Mitochondrial Genome of the Colorless Green Alga *Polytomella capuana*: A Linear Molecule with an Unprecedented GC Content

David R. Smith and Robert W. Lee

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

One common observation concerning mitochondrial genomes is that they have a low guanine and cytosine content (GC content); of the complete mitochondrial genome sequences currently available at the National Center for Biotechnology Information (NCBI) (July 2007), the GC content ranges from 13.3% to 53.2% and has an average value of 38%. Here, we present the GC-rich mitochondrial genome (57% GC) of the colorless green alga *Polytomella capuana*. The disproportion of GC among the different regions of the *P. capuana* mitochondrial DNA (mtDNA) suggests that a neutral process is responsible for the GC bias. We propose that a biased gene conversion mechanism resulted in the GC-rich state of the *P. capuana* mtDNA. In addition, our analysis indicates that the *P. capuana* mitochondrial genome is a single 13-kb linear molecule with telomeres, which have a closed (hairpin-loop) conformation: a novel terminal structure among described linear green-algal mtDNAs. Furthermore, using a series of GC-rich inverted repeats found within the *P. capuana* mitochondrial genome, we describe recombination-based scenarios of how intact linear mtDNA conformations can be converted into the fragmented forms found in other *Polytomella* taxa.

Introduction

One of the most distinguishing characteristics of mitochondrial genomes as compared with nuclear genomes is their low GC content; of the 1,125 complete mitochondrial DNA (mtDNA) sequences available at NCBI as of July 2007, the GC content ranges from 13.3% to 53.2% and has an average value of 38%. Although sampling is highly biased toward animal mitochondrial genomes (1,015 out of the 1,125), the trend of having a low GC content is seen throughout other major groups, including fungi (17.1–43.2% GC), the Archaeplastida (Adl et al. 2005) (22.2–45.2% GC), and the heterogeneous group of eukaryotic unicells negatively defined as not belonging to animals, fungi, or the Archaeplastida (14.0–41.2% GC). Various hypotheses for explaining why mtDNA is GC poor have been proposed. For example, it has been suggested that the low levels of G and C are due to a mutational bias caused by the loss of DNA repair genes in the endosymbiotic genome that gave rise to the mitochondrial genome (Glass et al. 2000; Moran 2002; Burger and Lang 2003). Others contend that a low GC content correlates with adaptation to an intracellular lifestyle where high levels of ATP and UTP relative to GTP and CTP make replication and transcription of an AT-rich genome more efficient (Howe et al. 2002; Rocha and Danchin 2002). Convergent evolution toward a low GC content is also seen in the genomes of chloroplasts, symbionts, and endocellular parasites (Dybvig and Voelker 1996; Ogata et al. 2001; Kusumi and Tachida 2006).

The mitochondrial genome of the colorless green algae *Polytomella capuana* attracted our interest because of its potential for having a particularly high GC content relative to other available mitochondrial genome sequences, not only for green algae but also for eukaryotes in general. This view was motivated by the high GC content of a 768-nt segment of *cox1* from the mtDNA of *P. capuana* (GenBank accession number DQ221113; Mallet and Lee 2006), especially at sites expected to be under low selective constraint, such as 4-fold degenerate sites, which are third-position codon sites that can tolerate any of the 4 nt without altering the amino acid specified.

Polytomella is a group of wall-less and colorless unicells (Pringsheim 1955) belonging to the "Reinhardtii clade" (Gerloff-Elias et al. 2005) of chlorophycean green algae. Available *Polytomella* taxa fall into 3 lineages (Mallet and Lee 2006). Two of these lineages, represented by *Polytomella parva* and *Polytomella* strain 63-10 (Sammlung von Algenkulturen Göttingen [SAG]), possess fragmented mitochondrial genomes (Fan and Lee 2002; Mallet and Lee 2006). Whereas the third and earliest branching lineage, represented by *P. capuana*, appears to have an intact linear mitochondrial genome (Mallet and Lee 2006). Substantial sequence data exist only for the mitochondrial genome of *P. parva*, which is comprised of 2 linear fragments with estimated sizes of 13.5 and 3.5 kb (Fan and Lee 2002). Sequence information spanning 97% and 86%, of the 13.5- and 3.5-kb fragments, respectively, is available. The 3.5-kb fragment encodes only 1 gene (*nad6*), which is missing from the 13.5-kb fragment. The telomeres of both fragments contain virtually identical inverted repeats that are at least 1.3 kb in length; however, the extreme termini of both fragments still remain to be sequenced.

Chlamydomonas reinhardtii, a close relative to *Polytomella* taxa (Nakayama et al. 1996; Pröschold et al. 2001; Gerloff-Elias et al. 2005), has a completely sequenced linear mtDNA molecule of 15.8 kb with terminal inverted repeats of 531 or 532 nt including a 39- to 41-nt noncomplementary 3′ extension (Gray and Boer 1988; Michaelis et al. 1990; Vahrenholz et al. 1993). Based on the potential interaction of the telomeres with internal repeats, 2 models of replication have been proposed for the *C. reinhardtii* mtDNA (Vahrenholz et al. 1993). There are no apparent similarities between the telomeric sequences of *P. parva* and *C. reinhardtii* (Fan and Lee 2002), and a model describing how the *P. parva* mtDNA may replicate has not yet been proposed.

Taken as a whole, sequence data from the *P. capuana* mtDNA will provide useful information on nucleotide composition biases in mitochondrial genomes as well as knowledge about the evolution of mitochondrial telomeres and the mechanisms through which intact linear mtDNA

Key words: *Polytomella capuana*, green algae, mitochondrial DNA (mtDNA), GC content, biased gene conversion, telomeres.

E-mail: robert.lee@dal.ca.

Mol. Biol. Evol. 25(3):487–496. 2008
doi:10.1093/molbev/msm245
Advance Access publication November 12, 2007

conformations were converted into fragmented forms. With these motives in mind, we sequenced to completion the *P. capuana* mitochondrial genome.

Materials and Methods

Strain, Culture Conditions, Mitochondrial Enrichment, and DNA Extraction

We used a stock of *P. capuana* (SAG strain 63-5) made axenic by Mallet and Lee (2006). Cells were cultured at 22 °C in *Polytomella* medium (0.1% tryptone, 0.2% yeast extract, and 0.2% sodium acetate) and harvested in the late logarithmic phase of growth ($OD_{750\ nm} = 0.4$; determined with a Bausch and Lomb Spectronic 20 spectrophotometer) by centrifugation (1,000 × g) at 4 °C. Cells were disrupted with a Dounce homogenizer. The mitochondrial-enriched fraction was prepared and treated with DNase I following procedure B of Ryan et al. (1978). Isolation of DNA followed the method of Ryan et al. (1978), with the exception that there was no further DNA purification step employing preparative CsCl gradient centrifugation.

DNA Amplification

Polymerase chain reaction (PCR) experiments were performed in High Fidelity Platinum SuperMix (Invitrogen, Carlsbad, CA) using DNA from a mitochondrial-enriched fraction as the template. DNA was initially denatured at 94 °C for 3 min, then amplified by 35 cycles of denaturation at 94 °C for 45 s, annealing at 50–60 °C (depending on the melting temperature of the primers), and extension at 72 °C; there was a final extension at 72 °C for 10 min. The telomeric regions of the *P. capuana* mtDNA were amplified using: 1) the long-walk PCR method of Katz et al. (2000), 2) terminal deoxynucleotidyl transferase (TdT) tailing as described by Förstemann et al. (2000) and Bah et al. (2004), and 3) standard PCR amplification (as described above).

DNA Blotting and Hybridization

Blotting of agarose gels onto Hybond-N+ membranes (Amersham, Buckinghamshire, UK) was performed using the Vacublot XL system (Amersham). Probes used in this study were labeled and hybridized to samples with the AlkPhos Direct Labelling and Detection System (Amersham) following the manufacturer's instructions. Label was detected by exposing the membranes to Fuji Super RX medical X-ray film.

Cloning and Sequencing of DNA Fragments

PCR-, long-walk PCR-, and TdT-tailing products were separated by agarose gel electrophoresis, purified with the QIAquick Gel Extraction Kit (Qiagen, Germantown, MD), and then cloned using the TOPO TA Cloning Kit (Invitrogen). Plasmid DNA was extracted with the QIAquick Spin Miniprep Kit (Qiagen). PCR products and the clones derived from PCR products were sequenced on both strands at the Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada.

Sequence Analysis

Sequences were edited and assembled using CodonCode Aligner (Version 1.5.2). The Blast network services (Altschul et al. 1990) were employed for sequence similarity searches. Protein-coding regions not initially detected in the Blast search were identified with the Open Reading Frame Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The boundaries of the mitochondrial rRNA-coding modules were estimated by sequence comparisons with their counterparts in *P. parva* and *C. reinhardtii* mtDNA. Genes encoding tRNAs were located using tRNAScan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/; Lowe and Eddy 1997). Repeated regions were identified using Repeat Finder (http://www.proweb.org/proweb/Tools/selfblast.html) and REPuter (Kurtz et al. 2001). Secondary structures and folding energies of the inverted repeat sequences were predicted with Mfold (Zucker 2003). The equation used for calculating the GC-skew value was $(G - C)/(G + C)$ and that for the AT skew was $(A - T)/(A + T)$. The cumulative GC-skew plot was formulated using the GenSkew software program (http://mips.gsf.de/services/analysis/genskew).

Nucleotide Sequence Accession Number

The complete sequence of the *P. capuana* mtDNA is deposited in GenBank under the accession number EF645804.

Results

General Features

The *P. capuana* mitochondrial genome is a single linear molecule of 12,998 nt with terminal inverted repeats (i.e., the sequence of one terminus is present in an inverted orientation relative to the other terminus) (fig. 1). The size and conformation of the *P. capuana* mtDNA were confirmed by pulse-field gel electrophoresis and restriction endonuclease digestion patterns (data not shown). The coding regions in this mtDNA are arranged into 2 unequally sized clusters with opposite transcriptional polarities, which proceed outwards toward the ends of the genome. The region separating the 2 transcriptional orientations shows no similarity to the potential promoter sequence identified in the corresponding region of the *P. parva* (Fan and Lee 2002) and *C. reinhardtii* mtDNAs (Duby et al. 2001). The complement and arrangement of genes in the *P. capuana* mtDNA parallels that of *P. parva*, except for *nad6*, which is located internal to the left terminal repeat in *P. capuana* (fig. 1) but on a separate chromosome in *P. parva* (fig. 1). The sole tRNA gene, *trnM*(*cau*), has characteristics that are consistent with a role in elongation rather than initiation, like the mitochondrial-encoded *trnM*(*cau*) of other Reinhardtii-clade taxa (Boer and Gray 1988; Denovan-Wright et al. 1998).

Of the 12,998 nt comprising the *P. capuana* mtDNA, 10,662 nt (82%) code for proteins and functional RNAs, and 2,336 nt (18%) represent noncoding DNA. The latter can be subdivided into intergenic regions and terminal repeats, which constitute 555 nt and 1,780 nt (890 nt at each

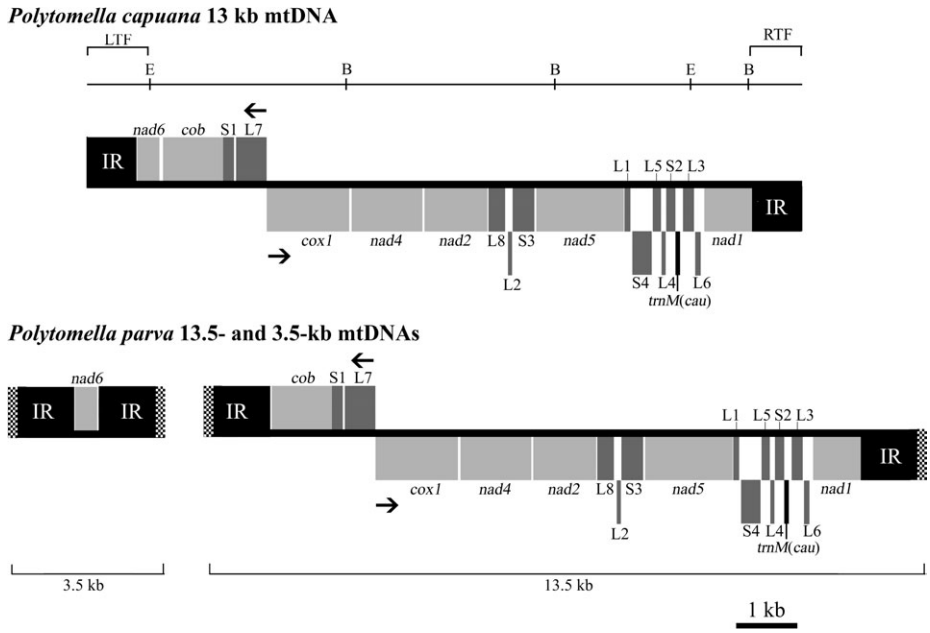

FIG. 1.—Genetic maps of the *Polytomella capuana* and the *Polytomella parva* mtDNAs. Protein-coding, rRNA-coding, and terminal inverted repeat (IR) regions are shown in light gray, dark gray, and black, respectively. S1–S4 and L1–L8 denote small-subunit and large-subunit rRNA–coding modules, respectively; *trnM*(*cau*) signifies the gene for $tRNA^{met}$. Thick solid arrows indicate transcriptional polarities. Restriction sites for *Eco*RI (E) and *Bam*H1 (B) are shown on the *P. capuana* mtDNA map, and the fragments identified by brackets above this restriction map (left-terminal restriction fragment [LTF] and right-terminal restriction fragment [RTF]) represent the location of molecular probes used in this work. The *P. parva* mtDNA maps are reproduced with modifications from Fan and Lee (2002); the checkered regions on these maps are at present unsequenced.

terminus), respectively. The 19 intergenic regions identified in the *P. capuana* mtDNA range in size from 3 to 62 nt and have an average length of 29 nt.

The *P. capuana* has a genome-wide GC bias; the overall GC content is 57.7%. The allocation of G versus C (GC skew) on the main sense strand (the strand encoding the gene for *cox1*) is negligible with a value of only 0.006. The distribution of A versus T is slightly more skewed at −0.09, reflecting a slight tendency toward T on the main sense strand.

GC Bias by Region

The GC content of the *P. capuana* mtDNA differs considerably among the various regions of the genome. Table 1 shows that the GC values in *P. capuana* mtDNA exceed those of *P. parva* and *C. reinhardtii* for all defined regions.

Base composition of the coding mtDNA in *P. capuana* has an average GC content of 56.4% (table 1). This value reflects both rRNA- and protein-coding regions evenly. In opposition, the *trnM* gene is slightly GC poor (47.9%). The inflated GC content of the protein-coding genes in the *P. capuana* mtDNA comes from a large number of codons ending in G or C (76%) (table 1); when considering only 4-fold degenerate sites, this value is even higher (85%). In both of these cases, the occurrence of G versus C is approximately equal. At the more functionally constrained first and second codon positions, the base compositions are less GC rich, with values of 52% and 41%, respectively (table 1). The protein-coding genes in the mitochondrial genomes of *P. parva* and *C. reinhardtii* show a tendency toward A

Table 1
Base Composition by Region of the *Polytomella capuana* mtDNA

	Overall			Coding[a]			Intergenic			Telomeres			Codon Site Position								
													Site 1			Site 2			Site 3		
	Pc	*Pp*	*Cr*	*Pc*	*Pp*	*Cr*	*Pc*	*Pp*	*Cr*	*Pc*	*Pp*[b]	*Cr*	*Pc*	*Pp*	*Cr*	*Pc*	*Pp*	*Cr*	*Pc*	*Pp*	*Cr*
%A	19.5	27.6	27.2	18.2	25.9	22.7	15.0	38.4	30.0	25.6	24.9	26.7	22.9	26.2	24.7	17.5	18.9	16.4	8.1	27.5	15.5
%T	23.3	27.2	31.8	25.4	34.4	32.2	17.1	32.7	30.5	15.7	31.2	18.8	24.8	29.5	27.2	41.2	41.8	42.8	15.9	37.4	38.4
%C	28.4	22.4	19.9	28.4	19.4	21.9	35.0	15.0	20.7	27.8	22.9	31.2	20.3	16.9	17.5	22.8	20.9	21.5	42.8	20.9	26.3
%G	28.8	22.8	21.1	28.0	20.3	23.2	33.0	13.9	18.8	30.9	21.1	23.3	31.9	27.3	30.5	18.5	18.4	19.3	33.2	14.2	19.8
%GC	57.2	45.2	41.0	56.4	39.7	45.1	68.0	28.9	39.5	58.7	44.0	54.5	52.2	44.2	48.0	41.3	39.3	40.8	76.0	35.1	46.1

NOTE.—Pc, *Polytomella capuana*; Pp, *Polytomella parva*; Cr, *Chlamydomonas reinhardtii*.

[a] Based on protein-, rRNA-, and tRNA-coding regions.

[b] Based on incomplete sequence data.

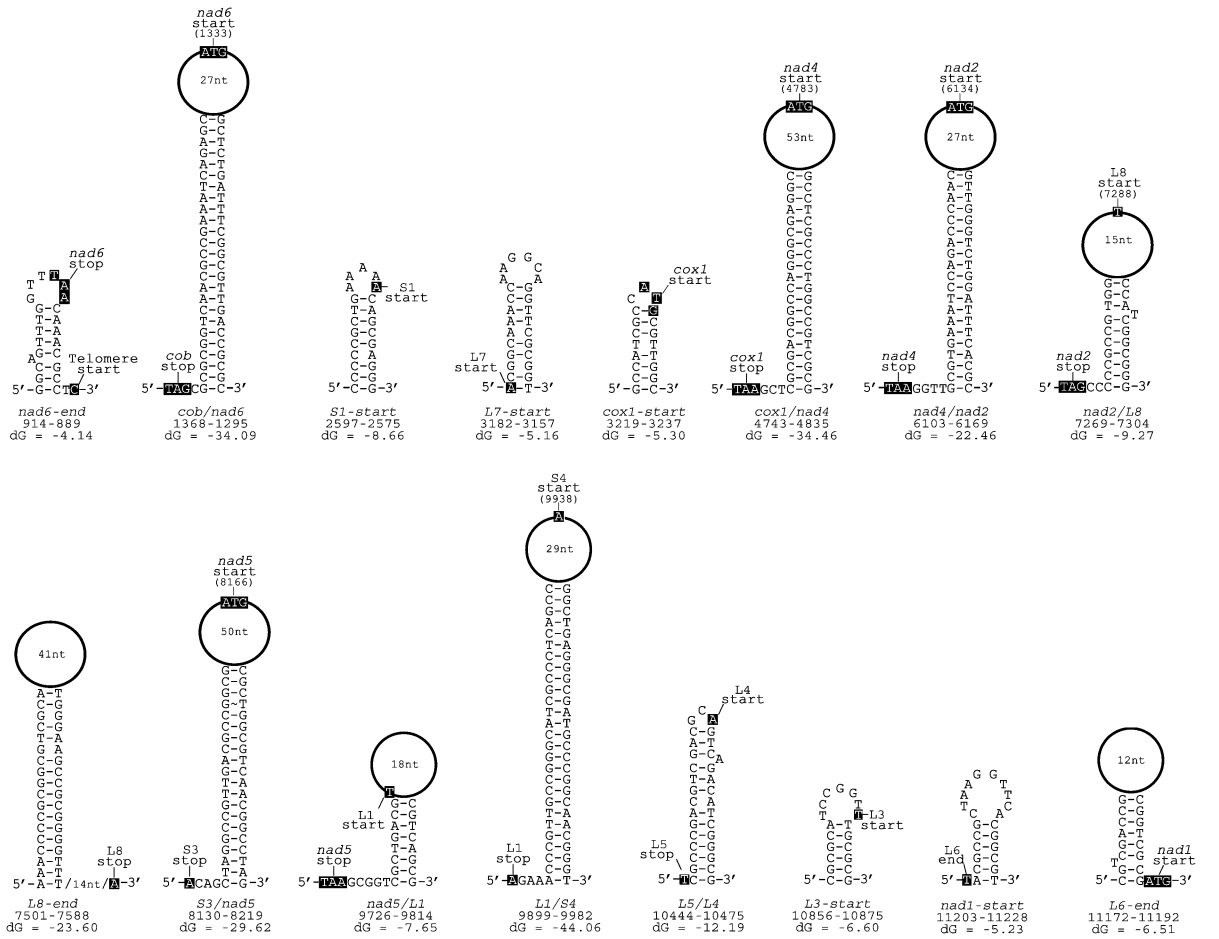

FIG. 2.—Putative stem-loop structures inferred from the *Polytomella capuana* mtDNA sequence. Name (based on genome location), genome coordinates (nt), and folding energy (dG) at 37 °C are given beneath each secondary structure, respectively. If the loop component of the secondary structure is larger than 10 nt, then it is depicted by a hollow black circle with its size shown in the center. The predicted start and stop sites of coding regions are shaded; when these sites occur within 1 of the depicted loop structures their position in the genome is shown.

and T at all 3 codon positions (table 1). The *P. capuana* protein-coding genes, in spite of being GC rich, show a derived amino acid composition similar to that of the *P. parva* and *C. reinhardtii* protein-coding genes (supplementary fig. 1, Supplementary Material online). Alanine is the only significant exception to this trend, encoded 308 times in the *P. capuana* mtDNA but only 257 and 215 times in that of *C. reinhardtii* and *P. parva*, respectively. It is noteworthy that alanine is encoded by the GC-rich codon family GCN.

Noncoding DNA in the *P. capuana* mitochondrial genome is more GC rich than the coding DNA (table 1); the average GC content of the telomeres is 58.7% and that of the intergenic spacer regions is 68% (table 1), with the individual intergenic regions ranging from 55% to 100%. Noncoding DNA in both the *P. parva* and *C. reinhardtii* mitochondrial genomes is GC poor (<50%), except for the telomeres of *C. reinhardtii*, which are slightly GC rich (54%) (table 1).

Repetitive Elements

Repetitive elements in the *P. capuana* mtDNA can be divided into 2 categories: short inverted repeats, which can be folded into stem-loop structures and long-terminal inverted repeats, which make up the telomeric regions.

Short Inverted Repeat Elements

Sixteen pairs of inverted repeats were identified throughout the coding and noncoding regions of the *P. capuana* mtDNA (fig. 2). The inverted repeats vary in length from 5 to 27 nt, and all 16 pairs can be folded into stem-loop (hairpin) structures (fig. 2). Although there is no sequence identity between pairs of inverted repeats and though the size of stems and loops among their different predicted secondary structures varies considerably, 2 trends are apparent: 1) the stems are GC rich (>60%), whereas the loops are less so (~50%); and 2) the location of the stem-loop structures correlate with the start and end of coding regions.

Twelve of the inverted repeat pairs have an arrangement where one inverted repeat is found in intergenic DNA and the other (matching) inverted repeat is located in an adjacent coding region, thereby resulting in a stem-loop structure that spans both coding and noncoding DNA. In 9 of 12 cases where this occurs, the "loop" portion

of the hairpin contains the start of a coding region and the "stem" component is adjacent to the end of a coding region (fig. 2). Every protein-coding gene and all but 2 of the rRNA-coding modules found in the *P. capuana* mtDNA are bordered by potential stem-loop structures.

A few of the inverted repeats share sequence identity with other parts of the genome. For example, a 20-nt perfect-match sequence corresponding to the stem of the *cox1*/*nad4* hairpin structure (fig. 2) was found inserted into *nad6*. Similarly, a 12-nt portion of the *cob*/*nad6* stem (fig. 2) was found inserted into *nad5*. The location of the inverted repeat between the *nad6* and *cob* genes (comprising one-half of the stem in the *cob*/*nad6* stem-loop structure) corresponds to the region that is fragmented in other *Polytomella* lineages.

The "global minimum" of a cumulative GC-skew plot (a plot that measures the change in G vs. C over a moving window; supplementary fig. 2, Supplementary Material online) of the main sense strand in the *P. capuana* mtDNA occurs at the apex (nucleotide 9,934) of the largest stem-loop structure in the genome—that between the regions encoding the *rrnl-L1* and *rrns-S4* gene fragments (fig. 2). The global minimum of a GC-skew plot is often used to predict the origin of replication in bacterial and mitochondrial genomes (Grigoriev 1998). A similar but much weaker stem-loop structure was found in the mitochondrial genome of *P. parva* at the corresponding region.

Telomeric Repeats

The terminal regions (telomeres) of the *P. capuana* mitochondrial genome proved unamenable to standard cloning techniques. Sequencing of the telomeres was thus achieved using terminal TdT-tailing (Förstemann et al. 2000; Bah et al. 2004) and long-walk PCR methods (Katz et al. 2000). The nature of these protocols is that the TdT-tailing method allows access to the 3′ end of a telomere, whereas the long-walk PCR approach works outwards on the strand containing the 5′ end; by using each of these techniques, one can sequence both strands of a telomere independently.

Sequencing results from TdT-tailing and long-walk PCR (fig. 3) suggest that the terminal regions of the *P. capuana* mtDNA exist (in vitro) in 2 separate conformations: a closed (hairpin-loop) conformation and an open (nicked-loop) conformation. The nucleotide sequence from both these conformations appears to be identical; however, that from the closed conformation seems to terminate with a 220-nt single-stranded loop, whereas in the open conformation this loop appears nicked (fig. 3). The location of the nick was shown to vary but was most often observed at the apex of the loop. Further experiments using standard PCR techniques were performed to confirm the sequence of the telomeric regions. As expected, PCR reactions were able to proceed through the terminal loop structure. Gel analyses and restriction digest results also support the idea that the *P. capuana* mtDNA telomeres exists in both an open and a closed conformation (supplementary figs. 3 and 4, Supplementary Material online).

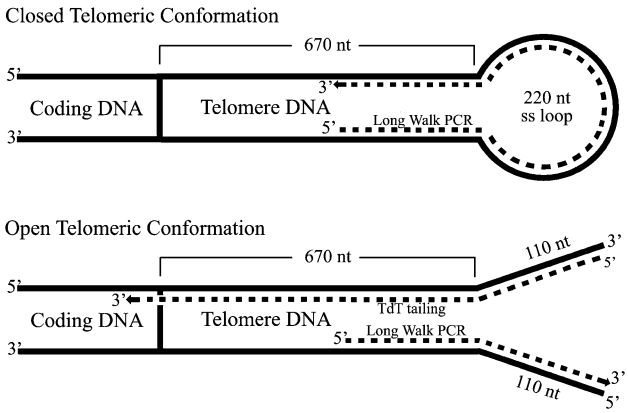

FIG. 3.—TdT-tailing and long-walk PCR reactions were independently performed on terminal restriction fragments (LTF and RTF from fig. 1) coming from the left and right ends of the *Polytomella capuana* mtDNA. Sequencing of the long-walk PCR and TdT-tailing products suggest that the telomeres can exist in either a closed (hairpin-loop) or an open (nicked-loop) conformation. Products obtained by either long-walk PCR or TdT tailing are depicted with a dashed line.

Discussion

Nucleotide Composition Bias

The *P. capuana* mtDNA has the highest GC content (57.7%) of any completely sequenced mitochondrial (or organelle) genome currently deposited in the NCBI organelle genome data bank (http://www.ncbi.nlm.nih.gov/genomes/static/euk_o.html; July 2007). Heretofore, all completely sequenced green-algal mitochondrial genomes have had a GC content below 46%, which leads to the question: why is the mtDNA of *P. capuana* GC rich? Or alternatively, why is it so AT poor? Hypotheses about biased nucleotide composition fall into 2 categories, which we will call: selectionist based and neutralist based. The former suggests that nucleotide bias is the result of natural selection; for example, this approach is often used to argue that GC richness is an adaptation to homeothermy (Jabbari and Bernardi 2004) or UV tolerance (Singer and Ames 1970; Kellogg and Paul 2002). In opposition, neutralist-based hypotheses posit that nucleotide inequities arise by either biased mutation pressure or biased gene conversion (BGC). Biased mutational pressure is often applied to the low GC content of organelle and endocellular parasite genomes (Dybvig and Voelker 1996; Ogata et al. 2001; Kusumi and Tachida 2006), whereas BGC, which results from biased repair of mismatches in heteroduplexed recombination intermediates (Holmquist 1992; Eyre-Walker 1993; Galtier et al. 2001; Galtier and Duret 2007), has been correlated to GC richness. Furthermore, gene conversion, which is believed to occur in organelle genomes (Birky and Walsh 1992; Walsh 1992; Khakhlova and Bock 2006; Tatarenkov and Avise 2007), may be GC biased in mitochondrial systems: a positive correlation between the GC content of repeat sequences and their recombinogenic properties in mitochondrial genomes has been observed in fungi and plants (Dieckmann and Gandy 1987; Clark-Walker 1989; Weiller et al. 1989; Nakazono et al. 1995), and has been suggested for certain Reinhardtii-clade green algae (Boer and Gray 1991; Nedelcu 1997, 1998; Nedelcu and Lee 1998).

Within the *P. capuana* mtDNA, the GC bias is highest at what are typically regarded as among the most neutrally evolving positions (intergenic and 4-fold degenerate sites) and it is lowest at what are generally considered more functionally constrained positions (first and second codon sites). This disproportion of GC content among the different regions of the *P. capuana* mitochondrial genome is best explained by the negative selection principle of the neutral theory of molecular evolution (Kimura 1983). If a neutral process is responsible for the GC bias of *P. capuana* mtDNA, is that process the effect of biased mutation pressure or BGC_{GC}?—or both? Any attempt to choose between these possibilities is purely speculative; however, we tend to favor a predominant role of BGC_{GC}, for reasons discussed below.

Duret et al. (2006) proposed that in the nuclear DNA of mammals the GC content of a given region reflects a balance between an AT-biased mutation process and BGC_{GC} (a GC-biased fixation process). They further suggest (following the theory of Galtier 2004) that GC-rich regions form rapidly during times when the recombination rate is high enough for BGC_{GC} to be effective; then, once the recombination rate decreases, the GC content declines slowly as a result of AT-mutation pressure. If we suppose that mitochondrial genomes have an AT-biased mutation process and that gene conversion in mtDNA is GC biased, we can apply the model of Duret et al. to mitochondrial systems. Under this premise, the history of most mitochondrial genomes would reflect an AT-biased mutation process and a recombination rate where BGC_{GC} is ineffective, thereby giving rise to AT-rich mitochondrial genomes. For the mtDNA of *P. capuana*, however, we suggest a recent history with high levels of recombination, thus shifting the nucleotide composition toward G and C. According to this hypothesis, any new mutation in one of the multiple copies of the *P. capuana* mtDNA that results in heteroplasmy, where one allele is A•T at a given site and the other allele is G•C at the same position, should be preferentially converted to the G•C allele via BGC_{GC}. Reasons why the *P. capuana* mtDNA may have undergone an increased rate of recombination would have to be entirely suppositional at this time. One intriguing observation, however, is that the isolate of *P. capuana* used for this research came from Italy (Capua), whereas the available mtDNA sequences of *P. parva* came from a strain isolated in the United Kingdom (Cambridge). Sun exposure between these 2 regions differs substantially, invoking the possibility that an elevated recombination rate may be a repair response to UV damage in the *P. capuana* mtDNA; but because little is known about the full habitat range of either *P. capuana* or *P. parva*, little weight can be placed on this observation. Moreover, we have been unable to get *P. capuana* to form cysts in the laboratory, unlike the other known species of *Polytomella*, which may make it more susceptible to UV damage.

Linear Mitochondrial Genomes

The *P. capuana* mitochondrial genome is one of several examples of linear-type mtDNA from the Reinhardtii clade of chlorophycean green algae (for a compilation see Laflamme and Lee 2003; Mallet and Lee 2006; Popescu and Lee 2006). When comparing the telomeres of the *P. capuana* mtDNA with those of other Reinhardtii-clade algae for which telomere data are available (namely *C. reinhardtii* and *P. parva*), no universal themes are apparent. Both the length and sequence of the telomeres in the *P. capuana* mtDNA differ substantially from those of *P. parva* and *C. reinhardtii*. Furthermore, the terminal structures of the *P. capuana* mtDNA, which appear to exist in either closed (hairpin-loop) or open (nicked-loop) conformations, are different from those of *C. reinhardtii*, which are made up of 3′ single-stranded extensions (Vahrenholz et al. 1993). We are unable to exclude the possibility, however, that the open telomeric conformation is the result of nicking during the DNA extraction process and does not normally exist in vivo. *Polytomella capuana* is not the first example of a linear mtDNA with terminal hairpins; they are found in the linear mitochondrial genome of the yeast *Pichia* (Dinouël et al. 1993), at one end of the *Paramecium* mtDNA (Pritchard and Cummings 1981), and also in the mitochondrial plasmid of the plant pathogenic fungus *Rhizoctonia solani* (Miyashita et al. 1990). Other examples of this telomeric structure come from viruses of eukaryotic cells (Baroudy et al. 1982; González et al. 1986), including a virus which infects certain species of the green-algal genus *Chlorella* (Rohozinski et al. 1989), and from the bacterial plasmids of the genus *Borrelia* (Hinnebusch and Barbour 1991). Although the terminal structures of the *P. parva* mitochondrial genome are not characterized, preliminary gel electrophoresis results (Smith and Lee, unpublished data) suggest that they are similar in conformation to those of *P. capuana*.

All linear genomes must develop a strategy to overcome the end replication problem, as defined by Olovnikov (1971) and Watson (1972). For the *C. reinhardtii* mitochondrial genome, 2 replication models have been proposed (Vahrenholz et al. 1993). One model involves reverse transcription of an internal repeat via a putative mitochondrial–encoded reverse transcriptase (*RTL*), whereas the second model takes into account that the reverse transcriptase gene may be nonfunctional. No open reading frames resembling a reverse transcriptase-like gene were found in the *P. capuana* (or *P. parva*) mitochondrial genome, and the fact that the structure of its telomeres depart from that of *C. reinhardtii* suggest that it uses a different replication strategy. Although in the case of *P. capuana* no strategy is apparent, many replication models for linear genomes with terminal hairpins have been suggested (Cavalier-Smith 1974; Bateman 1975; Pritchard and Cummings 1981; Baroudy et al. 1983; Dinouël et al. 1993; Traktman 1996).

Origin of Fragmentation

Although linear-fragmented mitochondrial genomes have been observed in other eukaryotic lineages, including 4 classes of *Cnidaria* (Warrior and Gall 1985; Bridge et al. 1992; Ender and Schierwater 2003) and the ichthyosporean *Amoebidium parasiticum* (Burger et al. 2003), the *Polytomella* genus represents a unique example in that substantial mtDNA sequence data exist for both a linear-fragmented and a linear-intact genome from 2 closely related taxa, thus allowing for comparative analyses.

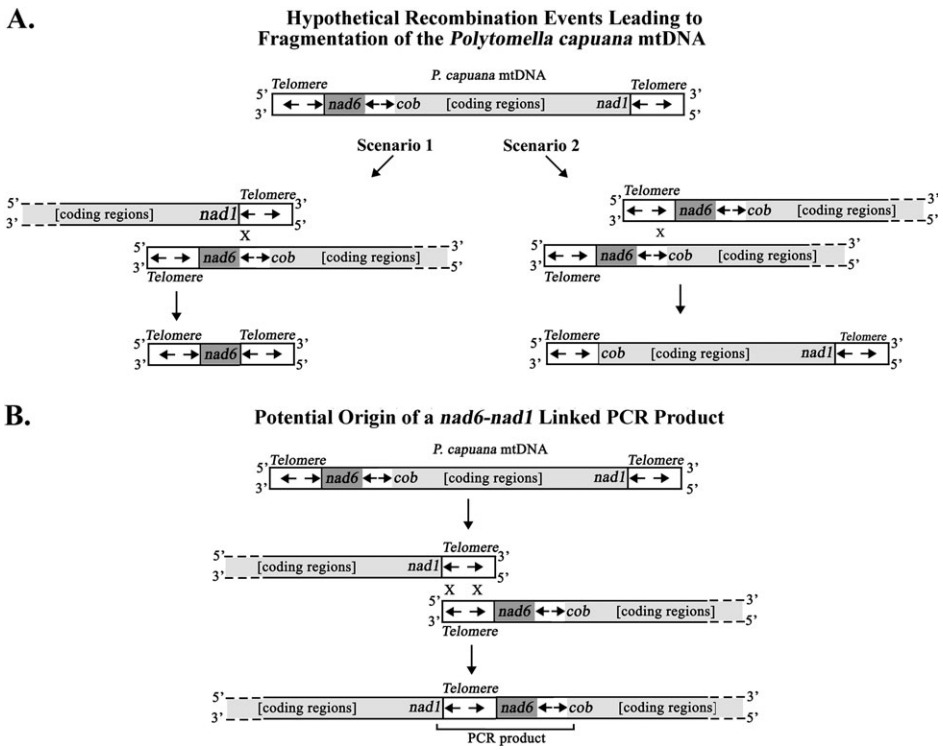

FIG. 4.—(*A*) Outline of 2 scenarios involving illegitimate recombination between short inverted repeats in the *Polytomella capuana* mtDNA, which produce products structurally similar to the 3.5-kb and 13.5-kb mtDNAs of *Polytomella parva*. (*B*) PCR product consistent with an illegitimate recombinational event involving the short inverted repeat sequences in the telomere regions resulting in a *nad6*/*nad1*-linked fragment. The short inverted repeats are represented by arrows.

Assuming that the linear bipartite mitochondrial genome of *P. parva* and *Polytomella* strain 63-10 were derived from an ancestral unfragmented linear molecule, we can posit that the ancestral mtDNA conformation may have been similar to that of *P. capuana*: a single linear chromosome with terminal inverted repeats. Furthermore, because the gene arrangement of the unfragmented *P. capuana* mitochondrial genome is parallel to that of the fragmented *P. parva* mitochondrial genome—fragmentation notwithstanding—we can consider the ancestral *Polytomella* gene arrangement to be equivalent to that of *P. capuana*, where the gene encoding *nad6* is found internal to the left telomere. Under these premises, the *P. capuana* mitochondrial genome can act as a model for understanding fragmentation of the ancestral *Polytomella* mtDNA.

Several features of the *P. capuana* mtDNA suggest that the *nad6* gene is in an unstable region. Its terminal position lends itself to recombination and possible fragmentation more readily than other internally located genes: recombination rates have been shown to be higher at the termini of linear chromosomes as compared with their more centrally located regions (Eichler and Sankoff 2003; See et al. 2006). Also, the intergenic sequence between the *nad6* and *cob* genes is comprised of a potentially unstable GC-rich inverted repeat: in the mitochondrial genomes of *Neurospora cerevisiae* and *Saccharomyces cerevisiae* GC-rich inverted repeat sequences were shown to have inflated rates of recombination causing genome rearrangements or deletion mutations that were maintained in the population of mtDNAs (Almasan and Mishra 1988; Clark-Walker 1989). In fact, portions of the *P. capuana nad6*/*cob* intergenic sequence were found inserted into both the gene for *nad5* and the telomeres, indicating that this sequence may have mobile properties, perhaps similar to those of the GC clusters found in the mtDNA of certain *Saccharomyces* species (de Zamaroczy and Bernardi 1986). By using a portion of the short inverted repeat sequences in the *nad6*/*cob* intergenic region and homologous sequences in the telomere regions, we were able to outline a scenario 1 and 2 involving illegitimate recombination between *P. capuana* mitochondrial genomes, which produce products structurally similar to the 3.5-kb and 13.5-kb mtDNAs of *P. parva*, respectively (fig. 4*A*). Lineages with such fragmented mtDNA forms may have become fixed by random genetic drift especially if *Polytomella* populations went through a bottleneck. According to this possibility, one might expect an ongoing, low-level production of such fragmented mtDNA forms from the intact mtDNA structure in *P. capuana*. Although we were not able to detect these forms by a PCR approach, we were able to reliably detect PCR products, using a wide range of primer combinations, that are consistent with other illegitimate recombination events involving the short inverted repeat sequences in the telomere regions as shown in figure 4*B*.

Short Inverted Repeats

The short GC-rich inverted repeat sequences in the *P. capuana* mtDNA evoke several questions regarding their

evolution and function, such as: 1) do they play a role in gene expression, 2) do they have mobile properties, and 3) are they related to the inverted repeats found in other mitochondrial genomes. Inverted repeat sequences capable of forming stem-loop structures have been described in the mitochondrial genomes of animals, fungi, plants, and a series of Reinhardtii-clade green algae including *C. reinhardtii* (Boer and Gray 1986, 1991) and 2 species of *Volvox* (Aono et al. 2002). In the above cases, the inverted repeat sequences are generally restricted to intronic or intergenic regions, and in many cases, they have been implicated in mobility or RNA processing (Boer and Gray 1986, 1991; Nedelcu and Lee 1998; Aono et al. 2002). What distinguishes the inverted repeats of the *P. capuana* mtDNA from those of most other mitochondrial systems is that they span both coding and noncoding DNA, often resulting in stem-loop structures that contain the start or the end of a gene; this arrangement suggests that the inverted repeats and their putative secondary structures may have a role in gene expression, perhaps akin to mammalian mitochondrial systems, where large polycistronic transcripts are processed by cleavage at the boundaries of tRNA sequences, which flank almost every gene (Clayton 1984). We are aware of only one other mitochondrial genome with a similar orientation of inverted repeats: the dinoflagellate *Amphidinium carterae* (Nash et al. 2007); however, the utility of the repeats within this taxon are also unknown. The absence of inverted repeat sequences from the *P. parva* mitochondrial genome indicates that the inverted repeats in the *P. capuana* mtDNA may be invasive elements, conceivably appearing in *P. capuana* after its divergence from *P. parva*; but we can not eliminate the possibility that the elements were lost from the *P. parva* mtDNA. Two of the inverted repeats share sequence identity with other parts of the genome; 20 nt of the *cox1/nad4* hairpin structure was found inserted into *nad6*, and 12 nt of the *cob/nad6* stem was inserted into *nad5*. This suggests that some of the inverted repeat sequences in the *P. capuana* mitochondrial genome may have mobile properties, but the lack of conservation in primary sequence among the different inverted repeats implies that if there is mobility it may be dependent on secondary structure rather than a specific nucleic acid sequence.

Supplementary Material

Supplementary figures 1–4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are indebted to Tudor Borza for his expert advice throughout the experimental and writing phases of this work. We also thank Gertraud Burger and Emmet O'Brien for graciously providing the nucleotide composition data of the currently available mitochondrial genome sequences. 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. received a Canada Graduate Scholarship from NSERC.

Literature Cited

Adl SM, Simpson AG, Farmer MA, et al. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol. 52:399–451.

Almasan A, Mishra NC. 1988. Molecular characterization of the mitochondrial DNA of *Schizosaccharomyces pombae* mutator strains. J Mol Biol. 202:725–734.

Altschul S, Gish W, Miller W, Myers E, Lipman D. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

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

Bah A, Bachand F, Clair E, Autexier C, Wellinger RJ. 2004. Human telomeres and an attempt to express a functional human telomerase in yeast. Nucleic Acids Res. 32:1917–1927.

Baroudy BM, Venkatesan S, Moss B. 1982. Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. Cell. 28:315–324.

Baroudy BM, Venkatesan S, Moss B. 1983. Structure and replication of vaccinia virus telomeres. Cold Spring Harbor Symp Quant Biol. 47:723–729.

Bateman AJ. 1975. Simplification of palindromic telomere theory. Nature. 253:379–380.

Birky CW, Walsh JB. 1992. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics. 130:677–683.

Boer PH, Gray MW. 1986. The URF 5 gene of *Chlamydomonas reinhardtii* mitochondria: DNA sequence and mode of transcription. EMBO J. 5:21–28.

Boer PH, Gray MW. 1988. Transfer RNA genes and the genetic code in *Chlamydomonas reinhardtii* mitochondria. Curr Genet. 14:583–590.

Boer PH, Gray MW. 1991. Short dispersed repeats localized in spacer regions of *Chlamydomonas reinhardtii* mitochondrial DNA. Curr Genet. 19:309–312.

Bridge D, Cunningham CW, Schierwater B, DeSalle R, Buss LW. 1992. Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc Natl Acad Sci USA. 89:8750–8753.

Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci USA. 100:892–897.

Burger G, Lang BF. 2003. Parallels in genome evolution in mitochondria and bacterial symbionts. IUBMB Life. 55:205–212.

Cavalier-Smith T. 1974. Palindromic base sequence and replication of eukaryote chromosome ends. Nature. 250:467–470.

Clark-Walker GD. 1989. In vivo rearrangement of mitochondrial DNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 86:8847–8851.

Clayton DA. 1984. Transcription of the mammalian mitochondrial genome. Annu Rev Biochem. 53:573–594.

Denovan-Wright EM, Nedelcu AM, Lee RW. 1998. Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. Plant Mol Biol. 36:2315–2323.

de Zamaroczy M, Bernardi G. 1986. The GC clusters of the mitochondrial genome of yeasts and their evolutionary origin. Gene. 41:1–22.

Dieckmann CL, Gandy B. 1987. Preferential recombination between GC clusters in yeast mitochondrial DNA. EMBO J. 6:4196–4203.

Dinouël N, Drissi R, Miyakawa I, Sor F, Rousset S, Fukuhara H. 1993. Linear mitochondrial DNAs of yeasts: closed-loop structure of the termini and possible linear-circular conversion mechanisms. Mol Cell Biol. 13:2315–2323.

Duby F, Cardol P, Matagne RF, Remacle C. 2001. The structure of the telomeric ends of mt DNA, transcriptional analysis and complex I assembly in the *dum24* mitochondrial mutant of *Chlamydomonas reinhardtii*. Mol Genet Genomics. 266:109–114.

Duret L, Eyre-Walker A, Galtier N. 2006. A new perspective on isochore evolution. Gene. 30:71–74.

Dybvig K, Voelker LL. 1996. Molecular biology of mycoplasmas. Annu Rev Microbiol. 50:25–57.

Eichler EE, Sankoff D. 2003. Structural dynamics of eukaryotic chromosome evolution. Science. 301:793–797.

Ender A, Schierwater B. 2003. Placozoa are not derived cnidarians: evidence from molecular morphology. Mol Biol Evol. 20:130–134.

Eyre-Walker A. 1993. Recombination and mammalian genome evolution. Proc R Soc Lond B Biol Sci. 252:237–243.

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.

Förstemann K, Hoss M, Lingner J. 2000. Telomerase-dependent repeat divergence at the 3′ ends of yeast telomeres. Nucleic Acids Res. 28:2690–2694.

Galtier N. 2004. Recombination, GC-content and the human pseudoautosomal boundary paradox. Trends Genet. 20:347–349.

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23:273–277.

Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics. 159:907–911.

Gerloff-Elias A, Spijkerman E, Pröschold T. 2005. Effect of external pH on the growth, photosynthesis and photosynthetic electron transport of *Chlamydomonas acidophila* Negoro, isolated from an extremely acidic lake (pH 2.6). Plant Cell Environ. 28:1218–1229.

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. Nature. 407:757–762.

González A, Talavera A, Almendral JM, Viñuela E. 1986. Hairpin loop structure of African swine fever virus DNA. Nucleic Acids Res. 14:6835–6844.

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.

Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 26:2286–2290.

Hinnebusch J, Barbour AG. 1991. Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. J Bacteriol. 173:7233–7239.

Holmquist GP. 1992. Chromosome bands, their chromatin flavors, and their functional features. Am J Hum Genet. 51:17–37.

Howe CJ, Barbrook AC, Koumandou VL, Nisbet RER, Symington HA, Wightman TF. 2002. Evolution of the chloroplast genome. Philos Trans R Soc Lond B Biol Sci. 358:99–107.

Jabbari K, Bernardi G. 2004. Body temperature and evolutionary genomics of vertebrates: a lesson from the genomes of *Takifugu rubripes* and *Tetraodon nigroviridis*. Gene. 333:179–181.

Katz LA, Curtis EA, Pfunder M, Landweber LF. 2000. Characterization of novel sequences from distantly related taxa by walking PCR. Mol Phylogenet Evol. 14:318–321.

Kellogg CA, Paul AJ. 2002. Degree of ultraviolet radiation damage and repair capabilities are related to G + C content in marine vibriophages. Aquat Microb Ecol. 27:13–20.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46:85–94.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kurtz S, Choundhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29:4633–4642.

Kusumi J, Tachida H. 2006. Compositional properties of green-plant plastid genomes. J Mol Evol. 60:417–425.

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

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Mallet M, Lee RW. 2006. Identification of three distinct *Polytomella* lineages based on mitochondrial DNA features. J Eukaryot Microbiol. 53:79–84.

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.

Miyashita S, Hirochika H, Ikeda J, Hashiba T. 1990. Linear plasmid DNAs of the plant pathogenic fungus *Rhizoctonia solani* with unique terminal structures. Mol Gen Genet. 220:165–171.

Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell. 108:583–586.

Nakayama T, Watanabe S, Mitsui K, Uchida H, Inouye I. 1996. The phylogenetic relationship between the Chlamydomonadales and Chlorococcales inferred from 18S rDNA sequence data. Phycol Res. 44:47–55.

Nakazono M, Tsutsumi N, Sugiura M, Hirai A. 1995. A small repeated sequence contains the transcription initiation sites for both *trnfM* and *rrn26* in rice mitochondria. Plant Mol Biol. 28:343–346.

Nash EA, Barbrook AC, Edwards-Stuart RK, Bernhardt K, Howe CJ, Nisbet RER. 2007. Organization of the mitochondrial genome in the dinoflagellate *Amphidinium carterae*. Mol Biol Evol. 24:1528–1536.

Nedelcu AM. 1997. Fragmented and scrambled mitochondrial ribosomal RNA coding regions among green algae: a model for their origin and evolution. Mol Biol Evol. 14:506–517.

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.

Ogata H, Audic S, Renesto-Audiffren P, et al. (11 co-authors). 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science. 293:2093–2098.

Olovnikov AM. 1971. Principle of marginotomy in template synthesis of polynucleotides. Dokl Akad Nauk SSSR. 201:1496–1499.

Popescu CE, Lee RW. 2006. Mitochondrial genome sequence evolution in *Chlamydomonas*. Genetics. 175:819–826.

Pringsheim EG. 1955. The Genus *Polytomella*. J Protozool. 2:137–145.

Pritchard AE, Cummings DJ. 1981. Replication of linear mitochondrial DNA from *Paramecium*: sequence and structure of the initiation-end crosslink. Proc Natl Acad Sci USA. 78:7341–7345.

Pröschold T, Marin B, Schlösser UG, Melkonian M. 2001. Molecular phylogeny and taxonomy revision of *Chlamydomonas* (Chlorophyta). I. Emendation of *Chlamydomonas*

Ehrenberg and *Chloromonas* Gobi, and description of *Oogamochlamys* gen. nov. and *Lobochlamys* gen. nov. Protist. 152:265–300.
Rocha EP, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. 18:291–294.
Rohozinski J, Girton LE, Van Etten JL. 1989. *Chlorella* viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. Virology. 168:363–369.
Ryan R, Grant D, Chiang KS, Swift H. 1978. Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA. 75:3268–3272.
See DR, Brooks S, Nelson JC, Brown-Guedira G, Friebe B, Gill BS. 2006. Gene evolution at the ends of wheat chromosomes. Proc Natl Acad Sci USA. 103:4162–4167.
Singer CE, Ames BN. 1970. Sunlight ultraviolet and bacterial DNA base ratios. Science. 170:822–826.
Tatarenkov A, Avise JC. 2007. Rapid concerted evolution in animal mitochondrial DNA. Proc R Soc B. 274:1795–1798.
Traktman P. 1996. Poxvirus DNA replication. In: DePamphilis ML, editor. DNA replication in eukaryotic cells. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 775–798.
Vahrenholz C, Riemen G, Pratje E, Dujon B, Michaelis G. 1993. Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. Curr Genet. 24:241–247.
Walsh JB. 1992. Intracellular selection, conversion bias, and the expected substitution rate of organelle genes. Genetics. 130:939–946.
Warrior R, Gall J. 1985. The mitochondrial DNA of *Hydra attenuata* and *Hydra littoralis* consists of two linear molecules. Arch Sci. 38:439–445.
Watson JD. 1972. Origin of concatemeric T7 DNA. Nat New Biol. 239:197–201.
Weiller G, Schueller CM, Schweyen RJ. 1989. Putative target sites for mobile G + C rich clusters in yeast mitochondrial DNA: single elements and tandem arrays. Genetics. 218: 272–283.
Zucker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.

Charles Delwiche, Associate Editor

Accepted November 1, 2007