sites with threshold ≥ 0.5 were considered tail variants and quantified (Figure 2F second panel, Supplementary Table 4). Eight internal control genes, Actin, mitochondrial ATP-synthase α subunit (mATP-A), mitochondrial

ATP-synthase β subunit mATP-B, α -tubulin, β -tubulin, β -2 tubulin, flagellar dynein heavy chain, chloroplast ferridoxin were identified using OrthoMCL (Supplementary Table S2).

Identification of gene count trends

Rate of gene number change was calculated by applying a linear regression model on the relationship between gene counts per orthologous group per species and distance of said species from their Last Common Ancestor (LCA, Figure 1A) as previously reported (Jiménez-Marín et al., 2021).

Identification of positively selected genes and generation of whole-genome phylogenetic tree

Positively selected genes in *Astrephomene* were identified using the PosiGene pipeline (Sahm et al., 2017). Two independent PosiGene analyses were performed using *Chlamydomonas* and *Volvox* genome annotations as anchors. Chlorophytes *Chlorella variabilis*, *Micromonas pusilla*, *Ostreococcus lucimarinus* and *Chromochloris zofingiensis* were used as outgroups. A gene was considered to be positively selected if the dN/dS value (HA foreground omega) was greater than 1 and the Bonferroni-corrected p-value was less than 0.05. As a part of the workflow, PosiGene generates a phylogenetic tree from the isoform assignments and calculates the distance to LCA (Figure 1A).

Characterization of gene loss patterns

Orthologous groups showing significant contraction (Figure 2A) were selected for gene loss analysis based on whether they included *Chlamydomonas* genes using custom python scripts as previously reported (Jiménez-Marín et al., 2021). *Chlamydomonas* loci that were not assigned to chromosomes were not analyzed further. Orthologous groups that included *Chlamydomonas* loci with the aforementioned characteristics (target loci) served as anchors for obtention of genes up

and downstream, called the ‘syntenic neighborhood’. Orthologs for loci in the syntenic neighborhood, as well as target loci for *Gonium*, *Astrephomene*, *Pandorina*, *Yamagishiella*, *Eudorina* and *Volvox* were retrieved from their respective orthologous groups. Instances of no ortholog found for target loci were subjected to tBLASTx (Altschul et al., 1990) 2.2.18+ searches using tailored databases based on syntenic neighborhoods for each species. Non-informative (loss of syntenic neighborhoods or no BLAST results) genes were omitted from further analyses. tBLASTx hits were filtered for quality (e-value < 1e-5). Hits passing quality threshold were considered to be evidence of a gene remnant suggestive of loss via decay. Genes with no hits passing quality threshold and conserving synteny are candidate deleted genes. Decayed and deleted genes were assigned to bins corresponding to their pattern of absence (e.g. a gene lost by whichever mechanism only in *Gonium*, but present in the other clades; see Figure 3C and Supplementary Table 5), or by type of event, where decayed gene clusters consist of orthologs with at least one signature of decay in multicellular species, and deleted gene clusters are those where all losses happen via deletion (Figure 3B). Percentages of decay and deletion cases were calculated.

Identification and analysis of known co-opted genes and prediction of lineage specific genes

Cyclin and cyclin dependent kinase (CDK) genes in *Astrephomene* genomes were annotated as previously described (Prochnik et al., 2010; Hanschen et al., 2016) using BLASTp and tBLASTn.

Figures and tables

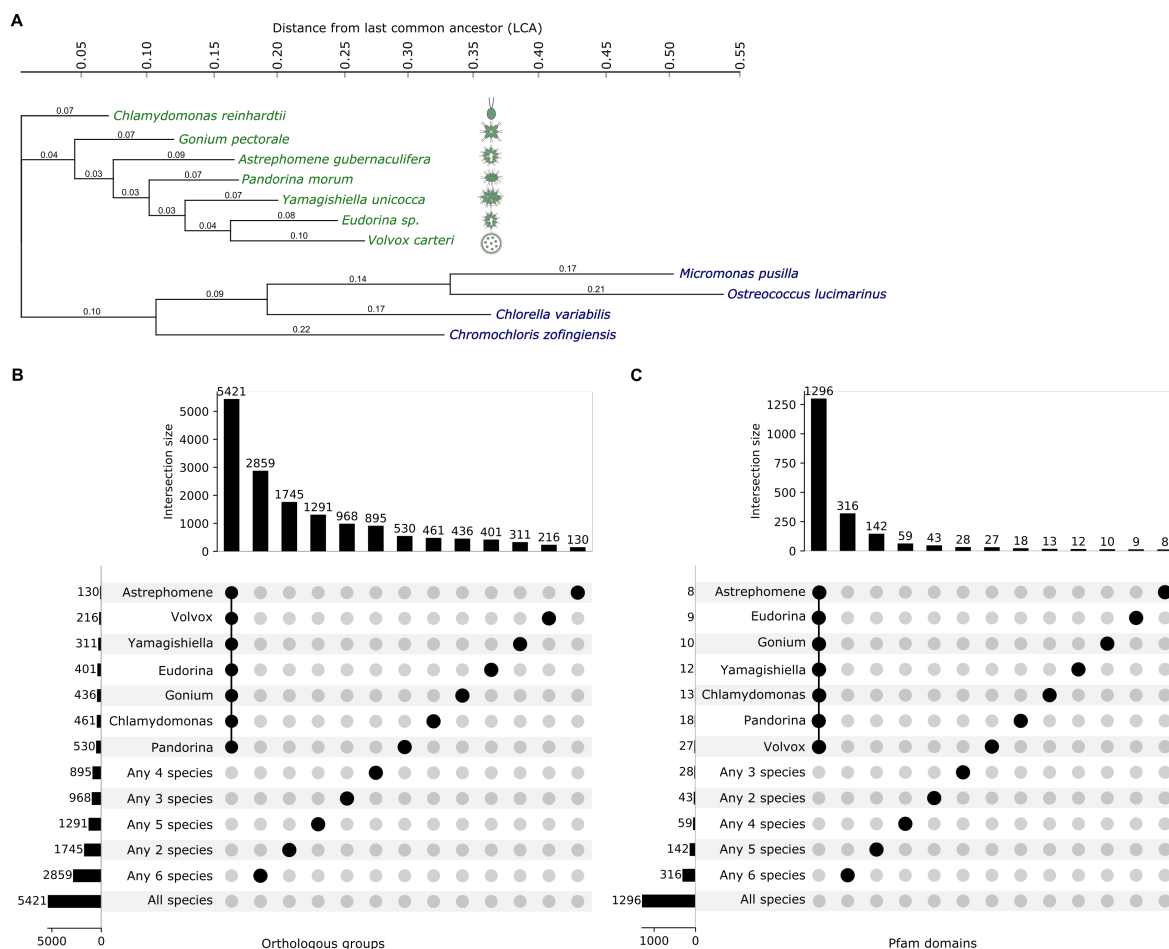


Figure 3.1. Volvocine genomes are similar.

A) Genome-wide phylogeny of the volvocine algae (green) compared to other chlorophyte outgroups (blue) inferred by PosiGene. Branch values indicate distance from the nearest node. Note *Chlamydomonas* is unicellular and closely related to colonial volvocines. B) Orthologous groups shared between the seven volvocine species. C), Pfam domains shared between seven volvocine species.

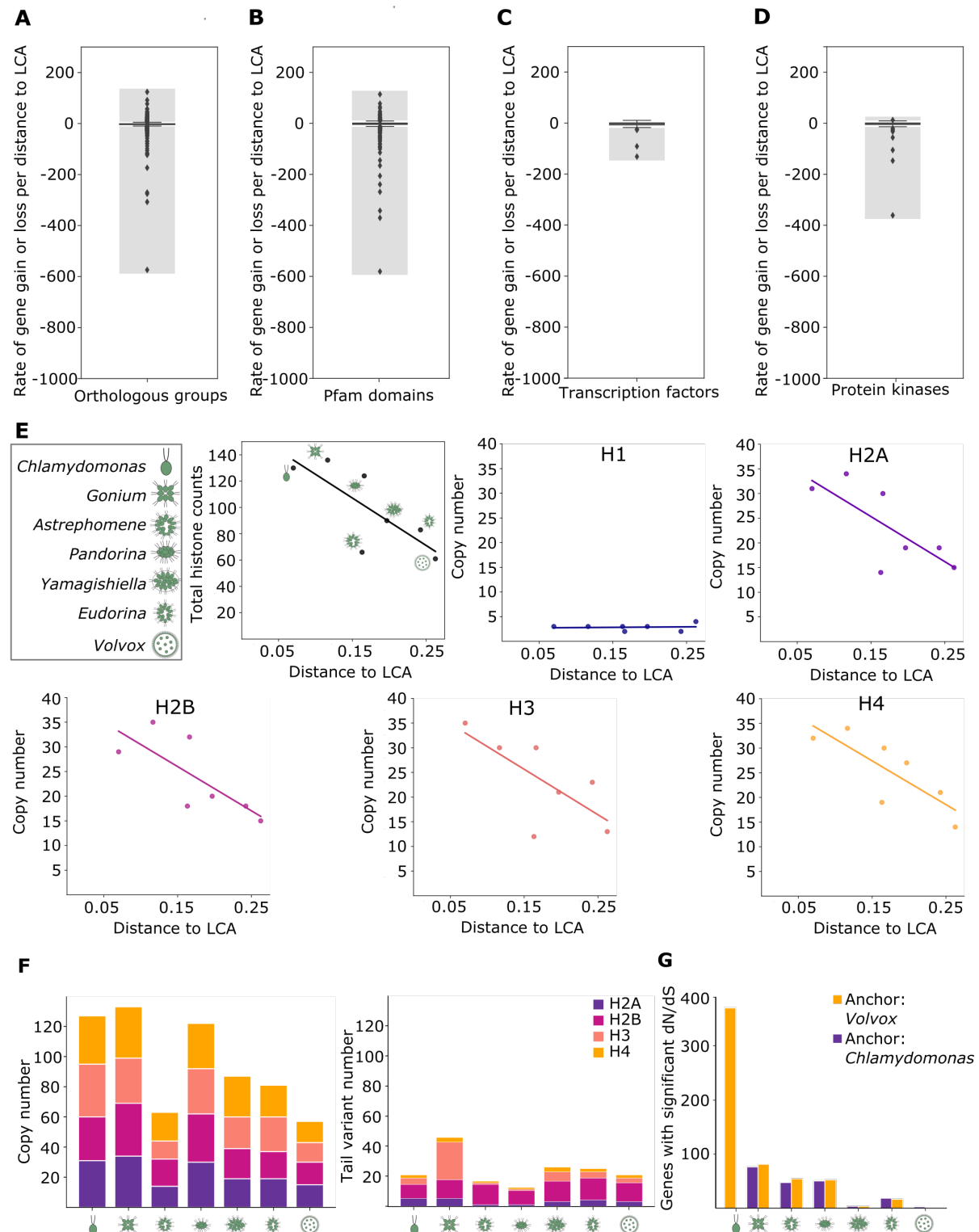


Figure 3.2. Gene loss outpaces gain in volvocine genomes.

Distribution of the gene loss and gain rates per distance to LCA in the volvocine genomes of A) orthologous groups, B) Pfam domains, C) transcription factors, and D) protein kinases. Outlier groups, which represent significant loss (negative values) or gain (positive values) rates, are within the shaded grey area. E) Histone copy number relative to distance to LCA shows a decreasing trend except for H1 in the volvocine genomes. Top left panel describes total histone counts per species. F) Histone copy numbers (left) are reduced with minimal loss of tail variants (right), and histone copy numbers converge between *Astrephomene* and *Volvox*. G) Quantification of positively selected genes per species using *Chlamydomonas* (purple) and *Volvox* (yellow) as reference genomes.

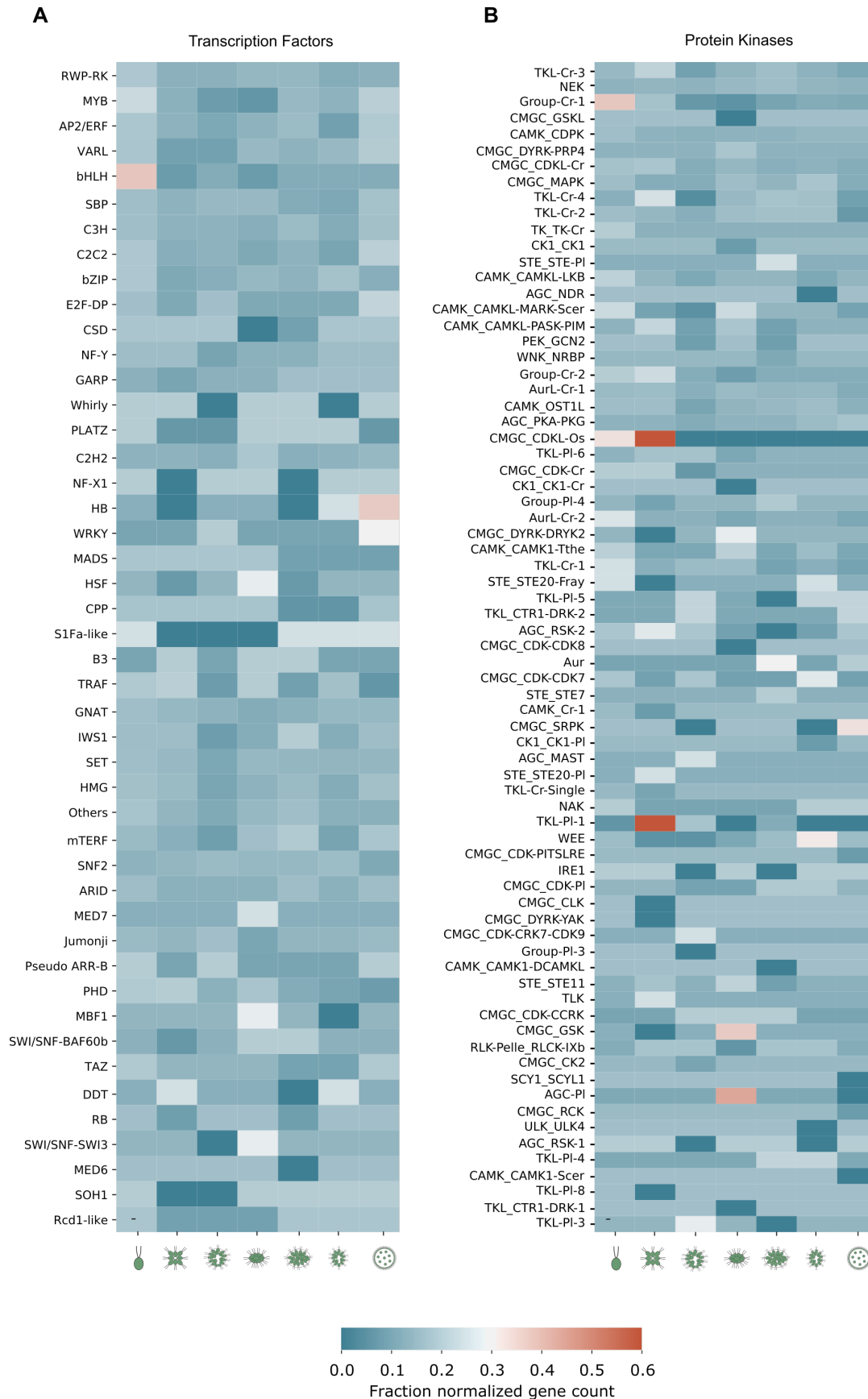


Figure 3.3. A) Transcription factor and B) protein kinase family loss and gain in seven volvocine genomes.

Relative gene counts represent the fraction normalized count relative to counts of the other species.

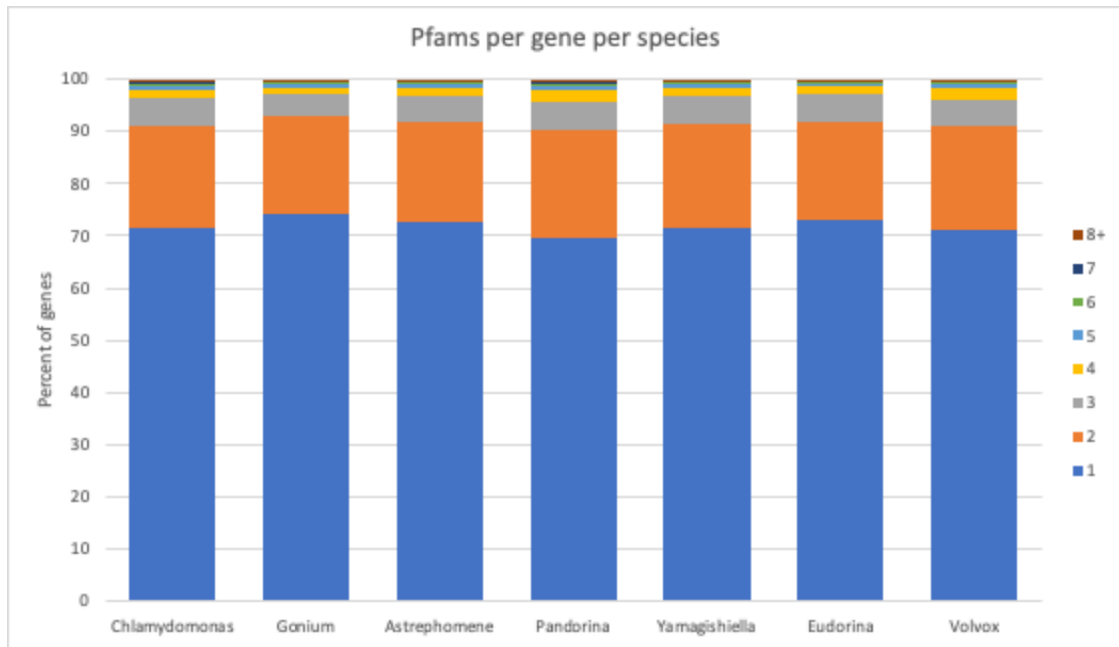


Figure 3.4. Number of Pfams per gene does not change in seven volvocine algae species.

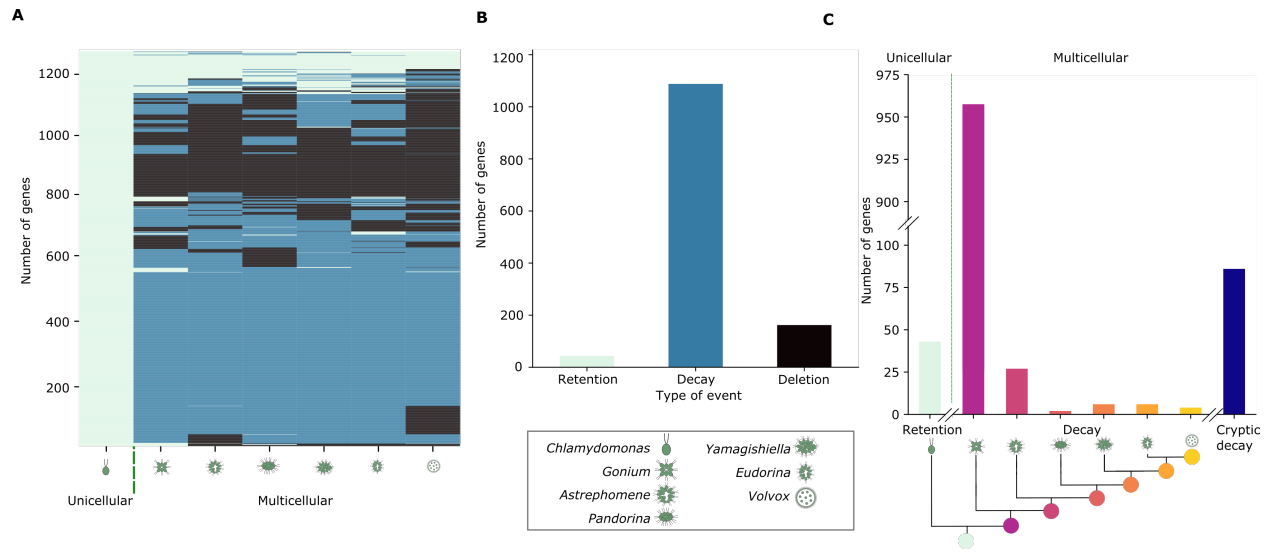


Figure 3.5. Gene loss in the volvocine algae occurs primarily by gradual decay.

A) Retention, decay, and deletion of 1,293 loci belonging to orthologous groups undergoing contraction in volvocine species. B) Cross species quantification of retention (light green), decay (blue), and deletion (black) of loci in contracting orthologous groups. C) Quantification of events of retention and loss by decay relative to evolutionary relationships between volvocine species. Losses that do not follow phylogenetic distributions are considered cryptic (dark blue).

Figure 3.6. Expansion of Cyclin D repertoire is conserved across multicellular volvocines.

A) Phylogeny of cyclin dependent kinases (CDKs). B-F) Phylogeny of cyclins (cyc). CycD1s are in green.

Present in	Number of orthologous groups	Number of Pfams
C	461	13
G	436	10
A	130	8
P	530	18
Y	311	12
E	401	9
V	216	27
CG	127	7
CA	43	3
CP	30	4
CY	82	1
CE	79	2
CV	55	4
GA	112	2
GP	58	1
GY	123	2
GE	105	7
GV	35	0
AP	41	1
AY	60	1
AE	91	0
AV	36	0
PY	75	0
PE	60	1
PV	26	1
YE	210	1
YV	68	2
EV	229	3
CGA	28	1
CGP	23	1
CGY	59	2
CGE	32	0
CGV	24	3
CAP	17	1
CAY	20	0
CAE	12	0
CAV	30	3
CPY	24	1

CPE	9	0
CPV	9	0
CYE	33	0
CYV	28	2
CEV	40	1
GAP	12	1
GAY	30	1
GAE	27	0
GAV	11	1
GPY	38	2
GPE	25	0
GPV	10	0
GYE	59	0
GYV	29	0
GEV	33	0
APY	18	0
APE	8	1
APV	10	1
AYE	28	1
AYV	26	1
AEV	23	1
PYE	49	0
PYV	20	0
PEV	36	0
YEV	88	3
CGAP	13	1
CGAY	25	2
CGAE	14	0
CGAV	14	3
CGPY	35	1
CGPE	24	2
CGPV	5	4
CGYE	80	3
CGYV	29	1
CGEV	47	4
CAPY	7	2
CAPE	1	1
CAPV	12	5
CAYE	19	2

CAYV	23	4
CAEV	16	0
CPYE	55	1
CPYV	17	1
CPEV	16	0
CYEV	70	5
GAPY	14	0
GAPE	13	1
GAPV	5	0
GAYE	27	1
GAYV	10	1
GAEV	11	1
GPYE	50	3
GPYV	15	3
GPEV	20	1
GYEV	61	0
APYE	20	1
APYV	12	0
APEV	11	1
AYEV	32	1
PYEV	72	3
CGAPY	32	7
CGAPE	21	0
CGAPV	22	7
CGAYE	61	4
CGAYV	45	8
CGAEV	33	5
CGPYE	138	17
CGPYV	45	10
CGPEV	46	9
CGYEV	303	27
CAPYE	29	2
CAPYV	21	6
CAPEV	19	4
CAYEV	97	7
CPYEV	141	15
GPYEV	71	3
GAPYE	38	2
GAPYV	16	0

GAYEV	51	4
GAPEV	14	0
APYEV	48	5
CGAPYE	223	27
CGAPYV	107	29
CGAPEV	72	21
CGAYEV	1056	84
CGPYEV	955	78
CAPYEV	306	64
GAPYEV	140	13
CGAPYEV	5421	1296

Table 3.1. Counts for shared orthologous groups and Pfams across seven volvocine species.

Species are codified as follows: C= *Chlamydomonas*, G= *Gonium*, A= *Astrephomene*, P=

Pandorina, Y= *Yamagishiella*, E= *Eudorina*, V= *Volvox*

Species	Actin	Alpha-Tubulin	Beta-tubulin	Chloroplast Ferredoxin	mtATP synthase- alpha subunits	mtATP synthase- beta subunit	Flagellar Dynein Heavy Chain
<i>Chlamydomonas</i>	2	2	2	1	2	1	16
<i>Gonium</i>	2	2	2	1	3	1	15
<i>Astrephomene</i>	2	2	2	1	2	1	23
<i>Pandorina</i>	2	3	2	1	2	1	22
<i>Yamagishiella</i>	2	2	2	1	2	1	18
<i>Eudorina</i>	2	2	2	1	2	1	16
<i>Volvox</i>	2	2	2	2	3	1	15

Table 3.2. Conserved gene counts for seven volvocine species.

Species	H1
<i>Chlamydomonas</i>	3
<i>Gonium</i>	3
<i>Astrephomene</i>	3
<i>Pandorina</i>	2
<i>Yamagishiella</i>	3
<i>Eudorina</i>	2
<i>Volvox</i>	4

Table 3.3. Histone H1 gene counts for seven volvocine species.

Histone	Species	Copy number	Tail var. number
H2A	<i>Chlamydomonas</i>	31	5
	<i>Gonium</i>	34	5
	<i>Astrephomene</i>	14	1
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	19	3
	<i>Eudorina</i>	19	4
	<i>Volvox</i>	15	3
H2B	<i>Chlamydomonas</i>	29	9
	<i>Gonium</i>	35	12
	<i>Astrephomene</i>	18	13
	<i>Pandorina</i>	32	9
	<i>Yamagishiella</i>	20	13
	<i>Eudorina</i>	18	14
	<i>Volvox</i>	15	12
H3	<i>Chlamydomonas</i>	35	4
	<i>Gonium</i>	30	24
	<i>Astrephomene</i>	12	1
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	21	6
	<i>Eudorina</i>	23	4
	<i>Volvox</i>	13	3
H4	<i>Chlamydomonas</i>	32	2
	<i>Gonium</i>	34	3
	<i>Astrephomene</i>	19	1
	<i>Pandorina</i>	30	1
	<i>Yamagishiella</i>	27	3
	<i>Eudorina</i>	21	2
	<i>Volvox</i>	14	2

Table 3.4. Histone family copy number and tail variant number for seven volvocine species.

Absence in	Total	Decay	Deletion
GAPYEV	1082	957	125
APYEV	31	27	4
GPYEV	3	3	0
GAYEV	4	4	0
GAPEV	8	6	2
GAPYV	8	7	1
GAPYE	0	0	0
PYEV	3	2	1
AYEV	3	1	2
APEV	7	5	2
APYV	10	8	2
APYE	3	2	1
GYEV	0	0	0
GPEV	2	2	0
GPYV	0	0	0
GPYE	0	0	0
GAEV	1	1	0
GAYV	0	0	0
GAYE	1	1	0
GAPV	2	2	0
GAPE	0	0	0
GAPY	0	0	0
YEV	6	6	0
PEV	5	4	1
PYV	2	2	0
PYE	0	0	0
AEV	3	3	0
AYV	4	3	1
AYE	0	0	0
APV	5	5	0
APE	2	1	1
APY	0	0	0
GEV	0	0	0
GYV	0	0	0
GYE	0	0	0
GPV	0	0	0
GPE	0	0	0
GPY	0	0	0

GAV	1	0	1
GAE	0	0	0
GAY	0	0	0
GAP	0	0	0
EV	6	6	0
YV	3	1	2
YE	0	0	0
PV	9	5	4
PE	0	0	0
PY	1	1	0
AV	8	5	3
AE	0	0	0
AY	1	1	0
AP	2	2	0
GV	0	0	0
GE	0	0	0
GY	0	0	0
GP	1	1	0
GA	2	2	0
V	10	4	6
E	1	1	0
Y	3	2	1
P	4	2	2
A	2	2	0
G	1	1	0

Table 3.5. Distribution and type of gene losses in multicellular volvocine algae.

Species names are encoded as follows: G= *Gonium*, A= *Astrephomene*, P= *Pandorina*, Y= *Yamagishiella*, E= *Eudorina*, V= *Volvox*

Chapter 4 - Transcriptional regulation of cellular differentiation by

rlsD* in *Volvox carteri

Berenice Jiménez-Marín, José Ortega Escalante, J. D. Seah, Bradley J.S.C. Olson,
Stephen Miller

To be submitted in November 2022

Author contributions: BJM: conceptualization, software, formal analysis, investigation, writing – original draft, data curation, visualization. JOE: Conceptualization, formal analysis, investigation, writing – original draft, visualization. JDS: investigation, formal analysis visualization, validation, data curation. BJSCO: conceptualization, methodology, writing – review & editing, supervision. SM: conceptualization, methodology, validation, data curation, writing – original draft, review & editing, supervision, project administration, funding acquisition.

Introduction

The evolution of multicellularity is among the most consequential events on the history of life itself (John Maynard Smith, 1995). It is also unusual in the sense that it has occurred independently across different domains and lineages within domains (Bonner, 1998; Grosberg and Strathmann, 2007). A transcendental consequence of multicellular life is the evolution of cell differentiation; however, the mechanisms and molecular players underlying cell differentiation are for the most part an enigma.

The volvocine algal model system is ideal to study the molecular underpinnings of multicellular and cell differentiation evolution (Nedelcu and Michod, 2003; Kirk, 2005; Niklas, 2014; Hanschen et al., 2017; Jimenez-Marin and Olson, 2022). This algal group evolved multicellularity fairly recently (~250 MYA) (Herron et al., 2009), key species have been sequenced and are experimentally tractable (Merchant et al., 2007; Lerche and Hallmann, 2009; Prochnik et al., 2010; Umen and Olson, 2012; Lerche and Hallmann, 2013; Hanschen et al., 2016; Featherston et al., 2017; Hamaji et al., 2018; Yamashita et al., 2021). Moreover, member species exhibit phenotypes that range from unicellular *Chlamydomonas reinhardtii* to multicellular undifferentiated e.g. *Gonium pectorale*, *Pandorina morum*, to multicellular with partial differentiation in genus *Pleodorina*, and to multicellular differentiated *Volvox carteri* (Coleman, 2012; Umen and Olson, 2012; Yamashita et al., 2016).

The life cycle of *Volvox carteri* lasts for 48 hours and is well characterized (Kirk, 1997; Kirk, 2000; Miller, 2002; Umen and Olson, 2012). During asexual development, *Volvox* gonidia undergo repeated rounds of mitosis, called multiple fission. At the 32-cell stage, a subset of cells undergo three rounds of asymmetric division, which marks the beginning of cell differentiation in the organism (Kirk, 2001;2005). Large cells produced by these asymmetric divisions are

primed to become gonidia, while all remaining small cells develop into flagellated soma that provides motility, nutrients, and deposition of most of the extracellular matrix (ECM) which holds the organism together and comprises the bulk of its volume (Umen and Olson, 2012). Young *Volvox* spheroids complete their juvenile development at about 36 hours after they started dividing and hatch from their parental spheroid (Ortega Escalante, 2018)

The reorganization of fitness at different levels is thought to be a necessity for stabilization of the multicellular state and organismal integration (Michod, 2007). One proposed mechanism of group integration into a new level of individuality is cell specialization, especially into reproductive (germ) and vegetative (soma) cell types, since specialization forces fitness transfers from individual cells to the group in such a way that no individual cell type can survive without the other (Michod, 2007). The *somatic regenerator A* (*regA*) gene encodes a volvocine algal RegA-like (VARL) domain (Duncan et al., 2006) that is related to the SAND domain, which exists in transcription factors of both plants and animals (Bottomley et al., 2001). Hence, *regA* is considered a putative transcription factor. *regA* maintains terminal differentiation of somatic initials into somatic cells in *Volvox* through repression of genes required for cell growth and reproduction (Tam and Kirk, 1991a; Kirk et al., 1999) and is expressed specifically in somatic cells (Kirk et al., 1999; Matt and Umen, 2018).

regA belongs to a family of VARL encoding genes that resulted from a tandem gene duplication event, and hence belongs to what is called the *reg* cluster (Hanschen et al., 2014). There are no orthologs of *regA* in undifferentiated *Chlamydomonas* or multicellular undifferentiated *Gonium*, but another VARL gene, *rlsD*, does have an ortholog in *Chlamydomonas*, *RLS1*. Thus, it has been hypothesized that a unicellular common ancestor of the volvocines contained an *RLS1/rlsD*-like gene with a role in repression of reproduction as a

stress response (Nedelcu, 2009) that was amplified to generate the *reg* cluster in the volvocine lineage after its divergence from *Gonium* but prior to the evolution of cell differentiation in *Volvox* (Hanschen et al., 2014; Grochau-Wright et al., 2017); then *regA* was co-opted for repression of reproduction as a developmental trait (Nedelcu, 2009; Olson and Nedelcu, 2016). In this work we functionally test the role of *rlsD* in *Volvox* development using an overexpression approach in wild-type strains. We found that *rlsD* impacts gonidial development and size. Moreover, high *rlsD* overexpression can lead to the development of cells with phenotypes that are intermediate between gonidial and somatic cells. Comparative RNA-sequencing was then used to understand the molecular basis of the developmental alterations that result from *rlsD* overexpression.

Summary of previous work (Ortega Escalante, 2018)

A construct containing the *rlsD* gene under control of a nitrate-inducible nitrate reductase gene (*nitA*) promoter was transformed into the wild-type (WT) *Volvox carteri* EVE strain. A selectable marker plasmid with a hygromycin resistance cassette was used for co-transformation. Transformants were generated by bombarding *Volvox* gonidia with the plasmids, and transformant lines were selected and evaluated by light microscopy. Nine hygromycin-resistant transformants were obtained, 3 of which showed growth and developmental defects upon a single generation of induction in medium containing nitrate as sole nitrogen source. The transformant line DOE-g-6 showed the strongest phenotype, and thus was selected for characterization and RNA-seq experiments.

RT-qPCR showed that *rlsD* transcript levels were ~24x and ~22x higher in nitrate-grown DOE-g-6 compared to nitrate grown EVE at 24 hrs and 48 hrs post induction, respectively. Moreover, *rlsD* transcript levels were ~6.2x and ~3.1x higher in ammonia-grown DOE-g-6

compared to ammonia-grown EVE at 24 hrs and 48 hrs post induction, respectively, suggesting that the *nitA* promoter fragment in this construct drives some *rlsD* expression even under presumed repressing conditions. Hence, we refer to nitrate- and ammonia-grown DOE-g-6 cultures as “highly induced” and “moderately induced”, respectively.

Parental spheroid diameter was not affected significantly by *rlsD* overexpression until the end of the second generation. However, gonidia size of daughter spheroids was affected as early as 24 hours post-induction in DOE-g-6. At 48 hours after induction, parents contained a mixture of normal developing embryos and presumptive gonidia that display slower growth that approximates 2-3x the size of soma. More importantly, these presumptive gonidia developed eye spots, which are somatic traits, and so are at least partially differentiated as somatic cells. The aberrant ‘soma-gonidia’ were not observed in DOE-g-6 under moderately inducing conditions, and they did not de-differentiate or reproduce, even when individuals containing them were shifted back to moderately inducing conditions. These results suggest that *rlsD* overexpression affects growth, and that certain targets of *rlsD* might also have roles in cell differentiation. In light of this possibility, the effects of *rlsD* on the *Volvox* transcriptional landscape were assayed.

Results

***rlsD* acts as both as a gene activator and repressor**

As means to evaluate the overall impact of *rlsD* overexpression on the *Volvox* transcriptomic landscape, triplicate cultures of EVE and DOE-g-6 under highly inducing and moderately inducing conditions were subjected to RNA sequencing. The resulting reads counts were compared with a previously published cell-type RNA-seq dataset from *Volvox carteri* (Matt and Umen, 2018) were mapped to a tailored merge of *Volvox* genome assembly versions 1, 2, and 2.1. A total of 22,253 transcripts were mapped, median normalized, and filtered for differential

expression under a false discovery rate (FDR) < 0.05. This analysis yielded 17,166 differentially expressed transcripts (77.14% of the total) representing 15,833 genes.

Changes to the nitrogen source in *Volvox* growth media for the purpose of inducing (or not) *rlsD* may result in differential expression of genes unrelated to *rlsD* function itself. Given that *rlsD* is somewhat overexpressed in DOE-g-6 even under the most repressing conditions, all subsequent transcript level comparisons were performed on DOE-g-6 against EVE grown under the same nitrogen source. The genes whose differential expression was two-fold or greater between WT and DOE-g-6 were classified as significantly differentially expressed, further analyzed, and suggest that *rlsD* overexpression in highly inducing conditions results in both gene expression and repression (1,515 upregulated and 2,133 downregulated genes) (Figure 4.1). This dual role is also observable to a smaller extent in moderately inducing conditions, where 439 genes were upregulated and 633 were downregulated in DOE-g-6 compared to WT (Figure 4.1). These results match previously obtained RT qPCR data, suggesting the presence of leaky *rlsD* overexpression under the *nitA* promoter in DOE-g-6.

Genes involved in ribosomal biogenesis, photosynthesis and ECM modification are *rlsD* targets

To test for the functional profile of *rlsD* targets (direct and indirect), PFAM enrichment analysis of significantly differentially expressed genes was performed on DOE-g-6 up- and down-regulated genes under highly inducing conditions as well as on genes that are up- and down-regulated under both highly and moderately inducing conditions.

Genes overexpressed in DOE-g-6 under highly inducing conditions encode for 180 Pfams, of which 98 are enriched (Table 4.1), whereas genes upregulated under both highly inducing and moderately inducing conditions encode 62 Pfams, of which 12 are enriched (Table

4.1). These two datasets comprise 109 distinct enriched Pfams, of which 14 (12.8% of enriched Pfams in upregulated genes) are involved in carbon metabolism (Figure 4.2, Table 4.1).

Phosphorins, gametolysin peptidase, and fasciclin domains, which are found in proteins involved in ECM maintenance and modification, and in cell-cell adhesion molecules, are also enriched in upregulated genes (Figure 4.2, Table 4.1).

The downregulated genes in DOE-g-6 under highly inducing conditions encode for 904 Pfams, of which 279 are enriched (Table 4.1). Genes downregulated under both highly inducing and moderately inducing conditions contain 117 Pfams, of which 19 are enriched (Table 4.1). These two datasets consist of 285 enriched Pfams. Surprisingly, 62 Pfams (21.8% of enriched Pfams in downregulated genes) are related to ribosomal biosynthesis and processing (Figure 4.2, Table 4.1). Moreover, 21 (7.4%) Pfams belong to genes encoding for core photosynthetic apparatus proteins. DNA repair, carbon metabolism, and phosphorin domains are also enriched among downregulated genes (Figure 4.2, Table 4.1). Thus, it appears that *rlsD* targets key metabolic pathways that relate to cell metabolism and growth by altering the ECM, redirecting carbon, and reducing the photosynthetic and translational capacity of *Volvox* cells.

***rlsD* overexpression results in biased regulation of cell-type related genes**

Recent evidence suggests that germ and soma in *Volvox* have well-defined, distinct metabolic profiles (Matt and Umen, 2018). Considering the functional biases between *rlsD* activation and repression targets, and of facts that *regA* maintains somatic status and is a close *rlsD* paralog, we evaluated the cell bias of genes differentially regulated by *rlsD*. Cell-type differential expression information from our experimental conditions was mapped and normalized along with a published cell-type transcriptome dataset (Matt and Umen, 2018) as means to assign cell-type specificities to *rlsD* targets. Differentially expressed genes by cell type were then classified as

follows: differential expression of 5-fold or more constituted ‘cell-type specific’ genes, expression differences between 2-fold and 5-fold represented ‘cell-type biased’ genes, and genes that were differentially expressed by less than 2-fold are considered ‘non-specific’.

Non-specific genes (4,321 total) were over- and under-expressed in roughly the same proportions in DOE-g-6 under highly inducing conditions, with 496 (11.5%) genes upregulated and 452 (10.5%) downregulated (Figure 4.3). Soma-related genes (that is, soma-specific and soma-biased genes) showed a moderate expression bias. Of 4,981 soma-related genes, 660 (13.2%) are upregulated by two-fold or more in DOE-g-6 under highly inducing conditions, whereas 390 (7.8%) are downregulated by the same amount (Figure 4.3). The strongest expression bias, however, belongs to germ-related genes. Of 6,530 germ-related genes in *Volvox*, 1,291 (19.8%) are downregulated by two-fold or more, whereas only 358 (5.5%) are upregulated to the same extent (Figure 4.3).

As means to minimize the presence of transcriptional noise resulting from differences in the nitrogen source used for high induction experiments, cell-type bias analysis was also done for genes that are up- or downregulated by twofold or more in DOE-g-6 in both highly inducing and moderately inducing conditions (Figure 4.3, highlighted quadrants). Including *rlsD*, 170 genes are overexpressed in DOE-g-6, with a small majority of them (67 genes, 39.4% of upregulated set) being soma-related (Figure 4.3). Conversely, of 280 genes downregulated by two-fold or more in DOE-g-6 under both induction conditions, 166 (59.3% of downregulated set) are germ-related (Figure 4.3). These results strongly suggest that *rlsD* preferentially represses germ-related genes.

Growth, reproduction and stress related genes are regulated by *rlsD*

rlsD is believed to regulate the expression patterns of genes encoding for key pathways and processes related to growth and reproduction, as a response to stress (Nedelcu, 2009). Hence, the expression patterns for key sets of genes involved in metabolism, stress, cell growth and differentiation were queried. A dataset of 17 phosphorus availability genes and 27 sulfur availability genes (14 anabolic, 13 catabolic) were assayed as representatives of nutritional stress responses. Phosphorus metabolism genes show a small bias towards downregulation in highly inducing conditions, and two phosphorus metabolism genes are significantly downregulated in DOE-g-6 regardless of induction strength (Figure 4.4A). Conversely, sulfur metabolism genes (Figure 4.4B, C), specifically anabolism related genes, are biased towards upregulation, and two sulfur anabolism genes are upregulated by twofold or more in DOE-g-6 (Figure 4.4 C).

Oxidative stress genes were also analyzed. While the expression of most reactive oxygen species (ROS) response genes was minimally affected, two genes were significantly downregulated and two were significantly upregulated (Figure 4.4D). Interestingly, upon classifying the genes by their cell-type expression, both downregulated genes are germ-specific (Fig. 4.4D).

Regulation of reproduction could occur through changes to expression of ribosomal proteins or downregulation of ribosome biosynthesis. However, none of the 87 ribosomal proteins or 107 ribosomal biosynthesis appeared significantly dysregulated in DOE-g-6 under both induction conditions. Despite this, and in accordance to Pfam enrichment results, the overall trend for both gene sets was repression, especially under high induction (Figure 4.4E, F, respectively). Cell cycle related genes (anaphase checkpoint genes, CDCs, CDKs, cyclins, and cell-cycle regulator RB encoding genes) are likewise unaffected by *rlsD* overexpression (Figure 4.4 G-K).

Key structural and developmental genes were analyzed to address the potential regulatory role of *rlsD* on *Volvox* morphogenesis. Inversion gene (*inv*) expression was not affected (Figure 4.4 L), nor was expression of the asymmetric division gene *glsA* (Figure 4.5 A). Likewise, only one ciliary gene was significantly upregulated out of 201 tested genes (Figure 4.5 F).

Surprisingly, chloroplast maintenance and chlorophyll biosynthesis genes were unaffected by *rlsD* overexpression (Figure 4.5 B, C). Unlike those gene groups, electron transport and light harvesting complex II (LHC II complex) genes had a downregulation bias under high induction conditions (Figure 4.5 D, E). Two of 55 core photosynthesis genes are significantly downregulated in DOE-g-6 (Figure 4.5 D), and three of 27 LHC II complex genes are significantly downregulated (Figure 4.5 E) across both induction conditions. This is of special interest, as all significantly downregulated LHC II complex and electron transport genes are germ-related, and photosystem I and II Pfams were found to be enriched.

Given the possibility that *rlsD* impacts growth as a result of stress, carbon metabolism and ECM related gene expression was evaluated. Two of 139 carbon metabolism genes analyzed were significantly upregulated, while another 2 were downregulated (Figure 4.5 G). The downregulated genes are both germ-specific, while the upregulated genes are soma-specific. In agreement with Pfam enrichment analysis, 4 of 41 matrix metalloproteases (MMPs) are significantly upregulated, while 1 is downregulated in DOE-g-6 (Figure 4.5 H). Likewise, 5 of 91 pherophorin genes are upregulated, and 10 are downregulated (Figure 4.5 I). Interestingly, 8 out of the 10 downregulated pherophorins are germ-specific genes. Other candidate growth and metabolism genes (glycosyl transferases, nucleotide sugar metabolism genes, and autophagy related genes) were not affected (Figure 4.5 J, K, and L, respectively).

rlsD encodes for a VARL domain, and thus is a putative transcription factor (TF) (Duncan et al., 2006). Since *rlsD* overexpression could result in a cascade of regulatory outputs, we analyzed 133 TFs. Of these, 4 are significantly upregulated and 3 are significantly downregulated in DOE-g-6 under both induction conditions (Figure 4.6 A). Three of the 4 upregulated TFs are soma-specific, whilst 2 of the 3 downregulated TFs are germ-specific. One such upregulated gene, an Myb encoding TF, has a candidate role in cell differentiation for *Astrephomene gubernaculifera*, whose lineage underwent independent evolution of somatic cells (Yamashita et al., 2021) (Figure 4.6 B). Among the 14 VARL genes in the *Volvox* genome, only *rlsA* was significantly upregulated (besides *rlsD* itself) (Figure 4.6 C). Taken together, these results suggest that *rlsD* impacts the expression of a wide array of genes, especially among key functions involved in reproduction and growth and that have cell-type differential expression.

Discussion

At the phenotypic level, *rlsD* overexpression can result in the development of “somagonidia” – cells that are larger than somatic cells, smaller than germ cells, and have eyespots (Ortega Escalante, 2018). The existence of somagonidia suggests that *rlsD* can mimic the function of *regA* and enforce the somatic cell fate when highly overexpressed. Interestingly, ~25% of genes in *Volvox* were upregulated or downregulated twofold or more across the high and moderate overexpression conditions in DOE-g-6. However, there is a clear tendency for downregulation of germ-related genes, whereas upregulated genes were somewhat more likely to be soma-related (Figure 4.3).

Upon examining 27 different candidate gene sets, it would seem that *rlsD* mainly targets genes that relate to ribosome biosynthesis, photosynthesis, ECM biosynthesis, and macronutrient metabolism. Despite not being repressed significantly in most cases, ribosomal proteins as a

whole show a clear pattern of downregulation under highly inducing conditions (Figure 4.2, 4.4 E, F, Table 4.1). Similarly, LHC II genes (Figure 4.5 E, Table 4.1), RuBisCO subunit genes (Figure 4.5 G, Table 4.1), and electron transport genes (Figure 4.5 D, Table 4.1) show signatures of downregulation. These results suggest that *rlsD* plays a role in limiting ribosomal and photosynthetic activity in the cell, and, as a consequence, reducing cell growth. It is interesting to note that not all photosynthesis-related genes are affected by *rlsD* overexpression (Figure 4.5 B, C); these other genes could be regulated by other transcription factors (including *regA*) as means to provide robust cell-type specification, or could reflect that transcriptional patterns that are shared between distinct response pathways eventually diverge in their targets.

Several genes encoding products related to ECM biosynthesis and remodeling are impacted by *rlsD* overexpression. Interestingly, these classes have genes be both significantly up- and downregulated, with the regulatory biases following the repression of germ-related and expression of soma-related genes that was observed overall for *rlsD* targets (Figure 4.2, 4.5 H, I, Table 4.1). The distribution of expression profiles for the ECM related genes could be a consequence of different cell types tailoring their deposited ECM for developmental and/or metabolic purposes.

Sulfur and phosphorus metabolism are also impacted by *rlsD* overexpression, but they are affected in different ways. While phosphorus metabolism genes trend towards downregulation, sulfur catabolism gene expression is mostly unaffected, and sulfur anabolism genes trend towards upregulation (Figure 4.4 A-C, Table 4.1). The reason for this apparent interplay between sulfur and phosphorus metabolism regulation is unclear, but studies in *Arabidopsis* have shown that sulfur deprivation leads to replacement of sulfolipids by phospholipids, whereas phosphate deprivation leads to the replacement of phospholipids by

sulfolipids (Yu et al., 2002), leaving way for the possibility that sulfur and phosphorus metabolism interact in *Volvox* similarly.

Despite its wide range of targets, *rlsD* is by no means regulating the full spread of genes that play roles in cell-type specification in *Volvox*. For instance, most carbon metabolism genes (except RuBisCO subunit encoding genes), nucleotide sugar biosynthesis, and cilia genes, were minimally affected, if at all (Figure 4.5 F, G, J, Table 4.1). Similarly, genes involved in early *Volvox* development, including cell cycle, inversion, and asymmetric division genes, were unaffected (Figure 4.4 H-L, 4.5 A). Likewise, despite the fact that VARL genes have been proposed to have an ancestral function as stress-response genes, most ROS response genes were unaffected by *rlsD* overexpression (Figure 4.4 D), suggesting that *rlsD* likely doesn't play a substantial protective role against ROS.

These findings pose the possibility that *rlsD* is a repressor of cell growth and germ development that might act complementarily to *regA*. Indeed, both genes share a multitude of targets, especially among photosynthesis and ribosome related genes. While there is no evidence for what genes are specifically targeted by *regA*, it is true that the expression patterns between germ and soma suggest that *rlsD* is neither the only nor the main enforcer of cell identity maintenance in *Volvox*. Rather, it seems that *rlsD* acts as a fine-tuning modulator that slows growth of cells in response to environmental and perhaps developmental cues as well.

In light of these results, we propose that *rlsD* and *regA* might represent an instance of neofunctionalization; that is, that as *regA* diverged from *rlsD* it became co-opted for cell-type specification maintenance through mutation and changes to its regulatory pattern, while *rlsD* kept its ancestral function (to a greater or lesser degree) (Figure 4.7). Exactly how *regA* evolved

to function differently than *rIsD* remains to be determined, and should be a priority for future investigations.

Methods

Library preparation, RNA-sequencing and read assembly

Triplicate cultures of EVE and DOE-g-6 in highly inducing and moderately inducing conditions were harvested by centrifugation at 600g for 10 minutes with 0.005% Tween-20 and flash frozen in liquid nitrogen. Frozen cell pellets were ground to a fine powder using liquid nitrogen in a mortar and pestle. RNA was extracted using the Quick-RNA Plant Miniprep Kit (Zymo Research). cDNA libraries were prepared with the TruSeq Stranded mRNA kit (Illumina) according to manufacturer's instructions. Libraries were paired-end sequenced (2x100 bp) for 100 cycles using the Illumina HiSeq 2500 System. All sequencing was performed by Genewiz (South Plainfield, New Jersey, USA).

FASTQ files for this dataset and a published dataset for *Volvox* cell-type RNA-seq (Matt and Umen, 2018) were processed using sickle (v1.33) and scythe (v8.3.0). High quality reads were aligned to a merged *Volvox* genome using RSEM (v1.3.2) (Li and Dewey, 2011). Due to the high similarity between LHCII genes in clade I and IV (Matt and Umen, 2018), 5' and 3' UTR regions were extracted and aligned using MUSCLE (v3.8.31) (Edgar, 2004). Clade I genes were collapsed based on their UTR similarities into a total of four representative genes (Vocar.0006s0236, Vocar.0001s1775, Vocar.0025s0135, and Vocar.0002a0041), whereas clade IV genes were all kept for alignment against the merged *Volvox* genome.

Gene expression analysis and cell-type related expression

Gene counts were median normalized and analyzed for differential expression using EBSeq (v1.26.0) (Leng et al., 2013). Differential expression was determined under a False Discovery

Rate (FDR) < 0.05. Differentially expressed gene counts for isoforms were collapsed into single isoforms by addition, and the resulting normalized counts were log2 transformed.

Differentially expressed genes were classified as germ-related or soma-related in a similar fashion to what has been reported (Matt and Umen, 2018): expression ratios greater than fivefold for the cell-type dataset describe germ/soma specific genes, ratios between twofold and fivefold describe germ/soma biased genes, and genes with expression ratio smaller than twofold constitute non-specific genes. Genes with expression differences greater or equal than twofold between WT and DOE-g-6 were considered significantly differentially expressed and used for further analysis.

PFAM enrichment

Pfam domains were queried against Phytozome data for *Volvox* genomes version 1.0, 2.0, and 2.1. Enrichment analysis was performed on twofold (or more) differentially expressed genes between DOE-g-6 and EVE under highly inducing conditions, as well as on the genes that represent the intersection of two-fold (or more) differentially expressed between DOE-g-6 and EVE under both highly inducing and moderately inducing conditions. Statistical enrichment tests were performed via hypergeometric tests and Bonferroni correction with a threshold P-value = 0.01.

Assembly of candidate gene sets for differential expression analysis

Candidate gene sets for specific biological functions that might be under *rlsD* control were assembled as follows. Genes related to photosynthesis (LHC, electron transport, chloroplast maintenance, chlorophyll biosynthesis), cilia, cell-wall protein glycosylation (glycosyl transferases and nucleotide sugar metabolism), autophagy, and carbon metabolism were previously identified (Matt and Umen, 2018). Additional LHC and electron transport genes not

reported/analyzed in that study were identified by keyword or protein BLAST searches (using gene IDs listed in Matt and Umen) to query *V. carteri* genome assembly 2.1 at Phytozome (<https://phytozome-next.jgi.doe.gov/pz/portal.html>); subsequently “novel” members were verified as members of those functional gene families by BLASTp against NCBI genbank. Antennae protein encoding genes from clade I were treated as a collapsed set of four genes as described above. Cell cycle genes (CDKs, cyclins, CDC, and anaphase checkpoint were identified by using human cell cycle genes listed at the Harmonizome web portal (<https://ngdc.cncb.ac.cn/databasecommons/database/id/1957>) as queries to BLASTp the *Volvox* Phytozome genome assembly 2.1. Cell cycle regulator MAT3 sequence was obtained from published work (Ferris et al., 2010). Genes related to translation (ribosomal and ribosome biogenesis) were identified by BLASTp of candidates returned by keyword search of the *Volvox* assembly 2.1 against NCBI GenBank. Sulfur-responsive genes were identified by reciprocal BLASTp based on a set of *Chlamydomonas* S-responsive genes kindly provided by D. Strenkert. Pherophorin genes were from a previous study (von der Heyde and Hallmann, 2020), with that set augmented as described for LHC protein and electron transport genes. Genes related to key developmental steps in the *Volvox* life cycle (*invA/B/C*, *glsA*, VARL genes) were obtained from Phytozome (Cheng et al., 2003; Nishii et al., 2003; Duncan et al., 2006; Ueki and Nishii, 2009). Matrix metalloproteinases (MMPs) and transcription factor gene sets from the *Volvox* assembly version 2.1 were obtained from published work (Jiménez-Marín et al., 2021). Orthologs of key candidate genes of *Astrephomene gubernaculifera* differentiation were obtained from published work (Yamashita et al., 2021). Genes encoding gametolysin domains were obtained by querying Phytozome *Volvox* genome assembly versions 1.0, 2.0, and 2.1 for genes encoding each Pfam.

Orthologs of *Chlamydomonas* reactive oxygen species (ROS) detoxification genes (Blaby et al., 2015) were obtained by querying against OrthoMCL output.

Phosphate-responsive genes were identified by BLASTp based on a set of *Chlamydomonas* P-responsive genes from a previous study (Bajhaiya et al., 2016).

Figures and tables

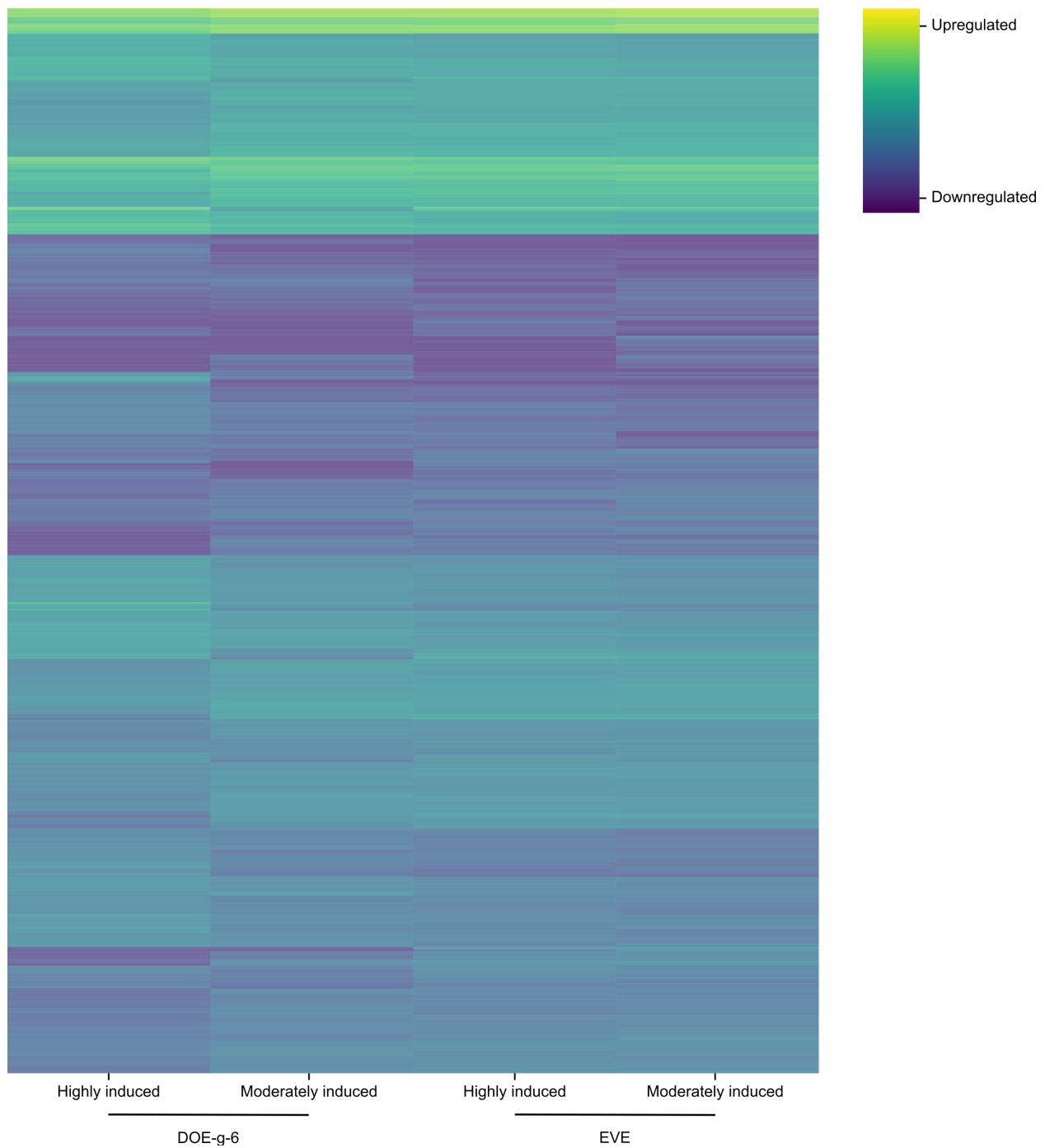


Figure 4.1. *rlsD* overexpression changes transcriptional landscape for a wide array of genes.

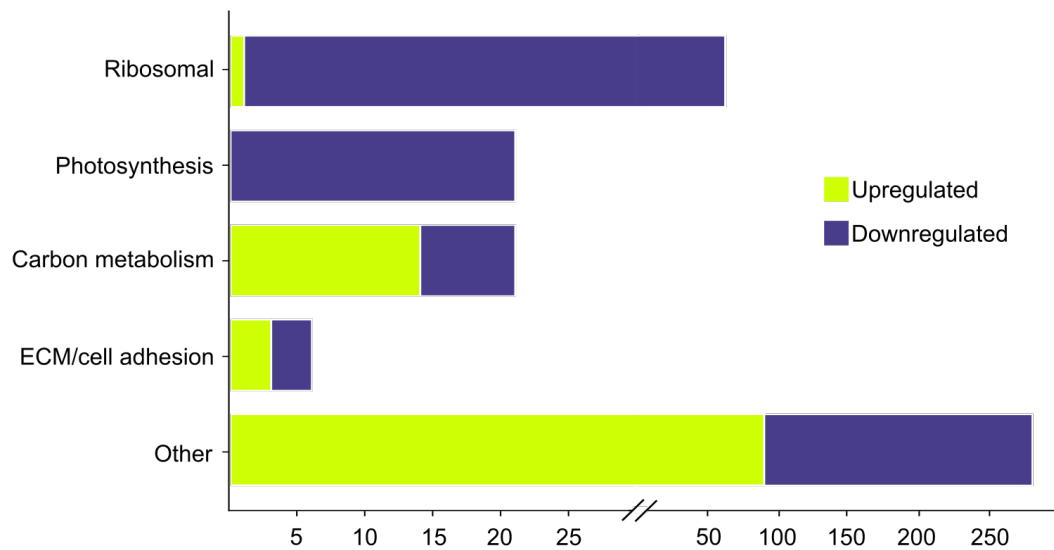


Figure 4.2. *rlsD* targets genes encoding ribosomal and photosynthetic Pfams for downregulation, as well as carbon metabolism and ECM Pfam encoding genes for both up- and downregulation, among others.

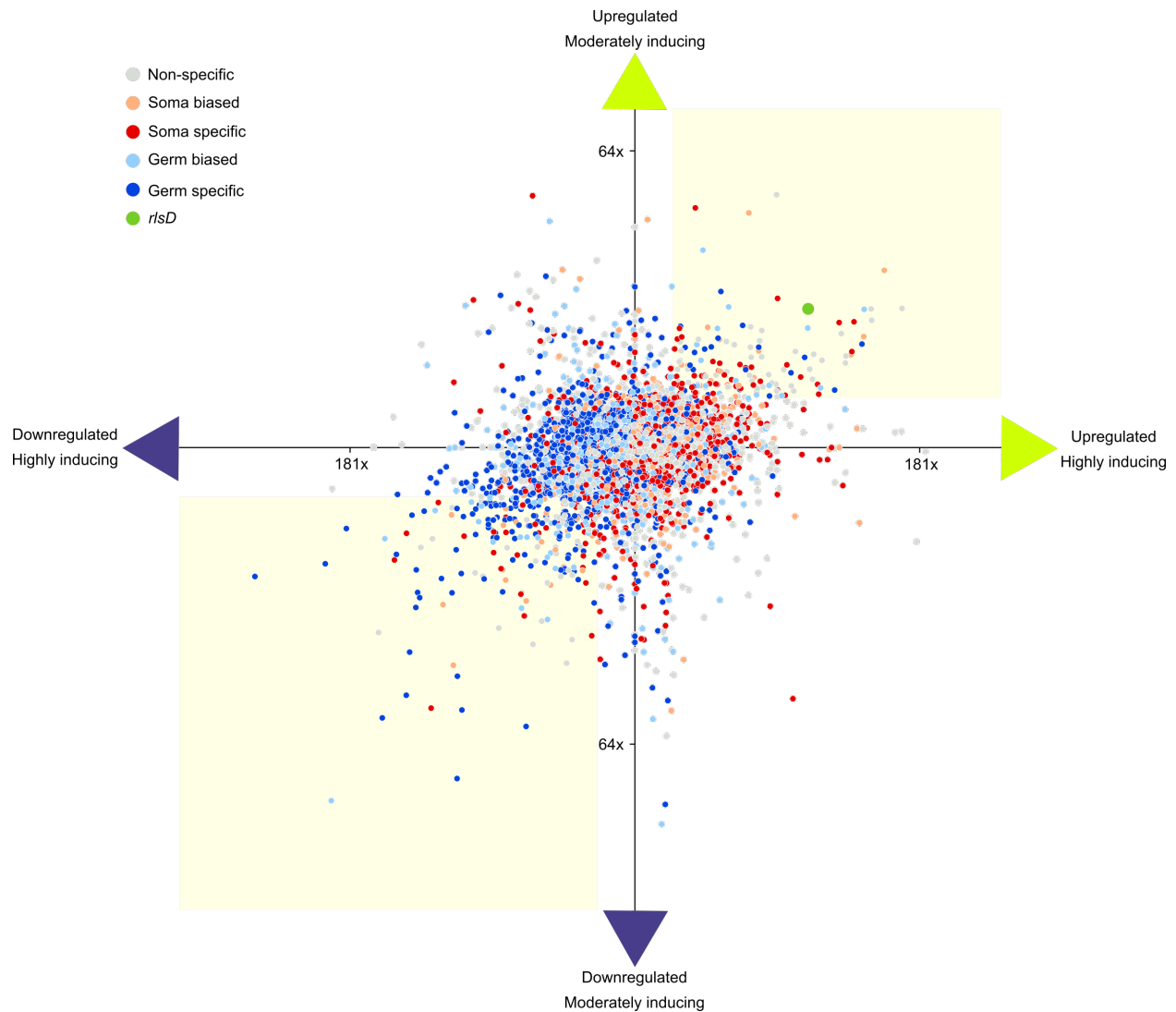


Figure 4.3. *rlsD* targets are biased towards downregulation of germ related genes and upregulation of soma related genes.

Points within highlighted quadrants represent genes that are significantly up- or downregulated in both highly and moderately inducing conditions.

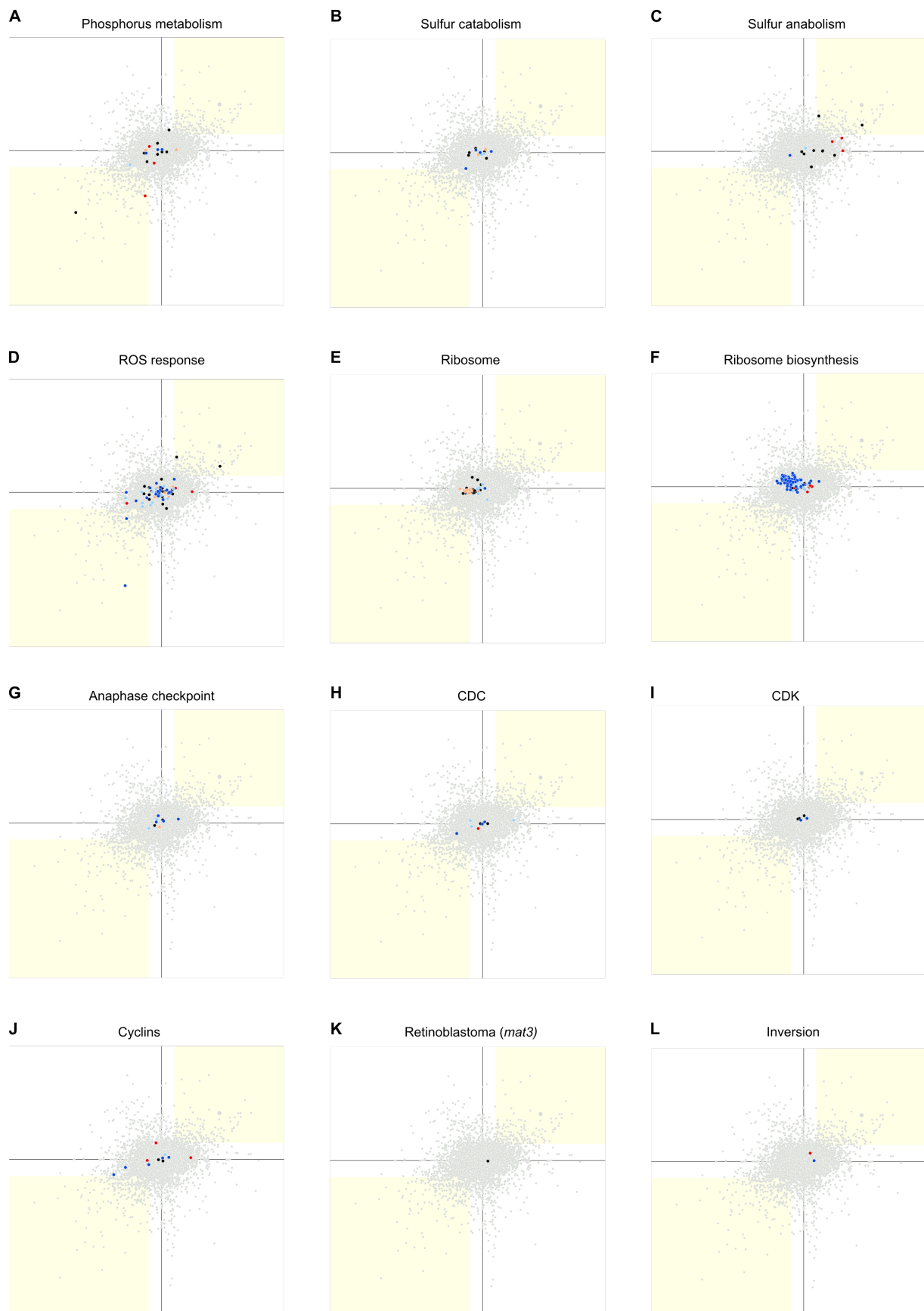


Figure 4.4. Expression patterns of candidate targets of *rlsD* (set 1-12).

Points within highlighted quadrants represent genes that are significantly up- or downregulated in both highly and moderately inducing conditions. A) Phosphorus metabolism, B) sulfur catabolism, C) sulfur anabolism, D) ROS response, E) ribosomal, F) ribosome biosynthesis, G) anaphase checkpoint, H) CDCs, I) CDKs, J) cyclins, K) MAT3, L) *inv* genes.

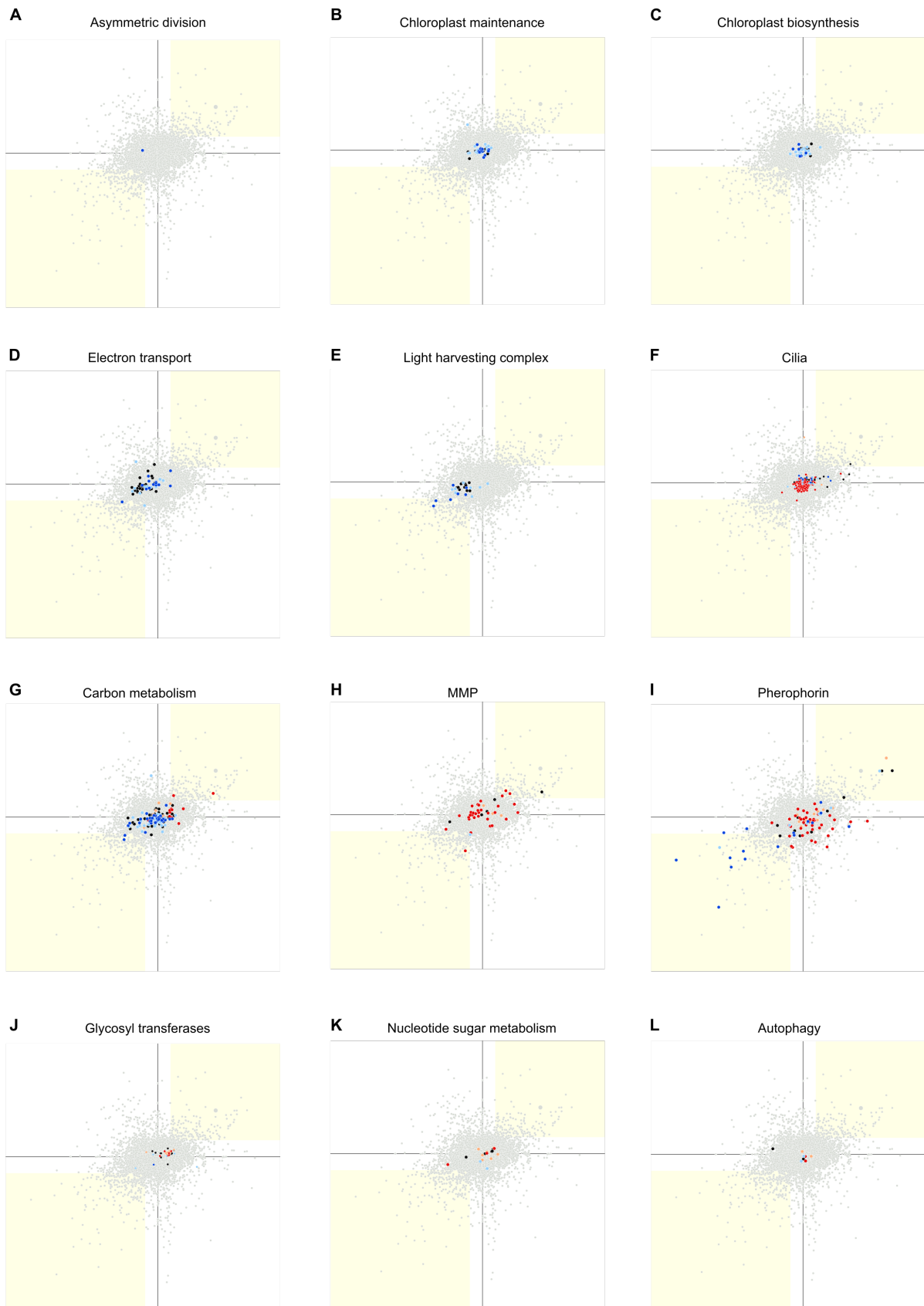


Figure 4.5. Expression patterns of candidate targets of *rlsD* (set 13-24).

Points within highlighted quadrants represent genes that are significantly up- or downregulated in both highly and moderately inducing conditions. A) Asymmetric division, B) chloroplast maintenance, C) chloroplast biosynthesis, D) electron transport, E) LHC II, F) cilia, G) carbon metabolism, H) MMPs, I) perophorins, J) glycosyl transferases, K) nucleotide sugar metabolism, L) autophagy.

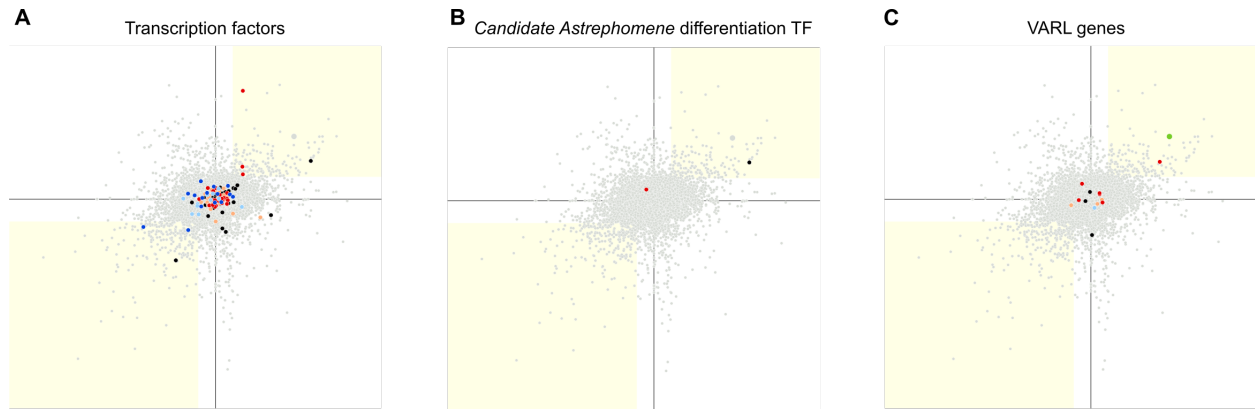


Figure 4.6. Expression patterns of candidate targets of *rlsD* (set 25-27).

Points within highlighted quadrants represent genes that are significantly up- or downregulated in both highly and moderately inducing conditions. A) Transcription factors, B) *Astrephomene* candidate cell differentiation transcription factors, C) VARL genes

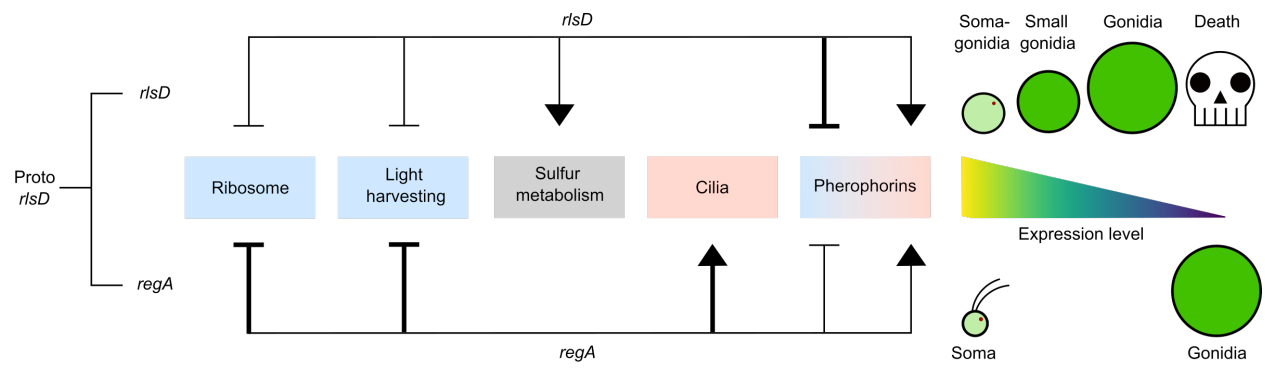


Figure 4.7. *rlsD* and *regA* share multiple targets, but phenotypic output differs.

		Upregulated		Downregulated	
PFAM_name	PFAM	Pval high inducing	Pval both inducing	Pval high inducing	Pval both inducing
Pherophorin	PF12499	2.00E-07	0.003200408		7.95E-06
Bax inhibitor 1 like	PF12811	0	0		
Proteasome-substrate-size regulator, mid region	PF16507	0	0		
Uncharacterized protein family UPF0564	PF10595	0	0		
Gametolysin peptidase M11	PF05548	0.008018727	0.000208254		
Kelch motif	PF13964	0	0.003551731		
PIN domain	PF13638		0.003551731		
Cysteine dioxygenase type 1	PF05995		0.003551731		
Alpha amylase C-terminal beta-sheet domain	PF07821		0.003551731		
Acyl transferase domain	PF00698		0.003551731		
Protein of unknown function (DUF3641)	PF12345		0.003551731		
YCII-related domain	PF03795		0.003551731		
Dicarboxylate carrier protein MatC N-terminus	PF07158	0			
Niemann-Pick C1 N terminus	PF16414	0			
Ferric reductase NAD binding domain	PF08030	0			
Vasohibin	PF14822	0			
PI31 proteasome regulator N-terminal	PF11566	0			
Hydantoinase/oxoprolinase N-terminal region	PF05378	0			
Bacterial extracellular solute-binding protein	PF13416	0			
Microtubule-associated protein 70	PF07058	0			
Low iron-inducible periplasmic protein	PF07692	0			
OHCU decarboxylase	PF09349	0			
HNH endonuclease	PF01844	0			
Bacterial extracellular solute-binding proteins, family 3	PF00497	0			
L-lysine 6-monooxygenase/L-ornithine 5-monooxygenase	PF13434	0			

Alpha adaptin AP2, C-terminal domain	PF02296	0			
First C2 domain of RPGR-interacting protein 1	PF11618	0			
Cytidine and deoxycytidylate deaminase zinc-binding region	PF08211	0			
Centrosomal spindle body, CEP44	PF15007	0			
De-etiolated protein 1 Det1	PF09737	0			
ASCH doman	PF04266	0			
PhoD-like phosphatase	PF09423	0			
Uncharacterized protein family (DUF2045)	PF09741	0			
Core-2/I-Branching enzyme	PF02485	0			
ATP-dependent helicase C-terminal	PF08482	0			
Ligand-gated ion channel	PF00060	0			
Brain and reproductive organ-expressed protein (BRE)	PF06113	0			
UDP-glucuronosyl and UDP-glucosyl transferase	PF00201	0			
USO1/p115 like vesicle tethering protein, C terminal region	PF04871	0			
CC2D2A N-terminal C2 domain	PF15625	0			
Transcription factor TFIID cmplex subunit 8 C-term	PF10406	0			
Alpha galactosidase A	PF16499	0			
DDE superfamily endonuclease	PF13358	0			
Acetamidase/Formamidase family	PF03069	0			
PurA ssDNA and RNA-binding protein	PF04845	0			
ELMO/CED-12 family	PF04727	0			
Nucleotide-sugar transporter	PF04142	0			
Glycosyl transferase family 4	PF00953	0			
Melibiose	PF02065	0			
ChaC-like protein	PF04752	0			
Carbohydrate binding domain (family25)	PF03423	0			
Ferric reductase like transmembrane component	PF01794	0			
Isoprenylcysteine carboxyl methyltransferase (ICMT) family	PF04140	0			

Protein tyrosine and serine/threonine kinase	PF07714	0.007085528			
CD36 family	PF01130	0			
Adenylate and Guanylate cyclase catalytic domain	PF00211	0.000153588			
Rft protein	PF04506	0			
Transport and Golgi organization 2	PF05742	0			
ANTH domain	PF07651	0			
Piruvate ferredoxin/flavodoxin oxidoreductase	PF01558	0			
WSC domain	PF01822	2.94E-03			
Electron transfer flavoprotein FAD-binding domain	PF00766	0			
Cephalosporin hydroxylase	PF04989	0			
HIUase/Transthyretin family	PF00576	0			
Trehalase	PF01204	0			
Swachman-Bodian-Diamond syndrome (SBDS) protein	PF01172	0			
Mad3/BUB1 homology region 1	PF08311	0			
Beta-L-arabinofuranosidase, GH127	PF07944	0			
N-terminal C2 in EEIG1 and EHBPI proteins	PF10358	0			
Domain of unknown function (DUF1338)	PF07063	0			
FAD-binding domain	PF08022	0			
Protein O-mannosyl-transferase TMTC, DUF1736	PF08409	0			
Domain of unknown function (DUF4218)	PF13960	0			
Golgi to ER traffic protein 4	PF04190	0			
Helix-hairpin-helix motif	PF12826	0			
Fasciclin domain	PF02469	6.32E-06			
NAD-dependent DNA ligase adenylation domain	PF01653	0			
Africa swine fever virus J13L protein	PF05568	0			
Transposase family tnp2	PF02992	0			
DNA polymerase delta, subunit 4	PF04081	0			
Phosphoenolpyruvate phosphomutase	PF13714	0			

Serine-threonine protein kinase 19	PF10494	0			
PDZ-like domain	PF12812	0			
Bacterial protein of unknown function (DUF924)	PF06041	0			
Per1-like family	PF04080	0			
Bromodomain associated	PF07524	0			
B3 DNA binding domain	PF02362	0			
Alpha-L-arabinofuranosidase B (ABFB) domain	PF05270	0			
Zinc finger, C2H2 type	PF00096	0			
Domain of unknown function (DUF4485)	PF14846	0			
Ligated ion channel L-glutamate- and glycine-binding site	PF10613	0			
CEP-19-like protein	PF14933	0			
Protein of unknown function (DUF4497)	PF14924	0			
Pyruvate flavodoxin/ferredoxin oxidoreductase, thiamine diP-bdg	PF01855	0			
Domain of unknown function	PF10371	0			
Protein of unknown function (DUF679)	PF05078	0			
TAL effector repeat	PF03377	0			
Bacterial fructose-1,6-bisphosphatase, glpX-encoded	PF03320	0			
Protein of unknown function (DUF3445)	PF11927	0			
D-Tyr-tRNA(Tyr) deacylase	PF02580	0			
GLTT repeat	PF01744	0			
NAD-dependent DNA ligase OB-fold domain	PF03120	0			
Domain of unknown function (DUF4436)	PF14494	0			
SBDS protein, domain II	PF09377	0			
tRNase Z endonuclease	PF13691			0	0
Peptide-N-glycosidase F, C terminal	PF09113			0	0
Peptide-N-glycosidase F, N terminal	PF09112			0	0
PI31 proteasome regulator	PF08577			0	0
Bacterial protein of unknown function (DUF839)	PF05787			0	0

Calcium/calmodulin dependent protein kinase II association domain	PF08332			0	0
Ubiquitin-activating protein	PF02134			0	0
Protein of unknown function (DUF952)	PF06108			0	0
WH2 motif	PF02205			0	0
NAD synthase	PF02540			0	0
Profilin	PF00235			0	0
Copper binding proteins, plastocyanin/azurin family	PF00127			0	0
M protein repeat	PF02370			0	0
FATC domain	PF02260				0.003785768
PAN domain	PF00024				0.001531137
Domain of unknown function (DUF4470)	PF14737				0.000387041
Dihydroorotate dehydrogenase	PF01180				0.001531137
Leucine Rich Repeat	PF00560				0.007488363
Uncharacterized protein family UPF0014	PF03649			0	
Bystin	PF05291			0	
tRNA pseudouridylate synthase B C-terminal domain	PF16198			0	
tRNA (Uracil-5-)-methyltransferase	PF05958			0	
PAT complex subunit CCDC47	PF07946			0	
HIO933-like protein	PF03486			0	
Protein of unknown function (DUF3755)	PF12579			0	
THUMP domain	PF02926			0	
Ribosome 60S biogenesis N-terminal	PF11707			0	
Nucleolar pre-ribosomal-associated protein 1	PF16201			0	
Uncharacterized conserved protein	PF09749			0	
Formate--tetrahydrofolate ligase	PF01268			0	
Ubiquitin-like modifier-activating enzyme ATG7 N-terminus	PF16420			0	
MYB-CC type transfactor, LHEQLE motif	PF14379			0	

Bao31/Bap29 transmembrane region	PF05529			0	
Nuclear pore complex scaffold, nucleoporins 186/192/205	PF11894			0	
GNAT acetyltransferase 2	PF13718			0	
Transcriptional regulator	PF01709			0	
Porphyromonas-type peptidyl-arginine deiminase	PF04371			0	
Neurochondrin	PF05536			0	
Surfeit locus protein 6	PF04935			0	
Methylmalonic aciduria and homocystinuria type C family	PF16690			0	
Dos2-interacting transcription regulator of RNA-Pol-II	PF14500			0	
rRNA-Processing protein Efg1	PF10153			0	
Transferase family	PF02458			0	
Gar1/Naf1 RNA binding region	PF04410			0	
Transcription factor Tfb4	PF03850			0	
Transcriptional Coactivator p15 (PC4)	PF02229			0	
TGU ubiquitin-like domain	PF11470			0	
Urb2/Npa2 family	PF10441			0	
Ribosomal protein S30	PF04758			0	
Chlorophyll A-B binding protein	PF00504			1.67E-12	
Mur ligase family, catalytic domain	PF01225			0	
Photosystem II reaction center X protein (PsbX)	PF06596			0	
ARP2/3 complex ARPC3 (21kDa) subunit	PF04062			0	
Ribosomal protein s7e	PF01251			0	
DNA directed RNA polymerase, 7kDa subunit	PF03604			0	
RAVE protein 1 C terminal	PF12234			0	
tRNA (Guanine-1)-methyltransferase	PF01746			0	
BRCT domain, a BRCA1 C-terminus domain	PF16589			0	
Telomerase ribonucleoprotein complex - RNA binding domain	PF12009			0	
Ribosomal protein 50S-L18Ae/60S-L20/60S-L18A	PF01775			0	

Nidix N-terminal	PF14803			0	
Dicer dimerization domain	PF03368			0	
Peptidoglycan biosynthesis/recognition	PF04339			0	
WD40-like Beta Propeller Repeat	PF07676			0	
6,7-dimethyl-8-ribityllumazine synthase	PF00885			0	
BOP1NT (NUC169) domain	PF08145			0	
Ubiquinol-cytochrome C reductase hinge protein	PF02320			0	
Nucleoplasmin/nuceophosmin domain	PF03066			0	
Whirly transcription factor	PF08536			0	
Vps16, N-terminal region	PF04841			0	
Fumarase C-terminus	PF05683			0	
56kDa selenium binding protein (SBP56)	PF05694			0	
RNA polymerase Rpb8	PF03870			0	
Ribosomal protein S27	PF01667			0	
Macrophage migration inhibitory factor (MIF)	PF01187			0	
Sas10 C-terminal domain	PF09368			0	
Protein of unknown function (DUF262)	PF03235			0	
Ribosomal L27e protein family	PF01777			0	
BING4CT (NUC141) domain	PF08149			0	
YgbB family	PF02542			0	
Spb1 C-terminal domain	PF07780			0	
Mitochondrial ATP synthase epsilon chain	PF04627			0	
Photosystem I reaction center subunit N (PSAN or PSI-N)	PF05479			0	
MnmE helical domain	PF12631			0	
Ferredoxin chloroplastic transit peptide	PF11591			0	
Apoptosis-antagonizing transcription factor, C-terminal	PF08164			0	
Ribosomal L39 protein	PF00832			0	
GTP-binding protein TrmE N-terminus	PF10396			0	

RNA polymerase III RPC4	PF05132			0	
TspO/MBR family	PF03073			0	
Peptidase family M28	PF04389			0	
DNA polymerase subunit Cdc27	PF09507			0	
Putative diphthamide synthesis protein	PF01866			0	
Ribosomal L22e protein family	PF01776			0	
Ribosomal protein L23, N-terminal domain	PF03939			0	
Photosystem I reaction center subunit VIII	PF00796			0	
AT hook motif	PF02178			0	
Ribonuclease III domain	PF00636			0	
5'-3' exonuclease, N-terminal resolvase-like domain	PF02739			0	
Regulator of volume decrease after cellular swelling	PF03517			0	
Nitroreductase family	PF00881			0	
Septin	PF00735			0	
Ribosomal L15	PF00827			0	
Deoxynucleoside kinase	PF01712			0	
EGF-like domain	PF00008			0	
Ribosomal L29e protein family	PF01779			0	
Diaminopimelate epimerase	PF01678			0	
Oxygen evolving enhancer protein 3 (PsbQ)	PF05757			0	
DNA polymerase phi	PF04931			0	
Rio2, N-terminal	PF09202			0	
Cytochrome oxidase c subunit VIII	PF02285			0	
Flavin of succinate dehydrogenase	PF03937			0	
UTP15 C terminal	PF09384			0	
CreA protein	PF05981			0	
Chromatin associated protein KTI12	PF08433			0	
biquinol-cytochrome C reductase, UQCRX/QCR9 like	PF05365			0	

NusB family	PF01029			0	
Oxoglutarate and iron-dependent oxygenase degradation C-term	PF10637			0	
LYAR-type C2HC zinc finger	PF08790			0	
Rad4 beta-hairpin domain 1	PF10403			0	
RNA polymerases N/8 kDa subunit	PF01194			0	
AICARFT/IMPCHase bienzyme	PF01808			0	
U3 small nucleolar RNA-associated protein 10	PF12397			0	
Quinolate phosphoribosyl transferase, C-terminal domain	PF01729			0	
Protein SCA1	PF12070			0	
Gaa1-like, GPI transamidase component	PF04114			0	
DNA-directed RNA polymerase I subunit RPA34.5	PF08208			0	
Pentapeptide repeats (8 copies)	PF01469			0	
Photosystem II 10 kDa polypeptide PsbR	PF04725			0	
Adenosylmethionine decarboxylase	PF01536			0	
Zinc-ribbon	PF16503			0	
TIM21	PF08294			0	
ThiC-associated domain	PF13667			0	
DNA replication complex GINS protein SLD5 C-terminus	PF16922			0	
Utp21 specific WD40 associated putative domain	PF04192			0	
Ribulose biphosphate carboxylase, small chain	PF00101			0.000599654	
Myosin-like coiled-coil protein	PF09728			0	
Photosystem I reaction center subunit IV / PsaE	PF02427			0	
Galactokinase galactose-binding signature	PF10509			0	
Rad4 transglutaminase-like domain	PF03835			0	
IMP-specific 5'-nucleotidase	PF06437			0	
Nup93/Nic96	PF04097			0	
SRP40, C-terminal domain	PF05022			0	
Ribosomal S17 N-terminal	PF16205			0	

Lyase	PF00206			0	
Tyrosyl-DNA phosphodiesterase	PF06087			0	
Ribosomal protein L35A	PF01247			0	
Photosystem I psaG / psaK	PF01241			0	
ATP-grasp domain	PF02222			0	
Plectin/S10 domain	PF03501			0	
MgtE intracellular N domain	PF03448			0	
RanBP1 domain	PF00638			0	
N-terminal domain of 16S rRNA methyltransferase RsmF	PF17125			0	
Ribosomal protein S28e	PF01200			0	
Erg28 like protein	PF03694			0	
Ribosomal protein S26e	PF01283			0	
TonB-dependent Receptor Plug Domain	PF07715			0	
Radical SAM ThiC family	PF01964			0	
Photosystem II 10 kDa phosphoprotein	PF00737			0	
WPP domain	PF13943			0	
Exosome complex exonuclease RRP4 N-terminal region	PF14382			0	
E2F-associated phosphoprotein	PF10238			0	
Glucokinase	PF02685			0	
CYTH domain	PF01928			0	
Methyltransferase small domain	PF05175			0	
PA14 domain	PF07691			0	
Pep3/Vps18/deep orange family	PF05131			0	
Pescadilo N-terminus	PF06732			0	
Ribosomal protein S21e	PF01249			0	
GPI-GlcNAc transferase complex, PIG-H component	PF10181			0	
Ferritin-like domain	PF00210			0	
Ycf4 (photosystem I)	PF02392			0	

NGPINT (NUC091) domain	PF08153			0	
rRNA processing/ribosome biogenesis	PF08167			0	
Cold-shock' DNA-binding domain	PF00313			0	
tRNA (Met) cytidine acetyltransferase TmcA, N-terminal	PF08351			0	
UPF0113 PUA domain	PF03657			0	
Thymidylate kinase	PF02223			0	
Ribosomal protein L34e	PF01199			0	
Ran-interacting Mog1 protein	PF04603			0	
Ribosomal protein L21e	PF01157			0	
Glycine cleavage H-protein	PF01597			0	
Cytochrome b-c1 complex subunit 8	PF10890			0	
Nop14-like family	PF04147			0	
EutQ-like cupin domain	PF05899			0	
Monogalactosyldiacylglycerol (MGDG) synthase	PF06925			0	
Phosphomannose isomerase type I C-terminal	PF01238			0	
Ribosome associated membrane protein RAMP4	PF06624			0	
Nitrogen regulatory protein P-II	PF00543			0	
BP28CT (NUC211) domain	PF08146			0	
Photosystem I reaction center subunit XI	PF02605			0	
Possible tRNA binding domain	PF13725			0	
CRM1 C terminal	PF08767			0	
Neurotransmitter-gated ion-channel transmembrane region	PF02932			0	
N-(5'phosphoribosyl) anthranilate (PRA) isomerase	PF00697			0	
CutA1 divalent ion tolerance protein	PF03091			0	
Polypeptide deformylase	PF01327			0	
Fibrillarin	PF01269			0	
SOS response associated peptidase (SRAP)	PF02586			0	
NOP protein chaperone 1	PF15370			0	

PPIC-type PPIASE domain	PF00639			0	
Mre11 DIA-binding presumed domain	PF04152			0	
Photosystem I reaction center subunit VI	PF03244			0	
Putative transmembrane family 234	PF10639			0	
Cupin	PF00190			0	
Sas10/Utp3/C1D family	PF04000			0	
Mpp10 protein	PF04006			0	
Utp13 speciic WD40 associated domain	PF08625			0	
Sec20	PF03908			0	
TAP42-like family	PF04177			0	
PetM family of cytochrome b6f complex subunit 7	PF08041			0	
Ribosomal L38e protein family	PF01781			0	
Photosystem I psaA/psaB protein	PF00223			0	
hosphatidate cytidyltransferase, mitochondrial	PF09139			0	
Putative RNA methyltransferase	PF02598			0	
Formiminotransferase domain, N-terminal subdomain	PF07837			0	
MAPEG family	PF01124			0	
Rad4 beta-hairpin domain 2	PF10404			0	
Fumarate hydratae (Fumerase)	PF05681			0	
VRR-NUC domain	PF08774			0	
Urm1 (ubiquitin related modifier)	PF09138			0	
60s ribosome biogenesis protein Rrp14	PF15459			0	
asein kinase 2 substrate	PF15011			0	
Zc3h12a-like Ribonuclease NYN domain	PF11977			0	
Rad4 beta-hairpin domain 3	PF10405			0	
Sulphur transport	PF04143			0	
K ⁺ potassium transported	PF02705			0	
RNA methyltransferase complex GCD14 subunit	PF08704			0	

Ribosome biogenesis regulatory protein (RRS1)	PF04939			0	
Utp14 protein	PF04615			0	
Ribonuclease HII	PF01351			0	
Uncharacterized conserved domain (SAYSvFN)	PF10260			0	
D123	PF07065			0	
Adenosine-deaminase (editase) domainn	PF02137			0	
Ribosomal protein L14	PF01929			0	
Photosystem I PsaD	PF02531			0	
NMT1/THI5 like	PF09084			0	
2H2 type zinc-finger (2 copies)	PF12756			0	
Cupredoxin-like domain	PF13473			0	
Manganese-stabilising protein / photosystem II polypeptide	PF01716			0	
Quinolate phosphoribosyl transferase, N-terminal domain	PF02749			0	
Nop53 (60S ribosomal biogenesis)	PF07767			0	
NOP5NT (NUC127) domain	PF08156			0	
dUTPase	PF00692			0	
Argininosuccinate lyase C-terminal	PF14698			0	
Lumenal portion of Cytochrome b559, alpha (gene psbE) subunit	PF00284			0	
CAAD domains of cyanobacterial aminoacyl-tRNA synthetase	PF14159			0	
Origin recognition complex (ORC) subunit 5 C-terminus	PF14630			0	
Nucleoporin complex subunit 54	PF13874			0	
Ribosomal L29 protein	PF00831			0	
N-terminal beta-sandwich domain of polyadenylation factor	PF16573			0	
Focadhesin/RST1, DUF3730	PF12530			0	
Down-regulated in metastasis	PF07539			0	
MoaE protein	PF02391			0	
Noc2p family	PF03715			0	
Apoptosis antagonizing transcription factor	PF13339			0	

Helicase	PF05127			0	
ATP synthase, Delta/Epsilon chain, long alpha-helix domain	PF00401			0	
Ribosomal L37ac protein family	PF01780			0	
Domain of unknown function (DUF4399)	PF14347			0	
Ham1 family	PF01725			0	
ATP11 protein	PF06644			0	
Caleosn related protein	PF05042			0	
Nucleolin binding domain	PF16725			0	
Eukaryotic rRNA processing protein EBP2	PF05890			0	
Domain of unknown function (DUF3381)	PF11861			0	
Photosystem II reaction center W protein (PsbW)	PF07123			0	
tRNA pseudouridine synthase D (TruD)	PF01142			0	
Proteasome assembly chaperone 3	PF10178			0	
AARP2CN (NUC121) domain	PF08142			0	
GNL3L/Grn1 putative GTPase	PF08701			0	
NTPase	PF03266			0	
Pre-mRNA cleavage complex II protein Clp1	PF06807			0	
BCCIP	PF13862			0	
Ribosomal protein L37e	PF01907			0	
PetN	PF03742			0	
40s ribosome biogenesis protein Tsr1 and BMS1 C-terminal	PF04950			0	
Adenylosuccinate lyase C-terminal	PF08328			0	
Exosome component EXOSC1/CSL4	PF10447			0	
HIT zinc finger	PF04438			0	
4F5 protein related disordered region	PF04419			0	
SURF4 family	PF02077			0	
5'-3' exonuclease, C-terminal SAM fold	PF01367			0	
Rab3 GTPase-activating protein regulatory subunit N-terminus	PF14655			0	

Table 4.1. Pfam enrichment of differentially expressed genes across induction conditions pf *rlsD* overexpression.

Highlighted entries represent Pfams of special interest.

Chapter 5 - Cell adhesion gene *fsl1* is important for multicellularity in colonial alga *Gonium pectorale*

Berenice Jiménez-Marín, Gabriela Shipman, Tara Marriage, Courtney Hill, John Tumberger,
Bradley J.S.C. Olson

Author contributions: BJM: Conceptualization, formal analysis, investigation, writing – original draft, visualization. GS: formal analysis, investigation, visualization, writing – original draft. TM: conceptualization, formal analysis, investigation. CH: formal analysis, visualization. JT: investigation, formal analysis. BJSCO: conceptualization, methodology, formal analysis, writing – review & editing, supervision, project administration, funding acquisition

Introduction

The evolution of multicellular organisms from unicellular ancestors is a Major Evolutionary Transition (John Maynard Smith, 1995). Multicellularity differs from other Major Transitions in that it has occurred across at least two domains of life independently (Grosberg and Strathmann, 2007; Knoll, 2011). The recurrent nature of multicellular evolution in disparate lineages and through different geological eras (King, 2004; Grosberg and Strathmann, 2007; Herron et al., 2009; Knoll, 2011) strongly suggests that the molecular toolkits required to evolve multicellular lifestyles might have existed in ancestral unicellular organisms (Grosberg and Strathmann, 2007).

The volvocine algae are a model system that is well suited for studying the evolution of multicellularity. Multicellularity in the volvocine lineage occurred relatively recently (~240 MYA) (Herron et al., 2009), and member species exhibit phenotypes ranging from unicellular to multicellular undifferentiated and to multicellular differentiated. Multicellular volvocine species are thought to have evolved from a unicellular ancestor resembling extant *Chlamydomonas reinhardtii* (*Chlamydomonas* henceforth). Furthermore, genomic sequencing of *Chlamydomonas* and of key multicellular volvocines, such as undifferentiated multicellular *Gonium pectorale* (*Gonium*) and differentiated multicellular *Volvox carteri* (*Volvox*) among others (Merchant et al., 2007; Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2017; Hamaji et al., 2018; Jiménez-Marín et al., 2021) show that their genomes are similar and syntenic, further supporting the idea that the genetic tool kit required for multicellularity was present in the unicellular ancestors of multicellular volvocines, and that a small number of regulatory and/or genetic changes are required to generate novel phenotypes within this clade (Olson and Nedelcu, 2016).

Cell adhesion in multicellular volvocines is composed of two structures: cytoplasmic bridges – believed to result from incomplete cytokinesis and that possibly hold the organism through its early development (Iida et al., 2013), and the hydroxyproline-rich glycoprotein-based extracellular matrix (ECM), which is closely related to the cell wall of unicellular *Chlamydomonas* (Adair et al., 1987; Adair and Appel, 1989; Hallmann, 2006b). The biochemical conservation of the ECM, in addition to the high similarity between the life cycles of unicellular *Chlamydomonas* and undifferentiated multicellular *Gonium* (Umen and Olson, 2012; Hanschen et al., 2016), allow for the possibility that the molecular players underlying cell adhesion in basal multicellular *Gonium* were co-opted from their *Chlamydomonas*-like ancestor to be developmentally regulated.

Analysis of the *Gonium* genome sequence shows that a relatively small number of genes have signatures of positive selection in *Gonium* as compared to *Chlamydomonas* (Hanschen et al., 2016). *Fasciclin-like 1 (Fsl1)* is a gene encoding fasciclin domains, which are conserved across eukaryotic lineages and have a role in cell adhesion (Johnson et al., 2003; Packard et al., 2003). With orthologs in *Chlamydomonas* and *Volvox*, *Fsl1* is positively selected in *Gonium*, making it a candidate cell adhesion factor. The divergence resulting from positive selection between the *Chlamydomonas* and the *Gonium* orthologs of *Fsl1* allow for the possibility that each species has evolved differential uses for its product, such as maintenance of colonial integrity through adhesion. In this work, we use reverse genetics approaches in *Chlamydomonas* and, for the first time, in *Gonium*, to demonstrate that *Fsl1* was likely co-opted in the multicellular volvocines for cell-cell adhesion. A combination of gain-of-function and loss-of-function experiments in *Chlamydomonas* and *Gonium* respectively suggest that *Fsl1* in *Gonium* is necessary for multicellularity .

Results

fsII* is an ortholog of a cell adhesion molecule in *Volvox* with differential regulation in *Chlamydomonas

As means to identify candidate homologs of *Gonium*'s *FsII* (*GpFsII*), fasciclin-domain containing genes in *Chlamydomonas reinhardtii*, *Gonium pectorale*, *Yamagishiella unicocca*, *Eudorina* sp. and *Volvox carteri* were queried against available Pfam annotations (Jiménez-Marín et al., 2021) and queried against the NCBI conserved domain search web tool. All species have fasciclin encoding genes, with *Yamagishiella* having the highest number (28 genes), and *Volvox* the smallest number (14 genes, Table 5.1). Reciprocal BLASTp of the predicted amino acid sequence for *GpFsII* against the *Volvox* assembly version 2.1 showed that *GpFsII* is a candidate ortholog of *Volvox*'s algal cell adhesion molecule (algal-CAM). Algal-CAM is considered the first discovered plant homolog of animal cell-adhesion proteins (Huber and Sumper, 1994). Interestingly, the identity between *GpFsII* and algal-CAM is very low (32%); this divergence is likely the consequence of positive selection acting on each species' ortholog, as the gene for algal-CAM was found to also have signatures of positive selection relative to *Chlamydomonas* (Figure 5.2) (Hanschen et al., 2016). Similarly, the sequence identity between *GpFsII* and its *Chlamydomonas* ortholog, *CrFSLI*, is low (35%) (Figure 5.2).

To assess the possibility that *GpFsII* was co-opted for multicellular development in *Gonium*, the transcriptional patterns of *GpFsII* and its candidate *Chlamydomonas* ortholog (*CrFSLI*) were compared using transcriptomic data from the cell cycle of both species (personal communication with Rakijas, J.B.). While *GpFsII* is upregulated during G1 and then down-regulated during mitosis in *Gonium*, *CrFSLI* is upregulated during mitosis, and downregulated

during most of G1 (Figure 5.1). These results suggest that *GpFslI* might have become co-opted for cell-cell adhesion in *Gonium* through rewiring of its expression pattern and rapid divergence.

Antisense RNA of *GpFslI* results in loss of multicellularity in *Gonium*

To gain insight into the role of *GpfsII* in *Gonium*'s development, *GpFslI* expression was reduced with antisense RNA. Surprisingly, the resulting antisense *GpFslI* lines lost their multicellularity by 70% or more in most lines (Figure 5.3, 5.4). No other phenotypic alterations were identified in unicellular *Gonium* strains. This is the first instance to our knowledge of genetic manipulation of *Gonium* using RNA interference, as well as the first proof that *Chlamydomonas* promoters can be used to drive gene expression in *Gonium* (see methods). Furthermore, these results suggest that *GpFslI* is necessary for *Gonium* to develop as a multicellular organism.

To confirm that unicellular *Gonium* lines were *GpFslI* knockdowns, RNA samples were collected from three transformant lines and the WT strain. qPCR analysis determined that unicellular lines have reduced *GpFslI* transcript accumulation relative to the WT (Figure 5.5). These results show that *GpFslI* is necessary for multicellularity in *Gonium*.

Heterologous expression of *GpFslI* produces multicellular *Chlamydomonas*

As means to further understand the role of *GpFslI* in multicellularity, gain-of-function experiments using unicellular, wild-type (WT) *Chlamydomonas* strain 21gr were performed via transformation with *GpFSL1-C-3xHA*, which is under the putative native promoter for *GpFslI* (see Methods). In lines expressing *GpFslI*, we observe multi-celled individuals (Figure 5.6, 5.7), despite expression being low (data not shown). Upon inspection of the multicellular individuals, the *Chlamydomonas* cells appear to be connected rather than adhered to each other (Figure 5.6). Moreover, the connections between cells are reminiscent of cytoplasmic bridges, which are

structures that link adjacent cells in a variety of multicellular volvocines (Bisalputra and Stein, 1966; Green et al., 1981; Hoops et al., 2006; Iida et al., 2013).

Cytokinesis is disrupted in gain-of-function cell-wall compromised *Chlamydomonas* strains

The differences in sequence and regulation between *GpFslI* and *CrFSLI* suggest their respective products perform different functions. However, as means to eliminate the possibility of dosage effects on the *Chlamydomonas* gain-of-function phenotype, a *Chlamydomonas* knock-out strain of *CrFSLI* (*CrfsII*) was also co-transformed with *GpFSLI*-C-3xHA and pHyg3. The *CrfsII* strains were produced as part of the *Chlamydomonas* library project (CLiP), which generated insertional mutants for most of the genes in the genome of *Chlamydomonas* strain CC-5325 (5325 henceforth). The 5325 strain has a thinner, likely compromised cell wall compared to cell walled WT strains such as 21gr (Zhang et al., 2022). In transformant lines, large, multiflagellate individuals were observed (Figure 5.8, 5.9). Upon co-transforming 5325 cells, large, multiflagellated individuals were observed in comparable numbers to *GpFslI*-*CrfsII* gain-of-function lines (Figure 5.8, 5.9).

As part of their life cycle, *Chlamydomonas* and *Gonium* retract their flagella upon undergoing mitosis, since their basal bodies act as both flagella organizing centers and centrioles (Koufopanou, 1994; Cross and Umen, 2015). Furthermore, *Chlamydomonas* cultures can be synchronized to laboratory-generated day/night cycles as means to track their development. Thus, cultures of *GpFslI*-*CrfsII*, *GpFslI*-5325, and control *CrFslI* and 5325 strains were synchronized and their mitotic status was observed using light microscopy to evaluate whether multiflagellated cells were the product of fusion of distinct individuals or of failed cytokinesis. Post-mitotic clusters of both *GpFslI*-*CrfsII* and *GpFslI*-5325 display a mixture of WT daughter

cells, and larger daughter cells with distinct sets of organelles, suggesting these cells failed cytokinesis (Figure 5.8). qPCR analysis of transformant lines in both strains confirmed *GpFslI* is expressed, albeit minimally (data not shown). In light of the similar phenotype displayed by *GpFslI*-*CrfsII* and *GpFslI*-5325, and of the phenotype observed in *GpFslI*-21gr, *GpFslI* is unlikely to interact with *CrFSLI*. Moreover, *GpFslI* probably interacts with the cell wall in both *Gonium* and *Chlamydomonas*, such that upon undergoing cytokinesis, cells remain connected by a small structure (Figure 5.3), rather than fail to separate altogether (Figure 5.9).

Discussion

Modifications to the cell-cycle molecular machinery are known to have been co-opted for multicellularity in the volvocine lineage. One consequence of this is the differential disassociation of daughter cells between *Chlamydomonas* and *Gonium*, which ultimately result in the former generating individual cells and the latter maintaining clusters of cells that constitute daughter colonies (Olson and Nedelcu, 2016). The incomplete cytokinesis of *GpfsII* gain-of-function *Chlamydomonas* strains provides an avenue to infer how this differential disassociation might have occurred.

Incomplete cytokinesis is thought to be the mechanism underlying the formation of cytoplasmic bridges (Bisalputra and Stein, 1966; Green et al., 1981; Hoops et al., 2006; Iida et al., 2013). Interestingly, the cytoplasmic bridges appear to have a role in spheroidal volvocine development (e.g. *Volvox*) by maintaining the colony together as it reconfigures itself from its juvenile to its adult conformation – a process called inversion (Hoops et al., 2006). Shortly after inversion is completed, in most species the bridges disappear and colonial cohesion is maintained by ECM deposition (Hoops et al., 2006). It is unclear whether cytoplasmic bridges also play a role in *Gonium*'s inversion; however, *Gonium*'s cytoplasmic bridges also disappear after this

morphogenetic event is completed (Iida et al., 2013). The failed cytokinesis phenotype observed in *Chlamydomonas* *GpFsII* gain-of-function cell lines (Figure 5.6, 5.8) suggest that this gene might be involved in the formation of *Gonium*'s cytoplasmic bridges. Studies on the *Volvox* ortholog of *GpfsII*, algal-CAM, further support this possibility, as anti-algal-CAM antibodies given at embryogenesis can inhibit the number of cell divisions as well as inversion (Huber and Sumper, 1994). Moreover, cell-wall deficient gain-of-function *Chlamydomonas* transformed with algal-CAM have been observed to form large aggregates (Lowder and Herbert, 2014).

GpFsII might represent an instance of genetic co-option through differential regulation and later divergence that plays a role not only in the development of spheroidal volvocines, such as *Volvox*, but in volvocines with alternate body plans such as convex-sheet *Gonium*, which undergoes partial inversion (Kirk, 2005; Iida et al., 2013). Inversion, be it partial or complete, is a fundamental developmental process to multicellular volvocines, since it corrects for the maladaptive configuration of juvenile volvocine colonies (Hallmann, 2011). Should the product of *GpFsII* be involved in the biogenesis of cytoplasmic bridges, it will be a fundamental multicellularity gene with the multifaceted role of maintaining colonial integrity prior to inversion and deposition of the ECM, and in the establishment of organismal polarity as well through inversion.

The relationship between the cell wall/ECM and *GpFsII* associated phenotypes warrants further study. Algal-CAM is predicted to be membrane associated, but not transmembrane (Huber and Sumper, 1994), which might also be true for *GpFsII*. Should this be the case, it is likely that the disrupted cell wall from the *Chlamydomonas* 5325 and *CrfsII* backgrounds aggravates the cell-adhesion phenotype of *GpFsII* gain-of-function lines. Furthermore, the similarity in phenotype and phenotype frequency between *GpFsII*-5325 and *GpFsII*-*CrfsII*

suggest that *CrFSL1* is not interacting with *GpFsl1* nor contributing to the cell adhesion phenotypes observed in these strains (Figure 5.8), and likely under the 21gr background. Also, the low frequency at which multicellular and failed cytokinesis phenotypes are observed among gain-of-function *Chlamydomonas* are observed might be a consequence of low transcription of *GpFsl1*. Future work using this gene under constitutive expression promoters, such as the *Chlamydomonas* HSP70/RBCS2 promoter (Schroda et al., 2000) might result in more consistent multicellular phenotypes. Moreover, since *GpFsl1* might be involved in the formation of cytoplasmic bridges, cell localization and synchrony experiments in *Gonium* might provide valuable insight into the function and developmental role of *GpFsl1*. Regardless, these results as a whole show that *GpFsl1* is necessary for multicellularity in *Gonium*.

Methods

Strains and culture conditions

Gonium pectorale K3-F3-4 MT⁻ NIES-2863 and K4-F3-4 MT⁺ were grown in standard *Volvox* media supplemented with acetate (SVMA) (Starr, 1968; Kirk and Kirk, 1983; Kirk and Kirk, 1985), and bubbled with air under light at 400 $\mu\text{mol m}^{-2} \text{sec}$ photosynthetically active radiation. *Chlamydomonas* strains CC-5325 and *Crfs11* (LMJ.RY0402.163350) were obtained from the Chlamydomonas Library Project (CLiP). These strains were maintained on TAP-agar plates when not utilized. For experimental purposes, each strain was grown TAP media with air bubbling and under light at 400 $\mu\text{mol m}^{-2} \text{sec}$ photosynthetically active radiation.

Cell cycle synchrony assay

Chlamydomonas reinhardtii 5325 and *Crfs11* strains were assayed for synchrony as per previous published works (Fang et al., 2006; Olson et al., 2010; Zones et al., 2015). Synchronous culture

samples were aliquoted at roughly 10h into the light cycle, plated onto TAP plates and placed in the dark for overnight dark-shifting. Post-mitotic cells were observed under the light microscope after the dark shift.

Cloning

The coding sequence for *Gpfs11* was obtained as follows. Total RNA was extracted from asynchronous *Gonium* K4 cultures and then reverse transcribed with Superscript IV (ThermoFisher) using anchored oligo(dT). Reverse transcribed cDNA was then PCR amplified with primers for *Gpfs11* (Table 5.2), using Phusion polymerase (New England Biolabs). A single band was excised from an agarose gel, purified and cloned into CloneJET 1.2 (ThermoFisher). Identity of *Gpfs11* cDNA was verified by Sanger sequencing and named P3-3. The *Chlamydomonas* constitutive promoter HSP70/RBCS2 (Schroda et al., 2000) was PCR amplified from the psi103 construct using Phusion polymerase (New England Biolabs). The HSP70/RBCS2 amplicon was verified by gel electrophoresis and purified using the DNA clean and Concentrator Kit (Zymo).

GpFsl1-hi-anti was constructed as follows. The coding sequence for *Gpfs11* and the HSP70/RBCS2 promoter were PCR amplified using Phusion polymerase (New England Biolabs) from P3-3 and pClover-Hyg, a plasmid containing the HSP70/RBCS2 promoter originally from the psi103 plasmid. Both fragments were Gibson cloned into a PCR-linearized pHyg3 vector such that the *Gpfs11* sequence would be inverted (Figure 5.10 A) using the NEBuilder HiFi DNA assembly Kit (New England Biolabs) as per manufacturer's instructions. Successful insertion of the promoter and of the gene were tested by restriction digestion and Sanger sequencing.

The genomic sequence for *Gpfs11* was obtained by PCR from *Gonium* K3 genomic DNA using primers 1 kb upstream of the gene start and 500 bp downstream of the gene (table 5.2) end

to amplify the gene's putative native promoter and terminator (Figure 5.10 B). PrimeStar GXL polymerase (Takara) was used. The gene was cloned into CloneJET 1.2 (ThermoFisher). GpFSL1-C-3xHA was constructed as follows. P1-2 was linearized and tagged through inverse PCR with one of the primers encoding for a C-terminal triple haemagglutinin (3xHA) tag (Table 5.2) using PrimeStar polymerase (Takara). A single band was excised from an agarose gel, and purified using the Zymoclean Gel DNA recovery kit (Zymo). The PCR amplicon was digested with EcoRI and ligated using T4 DNA ligase (New England Biolabs). Successful insertion of the tag was confirmed by PCR and Sanger sequencing.

Nuclear transformation of *Gonium* and *Chlamydomonas*

Axenic cultures of WT *Gonium* were grown in flasks containing 300 mL SVMA (see culture conditions) for 4-5 days prior to transformation. Tween-20 was added to a concentration of 0.005% and cultures were centrifuged at 1,000*G* for 5 minutes. The bulk of the supernatant discarded and the pellet was then resuspended into 50 mL cultures using SVMA with sorbitol at a concentration of 50 mM. Chlorophyll concentration assays (Ni et al., 2009) were performed to approximate culture density and for calculation of resuspension volumes. Resuspended cultures were then centrifuged at 600*G* for 5 minutes, the supernatant was decanted, and the cells were resuspended into 10 mL of SVMA+sorbitol. Next, 10 mL of *Chlamydomonas*-derived autolysins were added to the cells for cell wall permeabilization. The cells were autolysin treated for 1 hour under light.

Autolysin-treated cells were centrifuged at 600*G* for 3 minutes and the supernatant was discarded. The pellet was gently resuspended and washed with 5 mL Max Efficiency Transformation Reagent for Algae (Invitrogen) and incubated for two minutes at room temperature. Then, the cells were centrifuged at 600*G* for 3 minutes and resuspended in Max

Efficiency Buffer to a concentration of 800 ug/mL of chlorophyll. Aliquots of treated cells (300uL) were then pipetted onto 4mm-wide electroporation cuvettes. Empty vector and GpFsl1-hi-anti plasmids were linearized using NdeI. A total of 1 ug of linearized plasmid was used per experimental sample and pipetted gently into the cell-containing cuvettes to be electroporated using a NEPA21 Super electroporator (Nepagene). Electroporation settings can be found in Table 5.3.

After electroporation, cells are left to rest for 10 minutes before getting transferred to culture tubes containing 10 mL of SVMA+Sorbitol. The cultures were then subjected to a 48-hour recovery period under ambient light and gentle shaking (100 rpm). Post-recovery cultures were then centrifuged at 600G for 3 minutes, the supernatant was decanted, and the pellet was resuspended in 2.1 mL SVMA. The resuspended cells then were plated onto SVMA-0.75% agar plates as well as into SVMA-0.75% agar plates with hygromycin at a final concentration of 2.5ug/mL for selection.

Transformation of *Chlamydomonas* cells by electroporation was done as per (Yamano et al., 2013).

RNA interference and gain-of-function experiments

Knock-down of GpFsl1 expression using antisense-RNA was performed as follows. The inverted GpFsl1 coding sequence was cloned into GpFsl1-hi-anti, which is under the *Chlamydomonas* constitutive promoters HSP70/RBCS2 (Schroda et al., 2000). This construct also contains a hygromycin-resistance cassette. WT *Gonium* K3 were transformed via electroporation and grown in standard *Volvox* media complemented with acetate (SVMA) and hygromycin at a concentration of 2.5 ug/mL.

Chlamydomonas strains 21gr, 5325, and *CrfsII* were transformed with construct GpFSL1-C-3xHA, which contains a recombinant Gp*FsII* sequence encoding for a C-terminal triple HA tagged GpFSL1 to be expressed under Gp*FsII*'s putative promoter and terminator region. This construct was co-transformed with a hygromycin-resistance conferring plasmid (pHyg3). All *Chlamydomonas* strains were transformed via electroporation and grown in Tris acetate phosphate media (TAP) with hygromycin at 10 ug/mL for selection.

RNA methods and qPCR

RNA was extracted from WT *Gonium*, knock-down *Gonium* strains *antifsl*-C12, *antifsl*-D8 and *antifsl*-H9, *Chlamydomonas* strains 21gr, 5325, *CrfsII*, and gain-of-function lines Gp*fsI*-5325-1A, Gp*fsI*-5325-1E, Gp*fsI*-*CrfsII*-4C, Gp*fsI*-21gr-1, Gp*fsI*-21gr-2, Gp*fsI*-21gr-4 using the Quick-RNA Plant Miniprep Kit (Zymo) in quintuplicate. A total of 200 ng of RNA per extraction were utilized for cDNA synthesis using the Protoscript II First Strand cDNA Synthesis Kit (New England Biolabs). All qPCR analyses were performed with Genious 2x SYBR Green Fast qPCR Mix (low ROX) (AB clonal) on an Applied Biosystems 7500 Real-Time PCR instrument. The qPCR was performed in quintuplicates. A melting curve for PCR was produced through the delta-delta CT method, and examined to verify that no primer-dimers or additional products were present.

Microscopy analysis

Light microscopy images of *Gonium* K3, Gp*fsII* knock-down *Gonium* lines, *Chlamydomonas* 21gr, 5325, *CrfsII*, Gp*fsII*-21gr, Gp*fsII*-5325, and Gp*fsII*-*CrfsII* were acquired on a Leica DM500 microscope with the 40x objective with a mounted Leica ICC50 HD camera. Phenotypic scoring was performed on samples from cultures that were immediately fixed in 3% glutaraldehyde, 1X PBS, 0.1% tween-20. Quantification of unicellular and multicellular/aberrant

phenotypes in *Gonium* and *Chlamydomonas* strains and cell lines were calculated by performing triplicate counts per line/strain using a hemocytometer.

Figures and tables

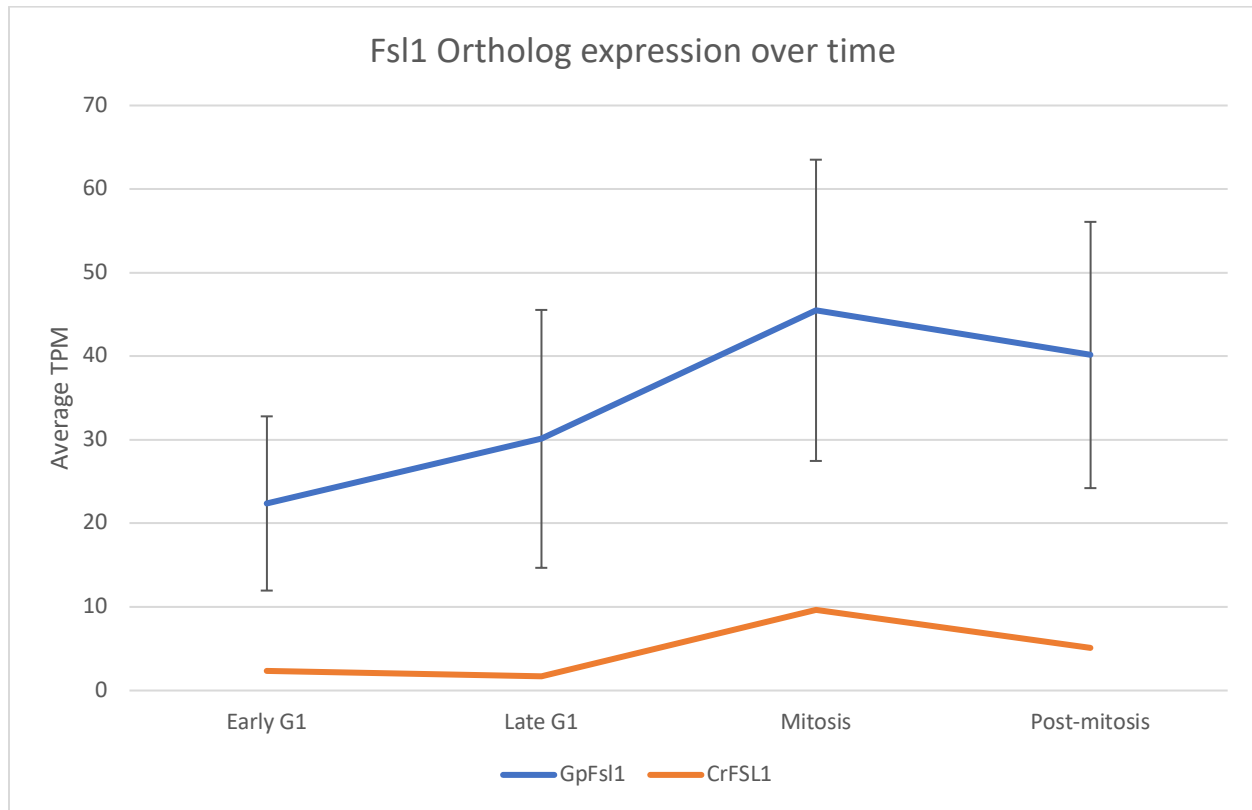


Figure 5.1. Expression patterns of *Gpfs11* and *CrFSL1* over the life cycle of *Chlamydomonas* and *Gonium*.

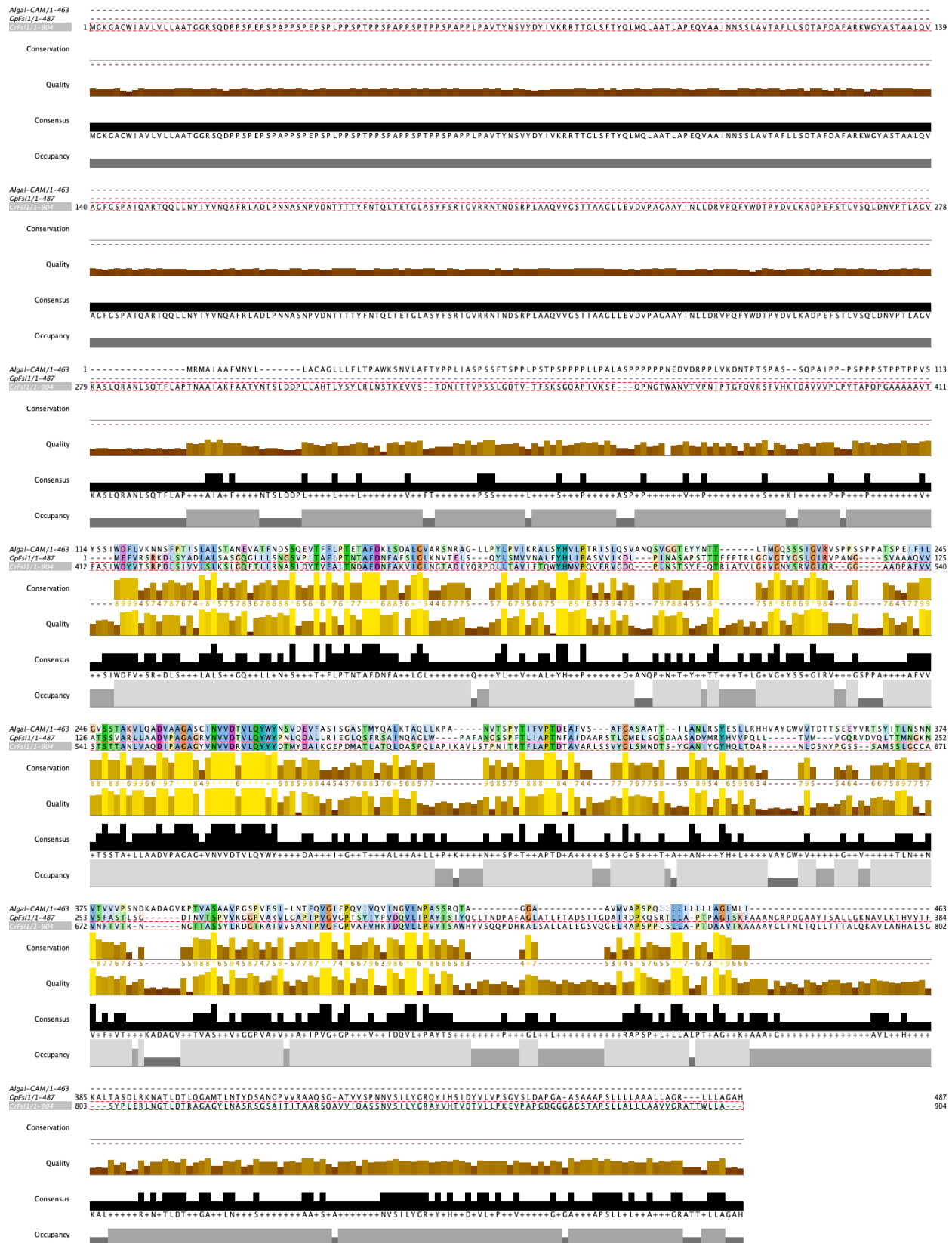


Figure 5.2. Amino acid conservation of orthologs of *GpFsl1* in *Volvox* and *Chlamydomonas*.

CrFSL1, GpFsl1, and Algal-CAM share little amino acid conservation.

Gpfs1 knock-down

WT

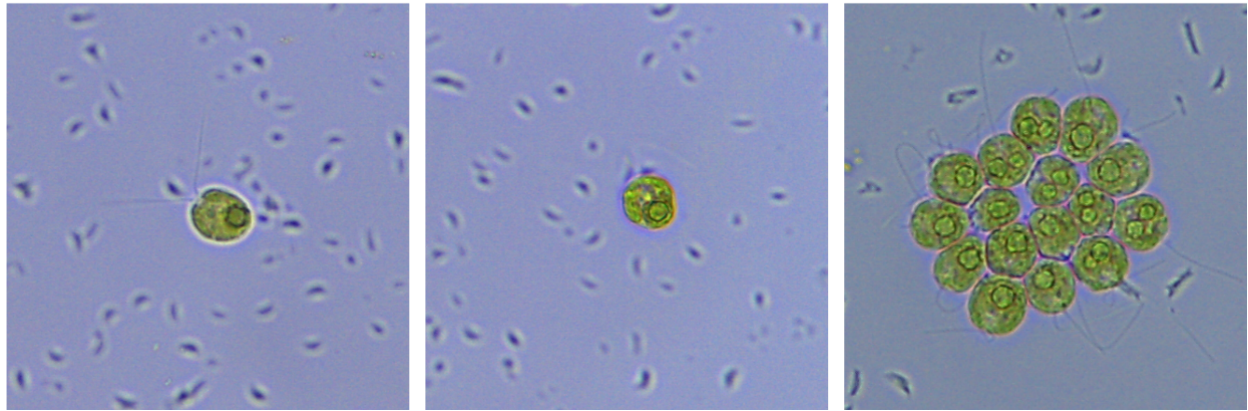


Figure 5.3. Gpfs1 knock-down *Gonium* are unicellular.

First and second panel shows knock-down unicellular *Gonium* cells, third panel shows a WT, 16-cell *Gonium* individual.

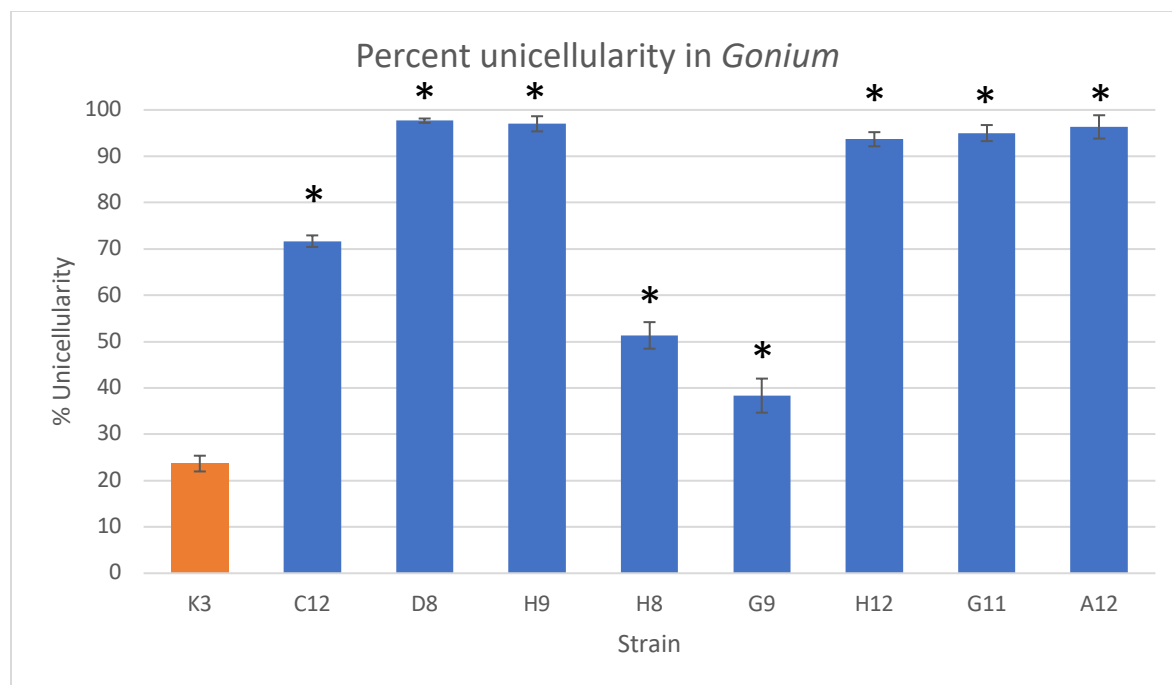


Figure 5.4. Percentage unicellularity in WT *Gonium* (K3) and *GpFslI* knock-down *Gonium* lines.

All loss-of-function cell lines (blue) are significantly more unicellular than the WT (blue)
($P < 0.05$ T-test)

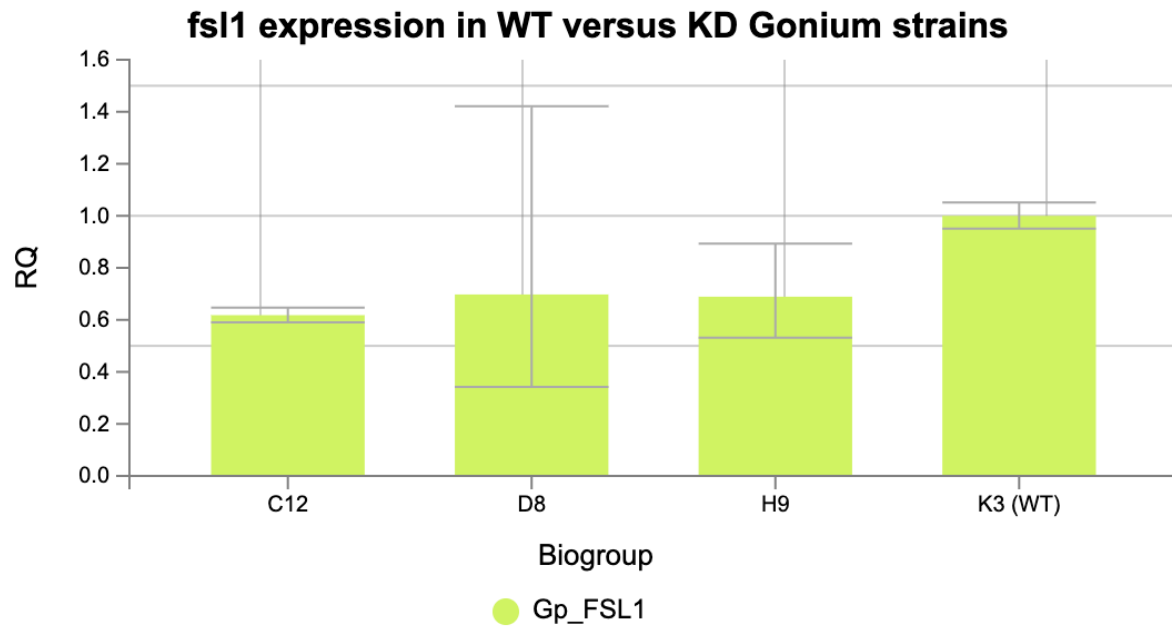


Figure 5.5. *GpFsl1* knock-down lines have reduced amounts of transcript relative to WT *Gonium* (K3).

RQ stands for relative quantity.

Gpfs1-21gr

WT

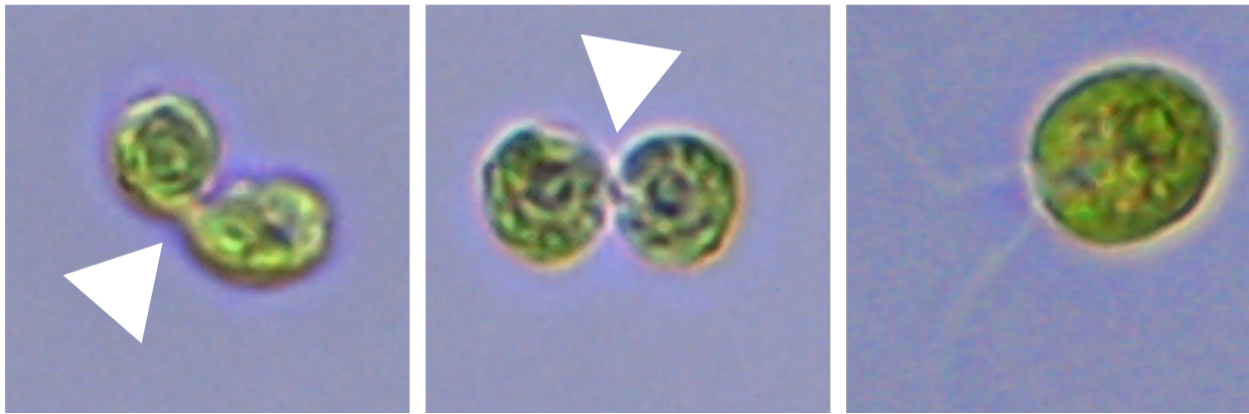


Figure 5.6. *Gpfs1* gain-of-function *Chlamydomonas* 21gr can be multicellular.

First and second panel shows *Chlamydomonas* gain-of-function multicellular phenotype under the 21gr background, third panel shows a WT, unicellular, *Chlamydomonas* cell. Cell-cell connections are indicated with white arrows.

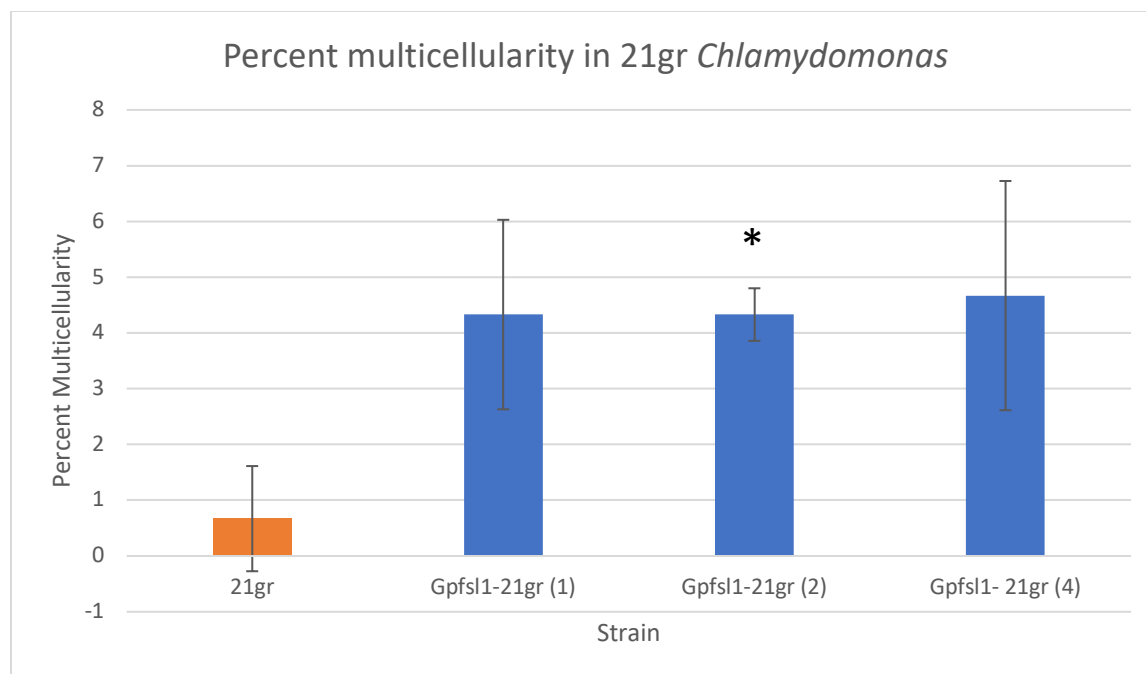


Figure 5.7. Percentage multicellularity in WT *Chlamydomonas* (21gr) (orange) and *Gpfs1* gain-of-function 21gr lines (blue).

One gain-of-function cell line was significantly more multicellular than the WT ($P < 0.05$ T-test)

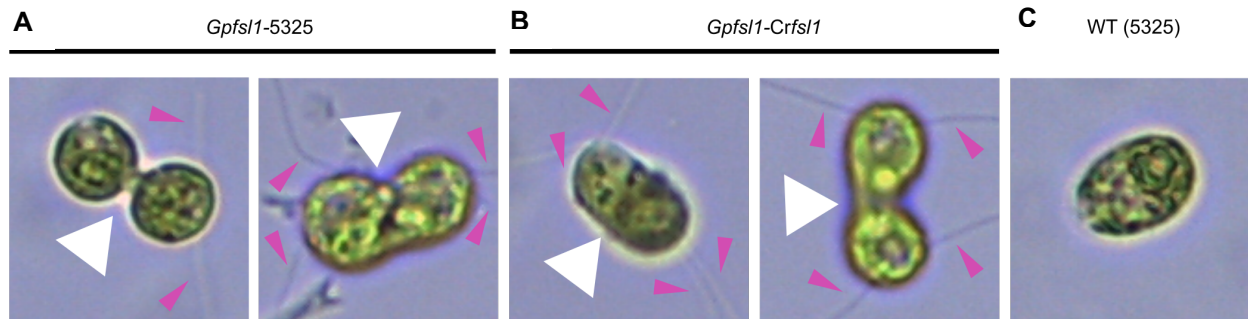


Figure 5.8. *Gpfs1* gain-of-function *Chlamydomonas* 5325 and *Crfs1* strains fail cytokinesis.

A) *Chlamydomonas* gain-of-function multiflagellated phenotype under the 5325 background. B)

Multiflagellated phenotype for *Crfs1*, a strain lacking the *Chlamydomonas* ortholog of *Gpfs1*.

C) Unicellular *Chlamydomonas* strain 5325. Cell-cell connections are indicated with white arrows, and magenta arrows indicate the position of flagella, when visible.

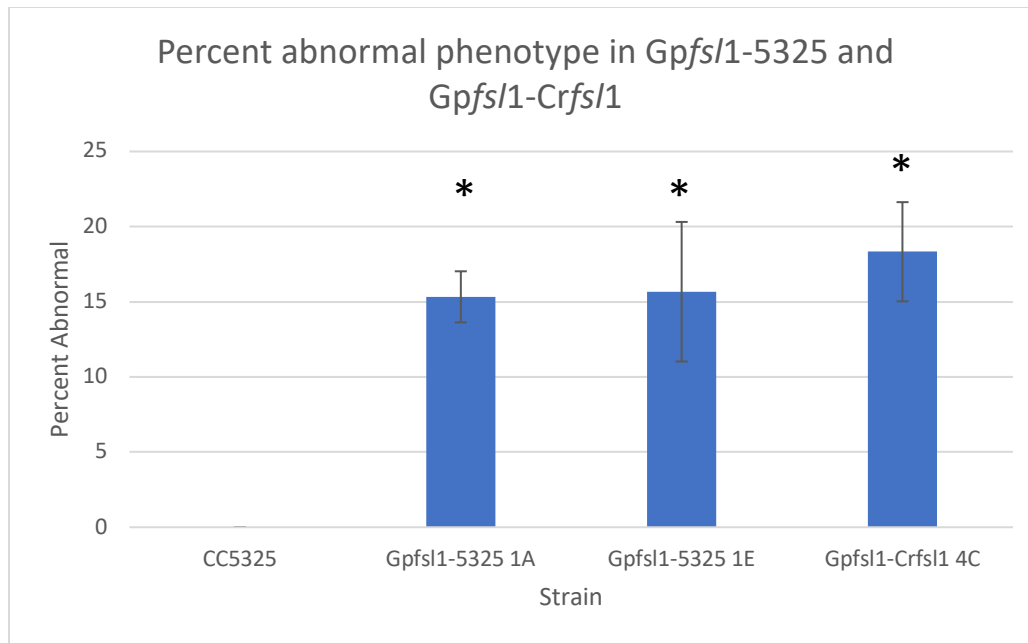


Figure 5.9. Percentage multicellularity in WT *Chlamydomonas* (5325) and Gp*Fsl1* gain-of-function 5325 and Cr*fs1* lines (blue).

All gain-of-function cell lines had significantly more abnormal cells than the WT ($P < 0.05$, T-test)

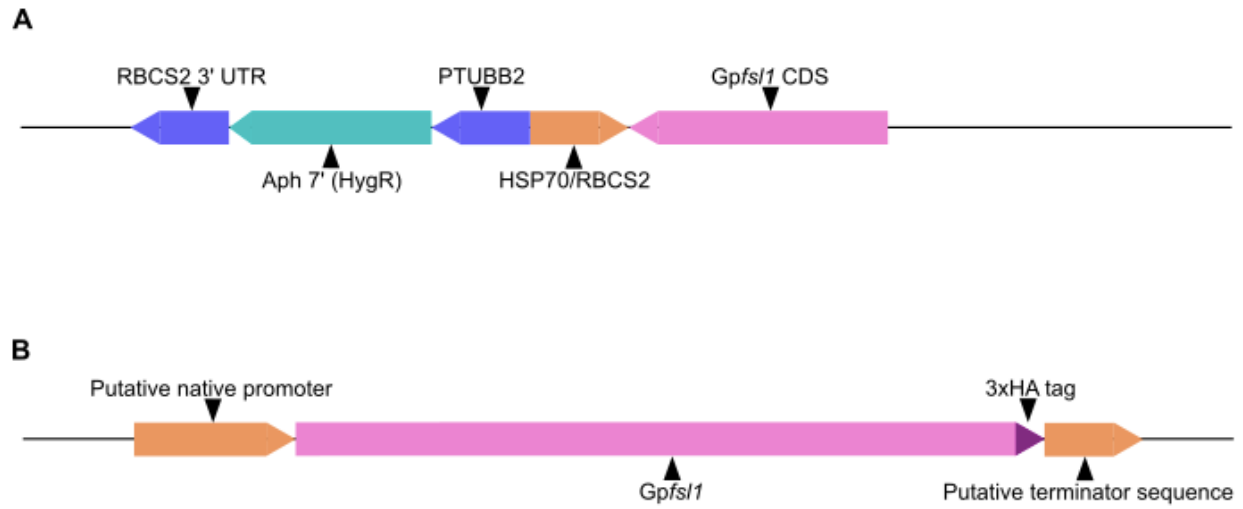


Figure 5.10. Construction of A) GpFsl1-hi-anti B) GpFSL1-C-3xHA.

Species	# genes
<i>Chlamydomonas</i>	21
<i>Gonium</i>	17
<i>Yamagishiella</i>	28
<i>Eudorina</i>	19
<i>Volvox</i>	14

Table 5.1. Fasciclin-encoding gene counts for five volvocine species.

Primer Name	Primer Sequence	Use of Primer
GpfsI1 qPCR F	CCAACACCGCCTTTGACAAC	qPCR for <i>fsI1</i> gene to test expression of antiFSL strains
GpfsI1 qPCR R	GTTGACCACCATGGACAGGT	qPCR for <i>fsI1</i> gene to test expression of antiFSL strains
GpFSL1-3XHA-F	AGAATTCGCGCGCGGGGACGGAGCC	Inverse PCR of GpFSL1 NA C-3xHA
GpFSL1-3XHA-R	GGAATTCTCAAGCGTAATCTGGAAC GTCATATGGATAGGATCCTGCATAG TCCGGGACGTCATAGGGATAGCCCG CATAGTCAGGAACATCGTATGGGT AGTGCGCCCCGGCCAGCAGGA	Inverse PCR of GpFSL1-C-3xHA with 3xHA tag included
GpFSL1 start F	GAACATCATGGAGTTTGTACGGAGCC	CDS obtention of <i>fsI1</i>
GpFSL1 stop R	CTCAGTGCGCCCGGCCAG	CDS obtention of <i>fsI1</i>
HA-forward	GACTATGCAGGATCCTATCCATATGACGT	For colony PCR of the 3xHA tag in the ligated GpFSL1-C-3xHA
GpScaffold0000.g730_500bp_down	CAAATCCAGGCCACAGGAAGAGC	For colony PCR of the 3xHA tag in the ligated GpFSL1-C-3xHA
GpFUS1AF	TACCCTGATGGTCGTGCGGAAC	Mating type for the FUS gene in <i>Gonium</i>
GpFUS1CR	GCGGGATCACTCAGAGGTGTT	Mating type for the FUS gene in <i>Gonium</i>
GpMID_Set4_Reverse	ATAGGAACTGATGTCTCGGCT	Mating type for the MID gene in <i>Gonium</i>
GpMID_Set4_Forward	ACGTCTTACGCTCGACCTTCA	Mating type for the MID gene in <i>Gonium</i>
MTP2F	GCTGGCATTCTGTATCCTTGACGC	Mating type “plus” primer for <i>Chlamydomonas</i>
MTP2R	GCGGCGTAACATAAAGGAGGGTCG	Mating type “plus” primer for <i>Chlamydomonas</i>
MTM3R	CTCGGCCAGAACCTTTCATAGGGTGG	Mating type “minus” primer for <i>Chlamydomonas</i>
MTM3F	CGACGACTTGGCATCGACAGGTGG	Mating type “minus” primer for <i>Chlamydomonas</i>
R7	CGGGGGAGGGTCAGAAAAAC	Genotyping for GpfsI1 <i>Chlamydomonas</i> transformants
F7	GGATGGATCAGGGCAGATTGTGAG	Genotyping for GpfsI1 <i>Chlamydomonas</i> transformants
Pjet_F	CGACTCACTATAGGGAGAGCGGC	Used for sequencing of the GpfsI1-3xHA tag junction
F14	GCCCTCCCTCGTCGCTCAG	Used for sequencing of the GpfsI1-3xHA tag junction
F15	GCCACCACAAACACAGTCG	Genotyping for GpfsI1 <i>Chlamydomonas</i> transformants
R15	CAACGACCCCTCTCTCC	Genotyping for GpfsI1 <i>Chlamydomonas</i> transformants
Hyg_HSP_70_Gibson_F	AAGCTTGATATCGAATTCGCT GAGGCTTGACAT	Amplification of HSP70/RBCS2 promoter region
RBCS_revFSL_Gibson_ant_R	GCGCACTGATTTAAGATGTT GAGTGACTTC	Amplification of HSP70/RBCS2 promoter region

229R	GTAATCATGGTCATAGCTGTT	Linearizing pHyg3
230F	GAATTCGATATCAAGCTTCTTTC	Linearizing pHyg3
AntiFSLHi_Seq_FSLend_F	GTCCACCACGACCTTCTT	Confirmation of correct insertion and cloning
AntiFSLHi_Seq_bbone_R	TACACCGAACTGAGATACC	Confirmation of correct insertion and cloning
AntiFSLHi_Seq_FSLstart_F	GTAGTCGATCGAGTGGAT	Confirmation of correct insertion and cloning
AntiFSLHi_Seq_FSLend_R	GTCCACCACGACCTTCTT	Confirmation of correct insertion and cloning
AntiFSLHi_Seq_Tub_F	GCTATAATGTCTTTGCAATC	Confirmation of correct insertion and cloning
AntiFSLHi_Seq_FSLstart_R	ATGACGCTGAACACGTAC	Confirmation of correct insertion and cloning

Table 5.2. Primers used for this work.

	Volts	Pulse Length	Pulse Interval	# of Pulses	Decay Rate
Poring	225V	8.0ms	50ms	2	40%
Transfer	20V	50ms	50ms	10	40%

Table 5.3. Settings for *Gonium* electroporation.

Chapter 6 - Conclusions

The grand challenge at the heart of studying the evolution of multicellularity and developmental complexity at the molecular level is one that has puzzled many an expert in a wide array of fields: the attempt at uncovering the path from genotypes to phenotypes. Indeed, the integration of functional units that *could* be independent into individuals of a higher order certainly appeals to the imagination, as the scale and extent of the integration provide fertile ground for diverse hypotheses placing greater or lesser roles at the genomic, transcriptomic, proteomic, or pathway levels. As it so often happens in biology, the answer might as well be ‘a little bit of all, and a little bit of none’.

While the shared genetic repertoire in the volvocine lineage certainly provides a scaffold for analyses such as the one in Chapter 5, the roles of distinct genes in the full genomic context is far from well understood. Works on genetic dispensability (Krylov et al., 2003; Albalat and Canestro, 2016) paint a complex picture in which adaptive and non-adaptive processes, ecological and even genetic context cues can affect the biological burdens placed on genes. For many of these cues we only have cursory information, and for others, less than that. Conversely, however, we have an enticing set of hints, especially from transcriptomic and protein-protein interaction analyses (Lindsey et al., 2021) (Chapter 2, 4), that illustrate familiar principles and functional players that serve as building blocks to our understanding of general principles governing developmental and evolutionary transitions.

This body of work highlights both the familiar and the unexpected in our understanding of the evolution of multicellularity. Core aspects of multicellular life, such as cell adhesion, are confirmed to play a pivotal role on colonial volvocine development that is linked to a genetic component. Moreover, the overlapping roles of transcription factors such as *regA* and *rlsD*

illustrate evolutionary processes (co-option, neofunctionalization) that are known to have played key roles in a wealth of lineages. The conserved expansion of Cyclin D1 genes across multicellular volvocines also reflects how co-option of cell cycle genes likely precipitated the evolution of multicellularity within a lineage. At the genomic level, through the comparative analysis of key volvocine genome species, the hypothesis that the developmental complexity portrayed by *Volvox* was made possible by the same mechanisms and tools visible in *Gonium* has been strengthened. Moreover, we have proposed that loss of gene functions that might be necessary for the unicellular state could have led to the establishment of conditions that stabilize undifferentiated multicellularity. Whereas gene loss is often associated with reversion to less developmentally complex lifestyles (e. g. loss of genes in symbionts and parasites), we propose that biological innovation can and does coincide with gene loss events. The corollary to this new model is that outside of a few key molecular players, there may not be such a thing as “multicellularity genes”; rather, the integration of cells into higher order individuals might be accompanied by the reorganization and integration of molecular players at the genomic, transcriptomic, proteomic and pathway level, including the switch from strictly temporal gene regulation to spatial or spatial temporal regulation patterns (Mikhailov et al., 2009).

Future directions for the work presented here span a wide array of techniques and experiments. On the subject of gene loss, the addition of outgroup members from genus *Chlamydomonas* to disentangle gene losses from Goniaceae and Volvocaceae from gene gains in *Chlamydomonas* can and should be performed. Follow-up studies on the impact of protein-protein interaction networks across the gamut of volvocine complexity are another avenue for exploration. Regarding the co-option of *FsII*, it will be interesting to see if other multicellular volvocines share the transcriptional pattern of *GpFsII*, as well as test for compromises to the

multicellular status in spheroidal species such as *Pandorina* and *Astrephomene*.

Immunofluorescence assays of the product of *GpFsl* will also undoubtedly provide valuable insight into how the protein localizes inside the cells; and if coupled with synchrony experiments, will likely aid in the determination of the role of *GpFsl* in *Gonium*'s development. Finally, evaluating whether *regA* and *rlsD* interact in any way could reform our understanding of cell differentiation in *Volvox carteri*. *regA*'s targets have not been elucidated yet, and so there is a possibility that knock-down of *regA* might induce the expression of *rlsD* among other transcription factors. Moreover, it will be interesting to see if in undifferentiated volvocines encoding for the *reg* cluster *rlsD* and *regA* share a greater or lesser number of direct and indirect targets.

The most important contributions to our knowledge of the volvocine algae are certainly the product of the recent and vertiginous development of molecular biology and genetic tools for this model system, as well as of the recent efforts to sequence key species. However, the implications from work done in this collection of emerging model organisms is sure to ripple across clades and fields. Future work integrating transcriptomics and proteomics, as well as follow up experiments on candidate genes of interest will undoubtedly shed light on known and unknown mechanisms acting on other Major Transitions, especially in eukaryotes.

References

- Adair, W.S., and Appel, H. (1989). Identification of a highly conserved hydroxyproline-rich glycoprotein in the cell walls of *Chlamydomonas reinhardtii* and two other Volvocales. *Planta* 179, 381-386.
- Adair, W.S., Steinmetz, S.A., Mattson, D.M., Goodenough, U.W., and Heuser, J.E. (1987). Nucleated assembly of *Chlamydomonas* and *Volvox* cell walls. *J Cell Biol* 105, 2373-2382.
- Adami, C., Ofria, C., and Collier, T.C. (2000). Evolution of biological complexity. *Proc Natl Acad Sci U S A* 97, 4463-4468.
- Adler, M., Anjum, M., Berg, O.G., Andersson, D.I., and Sandegren, L. (2014). High fitness costs and instability of gene duplications reduce rates of evolution of new genes by duplication-divergence mechanisms. *Mol Biol Evol* 31, 1526-1535.
- Albalat, R., and Canestro, C. (2016). Evolution by gene loss. *Nat Rev Genet* 17, 379-391.
- Alexander, K., Yang, H.S., and Hinds, P.W. (2003). pRb inactivation in senescent cells leads to an E2F-dependent apoptosis requiring p73. *Mol Cancer Res* 1, 716-728.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Arenas-Mena, C. (2017). The origins of developmental gene regulation. *Evol Dev* 19, 96-107.
- Bajhaiya, A.K., Dean, A.P., Zeef, L.A., Webster, R.E., and Pittman, J.K. (2016). PSR1 Is a Global Transcriptional Regulator of Phosphorus Deficiency Responses and Carbon Storage Metabolism in *Chlamydomonas reinhardtii*. *Plant Physiol* 170, 1216-1234.

- Benitez, M., Hernandez-Hernandez, V., Newman, S.A., and Niklas, K.J. (2018). Dynamical Patterning Modules, Biogeneric Materials, and the Evolution of Multicellular Plants. *Front Plant Sci* 9, 871.
- Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P., Arias, M.C., Henrissat, B., Coutinho, P.M., Krishnan, A., Zauner, S., Morath, S., Hilliou, F., Egizi, A., Perrineau, M.M., and Yoon, H.S. (2013). Genome of the red alga *Porphyridium purpureum*. *Nat Commun* 4, 1941.
- Bhattacharya, D., Qiu, H., Lee, J., Yoon, H.S., Weber, A.P.M., and Price, D.C. (2018). When Less is More: Red Algae as Models for Studying Gene Loss and Genome Evolution in Eukaryotes. *Critical Reviews in Plant Sciences* 37, 81-99.
- Biggar, K.K., Charih, F., Liu, H., Ruiz-Blanco, Y.B., Stalker, L., Chopra, A., Connolly, J., Adhikary, H., Frensemier, K., Galka, M., Fang, Q., Wynder, C., Stanford, W.L., Green, J.R., and Li, S.S.C. (2020). Proteome-wide Prediction of Lysine Methylation Reveals Novel Histone Marks and Outlines the Methyllysine Proteome. *bioRxiv* 274688.
- Bisalputra, T., and Stein, J.R. (1966). The Development of Cytoplasmic Bridges in *Volvox Aureus*. *Canadian Journal of Botany* 44, 1697-1702.
- Blaby, I.K., Blaby-Haas, C.E., Perez-Perez, M.E., Schmollinger, S., Fitz-Gibbon, S., Lemaire, S.D., and Merchant, S.S. (2015). Genome-wide analysis on *Chlamydomonas reinhardtii* reveals the impact of hydrogen peroxide on protein stress responses and overlap with other stress transcriptomes. *Plant J* 84, 974-988.
- Blaby, I.K., Blaby-Haas, C.E., Tourasse, N., Hom, E.F., Lopez, D., Aksoy, M., Grossman, A., Umen, J., Dutcher, S., Porter, M., King, S., Witman, G.B., Stanke, M., Harris, E.H.,

- Goodstein, D., Grimwood, J., Schmutz, J., Vallon, O., Merchant, S.S., and Prochnik, S. (2014). The Chlamydomonas genome project: a decade on. *Trends Plant Sci* 19, 672-680.
- Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., Lindquist, E., Lucas, S., Pangilinan, J., Polle, J., Salamov, A., Terry, A., Yamada, T., Dunigan, D.D., Grigoriev, I.V., Claverie, J.M., and Van Etten, J.L. (2010). The Chlorella variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* 22, 2943-2955.
- Bolotin, E., and Hershberg, R. (2015). Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. *Genome Biol Evol* 7, 2173-2187.
- Bonner, J.T. (1998). The origins of multicellularity. *Integrative biology : quantitative biosciences from nano to macro* 1, 27-36.
- Bonner, J.T. (2000). *First Signals*. Princeton University Press.
- Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., and Sattler, M. (2001). The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. *Nat Struct Biol* 8, 626-633.
- Bowles, A.M.C., Bechtold, U., and Paps, J. (2019). The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty. *Curr Biol*.
- Carroll, S.B. (2001). Chance and necessity: the evolution of morphological complexity and diversity. *Nature* 409, 1102-1109.
- Cheng, Q., Fowler, R., Tam, L.W., Edwards, L., and Miller, S.M. (2003). The role of GlsA in the evolution of asymmetric cell division in the green alga Volvox carteri. *Dev Genes Evol* 213, 328-335.

- Cohen, O., Ashkenazy, H., Belinky, F., Huchon, D., and Pupko, T. (2010). GLOOME: gain loss mapping engine. *Bioinformatics* 26, 2914-2915.
- Coleman, A.W. (1999). Phylogenetic analysis of "Volvocaceae" for comparative genetic studies. *Proc Natl Acad Sci U S A* 96, 13892-13897.
- Coleman, A.W. (2012). A Comparative Analysis of the Volvocaceae (Chlorophyta)(1). *J Phycol* 48, 491-513.
- Collen, J., Porcel, B., Carre, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., Elias, M., Artiguenave, F., Arun, A., Aury, J.M., Barbosa-Neto, J.F., Bothwell, J.H., Bouget, F.Y., Brillet, L., Cabello-Hurtado, F., Capella-Gutierrez, S., Charrier, B., Cladiere, L., Cock, J.M., Coelho, S.M., Colleoni, C., Czjzek, M., Da Silva, C., Delage, L., Denoeud, F., Deschamps, P., Dittami, S.M., Gabaldon, T., Gachon, C.M., Groisillier, A., Herve, C., Jabbari, K., Katinka, M., Kloareg, B., Kowalczyk, N., Labadie, K., Leblanc, C., Lopez, P.J., Mclachlan, D.H., Meslet-Cladiere, L., Moustafa, A., Nehr, Z., Nyvall Collen, P., Panaud, O., Partensky, F., Poulain, J., Rensing, S.A., Rousvoal, S., Samson, G., Symeonidi, A., Weissenbach, J., Zambounis, A., Wincker, P., and Boyen, C. (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A* 110, 5247-5252.
- Cross, F.R., and Umen, J.G. (2015). The *Chlamydomonas* cell cycle. *Plant J* 82, 370-392.
- Domazet-Loso, T., Brajkovic, J., and Tautz, D. (2007). A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* 23, 533-539.
- Duncan, L., Nishii, I., Howard, A., Kirk, D., and Miller, S.M. (2006). Orthologs and paralogs of *regA*, a master cell-type regulatory gene in *Volvox carteri*. *Curr Genet* 50, 61-72.

- Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J.E. (1998). Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*, 473-484.
- Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., and Uversky, V.N. (2008). The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* 9 Suppl 2, S1.
- Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., and Finn, R.D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res* 47, D427-D432.
- Fang, S.C., De Los Reyes, C., and Umen, J.G. (2006). Cell size checkpoint control by the retinoblastoma tumor suppressor pathway. *PLoS Genet* 2, e167.
- Featherston, J., Arakaki, Y., Hanschen, E.R., Ferris, P.J., Michod, R.E., Olson, B.J.S.C., Nozaki, H., and Durand, P.M. (2017). 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*, msx332-msx332.
- Fernandez, R., and Gabaldon, T. (2020). Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol* 4, 524-533.

- Ferris, P., Olson, B.J., De Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., Nishii, I., Hamaji, T., Nozaki, H., Pellegrini, M., and Umen, J.G. (2010). Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328, 351-354.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., and Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res* 42, D222-230.
- Fisher, R.M., Bell, T., and West, S.A. (2016). Multicellular group formation in response to predators in the alga *Chlorella vulgaris*. *J Evol Biol* 29, 551-559.
- Gaiti, F., Jindrich, K., Fernandez-Valverde, S.L., Roper, K.E., Degnan, B.M., and Tanurdzic, M. (2017). Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *Elife* 6, e22194.
- Go, Y., Satta, Y., Takenaka, O., and Takahata, N. (2005). Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates. *Genetics* 170, 313-326.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40, D1178-1186.
- Green, K.J., Viamontes, G.I., and Kirk, D.L. (1981). Mechanism of formation, ultrastructure, and function of the cytoplasmic bridge system during morphogenesis in *Volvox*. *J Cell Biol* 91, 756-769.
- Grochau-Wright, Z.I., Hanschen, E.R., Ferris, P.J., Hamaji, T., Nozaki, H., Olson, B., and Michod, R.E. (2017). Genetic basis for soma is present in undifferentiated volvocine green algae. *J Evol Biol* 30, 1205-1218.

- Grosberg, R.K., and Strathmann, R.R. (2007). The Evolution of Multicellularity: A Minor Major Transition? *Annual Review of Ecology, Evolution, and Systematics* 38, 621-654.
- Guijarro-Clarke, C., Holland, P.W.H., and Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol* 4, 519-523.
- Hallmann, A. (2003). Extracellular matrix and sex-inducing pheromone in Volvox. *Int Rev Cytol* 227, 131-182.
- Hallmann, A. (2006a). Morphogenesis in the family Volvocaceae: different tactics for turning an embryo right-side out. *Protist* 157, 445-461.
- Hallmann, A. (2006b). The pherophorins: common, versatile building blocks in the evolution of extracellular matrix architecture in Volvocales. *Plant J* 45, 292-307.
- Hallmann, A. (2011). Evolution of reproductive development in the volvocine algae. *Sex Plant Reprod* 24, 97-112.
- Hallmann, A., and Kirk, D.L. (2000). The developmentally regulated ECM glycoprotein ISG plays an essential role in organizing the ECM and orienting the cells of Volvox. *J Cell Sci* 113 Pt 24, 4605-4617.
- Hamaji, T., Kawai-Toyooka, H., Uchimura, H., Suzuki, M., Noguchi, H., Minakuchi, Y., Toyoda, A., Fujiyama, A., Miyagishima, S.Y., Umen, J.G., and Nozaki, H. (2018). Anisogamy evolved with a reduced sex-determining region in volvocine green algae. *Commun Biol* 1, 17.
- Hanschen, E.R., Davison, D.R., Grochau-Wright, Z.I., and Michod, R.E. (2017). Evolution of Individuality: A Case Study in the Volvocine Green Algae. *Philosophy, Theory, and Practice in Biology* 9.

- Hanschen, E.R., Ferris, P.J., and Michod, R.E. (2014). Early evolution of the genetic basis for soma in the volvocaceae. *Evolution* 68, 2014-2025.
- Hanschen, E.R., Herron, M.D., Wiens, J.J., Nozaki, H., and Michod, R.E. (2018a). Multicellularity Drives the Evolution of Sexual Traits. *Am Nat* 192, E93-E105.
- Hanschen, E.R., Herron, M.D., Wiens, J.J., Nozaki, H., and Michod, R.E. (2018b). Repeated evolution and reversibility of self-fertilization in the volvocine green algae. *Evolution* 72, 386-398.
- 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., Bakaric, R., Luria, V., Karger, A., Kirschner, M.W., Durand, P.M., Michod, R.E., Nozaki, H., and Olson, B.J. (2016). The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat Commun* 7, 11370.
- Helsen, J., Voordeckers, K., Vanderwaeren, L., Santermans, T., Tsontaki, M., Verstrepen, K.J., and Jelier, R. (2020). Gene Loss Predictably Drives Evolutionary Adaptation. *Mol Biol Evol* 37, 2989-3002.
- Herron, M.D., Borin, J.M., Boswell, J.C., Walker, J., Chen, I.K., Knox, C.A., Boyd, M., Rosenzweig, F., and Ratcliff, W.C. (2019). De novo origins of multicellularity in response to predation. *Sci Rep* 9, 2328.
- Herron, M.D., Desnitskiy, A.G., and Michod, R.E. (2010). Evolution of Developmental Programs in *Volvox* (Chlorophyta). *Journal of Phycology* 46, 316-324.
- Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proc Natl Acad Sci U S A* 106, 3254-3258.

- Herron, M.D., and Michod, R.E. (2008). Evolution of complexity in the volvocine algae: transitions in individuality through Darwin's eye. *Evolution* 62, 436-451.
- Herron, M.D., Rashidi, A., Shelton, D.E., and Driscoll, W.W. (2013). Cellular differentiation and individuality in the 'minor' multicellular taxa. *Biol Rev Camb Philos Soc* 88, 844-861.
- Hiraide, R., Kawai-Toyooka, H., Hamaji, T., Matsuzaki, R., Kawafune, K., Abe, J., Sekimoto, H., Umen, J., and 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, 1038-1040.
- Holmes, J.A., and Dutcher, S.K. (1989). Cellular asymmetry in *Chlamydomonas reinhardtii*. *J Cell Sci* 94 (Pt 2), 273-285.
- Hoops, H.J. (1993). Flagellar, Cellular and Organismal Polarity in *Volvox-Carteri*. *Journal of Cell Science* 104, 105-117.
- Hoops, H.J., Nishii, I., and Kirk, D.L. (2006). "Cytoplasmic Bridges in *Volvox* and Its Relatives," in *Cell-Cell Channels*. (New York, NY: Springer New York), 65-84.
- Huber, O., and 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, 4212-4222.
- Iida, H., Ota, S., and Inouye, I. (2013). Cleavage, incomplete inversion, and cytoplasmic bridges in *Gonium pectorale* (Volvocales, Chlorophyta). *J Plant Res* 126, 699-707.
- Jimenez-Marin, B., and Olson, B. (2022). The Curious Case of Multicellularity in the Volvocine Algae. *Front Genet* 13, 787665.

- Jiménez-Marín, B., Rakijas, J.B., Tyagi, A., Pandey, A., Hanschen, E.R., Anderson, J., Heffel, M.G., Platt, T.G., and Olson, B.J.S.C. (2021). Gene loss during a transition to multicellularity. *bioRxiv*.
- John Maynard Smith, E.S. (1995). "The Major Transitions in Evolution". (New York: Oxford University Press Inc.).
- Johnson, K.L., Jones, B.J., Bacic, A., and Schultz, C.J. (2003). The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiol* 133, 1911-1925.
- Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31, 857-863.
- King, N. (2004). The unicellular ancestry of animal development. *Dev Cell* 7, 313-325.
- Kirk, D.L. (1997). *Volvox*. Cambridge: Cambridge University Press.
- Kirk, D.L. (2000). Volvox as a model system for studying the ontogeny and phylogeny of multicellularity and cellular differentiation. *Journal of Plant Growth Regulation* 19, 265-274.
- Kirk, D.L. (2001). Germ-soma differentiation in volvox. *Dev Biol* 238, 213-223.
- Kirk, D.L. (2005). A twelve-step program for evolving multicellularity and a division of labor. *Bioessays* 27, 299-310.
- Kirk, D.L., Birchem, R., and King, N. (1986). The extracellular matrix of Volvox: a comparative study and proposed system of nomenclature. *J Cell Sci* 80, 207-231.
- Kirk, D.L., and Kirk, M.M. (1983). Protein synthetic patterns during the asexual life cycle of Volvox carteri. *Dev Biol* 96, 493-506.

- Kirk, D.L., and Nishii, I. (2001). *Volvox carteri* as a model for studying the genetic and cytological control of morphogenesis. *Development Growth & Differentiation* 43, 621-631.
- Kirk, M.M., and Kirk, D.L. (1985). Translational regulation of protein synthesis, in response to light, at a critical stage of *Volvox* development. *Cell* 41, 419-428.
- Kirk, M.M., Stark, K., Miller, S.M., Muller, W., Taillon, B.E., Gruber, H., Schmitt, R., and Kirk, D.L. (1999). *regA*, a *Volvox* gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. *Development* 126, 639-647.
- Knoll, A.H. (2011). The Multiple Origins of Complex Multicellularity. *Annual Review of Earth and Planetary Sciences, Vol 39* 39, 217-239.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27, 722-736.
- Koufopanou, V. (1994). The Evolution of Soma in the Volvocales. *American Naturalist* 143, 907-931.
- Krylov, D.M., Wolf, Y.I., Rogozin, I.B., and Koonin, E.V. (2003). Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res* 13, 2229-2235.
- Kvitek, D.J., and Sherlock, G. (2013). Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet* 9, e1003972.
- Leigh, E.G., Smith, J.M., and Szathmary, E. (1995). The Major Transitions of Evolution. *Evolution* 49, 1302.

- Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M., Haag, J.D., Gould, M.N., Stewart, R.M., and Kendzierski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035-1043.
- Lerche, K., and Hallmann, A. (2009). Stable nuclear transformation of *Gonium pectorale*. *BMC Biotechnol* 9, 64.
- Lerche, K., and Hallmann, A. (2013). Stable nuclear transformation of *Eudorina elegans*. *BMC Biotechnol* 13, 11.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y.L., Teslovich, T.M., May-Simera, H., Li, H.T., Blacque, O.E., Li, L.Y., Leitch, C.C., Lewis, R.A., Green, J.S., Parfrey, P.S., Leroux, M.R., Davidson, W.S., Beales, P.L., Guay-Woodford, L.M., Yoder, B.K., Stormo, G.D., Katsanis, N., and Dutcher, S.K. (2004). Comparative and basal genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 117, 541-552.
- Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13, 2178-2189.
- Lindsey, C.R., Rosenzweig, F., and Herron, M.D. (2021). Phylotranscriptomics points to multiple independent origins of multicellularity and cellular differentiation in the volvocine algae. *BMC Biol* 19, 182.
- Lopez-Escardo, D., Grau-Bove, X., Guillaumet-Adkins, A., Gut, M., Sieracki, M.E., and Ruiz-Trillo, I. (2019). Reconstruction of protein domain evolution using single-cell amplified

- genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philos Trans R Soc Lond B Biol Sci* 374, 20190088.
- Lowder, L.G., and Herbert, S.K. (2014). Heterologous expression of a Volvox cell adhesion molecule causes flocculation in *Chlamydomonas reinhardtii*. *Journal of Applied Phycology* 27, 721-731.
- Martin E. Boraas, D.B.S.a.J.E.B. (1998). Phagotrophy by a flagellate selects for colonial prey: A possible origin of multicellularity. *Evolutionary Ecology* 12, 153-164.
- Matt, G.Y., and 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 (Bethesda)* 8, 531-550.
- Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and Munos, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques* 61, 203-205.
- Mcshea, D.W. (1996). Metazoan Complexity and Evolution: Is There a Trend? *Evolution* 50, 477-492.
- 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., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernandez, E., Fukuzawa, H., Gonzalez-Ballester, D., Gonzalez-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag,

- M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riano-Pachon, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245-250.
- Michod, R.E. (2007). Evolution of individuality during the transition from unicellular to multicellular life. *Proc Natl Acad Sci U S A* 104 Suppl 1, 8613-8618.
- Michod, R.E., Nedelcu, A.M., and Michod, R.E. (2004). Evolvability, Modularity, and Individuality during the Transition to Multicellularity in Volvocalean Green Algae. 466-489.
- Mikhailov, K.V., Konstantinova, A.V., Nikitin, M.A., Troshin, P.V., Rusin, L.Y., Lyubetsky, V.A., Panchin, Y.V., Mylnikov, A.P., Moroz, L.L., Kumar, S., and Aleoshin, V.V. (2009). The origin of Metazoa: a transition from temporal to spatial cell differentiation. *Bioessays* 31, 758-768.
- Miller, D.J., and Technau, U. (2010). "Understanding the evolution of multicellularity: insights from basal metazoans.", in: *BioEssays : news and reviews in molecular, cellular and developmental biology*).
- Miller, S.M. (2002). Taming the fierce roller: an "enhanced" understanding of cellular differentiation in *Volvox*. *Bioessays* 24, 3-7.

- Nanjundiah, V., Ruiz-Trillo, I., and Kirk, D. (2018). "Protists and Multiple Routes to the Evolution of Multicellularity," in *Cells in Evolutionary Biology*, eds. B. Hall & S. Moody. CRC Press).
- Nedelcu, A.M. (2009). Environmentally induced responses co-opted for reproductive altruism. *Biol Lett* 5, 805-808.
- Nedelcu, A.M., and Michod, R.E. (2003). Cooperation and conflict during the unicellular-multicellular and prokaryotic-eukaryotic transitions. *Cooperation and conflict during the unicellular-multicellular and prokaryotic-eukaryotic transitions*, 195-208.
- Newman, S.A., and Bhat, R. (2009). Dynamical patterning modules: a "pattern language" for development and evolution of multicellular form. *Int J Dev Biol* 53, 693-705.
- Ni, Z., Kim, E.-D., and Chen, Z.J. (2009). Chlorophyll and starch assays.
- Niklas, K.J. (2014). The evolutionary-developmental origins of multicellularity. *Am J Bot* 101, 6-25.
- Nishii, I., Ogihara, S., and Kirk, D.L. (2003). A kinesin, *invA*, plays an essential role in volvox morphogenesis. *Cell* 113, 743-753.
- Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S., and Watanabe, M.M. (2000). Origin and evolution of the colonial volvocales (Chlorophyceae) as inferred from multiple, chloroplast gene sequences. *Mol Phylogenet Evol* 17, 256-268.
- Nozaki, H., Misumi, O., and Kuroiwa, T. (2003). Phylogeny of the quadriflagellate Volvocales (Chlorophyceae) based on chloroplast multigene sequences. *Mol Phylogenet Evol* 29, 58-66.
- Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., Takio, S., Tamura, K., Chung, S.J., Nakamura,

- S., Kuroiwa, H., Tanaka, K., Sato, N., and Kuroiwa, T. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biol* 5, 28.
- Olson, B.J., and Markwell, J. (2007). Assays for determination of protein concentration. *Curr Protoc Protein Sci* Chapter 3, Unit 3 4.
- Olson, B.J., and Nedelcu, A.M. (2016). Co-option during the evolution of multicellular and developmental complexity in the volvocine green algae. *Curr Opin Genet Dev* 39, 107-115.
- Olson, B.J., Oberholzer, M., Li, Y., Zones, J.M., Kohli, H.S., Bisova, K., Fang, S.C., Meisenhelder, J., Hunter, T., and Umen, J.G. (2010). Regulation of the *Chlamydomonas* cell cycle by a stable, chromatin-associated retinoblastoma tumor suppressor complex. *Plant Cell* 22, 3331-3347.
- Ortega Escalante, J.A. (2018). *Investigation of <i>Volvox carteri</i> Cell Differentiation and Its Evolution Through Functional Analysis of <i>regA</i> and <i>regA</i> Homologues*. Ph.D., University of Maryland, Baltimore County.
- Packard, M., Mathew, D., and Budnik, V. (2003). FAST remodeling of synapses in *Drosophila*. *Curr Opin Neurobiol* 13, 527-534.
- Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., Dupont, C., Jorgensen, R., Derelle, E., Rombauts, S., Zhou, K., Otiillar, R., Merchant, S.S., Podell, S., Gaasterland, T., Napoli, C., Gendler, K., Manuell, A., Tai, V., Vallon, O., Piganeau, G., Jancek, S., Heijde, M., Jabbari, K., Bowler, C., Lohr, M., Robbens, S., Werner, G., Dubchak, I., Pazour, G.J., Ren, Q., Paulsen, I., Delwiche, C., Schmutz, J., Rokhsar, D., Van De Peer, Y., Moreau, H., and Grigoriev, I.V. (2007). The tiny eukaryote

- Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A* 104, 7705-7710.
- 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., Shapiro, H., Schmutz, J., Grimwood, J., Lindquist, E., Lucas, S., Grigoriev, I.V., Schmitt, R., Kirk, D., and Rokhsar, D.S. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329, 223-226.
- Ratcliff, W.C., Herron, M.D., Howell, K., Pentz, J.T., Rosenzweig, F., and Travisano, M. (2013). Experimental evolution of an alternating uni- and multicellular life cycle in *Chlamydomonas reinhardtii*. *Nat Commun* 4, 2742.
- Rokas, A. (2008). The origins of multicellularity and the early history of the genetic toolkit for animal development. *Annu Rev Genet* 42, 235-251.
- Roth, M.S., Cokus, S.J., Gallaher, S.D., Walter, A., Lopez, D., Erickson, E., Endelman, B., Westcott, D., Larabell, C.A., Merchant, S.S., Pellegrini, M., and Niyogi, K.K. (2017). Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *Proc Natl Acad Sci U S A* 114, E4296-E4305.
- Ruiz-Trillo, I., and Nedelcu, A.M. (eds.). (2015). *Evolutionary Transitions to Multicellular Life*. Dordrecht: Springer Netherlands.
- Sahm, A., Bens, M., Platzer, M., and Szafranski, K. (2017). PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Res* 45, e100.

- Salazar-Ciudad, I. (2006). On the origins of morphological disparity and its diverse developmental bases. *Bioessays* 28, 1112-1122.
- Sathe, S., and Durand, P.M. (2015). Cellular aggregation in *Chlamydomonas* (Chlorophyceae) is chimaeric and depends on traits like cell size and motility. *European Journal of Phycology* 51, 129-138.
- Schroda, M., Blocker, D., and Beck, C.F. (2000). The HSP70A promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant Journal* 21, 121-131.
- Sepey M., M.M., Zdobnov E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology* 1962.
- Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., Langer, B.E., and Hiller, M. (2018). A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun* 9, 1215.
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Res* 19, 1117-1123.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.
- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* 33, W465-467.
- Starr, R.C. (1968). Cellular differentiation in *Volvox*. *Proc Natl Acad Sci U S A* 59, 1082-1088.
- Sumper, M., and Hallmann, A. (1998). Biochemistry of the extracellular matrix of *Volvox*. *Int Rev Cytol* 180, 51-85.
- Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6, e21800.

- Tam, L.W., and Kirk, D.L. (1991a). Identification of cell-type-specific genes of *Volvox carteri* and characterization of their expression during the asexual life cycle. *Dev Biol* 145, 51-66.
- Tam, L.W., and Kirk, D.L. (1991b). The program for cellular differentiation in *Volvox carteri* as revealed by molecular analysis of development in a gonidialess/somatic regenerator mutant. *Development* 112, 571-580.
- Trigos, A.S., Pearson, R.B., Papenfuss, A.T., and Goode, D.L. (2018). How the evolution of multicellularity set the stage for cancer. *Br J Cancer* 118, 145-152.
- True, J.R.a.H., Eric S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development* 3, 109-119.
- Ueki, N., and Nishii, I. (2009). Controlled enlargement of the glycoprotein vesicle surrounding a volvox embryo requires the InvB nucleotide-sugar transporter and is required for normal morphogenesis. *Plant Cell* 21, 1166-1181.
- Umen, J.G., and Olson, B.J. (2012). Genomics of Volvocine Algae. *Adv Bot Res* 64, 185-243.
- Von Der Heyde, B., and Hallmann, A. (2020). Targeted migration of pterophorin-S indicates extensive extracellular matrix dynamics in *Volvox carteri*. *Plant J* 103, 2301-2317.
- Wagner, A. (1994). Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci U S A* 91, 4387-4391.
- Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., and Jones, D.T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics* 20, 2138-2139.
- Waterborg, J.H., Robertson, A.J., Tatar, D.L., Borza, C.M., and Davie, J.R. (1995). Histones of *Chlamydomonas reinhardtii*. Synthesis, acetylation, and methylation. *Plant Physiol* 109, 393-407.

Wolf, Y.I., and Koonin, E.V. (2013). Genome reduction as the dominant mode of evolution.

Bioessays 35, 829-837.

Worden, A.Z., Lee, J.H., Mock, T., Rouze, P., Simmons, M.P., Aerts, A.L., Allen, A.E.,

Cuvelier, M.L., Derelle, E., Everett, M.V., Foulon, E., Grimwood, J., Gundlach, H.,

Henrissat, B., Napoli, C., McDonald, S.M., Parker, M.S., Rombauts, S., Salamov, A., Von

Dassow, P., Badger, J.H., Coutinho, P.M., Demir, E., Dubchak, I., Gentemann, C.,

Eikrem, W., Gready, J.E., John, U., Lanier, W., Lindquist, E.A., Lucas, S., Mayer, K.F.,

Moreau, H., Not, F., Otiillar, R., Panaud, O., Pangilinan, J., Paulsen, I., Piegu, B.,

Poliakov, A., Robbens, S., Schmutz, J., Toulza, E., Wyss, T., Zelensky, A., Zhou, K.,

Armbrust, E.V., Bhattacharya, D., Goodenough, U.W., Van De Peer, Y., and Grigoriev,

I.V. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324, 268-272.

Yamano, T., Iguchi, H., and Fukuzawa, H. (2013). Rapid transformation of *Chlamydomonas*

reinhardtii without cell-wall removal. *J Biosci Bioeng* 115, 691-694.

Yamashita, S., Arakaki, Y., Kawai-Toyooka, H., Noga, A., Hirono, M., and Nozaki, H. (2016).

Alternative evolution of a spheroidal colony in volvocine algae: developmental analysis of embryogenesis in *Astrephomene* (Volvocales, Chlorophyta). *BMC Evol Biol* 16, 243.

Yamashita, S., Yamamoto, K., Matsuzaki, R., Suzuki, S., Yamaguchi, H., Hirooka, S.,

Minakuchi, Y., Miyagishima, S.Y., Kawachi, M., Toyoda, A., and Nozaki, H. (2021).

Genome sequencing of the multicellular alga *Astrephomene* provides insights into convergent evolution of germ-soma differentiation. *Sci Rep* 11, 22231.

- Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep* 6, 21361.
- Yruela, I., Oldfield, C.J., Niklas, K.J., and Dunker, A.K. (2017). Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity. *Genome Biol Evol* 9, 1248-1265.
- Yu, B., Xu, C., and Benning, C. (2002). Arabidopsis disrupted in SQD2 encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proc Natl Acad Sci U S A* 99, 5732-5737.
- Zhang, N., Pazouki, L., Nguyen, H., Jacobshagen, S., Bigge, B.M., Xia, M., Mattoon, E.M., Klebanovych, A., Sorkin, M., Nusinow, D.A., Avasthi, P., Czymmek, K.J., and Zhang, R. (2022). Comparative Phenotyping of Two Commonly Used *Chlamydomonas reinhardtii* Background Strains: CC-1690 (21gr) and CC-5325 (The CLiP Mutant Library Background). *Plants (Basel)* 11.
- Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., Zhao, P.X., Rhee, S.Y., and Fei, Z. (2016). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol Plant* 9, 1667-1670.
- Zmasek, C.M.a.G., A. (2011). Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. *Genome Biology* 12:R4.

Zones, J.M., Blaby, I.K., Merchant, S.S., and Umen, J.G. (2015). High-Resolution Profiling of a Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous Cell and Metabolic Differentiation. *Plant Cell* 27, 2743-2769.