A Family of Selfish Minicircular Chromosomes with Jumbled Chloroplast Gene Fragments from a Dinoflagellate

Zhaoduo Zhang, T. Cavalier-Smith, and B. R. Green

Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Chloroplast genes of several dinoflagellate species are located on unigenic DNA minicircular chromosomes. We have now completely sequenced five aberrant minicircular chromosomes from the dinoflagellate Heterocapsa triqueta. These probably nonfunctional DNA circles lack complete genes, with each being composed of several short fragments of two or three different chloroplast genes and a common conserved region with a tripartite 9G-9A-9G core like the putative replicon origin of functional single-gene circular chloroplast chromosomes. Their sequences imply that all five circles evolved by differential deletions and duplications from common ancestral circles bearing fragments of four genes: psbA, psbC, 16S rRNA, and 23S rRNA. It appears that recombination between separate unigenic chromosomes initially gave intermediate heterodimers, which were subsequently stabilized by deletions that included part or all of one putative replicon origin. We suggest that homologous recombination at the 9G-9A-9G core regions produced a psbA/psbC heterodimer which generated two distinct chimeric circles by differential deletions and duplications. A 23S/16S rRNA heterodimer more likely formed by illegitimate recombination between 16S and 23S rRNA genes. Homologous recombination between the 9G-9A-9G core regions of both heterodimers and additional differential deletions and duplications could then have yielded the other three circles. Near identity of the gene fragments and 9G-9A-9G cores, despite diverging adjacent regions, may be maintained by gene conversion. The conserved organization of the 9G-9A-9G cores alone favors the idea that they are replicon origins and suggests that they may enable the aberrant minicircles to parasitize the chloroplast’s replication machinery as selfish circles.

Introduction

It has recently been discovered that chloroplast genes of typical peridinean dinoflagellates are on minute circular chromosomes, each containing only a single gene (Zhang, Green, and Cavalier-Smith 1999; Barbrook and Howe 2000). Each minicircular chromosome consists of one complete gene and a noncoding tripartite or bipartite region that is highly conserved among all genes of that species. Only 11 different chloroplast single-gene circles have been reported from dinoflagellates—9 in Heterocapsa (Zhang, Green, and Cavalier-Smith 1999) and 5 in Amphidinium (Barbrook and Howe 2000)—in strong contrast to approximately 100–200 genes found in all other photosynthetic chloroplast genomes (Turmel, Otis, and Lemieux 1999). Although direct physical proof that these minicircles are located in chloroplasts is lacking, they include the genes for chloroplast 23S and 16S rRNA, which are very large, highly charged molecules that have never been shown to be transported across a membrane and therefore are almost certainly in the chloroplast compartment. Since the other minicircular genes share the highly conserved noncoding region with the rRNA genes and undergo gene conversion with them, it is most unlikely that they are in a different compartment. If the other nine genes were in a different compartment, their encoded proteins would have to be imported into the chloroplast to carry out their functions, which would require conserved import signals, for which there is no evidence at all in their gene sequences.

Here we show that in addition to normal functional minicircles, Heterocapsa triqueta contains a family of related, probably nonfunctional minicircular chromosomes carrying fragments of chloroplast genes, and we present complete sequences of five such nongenic minicircles. Each minicircle consists of jumbled chloroplast gene fragments derived from four separate unigenic circles and the very conserved tripartite 9G-9A-9G region. Plasmid-like DNA has previously been found in algal chloroplasts but has not been well characterized: in the diatom Cylindrotheca fusiformis (Hildebrand et al. 1992; Jacobs et al. 1992) and the green alga Acetabularia (Ebert, Tymms, and Schweiger 1985), homology between parts of the plasmid sequence and chloroplast DNA has been shown by hybridization. The linear plasmids of the green alga Ernodesmis verticillata are localized around the chloroplast pyrenoid (La Claire and Wang 2000a) and have a unique hairpin-like structure (La Claire and Wang 2000b) but appear to contain only fragments of chloroplast genes (La Claire et al. 1998).

In the case of the aberrant H. triqueta minicircles we report here, it is unlikely that the fragments of chloroplast genes encode functional polypeptides, so the aberrant minicircles are probably nonfunctional selfish DNA like that widespread in nuclear genomes (Doolittle and Sapienza 1980; Orgel and Crick 1980). The very conserved tripartite 9G-9A-9G region of these minicircles may predispose them to form heterodimers by homologous recombination at these sites. We propose a model for the origin of all five circles from ancestral heterodimeric circles containing fragments of four genes—psbA, psbC, and 16S rRNA and 23S rRNA.
genes—followed by numerous deletions and some duplications.

Materials and Methods
DNA Isolation and Library Construction

Isolation of satellite DNA from the axenic dinoflagellate *H. triqueta* (CCMP 449) and the construction of three libraries for the satellite DNA were described elsewhere (Zhang, Green, and Cavalier-Smith 1999).

DNA Sequencing and Sequence Assembly

Plasmid DNA preparation and the sequencing were as described (Zhang, Green, and Cavalier-Smith 1999). Sequencing reactions used the Perkin-Elmer GeneAmp 9600 with the ABI cycle sequencing protocol: 94°C for 5 s, 50°C for 5 s, and 60°C for 4 min for 25 cycles. Each reaction contained 2–3 μl FS-Taq or Bigdye, 20–30 ng DNA (purified PCR product) or 100–150 ng DNA (plasmids), 3–5 pmol primers, and distilled water to 10 μl. The sequencing samples were precipitated by adding 1/10 volume 3 M sodium acetate (pH 5.2) and 2 volumes 95% ethanol, quenched on ice for 10 min, centrifuged for 20 min, air-dried, and analyzed by an ABI 373 or 377 automatic sequencer. Sequences were imported into Gap 4 of Staden and edited using Trev; contigs were generated from overlapped clone sequences using the options “shotgun assembly” and “find internal joins” of Gap 4 (http://www.mrc-lmb.cam.ac.uk/pubseq).

PCR Reactions

Specific primers to verify sequences and circularity were based on the sequences of normal *H. triqueta* *psbA*, *psbC*, 16S rRNA, and 23S rRNA genes. PCR reactions were for 35 cycles of 94°C for 30 s and 55°C for 30 s followed by 2 min at 72°C in a GeneAmp PCR system 9600 (Perkin-Elmer). Each reaction (50 μl) contained 0.2 mM dNTP, 1 × PCR buffer, 0.1–1.0 μg template DNA, 50–200 pmol primers, 2.0 or 2.5 mM MgCl₂, and 1.5–2.5 U Taq polymerase (Sigma or Rose). PCR products were purified from 1% agarose gels by a purification kit (Amersham-Pharmacia Biotech) and sequenced directly. Sequences of PCR products were integrated into appropriate contigs as above.

BLAST Searches

NCBI’s sequence similarity search tool BLAST (http://www.ncbi.nlm.nih.gov/BLAST) was used to analyze DNA sequences. Each sequence was compared against the entire GenBank database, which included the nine unigenic circles; homologs of these sequences were further characterized by carefully examining multiple alignments of related regions using the Staden package (http://www.mrc-lmb.cam.ac.uk/pubseq).

Results

Minicircles Bearing Multiple Chloroplast Gene Fragments

In addition to the nine DNA minicircles carrying complete chloroplast genes that we have previously reported (Zhang, Green, and Cavalier-Smith 1999), random sequencing of our plasmid libraries of Heterocapsa chloroplast DNA generated five additional circular contigs, each consisting of several overlapped clones. These five aberrant minicircles bore only fragments of four separate chloroplast genes: 16S rRNA, 23S rRNA, *psbA*, and *psbC* (fig. 1a). Like the unigenic circles, they all have very conserved tripartite 9G-9A-9G regions of ~600–800 nt. However, they differ dramatically from unigenic circles in that none contain a complete chloroplast gene. They are a little smaller (just over 2 kb) than the average size (about 2.6 kb) of known functional unigenic circles.

All of the gene fragments are identical or almost identical in sequence to the corresponding regions of the normal circles, suggesting that they arose by combining pieces from four separate circles and (as we argue below) that the essential identity of the gene fragments has been maintained by regular gene conversions, which earlier evidence suggested was frequent between normal circles (Zhang, Green, and Cavalier-Smith 1999). Base substitutions are very rare, with one example being two changes of an A to a T in the 16S rRNA fragment of circle 2 compared with circle 1 and the original 16S rRNA circle. These circular chimeras of unrelated gene fragments are unlikely to be cloning artifacts, since none of the fragment junctions have *Sau3A* sites as would be expected if that were so. The reality of the chimeric circles in uncloned DNA was directly demonstrated for circle 1 by sequencing the inverse PCR products obtained using primers from two different genes: primer pairs 23S3/16S1 and 23S4/16S2. Furthermore, PCR reactions performed using primer pairs 23ST/99C2 and 23ST/16S2 (circle 2), 116C1/116C3 (circle 3), and 99C2/99C5 (circle 4) all gave products of the predicted sizes.

Each gene fragment on an aberrant circle has the same orientation with respect to the 9G-9A-9G region as its homologous segment on the normal unigenic circle (fig. 1b). This regularity would not have been found had the fragments been generated artificially and randomly ligated during cloning, but it is exactly what would be expected if they evolved in vivo from naturally chimeric molecules through many deletions. The patterns of sequence identity shared between the aberrant and normal circles are complex and are shown in figure 1b. Comparison of the gene fragments with normal circles showed that all originate from only five regions: A (720 bp) and B (399 bp) of the 23S rRNA circle, C (181 bp) of 16S rRNA, D (322 bp) of *psbA*, and E (361 bp) of *psbC* circles. Fragments homologous to each region are present in at least three circles, often with identical boundaries indicative of a common origin (table 1); e.g., fragments from the A, B, and C regions are on circles 1, 2, and 3, fragments from the E region
Jumbled chloroplast gene fragments on five minicircles. Gene names and numbers in brackets are consistent with those in table 1 and in b; each pattern represents a fragment of a particular chloroplast gene. All fragments on each circle have the same order and orientation as their coordinates on the original unigenic circles (see b); 9G L and 9G R are related to 9G , and 9G is related to 9G ; 0 is the 9G start site for nucleotide numbering. Primer pairs 16S1/23S3 and 16S2/23S4 were used in inverse PCR to confirm that circle 1 is a genuine minicircle. Primer pairs 16S2/23S7, 99C2/23S7, 116C1/116C3, and 99C2/99C5 on circles 2–4 gave PCR products of the expected sizes. The large homologous segments revealing a family relationship among the five circles are shown by lines labeled by Roman numerals (continuous, dashed, or dotted) within each circle. b, Location of chloroplast gene fragments from circles 1–5 on the normal minicircular 23S rRNA, 16S rRNA, psbA, and psbC, genes. Five regions of homology (A–E) are indicated with dotted lines; the nucleotides included are numbered at each end. Numbers next to the fragments correspond to those on the four aberrant circles (see a and table 1). The short fragments labeled by Greek letters (e.g., α, β, γ) next to large ones represent duplications (repeats) (see also fig. 2); gaps within the large fragments represent deletions. Solid arrows on the fragments of circles 1–3 homologous to the B region of 23S rRNA show the insertion sites of the 16S fragments (C region). Arrows on the fragments homologous to the E region represent the 11-bp insertions on circles 3–5.

are on circles 3, 4, and 5, and fragments from the D region are on circles 1, 2, 4, and 5.

No terminal repeats suggestive of transposon activity were observed. However, the fragment junction between 23S rRNA and 16S rRNA gene segments in circles 1–3 has three identical nucleotides, TTC (table 1), suggesting that illegitimate recombination may have occurred, integrating the 16S fragment into the 23S rRNA segments (fig. 1b, arrows in B region).

Not only the gene fragments but also their arrangement are strikingly shared between the five circles, apart from many insertions and deletions, suggesting that similar chimeric regions from the separate circles had a common evolutionary origin. The observed pattern makes it highly improbable that the circles were assembled directly from the gene fragments; instead, almost all regions outside of the 9G-9A-9G region can be derived from just five ancestral segments, each with three, four, or five gene fragments. Those segments shown by the same Roman numerals in figure 1a are essentially identical between different circles: segment I is shared by circles 1 and 2 (part of it also occurs in circles 3–5), segment II is shared by circles 2 and 3, and segment IV is shared by circles 3–5. This is true not only of genic regions shared with the unigenic circles, but also of the intervening “nongenic” regions (e.g., S2; see below).

Deletions and Duplications Within Fragments

All five sets of homologous gene fragments have suffered deletions and duplications relative to the putatively ancestral single-gene circles (fig. 1). For example, region A or segment 23S(5), starting from *9G to 9G of circle 1 (fig. 1a), has three gaps of 10, 5, and 23 bp compared with the 23S rRNA gene (fig. 1b), and there are also two 116-bp repeats, with a deletion of 24 bp in the second (α in figs. 1b and 2a). Segment 23S(5) is 98% identical to the corresponding segment of the complete 23S rRNA gene apart from gaps and duplications (table 1). Fragment 23S(6) (337 bp) on circle 2 is 99% identical to the B region of the 23S rRNA gene and has three 29-bp repeats, the second with an internal repeat of 6 bp (β, γ in figs. 1b and 2b). Similarly, fragments homologous to region B on circle 3 have a gap of 26 bp, two 39-bp repeats (δ in fig. 1b), and two repeats of 25 bp (ε in fig. 1b).

Segments IV, IV′, and IV″ (fig. 1a) are each essentially a chimera of the long nongenic region (S2; see below) and region E of the psbC minicircle that encodes part of the fifth transmembrane helix and the following loop region of the PsbC protein (fig. 2c). It seems that this region, with the central part of psbC juxtaposed to a noncoding region of unknown origin, had already evolved prior to the divergence of circles 3–5 from a common ancestor. Sequences of psbC fragments on circles 3–5 are identical to the E region of the normal psbC gene apart from indels (figs. 1 and 2c). The psbC fragment on circle 3 has a 165-bp deletion; psbC fragments on circles 4 and 5 have two and three 33-bp repeats respectively, which encode 11 amino acids of PsbC protein (κ in figs. 1b and 2c). Sequence comparison showed an insertion of 11 bp at the 40th bp from the 5′ end of the psbC fragments on the three circles (arrow in fig. 1b). The translation products of the psbC segments on the three circles gave two significant BLAST hits on PsbC protein: a small peptide of 13 amino acids and, in
Table 1  
Characteristics of Chloroplast Gene Fragments on Circles 1–5 and Their Positions on Single-Gene Circles of Heterocapsa triquetra

<table>
<thead>
<tr>
<th>ABBERRANT CIRCLE</th>
<th>COORDINATED ON GENE CIRCLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle</td>
<td>Fragment Position</td>
</tr>
<tr>
<td>1 (2,249 bp 6 clones)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>2 (2,147 bp, 6 clones)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>3 (2,132 bp, 4 clones)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>4 (2,045 bp, 3 clones)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>5 (2,099 bp, 5 clones)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Note.—# indicates fragments that are almost identical between related circles if deletions and duplications are not taken into account. The position of each fragment on the single-gene circles is shown in parentheses if it is not in a coding region. Sequences (10 bp) at each fragment end are shown. Indels were not used in calculated identity between a chloroplast gene fragment and its homolog on a normal gene circle. Repeated fragments are underlined. * indicates sequences that are shown in figure 2. Bold numbers are inserts in psbC fragments.
Shown are 116-bp repeats on circle 1; the second repeat (a deletion of 24 bp.

The S2 nongenic region present on circles 2±5 is identical to the 9GR of the 23S rRNA circle, and the extra 9G of circle 2 is identical to the 9GR of the 16S rRNA and psbC circles except for two base substitutions. On circle 1, apart from the two small indels, a region of 791 bp (23S(5)) is 98% identical to the 9GR region and the following region of the normal 23S rRNA gene circle, supporting the idea that this whole segment originated from a 23S RNA circle. However, the extra 9GR of circle 3 (#9G in fig. 1a) is identical to the 9GR of circles 1 and 5, as well as to that of the four normal chloroplast gene circles, except for three base substitutions. Therefore, the extra 9G regions on circles 1–3 may have originated independently from the 9GR or 9GL regions of different minicircles. Overall, none of the 9G-9A-9G regions of the chimeric circles are identical to each other or to any of the putatively ancestral single gene circles, although they are all very closely related to each other.

Shared variants indicative of gene conversion were found in the D2 and D3 regions of normal gene minicircles (Zhang, Green, and Cavalier-Smith 1999), and similar shared sequences were found in the five chimeric minicircles (fig. 3). The D2 region of circles 3 and 4 share 148 bp containing two 41-bp repeats with a few substitutions, making it longer than that of other minicircles; 65 of the 148 bp shared by circles 3 and 4 are present in circle 2. Gene conversions were also found between circle 1 and the 23S rRNA circles (20 bp) and between circle 5 and the 16S rRNA circle (30 bp) (fig. 3). The D3 and D4 regions (downstream of 9GR) are more conserved than the D2 region among both chimeric and unigenic circles.

Circles 1 and 2 have a region labeled S1 (fig. 1a) with no significant homology with normal single gene circles or any sequence in the databases. S1 on circle 1 (94 bp) consists of two direct repeats of 33 bp and a third one that retained only 18 bp at the 5′ end. The S1 sequence on circle 2 (76 bp) is identical to that on circle 1 except that its second 33-bp repeat has an 18-bp deletion at the 5′ end.

Nongenic Regions

We have previously shown (Zhang, Green, and Cavalier-Smith 1999) that the 9G-9A-9G noncoding region on the nine unigenic circles consists of three very conserved regions, the 9G1 (135 bp), 9A (188 bp), and 9G2 (135 bp) cores (named after their central mononucleotide tracts), with less conserved D2 and D3 intervening regions (fig. 1). The cores are almost identical among all normal gene circles, and the 9G1 and 9G2 sequences are highly related to each other. The 9G-9A-9G regions on the aberrant circles are very closely related to those of the normal single-gene circles. The 9G cores share 108 identical base pairs with those of the normal circles, while the 9A cores differ by very few substitutions. However, the 9G1 sequences of circles 2–4 have several mutations and are the most divergent of all 9G1 sequences. The presumed 9G1 of circle 3 is a core with a run of 8G's instead of 9G's; sequence alignment indicated that the 8G core was highly related to a conventional 9GR core except for a few base substitutions.

A notable feature of circles 1–3 is that each has an extra 9G core segment lying between other gene fragments instead of within a tripartite core region (fig. 1a). Sequence comparison showed that the extra 9G of circle 1 is identical to the 9GR of the 23S rRNA circle, and the extra 9G of circle 2 is identical to the 9GR of the 16S rRNA and psbC circles except for two base substitutions. On circle 1, apart from the two small indels, a region of 791 bp (23S(5)) is 98% identical to the 9GR region and the following region of the normal 23S rRNA gene circle, supporting the idea that this whole segment originated from a 23S RNA circle. However, the extra 9GR of circle 3 (#9G in fig. 1a) is identical to the 9GR of circles 1 and 5, as well as to that of the four normal chloroplast gene circles, except for three base substitutions. Therefore, the extra 9G regions on circles 1–3 may have originated independently from the 9GR or 9GL regions of different minicircles. Overall, none of the 9G-9A-9G regions of the chimeric circles are identical to each other or to any of the putatively ancestral single gene circles, although they are all very closely related to each other.

Shared variants indicative of gene conversion were found in the D2 and D3 regions of normal gene minicircles (Zhang, Green, and Cavalier-Smith 1999), and similar shared sequences were found in the five chimeric minicircles (fig. 3). The D2 region of circles 3 and 4 share 148 bp containing two 41-bp repeats with a few substitutions, making it longer than that of other minicircles; 65 of the 148 bp shared by circles 3 and 4 are present in circle 2. Gene conversions were also found between circle 1 and the 23S rRNA circles (20 bp) and between circle 5 and the 16S rRNA circle (30 bp) (fig. 3). The D3 and D4 regions (downstream of 9GR) are more conserved than the D2 region among both chimeric and unigenic circles.

Circles 1 and 2 have a region labeled S1 (fig. 1a) with no significant homology with normal single gene circles or any sequence in the databases. S1 on circle 1 (94 bp) consists of two direct repeats of 33 bp and a third one that retained only 18 bp at the 5′ end. The S1 sequence on circle 2 (76 bp) is identical to that on circle 1 except that its second 33-bp repeat has an 18-bp deletion at the 5′ end.

The S2 nongenic region present on circles 2–5 is 200–400 nt long (236, 307, 422, and 499 bp on circles 2–5). Sequence analysis showed that S2 on each circle
consisted of four motifs (fig. 4), one (the 19-bp motif 3) present in variable numbers, accounting for the length difference. Motifs 1 and 2 lack homologs in the databases. Motif 1 is almost identical on circles 3–5 but is totally different from motif 1 on circle 2 (fig. 4a). The sequences of motifs 2–4 are almost identical among the four circles except for several deletions (3, 1, and 6 bp) and substitutions on circle 2 (fig. 4). Motif 4 is immediately upstream of 9G1; its sequence is homologous to the sequence upstream of 9G on circle 1 and the 16S and 23S rRNA and psbA unigenic circles (fig. 4b). It is also related to the sequence upstream of the extra 9G of circles 1–3 (fig. 4b). The similarities of nongenic regions among the aberrant circles show that they did not originate independently. Both chloroplast gene fragments and nongenic regions on the aberrant circles must be evolutionarily related, as is particularly clearly indicated by the sharing of segment IV (psbC plus S2) by circles 3, 4, and 5.

Discussion

Are Fragmented Gene Chimeric Circles an Example of Selfish DNA?

The five chimeric circles were discovered in the process of sequencing random genomic clones from the same library of AT-rich dinoflagellate satellite DNA which yielded the complete minicircular chloroplast genes (Zhang, Green, and Cavalier-Smith 1999). Sequence analysis of the chimeric circles showed that there were no Sau3A recognition sites at the borders between different gene fragments, making a cloning artifact unlikely. The sequences of the PCR products generated using both inward- and outward-directed primers for circle 1 were identical to the sequences derived from the contig made from several overlapping genomic clones. This implies that circle 1 does indeed exist in the normal population of molecules within the dinoflagellate. PCR products from the other four circles were of the predicted sizes, implying that they, too, exist in the normal population.

It is most improbable that any of the gene fragments on these circles are functional. Most of the protein gene fragments lack start codons or have internal stop codons. It would be difficult to detect transcription of these fragments against the background of transcription from normal single-gene circles, which give products of the length expected for monocistronic transcripts (Zhang, Green, and Cavalier-Smith 1999), but it is hard to imagine that fragmentary genes would be functional in the presence of the corresponding complete genes on the unigenic circles. The coexistence of normal and fragmentary genes in H. triquetra chloroplasts is not comparable to the fragmented mitochondrial rRNA genes in the green alga Chlamydomonas (Nedelcu and Lee 1998) and in apicomplexans (Wilson and Williamson 1997), where the normal full-length genes no longer exist in the mitochondrial genomes.

The great conservation of the 9G-9A-9G regions, in marked contrast to the fragmented and jumbled gene fragments of these chimeric circles, strongly suggests that such conservation is essential for their replication and persistence. This conservation is probably a sufficient explanation for their replication and transmission from generation to generation, even if the jumbled “coding” regions are entirely devoid of function. If so, then these tiny circles are genetic parasites of the chloroplast replication machinery, and it is reasonable to regard them as selfish DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). The small size of both the functional unigenic circles and the aberrant circles, along with the high frequency of deletions in the latter, suggests that both kinds of circles are under strong selection for small size and may imply that smaller molecules have a replicative advantage over larger ones in dinoflagellate chloroplasts.

This conservation of the noncoding 9G-9A-9G regions of the five selfish circles provides a very strong argument that they must be located in the same cell compartment as the functional minicircles. We know that the noncoding regions diverge exceedingly rapidly and are totally different between different species of the
same genus (Zhang, Green, Cavalier-Smith 1999). This means that the 9G-9A-9G regions of both the normal circles and the chimeric selfish circles must be kept similar in both length and sequence by a combination of gene conversion and positive selection. Positive selection for retaining the 9G-9A-9G regions in both kinds of molecules is likely to arise because of their interactions with the replication initiation machinery and/or the DNA segregation machinery, with which they must coevolve. Both kinds of machinery must be very different between chloroplasts and other cell compartments, such as the nucleus or the mitochondria. Because of this, and because gene conversion between normal and selfish circles could occur only if they were in the same compartment, we conclude that the five selfish circles must have been present in the same compartment as the normal circles for most, and probably all, of their evolutionary history during which they diverged so radically from each other and from the normal circles. As explained in the introduction, the arguments for the chloroplast location of the normal minicircles are very strong. Therefore, although direct evidence is lacking for either, it is almost certain that both kinds of minicircles are actually located within the chloroplasts.

Origin of Chimeric Circles by Recombination and Multiple Deletions

It is clear from our data that all five circles are related and carry gene fragments of four nonhomologous genes that originated from four separate unigenic circles. We suggest that the sharing of the 9G-9A-9G regions by all of the unigenic circles in H. triquetra, kept similar by gene conversion, predisposes normal circles to form dimers by homologous recombination between their almost-identical 9G-9A-9G regions. Normally, such dimers (both homodimers and heterodimers) would be resolved into separate unigenic circles by a DNA topoisomerase II, but if by chance one 9G-9A-9G region were accidentally deleted before this happened, the dimer could persist for a while. Given the apparent selection for smaller DNA circles in dinoflagellate chloroplasts, variants with partial deletions would remain longer, perhaps indefinitely if they were smaller than unigenic circles. Such chimeric circles could also undergo recombination with other circles to form higher-order chimeras bearing more gene fragments.

We propose a model (fig. 5b) where a heterodimer of \textit{psb}A and \textit{psb}C minicircles was formed as just suggested and stabilized and compressed by one or more large deletions including a 9G-9A-9G region (e.g., double-headed arrow in fig.5b). This compressed \textit{psb}A/\textit{psb}C dimer could have given rise directly to circles 4 and 5 with more small deletions and small duplications of the D4 region of \textit{psb}A (fig. 1b) and a duplication of 33 bp within the \textit{psb}C fragments (k). Circle 5 would have had a further duplication of 63 bp near the N terminus of the \textit{psb}A gene (q). A recombinant 23S/16S rRNA minicircle might have been formed in a similar fashion. However, this mechanism would not have placed a 16S gene fragment between two 23S rRNA fragments, as is observed in circle 3. Therefore, it is more likely that a 16S/23S rRNA heterodimer was generated by illegitimate recombination (fig. 5a) at the TTC sites found in both molecules at the exact point of the
present 16S/23S chimeric boundary. The 3’ end of the 16S fragment (TTC) could have formed a heteroduplex with the TTC of the 23S gene at the integration site (table 1). Following deletion of most of the 16S circle, this chimeric 16S/23S rRNA circle (fig. 5a) could then have recombined homologously with the compressed psbA/psbC dimer at the 9G-9A-9G region to yield an intermediate circle with all five gene regions (A–E). Further deletions and duplications of this multiply chimeric circle could have generated circles 1–3. Deletions and duplications in other regions, such as D2 and S2, possibly stabilized these circles at an appropriate size (∼2 kb).

This recombination/deletion model could explain why all gene fragments have the same polarity and relative order as in the original circles; such consistency would be surprising if they had been transferred from one circle to another by transposition rather than dimerization of the whole circle. A role for transposons is highly improbable, as they are typically larger than the entire unigenic circles, which would be very unlikely to be able to harbor them. There is also no evidence at all at the fragment junctions for repeated sequences of a type generated by transposition.

Gene Conversion Between Aberrant and Normal Minicircles

Interpretation of the phylogenetic origins of the extra 9G region on the aberrant minicircles is complicated by the prevalence of gene conversion among dinoflagellate minicircles. The 9G-9A-9G regions are strongly conserved among the genic minicircles and provide compelling evidence for extensive gene conversion between 9G-9A-9G regions of different gene circles (Zhang, Green, and Cavalier-Smith 1999). All the patterns of identity or near identity between different parts of the 9G-9A-9G regions of the chimeric circles can also be explained as the result of mutation and gene conversion, making gene conversion equally important in the evolution of the chimeric circles.

Functional Implications

The 9G-9A-9G region was proposed to be the replication origin based on its high degree of conservation across all single gene circles and its propensity to fold into complicated hairpin structures (Zhang, Green, and Cavalier-Smith 1999). The fact that only this region is also well conserved in the selfish circles strongly supports this proposal. We argued above that a circle with two 9G-9A-9G regions would be quickly resolved into two separate ones. It is interesting that three of the five selfish circles have an extra 9G core, indicating that partial deletion of the 9G-9A-9G region is sufficient to stabilize a dimer. It is tempting to suggest that the 9A region, which would more easily undergo strand separation, is at the heart of the replicon origin and that the extra 9G core fragment of circles 1–3 does not function in replication.

Supplementary Material

The sequences reported in this paper have been deposited in the GenBank database with accession numbers AF004267–AF004271.

Acknowledgments

B.R.G. and T.C.-S. thank NSERC (Canada) for research grants, and T.C.-S. thanks the Canadian Institute for Advanced Research and NERC (U.K.) for fellowship support.

LITERATURE CITED


WILLIAM MARTIN, reviewing editor

Accepted April 18, 2001