Protist, Vol. 163, 105–115, January 2012
http://www.elsevier.de/protis
Published online date 28 May 2011

Protist

ORIGINAL PAPER

Similar Relative Mutation Rates in the Three Genetic Compartments of Mesostigma and Chlamydomonas

Jimeng Hua[a], David Roy Smith[a,b], Tudor Borza[a], and Robert W. Lee[a,1]

[a]Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4R2, Canada
[b]Current address: Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Submitted December 5, 2010; Accepted April 2, 2011
Monitoring Editor: Michael Melkonian.

Levels of nucleotide substitution at silent sites in organelle versus nuclear DNAs have been used to estimate relative mutation rates among these compartments and explain lineage-specific features of genome evolution. Synonymous substitution divergence values in animals suggest that the rate of mutation in the mitochondrial DNA is 10-50 times higher than that of the nuclear DNA, whereas overall data for most seed plants support relative mutation rates in mitochondrial, plastid, and nuclear DNAs of 1:3:10. Little is known about relative mutation rates in green algae, as substitution rate data is limited to only the mitochondrial and nuclear genomes of the chlorophyte *Chlamydomonas*. Here, we measure silent-site substitution rates in the plastid DNA of *Chlamydomonas* and the three genetic compartments of the streptophyte green alga *Mesostigma*. In contrast to the situation in animals and land plants, our results support similar relative mutation rates among the three genetic compartments of both *Chlamydomonas* and *Mesostigma*. These data are discussed in relation to published intra-species genetic diversity data for the three genetic compartments of *Chlamydomonas* and are ultimately used to address contemporary hypotheses on the organelle genome evolution. To guide future work, we describe evolutionary divergence data of all publically available *Mesostigma viride* strains and identify, for the first time, three distinct lineages of *Mesostigma*.

Key words: *Chlamydomonas*; genome architecture; *Mesostigma*; mutation rate; silent substitution rate.

Introduction

Knowledge of mutation rate is essential for understanding biological evolution. Although difficult to estimate, insights into this fundamental parameter can be gained by measuring the rate of nucleotide substitution at silent sites (defined as noncoding sites and the synonymous positions of protein-coding DNA) between closely related species (Kimura 1983). Significant advances in our understanding of mutation and genome evolution have come from nucleotide substitution rate data, especially with respect to the evolution of nuclear versus organelle genomes (Graur and Li 2000; Lynch 2007).

Wolfe and colleagues (1987, 1989) reported that for angiosperms the silent-site substitution rates of genes in the mitochondrial DNA (mtDNA) were three-fold lower than those in the plastid DNA (ptDNA) and 12-times lower than those in the nuclear DNA (nucDNA) — more recent

[1]Corresponding author; fax +1 902 494 3736
e-mail robert.lee@dal.ca (R.W. Lee).

doi:10.1016/j.protis.2011.04.003

calculations using an expanded data set and the maximum likelihood (ML) method reveal an even greater relative synonymous substitution rate for the nuclear genome of angiosperms and an overall relative rate for seed plant mitochondrial, plastid, and nuclear genes of 1:3:10 (Drouin et al. 2008). This trend, although having some conspicuous exceptions (Mower et al. 2007 and references therein), is in sharp contrast to the situation for most animals where the silent-site substitution rate of the mtDNA is estimated to be 10-50 times that of the nucDNA (Brown et al. 1979, 1982; Lynch et al. 2006; Miyata et al. 1982). The disparity in relative rates between these two lineages appears to be a reflection of differing organelle DNA mutation rates, as both groups are believed to have similar nucDNA mutation rates (Lynch 2007; Lynch et al. 2006).

The above data have been used to help explain why land plants and animals have comparable nuclear genome architectures but organelle genomes with opposing levels of genomic complexity. Land plant mitochondrial genomes are capacious, intron- and repeat-dense, and undergo high levels of post-transcriptional editing (i.e., RNA editing), whereas animal mtDNA is compact, intron- and gene-poor, and with few exceptions is not post-transcriptionally edited (e.g., Janke and Pääbo 1993; Yokobori and Pääbo 1995). The structure of land plant plastid genomes, which generally have a moderate noncoding DNA content and experience minor levels of RNA editing, is intermediate to that of land plant and animal mtDNAs (Table 1). It is argued that the mutationally active mtDNA of animals makes for a less permissive environment for the accumulation of genomic embellishments than the mutationally quiescent land plant mtDNA. The argument being that escalations in genomic complexity, such as the addition of introns or regulatory sites for RNA editing, increase the mutational liability of the genome because they represent targets for potentially deleterious mutations, where the higher the mutation rate the greater the burden of the adornment (Lynch 2007). It is also suggested that high organelle DNA mutation rates promote the migration of organelle-DNA-encoded genes to the nuclear genome (e.g., Berg and Kurland 2000).

Green algae are an interesting case study for the evolution of genome architecture because they harbor a diverse array of genomic structures. Little is known, however, about either the absolute or relative silent-site substitution rates in this important group, which includes the Chlorophyta (a lineage containing most of the green algal diversification) and the algal members of its sister group, the Streptophyta (a lineage also containing the land plants) (Becker and Marin 2009).

In a broad sense, the organelle and nuclear genome architectures of both chlorophyte and streptophyte green algae are distinct from those of land plants and animals. Green algae tend to have relatively streamlined, intron-poor mitochondrial genomes, which are more compact than those of land plants but more complex than animal mtDNAs. Green algal plastid genomes are generally larger and more bloated than their mitochondrial counterparts, but for land plants the opposite is true. Moreover, there is no evidence of RNA editing in green algal organelle DNA (Lenz et al. 2010 and references therein). The nuclear genomes of green algae, with some exceptions, are smaller, have fewer genes, and less noncoding DNA than those of land plants and animals (Merchant et al. 2007; Prochnik et al. 2010; Worden et al. 2009).

To the best of our knowledge, the only studies to investigate micro-evolutionary rates in both the nuclear and organelle genomes of green algae (or any other algal group in the Plantae) are those of Popescu et al. (2006) and Popescu and Lee (2007), which compared the gene sequences of *Chlamydomonas reinhardtii* and *Chlamydomonas incerta* (model unicellular species belonging to the chlorophycean class of the Chlorophyta) and showed that the silent-site substitution rates of the mitochondrial and nuclear compartments are approximately the same. Here we build upon these works and analyze the silent-site substitution rate between the plastid compartments of *C. reinhardtii* and *C. incerta*, allowing for a complete picture of relative evolutionary rates in this lineage. *Mesostigma viride* is a model unicellular streptophyte green alga (Lemieux et al. 2007 and references therein; Marin and Melkonian 1999) and our preliminary analysis of published EST sequence data from two different *Mesostigma viride* strains indicate that these likely originate from deeply divergent populations or correspond to distinct biological species. Therefore, to broaden our understanding of evolutionary rates in green algae beyond *Chlamydomonas*, we analyzed the rates of silent-site substitution in the mitochondrial, plastid, and nuclear genomes between these two *Mesostigma* lineages and to guide future studies on genetic divergence in this taxon we compared homologous sequences from all publically available *Mesostigma viride* strains and prepared a phylogenetic tree of these sequences. Altogether, these data provide novel insights into the forces governing genome evolution in green algae and green plants as a whole.

Table 1. Mean architectural features of organelle genomes for green algae and land plants.

Phylogenetic Group	Genome Size (kb)	% Noncoding	# of introns	# of genes	# of introns/ gene	*n*
MITOCHONDRIAL DNA						
Chlorophyta	35.6	35.3	8	33	0.21	19
Chlorophyceae	23.3	39.6	5	15	0.36	10
Prasinophyceae	40.3	17.6	1	57	0.02	4
Trebouxiophyceae	43.2	37.7	3	48	0.06	3
Ulvophyceae	76.3	45.9	5	56	0.09	2
Streptophyte green algae	92.1	47.2	16	68	0.23	4
Land plants	431.2	84.3	24	58	0.43	22
PLASTID DNA						
Chlorophyta	179.3	41.3	8	98	0.09	22
Chlorophyceae	300.1	60.3	18	97	0.18	7
Prasinophyceae	106.9	26.2	2	100	0.02	6
Trebouxiophyceae	116.9	31.6	2	96	0.02	6
Ulvophyceae	167.1	49.4	16	104	0.15	3
Streptophyte green algae	151.5	41.2	10	129	0.08	6
Land plants[a]	147.8	41.9	24	107	0.39	107

n, sample size. Data used to calculate means come from completely sequenced organelle genomes deposited in GenBank.
[a]For land plant plastid genomes, many of the GenBank entries were incompletely or incorrectly annotated; thus, intron and gene contents were based on the 24 entries of which we were most confident.

Results

Sequence Analysis of all Available *M. viride* Strains

Two large EST datasets in GenBank, one from NIES-296 (EC726859- EC732477) (Nedelcu et al. 2006) and the other allegedly from NIES-476 (DN254242-DN264595) (Simon et al. 2006) have considerable sequence divergence between homologous sequences (discussed further below). In preliminary experiments using our own DNA sequence data we confirmed the origin of the published EST sequences for NIES-296 but could not do so for NIES-476; a number of mitochondrial DNA sequences that we prepared from our stock of NIES- 476 (unpublished data) were identical to published NIES-296 values (NC008240) (Turmel et al. 2002) and did not match mitochondrial cDNA sequences we found in the EST library of Simon et al. (2006). With the goal of identifying the *Mesostigma* strain employed by Simon et al. (2006), we obtained fresh stocks of all six unique strains of *M. viride* available in culture collections as of September 2010 (Supplementary Table S1) and obtained *GapA* DNA sequence from each strain (GenBank accession numbers: NIES-296, HQ668005; NIES-475, HQ668008; NIES-476, HQ668009; NIES-477, HQ668010; NIES-995, HQ668011; SAG 50-1, HQ668004), a gene locus that is represented in the library of Simon et al. (2006). The results showed that the *GapA* cDNA consensus sequence we obtained from the Simon et al. (2006) ESTs was identical to the *GapA* DNA sequence we prepared from SAG 50-1 but had a dissimilarity value with the counterparts from NIES-296, NIES-475, NIES-476, and NIES-477 of about 5% each and with the counterpart from NIES-995 of about 8%. Based on these analyses we conclude that SAG 50-1 rather than NIES-476 was used in the study reported by Simon et al. (2006).

Nucleotide Substitution Rates in *Mesostigma* and *Chlamydomonas*

Protein-coding Genes

To estimate synonymous-site substitution rates for *Mesostigma*, 31 mitochondrial, 44 plastid, and 67 nuclear protein-coding genes (Supplementary Table S2) were compared between NIES-296 and SAG 50-1. On average, the number of synonymous substitutions per synonymous site of protein-coding genes (dS) for the mtDNA, ptDNA and nucDNA loci were 0.17, 0.11, and 0.27, respectively — these values are all significantly different from one another ($p < 0.05$). Similar substitution rate estimates for the three compartments were obtained when dS was calculated from a concatenation of the coding regions (Table 2). The variation in dS was ~10-fold among mitochondrial-encoded

Table 2. Comparison of silent substitution rates between *Mesostigma viride* strains NIES-296 and SAG 50-1, and between *Chlamydomonas incerta* and *Chlamydomonas reinhardtii*.

	Synonymous sites [a] (dS)	Intronic sites		Intergenic sites
		ORF [e]	Non-ORF [f]	
Mesostigma				
mtDNA				
Average (sd)	0.17 (0.11)	0.42(0.10)	0.13 (0.08)	0.30 (0.28)
Maximum	0.58	0.51	0.29	0.89
Minimum	0.06	0.31	0.04	0.03
Concatenation [g]	0.16 (7476)	0.42 (749)	0.13 (6791)	0.31 (5599)
ptDNA				
Average (sd)	0.11 (0.06)	/	/	0.03 (0.02)
Maximum	0.27	/	/	0.08
Minimum	0.04	/	/	0.00
Concatenation [g]	0.10 (11047)	/	/	0.03 (3318)
nucDNA				
Average (sd)	0.27 (0.18)	/	0.37 (0.19)	-
Maximum	0.83	/	0.73	-
Minimum	0.01	/	0.17	-
Concatenation [g]	0.21 (11875)	/	0.34 (1954)	-
Chlamydomonas				
mtDNA				
Average (sd)	0.29 (0.05) [b]	0.85 [d]	0.24 [d]	0.67 (0.35)
Maximum	0.38 [b]	0.85 [d]	0.24 [d]	1.01
Minimum	0.22 [b]	0.85 [d]	0.24 [d]	0.21
Concatenation [g]	0.30 (2566)	0.85 (237) [d]	0.24 (364) [d]	0.56 (1064)
ptDNA				
Average (sd)	0.30 (0.11)	/	0.11 (0.06)	0.38 (0.14)
Maximum	0.41	/	0.16	0.55
Minimum	0.09	/	0.07	0.19
Concatenation [g]	0.23 (1879)	/	0.12 (2127)	0.37 (3511)
nucDNA[c]				
Average (sd)	0.37 (0.29) [c]	/	1.48 (0.86)	-
Maximum	1.68 [c]	/	3.57	-
Minimum	0.03 [c]	/	0.37	-
Concatenation [g]	0.37 (14206)	/	1.05 (6364)	-

[a]Standard mitochondrial, plastid, and nuclear protein-coding genes. [b]Data were obtained from Popescu and Lee 2007. [c]Data were obtained from Popescu et al. 2006. [d]Only one intron (*cob*). [e]Intronic ORFs (open reading frames) were obtained according to the annotation of published records and the rates are for substitutions per synonymous site. [f]Intronic sites without ORFs. [g]Numbers in parentheses indicate the length of alignment, for protein-coding genes and ORFs, length is number of codons, for other sequences, length is number of nucleotides. "/" indicates absence from the genes examined. "–" indicates that we don't have intergenic sequences from the nuclear genome that can be compared. Note that in the standard laboratory strains of *C. reinhardtii* (e.g., CC-277) there are no introns in the mtDNA (NC_001638), however, another geographical isolate of *C. reinhardtii* (CC-1373) contains an intron (EU_306617) at the same position as the one in *C. incerta* mtDNA and this was used to calculate substitution rates in intronic DNA.

genes, 7-fold for plastid-encoded genes, and about 75-fold for nuclear-encoded genes (Table 2 and Supplementary Table S2).

Alignments of EST sequence data from the mtDNA, ptDNA, and nucDNA of either *M. viride* NIES-296 or SAG 50-1 to the corresponding organelle and nuclear genome sequences (Supplementary Table S3) revealed no signs of post-transcriptional editing, further supporting the idea that the occurrence of RNA editing within the Viridiplantae is limited to land plants and that our nucleotide substitution rate analyses employ-

ing EST sequences can be used to estimate the substitutions at the DNA level.

For *Chlamydomonas*, 7 mitochondrial, 9 plastid, and 67 nuclear genes were used to estimate dS in *C. reinhardtii* vs. *C. incerta* (Supplementary Table S2). This dataset includes the same genes that were used previously by Popescu and colleagues (2006, 2007) to estimate evolutionary rates in the mitochondrial and nuclear compartments of *Chlamydomonas*, save for the mtDNA-located reverse transcriptase-like coding sequence *rtl*. All of the ptDNA sequence analyses were generated in this study. The average dS for the mitochondrial, plastid, and nuclear genes were 0.29, 0.30, and 0.37, respectively — these values are not statistically different between pairs ($p > 0.05$). Similar results were obtained when dS was calculated from a concatenation of the same coding regions (Table 2). The dS values varied by about 2-fold among genes in the mitochondrial compartment, by 5-fold within the plastid compartment, and by about 70-fold within the nuclear compartment.

Intronic and Intergenic Regions

We made an effort to use both intronic and intergenic sites to estimate the silent-site substitution rate of the three genetic compartments of *Mesostigma* and *Chlamydomonas*. However, due to either a lack of available sequence data and/or the absence of intronic DNA, it was not always possible to employ both types of silent-site in our analyses (Table 2). For both *Mesostigma* and *Chlamydomonas*, the substitution rate of the mtDNA intergenic regions were on average two-times larger than the corresponding mtDNA synonymous substitution rate (Table 2). For *Chlamydomonas*, the average substitution rate for the ptDNA intergenic regions was greater than the average rate of synonymous ptDNA substitutions, but for *Mesostigma* the opposite was true (Table 2, Supplementary Table S2). The average synonymous substitution rates of the intronic open-reading-frames (ORFs) in the mtDNA of *Mesostigma* and *Chlamydomonas* were about double that of the standard protein-coding genes from the same compartment. Interestingly, the substitution rates for the non-ORF intronic mtDNA sequences were substantially lower than dS of the ORFs. In both *Mesostigma* and *Chlamydomonas* nucDNAs, substitution rates in the intronic sequences were statistically higher than the average for the standard protein-coding genes from the same compartments, especially in the case of *Chlamydomonas.* Finally, the substitution rates calculated from concatenations of the organelle intronic and intergenic sequences, when available for *Mesostigma* and *Chlamydomonas*, gave values similar to the averages of the unconcatenated intronic and intergenic sequences.

Evolutionary Divergence among the *Mesostigma* Lineages

Consistent with the analysis of *GapA* sequence described above, each of five gene loci including *GapA* showed a greater dS between homologs in NIES-995 and SAG 50-1 and between NIES-995 and NIES-296 than do the homologs of SAG 50-1 and NIES-296 compared with each other (Table 3). Not surprisingly, based on phylogenetic analysis with *GapA* and *cox1* sequence, the NIES-995 lineage is the sister to the SAG 50-1/NIES 296 pair (Supplementary Fig. S1). NIES- 296, NIES-475, NIES-476, and NIES-477 showed the same *cox1* sequence (GenBank accession numbers: NIES-296, NC008240; NIES-475, HQ668012; NIES-476, HQ668013; NIES-477, HQ668014) in agreement with the *GapA* results presented earlier.

Discussion

Our silent-site substitution rate data between the *Mesostigma* SAG 50-1/NIES-296 pair and the *C. reinhardtii*/*C. incerta* pair are most comprehensive for synonymous positions in standard protein coding genes, so we chose to focus first on these data. For both *Mesostigma* and *Chlamydomonas*, the relative average dS values were similar (within a factor of two) among the three genetic compartments with the highest and lowest values being observed in the nuclear and plastid compartments, respectively. A drawback to using synonymous sites for estimating neutral substitution rates is that they are sometimes impacted by natural selection for codons that optimize translational efficiency — this is especially true for highly expressed genes in unicellular organisms (Hershberg and Petrov 2008; Sharp et al. 2010). Evidence consistent with translational selection shaping codon usage has been reported for the three genetic compartments of *Chlamydomonas* (Morton 1996; Naya et al. 2001; Popescu et al. 2006; Popescu and Lee 2007), the cyanelle and plastid genes of different algal and land plant lineages (Morton 1998; Morton et al. 2002), and mitochondrial genes in land plants (Liu et al. 2004) (but see Sloan and Taylor 2010, who argue that the inferred strength of selection appears too weak to account for the variation in substitution rates between the mitochondrial genomes of plants

Table 3. Synonymous substitution rates of five genes between *M. viride* NIES-296, SAG 50-1, and NIES-995.

Gene product	Number of codons compared	dS_{AB}[a]	dS_{AC}[a]	dS_{BC}[a]
Nitrate transporter[b]	191	0.69	1.70	1.42
Elongation factor EF-3[b]	154	0.61	0.80	0.80
60S ribosomal protein L12[b]	125	0.04	0.21	0.24
Cytochrome c oxidase subunit 1[b]	158	0.22	0.43	0.50
Glyceraldehyde-3-phosphate dehydrogenase subunit A[b]	111	0.14	0.36	0.45

[a]dS_{ij}, synonymous substitution rate between taxon i and j. Taxon A, NIES-296; taxon B, SAG 50-1; taxon C, NIES-995. [b]GenBank accession numbers of sequences from NIES-296 and SAG 50-1 are listed in Supplementary Table S2, and those from NIES-995 are HQ668016, HQ668017, HQ668018, HQ668016, and HQ668011.

and other multicellular eukaryotes). Because we employed an eclectic mix of genes, some of which are highly expressed, from the three genetic compartments of *Chlamydomonas* and *Mesostigma*, translational selection is probably affecting our dS measurements from each compartment. However, the fact that we are comparing the relative rates between the three compartments may to some extent alleviate these affects. We note that when these analyses were restricted to the same set of genes in both *Chlamydomonas* and *Mesostigma* there was no significant change in the mitochondrial and plastid substitution rates; however, for the nuclear genes the average dS was reduced in half, likely due to an enrichment of loci that are highly expressed; for a nuclear gene locus to be represented in the restricted dataset required in most cases that it be represented among the cDNA sequences in EST libraries from all four taxa employed.

Although incomplete, the substitution-rate statistics for introns and intergenic spacers in the various genomes of *Mesostigma* and *Chlamydomonas* are in many instances 2- to 4-fold higher than the average dS from the same compartment. This suggests that intronic and intergenic sites in these cases may be better than synonymous sites for estimating the neutral substitution rate, supporting previous conclusions for the nuclear compartment of *Chlamydomonas* (Popescu et al. 2006). However, due to the incompleteness of the data from intronic and intergenic regions in the different genomes of *Chlamydomonas* and *Mesostigma*, we are reluctant to modify our estimates of the relative neutral rates beyond what is suggested from the average dS values.

Based on the above discussions, we argue that for both *Mesostigma* and *Chlamydomonas* the neutral substitution rates of the mitochondrial, plastid, and nuclear compartments are similar to one another (within a factor of two), with the nuclear rate being the highest and that of the plastid being the lowest. Given that *Mesostigma* and *Chlamydomonas* occupy deeply divergent positions within the Viridiplantae, we suggest that these relative rates represent the ancestral substitution-rate patterns for green plants. These predicted rates are in contrast to land plants where the relative synonymous substitution rates of the mitochondrial, plastid, and nuclear genomes are estimated to be 1:3:10 (Drouin et al. 2008; Wolfe et al. 1987, 1989).

For unicellular eukaryotes, data on relative silent-site substitution rates for the organelle and nuclear compartments are limited in the number of taxa examined and in many cases the number loci sampled (Lynch et al. 2006; Sorhannus and Fox 1999) and the data that are available sometimes disagree, especially when comparing estimates gleaned from comparative genomics versus those from the direct screening of accumulated mutations (Lynch et al. 2008). Nonetheless, in a very general sense the existing evolutionary rate data suggest that the organelle and nuclear genomes of most unicellular eukaryotes have similar silent-site substitution rates, thus, supporting our findings for *Mesostigma* and *Chlamydomonas*.

It has been argued that the primary forces governing the evolution of genome architecture are mutation and random genetic drift (Lynch and Conery 2003; Lynch et al. 2006), and that any increase in noncoding DNA exposes a genome to potentially deleterious mutations, where the higher the mutation rate the greater the burden of the embellishment. Support for this view of genome evolution has come from the observations that the large and elaborate mitochondrial genomes of land plants have, with few exceptions, a low mutation rate [mean = 0.36×10^{-9} base substitu-

tions per site per year (Lynch 2007)], much lower than that predicted for mammals (mean = 34×10^{-9} base substitutions per site per year (Lynch 2007)], whose mitochondrial genomes are paragons of compactness. Moreover, relative rates of synonymous substitution between the mitochondrial and nuclear genomes of land plants average 1:10 (Drouin et al. 2008; Wolfe et al. 1987, 1989), whereas for animals they average 24:1 (Lynch et al. 2006), supporting the idea that the opposing organelle DNA architectures between these two groups [both of which have relatively small effective population sizes and similar nuclear genome structures (Lynch and Conery 2003)] is a consequence of differing mitochondrial mutation rates. Consistent with this interpretation is the finding that land plant plastid genomes, which are smaller and more compact than their mitochondrial counterparts (but still more bloated than animal mtDNA), have a synonymous substitution rate that is on average 3-times that of land plant mtDNA (Drouin et al. 2008; Wolfe et al. 1987, 1989).

If our relative-rate data for *Mesostigma* and *Chlamydomonas* are representative of green algae in general, they may help us understand the evolution of organelle genome architecture within green plants. Broadly speaking, there is less disparity in genomic complexity between the mitochondrial and plastid genomes of green algae than there is for land plant mitochondrial and plastid DNAs. Indeed, the average fraction of noncoding DNA in the available mtDNA and ptDNA sequences from chlorophytes is 0.35 and 0.41, respectively, and for streptophyte green algae the corresponding values are 0.47 (mtDNA) and 0.41 (ptDNA). Conversely, the proportion of noncoding nucleotides in the mtDNA of land plants averages 0.84 but is 0.42 for the ptDNA (Table 1). We propose that the comparable architectures of green algal plastid and mitochondrial DNAs are a consequence of these genomes having similar mutation rates, as reflected in their similar rates of silent-site substitution. The fact that we found the silent-site substitution rate of ptDNA for both *Mesostigma* and *Chlamydomonas* to be slightly lower than that of the mtDNA may help explain why green algal plastid genomes tend to be a little bit larger and more bloated than their mitochondrial counterparts.

When combining our relative mutation rate estimates for *Chlamydomonas* with published nucleotide diversity data from *C. reinhardtii* (Smith and Lee 2008, 2009), we can gain insights into the effective number of genes per locus — a parameter that is inversely related to the power of random genetic drift — for the three genetic compartments of this alga. According to population genetics theory, nucleotide diversity at neutral sites is equivalent to $2N_g\mu$ where N_g is the effective number of genes per locus in the population and μ is the mutation rate per nucleotide site per generation (Lynch 2007). Assuming similar mutation rates in the three genetic compartments of *Chlamydomonas*, as our data suggest, one would expect, after correcting for uniparental inheritance of the organelle genes, that the silent-site nucleotide diversity of both the mitochondrial and plastid compartments would be approximately ½ that of the nucleus. The observed silent-site diversity values for the mitochondrial and plastid genomes at ¼ and ½ that of the nucleus, respectively, are close to these expectations. The slightly lower than expected nucleotide diversity for the mitochondrial compartment may reflect an overestimate of the relative mutation rate in this compartment or a greater incidence of selection on linked variation as discussed by Smith and Lee (2008). Interestingly, in both outcrossing and highly inbreeding species of *Arabidopsis*, the expected effective population sizes of both organelle and nuclear genes are consistent with neutral expectations or similar interference effects across the genomes (Wright et al. 2008).

Mitochondrial and plastid RNA editing, as discussed earlier, occurs in land plants but has not been observed in green algae. In the present study, therefore, it is not surprising that we found no evidence of RNA editing in either the organelle or nuclear genomes of *Chlamydomonas* and *Mesostigma*. Absolute mutation rates will have to be determined for these lineages and other related lineages, before one can argue for or against the possibility that the absence of organelle RNA editing in green algae is due to elevated organelle DNA mutation rates as predicted by the mutational hazard hypothesis (Lynch et al. 2006; Sloan and Taylor 2010).

Various hypotheses have tried to explain the wide variation in protein-coding gene content among mitochondrial genomes (Adams and Palmer 2003; Berg and Kurland 2000; Brandvain and Wade 2009). It has been argued that higher rates of mutation in the mtDNA relative to that of the nucDNA favours mitochondrial-to-nucleus gene migrations. Consistent with this view is the observation that the gene-rich mtDNAs of land plants tend to have low mutation rates, whereas the gene-poor mtDNAs of animals generally have high mutation rates. The mitochondrial genomes of *Mesostigma* and *Chlamydomonas* are good models for addressing this issue because they have opposing

gene contents, harbouring 36 vs. 8 protein-coding genes, respectively. Given that we predict approximately equal mitochondrial and nuclear mutation rates in both *Mesostigma* and *Chlamydomonas*, our data do not support the above hypothesis and thus do not explain the ~5-fold difference in mitochondrial gene content between these lineages.

The two pairs of lineages examined here appear to represent deeply isolated lines within *Mesostigma* and *Chlamydomonas*. In both instances the average level of silent site divergence between homologous loci in the three genetic compartments is typically an order of magnitude or more greater than the average frequency of polymorphisms at silent sites (nucleotide diversity) in geographic isolates of *Chlamydomonas reinhardtii* (Smith and Lee 2008, 2009) and the multicellular chlorophyte *Volvox carteri* (Smith and Lee 2010), and at least several fold more than most other unicellular/oligocelluar non-green algal alleged species (Lynch 2006). Moreover, the four isolates of *M. viride* from the NIES-296 lineage (NIES-296, NIES-475, NIES-476 and NIES-477) have no nucleotide diversity for the combined 800 nt of *cox1* and *GapA* sequence examined, but these strains were isolated from the same region of Japan (Supplementary Table S1). The reproductive compatibility of *C. reinhardtii* geographical isolates (see Smith and Lee 2008 and references therein) and the reproductive incompatibility of *C. reinhardtii* with the one available strain of *C. incerta* (see Popescu et al. 2006 and references therein), supports the separate biological species status of *C. reinhardtii* and *C. incerta*. However, in the case of *Mesostigma viridie* strains NIES-296 and SAG 50-1, there is no parallel knowledge on this matter. In fact, information on sexual reproduction in *M. viride* is extremely limited. Strains NIES-475, NIES-476, and NIES-477 were characterized as having a heterothallic mode of sexual reproduction with mating and zygote formation between strains NIES-475 and NIES-476 and between NIES-475 and NIES-477 (see NIES website: http://mcc.nies.go.jp/genusList.do?genus=Mesostigma), but after a few years in culture these strains have lost their fertility (Professor Shoichiro Suda, personal communication) and, therefore, cannot be tested against other strains currently in culture such as NIES-296 (from Japan) and SAG 50-1 (from England), which have probably also lost fertility. DNA sequences should be obtained from new geographical isolates of *M. viride* and these data should be compared to the sequences reported here, and strain combinations should be identified that mate and form viable zygotes.

Nevertheless, we should consider the fact that recently diverged lineages can differ for alleles that existed prior to the separation of these lineages and that divergence between these ancient alleles can lead to overestimates of the substitution rate (Charlesworth 2010; Peterson and Masel 2009). The existence of such early diverging alleles could bias our estimates of relative rates of substitution across the genetic compartments of both *Chlamydomonas* and *Mesostigma* if one or more of these compartments have a different proportion of ancient polymorphisms. In an extreme case, for example, consider two lineages that each contain one of a pair of distinct mitochondrial or plastid genomes that existed prior to the lineage divergence and these underwent no recombination because of uniparental inheritance. In these instances the relative substitution rates in the affected genome would be over estimated. Although we cannot dismiss such possibilities, the common trend for similar average relative silent substitution rates across the three compartments in both *Chlamydomons* and *Mesostigma* seems to make this possibility unlikely.

Another related potential source of error in our studies of substitution rates in both *Mesostigma* and *Chlamydomonas* is that some of our nuclear gene divergences are between paralogous genes. The wide range of dS values among nuclear genes in the *C. reinhardtii*/*C. incerta* (Popescu et al. 2006) and *M. viride* NIES-296 / SAG 50-1 comparisons might suggest such a possibility for some of the more diverged examples. Paralogous pairs if common could artificially increase the apparent average substitution rate in the nuclear compartment relative to the organelle ones. However, wide ranges in silent substitution rates in nuclear genes have been reported in other systems (see discussion in Popescu et al. 2006) and it seems unlikely that paralogy is an important issue in the *Chlamydomonas* sequences compared in our data set because none of these gene sequences when blasted against the complete nuclear genome of *C. reinhardtii* (NZ_ABCN00000000) identified multigene families. For *Mesostigma* there is currently no nuclear genome sequence against which we can blast the *Mesostigma* gene sequences employed; however, based on blast analysis against all GenBank entries we excluded from our dataset any genes that appeared to be part of a multigene family. For this and other potential sources of error discussed above, relative rates of mutation in the three genetic compartments of *Chlamydomonas*

and *Mesostigma* should be determined directly by sequencing DNA from mutation accumulation lines.

Finally, although not a goal of this study our results open up the possibility that *M. viride* is a cryptic species complex containing at least three distinct lineages represented by NIES-995, SAG 50-1, and NIES-296. *M. viride* strain NIES-995 is an attractive candidtate for obtaining plastid DNA sequence and additional nuclear and mitochondrial DNA sequence in order to see if the trends across genetic compartments observed with the SAG 50-1/NIES-296 pair are maintined in comparisions between these strains and NIES-995.

Methods

Strains and growth conditions: *M. viride* and *C. incerta* were obtained from the Microbial Culture Collection at the National Institute for Environmental Studies (NIES) and from the Sammlung von Algenkulturen, Göttingen (SAG). See Supplementary Table S1 for a complete list of strains employed in this study. *M. viride* was grown in 75 ml of medium CT (http://mcc.nies.go.jp/medium/en/ct.pdf) in 250 ml Erlenmeyer flasks at 22 °C on a 12 h light (20 μmoles/m^{-2}/s): 12 h dark cycle. *C. incerta* was grown in Tris-acetate-phosphate (TAP) medium as described previously (Popescu et al. 2006).

DNA amplification, sequencing, and accession numbers: DNA was extracted using a CTAB-based method (Reineke et al. 1998) and amplified with Platinum PCR SuperMix (Invitrogen, San Diego) using primers designed from publically available *M. viride* and *C. reinhardtii* sequence data. PCR products were sequenced on both strands at the Macrogen Sequencing Facility in Rockville, MD, and the resulting sequences were deposited in GenBank under the accession numbers HQ667981-HQ668018.

Sequence data used to measure substitution rate: For the pairs *M. viride* SAG 50-1/NIES 296 and *C. reinhardtii*/*C. incerta*, the mitochondrial, plastid and nuclear DNA sequences compared were obtained from i) genomic sequences in GenBank, or ii) EST data in Genbank and/or PCR products obtained in this study; see Supplementary Table S2 for details on the sources of sequences utilized.

Data analysis: Sequences were aligned with MUSCLE (Edgar 2004). The number of substitutions per synonymous site (dS) was calculated using the maximum likelihood (ML) method (Goldman and Yang 1994) and F3X4 model in the CODEML program of PAML 4.3 (Yang 2007). Only protein-coding regions and intronic open reading frames (ORFs) that had at least 100 codons were used. The number of substitutions per noncoding site was estimated using the HKY85 model of the BASEML program of PAML. Noncoding regions less than 100 nt were ignored. ML phylogenies were performed with the PHYML web server (Guindon et al. 2005) under the GTR model. Minimum evolution (ME) and maximum parsimony (MP) phylogenies were performed by MEGA 4.0 (Tamura et al. 2007) and PAUP* 4.0 (Swofford 2003), respectively. For all of the phylogenetic analyses, node support values were assessed using 100 bootstrap replicates. *P* values of between group differences were calculated using the t-test as implemented in Excel 2003.

Acknowledgements

This work was supported by a grant to R.W.L. from the Natural Sciences and Engineering Research Council of Canada and was part of the Protist EST Program (PEP) funded by Genome Canada and the Atlantic Canada Opportunities Agency (Atlantic Innovation Fund). We thank the anonymous reviewers for their helpful comments.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.protis.2011.04.003.

References

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol **29**:380–395

Becker B, Marin B (2009) Streptophyte algae and the origin of embryophytes. Ann Bot **103**:999–1004

Berg OG, Kurland CG (2000) Why mitochondrial genes are most often found in nuclei. Mol Biol Evol **17**:951–961

Brandvain Y, Wade MJ (2009) The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. Genetics **182**:1129–1139

Brown WM, George Jr M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA **76**:1967–1971

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol **18**:225–239

Charlesworth D (2010) Don't forget the ancestral polymorphisms. Heredity **105**:509–510

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

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**:1792–1797

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol **11**:725–736

Graur D, Li W (2000) Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, Mass

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online–a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **33**:W557–559

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

Janke A, Pääbo S (1993) Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. Nucleic Acids Res **21**:1523–1525

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK

Lemieux C, Otis C, Turmel M (2007) A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol **5**:2

Lenz H, Rudinger M, Volkmar U, Fischer S, Herres S, Grewe F, Knoop V (2010) Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. Curr Genet **56**:189–201

Liu Q, Feng Y, Xue Q (2004) Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. Mitochondrion **4**:313–320

Lynch M (2006) The origins of eukaryotic gene structure. Mol Biol Evol **23**:450–468

Lynch M (2007) The Origins of Genome Architecture. Sinauer Assocs., Inc, Sunderland, MA

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

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, Thomas WK (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA **105**:9272–9277

Marin B, Melkonian M (1999) Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequences comparisons. Protist **150**:399–417

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WC, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318**:245–250

Miyata T, Hayashida H, Kikuno R, Hasegawa M, Kobayashi M, Koike K (1982) Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. J Mol Evol **19**:28–35

Morton BR (1996) Selection on the codon bias of *Chlamydomonas reinhardtii* chloroplast genes and the plant *psbA* gene. J Mol Evol **43**:28–31

Morton BR (1998) Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. J Mol Evol **46**:449–459

Morton BR, Sorhannus U, Fox M (2002) Codon adaptation and synonymous substitution rate in diatom plastid genes. Mol Phylogenet Evol **24**:1–9

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol **7**: 135

Naya H, Romero H, Carels N, Zavala A, Musto H (2001) Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. FEBS Lett **501**:127–130

Nedelcu AM, Borza T, Lee RW (2006) A land plant-specific multigene family in the unicellular *Mesostigma* argues for its close relationship to Streptophyta. Mol Biol Evol **10**:819–831

Peterson GI, Masel J (2009) Quantitative prediction of molecular clock and K_a/K_s at short timescales. Mol Biol Evol **26**:2595–2603

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

Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, Hellsten U, Chapman J, Simakov O, Rensing SA, Terry A, Pangilinan J, Kapitonov V, Jurka J, Salamov A, Shapiro H, Schmutz J, Grimwood J, Lindquist E, Lucas S, Grigoriev IV, Schmitt R, Kirk D, Rokhsar DS (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science **329**:223–226

Reineke A, Karlovsky P, Zebitz CP (1998) Preparation and purification of DNA from insects for AFLP analysis. Insect Mol Biol **7**:95–99

Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. Philos Trans R Soc Lond B Biol Sci **365**:1203–1212

Simon A, Glöckner G, Felder M, Melkonian M, Becker B (2006) EST analysis of the scaly green flagellate *Mesostigma viride* (Streptophyta): implication for the evolution of green plants (Viridiplanta). BMC Plant Biol **6**:2

Sloan DB, Taylor DR (2010) Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. J Mol Evol **70**:479–491

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 (2009) 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

Sorhannus U, Fox M (1999) Synonymous and nonsynonymous substitution rates in diatoms: a comparison between chloroplast and nuclear genes. J Mol Evol **48**:209–212

Swofford DL (2003) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Assocates, Sunderland, MA

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol **24**:1596–1599

Turmel M, Otis C, Lemieux (2002) The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. Mol Biol Evol **19**:24–38

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA **84**:9054–9058

Wolfe KH, Sharp PM, Li WH (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol **29**:208–211

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, Von Dassow P, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist EA, Lucas S, Mayer KF, Moreau H, Not F, Otillar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov A, Robbens S, Schmutz J, Toulza E, Wyss T, Zelensky A, Zhou K, Armbrust EV, Bhattacharya D, Goodenough UW, Van de Peer Y, Grigoriev IV (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. Science **324**:268–272

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

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

Yokobori S, Pääbo S (1995) Transfer RNA editing in land snail mitochondria. Proc Natl Acad Sci USA **92**:10432–10435