Title: Gene loss during the transition to multicellularity

Authors: Berenice Jiménez-Marín^{1,2}, Jessica B. Rakijas¹, Antariksh Tyagi¹, Aakash Pandey¹, Erik R. Hanschen³, Jaden Anderson¹, Matthew G. Heffel^{1,2}, Thomas G. Platt¹, Bradley J. S. C. Olson^{1*}

Affiliations:

1. Division of Biology, Kansas State University, Manhattan, KS, USA

2. Interdepartmental Genetics Graduate Program, Kansas State University, Manhattan, KS, USA

3. Los Alamos National Lab, Los Alamos, NM, USA

*Corresponding author

Summary: 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 highly similar, suggesting multicellular evolution requires few genetic changes to undergo dramatic shifts in developmental complexity. Here, the evolutionary dynamics of five 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 five species exhibited a high degree of novel interactions, suggesting that gene loss promotes 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 developmental complexity may be driven by gene loss rather than gene gain.

The evolution of multicellular organisms is a prerequisite for the emergence of complex organismal body plans (Grosberg & Strathmann, 2007; Knoll, 2011). However, the genomic basis for multicellular evolution is poorly understood (Olson & 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, Hackett, Aylward, & Michod, 2009), and the ~100 member species span a wide range of developmental complexity within a size range from 10 μ m to 3 mm (Hanschen, Herron, Wiens, Nozaki, & Michod, 2018a). This group includes unicellular *Chlamydomonas*, undifferentiated multicellular *Gonium* (8-16 cells), multicellular isogamous *Yamagishiella* (32 cells), multicellular anisogamous *Eudorina* (32 cells), and multicellular differentiated *Volvox* (>500 cells)(Coleman, 2012; D. L. Kirk, 2005).

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 (D. L. Kirk, 2005)(Figure 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 suggesting that genetic gains for each step might not fully explain the organismal complexity of the volvocines (Hanschen et al., 2016). Indeed, the genomes of *Chlamydomonas reinhardtii*, *Gonium pectorale*, and *Volvox carteri*^a are highly similar and syntenic (Blaby et al., 2014; Hanschen et al., 2016; Merchant et al., 2007). Moreover, a homolog of the cell cycle regulator and tumor suppressor *Retinoblastoma* (RB) is sufficient for undifferentiated multicellularity (Hanschen et al., 2016),

^aHereafter referred to by genus

suggesting co-option of a developmental regulator can result in complex developmental patterns required for multicellularity (Olson & Nedelcu, 2016). Thus, the evolution and diversification of the volvocine algae is thought to have heavily relied on duplication, divergence and co-option of key functional elements.

Co-option, however, is unlikely to be the only evolutionary process behind volvocine multicellularity. Gene loss is found in many lineages undergoing evolutionary change (Albalat & Canestro, 2016; Fernandez & Gabaldon, 2020; Go, Satta, Takenaka, & Takahata, 2005; Guijarro-Clarke, Holland, & Paps, 2020; Kvitek & Sherlock, 2013); nevertheless, its action in concert with other evolutionary drivers is not well understood (Bhattacharya et al., 2018; Bolotin & Hershberg, 2015; Bowles, Bechtold, & Paps, 2019; Lopez-Escardo et al., 2019). To investigate the role of gene loss during the evolution of multicellularity, we report the annotation of genomes from two volvocine algae (*Yamagishiella unicocca* and *Eudorina sp.*)^a. These newly analyzed genomes, in addition to those of *Chlamydomonas*, *Gonium*, and *Volvox* (Hanschen et al., 2016; Merchant et al., 2007; Prochnik et al., 2010), 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 observable between species. These results suggest that elimination of non-essential functions at the transition to multicellularity likely drives changes to molecular networks that result in novel functions associated to changes in developmental complexity.

Results

67 We compared the genomes of five closely related volvocine algae exhibiting the major changes
68 in morphological complexity (D. L. Kirk, 2005)(Figure 1A cartoons, Table S1) and validated

Figure 1. Volvocine algal genomes are highly similar. **A**, Genome-wide phylogeny of the volvocine algae (green) compared to other chlorophycean 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 five species (Supplementary File 1). **C**, Pfam domains shared between five species studied (Supplementary File 2).

69 their annotation using BUSCO (Seppey M., 2019). BUSCO inferences suggest that all five
70 annotated genomes are of high quality (Table S2). Orthologous group and protein family domain

(Pfams) catalogs for the analyzed species and their chlorophyte outgroups were compiled, and a genome-wide reconstruction of volvocine phylogeny was generated (Figure 1A) based on orthologous groups, which is in close agreement with previously determined phylogenies based on smaller gene sets (Hanschen et al., 2016; Herron & Michod, 2008; Nozaki, Misumi, & Kuroiwa, 2003). Overall, 13,424 orthologous groups were identified; of these, 7,163 (53.3%) are shared among the five genomes, and 1,820 are species specific (13.6%, Figure 1B). Analysis of the evolutionary trajectory of Pfams (El-Gebali et al., 2019) in the five species shows 1,725 Pfam domains. Of these, 1,275 (73.9%) are shared in all five species, and 88 (5.1%) are species specific. Notably, 8,958 (66.7%) orthologous groups, and 1,515 (87.8%) Pfams are shared by at least 4 species (Figure 1C). Pairwise comparisons of orthologous group and Pfam domains between recently diverged species further highlight this strong conservation (Supplementary Figure S1A, B). These results support previous findings of the remarkable similarity found between the volvocine genomes (Hanschen et al., 2016; Prochnik et al., 2010), despite their marked developmental and morphological differences.

Given the high degree of similarity between volvocine algae genomes (Figure 1B, C), we hypothesized that small differences between the *Chlamydomonas*, *Gonium*, *Yamagishiella*, *Eudorina* and *Volvox* genomes contribute to the evolution of multicellularity. Rates of evolution of shared genes and protein domains (orthologous groups and Pfam domains) for the five 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 375 (5.1%) orthologous groups have undergone significant changes, with a bias towards gene loss (146 expanding orthologous groups, 229 contracting, Figure 2A outliers, shaded grey area). Surprisingly, while most Pfam domains were conserved, significantly expanding domains were

94 less abundant than their contracting counterparts (39 and 96, respectively, Figure 2B outliers,

Figure 2. Gene loss outpaces gain in volvocine algae genomes. Distribution of the loss and gain rates per distance to LCA in volvocine and red algae for **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 in shaded grey areas. **E**, Histone copy number per distance to LCA decreases except for H1 in the volvocine genomes. **F**, Histone copies (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.

shaded grey area). Pfam domain counts per gene did not show significant variation between the five species (Supplementary Figure S2), and contraction trends were not observed in highly conserved genes Actin, β 1-Tubulin, β 2-Tubulin, α 2-Tubulin, ferridoxin, ATP-synthase α , ATP-synthase β , dynein heavy chain 1 α , and dynein heavy chain 1 β (Supplementary Figure S3, TableS3).

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 (49 and 84 respectively) revealed that 7 families of TFs (14.3%) underwent significant contraction, while none expanded significantly (Figure 2C, Supplementary Figure S4A). Similarly, 16 families of PKs (19.0%) underwent significant change, where 3 of the PKs expanded and 13 contracted (Figure 2D, Supplementary Figure S4B). These data suggest that volvocine algae have lost genes with conserved functions. Conversely, gene gain occurred at a much lower rate and frequency (Fig 2A-F, Supplementary Figure S4).

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, Robertson, Tatar, Borza, & Davie, 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 2E, Table S4 – S5). Analysis of histone N-termini methylation and acetylation site variation for H2A, H2B, H3, and H4 shows that H2A and H4 underwent copy number reduction (Figure 2F, Table S4). However, Histone H2B and H3 exhibited greater N-terminal variant diversity in

Gonium (Figure 2F, Table S5). 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, their orthologous groups (Figure 2A), Pfam domains (Figure 2B), TFs (Figure 2C), PKs (Figure 2D) and histone (Table S6) 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).

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). To assess the impact of positive selection on increases to developmental complexity, orthologous groups of genes with representatives in all five species were subjected to dN/dS analysis. Using *Chlamydomonas* as the reference for ratio calculation 219, 56, 77 and 43 genes exhibited positive selection in *Gonium*, *Yamagishiella*, *Eudorina*, and *Volvox*, respectively (Figure 2G). Using *Volvox* as the reference, there were 74, 248, 67 and 83 genes with a significant dN/dS value in *Chlamydomonas*, *Gonium*, *Yamagishiella*, and *Eudorina*, respectively (Figure 2G). *Gonium* has a larger number of genes under positive selection, with the other volvocine genomes 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 processes of gene loss coupled to key co-option events might have set the stage for instances of positive selection, especially upon the transition to undifferentiated multicellularity.

The genomes of the volvocines are highly syntenic (Hanschen et al., 2016; Prochnik et al., 2010), thus allowing the examination of the mechanism underlying gene loss for the contracting orthologous groups at the sequence level (Figure 2A). Determination of the mechanisms behind gene loss at the sequence level is difficult to do in other phyla, where it is inferred indirectly through phylogenetic analysis (Fernandez & Gabaldon, 2020; Guijarro-Clarke et al., 2020). The contracting groups contain 1,560 genes assembled into *Chlamydomonas* chromosomes, of which 1,218 (78.1%) in 196 (85.6%) orthologous groups had loci that were adequately assembled in all genomes for further analysis. This analysis demonstrated 28 of 1,218 genes in contracting orthologous groups are retained (preservation of orthologs across all

Figure 3. Gene loss in the volvocine algae occurs primarily by gradual decay. **A**, Retention, decay, and deletion of 1,218 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).

species), while 1,190 loci exhibit gene loss. 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 except for *Volvox*; 56.5% of *Gonium*, 60.7% of *Yamagishiella*, 57.3% of *Eudorina* and 17.4% of *Volvox* gene losses show decay signatures (Figure 3A). 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 continued as organismal complexity increased.

Cross-species comparisons of retention, decay, and deletion events show that most candidate gene losses, 960 loci (78.8%), are due to decay (Figure 3B). Surprisingly, our analysis also suggests 797 (65.4%) loci are gene loss candidates in all four multicellular species. Of these, 674 (55.3%) occurred by gradual decay (Figure 3C, Table S7). The presence of decay signatures eliminates the possibility that these loci represent gene gains of *Chlamydomonas*. In the case of deletions (18.9%), only 123 loci (10.1%) were completely absent in all multicellular species, which could either represent complete gene loss in the multicellular lineage or *Chlamydomonas*-specific gene gains. The predominance and 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.

Previously, genes important for *Volvox* development were hypothesized to involved in the evolution of multicellularity (Hanschen et al., 2016; Olson & Nedelcu, 2016; Prochnik et al., 2010). used phylogenetic analysis to gain deeper understanding of co-option in the volvocine algae. Cyclin D1 genes (Supplementary Figure S5), certain extracellular matrix (ECM) related

genes (Table S8, Supplementary Figure S6, S7), and the *Volvox* cell differentiation gene *regA* (Supplementary Figure S8) follow a duplication and divergence pattern that coincides with increases in complexity. However, other developmentally relevant genes such as *inv*, (Supplementary Figure S9) and *glsA*, (Supplementary Figure S10) 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). However, no significant change in the IDP content, frequency or distribution was observed between the volvocine algae genomes (Supplementary Figure S11).

To examine the role of genetic innovation in volvocine evolutionary history, their genes were assigned to nine phylostrata (Supplementary Figure S12). This analysis revealed that 53.7 to 63.0% of the genes fell under phylostratum 1 (PS-1, cellular organisms) and PS-2 (Eukaryotes) (Supplementary Figure S12, dark grey). Conversely, 0 to 4.4% of the genes were found in PS-7 (family). Pfam phylostrata also fail to coincide with the evolution of multicellularity (Supplementary S12, light grey). These results suggest that co-option and expansion of key genes is a relevant, but does not coincide with stepwise increases in developmental complexity as previously thought (D. L. Kirk, 2005).

To understand how developmental complexity could evolve in the context of significant 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 changes in volvocine developmental complexity, even though few genes are undergoing positive selection or co-option (Figure 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 4A, Supplementary Figure S13).

197 Changes in morphological and developmental complexity in the volvocine algae are
198 thought to be related to the cytoskeleton (D. L. Kirk & Nishii, 2001). This is because cytoskeletal
199 elements common to *Chlamydomonas* and *Volvox* are involved in cell division in both species,
200 and in the latter also in inversion (D. L. Kirk & Nishii, 2001). Hence, PPIs of the actin and
201 tubulin cytoskeleton were examined. Tubulin and actin proteins are highly conserved in the
202 volvocine algae (Supplementary Figure S14), and as expected, complexes of similar size were
203 found in *Chlamydomonas*, *Gonium*, and *Eudorina* 2D-PAGE immunoblots (Figure 4B, C,
204 Supplementary Figure S13). Minor differences in migration patterns for both proteins in the
205 denaturing dimension are likely caused by post-translational modifications. Notably, novel α -
206 tubulin complexes were shifted along the native axis, suggesting the presence of larger, species-
207 specific complexes containing tubulin (Figure 4B, arrows). Similarly, 2D-PAGE immunoblots
208 for β -actin showed a common streak in all three species, but also displayed species-specific
209 actin-containing complexes that are larger than the shared complex (Figure 4C, arrows). Taken

Figure 4. 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.

210 together, these results indicate that while the genomes of the volvocine algae are highly similar
 211 and experiencing gene loss, PPIs between conserved gene products are forming species-specific
 212 complexes, which could account for the observed differences in biological complexity between
 213 the volvocines.

Figure 5. 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.

214 To evaluate the potential of gene loss as a mechanism of evolutionary transitions, gene
 215 loss and gain were modeled using an adaptation of the Wagner gene regulatory network model
 216 (Wagner, 1994). Our model mirrors the effects of gene gain in Wagner's original model, which
 217 demonstrated that networks more often reach a new equilibrium state following an intermediate
 218 number of gene duplications than when there are many or few duplications (Wagner, 1994). Our
 219 model predicts similar outcomes following gene loss with the proportion of novel stable gene
 220 network configurations being higher following gene deletions as compared to an equal number
 221 of duplications (Figure 5). Upon iterating the model over a broad range of network

connectivities, similar results were obtained (Supplementary Figure S15). 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 challenges the common assumption that biological and genomic complexity are positively correlated (Adami, Ofria, & Collier, 2000; Trigos, Pearson, Papenfuss, & Goode, 2018). If this were the case, the developmental gains in *Volvox* and species that exhibit seemingly intermediate steps of developmental complexity (typified by *Gonium*, *Yamagishiella* and *Eudorina*) might be the consequence of sequential acquisition of novel genes and/or the result of co-option events correlating to each developmental gain (Cheng, Fowler, Tam, Edwards, & Miller, 2003; Grochau-Wright et al., 2017; Nishii, Ogiwara, & Kirk, 2003; Olson & Nedelcu, 2016). However, the five analyzed genomes did not show evidence of extensive gene duplication, subspecialization or neofunctionalization. Rather, we discovered that a very small set of expanding and co-opted genes are found in the five species (Table S8, Supplementary Figure S5 – S10), 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 obvious similarities between volvocine genomes (Figure 1B, C), genetic losses in conserved genomic regions have occurred in volvocine history (Figure 3). 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 2A, B), even in the absence of significant variation in the number of Pfam domain counts per gene between the five species (Supplementary Figure S2). Regulatory and histone genes, which are important for development (Carroll, 2001; Gaiti et al., 2017; Hanschen et al., 2016), are not exceptions to this trend (Figure 2C, D, E). The non-uniform distribution of gene loss events across the five species (Figure 3A, C), as well as the signatures of decay found for many of these genes (Figure 3B), 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 (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 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 & Gabaldon, 2020). This supports the adaptive role of gene loss for phenotypic innovation.

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, Anjum, Berg, Andersson, & Sandegren, 2014). Moreover, novel stable regulatory networks can be readily generated via gene loss (Figure 5, Supplementary Figure S15). Hence, both theoretical and biological systems suggest that gene gains are neither the only nor the fastest route to biological innovation. Our data supports the idea that the transition to multicellularity is

driven by re-purposing existing genes for new functions while eliminating others as opposed to by generating of extensive genetic novelty.

The patterns of gene loss and PPI diversification in the volvocine algae illustrate vigorous genomic and proteomic shifts that likely 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 four colonial species, whereas other gene loss patterns were less numerous (Figure 3C). The ancestor of undifferentiated *Gonium* (and the other multicellular species studied) would then appear to be the 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 2G), and it appears to have undergone diversification of histone H3 N-terminal tails (Figure 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 & Nedelcu, 2016). Although the volvocine transition to undifferentiated multicellularity is a single event, differentiated multicellularity has arisen multiple times (Albalat & Canestro, 2016; Nozaki et al., 2000). The volvocine spherical colony morphology has also evolved at least twice (Yamashita et al., 2016); *Astrephomene* is notably similar to *Eudorina* phenotypically, it is closely related to *Gonium* (Goniaceae) and its embryogenesis and development is markedly different from that of other spheroidal volvocines (Yamashita et al., 2016). These examples support the possibility that complex multicellularity is a relatively straightforward consequence of simple multicellularity. Should this be the case, 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 is driven

primarily by gene loss events, allowing for reconfiguration of gene networks, differential use of existing functional repertoires, and reducing the impact of genetic gains.

Materials and 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)(Goodstein et al., 2012; Prochnik et al., 2010) 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* (Brawley et al., 2017)(GCA_002049455.2), *Pyropia yezoensis* (Nakamura et al., 2013)(GCA_009829735.1), *Chondrus crispus* (Collen et al., 2013)(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 zongingiensis* (available on Phytozome)(Goodstein et al., 2012; 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

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 in Standard Volvox Media (SVM) with 0.5% CO₂ bubbling at 30°C, 350 $\mu\text{mol m}^{-1}\text{sec}^{-1}$ light intensity on a 14:10 day:night cycle. Triplicate cultures were harvested by centrifugation at 600g 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 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.

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 & Morgenstern, 2005), in a manner similar to gene predictions for *Volvox* and *Gonium* (Hanschen et al., 2016; Prochnik et al., 2010). The gene models for *Yamagishiella* and *Eudorina* genomes were generated using RNA-seq data and the AUGUSTUS 3.2.3 software package (Stanke & Morgenstern, 2005), in a manner similar to gene predictions for *Volvox* and *Gonium* (Hanschen, Herron, Wiens, Nozaki, & Michod, 2018b;

Hanschen et al., 2016; Prochnik et al., 2010). 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, Stoeckert, & Roos, 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 (Supplementary Figure S16). The inflation value of 1.9 was used for all analyses (see Supplementary File 1).

Identification of Pfam domains

The diversity and abundance of known Pfam domains in the volvocine algae and red algae genomes were identified using the PFAM (Finn et al., 2014) database version 31 and associated Perl scripts available at: <ftp://ftp.ebi.ac.uk/pub/databases/Pfam/Tools/> (El-Gebali et al., 2019)(see Supplementary File 2).

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 2E second panel, Table S4). 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 2E panels 3-6, Figure 2F first panel, Table S5). The first 30 amino acids from each histone were used for analysis of duplication and presence of predicted post-translational modification sites (lysine methylation and acetylation,

arginine methylation) using PmeS (Shi et al., 2012b) and PLMLA (Shi et al., 2012a) respectively. N-tails per histone per species that had different predicted sites with SVM ≥ 0.6 were considered tail variants and quantified (Figure 2F second panel, Table S5). Histone genes were also identified in the five red algae genomes using gene definitions from the *Chondrus* genome annotation (Table S6). Nine internal control genes, Actin, α -2 Tubulin, ATP-A, ATP-B, β -1 Tubulin, β -2 Tubulin, Dynein heavy chain 1 α , Dynein heavy chain 1 β and Ferridoxin were also identified in the volvocine algae genomes using the reciprocal best BLAST hit method (Table S3).

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). 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 2B), transcription factors (Figure 2C), and protein kinases (Figure 2D) for volvocine species and for red algae species. Regression and quartile analyses were performed using custom Python scripts.

Identification of positively selected genes

Genes with dN/dS value > 1 were identified using the PosiGene pipeline (Sahm, Bens, Platzer, & Szafranski, 2017). Two independent PosiGene analyses were performed using *Chlamydomonas* and *Volvox* genome annotations as anchors. Chlorophytes *Chlorella variabilis*, *Micromonas pusilla*, *Ostreococcus tauri* 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,

Yang, & Hinds, 2003; Yang et al., 2016). As a part of the workflow, PosiGene generates a phylogenetic tree from the isoform assignments (Figure 1A, Supplementary figure S17). Distances from LCA were calculated using these trees.

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.

Orthologous groups showing significant contraction (Figure 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*, *Yamagishiella*, *Eudorina* and *Volvox* were retrieved from their respective orthologous groups. Instances of no ortholog found for target loci were subjected to tBLASTx (Altschul, Gish, Miller, Myers, & Lipman, 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, but only comprised 21.9% of the total dataset. 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 Table S7), 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, and Chi square goodness-of-fit tests were applied to determine statistical significance of observed patterns using a uniform distribution as expected probability distribution (Table S7). All analyses except for tBLASTx search were performed using custom Python scripts.

Identification and analysis of known co-opted genes

VARL genes from *Yamagishiella* and *Eudorina* were previously described (Grochau-Wright et al., 2017; M. M. Kirk et al., 1999). Cell cycle genes in *Yamagishiella* and *Eudorina* genomes were annotated as previously described (Hanschen et al., 2016; Prochnik et al., 2010) using tBLASTn. Transcriptome data and manual curation (Hanschen et al., 2016; Prochnik et al., 2010) was used to prepare gene models. Known pherophorins are composed of two pherophorin domains (Pfam DUF3707) that are connected by a variable length hydroxyproline-rich repeat region (D. L. Kirk, Birchem, & King, 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). To identify the MMP genes in the five genomes, the genes containing this Pfam domain were selected for the existence of the [HQ]EXXHXXGXXH motif in the gene model (Hallmann, 2003). The *invA*, *invB*, *invC* and *glsA* genes were identified in *Yamagishiella* and *Eudorina* by the presence of kinesin (PF00225) and DnaJ (PF00226) domains respectively (Cheng et al., 2003; Nishii et al., 2003). To identify the MMP genes in the five genomes, the genes containing this Pfam domain were selected for the existence of the [HQ]EXXHXXGXXH motif in the gene model (Hallmann, 2003). The *invA*, *invB*, *invC* and *glsA* genes were identified in *Yamagishiella* and *Eudorina* by the presence of

kinesin (PF00225) and DnaJ (PF00226) domains respectively (Cheng et al., 2003; Nishii et al., 2003).

Prediction of lineage specific genes

Phylostratigraphy (Domazet-Loso, Brajkovic, & Tautz, 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. 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 (Jones & Cozzetto, 2015; Ward, McGuffin, Bryson, Buxton, & Jones, 2004), 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 10^6 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 Na_3VO_4 , 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 & Markwell, 2007). For blue native PAGE, 20 μg 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 μm 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 five 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.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Nos. MCB-1715894 and MCB-1412738. We wish to acknowledge Professors Hisayoshi Nozaki and Takashi Hamaji for kindly sharing valuable data with us, Hunter Goddard for providing

insight on custom scripts, and Professors Stephen Miller, Christopher Toomajian, and Stephanie Shames for their feedback on early drafts of the manuscript.

Competing interests

The authors declare no competing interests.

References

- Adami, C., Ofria, C., & Collier, T. C. (2000). Evolution of biological complexity. *Proc Natl Acad Sci U S A*, 97(9), 4463-4468. doi:10.1073/pnas.97.9.4463
- Adler, M., Anjum, M., Berg, O. G., Andersson, D. I., & 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(6), 1526-1535. doi:10.1093/molbev/msu111
- Albalat, R., & Canestro, C. (2016). Evolution by gene loss. *Nat Rev Genet*, 17(7), 379-391. doi:10.1038/nrg.2016.39
- Alexander, K., Yang, H. S., & Hinds, P. W. (2003). pRb inactivation in senescent cells leads to an E2F-dependent apoptosis requiring p73. *Mol Cancer Res*, 1(10), 716-728.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol*, 215(3), 403-410. doi:10.1016/S0022-2836(05)80360-2
- Bhattacharya, D., Price, D. C., Chan, C. X., Qiu, H., Rose, N., Ball, S., . . . Yoon, H. S. (2013). Genome of the red alga *Porphyridium purpureum*. *Nat Commun*, 4, 1941. doi:10.1038/ncomms2931
- Bhattacharya, D., Qiu, H., Lee, J., Yoon, H. S., Weber, A. P. M., & 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(1), 81-99. doi:10.1080/07352689.2018.1482364
- Blaby, I. K., Blaby-Haas, C. E., Tourasse, N., Hom, E. F., Lopez, D., Aksoy, M., . . . Prochnik, S. (2014). The *Chlamydomonas* genome project: a decade on. *Trends Plant Sci*, 19(10), 672-680. doi:10.1016/j.tplants.2014.05.008
- Blanc, G., Duncan, G., Agarkova, I., Borodovsky, M., Gurnon, J., Kuo, A., . . . Van Etten, J. L. (2010). The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell*, 22(9), 2943-2955. doi:10.1105/tpc.110.076406
- Bolotin, E., & Hershberg, R. (2015). Gene Loss Dominates As a Source of Genetic Variation within Clonal Pathogenic Bacterial Species. *Genome Biol Evol*, 7(8), 2173-2187. doi:10.1093/gbe/evv135
- Bowles, A. M. C., Bechtold, U., & Paps, J. (2019). The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty. *Curr Biol*. doi:10.1016/j.cub.2019.11.090
- Brawley, S. H., Blouin, N. A., Ficko-Blean, E., Wheeler, G. L., Lohr, M., Goodson, H. V., . . . Prochnik, S. E. (2017). Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc Natl Acad Sci U S A*, 114(31), E6361-E6370. doi:10.1073/pnas.1703088114
- Carroll, S. B. (2001). Chance and necessity: the evolution of morphological complexity and diversity. *Nature*, 409(6823), 1102-1109. doi:10.1038/35059227
- Cheng, Q., Fowler, R., Tam, L. W., Edwards, L., & Miller, S. M. (2003). The role of GlcA in the evolution of asymmetric cell division in the green alga *Volvox carteri*. *Dev Genes Evol*, 213(7), 328-335. doi:10.1007/s00427-003-0332-x

- 561 Coleman, A. W. (2012). A Comparative Analysis of the Volvocaceae (Chlorophyta)(1). *J Phycol*,
562 48(3), 491-513. doi:10.1111/j.1529-8817.2012.01168.x
- 563 Collen, J., Porcel, B., Carre, W., Ball, S. G., Chaparro, C., Tonon, T., . . . Boyen, C. (2013).
564 Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed
565 light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A*, 110(13), 5247-5252.
566 doi:10.1073/pnas.1221259110
- 567 Domazet-Loso, T., Brajkovic, J., & Tautz, D. (2007). A phylostratigraphy approach to uncover
568 the genomic history of major adaptations in metazoan lineages. *Trends Genet*, 23(11),
569 533-539. doi:10.1016/j.tig.2007.08.014
- 570 Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., . . . Villafranca, J. E.
571 (1998). Protein disorder and the evolution of molecular recognition: theory, predictions
572 and observations. *Pac Symp Biocomput*, 473-484.
- 573 Dunker, A. K., Oldfield, C. J., Meng, J., Romero, P., Yang, J. Y., Chen, J. W., . . . Uversky, V. N.
574 (2008). The unfoldomics decade: an update on intrinsically disordered proteins. *BMC*
575 *Genomics*, 9 Suppl 2, S1. doi:10.1186/1471-2164-9-S2-S1
- 576 Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and
577 space complexity. *BMC Bioinformatics*, 5(1), 113. doi:10.1186/1471-2105-5-113
- 578 El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., . . . Finn, R. D.
579 (2019). The Pfam protein families database in 2019. *Nucleic Acids Res*, 47(D1), D427-
580 D432. doi:10.1093/nar/gky995
- 581 Fernandez, R., & Gabaldon, T. (2020). Gene gain and loss across the metazoan tree of life. *Nat*
582 *Ecol Evol*, 4(4), 524-533. doi:10.1038/s41559-019-1069-x
- 583 Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., . . . Punta, M.
584 (2014). Pfam: the protein families database. *Nucleic Acids Res*, 42(Database issue),
585 D222-230. doi:10.1093/nar/gkt1223
- 586 Gaiti, F., Jindrich, K., Fernandez-Valverde, S. L., Roper, K. E., Degnan, B. M., & Tanurdzic, M.
587 (2017). Landscape of histone modifications in a sponge reveals the origin of animal cis-
588 regulatory complexity. *Elife*, 6, e22194. doi:10.7554/eLife.22194
- 589 Go, Y., Satta, Y., Takenaka, O., & Takahata, N. (2005). Lineage-specific loss of function of bitter
590 taste receptor genes in humans and nonhuman primates. *Genetics*, 170(1), 313-326.
591 doi:10.1534/genetics.104.037523
- 592 Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., . . . Rokhsar, D. S.
593 (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids*
594 *Res*, 40(Database issue), D1178-1186. doi:10.1093/nar/gkr944
- 595 Grochau-Wright, Z. I., Hanschen, E. R., Ferris, P. J., Hamaji, T., Nozaki, H., Olson, B., &
596 Michod, R. E. (2017). Genetic basis for soma is present in undifferentiated volvocine
597 green algae. *J Evol Biol*, 30(6), 1205-1218. doi:10.1111/jeb.13100
- 598 Grosberg, R. K., & Strathmann, R. R. (2007). The Evolution of Multicellularity: A Minor Major
599 Transition? *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 621-654.
600 doi:10.1146/annurev.ecolsys.36.102403.114735
- 601 Guijarro-Clarke, C., Holland, P. W. H., & Paps, J. (2020). Widespread patterns of gene loss in
602 the evolution of the animal kingdom. *Nat Ecol Evol*, 4(4), 519-523. doi:10.1038/s41559-
603 020-1129-2
- 604 Hallmann, A. (2003). Extracellular matrix and sex-inducing pheromone in *Volvox*. *Int Rev Cytol*,
605 227, 131-182. doi:10.1016/S0074-7696(03)01009-X
- 606 Hamaji, T., Kawai-Toyooka, H., Uchimura, H., Suzuki, M., Noguchi, H., Minakuchi, Y., . . .
607 Nozaki, H. (2018). Anisogamy evolved with a reduced sex-determining region in
608 volvocine green algae. *Commun Biol*, 1, 17. doi:10.1038/s42003-018-0019-5
- 609 Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., & Michod, R. E. (2018a).
610 Multicellularity Drives the Evolution of Sexual Traits. *Am Nat*, 192(3), E93-E105.
611 doi:10.1086/698301

- 612 Hanschen, E. R., Herron, M. D., Wiens, J. J., Nozaki, H., & Michod, R. E. (2018b). Repeated
613 evolution and reversibility of self-fertilization in the volvocine green algae. *Evolution*,
614 72(2), 386-398. doi:10.1111/evo.13394
- 615 Hanschen, E. R., Marriage, T. N., Ferris, P. J., Hamaji, T., Toyoda, A., Fujiyama, A., . . . Olson,
616 B. J. (2016). The *Gonium pectorale* genome demonstrates co-option of cell cycle
617 regulation during the evolution of multicellularity. *Nat Commun*, 7, 11370.
618 doi:10.1038/ncomms11370
- 619 Herron, M. D., Hackett, J. D., Aylward, F. O., & Michod, R. E. (2009). Triassic origin and early
620 radiation of multicellular volvocine algae. *Proc Natl Acad Sci U S A*, 106(9), 3254-3258.
621 doi:10.1073/pnas.0811205106
- 622 Herron, M. D., & Michod, R. E. (2008). Evolution of complexity in the volvocine algae: transitions
623 in individuality through Darwin's eye. *Evolution*, 62(2), 436-451. doi:10.1111/j.1558-
624 5646.2007.00304.x
- 625 Jones, D. T., & Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with
626 annotated protein-binding activity. *Bioinformatics*, 31(6), 857-863.
627 doi:10.1093/bioinformatics/btu744
- 628 Kirk, D. L. (2005). A twelve-step program for evolving multicellularity and a division of labor.
629 *BioEssays*, 27(3), 299-310. doi:10.1002/bies.20197
- 630 Kirk, D. L., Bircham, R., & King, N. (1986). The extracellular matrix of *Volvox*: a comparative
631 study and proposed system of nomenclature. *J Cell Sci*, 80, 207-231.
- 632 Kirk, D. L., & Nishii, I. (2001). *Volvox carteri* as a model for studying the genetic and cytological
633 control of morphogenesis. *Development Growth & Differentiation*, 43(6), 621-631.
634 doi:10.1046/j.1440-169X.2001.00612.x
- 635 Kirk, M. M., Stark, K., Miller, S. M., Muller, W., Taillon, B. E., Gruber, H., . . . Kirk, D. L. (1999).
636 *regA*, a *Volvox* gene that plays a central role in germ-soma differentiation, encodes a
637 novel regulatory protein. *Development*, 126(4), 639-647.
- 638 Knoll, A. H. (2011). The Multiple Origins of Complex Multicellularity. *Annual Review of Earth and*
639 *Planetary Sciences*, Vol 39, 39(1), 217-239. doi:10.1146/annurev.earth.031208.100209
- 640 Kvitek, D. J., & Sherlock, G. (2013). Whole genome, whole population sequencing reveals that
641 loss of signaling networks is the major adaptive strategy in a constant environment.
642 *PLoS Genet*, 9(11), e1003972. doi:10.1371/journal.pgen.1003972
- 643 Li, L., Stoeckert, C. J., Jr., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for
644 eukaryotic genomes. *Genome Res*, 13(9), 2178-2189. doi:10.1101/gr.1224503
- 645 Lopez-Escardo, D., Grau-Bove, X., Guillaumet-Adkins, A., Gut, M., Sieracki, M. E., & Ruiz-Trillo,
646 I. (2019). Reconstruction of protein domain evolution using single-cell amplified
647 genomes of uncultured choanoflagellates sheds light on the origin of animals. *Philos*
648 *Trans R Soc Lond B Biol Sci*, 374(1786), 20190088. doi:10.1098/rstb.2019.0088
- 649 Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., . . .
650 Grossman, A. R. (2007). The *Chlamydomonas* genome reveals the evolution of key
651 animal and plant functions. *Science*, 318(5848), 245-250. doi:10.1126/science.1143609
- 652 Nakamura, Y., Sasaki, N., Kobayashi, M., Ojima, N., Yasuike, M., Shigenobu, Y., . . . Ikeo, K.
653 (2013). The first symbiont-free genome sequence of marine red alga, *Susabi-nori*
654 (*Pyropia yezoensis*). *PLoS One*, 8(3), e57122. doi:10.1371/journal.pone.0057122
- 655 Nishii, I., Ogihara, S., & Kirk, D. L. (2003). A kinesin, *invA*, plays an essential role in *volvox*
656 morphogenesis. *Cell*, 113(6), 743-753. doi:10.1016/s0092-8674(03)00431-8
- 657 Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S., & Watanabe, M. M. (2000). Origin and
658 evolution of the colonial volvocales (Chlorophyceae) as inferred from multiple,
659 chloroplast gene sequences. *Mol Phylogenet Evol*, 17(2), 256-268.
660 doi:10.1006/mpev.2000.0831

- 661 Nozaki, H., Misumi, O., & Kuroiwa, T. (2003). Phylogeny of the quadriflagellate Volvocales
662 (Chlorophyceae) based on chloroplast multigene sequences. *Mol Phylogenet Evol*,
663 29(1), 58-66. doi:10.1016/S1055-7903(03)00089-7
- 664 Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., . . . Kuroiwa,
665 T. (2007). A 100%-complete sequence reveals unusually simple genomic features in the
666 hot-spring red alga Cyanidioschyzon merolae. *BMC Biol*, 5, 28. doi:10.1186/1741-7007-
667 5-28
- 668 Olson, B. J., & Markwell, J. (2007). Assays for determination of protein concentration. *Curr*
669 *Protoc Protein Sci, Chapter 3*(September), Unit 3 4.
670 doi:10.1002/0471140864.ps0304s48
- 671 Olson, B. J., & Nedelcu, A. M. (2016). Co-option during the evolution of multicellular and
672 developmental complexity in the volvocine green algae. *Curr Opin Genet Dev*, 39, 107-
673 115. doi:10.1016/j.gde.2016.06.003
- 674 Olson, B. J., Oberholzer, M., Li, Y., Zones, J. M., Kohli, H. S., Bisova, K., . . . Umen, J. G.
675 (2010). Regulation of the Chlamydomonas cell cycle by a stable, chromatin-associated
676 retinoblastoma tumor suppressor complex. *Plant Cell*, 22(10), 3331-3347.
677 doi:10.1105/tpc.110.076067
- 678 Palenik, B., Grimwood, J., Aerts, A., Rouze, P., Salamov, A., Putnam, N., . . . Grigoriev, I. V.
679 (2007). The tiny eukaryote Ostreococcus provides genomic insights into the paradox of
680 plankton speciation. *Proc Natl Acad Sci U S A*, 104(18), 7705-7710.
681 doi:10.1073/pnas.0611046104
- 682 Prochnik, S. E., Umen, J., Nedelcu, A. M., Hallmann, A., Miller, S. M., Nishii, I., . . . Rokhsar, D.
683 S. (2010). Genomic analysis of organismal complexity in the multicellular green alga
684 Volvox carteri. *Science*, 329(5988), 223-226. doi:10.1126/science.1188800
- 685 Roth, M. S., Cokus, S. J., Gallaher, S. D., Walter, A., Lopez, D., Erickson, E., . . . Niyogi, K. K.
686 (2017). Chromosome-level genome assembly and transcriptome of the green alga
687 Chromochloris zofingiensis illuminates astaxanthin production. *Proc Natl Acad Sci U S*
688 *A*, 114(21), E4296-E4305. doi:10.1073/pnas.1619928114
- 689 Sahm, A., Bens, M., Platzer, M., & Szafranski, K. (2017). PosiGene: automated and easy-to-use
690 pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Res*,
691 45(11), e100. doi:10.1093/nar/gkx179
- 692 Seppey M., M. M., Zdobnov E.M. . (2019). BUSCO: Assessing Genome Assembly and
693 Annotation Completeness. *Methods in Molecular Biology*, 1962.
- 694 Shi, S. P., Qiu, J. D., Sun, X. Y., Suo, S. B., Huang, S. Y., & Liang, R. P. (2012a). PLMLA:
695 prediction of lysine methylation and lysine acetylation by combining multiple features.
696 *Mol Biosyst*, 8(5), 1520-1527. doi:10.1039/c2mb05502c
- 697 Shi, S. P., Qiu, J. D., Sun, X. Y., Suo, S. B., Huang, S. Y., & Liang, R. P. (2012b). PMeS:
698 prediction of methylation sites based on enhanced feature encoding scheme. *PLoS One*,
699 7(6), e38772. doi:10.1371/journal.pone.0038772
- 700 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of
701 large phylogenies. *Bioinformatics*, 30(9), 1312-1313. doi:10.1093/bioinformatics/btu033
- 702 Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in
703 eukaryotes that allows user-defined constraints. *Nucleic Acids Res*, 33(Web Server
704 issue), W465-467. doi:10.1093/nar/gki458
- 705 Trigoso, A. S., Pearson, R. B., Papenfuss, A. T., & Goode, D. L. (2018). How the evolution of
706 multicellularity set the stage for cancer. *Br J Cancer*, 118(2), 145-152.
707 doi:10.1038/bjc.2017.398
- 708 Wagner, A. (1994). Evolution of gene networks by gene duplications: a mathematical model and
709 its implications on genome organization. *Proc Natl Acad Sci U S A*, 91(10), 4387-4391.
710 doi:10.1073/pnas.91.10.4387

- Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., & Jones, D. T. (2004). The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, 20(13), 2138-2139. doi:10.1093/bioinformatics/bth195
- Waterborg, J. H., Robertson, A. J., Tatar, D. L., Borza, C. M., & Davie, J. R. (1995). Histones of *Chlamydomonas reinhardtii*. Synthesis, acetylation, and methylation. *Plant Physiol*, 109(2), 393-407. doi:10.1104/pp.109.2.393
- Worden, A. Z., Lee, J. H., Mock, T., Rouze, P., Simmons, M. P., Aerts, A. L., . . . Grigoriev, I. V. (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*, 324(5924), 268-272. doi:10.1126/science.1167222
- Yamashita, S., Arakaki, Y., Kawai-Toyooka, H., Noga, A., Hirono, M., & Nozaki, H. (2016). Alternative evolution of a spheroidal colony in volvocine algae: developmental analysis of embryogenesis in *Astrephomene* (Volvocales, Chlorophyta). *BMC Evol Biol*, 16(1), 243. doi:10.1186/s12862-016-0794-x
- Yang, E. C., Boo, S. M., Bhattacharya, D., Saunders, G. W., Knoll, A. H., Fredericq, S., . . . Yoon, H. S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep*, 6, 21361. doi:10.1038/srep21361
- Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., . . . Fei, Z. (2016). iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol Plant*, 9(12), 1667-1670. doi:10.1016/j.molp.2016.09.014