Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 7737–7742, July 1996
Evolution

The miniaturized nuclear genome of a eukaryotic endosymbiont contains genes that overlap, genes that are cotranscribed, and the smallest known spliceosomal introns

(eukaryotic operons/S13/S4/small nuclear RNP E/clp protease)

PAUL R. GILSON* AND GEOFFREY I. MCFADDEN

Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Parkville, 3052 Victoria, Australia

Communicated by Adrienne E. Clarke, University of Melbourne, Parkville, Victoria, Austrialia, February 12, 1996 (received for review December 1, 1995)

ABSTRACT **Chlorarachniophyte algae contain a complex, multi-membraned chloroplast derived from the endosymbiosis of a eukaryotic alga. The vestigial nucleus of the endosymbiont, called the nucleomorph, contains only three small linear chromosomes with a haploid genome size of 380 kb and is the smallest known eukaryotic genome. Nucleotide sequence data from a subtelomeric fragment of chromosome III were analyzed as a preliminary investigation of the coding capacity of this vestigial genome. Several housekeeping genes including U6 small nuclear RNA (snRNA), ribosomal proteins S4 and S13, a core protein of the spliceosome [small nuclear ribonucleoprotein (snRNP) E], and a clp-like protease (clpP) were identified. Expression of these genes was confirmed by combinations of Northern blot analysis, *in situ* hybridization, immunocytochemistry, and cDNA analysis. The protein-encoding genes are typically eukaryotic in overall structure and their messenger RNAs are polyadenylylated. A novel feature is the abundance of 18-, 19-, or 20-nucleotide introns; the smallest spliceosomal introns known. Two of the genes, U6 and S13, overlap while another two genes, snRNP E and clpP, are cotranscribed in a single mRNA. The overall gene organization is extraordinarily compact, making the nucleomorph a unique model for eukaryotic genomics.**

Chlorarachniophytes are unicellular, amoeboflagellate protists with unusual plastids (1) derived from the endosymbiosis of a eukaryotic alga through a process termed secondary endosymbiosis (2). Phylogenetic analyses indicate the host is related to filose amoebae (3) while the endosymbiont is most probably related to a green alga (4, 5), with *Chlorella* and *Chlamydomonas* being the closest relatives characterized thus far (6). Thus, a phagotrophic host is envisaged to have engulfed a green algal cell and retained it as a permanent endosymbiont (7). The endosymbiont is much reduced and has lost most cellular structures such as the cytoskeleton, wall, endomembrane system, and mitochondria. Only the plastid, the plasma membrane, a small nucleus (known as the nucleomorph), and a very few ribosomes in the vestigial cytoplasm are identifiable (1, 4). In addition to the two chloroplast envelopes and the remnant plasma membrane, a fourth membrane—believed to derive from the food vacuolar membrane of the host cell (see ref. 7 for a review)—also surrounds the plastid/endosymbiont complex. Essentially, the chlorarachniophyte is a cell within a cell.

The endosymbiont's nuclear genome, the nucleomorph, is very much reduced with a haploid genome size of only 380 kb contained in three small chromosomes of sizes 145, 140, and 95 kb (2). Despite their tiny size, these chromosomes are nonetheless eukaryotic in nature, being linear in conformation and furnished with eukaryote-like telomere motifs at their termini (8). The nucleomorph telomere motif ($TCTAGGG_n$) is different to that of the host nucleus chromosomes ($TTAGGG_n$) (8), which is consistent with the nucleomorph being the genome of a phylogenetically unrelated endosymbiont.

Previously, chlorarachniophyte nucleomorph DNA has only been shown to encode eukaryotic rRNAs that are incorporated into the ribosomes in the vestigial cytoplasm surrounding the nucleomorph (2). We earlier speculated the nucleomorph's *raison d'être* is to provide proteins for the maintenance of the chloroplast (2, 7). To synthesize these chloroplast proteins, the nucleomorph may also have to maintain genes that encode expression, translation and self-replication machinery (2, 7). The nucleomorph's diminutive size alone suggests that many genes have been lost during the establishment of the endosymbiosis, either because these genes were no longer required, were transferred to the nucleus from the nucleomorph, or even had their functions entirely assumed by nuclear-encoded equivalents (7). The nucleomorph may thus represent a pared-down eukaryotic nucleus with only a basic set of self-replicating and protein synthesis genes. Characterization of nucleomorph genomic information content should not only reveal why the nucleomorph persists, but could also provide valuable insights into the core housekeeping functions of the eukaryotic cell. This paper reports the initial characterization and unusual organization of several nucleomorph genes involved in splicing of mRNAs and the translation process.

MATERIALS AND METHODS

Cells, Gels, Blots, and Sequencing. Chlorarachniophyte strain CCMP 621 was grown and prepared for pulsed field gel electrophoresis, Northern analysis, and blotting as described (8, 9). A plasmid containing a 13.2-kb *Spe*I fragment from nucleomorph chromosome III (8) was sequenced on an automated sequencer (model 373a; Applied Biosystems) by using a Dye terminator sequencing kit (Applied Biosystems). Complete, double-stranded sequence was obtained through a combination of the Erase-a-Base system (Promega) and primer walking. Contiguous traces were assembled and edited by using SEQUENCHER version 2.1 (Gene Codes, Ann Arbor, MI). Sequence data were analyzed by FASTA and BLAST searches (10, 11) of GenBank. Initial gene identification of nucleomorph open reading frames (ORFs) by FASTA and BLAST were substantiated by multiple alignments of each nucleomorph gene to putative homologues from other organisms using CLUSTALW and MACAW (Multiple Alignment Construction & Analysis Workbench).

Reverse Transcription–PCR (RT–PCR) cDNA Synthesis. Single-stranded cDNAs of nucleomorph genes were synthe-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RT–PCR, reverse transcription–PCR; snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein.

*To whom reprint requests should be addressed. e-mail: U6068083@ucsvc.ucs.unimelb.edu.au.

sized by using avian myeloblastosis virus (AMV) reverse transcriptase (Promega) with total chlorarachniophyte RNA as a template primed with oligonucleotides specific for the 3′ ends of each gene. Reactions were performed in a volume of 20 μl containing 2 pmol of oligonucleotide, 1 μg RNA, 8 units AMV reverse transcriptase, 1 mM dNTPs, 20 units RNAsin (Promega), 5 mM $MgCl_2$, and 1× standard PCR buffer (no $MgCl_2$; Perkin–Elmer) at 42°C for 1 hr. PCR incorporating oligonucleotides specific for the 5′ and 3′ ends of nucleomorph genes was then used to amplify nucleomorph cDNAs from single-stranded cDNA template (12).

The synthesis of cDNAs of the 3′ untranslated regions of chlorarachniophyte genes was primed with an oligo deoxythymidine oligo(dT)-adaptor primer and reverse transcriptase reactions were performed as above. PCR experiments incorporating one gene-specific oligonucleotide and an adaptor primer specific for the oligo(dT) end were used to amplify the 3′ untranslated regions of nucleomorph genes as per the method of Frohman (12). RT–PCR products were cloned in pGEM-T (Promega) and sequenced as above.

***In Situ* Hybridization and Immunolabeling.** Cells for *in situ* hybridization were fixed and embedded as described (2). Biotinylated sense and antisense RNA probes were synthesized and hybridized to ultra-thin sections at 60°C in a buffer containing 50% formamide and 0.1× standard saline citrate (2). Immunolabeling experiments were performed as per ref. 13 with a mouse monoclonal antibody (clone K121; ref. 14) specific for the 2,2,7-trimethylguanosine cap structures of spliceosomal RNAs (Oncogene Science) on cells prepared as above.

RESULTS AND DISCUSSION

Nucleotide sequence from a 13.2-kb *Spe*I fragment from nucleomorph chromosome III (Fig. 1*A*, S^2–S^3) (8) was determined (accession no. U58510) and submitted for data base searches. In addition to the rRNA genes already identified on this clone (8), the searches identified one putative gene for a spliceosomal structural RNA (U6 snRNA) (Fig. 1*A*). Nucleomorph U6 was 70–80% identical to the U6 snRNAs of other eukaryotes and contained all highly conserved sequence elements.

Two of the putative nucleomorph protein-encoding genes identified by data base searches (Fig. 1*A*), possessed some identity to highly conserved proteins (S4 and S13) of the cytoplasmic-type small ribosomal subunit (nucleomorph S13 protein was 42% identical to the homologue from pea, and nucleomorph S4 was 26% identical to potato S4). Other nucleomorph genes belong to less conserved families, and gene identification should be considered tentative until more genes from other organisms are lodged in the data base. These putative protein genes consist of (*i*) a shared core protein (protein E) found in several different small nuclear ribonucleoprotein (snRNP) particles involved in splicing, (*ii*) the catalytic subunit of the ATP-dependent Clp protease (ClpP), (*iii*) a gene similar to yeast splicing factor PRP6 (GenBank accession no. X53465), (*iv*) a subunit of RNA polymerase (human homologue, GenBank accession no. Z49199), and (*v*) the 3′ portion of an RNA helicase. Those genes for which expression is demonstrated are depicted in Fig. 1*A* as filled boxes; open boxes are putative genes.

Expression of U6 and Other snRNAs in the Nucleomorph. Probing of a Northern blot of total chlorarachniophyte RNA with a 233-bp *Hinc*II/*Nco*I fragment containing the putative U6 gene labeled abundant transcripts of ≈100 nt (data not shown). As U6 genes are highly conserved, cross-hybridization between nuclear and nucleomorph transcripts (if they exist) is likely; therefore, we undertook *in situ* hybridization studies to localize U6 transcripts at the subcellular level.

The putative nucleomorph U6 gene was used to prepare biotinylated antisense RNA probes that were hybridized to ultrathin sections and detected with colloidal gold markers

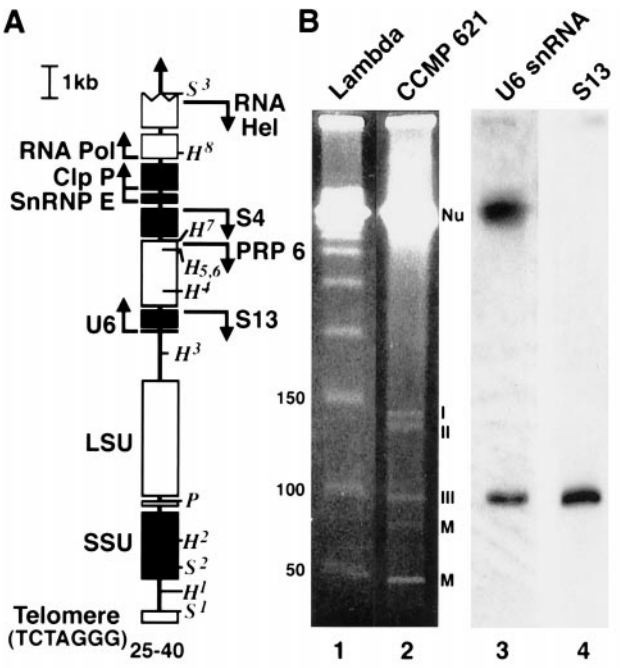

FIG. 1. Southern analysis and mapping of nucleomorph DNA. (*A*) Map of one end of nucleomorph chromosome III. Note that arrows indicate direction of gene transcription only and are not indicative of transcript length. (*B*) Ethidium bromide-stained pulsed-field agarose gel containing λ DNA molecular weight markers (in kilobases) (lane 1) and total chlorarachniophyte strain CCMP 621 DNA (lane 2). An autoradiograph of a Southern blot of total chlorarachniophyte DNA probed with U6 small nuclear RNA (snRNA) (lane 3) and S13 ribosomal protein (lane 4) genes. Nu, nuclear chromosomes; I, II, and III, nucleomorph chromosomes I, II and III; M, mitochondrial chromosomes; LSU and SSU, large and small subunit ribosomal RNA genes, respectively; and H, P, and S, *Hind*III, *Pst*I, and *Spe*I sites.

(Fig. 2*A*). The nucleomorph was strongly labeled with an average density of 65.5 gold particles/μm^2 over the nucleomorph (n = 9; range, 39.8–114.3 gold particles/μm^2). The probe also labeled the host nucleus, but to a lesser extent (average density of 21.2 gold particles/μm^2; n = 8; range, 1.7–50.0 gold particles/μm^2). Other structures in the cell were not labeled (Fig. 2*A*) and the background labeling averaged only 0.41 gold particles/μm^2 (n = 11; range, 0.1–0.9 gold particles/μm^2). Control experiments using a sense probe did not label any structures.

To demonstrate that nucleomorph U6 transcripts could be encoded by the U6 gene depicted in Fig. 1*A*, we examined nucleomorph DNA for other U6 genes. Probing of a pulsed-field gel blot with U6 demonstrates that U6 is confined to chromosome III in the nucleomorph (Fig. 1*B*, lane 3). Further, probing of gel-isolated nucleomorph chromosome III DNA (8) digested with *Hind*III only labeled a 1.8-kb fragment (data not shown), suggesting that chromosome III, and therefore the entire nucleomorph, only carries the single U6 gene depicted on the 1.8-kb *Hind*III fragment (Fig. 1*A*, H^3–H^4).

The *in situ* hybridization data strongly suggest that the U6 transcripts in the nucleomorph are nucleomorph-encoded, but, since the nuclear gene is very similar, we cannot exclude the possibility that nuclear-encoded U6 transcripts accumulate in the nucleomorph.

U6 is one of a family of U-rich snRNAs that are components of ribonucleoprotein structures known as spliceosomes involved in transcript processing (15). Presence of U6 transcripts in the nucleomorph implies presence of a spliceosome. We used an antibody directed against spliceosomes to immunolocalize these structures in chlorarachniophytes (Fig. 2*B*). The antibody recognizes the trimethylguanosine cap present on

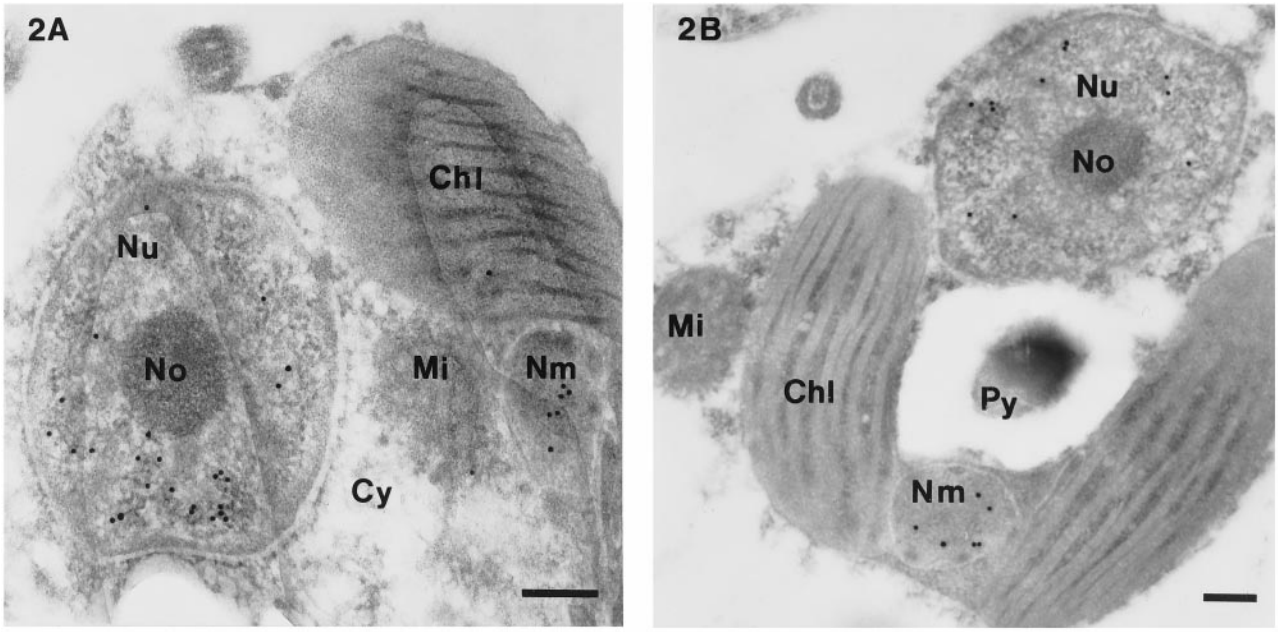

FIG. 2. Intracellular localization of spliceosomes. (*A*) Transmission electron micrograph of a chlorarachniophyte cell labeled by *in situ* hybridization with a probe for U6 snRNA transcripts. Colloidal gold particles are concentrated in the nucleoplasm of the host nucleus (Nu) and the nucleomorph (Nm). The nucleolus of the host nucleus (No), the chloroplast (Chl), mitochondrion (Mi), and the host cytoplasm (Cy) are not labeled. (*B*) Chlorarachniophyte cell labeled with immunogold using an antibody to the trimethyl guanosine cap of spliceosomal snRNAs. The nucleoplasm of the host nucleus and the nucleomorph are both heavily labeled while other components within the cell including the pyrenoid (Py) are not. (Bars = 200 nm)

several U snRNAs (U1–U5, U7, U8, U11, and U12) but absent from U6 (16). The antibody clearly labeled the nucleomorph (average density of 33.4 gold particles/μm^2; $n = 7$; range 17–41.4 gold particles/μm^2) as well as the host nucleus (average density of 20.3 gold particles/μm^2; $n = 6$; range, 3.8–45.8 gold particles/μm^2) (Fig. 2*B*). There was no labeling of other structures within the algal cells, and background label averaged only 0.05 gold particles/μm^2. Control experiments omitting the primary antibody did not label any structures. These data suggest that both genomes contain other U snRNAs in addition to U6, further supporting the presence of a spliceosomal apparatus in both nucleomorph and nucleus.

Expression of Nucleomorph Genes. Translation of nucleomorph chromosome III sequence revealed very few large ORFs. Nevertheless, data base searches indicated matches to several known proteins, but frame shifts were apparent in each match. The reading frames appeared to be interrupted by short, AT-rich stretches of DNA. Intriguingly, these stretches of DNA possessed similarity to spliceosomal introns in that they began with the dinucleotide GT and ended with the dinucleotide AG, but they were only 18, 19, or 20 bp in length. When these very short sequences were removed from the putative gene sequences, large ORFs with far more robust matches to known genes were created. We reasoned that the intervening sequences were likely introns, and we decided to characterize cDNA clones of the putative S13, S4, snRNP E, and clpP genes to demonstrate expression and verify removal of the putative introns *in vivo*.

Expression of putative nucleomorph genes was confirmed by isolation of cDNA clones by using RT–PCR (Fig. 3). Probing of a pulsed-field electrophoresis gel with the S13 cDNA indicated that this gene is unique to nucleomorph chromosome III (Fig. 1*B*, lane 4). Probing poly(A)-selected RNA from chlorarachniophyte cells with the S13 cDNA identified a low abundance 1100-nt transcript (data not shown). The four protein genes characterized thus far appear to use AUG as a start codon, but our cDNAs (derived from RT–PCR) do not all include the 5′ end of the gene, and the start codon is inferred in most cases from genomic sequence (Fig. 3). Stop codons are inferred on the basis of comparative alignments with known proteins. Stop codon UAA is utilized by S13 and snRNP E, UAG by S4, and UGA by clpP (Fig. 3). Codon usage shows a strong A+U bias. The messengers for S4 and S13 are polyadenylylated ≈50 nt downstream of the motif AUUAAA (Fig. 3), a variant of the standard AAUAAA polyadenylylation signal (17). Polyadenylylation of snRNP E and clpP genes is still under investigation.

Intriguingly, the 3′ untranslated region of the S13 gene transcript completely overlaps the U6 gene, which is encoded on the opposite strand (Figs. 1*A* and 3*A*). Overlapping genes occur sparingly in eukaryotes (18), and, to our knowledge, ours is the first example of a mRNA that contains an antisense copy of an abundant structural RNA like U6. How, or if, the messenger can function in the presence of an apparent abundance of U6 transcripts (Fig. 2*A*) remains unclear.

Possible CAAT and TATA boxes are apparent in appropriate locations upstream of the S13 gene (Fig. 3*A*). Similar motifs were also observed upstream of the other two genes (Fig. 3 *B* and *C*). As the S4 and snRNP E genes are transcribed divergently with start codons separated by only 54 bases (Fig. 1*A*), the putative CAAT boxes for these two genes lie within the coding/intron domain of the adjacent gene.

All 12 predicted introns were absent from cDNAs (Fig. 3), thus proving their authenticity. A number of transcript processing intermediates were also encountered among the cDNAs of each gene, and a selection of S4 intermediates is presented in Fig. 4. From a total of 25 S4 cDNAs sequenced at random, 14 were fully processed and 9 were partially processed (Fig. 4), perhaps suggesting nucleomorph transcript processing is inefficient. As these introns introduce stop codons in all three reading frames, it is unlikely that these partially processed transcripts produce alternative gene products. Inspection of the processing intermediates (Fig. 4) suggests nucleomorph introns are not removed in any orderly fashion. Similar observations have been made for ovalbumin transcript processing in humans (19).

snRNP E and clpP Genes Are Cotranscribed. The snRNP E and clpP genes are encoded on the same DNA strand and

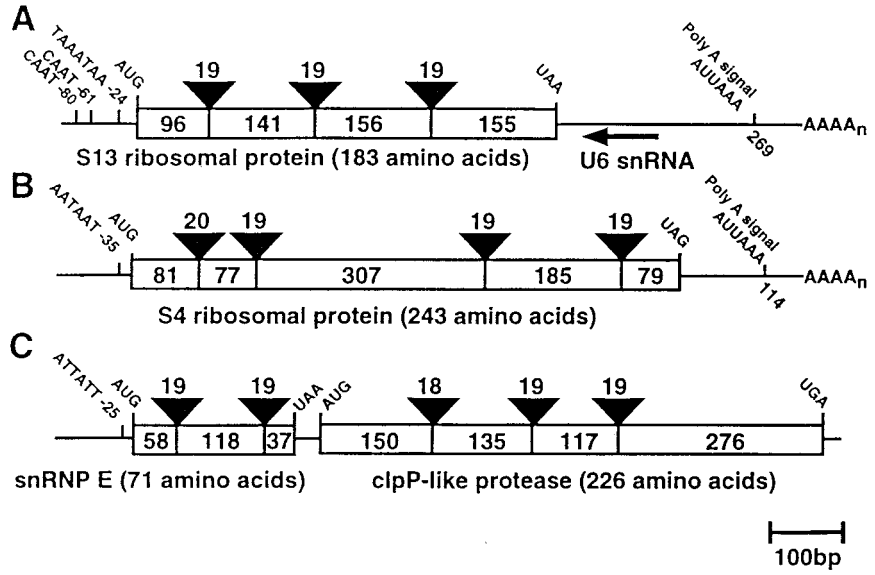

FIG. 3. Gene maps for nucleomorph S13 (*A*), S4 (*B*), and snRNP E/clpP (*C*) inferred from genomic and cDNA clone sequences. The maps assume a standard genetic code. Exons (size in bp) are indicated by boxes, introns by triangles, and untranslated regions by lines. Putative start and stop codons, 5′ transcription signals, and polyadenylylation signals are shown. The Poly(A) tails shown were determined through RT–PCR using oligo(dT) priming.

separated by only 49 bp (Fig. 1*A*). This proximity, and the absence of a polyadenylylation signal downstream of snRNP E, prompted us to test if snRNP E and clpP might be cotranscribed. Cotranscription was confirmed by RT–PCR amplification between the start of the snRNP E gene and the end of the clpP gene, which yielded a cDNA devoid of introns and containing both ORFs (Fig. 3*C*). Since the snRNP E and clpP gene products are not known to be functionally related, the dicistronic transcription is unlikely to represent an ancient operon in the classic prokaryotic sense. Rather, it appears more likely to have arisen *de novo* during genome compression, as has been suggested for the polycistronic gene clusters in *Caenorhabditis elegans* (20). The possibility that other nucleomorph genes could also be cotranscribed is currently being examined (e.g., clpP and RNA polymerase subunit).

Translation of nucleomorph snRNP E and clpP has not yet been demonstrated. We are currently examining whether the dicistronic message is translated in tandem (21) or undergoes further processing, perhaps to insinuate a trans-spliced leader ahead of the downstream ORF, as shown for the polycistronic messages of *C. elegans* (20).

Nucleomorph Introns Are the Smallest Known Spliceosomal Introns. Several lines of evidence suggest that the 18-, 19-, and 20-nt introns of nucleomorph protein genes are spliceosomal-type introns. First and foremost, all the introns have the consensus spliceosomal boundaries ($^{5'}$GT. . . . AG$^{3'}$) (Fig. 5). Second, the nucleomorph encodes and expresses both U6 snRNA and the protein snRNP E, two components of the spliceosome. Moreover, a putative gene for the homologue of the yeast splicing factor PRP6 (15) has also been identified in nucleomorph DNA (Fig. 1*A*). Finally, antibody labeling suggests the presence of abundant trimethyl guanosine-capped RNAs, such as occur in spliceosomes, in the nucleomorph. These data strongly suggest that the tiny introns of the nucleomorph are spliceosomal introns. Previously, the smallest known spliceosomal intron was a 20-nt intron in the ciliate *Paramecium* (22).

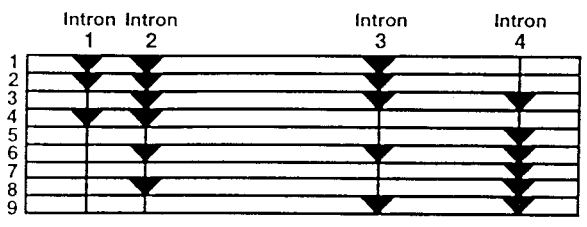

FIG. 4. Maps of partially processed S4 ribosomal protein cDNAs. Unexcised introns are shown as triangles for each clone.

None of the nucleomorph intron positions are conserved in homologous genes characterized from other organisms, but the data are sparse, particularly with respect to the green algal lineage from which the endosymbiont is thought to derive (4–6). One nucleomorph intron (intron 2 in S13) is located one nucleotide upstream of intron 2 in the *Schizosaccharomyces pombe* S13 gene (23) and is perhaps an example of a "slid intron" (24).

Sequence information that determines the fidelity of intron removal and exon splicing is encoded within the intron and surrounding exonic sequences (15). Specifically, this information must establish the 5′ and 3′ exon/intron boundaries and the branch point for the formation a splicing intermediate called the lariat structure (15). The diminutive size of nucleomorph introns potentially offers great opportunity to characterize such informational elements, and to this end we have sought to identify conserved elements within nucleomorph

		Exon	Intron	Exon
SnRNP E	intron 1	5' GAAATTCAAA	**GTAAAATAATTCA-CAATAG**	TTCAATTAAA 3'
SnRNP E	intron 2	5' TGATATTCTA	**GTATGATCTACAT-TAATAG**	ATTGTTAGAG 3'
S4 ribo	intron 1	5' TGGCATTTTC	**GTAAATATTAATTAAAATAG**	AACACAAAAA 3'
S4 ribo	intron 2	5' AATTTTTCCA	**GTAATAATAACAA-AAATAG**	TTTGGCTGAT 3'
S4 ribo	intron 3	5' GAAAATTAAA	**GTAATATATTGAA-TTTCAG**	GTTGAAGATT 3'
S4 ribo	intron 4	5' ACGGTAAAAA	**GTAAAAATAGATT-TACCAG**	ACTAACTAAA 3'
S13ribo	intron 1	5' TTCCCAAAAC	**GTAATTCTTAAAG-AAATAG**	CTAATTAAAA 3'
S13ribo	intron 2	5' CAGACTTAAG	**GTAATAATTATTA-AAAAAG**	GGCCTGGTGC 3'
S13ribo	intron 3	5' ATATTATAAA	**GTATTATCTCAAA-GAATAG**	CGAATTTTTC 3'
Clp P	intron 1	5' CCAAAAAATG	**GTAAAACACAAT--GAATAG**	TCTAAGACTA 3'
Clp P	intron 2	5' TGGATCTAAT	**GTATAAATGTAAA-AACTAG**	ATAACAGATT 3'
Clp P	intron 3	5' TGGAAAGAAA	**GTATGTATATAAA-ATAAAG**	GATATGAGAG 3'
			NNNNHNHHN **GTAWDWHHNNNHD NWHHAG**	NNHVNNDNDN

FIG. 5. Alignment of nucleomorph introns (boldface type) and bordering 10 bases of 5′ and 3′ exon sequence. To align 18- and 19-base introns to the 20-base intron, gaps were arbitrarily inserted into the shorter introns. A strict consensus sequence of the alignment in IUPAC (International Union of Pure and Applied Chemistry) format is shown below. Gene name and intron number are indicated on the left.

introns and exons. An alignment of 12 nucleomorph introns bordered by 10 bases of 5′ and 3′ exon sequence and a strict consensus are presented in Fig. 5. Apart from the GT dinucleotide at the 5′ end of the intron and the AG dinucleotide at the 3′ end, the only other absolutely conserved nucleotide is an A occupying the third intron position (Fig. 5). The only other patterns that could be identified are the paucity of G residues in 5′ exons (Fig. 5), particularly at the upstream exon junction site (which is predominantly a G in other eukaryotes), and the general A/T richness of intron sequence (80%) compared with cDNA sequence (A/T content of 73%). No polypyrimidine tract is present in any intron (Fig. 5). No conserved A residue offered itself as a branch site for a lariat structure (15). Identification of further nucleomorph intron sequences and other components of the spliceosome will hopefully establish if intron excision and exon splicing follows standard rules.

Evolution of Nucleomorph Introns. In a total of 2168 bases of nucleomorph coding sequence we have identified 12 introns, resulting in an average density 5.5 introns/kb, which is remarkably high (see ref. 25). It is not known if these introns were present in the endosymbiont genes prior to endosymbiosis or if they proliferated postendosymbiosis, as appears to have happened with group II introns in mitochondria and chloroplasts (25, 26). We note that *Chlamydomonas*—the nearest living relative of the endosymbiont (6) for which sufficient intron data are available—has a comparable 5.8 introns/kb in the nuclear genome (J. Logsdon and J. Palmer, personal communication). However, since intron positions for S4, S13, snRNP E, and nuclear-encoded clpP (if one exists) of *Chlamydomonas* and other green algae are yet to be determined, it is impossible to decide if this equivalence of intron density points to a preendosymbiotic origin for nucleomorph introns or is merely coincidental.

Nucleomorph introns are of particular interest to the "introns-early" versus "introns-late" hypotheses (summarized in ref. 27). Introns of the nucleomorph could be primordially small, in which case they could tell us a great deal about the primitive state of spliceosomal introns. Alternatively, they could be secondarily reduced to a minimum spliceable size during extreme compression of the endosymbiont genome. The reduction, but retention, of so many introns would be a repudiation of the "introns-early" hypothesis which posits frequent intron loss during genomic streamlining (28, 29).

ClpP, a Putative Nucleomorph-Encoded Plastid Protein. Nucleomorph clpP is a possible candidate for a nucleomorph-encoded plastid protein. ClpP occurs in all plastids examined to date (30) and is usually encoded by the plastid genome (31). Whether or not the chlorarachniophyte plastid contains ClpP has not been determined, but the nucleomorph clpP gene, which has an apparent N-terminal extension compared with plastid-encoded clpPs, might encode a protein targeted into the plastid. Alternatively, it could be a cytoplasmic form of the protease occurring in the endosymbiont's nucleocytoplasmic compartment. Demonstration of clpP translation, followed by subcellular localization, will be necessary to resolve the function of nucleomorph clpP.

CONCLUDING REMARKS

Many photosynthetic eukaryotes have acquired their chloroplasts by endosymbioses of eukaryotic algae (7). The success of this evolutionary strategy is evidenced by the diversity of eukaryotes that possess second-hand chloroplasts. To begin unraveling the process of endosymbiosis, we have initiated studies of the reduced nuclear genome—the nucleomorph—of chlorarachniophyte endosymbionts. We have demonstrated the nucleomorph genome is typically eukaryotic but remarkably compact. It contains a high density of genes (average intergenic spacers only 65 bp), some of which overlap, and some of which are cotranscribed. We have also discovered numerous spliceosomal introns that are smaller than any previously known. The picture is undoubtedly one of extreme pressure for miniaturization and is reminiscent of the situation in organelle genomes. The nucleomorph thus presents itself as an interesting and tractable model for investigating eukaryotic genome architecture.

Our hypothesized *raison d'être* for the chlorarachniophyte nucleomorph is a vestigial nuclear genome existing merely to express some proteins essential to the function of the endosymbiotic plastid (2). Characterization of several nucleomorph genes encoding housekeeping functions such as transcript processing and translation is consistent with this hypothesis. To provide a relatively few plastid proteins, the nucleomorph may need to encode an ensemble of transcription and translation genes. Although we have yet to identify nucleomorph-encoded plastid proteins, nucleomorph ClpP is a possible candidate.

We thank Jeff Palmer, John Logsdon, Arlin Stoltzfus and Sue Douglas for advice and J. Lafontaine for antibodies. P.G. is supported by an Australian Postgraduate Award and G.M. is the recipient of an Australian Research Council Senior Research Fellowship. The project is supported by the Australian Research Council. G.M. is grateful for the hospitality of the Institute for Marine Biosciences, National Research Council of Canada, who hosted a sabbatical visit.

1. Hibberd, D. J. & Norris, R. E. (1984) *J. Phycol.* **20,** 310–330.
2. McFadden, G. I., Gilson, P. R., Hofmann, C. J. B., Adcock, G. J. & Maier, U. G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3690–3694.
3. Bhattacharya, D., Helmchen, T. & Melkonian, M. (1995) *J. Eukaryotic Microbiol.* **42,** 65–69.
4. Ludwig, M. & Gibbs, S. P. (1989) *J. Phycol.* **25,** 385–404.
5. McFadden, G. I., Gilson, P. R. & Waller, R. F. (1995) *Arch. Protistenk.* **145,** 231–239.
6. Van der Peer, Y. Rensing, S. Maier, U.-G. & De Wachter, R. (1995) *Proc. Natl. Acad. Sci. USA* **93,** 7734–7738.
7. McFadden, G. I. & Gilson, P. R. (1995) *Trends Ecol. Evol.* **10,** 12–17.
8. Gilson, P. R. & McFadden, G. I. (1995) *Chromosoma* **103,** 635–641.
9. Gilson, P. R., Waller, R. F. & McFadden, G. I. (1995) *J. Eukaryotic Microbiol.* **42,** 696–701.
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
11. Gish, W. & States, D. J. (1993) *Nat. Genet.* **3,** 266–272.
12. Frohman, M. A. (1990) in *PCR Protocols: A Guide to Methods and Applications*, eds. Innis, M. A., Gelfland, D. H., Sninsky, J. J. & White, T. J. (Academic, San Diego, CA), pp. 28–38.
13. Chamberland, H. & Lafontaine, J. G. (1993) *Chromosoma* **102,** 220–226.
14. Krainer, A. R. (1988) *Nucleic Acids Res.* **16,** 9415–9428.
15. Moore M. J., Query, C. C. & Sharp, P. A. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 303–357.
16. Baserga, S. J. & Steitz, J. A. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 359–381.
17. Wickens, M. (1990) *Trends Biochem. Sci.* **15,** 277–281.
18. Merino, E., Balbás, P., Puente, J. L. & Bolívar, F. (1994) *Nucleic Acids Res.* **22,** 1903–1908.
19. Roop, D. R., Nordstrom, J. L., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. (1978) *Cell* **15,** 671–685.
20. Zorio, D. A. R., Cheng, N. N., Blumethal, T. & Spieth, J. (1994) *Nature (London)* **372,** 270–272.
21. Kozak, M. (1992) *Crit. Rev. Biochem. Mol. Biol.* **27,** 385–402.
22. Russell, C. B., Fraga, D. & Hinrichsen, R. D. (1994) *Nucleic Acids Res.* **22,** 1221–1225.
23. Marks, J. & Simanis, V. (1992) *Nucleic Acids Res.* **20,** 4094–4096.
24. Craik, C. S., Rutter, W. J. & Fletterick R. (1983) *Science* **220,** 1125–1129.
25. Palmer, J. D. & Logsdon, J. M. (1991) *Curr. Opin. Genet. Dev.* **1,** 470–477.
26. Grivell, L. A. (1994) *Curr. Biol.* **4,** 161–164.
27. Hurst, L. (1994) *Nature (London)* **337,** 381–382.
28. Doolittle, W. F. (1991) *Curr. Biol.* **1,** 145–146.

29. Tittiger, C., Whyard, S. & Wlaker, V. K. (1993) *Nature (London)* **361,** 470–472.
30. Gottesman, S., Squires, C., Pichersky, E., Carrington, M., Hobbs, M., Mattick, J. S., Dalrymple, B., Kuramitsu, H., Shoroza, T., Foster, T., Clark, W. P., Ross, B., Squires, C. L. & Maurizi, M. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3513–3517.
31. Huang, C., Wang, S., Chen, L., Lemieux, C., Otis, C., Turmel, M. & Liu, X. Q. (1994) *Mol. Gen. Genet.* **244,** 151–159.