Molecular Biology and Evolution

Low Nucleotide Diversity for the Expanded Organelle and Nuclear Genomes of *Volvox carteri* Supports the Mutational-Hazard Hypothesis

David Roy Smith* and Robert W. Lee

Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

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

Associate editor: Richard Thomas

Sequence data from this article have been deposited in GenBank under the accession numbers GU084820, GU048821, and GU169471-GU169689.

Abstract

The noncoding-DNA content of organelle and nuclear genomes can vary immensely. Both adaptive and nonadaptive explanations for this variation have been proposed. This study addresses a nonadaptive explanation called the mutational-hazard hypothesis and applies it to the mitochondrial, plastid, and nuclear genomes of the multicellular green alga *Volvox carteri*. Given the expanded architecture of the *V. carteri* organelle and nuclear genomes (60–85% noncoding DNA), the mutational-hazard hypothesis would predict them to have less silent-site nucleotide diversity (π_{silent}) than their more compact counterparts from other eukaryotes—ultimately reflecting differences in $2N_g\mu$ (twice the effective number of genes per locus in the population times the mutation rate). The data presented here support this prediction: Analyses of mitochondrial, plastid, and nuclear DNAs from seven *V. carteri* forma *nagariensis* geographical isolates reveal low values of π_{silent} (0.00038, 0.00065, and 0.00528, respectively), much lower values than those previously observed for the more compact organelle and nuclear DNAs of *Chlamydomonas reinhardtii* (a close relative of *V. carteri*). We conclude that the large noncoding-DNA content of the *V. carteri* genomes is best explained by the mutational-hazard hypothesis and speculate that the shift from unicellular to multicellular life in the ancestor that gave rise to *V. carteri* contributed to a low *V. carteri* population size and thus a reduced $2N_g\mu$. Complete mitochondrial and plastid genome maps for *V. carteri* are also presented and compared with those of *C. reinhardtii*.

Key words: genetic diversity, mitochondrial genome, plastid genome, *Chlamydomonas reinhardtii*, green algae, genome architecture.

Introduction

A striking observation to come from the genomics era is that the amount of noncoding DNA (defined here as intronic and intergenic DNA) in eukaryotic nuclear genomes varies by five orders of magnitude (Lynch and Conery 2003; Gregory 2005a). Similar findings are also observed for organelle genomes (Palmer 1991; Burger et al. 2003; Lynch et al. 2006), for which the noncoding-DNA content can range from 1% to 95% for mitochondrial DNA (mtDNA) and from 5% to 80% for plastid DNA (ptDNA) (fig. 1). The reason(s) why some eukaryotic genomes abound with seemingly useless DNA whereas others are paragons of compactness has proven to be a difficult question to answer, a question that has been dubbed the "C-value enigma" (Gregory 2001), replacing the earlier "C-value paradox," which was concerned with the discordance between genome size and organismal complexity (Mirsky and Ris 1951; Thomas 1971).

There are many hypotheses that confront the C-value enigma; for reviews, see Gregory (2005b), Bennett and Leitch (2005), and Lynch (2007). Some suggest that noncoding DNA is the direct product of natural selection (e.g., Cavalier-Smith 1982), whereas others, such as the selfish DNA hypothesis (Doolittle and Sapienza 1980; Orgel and Crick 1980), maintain that noncoding DNA arises primarily through neutral processes. This study addresses a contemporary nonadaptive hypothesis for the evolution of noncoding DNA called the mutational-hazard hypothesis (Lynch and Conery 2003; Lynch 2006; Lynch et al. 2006) and applies it to the mitochondrial, plastid, and nuclear genomes of the multicellular green alga *Volvox carteri* forma *nagariensis* (Chlorophyceae, Chlorophyta).

The mutational-hazard hypothesis is based on the premise that noncoding DNA is a mutational liability (i.e., more noncoding DNA means more chances for harmful mutations, such as those causing gene expression problems). The hypothesis argues that "the tendency for mutationally hazardous 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). Therefore, organisms with large effective population sizes and high mutation rates are predicted to have more compact genomes than those with small effective population sizes and low mutation rates.

Insights into effective population size (represented in this study as the effective number of gene copies at a locus [N_g],

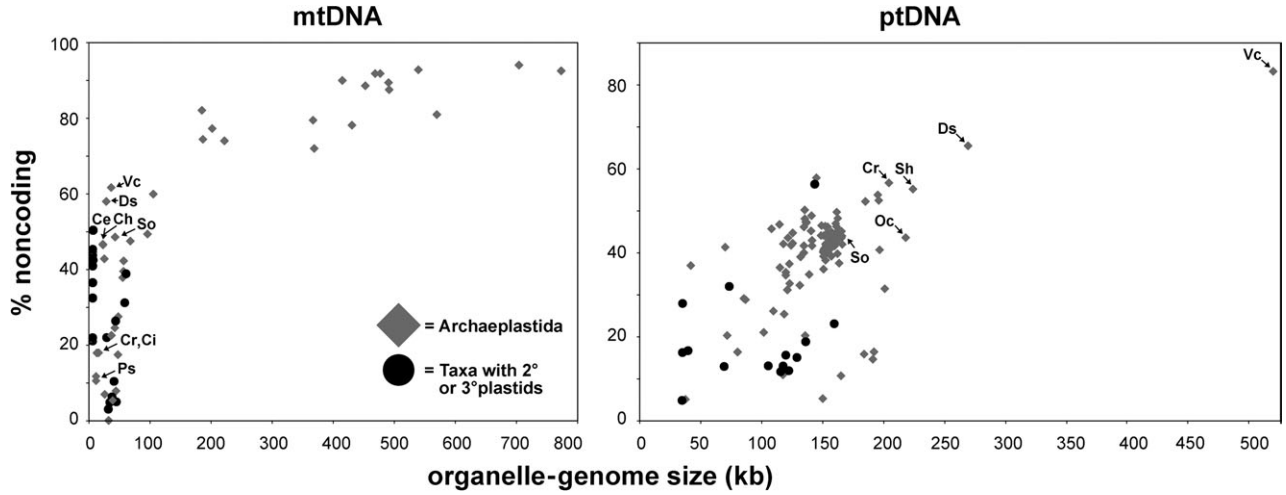

FIG. 1. Scaling of noncoding-DNA content with genome size for completely sequenced organelle DNAs of photosynthetic eukaryotes. Chlorophycean algae are labeled as follows: Ce = *Chlamydomonas eugametos*; Ch = *Chlorogonium elongatum*; Ci = *Chlamydomonas incerta*; Cr = *Chlamydomonas reinhardtii*; Ds = *Dunaliella salina*; Oc = *Oedogonium cardiacum*; Ps = *Polytomella capuana*, *Polytomella parva*, and *Polytomella* sp. 63-10 (SAG); Sh = *Stigeoclonium helveticum*; Sc = *Scenedesmus obliquus*; Vc = *Volvox carteri*. See Smith and Lee (2009b) for a compilation of the noncoding-DNA contents from which this graph was plotted.

not individuals) and mutation rate (μ, defined as the number of mutations per nucleotide site per generation) can be acquired by measuring the nucleotide diversity of silent sites (π_{silent}), which include intergenic and intronic sites and the synonymous sites of protein-coding DNA. Population genetic theory tells us that at mutation-drift equilibrium, the nucleotide diversity at neutral sites ($\pi_{neutral}$) is equal to $2N_g\mu$ (where N_g of uniparentally inherited organelle genes is thought to be about half that of haploid nuclear genes; Birky et al. 1989). Because silent sites are typically regarded as among the most neutrally evolving positions in a genome, measures of π_{silent} can provide an estimate of $2N_g\mu$.

As predicted by the mutational-hazard hypothesis, large-scale studies have found a positive correlation between π_{silent} and genome compactness, with data sets that include bacterial, organelle, and nuclear DNAs (Lynch and Conery 2003; Lynch 2006; Lynch et al. 2006). Opponents of the mutational-hazard hypothesis have argued that the data from which these correlations were derived are weak (Daubin and Moran 2004), and some studies have found that the connection between π_{silent} and genome compactness does not hold up for some organisms or for some genetic compartments (Gregory and Witt 2008; Smith and Lee 2008a, 2009a). However, a general lack of available π_{silent} data has prevented a thorough investigation of the mutational-hazard hypothesis, especially in relation to non-metazoan eukaryotic species. Moreover, the π_{silent} data that are available are often derived from a single locus, and only a few studies have measured π_{silent} for the organelle DNAs and nuclear DNAs (nucDNAs) of the same species.

Already a model for studying the origins of multicellularity and cellular differentiation (Kirk 1998), *V. carteri* is also an ideal species for addressing the mutational-hazard hypothesis. This sexually active alga has three distinct genomes (a haploid nuclear genome and uniparentally inherited mitochondrial and plastid genomes; Adams et al. 1990), all of which contain an abundance of noncoding DNA. Previously, we presented partial mtDNA and ptDNA sequences for *V. carteri* and showed that both these genomes abound with noncoding DNA (~60–80%) and are the most inflated organelle DNAs available from the Chlorophyta (Smith and Lee 2009b). Similarly, the *V. carteri* nuclear genome, sequenced in 2007 by the US Department of Energy Joint Genome Institute (DOE JGI), is ~140 megabases (Mb), ~85% noncoding, and has an average intron-to-gene density of 7.8, placing it among the most bloated nucDNAs observed from a photosynthetic protist. Moreover, when nucDNA-encoded genes from *V. carteri* are compared with their homologues from land plants and animals, *V. carteri* generally has a greater intron/gene ratio (Kirk 1998).

Given the expanded architecture of the *V. carteri* organelle and nuclear genomes, the mutational-hazard hypothesis would predict them to have less silent-site nucleotide diversity than their more compact counterparts from other eukaryotes—ultimately reflecting differences in $2N_g\mu$. Although data on π_{silent} are limited, especially for green algae, we recently measured π_{silent} from the mitochondrial, plastid, and nuclear compartments of the unicellular green alga *Chlamydomonas reinhardtii* (Smith and Lee 2008a, 2009a), a close relative of *V. carteri*. Our π_{silent} data for *C. reinhardtii* were difficult to interpret with respect to the mutational-hazard hypothesis because no other reliable π_{silent} estimates from green algae were available for comparison. Nucleotide diversity data from *V. carteri* would allow for a more thorough evaluation of the forces affecting genome architecture in these two model species. *Volvox carteri* and *C. reinhardtii* make a good pair for investigating the evolution of noncoding DNA because although both species have similar organelle DNA and nucDNA gene numbers, the genomes of *V. carteri* have more noncoding

Table 1. *Volvox carteri* f. *nagariensis* Strains Employed in Study.

Strain	Mating Type	Geographical Origin of Isolation[a]
UTEX 2908[b]	Female	Kobe, Hyogo, Japan
UTEX 1885	Female	Kobe, Hyogo, Japan
UTEX 2864	Male	Kobe, Hyogo, Japan
UTEX 2903	Female	Poona, India
NIES 397	Female	Ichinomiya, Aichi, Japan
NIES 398	Male	Ichinomiya, Aichi, Japan
Isanuma F[c]	Female	Kawagoe, Saitama, Japan
Isanuma M[c]	Male	Kawagoe, Saitama, Japan

[a] See figure 4 for geographical maps of Japan and India highlighting the origins of isolation for the different *V. carteri* strains.
[b] 72–52 dissociator mutant derived from UTEX 1885; this strain was used only to verify our sequences from UTEX 1885—that is, it was not included in our nucleotide diversity analyses.
[c] These strains were kindly provided to us by Stephen Miller of the University of Maryland, Baltimore County. They were originally collected from the Isanuma Pond in the summer of 2005 by Atsushi Nakazawa of Ichiro Nishii's laboratory during an excursion with a group from Hisayoshi Nozaki's laboratory.

DNA than those of *C. reinhardtii* (Merchant et al. 2007; Smith and Lee 2009b and see the *V. carteri* DOE JGI nuclear genome portal: http://genome.jgi-psf.org/Volca1/Volca1.info.html). This is especially true for the mitochondrial and plastid genomes of *V. carteri*, which are more than twice the size of their *C. reinhardtii* counterparts.

Here, we investigate the correlation between $2N_g\mu$ and genome compactness by presenting π_{silent} estimates for the mitochondrial, plastid, and nuclear genomes from seven geographical isolates of *V. carteri* f. *nagariensis*. These data are compared with our previously published π_{silent} values for *C. reinhardtii* and then placed in a broad phylogenetic context. Moreover, by building upon our earlier work (Smith and Lee 2009b), we report complete mitochondrial and plastid genome maps for *V. carteri* (UTEX 2908).

Materials and Methods

Strains and Culture Conditions

The *V. carteri* strains employed in this study (table 1) were obtained from either the Culture Collection of Algae at the University of Texas at Austin (UTEX) or the Microbial Culture Collection at the National Institute for Environmental Studies (NIES) in Ibaraki, Japan, with the exception of the *V. carteri* Isanuma male and female strains, which were graciously provided to us by Stephen Miller of the University of Maryland, Baltimore County. The *V. carteri* strains were grown in Volvox medium (Provasoli and Pintner 1960) at 28 °C under a 16-h light/8-h dark cycle; the illumination at 35 µmol photons m^{-2} s^{-1} photosynthetically active radiation was provided by "cool-white" fluorescent bulbs.

DNA Extraction, Amplification, and Sequencing

For each *V. carteri* strain, total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Germantown, MD) following the manufacturer's protocol. The mtDNA, ptDNA, and nucDNA sequences presented in this study were amplified using a polymerase chain reaction (PCR)-based approach. PCR reactions were performed with High Fidelity Platinum SuperMix (Invitrogen, Carlsbad, CA) using total genomic DNA as the template. PCR products corresponding to regions with extensive DNA repeats were cloned using the TOPO TA Cloning Kit (Invitrogen). Purified PCR products and isolated plasmids were sequenced on both strands using a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA) at the Macrogen Sequencing Facility in Rockville, MD.

Sequencing and Assembling the *V. carteri* (UTEX 2908) Mitochondrial and Plastid Genomes

The complete mitochondrial and plastid genome maps for *V. carteri* strain UTEX 2908 were generated using a two-pronged approach. First, mtDNA and ptDNA trace files generated by the DOE JGI *V. carteri* nuclear genome sequencing project (http://genome.jgi-psf.org/Volca1/Volca1.home.html) were data mined from the National Center for Biotechnology Information (NCBI) *V. carteri* Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/) using *V. carteri* organelle-DNA sequences (Smith and Lee 2009b) as Blast (blastn 2.2.21+) queries. The Blast parameters were as follows: an expectation value (*E* value) of 10; a word size of 11; match and mismatch scores of 2 and −3, respectively; and gap cost values of 5 (existence) and 2 (extension). Trace files showing >90% sequence identity to *V. carteri* organelle DNA in Blast alignments were downloaded and then scanned for trace files originating from inserts (i.e., shotgun sequencing fragments) with lengths greater than 7 kb. Trace files from large inserts were pooled and used to arrange in the proper orientation, the mtDNA and ptDNA contigs generated from our previous work (Smith and Lee 2009b). Long-range PCR was then used to fill the gaps between the organelle-DNA contigs. PCR reactions were performed with the LongRange PCR Kit (Qiagen) using total genomic DNA from *V. carteri* UTEX 2908 as the template. Sequences were assembled using CodonCode Aligner version 3.0.1 (CodonCode Corporation, Dedham, MA), which employs the Phred, Cross-match, and Phrap algorithms for base calling, sequence comparison, and sequence assembly, respectively. Assemblies were performed with a minimum-percent-identity score of 98, a minimum overlap length of 500 nt, a match score of 1, a mismatch penalty of −2, a gap penalty of −2, and an additional first gap penalty of −3.

DNA Sequence Analyses

Nucleotide diversity and its standard deviation were calculated with DnaSP 4.5 (Rozas et al. 2003). Two different methods for calculating silent-site nucleotide diversity were employed: one that excludes indels (indels out), which was used for calculating π_{silent}, and another that considers indels as polymorphic sites (indels in), which was used for measuring $\pi_{silent+}$. For our estimates of $\pi_{silent+}$, indels involving more than one nucleotide were considered to be a single polymorphic site. To ensure that the six nucDNA-encoded genes employed in this study are each present only once in the *V. carteri* nuclear genome, we blasted (blastn 2.2.21+) the six sequences against the DOE JGI *V. carteri* draft nuclear genome sequence (v.1.0., http://genome.jgi-psf.org/Volca1/Volca1.home.html). All six genes returned a single hit, which is consistent with the hypothesis

that these genes are in single copies in the *V. carteri* nuclear genome. The mtDNA and ptDNA sequences obtained from each of the seven strains were also blasted against the *V. carteri* draft nuclear genome sequence to confirm that they were not generated from either nuclear mitochondrial DNAs or nuclear plastid DNAs.

The Fraction of Noncoding DNA in Completely Sequenced Organelle Genomes

Completely sequenced organelle genomes were downloaded from the NCBI Reference Sequence (RefSeq) collection (http://www.ncbi.nlm.nih.gov/projects/RefSeq/) on 1 June 2009. The coding- and noncoding-DNA contents of these sequences were calculated using the following methods and definitions: 1) The number of coding nucleotides in the genome is equal to the collective length of all annotated protein-, rRNA-, and tRNA-coding regions, not including the portions of these regions that are also annotated as introns; 2) the amount of noncoding DNA is the genome length minus the number of coding nucleotides; 3) the number of intergenic nucleotides is equal to the genome length minus the collective length of regions annotated as genes (including their introns and intronic open reading frames); and 4) the amount of intronic DNA is equivalent to the number of noncoding nucleotides minus the number of intergenic nucleotides. The above methods and definitions are contingent on the authors of the GenBank records having properly annotated their entry. If coding regions or introns have been ignored or inaccurately annotated, this will result in incorrect coding- and noncoding-DNA content values. All records were quickly scanned for major errors, but due to the large number of organelle genomes deposited in GenBank, it was unfeasible to thoroughly review every record.

V. carteri Nuclear-Genome Statistics

Length, noncoding-DNA content, and average intron density estimates for the *V. carteri* nucDNA came from the DOE JGI *V. carteri* genome portal information page (accessed on 1 December 2009): http://genome.jgi-psf.org/Volca1/Volca1.home.html.

Nucleotide Sequence Accession Numbers

The mtDNA, ptDNA, and nucDNA sequence data that were used to measure nucleotide diversity are deposited in GenBank under the following accession numbers: GU169471–GU169512 (mtDNA), GU169513–GU169643 (ptDNA), and GU169644–GU169689 (nucDNA). The GenBank accession numbers for the updated mitochondrial and plastid genome sequences of *V. carteri* are GU048821 (mtDNA) and GU084820 (ptDNA), respectively.

Results and Discussion

The *V. carteri* Organelle Genomes

Complete genetic maps of the *V. carteri* UTEX 2908 organelle genomes are shown in figures 2 (mtDNA) and 3 (ptDNA). For comparison, these figures also contain the corresponding genetic maps from *C. reinhardtii* (Gray and Boer 1988; Michaelis et al. 1990; Maul et al. 2002). Previously, we (Smith and Lee 2009b) described partial fragmented genome assemblies of the *V. carteri* mtDNA (7 fragments) and ptDNA (34 fragments). In the current work, we were able to overcome earlier difficulties (Smith and Lee 2009b) and bridge the different organelle genome fragments (i.e., contigs) by data mining the *V. carteri* Trace Archive for organelle-DNA trace files generated from large 8–38 kb shotgun sequencing fragments. These trace files allowed us to position the mtDNA and ptDNA fragments in the correct orientation and from this perform long-range PCR reactions, linking the various organelle-DNA contigs. Although the *V. carteri* organelle genome maps are complete, there still exist several irresolvable gaps within the intergenic and intronic regions of the mtDNA and ptDNA sequence data. These gaps, which are annotated using strings of *n*'s on the *V. carteri* mtDNA and ptDNA GenBank entries, are due to complications with sequencing and assembling the repetitive elements found throughout the *V. carteri* organelle DNA—discussed in Smith and Lee (2009b) and Aono et al. (2002).

The mitochondrial and plastid genome maps of *V. carteri* assemble as circular molecules of ~35 and ~525 kb, respectively (figs. 2 and 3). Whether or not these maps (which are supported by PCR analyses) reflect in vivo unit genome-sized circular structures needs to be confirmed by gel electrophoresis, especially for the mitochondrial genome as all other characterized mtDNAs from *Reinhardtinia*-clade algae (sensu; Nakada et al. 2008) appear to be linear genome-sized or subgenomic-sized structures (Fan and Lee 2002; Laflamme and Lee 2003; Smith and Lee 2008b). Both organelle DNAs are dense with noncoding nucleotides, ~60% (mtDNA) and ~80% (ptDNA), and are among the most bloated organelle genomes observed from photosynthetic eukaryotes (fig. 1). In fact, *V. carteri* has the most expanded mitochondrial genome examined from the Chlorophyta, and its ptDNA has more noncoding DNA, both fractionally and in sheer amount than any other plastid genome sequenced to date—its closest rivals, in this regard, are the plastid genomes of the chlamydomonadalean green algae *Dunaliella salina* (~65.5% noncoding) (Smith et al. 2010) and *C. reinhardtii* (~57% noncoding) and the legume *Trifolium subterraneum* (~58% noncoding) (Maul et al. 2002; Cai et al. 2008). That said, available data suggest that the plastid genomes of some *Acetabularia* species may be upward of 2 Mb (reviewed by Palmer 1985), which implies that they have a noncoding-DNA composition greater than 90%. It should be noted that the *V. carteri cox1* has an optional group-I intron (fig. 2); this intron is present in UTEX 1885 but absent from UTEX 2908 (Smith and Lee 2009b). Finally, the *V. carteri* mtDNA sequence structure appears to be more dynamic than initially thought (Smith and Lee 2009b). Long-range PCR analyses and mtDNA trace file data, gleaned from the *V. carteri* Trace Archive, revealed two mitochondrial genome isomers (A and B), which differ from one another in gene arrangement. Isomer A, depicted in figure 2, was readily amplified by PCR and was the most abundant assembly

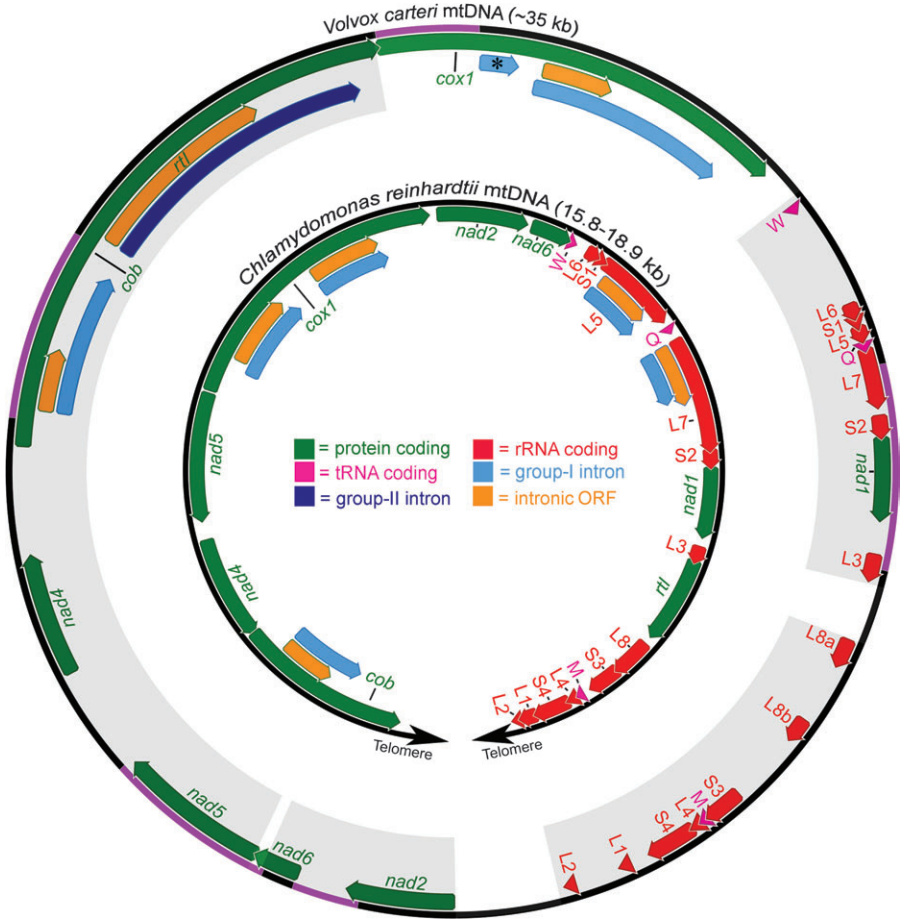

Fig. 2. Complete mitochondrial genome maps for *Volvox carteri* (outer) and *Chlamydomonas reinhardtii* (inner). S1–S4 and L1–L8 represent the small-subunit and large-subunit rRNA-coding modules, respectively. Portions of the *V. carteri* mitochondrial genome map that are shaded gray represent regions of gene colinearity (not including introns) with the *C. reinhardtii* mtDNA, and portions that are pink correspond to the regions that were used for measuring nucleotide diversity. The *V. carteri cox1* group-I intron with an asterisk is optional. The GenBank accession number for the *V. carteri* (UTEX 2908) mitochondrial genome is GU048821 and those for the *C. reinhardtii* mtDNA are EU306617–EU306623. Note that the *C. reinhardtii* mtDNA is a linear molecule, and it can range in size from 15.8 to 18.9 kb, depending on the presence or absence of optional introns (Smith and Lee 2008a).

product from the mtDNA trace file data. Isomer B, shown in supplementary figure S1, Supplementary Material online, was also attainable through PCR but was much less abundant in the mtDNA trace file data; we were not able to confirm circularity or linearity of the isomer B map. Illegitimate recombination between repetitive DNA elements has been suggested as a mechanism for the rearrangement of green algal mitochondrial genomes (Nedelcu 1998; Nedelcu and Lee 1998; Smith and Lee 2008b); such a mechanism might explain the origin of the different isomers of *V. carteri* mtDNA.

V. carteri Strains and their Genetic Loci

We used seven geographical isolates of *V. carteri* f. *nagariensis* to measure nucleotide diversity (table 1). Six of the seven isolates can be divided into three sets of male/female interfertile pairs, which were, respectively, collected close to the Japanese cities of Kobe, Ichinomiya, and Kawagoe; the remaining isolate is a female from Poona, India (table 1 and fig. 4). We could find no published reports about fertility between the different Japanese geographical isolates of *V. carteri*; however, there is evidence of reduced fertility between Indian and Japanese isolates; laboratory-generated hybrid zygotes of the Poona female (UTEX 2903) × Kobe male (UTEX 2864), and the reciprocal cross, had less than 10% of the postzygotic survival compared with zygotes from Kobe × Kobe or Poona × Poona crosses (Adams et al. 1990). These later results suggest that UTEX 2903 is a subspecies within the *V. carteri* f. *nagariensis* complex. As far as we are aware, the seven strains employed here represent the only distinct geographical isolates of *V. carteri* f. *nagariensis* that are currently maintained in public or private culture collections. Although eight strains of *V. carteri* are listed in table 1, UTEX 2908 is derived from UTEX 1885 (Sessoms and Huskey 1973) and is therefore not a distinct geographical isolate. For this reason, the DNA sequences that we generated from UTEX 2908 were used only to verify our UTEX 1885 DNA sequence data and were not included in the nucleotide diversity analyses.

The regions of the *V. carteri* mitochondrial, plastid, and nuclear genomes that were sequenced and used to calculate nucleotide diversity are listed in table 2 and highlighted

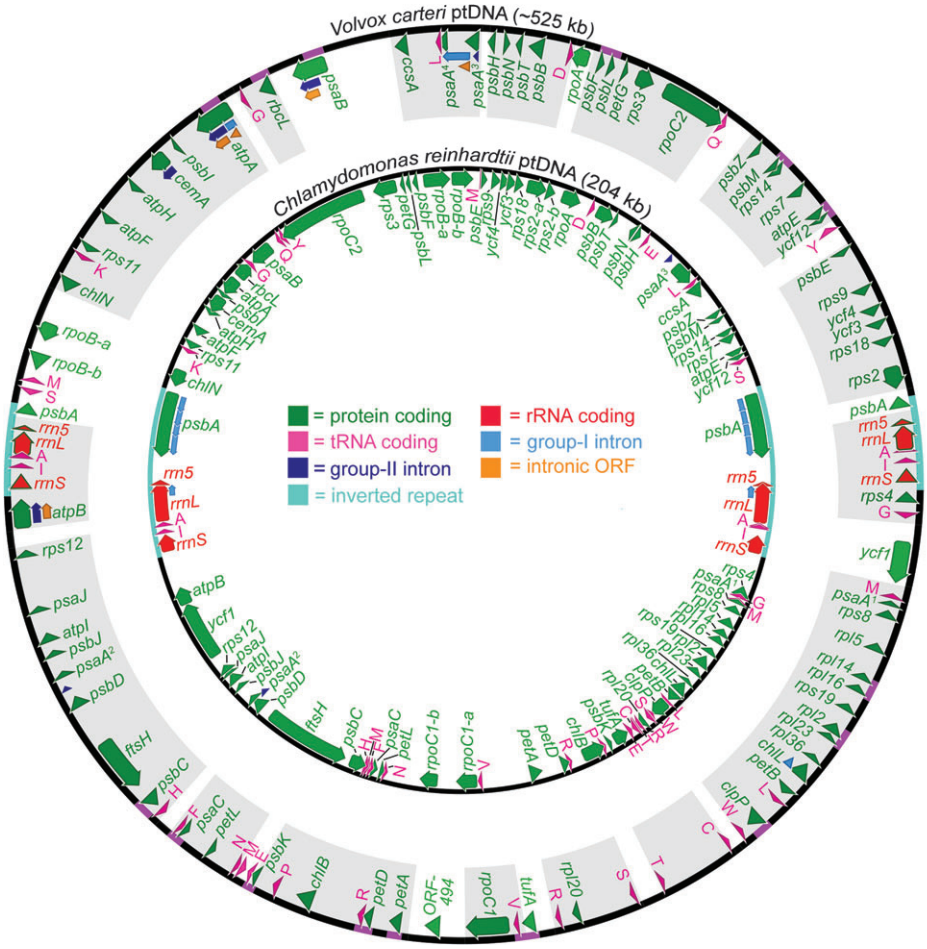

Fig. 3. Complete plastid genome maps for *Volvox carteri* (outer) and *Chlamydomonas reinhardtii* (inner). Portions of the *V. carteri* plastid genome map that are shaded gray represent regions of gene colinearity (not including introns) with the *C. reinhardtii* plastid genome, and portions that are pink correspond to the regions that were used for measuring nucleotide diversity. GenBank accession numbers for the *V. carteri* and *C. reinhardtii* plastid genomes are GU084820 and FJ423446, respectively. For both genomes, the *psaA* gene is fragmented—the translational order of these fragments is signified with superscript numbers.

in pink on the mitochondrial (fig. 2) and plastid (fig. 3) genome maps. Altogether, we sequenced 6 kb of mtDNA, 17.6 kb of ptDNA, and 8 kb of nucDNA from each of the seven *V. carteri* isolates. We made an effort to use intergenic, intronic, and synonymous sites in our calculations of π_{silent}. Other studies on genetic diversity in organelle and nuclear genomes, because of limited intraspecific sequence data, have used mostly synonymous sites for estimating π_{silent}—because silent sites can be under selective constraints (Hershberg and Petrov 2008), it is best to use more than one type of silent site to approximate $2N_g\mu$. We also tried to measure the silent-site nucleotide diversity of *V. carteri* using the same genetic loci that were employed previously for estimating π_{silent} from *C. reinhardtii* (Smith and Lee 2008a, 2009a), thereby allowing for a more meaningful comparison of genetic diversity between these two species. To this end, we were able to sequence the same nucDNA loci from *V. carteri* that were sequenced from *C. reinhardtii*. However, because of differences in gene order between the *V. carteri* and the *C. reinhardtii* organelle genomes and because of the very large intergenic regions found in the *V. carteri* organelle DNAs, we were only able to sequence some of the same organelle-DNA loci.

Nucleotide Diversity of *V. carteri*

Nucleotide diversity measurements for the three genetic compartments of *V. carteri* are summarized in tables 2 and 3. When indels are removed from the aligned DNA sequences, net values of π_{silent} are 0.38×10^{-3} (mtDNA), 0.65×10^{-3} (ptDNA), and 5.28×10^{-3} (nucDNA). Thus, π_{silent} of the nuclear genome is approximately 14 and 8 times that of the mitochondrial and plastid genomes, respectively, and the silent-site diversity of the ptDNA is 1.7 times that of the mtDNA. When our methods for calculating π_{silent} are modified to incorporate indels ($\pi_{silent+}$), by counting each insertion and deletion in the alignment as a nucleotide change (with indels longer than one nucleotide considered as a single polymorphic site), the overall silent-site nucleotide diversity of the mitochondrial compartment remains unchanged (0.38×10^{-3}) and that of nuclear compartment is only slightly higher (5.83×10^{-3} vs. 5.28×10^{-3}); this is because the mtDNA data

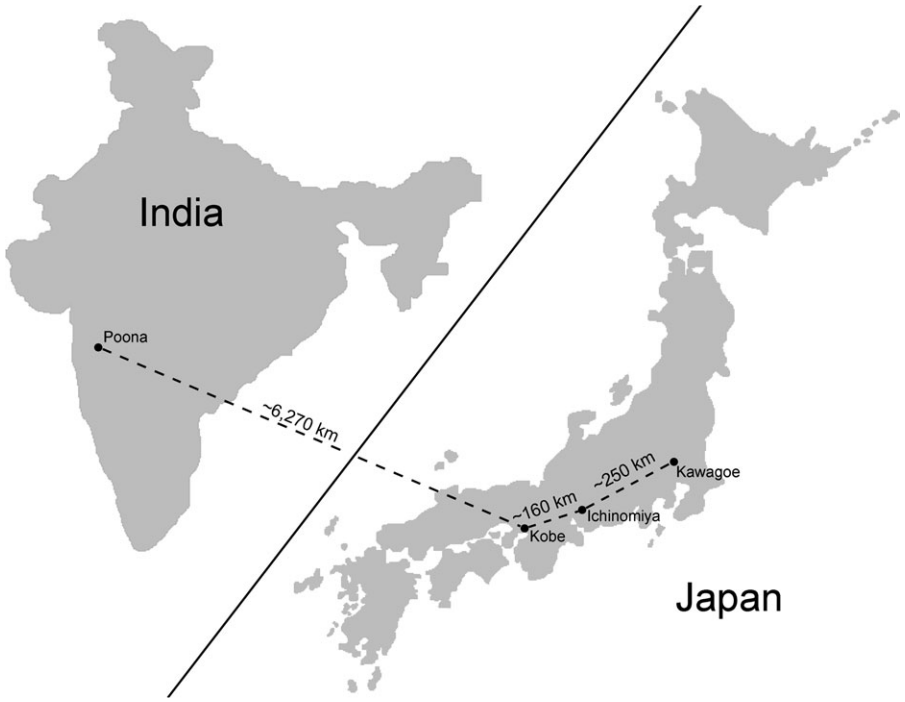

FIG. 4. Geographical maps of Japan and India highlighting the origins of isolation of the *Volvox carteri* strains used in this study. The distances in kilometers between origins of isolation are shown on the map. Refer to table 1 for a list of the *V. carteri* strains.

set contains no indels and the nucDNA sequence data have only 12 indels. For the ptDNA sequence data, $\pi_{silent+}$ was almost twice that of π_{silent} (1.23×10^{-3} vs. 0.65×10^{-3}), reflecting the large number of indels in the sequenced ptDNA regions (tables 2 and 3).

Within each genetic compartment, π_{silent} was relatively constant among the various loci and silent sites that were examined (table 2). Alignments of the sequenced mtDNA loci revealed only four polymorphic positions, all of which are silent: two and one polymorphic synonymous site in *cox1* and *nad6*, respectively, and one polymorphic site in the group-I intron of *cob*. For the ptDNA, polymorphisms are observed in four of the nine intergenic regions and in the group-I intron of *atpA* (table 2). No polymorphic sites were found in the protein-coding ptDNA. And for the nucDNA data, polymorphic sites are distributed almost equally both among the various loci and between synonymous sites and intronic regions. We examined the organelle DNA and nucDNA loci for traces of selection using Tajima's *D*-test (Tajima 1989), which compares the average number of nucleotide differences between pairs of sequences (i.e., π) to the total number of segregating sites (S). Tajima's *D* was not statistically significant when considering the individual loci (table 2) but was significant for the concatenated intron sequences of the nuclear genes *PETC*, *SFA*, and *YPT4* (*D* values of −1.69, −1.71, and −1.7, respectively), which could be a sign of purifying selection. That said, our sample size is most likely too small to critically test for neutrality.

When the Poona strain of *V. carteri* (UTEX 2903) is removed from our nucleotide diversity analyses, there is a substantial decrease in π_{silent} and $\pi_{silent+}$ for both the organelle DNA and the nucDNA data sets: For the mtDNA, π_{silent} and $\pi_{silent+}$ drop to 0.27×10^{-3}; the silent-site nucleotide diversity of ptDNA becomes 0.12×10^{-3} ($\pi_{silent+} = 0.29 \times 10^{-3}$); and for the nucDNA, π_{silent} and $\pi_{silent+}$ drop to 0.10×10^{-3}. This reduction in nucleotide diversity upon the removal of UTEX 2903 is not surprising because gene flow could have been restricted between this strain and that of the other isolates for the following reasons. UTEX 2903 was collected in India at a location, which is more than 6,000 km away from the collection sites of the other six isolates, all of which are located in Japan within ~400 km of each other (fig. 4). More importantly, the reduced fertility between Poona and Kobe isolates (Adams et al. 1990) suggest that there is limited opportunity for gene flow between Indian and Japanese strains. Keep in mind, therefore, that in the following section when we compare the π_{silent} data for *V. carteri* to those from other eukaryotes, the *V. carteri* nucleotide diversity estimates include UTEX 2903 and, thus, may be somewhat inflated as compared with a situation where Japanese and Indian strains where treated as being from separate populations.

π_{silent} of *V. carteri* versus that of *C. reinhardtii* and Other Eukaryotic Species

How do the estimated levels of silent-site nucleotide diversity for *V. carteri* compare with those from other photosynthetic eukaryotes? The only photosynthetic eukaryotes for which we were able to find published π_{silent} data from all three genetic compartments are *C. reinhardtii* and *Arabidopsis lyrata* (Smith and Lee 2008a; Wright et al. 2008; Smith and Lee 2009a). The silent-site nucleotide diversities

Table 2. Nucleotide Diversity Estimates by Region for the *Volvox carteri* Mitochondrial, Plastid, and Nuclear Genomes.

Loci	Number of Sites[a]	S	Number of Indels[b] (length in nucleotides)	$\pi^c \times 10^{-3}$ (SD $\times 10^{-3}$)	$\pi_{syn} \times 10^{-3}$	$\pi_{nsyn} \times 10^{-3}$	Tajima's *D*-test (*P* value)
mtDNA							
cox1	915	2	0	2.19 (1.09)	9.13	0	−1.23 (>0.1)
cob	225	0	0	0	0	0	—
cob (group-I intron)	1,138	1	0	0.42 (0.15)	—	—	0.55 (>0.1)
nad2/nad6	430	0	0	0	—	—	—
nad6	39	1	0	3.6 (1.4)	10.9	0	—
nad6/nad5	21	0	0	0	—	—	—
nad5	1,341	0	0	0	0	0	—
L7	412	0	0	0	—	—	—
L7/S2	113	0	0	0	—	—	—
S2	220	0	0	0	—	—	—
S2/*nad1*	15	0	0	0	—	—	—
nad1	882	0	0	0	0	0	—
nad1/L3	145	0	0	0	—	—	—
L3	133	0	0	0	—	—	—
RNA coding	765	0	0	0	—	—	—
Protein coding	3,420	3	0	0.14 (0.08)	0.48	0	—
Silent sites	2,984	4	0	0.38 (0.17)	—	—	—
ptDNA							
atpA	42	0	0	0	0	0	—
atpA (group-I intron)	1,816	8	5 (55)	1.41 (0.65)	—	—	−1.26 (>0.1)
atpE	138	0	0	0	0	0	—
atpE/ycf12	776	0	0	0	—	—	—
clpP	1,140	0	0	0	0	0	—
petA	657	0	0	0	0	0	—
psaB	1,374	0	0	0	0	0	—
psaC	81	0	0	0	0	0	—
psaC/trnF	776	9	1 (1)	3.32 (2.28)	—	—	−1.59 (>0.05)
trnF	54	0	0	0	—	—	—
psbC	1,230	0	0	0	0	0	—
psbF	54	0	0	0	0	0	—
psbF/psbL	1,161	6	1 (16)	1.05 (0.31)	—	—	−1.40 (>0.1)
psbL	63	0	0	0	0	0	—
rpl16	168	0	0	0	0	0	—
rpl16/rpl14	1,138	0	0	0	—	—	—
rpl14	51	0	0	0	0	0	—
rpl23	33	0	0	0	0	0	—
rpl23/rpl2	1,055	0	2 (364)	0	—	—	—
rpl2	102	0	0	0	0	0	—
tufA/trnV	1,294	1	2 (145)	4.1 (1.5)	—	—	0.55 (>0.1)
trnR/petD	964	6	3 (11)	1.80 (1.24)	—	—	−1.52 (>0.05)
petD	75	0	0	0	0	0	—
psbK	18	0	0	0	0	0	—
psbK/trnE	838	0	1 (3)	0	—	—	—
rps14	132	0	0	0	0	0	—
rps14/psbM	1,239	0	1 (1)	0	—	—	—
rbcL	1,185	0	0	0	0	0	—
Protein coding	6,543	0	0	0	0	0	—
Silent sites	13,238	30	16 (596)	0.65 (0.32)	—	—	—
nucDNA							
ACTIN e8–e10	411	3	0	2.09 (1.43)	8.90	0	−1.32 (>0.1)
ACTIN i8—i9	419	8	0	5.91 (3.22)	—	—	−1.26 (>0.1)
PETC e2–e4	276	0	0	0	0	0	—
PETC i1–i4	901	22	4 (13)	7.08 (4.86)	—	—	−1.69 (<0.05)
SFA e5–e9	450	3	0	1.90 (1.31)	7.87	0	−1.35 (>0.1)
SFA i5–i9	2,363	30	0	3.63 (2.49)	—	—	−1.71 (<0.05)
YPT4 e5–e7	273	1	0	1.05 (0.72)	4.72	0	−1.00 (>0.1)
YPT4 i4–i7	993	26	2 (6)	7.53 (5.17)	—	—	−1.70 (<0.05)
PDK e3	51	0	0	0	0	0	—

Table 2. Continued.

Loci	Number of Sites[a]	S	Number of Indels[b] (length in nucleotides)	$\pi^c \times 10^{-3}$ (SD $\times 10^{-3}$)	$\pi_{syn} \times 10^{-3}$	$\pi_{nsyn} \times 10^{-3}$	Tajima's *D*-test (*P* value)
PDK i3	359	5	5 (18)	4.19 (2.88)	—	—	−1.48 (>0.1)
MAT3 e4–e8	953	1	0	0.52 (0.28)	0	0.69	−0.61 (>0.1)
MAT3 i4—i7	679	5	1 (2)	3.69 (1.96)	—	—	−0.79 (>0.1)
Protein coding	2,415	8	0	1.37 (0.94)	5.82	0	—
Silent sites	6,519	96	12 (39)	5.28 (3.24)	—	—	—

NOTE.—S, number of segregating sites (i.e., polymorphic sites); indels, insertion–deletion events; π, nucleotide diversity; π_{syn}, nucleotide diversity at synonymous sites; π_{nsyn}, nucleotide diversity at nonsynonymous sites; SD, standard deviation. Intergenic regions are labeled using their bordering genes and a dash. For the nuclear loci, exons and introns are, respectively, labeled with an "e" and "i" followed by a number denoting their position within the gene.

[a] Comprises all sites in the nucleotide alignment, including those with indels.

[b] Indels involving more than one nucleotide are counted as a single event. Indel length includes the sum of all indels and includes consecutive indel events.

[c] Only includes sites without indels.

of the *V. carteri* mitochondrial and plastid genomes are both 0.044 times the corresponding values from *C. reinhardtii*, and the silent-site nucDNA diversity of *V. carteri* is approximately 0.16 times that of *C. reinhardtii*—similar proportional relationships are obtained when indels are included in the nucleotide diversity calculations (table 3). Therefore, based on our analyses, the overall π_{silent} estimates for the mitochondrial, plastid, and nuclear genomes of *V. carteri* are much smaller than those of *C. reinhardtii*. Interestingly, π_{silent} for *V. carteri* follows the same general trends that were observed for *C. reinhardtii*: $\pi_{silent(mtDNA)} < \pi_{silent(ptDNA)} < \pi_{silent(nucDNA)}$. Moreover, for both *V. carteri* and *C. reinhardtii*, π_{silent} of mtDNA is half that of the ptDNA. Although for *V. carteri*, π_{silent} of the nucDNA is approximately eight times greater than that of the ptDNA, whereas for *C. reinhardtii*, π_{silent} of the nucDNA is only about twice that of the ptDNA.

Surprisingly, the π_{silent} estimates of *V. carteri* are more similar to those of the land plant *A. lyrata* than they are to those of its close relative *C. reinhardtii*. For example, *V. carteri* and *A. lyrata* have comparable levels of silent-site mtDNA (0.38×10^{-3} vs. 0.35×10^{-3}) and ptDNA (0.65×10^{-3}–1.23×10^{-3} vs. 1.0×10^{-3}) diversities. And the silent-site nucDNA diversity of *V. carteri* is about 0.3 times that of *A. lyrata* (5.28×10^{-3}–5.83×10^{-3} vs. 20×10^{-3}) (Wright et al. 2008); however, most of this difference can be explained by the nuclear genome of *A. lyrata* being diploid and that of *V. carteri* being haploid (N_g of diploid loci is thought to be about twice that of haploid loci). Furthermore, as observed for *V. carteri* and *C. reinhardtii*, the silent-site nucleotide diversity of *A. lyrata* follows the trend: $\pi_{silent(mtDNA)} < \pi_{silent(ptDNA)} < \pi_{silent(nucDNA)}$. Note that no indels were observed in the *A. lyrata* organelle DNA and nucDNA sequences (Wright et al. 2008).

Silent-site nucleotide diversity values for only the mtDNA, ptDNA, or nucDNA of eukaryotic taxa, unaccompanied by data from the neighboring genetic compartments, are more readily available. Many of these π_{silent} estimates are listed in the supplementary materials of Lynch and Conery (2003) and Lynch et al. (2006), who compiled a summary of average π_{silent} values for some major eukaryotic groups and found the following. For the mitochondrial and nuclear genomes of protists, mean values of π_{silent} were 17.5×10^{-3} and 57.3×10^{-3}, respectively. And for land plants, average nucleotide diversity estimates were 0.4×10^{-3} (mtDNA), 3.7×10^{-3} (ptDNA), and 15.2×10^{-3} (nucDNA), which are more or less in accordance with the findings of Wright et al. (2008) for *A. lyrata* and in a general sense reflect the π_{silent} values (and their overall trends) that we observe for *V. carteri*. More recently published π_{silent} data from the ptDNA and nucDNA of land plants (Chen et al. 2008; Breen et al. 2009; Quang et al. 2009) are concordant with earlier approximations. Although reliable silent-site nucleotide diversity data are lacking for land plant mtDNA, it appears to be $<0.4 \times 10^{-3}$ (Lynch et al. 2006), which is in stark contrast to the $\pi_{silent(mtDNA)}$ estimates from animals, which are upward of 35×10^{-3}. That said, both animals and land plants appear to have similar mean $\pi_{silent(nucDNA)}$ estimates (13.4×10^{-3} vs. 15.2×10^{-3}) (Lynch and Conery 2003; Lynch 2006). If we take the above comparisons as a whole, the following conclusions can be made regarding the silent-site nucleotide diversity measurements for the organelle and nucDNAs of *V. carteri*: 1) They are generally lower than the corresponding π_{silent} values from other eukaryotes, 2) they are more similar to those of land plants than they are to those of *C. reinhardtii* and other protists, and 3) they follow the same trend that has been observed for other photosynthetic eukaryotes: $\pi_{silent(mtDNA)} < \pi_{silent(ptDNA)} < \pi_{silent(nucDNA)}$.

Interpreting π_{silent} for *V. carteri*

Do the relative π_{silent} values that we observed across the genetic compartments of *V. carteri* make sense from a population genetic standpoint? When a population is at mutation-drift equilibrium, the nucleotide diversity at silent sites should reflect $2N_g\mu$. In order to interpret our π_{silent} estimates for *V. carteri*, we must therefore consider, for each genetic compartment, the mutation rate and the effective number of genes per locus in the population. Recall that the relative differences of π_{silent} for the mitochondrial, plastid, and nuclear compartments of *V. carteri* are about 1:2:14 when indels are ignored and approximately 1:3:15 when indels are factored in. We can account for at least half of the difference between $\pi_{silent(organelle\ DNA)}$ and $\pi_{silent(nucDNA)}$ if the mitochondrial and plastid genomes of *V. carteri* are transmitted maternally as indicated by

Table 3. Silent-Site Nucleotide Diversity for *Volvox carteri* Compared with that of *Chlamydomonas reinhardtii*.

	mtDNA		ptDNA		nucDNA	
	Vc	Cr	Vc	Cr	Vc	Cr
Number of sites[a]	2,984	5,550	13,238	16,710	6,519	5,051
S	4	104	30	321	96	377
Number of indels[b] (length in nucleotides)	0	11 (31)	16 (596)	86 (1,679)	12 (39)	48 (222)
π_{silent}[c] $\times 10^{-3}$ (SD $\times 10^{-3}$)	0.38 (0.17)	8.51 (1.03)	0.65 (0.32)	14.53 (1.18)	5.28 (3.24)	32.29 (3.01)
$\pi_{silent+}$[d] $\times 10^{-3}$ (SD $\times 10^{-3}$)	0.38 (0.17)	9.23 (1.96)	1.23 (0.75)	18.36 (1.40)	5.83 (3.56)	36.00 (3.51)

NOTE.—Vc, *Volvox carteri*; Cr, *Chlamydomonas reinhardtii*; S, number of segregating sites (i.e., polymorphic sites); indels, insertion–deletion events; π_{silent}, nucleotide diversity at silent sites; $\pi_{silent+}$, nucleotide diversity at silent sites including both polymorphic sites and insertion–deletion events; SD, standard deviation. The *C. reinhardtii* data come from Smith and Lee (2008a, 2009a).

[a] Comprises all sites in the nucleotide alignment, including those with indels.

[b] Indels involving more than one nucleotide are counted as a single event. Indel length includes the sum of all indels and includes consecutive indel events.

[c] Only includes sites in the alignment without indels.

[d] Considers all sites, including those with indels. Consecutive indels are counted as a single polymorphic event. Indel states were measured using a multiallelic approach.

laboratory experiments (Adams et al. 1990), whereas the nuclear genome shows biparental inheritance (Starr 1969). Thus, we would expect N_g of the uniparentally inherited organelle DNAs to be about half that of the nucDNA. Even if we artificially reduce $\pi_{silent(nucDNA)}$ for *V. carteri* by one-half to account for this difference in N_g, we are still left with relative silent-site mtDNA, ptDNA, and nucDNA diversities of about 1:2:7 (indels out) and 1:3:8 (indels in). This remaining variation in π_{silent} among the three genetic compartments of *V. carteri* could be due to a further reduction in N_g as a result of natural selection on linked variation in the organelle genomes, which, because of uniparental inheritance, have less opportunity for recombination during sexual reproduction compared with the nucDNA (Birky et al. 1989). However, the only study to seriously investigate this issue in a photosynthetic eukaryote, namely *A. lyrata*, concluded that N_g of the organelle DNA and nucDNA did not depart significantly from neutral expectations (Wright et al. 2008). Differences in the mutation rates among the three genetic compartments could also contribute to the range of π_{silent} values that are observed for *V. carteri*. Unfortunately, there are no available data for this taxon that bear on this possibility. In Chlamydomonas, however, which also shows lower π_{silent} values in the organelle DNAs, silent-site substitution rate analyses for mtDNA and nucDNA suggest equal rates in these two genetic compartments (Popescu et al. 2006; Popescu and Lee 2007). Alternatively, a recent broad study of synonymous substitutions rates in the mitochondrial, plastid, and nuclear genes of seed plants suggests that the average relative mutation rates of these compartments is 1:3:10 (Drouin et al. 2008), which is almost exactly what we would predict for *V. carteri* if trying to account for the relative values of π_{silent} (after adjusting for differences in N_g).

Finally, why are the π_{silent} values for *V. carteri* significantly lower than those of *C. reinhardtii* (and other protists)? One possibility is that we have underestimated π_{silent} of the *V. carteri* f. *nagariensis* population, potentially because most of the strains that we employed were isolated in Japan within 400 km of each other and the true range of this population is much larger. Indeed, the one strain included in this study that came from outside Japan, the Indian one, showed greater genetic distance to the Japanese isolates than the Japanese isolates did to each other; but this Indian isolate, as mentioned previously, is also known to have reduced sexual compatibility with at least some of the Japanese strains and may not be a legitimate member of the Japanese population. Moreover, in the case of *C. reinhardtii*, which has a population of interfertile members that extend across eastern North America (Harris 2009, p. 16), we found no simple and obvious relationship between genetic distance and the proximity of sites where isolates were recovered (Smith DR and Lee RW, unpublished data), and the same seems to be true for the volvocacean *Pandorina morum* (Kirk 1998). A second possibility is that *V. carteri* has a lower rate of mutation per generation in all three genetic compartments or has fewer numbers of generations per year on average relative to *C. reinhardtii*. A third explanation is that the effective size of the *V. carteri* f. *nagariensis* population is considerably smaller than that of its unicellular cousin *C. reinhardtii* because of a smaller geographical range and/or a lower population density; the later would be expected because *V. carteri* individuals are approximately six orders of magnitude larger than those of *C. reinhardtii*, and both algae can be found in similar habitats (Kirk 1998).

V. carteri Genome Architecture and the Mutational-Hazard Hypothesis

As predicted by the mutational-hazard hypothesis, the *V. carteri* mitochondrial, plastid, and nuclear genomes have less silent-site nucleotide diversity than their more compact counterparts from *C. reinhardtii* and other eukaryotes and similar levels of diversity as those from eukaryotic species with comparable genome architectures, such as land plants (with some exceptions). In other words, our approximations of $2N_g\mu$ for the *V. carteri* organelle and nuclear genomes reflect their noncoding-DNA contents, in that $2N_g\mu$ positively correlates with genome compactness. For instance, the noncoding-DNA densities of the *V. carteri* mitochondrial, plastid, and nuclear genomes are approximately 60–65%, 80%, and 85%, respectively, and the corresponding values for *C. reinhardtii* are about 20–30%, 57%, and 80%. Given these noncoding-DNA contents, the mutational-hazard hypothesis would predict the difference in π_{silent} between these two algae to be greatest for their

organelle genomes. This is indeed the case, the relatively compact mitochondrial and plastid genomes of *C. reinhardtii* each have 22 times more silent-site nucleotide diversity than their more expanded *V. carteri* homologues, whereas the *C. reinhardtii* nuclear genome, which is only slightly more compact than that of *V. carteri*, has only around six times more silent-site nucleotide diversity than the *V. carteri* nucDNA. Similarly, the highly compact mtDNAs of animals (~10% noncoding) have, on average, 30–90 times more silent-site nucleotide diversity than the organelle DNAs of *V. carteri* (Lynch et al. 2006). Yet π_{silent} for the bloated nuclear genomes of animals is only around 2–3 times that of the *V. carteri* nucDNA (Lynch et al. 2006).

Now, let us consider the genome architectures and nucleotide diversities of *V. carteri* in relation to those of land plants. Based on data from completely sequenced land plant genomes, average noncoding-DNA contents are around 84% (mtDNA), 42% (ptDNA), and >80% (nucDNA). Thus, *V. carteri* and land plants have similar mtDNA and nucDNA genome architectures, but the *V. carteri* plastid genome is more expanded than land plant ptDNA (~80% vs. ~42% noncoding). Now, recall that π_{silent} measurements for land plants are in the range of 0.4×10^{-3} (mtDNA), $1–10 \times 10^{-3}$ (ptDNA), and $0–20 \times 10^{-3}$ (nucDNA), with means of 3.7×10^{-3} and 15.2×10^{-3} for the ptDNA and nucDNA, respectively (Lynch et al. 2006). As forecasted by the mutational-hazard hypothesis, $\pi_{silent(mtDNA)}$ for *V. carteri* and land plants are essentially the same ($\sim 0.4 \times 10^{-3}$), $\pi_{silent(ptDNA)}$ for *V. carteri* is around 0.02–0.35 times the mean value for land plants, and $\pi_{silent(nucDNA)}$ for *V. carteri* falls within the range of that for land plants (but is below the mean). For both *V. carteri* and land plants, $\pi_{silent(mtDNA)}$ and $\pi_{silent(ptDNA)}$ are between 0.02 and 0.5 times smaller than $\pi_{silent(nucDNA)}$, despite the fact that the organelle DNAs of *V. carteri* and land plants are equally or more compact than their nuclear counterparts. This suggests that noncoding organelle DNA carries a greater burden than noncoding nucDNA, meaning that the $2N_g\mu$ threshold above which noncoding nucleotides can be perceived and eliminated by natural selection may be lower for organelle genomes than for nuclear genomes. This makes sense if we consider that organelle genomes are typically several orders of magnitude smaller than nuclear genomes. So even though organelle and nuclear genomes can have comparable fractional noncoding-DNA compositions, the sheer volume of noncoding DNA in nuclear genomes will, generally speaking, always be much greater ($10–10^6$ times) than that of organelle genomes. This implies that the number of noncoding nucleotides that are crucial to gene function relative to the number that are inert will be higher for organelle DNA than for nucDNA. This is pertinent to our study if we remember that the central premise of the mutation-hazard hypothesis is that noncoding DNA increases the susceptibility of a genome to degenerative changes and that the mutational disadvantage of noncoding DNA is determined by the number of noncoding nucleotides that are associated with gene function (n), and the mutation rate (μ), where the overall mutational disadvantage (s) of an intron or intergenic region is $n\mu$ (Lynch 2002; Lynch et al. 2006). It is predicted that the threshold in a genome below which stretches of noncoding DNA can expand is $2N_gs < 1$ or alternatively $2N_g\mu < 1/n$ (Lynch 2002; Lynch et al. 2006). Although difficult to estimate, it has been argued that n for the intergenic and intronic regions of organelle DNA is larger than that of nucDNA (Lynch 2007; Smith and Lee 2008a, 2009a). Based on this and earlier studies, it appears that the threshold for noncoding-DNA proliferation may lie somewhere near $2N_g\mu < 0.002$ for organelle DNA and $2N_g\mu < 0.05$ for nucDNA. If the above suppositions are true, it would mean that when addressing the mutational burden of noncoding DNA, it is best to compare like with like: organelle DNA versus organelle DNA and nucDNA versus nucDNA, and so forth. Again, it is important to reemphasize that for all the above π_{silent}/genome compactness comparisons, the silent-site nucleotide diversity values for *V. carteri* include the Poona strain (UTEX 2903). As previously mentioned, if we remove UTEX 2903 from our π_{silent} calculations, it would only make stronger the positive correlations between π_{silent} and genome compactness that we observe.

To summarize, it appears that the large quantities of noncoding DNA in the *V. carteri* organelle and nuclear genomes can be explained by a reduced ability to eradicate excess DNA, caused by a low $2N_g\mu$. If in fact a reduced $2N_g\mu$ is responsible for the expanded architecture of the *V. carteri* organelle and nuclear genomes, could it mean that in an indirect way the evolution of multicellularity in the lineage that gave rise to *V. carteri* was a major contributor to genome expansion? As we have already discussed, the evolution of multicellularity could cause a reduction in N_g. Lynch (2006) proposes that "a central challenge for evolutionary genomics is to determine the extent to which the expansion of eukaryotic gene and genomic complexity was a necessary prerequisite or an indirect consequence of the evolution of complex morphologies." It will be interesting to investigate other volvocine algae with ranging levels of complexity to see if there is a continuum from relatively compact genomes in unicellular species, moderately expanded genomes in the smaller multicelluar forms, to highly bloated genomes in macroscopic multicellular species.

Soon, the DOE JGI will be releasing the nuclear genome sequence from a male isolate of *V. carteri* f. *nagariensis* (presumably UTEX 2864) to complement the nucDNA sequence of the female isolate UTEX 1885 (the plans to sequence this genome are listed under the DOE JGI Community Sequencing Program FY2009). The data generated by this endeavor will allow for a complete view of π for the *V. carteri* organelle and nuclear genomes. It will be interesting to see if the data presented here will be consistent with whole-genome analyses.

Supplementary Material

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

Acknowledgments

We thank Stephen Miller for graciously supplying the Isanuma male and female strains of *V. carteri*. We also thank Patrick Ferris, David Kirk, and Ichiro Nishii for their helpful advice on obtaining geographical isolates of *V. carteri* and for providing detailed histories of certain strains. This work was supported by a grant to R.W.L. from the Natural Sciences and Engineering Research Council (NSERC) of Canada. D.R.S. is an Izaak Walton Killam Memorial Scholar and holds a Canada Graduate Scholarship from NSERC.

References

Adams CR, Stamer KA, Miller JK, McNally JG, Kirk MM, Kirk DL. 1990. Patterns of organellar and nuclear inheritance among progeny of two geographically isolated strains of *Volvox carteri*. *Curr Genet.* 18:141–153.

Aono N, Shimizu T, Inoue T, Shiraishi H. 2002. Palindromic repetitive elements in the mitochondrial genome of *Volvox*. *FEBS Lett.* 521:95–99.

Bennett MD, Leitch IJ. 2005. Genome size evolution in plants. In: Gregory TR, editor. The evolution of the genome. San Diego (CA): Elsevier. p. 89–162.

Birky CW, Fuerst P, Maruyama T. 1989. Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* 121:613–627.

Breen AL, Glenn E, Yeager A, Olson MS. 2009. Nucleotide diversity among natural populations of a North American poplar (*Populus balsamifera*, Salicaceae). *New Phytol.* 182:763–773.

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet.* 19:709–716.

Cai Z, Guisinger M, Kim HG, Ruck E, Blazier JC, McMurtry V, Kuehl JV, Boore J, Jansen RK. 2008. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol.* 67:696–704.

Cavalier-Smith T. 1982. Skeletal DNA and the evolution of genome size. *Annu Rev Biophys Bioeng.* 11:273–302.

Chen H, Morrell PL, de la Cruz M, Clegg MT. 2008. Nucleotide diversity and linkage disequilibrium in wild avocado (*Persea americana* Mill.). *J Hered.* 99:382–389.

Daubin V, Moran NA. 2004. Comment on "The origins of genome complexity". *Science* 306:978.

Doolittle WF, Sapienza C. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284:601–603.

Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol.* 49:827–831.

Fan J, Lee RW. 2002. Mitochondrial genome of the colorless green alga *Polytomella parva*: two linear DNA molecules with homologous inverted repeat Termini. *Mol Biol Evol.* 19: 999–1007.

Gray MW, Boer PH. 1988. Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA. *Philos Trans R Soc Lond B Biol Sci.* 319:135–147.

Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc.* 76:65–101.

Gregory TR. 2005a. Genome size evolution in animals. In: Gregory TR, editor. The evolution of the genome. San Diego (CA): Elsevier. p. 89–162.

Gregory TR. 2005b. Synergy between sequence and size in large-scale genomics. *Nat Rev Genet.* 6:699–708.

Gregory TR, Witt JD. 2008. Population size and genome size in fishes: a closer look. *Genome*. 51:309–313.

Harris EH. 2009. The Chlamydomonas sourcebook, 2nd ed. San Diego (CA): Elsevier.

Hershberg R, Petrov DA. 2008. Selection on codon bias. *Annu Rev Genet.* 42:287–299.

Kirk DL. 1998. Volvox: molecular-genetic origins of multicellularity and cellular differentiation. New York: Cambridge University Press.

Laflamme M, Lee RW. 2003. Mitochondrial genome conformation among CW-group chlorophycean algae. *J Phycol.* 39: 213–220.

Lynch M. 2002. Intron evolution as a population-genetic process. *Proc Natl Acad Sci U S A.* 99:6118–6123.

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.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311:1727–1730.

Maul JE, Lilly JW, Cui L, de Pamphilis CW, Miller W, Harris EH, Stern DB. 2002. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *Plant Cell.* 14:2659–2679.

Merchant SS, Prochnik SE, Vallon O, et al. (117 co-authors). 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318:245–250.

Michaelis G, Vahrenholz C, Pratje E. 1990. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the gene for apocytochrome b and the complete functional map of the 15.8 kb DNA. *Mol Gen Genet.* 223:211–216.

Mirsky AE, Ris H. 1951. The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J Gen Physiol.* 34: 451–462.

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

Nedelcu AM. 1998. Contrasting mitochondrial genome organizations and sequence affiliations among green algae: potential factors, mechanisms, and evolutionary scenarios. *J Phycol.* 34:16–28.

Nedelcu AM, Lee RW. 1998. Short repetitive sequences in green algal mitochondrial genomes: potential roles in mitochondrial genome evolution. *Mol Biol Evol.* 15:690–701.

Orgel LE, Crick FH. 1980. Selfish DNA: the ultimate parasite. *Nature* 284:604–607.

Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annu Rev Genet.* 19:325–354.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. In: Bogorad LK, Vasil I, editors. Cell culture and somatic cell genetics of plants. San Diego (CA): Elsevier. Vol. 7. p. 5–53.

Popescu CE, Borza T, Bielawski JP, Lee RW. 2006. Evolutionary rates and expression level in Chlamydomonas. *Genetics* 172:1567–1576.

Popescu CE, Lee RW. 2007. Mitochondrial genome sequence evolution in Chlamydomonas. *Genetics* 175:819–826.

Provasoli L, Pintner IJ. 1960. Artificial media for fresh-water algae: problems and suggestions. In: Tyron CA, Hartman RT, editors. The ecology of algae. Pittsburgh (PA): University of Pittsburgh Press. p. 84–96.

Quang ND, Ikeda S, Harada K. 2009. Patterns of nucleotide diversity at the methionine synthase locus in fragmented and continuous populations of a wind-pollinated tree, *Quercus mongolica* var. *crispula*. *J Hered.* 100:762–770.

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.

Sessoms AH, Huskey RJ. 1973. Genetic control of development in Volvox: isolation and characterization of morphogenetic mutants. *Proc Natl Acad Sci U S A*. 70:1335–1358.

Smith DR, Lee RW. 2008a. 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. 2008b. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol Biol Evol.* 25:487–496.

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

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

Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, Polle JEW. 2010. The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biol*. 10:83.

Starr RC. 1969. Structure, reproduction, and differentiation in *Volvox carteri* f. *nagariensis* Iyengar, strains HK 9 and 10. *Arch Protistenkunde*. 111:204–222.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Thomas FC. 1971. The genetic organization of chromosomes. *Annu Rev Genet.* 5:237–256.

Wright SI, Nano N, Foxe JP, Dar VU. 2008. Effective population size and tests of neutrality at cytoplasmic genes in *Arabidopsis*. *Genet Res.* 90:119–128.