Organelle Genome Complexity Scales Positively with Organism Size in Volvocine Green Algae

David Roy Smith,*[1] Takashi Hamaji,[2] Bradley J.S.C. Olson,[3] Pierre M. Durand,[4] Patrick Ferris,[5] Richard E. Michod,[5] Jonathan Featherston,[4] Hisayoshi Nozaki,[6] and Patrick J. Keeling[1]

[1]Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

[2]Department of Botany, Graduate School of Science, Kyoto University, Oiwake-cho, Kita-shirakawa, Sakyo-ku, Kyoto, Japan

[3]Division of Biology, Kansas State University

[4]Evolutionary Medicine Unit, University of the Witwatersrand, Johannesburg, South Africa

[5]Department of Ecology and Evolutionary Biology, University of Arizona

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

***Corresponding author:** E-mail: smithdr@dal.ca.

Associate editor: Andrew Roger

Abstract

It has been argued that for certain lineages, noncoding DNA expansion is a consequence of the increased random genetic drift associated with long-term escalations in organism size. But a lack of data has prevented the investigation of this hypothesis in most plastid-bearing protists. Here, using newly sequenced mitochondrial and plastid genomes, we explore the relationship between organelle DNA noncoding content and organism size within volvocine green algae. By looking at unicellular, colonial, and differentiated multicellular algae, we show that organelle DNA complexity scales positively with species size and cell number across the volvocine lineage. Moreover, silent-site genetic diversity data suggest that the volvocine species with the largest cell numbers and most bloated organelle genomes have the smallest effective population sizes. Together, these findings support the view that nonadaptive processes, like random genetic drift, promote the expansion of noncoding regions in organelle genomes.

***Key words:* *Chlamydomonas*, *Gonium*, *Pleodorina*, *Volvox*, mitochondrion, chloroplast.**

Letter

Introduction

Why do some genomes contain a large number of noncoding nucleotides while others do not? One explanation might be that increases in noncoding DNA "emerged passively in response to the long-term population size reductions that accompanied increases in organism size" (Lynch and Conery 2003). Support for this hypothesis has come from the observed negative scaling of noncoding DNA with effective population size across major groups on the tree of life (Lynch 2006). Few studies, however, have explored this correlation within groups. This is because there are limited genomic and population genetic data from lineages whose members exhibit dramatic differences in size.

One eukaryotic group that spans the gamut of size and complexity and for which there are accumulating genetic data is the volvocine green algal lineage (Chlorophyceae, Chlorophyta). This monophyletic clade of freshwater flagellates contains unicellular species (e.g., *Chlamydomonas*), colonial forms (e.g., *Gonium*), all the way to complex multicellular taxa with cell differentiation and specialization (e.g., *Volvox*) (Herron et al. 2009; Coleman 2012). Recently, volvocine green algae have been the focus of organelle and nuclear genome sequencing projects, some of which have uncovered highly expanded genomic architectures (Smith and Lee 2009a; Ferris et al. 2010; Prochnik et al. 2010). For instance, the mitochondrial and plastid genomes of the multicellular *Volvox carteri* have massive amounts of noncoding DNA and are more than twice the size of those from the unicellular *Chlamydomonas reinhardtii* (Smith and Lee 2010). The *V. carteri* organelle genomes also harbor very low levels of silent-site genetic diversity, suggesting that they have small effective population sizes—much smaller than those predicted for the *C. reinhardtii* organelle DNAs (Smith and Lee 2010). These findings hint at a potential relationship between organelle genome noncoding content and organism size within the volvocine lineage. But more data from volvocine species with differing levels of cellular complexity are needed to further investigate this trend. Here, by employing data from the Volvocales Genome Initiative, we examine organelle genome evolution within a range of volvocine species.

More Cells, More Noncoding DNA

Using both newly generated and published data, we obtained complete mitochondrial and plastid DNA (mtDNA and ptDNA) sequences from four volvocine species with different levels of organizational complexity, including the single-celled *C. reinhardtii* CC-503, the 8- to 16-celled colonial *Gonium pectorale* NIES-2863, and the partially and fully differentiated multicellular taxa *Pleodorina starrii* NIES-1363 (32–64 cells) and *V. carteri* forma *nagariensis* UTEX 2908 (~2000 cells).

The organelle genomes from these four algae have almost identical gene complements, but they differ greatly in length and compactness (table 1).

We observed that within the volvocine lineage, organelle DNA size and noncoding content scale positively with cell number and organism size (fig. 1 and table 1). For the mitochondrial genomes, the proportion of noncoding nucleotides rises incrementally when moving from the lab strain of *C. reinhardtii* (~20%), to *G. pectorale* (~25%), to *P. starrii* (~40%), to *V. carteri* (>60%), as does the quantity of mitochondrial repeats, which are low in *C. reinhardtii* and *G. pectorale*, moderate in *P. starrii*, and very high in *V. carteri* (fig. 2). Moreover, mitochondrial genome size estimates for the 16-celled colonial alga *Pandorina morum* (~20 kb; Moore and Coleman 1989) reinforce that there is an escalation in mtDNA size across the volvocine line. Similar trends were observed for the plastid genomes, which increase in length as the cell number gets larger: 204 kb (*C. reinhardtii*), 223 kb (*G. pectorale*), 270 kb (*P. starrii*), and ~525 kb (*V. carteri*) (fig. 1 and table 1). The ptDNA repeat content steadily climbed along the unicellular to multicellular continuum (fig. 2), eventually ballooning in *V. carteri*, which has the most expanded ptDNA sequenced to date, from all eukaryotes.

We also found that other types of organelle genome embellishments go up in abundance relative to volvocine organismal complexity (table 1). For instance, the organelle DNAs of *P. starrii* and *V. carteri* have, on average, more introns, nonstandard open reading frames (NSORFs), and pseudogenes than those of the unicellular *C. reinhardtii* and colonial *G. pectorale* (table 1); however, a few strains of *C. reinhardtii* do contain mitochondrial introns (Smith and Lee 2008). Similarly, the intergenic regions of the *P. starrii* ptDNA harbor the decaying remnants of what appear to be group II introns, and in *V. carteri*, unlike the other three algae, palindromic repeats have invaded various organelle DNA coding and intronic regions (fig. 2) (Smith and Lee 2009a).

Species Size, Genetic Drift, and Genome Expansion

Does the size and cellular organization of volvocine algae influence their organelle genome architectures? There is an approximately 10-, 100-, and 1000-fold difference in cell number when comparing *C. reinhardtii* to *G. pectorale*, *P. starrii*, and *V. carteri*, respectively (Hallmann 2011). The doubling time of these algae also increases with cell number, as does their cellular complexity, with *C. reinhardtii* and *G. pectorale* possessing undifferentiated cells and *P. starrii* and *V. carteri* showing partial and full germ-soma differentiation (Hallmann 2011). Moreover, during their sexual cycles, *C. reinhardtii* and *G. pectorale* produce equal-sized gametes, whereas *P. starrii*

Table 1. Organismal and Organelle Genome Complexity within the Volvocine Lineage.

	Chlamydomonas reinhardtii[a]	*Gonium pectorale*	*Pleodorina starrii*	*Volvox carteri* f. *nagariensis*[a]
Organismal features				
Cell number	1	8–16	32–64	~2000
Size (μm)	~10	~30	~150	500–1500
Sex	Isogamy	Isogamy	Anisogamy	Oogamy
Division of labor	None	None	Partial	Full
Mitochondrial inheritance	MT minus	MT minus	n/a	Maternal
Plastid inheritance	MT plus	MT plus	n/a	Maternal
Silent-site genetic diversity	0.015	0.006	<0.001	<0.001
MtDNA				
Size (kb)	15.8	16.0	20.4	~35
AT content (%)	55	61	62	~66
Noncoding (%)	19	27	43	>60
Gene number	12	12	12	12
NSORFs	1	1	3	3
Intron number	0	1	3	3
PtDNA				
Size (kb)	204.2	222.6	269.9	~525
AT content (%)	66	70	65	~57
Noncoding (%)	56	56	63	>80
Gene number	98	97	96	~96
NSORFs	4	1	13	6
Intron number	7	3	15	~9

NOTE.—MT: mating type; n/a: data not available. Organelle genome statistics are based on the following volvocine strains: *C. reinhardtii* CC-503, *G. pectorale* NIES-2863, *P. starrii* NIES-1363, and *V. carteri* UTEX 2908. Cell size statistics come from Hallmann (2011). Intergenic regions, telomeres, and introns were treated as noncoding DNA. Gene number includes only standard protein-, rRNA-, and tRNA-coding genes. Duplicate genes and introns were counted only once. NSORFs include intronic ORFs and other nonstandard genes, such as the *rtl* gene in the *C. reinhardtii* mtDNA. Silent-site genetic diversity statistics are based on ptDNA and were calculated using synonymous, intergenic, and/or intronic nucleotide sites.

[a]The mitochondrial genome size, intron number, and noncoding content for *C. reinhardtii* and *V. carteri* can vary due to optional introns in some strains (Smith and Lee 2008, 2010). Statistics are shown for the strains with the lowest intron contents and for which complete ptDNA sequences are available (i.e., *C. reinhardtii* CC-503 and *V. carteri* UTEX 2908).

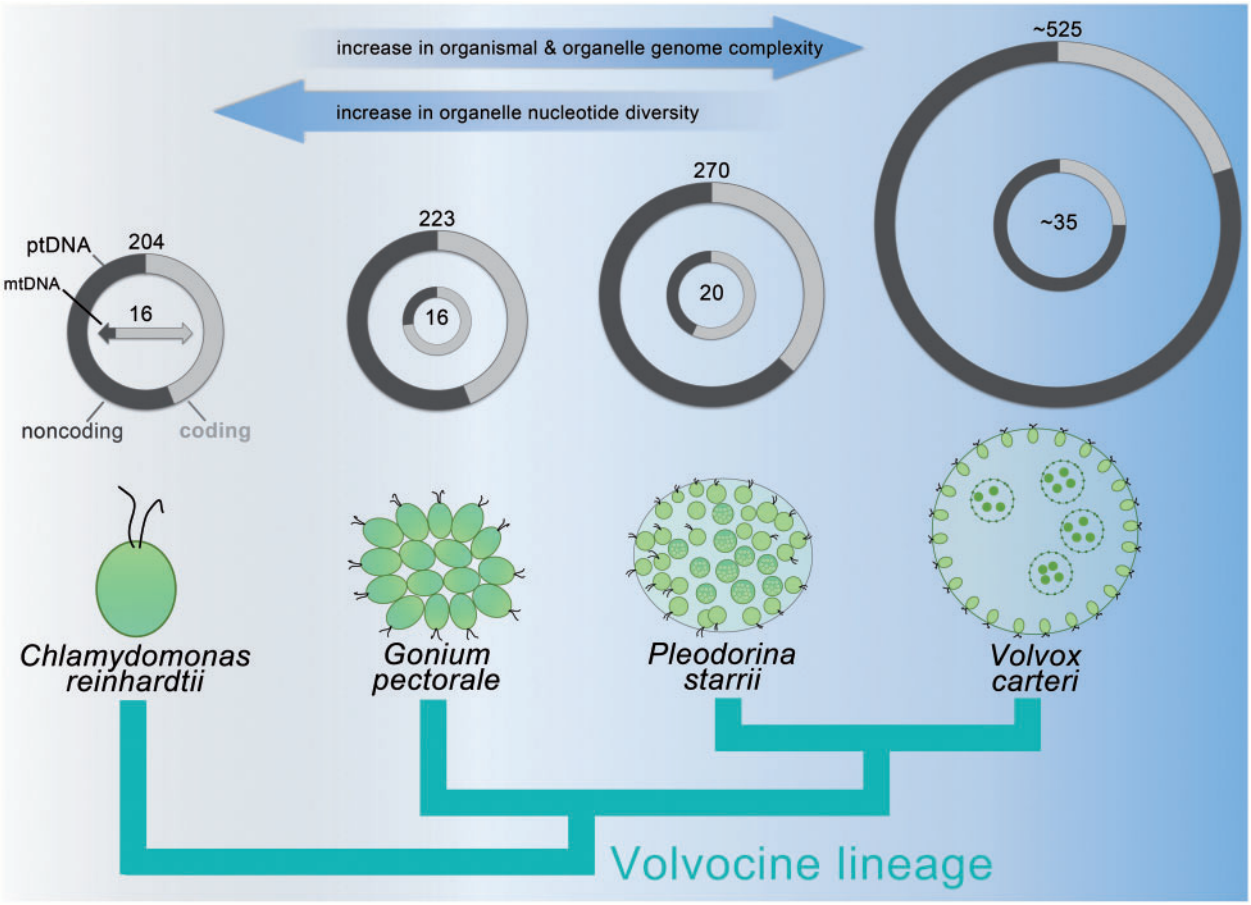

FIG. 1. Scaling of organelle genome size and noncoding content across the volvocine lineage. Mitochondrial and plastid genome maps (outside and inside, respectively) are shown for four volvocine species. Genomes are drawn to scale within categories (size marked in kilobases) but not when comparing between categories (i.e., ptDNA vs. mtDNA). All genomes are circular mapping (supplementary figure S1, Supplementary Material online), with the exception of the *Chlamydomonas reinhardtii* mtDNA, which is linear with terminal inverted repeats. The proportion of coding (light gray) versus noncoding (dark gray) nucleotides is marked on the genome maps. Tree topology is based on phylogenetic analyses (Nozaki et al. 2006; Herron et al. 2009). Organelle genome and organismal features are highlighted in table 1. Genome statistics are based on the following volvocine strains: *C. reinhardtii* CC-503, *Gonium pectorale* NIES-2863, *Pleodorina starrii* NIES-1363, and *Volvox carteri* UTEX 2908.

and *V. carteri* generate dimorphic ones (table 1) (Hallmann 2011).

Given these differences, one might predict that the effective population sizes of *C. reinhardtii* and *G. pectorale* are larger than those of *P. starrii* and *V. carteri*, especially when considering that all four of these algae inhabit similar environments (freshwater ponds and damp soil). Analyses of organelle nucleotide diversity at noncoding and synonymous sites (π_{silent}) support this hypothesis (fig. 1 and table 1). At mutation-drift equilibrium, the silent-site variation of uniparentally inherited organelles genes should approximate $N_e\mu$: the product of the effective population size and the mutation rate per nucleotide site per generation (Lynch 2007). When looking at ptDNA, we found that $N_e\mu$ differed by more than an order of magnitude and scaled inversely with cell number along the volvocine line: >0.01 in *C. reinhardtii*, ~0.005 in *G. pectorale*, and <0.001 in *P. starrii* and *V. carteri* (table 1). $N_e\mu$ estimates for the mtDNA of *C. reinhardtii* and *V. carteri* suggest a similar pattern for the mitochondrial compartment (Smith and Lee 2010).

There is accumulating evidence that the tendency for noncoding DNA to accumulate "depends on both the population size and the mutation rate: the latter defines the burden of excess DNA, while the former defines the ability of natural selection to eradicate it" (Lynch 2007, p. 40). According to this theory, species with a small $N_e\mu$ should be more susceptible to genome expansion than those with a large $N_e\mu$. This is in line with the data presented here and could help explain why multicellular volvocine algae have more exaggerated organelle genomes than their smaller-sized close relatives. Studies on other plastid-bearing lineages, including certain land plants and red algae, suggest that there are a diversity of forces, in addition to mutation rate and population size, impacting organelle genome compactness, and that some of these forces may differ among lineages (Sloan et al. 2012; Smith et al. 2012).

Nuclear Genome Expansion

If the population genetic consequences of multicellularity are contributing to volvocine organelle DNA expansion then

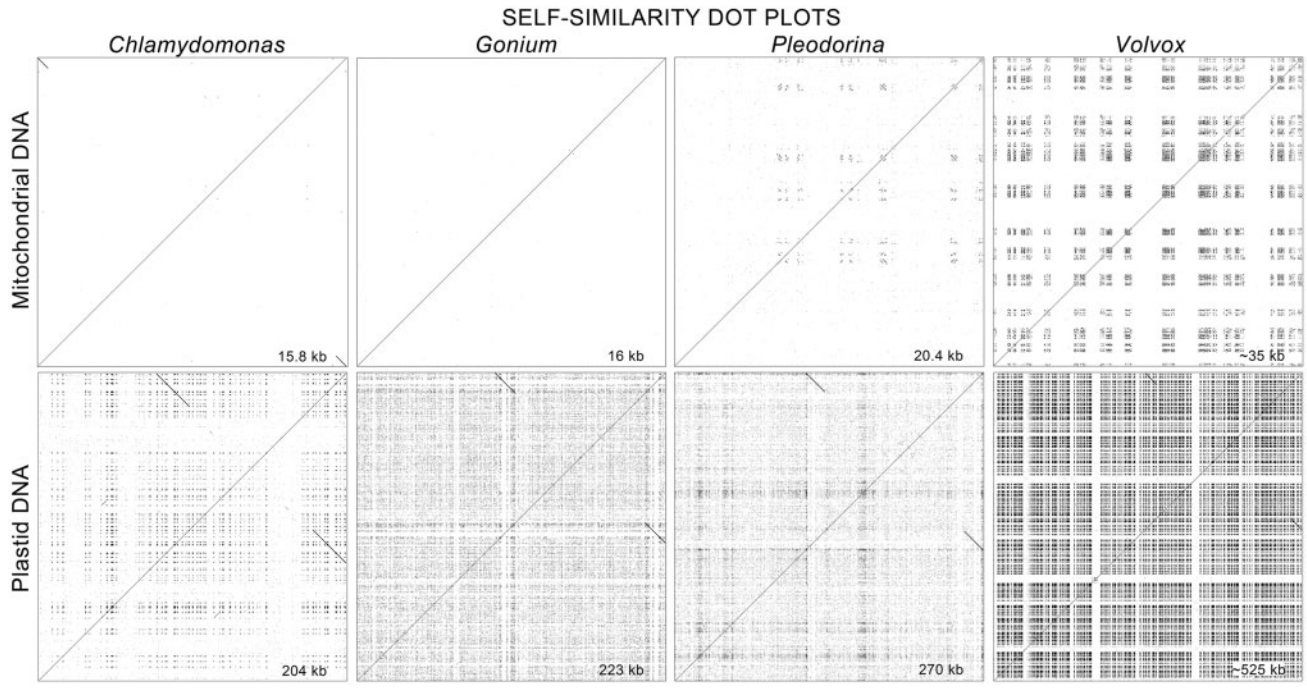

FIG. 2. Dot plot similarity matrices of volvocine organelle genomes. Each matrix contains an organelle genome sequence plotted against itself; the size of the genome is marked in the bottom right corner. Dots within the matrix highlight regions of nucleotide sequence similarity. The main diagonal represents the organelle genome on the *x* axis matching against its partner of the *y* axis. Dots adjacent to the main diagonal correspond to repetitive DNA. The ptDNA inverted repeats and the *Chlamydomonas reinhardtii* mtDNA telomeres are represented by long lines perpendicular to the main diagonal. Dot plots were generated with JDotter (Brodie et al. 2004), using a plot size of 1,000 bases/pixel and a sliding window size of 45.

they may also be impacting nuclear genome architecture. This is precisely what available data suggest. The *V. carteri* nuclear DNA has ~15% more repeats and, on average, 50% longer introns than that of *C. reinhardtii* (Prochnik 2010). Moreover, $N_e\mu$ for the nuclear compartment of *V. carteri* is estimated to be about 20 times lower than that of *C. reinhardtii* (Smith and Lee 2010). Volvocine nuclear DNA expansion is even more pronounced within the mating locus (*MT*)—a multigenic, recombinationally repressed region involved in sex determination (Umen 2011). Like the organelle DNAs, *MT* size and noncoding density are positively related to cell number: 200–300 kb (*C. reinhardtii*), 400–500 kb (*G. pectorale*), and >1000 kb (*V. carteri*) (Ferris et al. 2010; Hamaji et al. 2011). Interestingly, the *V. carteri* mating locus appears to be more recombinationally repressed than that of *C. reinhardtii* (Umen 2011), a feature that is consistent with it having a low N_e and one that may have fostered its bloated architecture (~8.5 introns per gene and ~70% repeats).

Conclusion

We described the gradual inflation in organelle genome noncoding DNA content along the volvocine line from *C. reinhardtii* to *G. pectorale* to *P. starrii* to *V. carteri* and suggest that there is a similar pattern in the nuclear compartment. These observations, combined with data on silent-site nucleotide diversity, raise interesting questions about how differences in organismal size and complexity among closely related species influence population genetic dynamics and consequently genomic architecture. Researchers are only just beginning to understand the population genetics of photosynthetic protists and appreciate the complexities associated with estimating fundamental evolutionary parameters, such as N_e and μ, within these taxa. It is important to stress that the nucleotide diversity and genomic architecture statistics presented here are based on a small sample size and in some cases a few loci. We predict that as more volvocine genome sequences become available, they will bear out the trends described here and provide further insights into how adaptive and nonadaptive processes shape genome structure.

Materials and Methods

P. starrii NIES-1363 (male) was grown in VTAC medium (Nozaki et al. 1989) at 20°C on a 14:10 h light–dark cycle (150–200 µmol photons/m^2/sec). Total DNA was isolated following the protocol of Nozaki et al. (2000), processed into libraries with the Nextera sample prep kit (Epicentre, Madison, Wisconsin, USA), and subjected to paired-end Illumina HiSeq 2000 sequencing (100 nt reads). The *P. starrii* sequence data (~100 Gb) were assembled de novo with Ray v1.2.1 (Boisvert et al. 2010), using a *k*-mer of 21. The resulting contigs were scanned for organelle sequences using BLAST (Altschul et al. 1990) and volvocine organelle genomes as search queries. Two large contigs matching to mtDNA and ptDNA were identified. These contigs were extended using the paired-end data and the "Map to reference" program in Geneious v5.5.8 (Biomatters Ltd., Auckland, NZ), giving complete, circular-mapping mitochondrial and plastid genomes (GenBank accession JX977845 and JX977846,

respectively). Southern blot and restriction digest analyses support the circular conformation of the *P. starrii* mtDNA (supplementary fig. S1, Supplementary Material online). The *G. pectorale* organelle genome data come from strain NIES-2863 (mating type *minus*) and are available in the DNA Data Bank of Japan: AP012493 (mtDNA) and AP012494 (ptDNA). We used the most recent assemblies of the *C. reinhardtii* and *V. carteri* organelle DNAs, which come from strains CC-503 (mating type *plus*) and UTEX 2908 (female), respectively (GenBank accessions EU306622, FJ423446, and GU048820–1) (Smith and Lee 2008, 2009b, 2010). Mitochondrial introns were identified with RNAweasel (Lang et al. 2007). Dot plots were generated with JDotter (Brodie et al. 2004), using a plot size of 1,000 bases/pixel and a sliding window size of 45.

Genetic diversity was calculated with DnaSP v5 (Librado and Rozas 2009), using the Jukes and Cantor correction. The plastid genome π_{silent} estimates for *C. reinhardtii* and *V. carteri* come directly from the literature (Smith and Lee 2008, 2009b, 2010). *G. pectorale* and *P. starrii* ptDNA nucleotide diversity is based on 6,016 and 6,021 protein-coding nucleotide sites from *atpB*, *psaA*, *psaB*, *psbC*, and *rbcL*. *G. pectorale* NIES-2863 (Okinawa Prefecture, Japan) was compared with NIES-569, which originates from Korakuen, Okayama, Japan (GenBank accessions AB014016, AB014017, AB044242, AB044463, AB044521, GPUCPRBCLJ). *P. starrii* NIES-1363 (male) was compared with NIES-1362 (female); both strains represent distinct isolates from Lake Sagami, Kanagawa, Japan (GenBank accessions AB214424, AB214427, AB214430, AB214432, AB214434).

Supplementary Material

Supplementary figure S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Masahiro Suzuki and Hiroko Kawai-Toyooka for help with the *P. starrii* Southern blot analyses. This work was supported by postdoctoral fellowships to D.R.S. from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Izaak Walton Killam Memorial Trusts. P.M.D. is supported by Medical Research Council (SA) and National Health Laboratory Service (SA) grants. P.J.K. is funded by a grant from NSERC (227301) and is a Fellow of the Canadian Institute for Advanced Research.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Boisvert S, Laviolette F, Corbeil J. 2010. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol.* 17:1519–1533.

Brodie R, Roper RL, Upton C. 2004. JDotter: a Java interface to multiple dotplots generated by dotter. *Bioinformatics* 22:279–281.

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

Ferris P, Olson BJ, De Hoff PL, et al. (12 co-authors). 2010. Evolution of an expanded sex-determining locus in *Volvox*. *Science* 328: 351–354.

Hallmann A. 2011. Evolution of reproductive development in the volvocine algae. *Sex Plant Reprod.* 24:97–112.

Hamaji T, Mogi Y, Ferris P, Umen J, Nishii I, Nishimura Y, Nozaki H. 2011. "Mating type locus of *Gonium pectorale*." Abstract presented as part of International *Volvox* Conference. Biosphere 2, Arizona, USA, 2011 December 1–4.

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

Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23:119–125.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Lynch M. 2006. Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol.* 60:327–349.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.

Moore LJ, Coleman AW. 1989. The linear 20 kb mitochondrial genome of *Pandorina morum* (Volvocaceae, Chlorophyta). *Plant Mol Biol.* 13: 459–465.

Nozaki H, Kuroiwa H, Mita T, Kuroiwa T. 1989. *Pleodorina japonica* sp. nov. (Volvocales, Chlorophyta) with bacteria-like endosymbionts. *Phycologia* 28:252–267.

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

Nozaki H, Ott FD, Coleman AW. 2006. Morphology, molecular phylogeny and taxonomy of two new species of *Pleodorina* (Volvoceae, Chlorophyceae). *J Phycol.* 42:1072–1080.

Prochnik SE, Umen J, Nedelcu AM, et al. (25 co-authors). 2010. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329:223–226.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 10:e1001241.

Smith DR, Hua J, Lee RW, Keeling PJ. 2012. Relative rates of evolution among the three genetic compartments of the red alga *Porphyra* differ from those of green plants and do not correlate with genome architecture. *Mol Phylogenet Evol.* 65:339–344.

Smith DR, Lee RW. 2008. Nucleotide diversity in the mitochondrial and nuclear compartments of *Chlamydomonas reinhardtii*: investigating the origins of genome architecture. *BMC Evol Biol.* 8:156.

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

Smith DR, Lee RW. 2009b. Nucleotide diversity of the *Chlamydomonas reinhardtii* plastid genome: addressing the mutational-hazard hypothesis. *BMC Evol Biol.* 9:120.

Smith DR, Lee RW. 2010. Low nucleotide diversity for the expanded organelle and nuclear genomes of *Volvox carteri* supports the mutational-hazard hypothesis. *Mol Biol Evol.* 27: 2244–2256.

Umen JG. 2011. Evolution of sex and mating loci: an expanded view from Volvocine algae. *Curr Opin Microbiol.* 14:634–641.