Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Organization and Expression of Algal (Chlamydomonas reinhardtii) Mitochondrial DNA

M. W. Gray and P. H. Boer

Phil. Trans. R. Soc. Lond. B 1988 **319**, 135-147
doi: 10.1098/rstb.1988.0038

References Article cited in:
http://rstb.royalsocietypublishing.org/content/319/1193/135#related-urls

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: **http://rstb.royalsocietypublishing.org/subscriptions**

This journal is © 1988 The Royal Society

Downloaded from rstb.royalsocietypublishing.org

Phil. Trans. R. Soc. Lond. B 319, 135–147 (1988)
Printed in Great Britain

Organization and expression of algal (*Chlamydomonas reinhardtii*) mitochondrial DNA

By M. W. Gray and P. H. Boer†

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

The mitochondrial genome of *Chlamydomonas reinhardtii*, a unicellular green alga, is a linear 15.8 kilobase pair (kbp) molecule. In gene arrangement and mode of expression, as well as in size, it differs radically from the large (200–2400 kbp) mitochondrial genomes of higher plants. Heterologous hybridization experiments and nucleotide sequence analysis have revealed that *C. reinhardtii* mitochondrial DNA (mtDNA) is a compactly organized genome specifying at least eight proteins, a minimum of three transfer RNAs, and large subunit (LS) and small subunit (SS) ribosomal RNAs. Both strands of the mtDNA encode genetic information, with genes organized into perhaps a single transcriptional unit on each strand. Stable transcripts have been identified by Northern hybridization analysis, and transcript termini have been mapped by primer extension and S1 nuclease protection experiments. The results suggest that mature RNAs, which virtually saturate the genome, are generated by precise endonucleolytic cleavage of long precursors, with specific motifs (both primary sequence and secondary structure) implicated as processing signals. Codon usage in *C. reinhardtii* mitochondria is highly biased, with eight codons entirely absent from all protein-coding genes; however, even though codon usage is restricted, it appears that *C. reinhardtii* mtDNA cannot encode the minimum number of tRNAs needed to support mitochondrial protein synthesis. The most striking feature of *C. reinhardtii* mtDNA is the division of SS and LS rRNA genes into a number of separate subgenic coding segments ('modules') that are interspersed with one another and with protein-coding and tRNA genes. We have identified abundant small RNAs, transcribed from these modules, that approximate to the latter in size. This indicates that splicing of rRNA 'pieces' does not occur in this system. Rather, the mature rRNAs apparently exist and function as non-covalent complexes of small RNAs (four in SS rRNA, at least eight in LS rRNA), held together by intermolecular base pairing. These complexes contain all the conserved elements of the minimal secondary structures that define the functional core of conventional LS and SS rRNAs.

1. Mitochondrial DNA in algae and plants

Higher plants have the largest and structurally most complex mitochondrial genome known, with estimated genome sizes ranging from 200 to 2400 kilobase pairs (kbp) (Pring & Lonsdale 1985). Mitochondrial DNA (mtDNA) from a number of angiosperms, both monocotyledons and dicotyledons, has been studied in detail; however, little or nothing is known about mitochondrial DNA in non-angiosperm members of the Kingdom Plantae, or in those unicellular, photosynthetic eukaryotes (the Chlorophyceae) with which plants are thought to share a recent common ancestor (Pickett-Heaps 1975; Ragan & Chapman 1978). In view of the markedly different patterns of mitochondrial genome arrangement and expression in the major groups of eukaryotes studied to date (Gray 1982; Sederoff 1984), and our interest in

† Present address: Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada K1H 8M5.

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to
Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. www.jstor.org

Downloaded from rstb.royalsocietypublishing.org

tracing the evolutionary history of mtDNA within the lineage leading to higher plants, we have undertaken a detailed study of the mtDNA of *Chlamydomonas reinhardtii*, a unicellular green alga. We particularly wished to know whether algal mtDNA and mitochondrial genes would show a specific resemblance to their counterparts in higher plants.

2. The mitochondrial genome of *Chlamydomonas reinhardtii*

Our interest in *C. reinhardtii* mtDNA was stimulated by early reports of its relatively small size (Ryan *et al.* 1978; Grant & Chiang 1980), which contrasts sharply with the size of a typical plant mitochondrial genome. The linear, 16 kbp mitochondrial genome of *C. reinhardtii* is close to animal mtDNA in size, prompting the question of whether it shows a similar, economical arrangement of genetic information (Clayton 1984; Attardi 1985). Because it had not been formally established that the 16 kbp DNA species found in *C. reinhardtii* mitochondria was, in fact, the functional equivalent of mtDNA in other eukaryotes, our initial work (Boer *et al.* 1985*a*) used heterologous hybridization and DNA sequence analysis to demonstrate that genes for repiratory chain proteins and ribosomal RNA are indeed encoded in this DNA.

In the course of this work, a continuous internal stretch of 12.4 kbp, accounting for almost 80% of the estimated 15.8 kbp mtDNA, was cloned and mapped; only the leftward 2.8 kbp and rightward 0.6 kbp have not yet been cloned and so remain to be investigated. The complete sequence of the interior 12.4 kbp has now been determined (Boer *et al.* 1985*a*; Boer

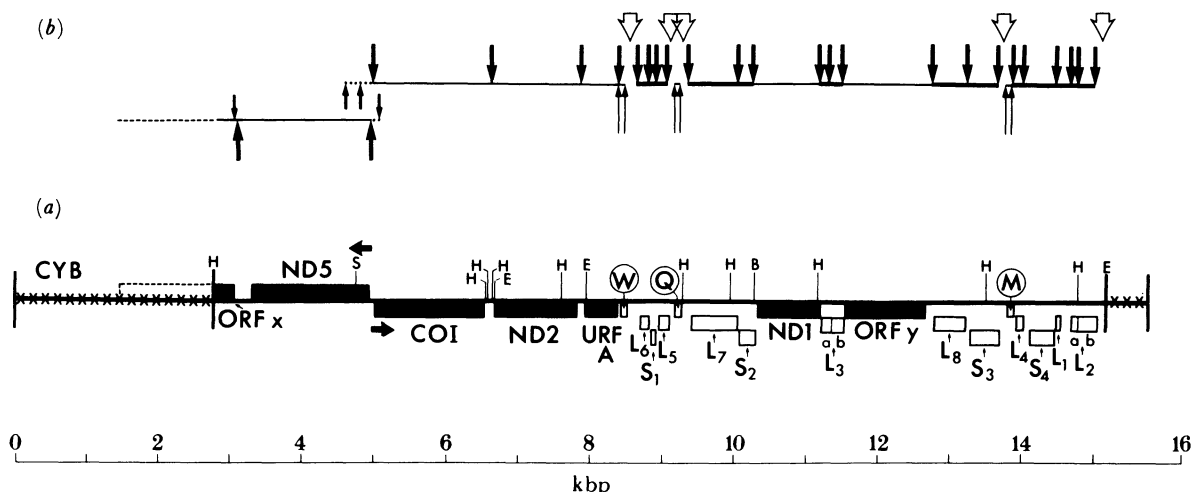

FIGURE 1. (*a*) Genetic map of the 15.8 kbp mtDNA of *Chlamydomonas reinhardtii*. Crosses mark uncloned terminal regions. Solid rectangles indicate protein-coding genes: *cyb*, apocytochrome *b*; *cox*I, cytochrome oxidase subunit I; *nd*1, *nd*2 and *nd*5, subunits of NADH dehydrogenase; *urf*A, unassigned reading frame A; *orf*s x and y, unidentified open reading frames x and y. The dashed line indicates the extent of the *orf*x gene, as estimated from the size of its transcript. Circled letters denote transfer RNA genes: W, tryptophan; Q, glutamine; M, methionine (elongator). Open rectangles indicate small subunit (S) and large subunit (L) ribosomal RNA-coding segments ('modules'), numbered in the 5'→3' order in which the homologous segments occur in *E. coli* 16*S* and 23*S* rRNAs, respectively (segments L_{2b} and L_{3a} have not been identified with particular portions of *E. coli* 23*S* rRNA, but give rise to abundant transcripts). Restriction sites are: H, *Hin*dIII; S, *Sal*I; E, *Eco*RI; B, *Bam*HI. Horizontal arrows indicate direction of transcription. (*b*) Transcript map of *C. reinhardtii* mtDNA. Low-abundance and high-abundance transcripts are denoted by thin and thick lines, respectively. Vertical thick arrows indicate endonucleolytic processing sites as determined by primer extension and S1 nuclease protection experiments to map transcript termini. Vertical thin arrows delineate tRNA-coding regions. Open vertical arrows mark apparent short spacers for which no transcripts have been identified.

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

& Gray 1986*a*, *b*; P. H. Boer, unpublished work), and a detailed transcript map has been assembled from Northern hybridization, S1 nuclease protection, and primer extension experiments (Boer & Gray 1986*a*; P. H. Boer, unpublished work). The results are summarized in figure 1, which shows that *C. reinhardtii* mtDNA, like that of animals (and unlike that of higher plants), is a compactly organized and expressed genome. Our work has utilized the CW-15-2 strain of *C. reinhardtii* (No. 277, Duke University Culture Collection). The sequence of part of the mtDNA of CW-15 strains CCAP 11/32 (Pratje *et al.* 1984; Vahrenholz *et al.* 1985) and CC406 (Kück & Neuhaus 1986) has also been reported, with identification of the same genes in the same arrangement between the *Hin*dIII site at 2.8 kbp and the *Eco*RI site at 8 kbp (figure 1). However, some differences between the CCAP 11/32 results and our sequence for the same region have been noted (Boer & Gray 1986*b*).

3. Coding function of *C. reinhardtii* mitochondrial DNA

Table 1 lists the protein-coding genes identified to date in *C. reinhardtii* mtDNA. Of the 13 protein-coding genes found in human mtDNA, only 5 (*cox*I, *cyb*, *nd*1, *nd*2 and *nd*5) are present in the algal mtDNA (the *cyb* gene having been identified tentatively on the basis of heterologous hybridization with a yeast (*Kluyveromyces lactis*) *cyb* gene-specific probe (Boer *et al.* 1985*a*)). A sixth gene, *urf*A, is homologous with one found in the mtDNA of *Aspergillus nidulans* (Netzker *et al.* 1982) and *Neurospora crassa* (U. L. RajBhandary, personal communication); the latter has been identified as the homologue of the *nd*6 gene (Chomyn *et al.* 1986) of human mtDNA.

In addition to these known genes, *C. reinhardtii* mtDNA contains two open reading frames (ORFs x and y) that potentially encode proteins, and for which discrete transcripts have been

Table 1. Protein-coding function of mitochondrial DNA

(Symbols: +, gene or portion of gene identified by sequence analysis; (+), gene tentatively identified by heterologous hybridization; *, gene whose presence in wheat mtDNA has not yet been established, but which has been found in some other plant mtDNAs. Abbreviations for genes are listed in the legend to figure 1, except that *atp* specifies an ATPase subunit and *rps*13 is the homologue of *E. coli* ribosomal protein S13. Data for human mtDNA are from Chomyn *et al.* (1985, 1986); data for wheat mtDNA are from Bonen *et al.* (1984), Boer *et al.* (1985*b*), and L. Bonen (unpublished work).)

gene	human	*C. reinhardtii*	wheat
cyb	+	(+)	+
*cox*I	+	+	+
*cox*II	+	.	+
*cox*III	+	.	*
*nd*1	+	+	+
*nd*2	+	+	(+)
*nd*3	+	.	.
*nd*4	+	.	.
*nd*4L	+	.	.
*nd*5	+	+	+
*nd*6	+	+	.
*atp*6	+	.	+
*atp*8	+	.	.
*atp*9	−	.	+
*atp*α	−	.	*
*rps*13	−	.	+
*orf*x	−	+	.
*orf*y	−	+	.

Downloaded from rstb.royalsocietypublishing.org

detected (see §4). However, computer searches have so far failed to uncover any strong homologies between the derived amino-acid sequences of these ORFs and the sequences of known proteins, mitochondrial or otherwise. A notable feature of the putative ORF y protein is its lower content of hydrophobic amino acids (39%) relative to that of the assigned proteins encoded by *C. reinhardtii* mtDNA (46–53%; average 48%).

Although the absence of the other *nd* genes found in animal mtDNA is not particularly surprising (*nd* genes are completely lacking in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* mtDNA (Chomyn *et al.* 1985)), the absence of a gene for COXII is unprecedented, and there is only one other report of a possible 'missing' COXIII gene in mtDNA (that of *Trypanosoma brucei* (Benne 1985)). Also surprising is the apparent absence of any gene(s) encoding subunit(s) of the F_1 ATPase, although it is possible that additional, small proteins (such as ATPase subunits 8 or 9) may be encoded in the rightward uncloned 0.6 kbp region.

In the 80% of the *C. reinhardtii* mitochondrial genome sequenced to date, we have found only three tRNA genes (specific for Trp, Gln and Met). The $tRNA^{Met}$ sequence has the characteristics of an elongator rather than an initiator species. When arranged in the standard cloverleaf configuration, these tRNA sequences have all of the expected primary sequence and secondary structure characteristics of a normal tRNA. Although it is possible that additional tRNA genes are present in the uncloned terminal regions, their number cannot be large. The remaining portion of the *orf*x gene (which specifies a 1.6 kb transcript) should occupy at least 1.3 kbp of the leftward 2.8 kbp terminal region; this leaves about 1.5 kbp, just sufficient to accommodate a putative *cyb* gene (which in other mitochondrial systems encodes a protein of about 380 amino acids). The rightward 0.6 kbp terminal region could maximally encode 8 transfer RNAs of length 75 nucleotides, if these were tightly clustered. It is unlikely that genes specifying structurally peculiar tRNAs (see, for example, Wolstenholme *et al.* 1987) have been overlooked because (i) the three identified tRNA genes are not structurally abnormal, and (ii) within the internal 12.4 kbp region, very little of the sequence remains without an assigned function and/or associated transcript. The available data strongly suggest that *C. reinhardtii* mtDNA does not encode the minimum number of tRNAs (24) required for protein synthesis in a system using an expanded ('two out of three') codon recognition pattern (Attardi 1985; Breitenberger & RajBhandary 1985).

The way in which rRNA coding sequences are arrayed in *C. reinhardtii* mtDNA is both novel and bizarre. In this genome, large subunit (LS) and small subunit (SS) rRNA genes are each discontinuous, with subgenic coding segments (S and L 'modules') interspersed with one another and with protein-coding and tRNA genes throughout a 6.5 kbp stretch of the mtDNA (between 8.5 and 15 kbp on the map) (figure 1). The S and L modules represent regions of the mtDNA that are homologous with particular regions in *E. coli* 16*S* or 23*S* rRNAs, respectively, in primary sequence and/or secondary structure. As discussed in §7, the rRNA modules (4 in SS rRNA, 8–10 in LS rRNA) are able to form non-covalent complexes that contain those structural elements that constitute the functional core of conventional, covalently continuous SS and LS rRNAs.

Together, sequences encoding proteins, tRNAs, and rRNA 'modules' account for about 80% of the total sequenced length of *C. reinhardtii* mtDNA. Detailed transcript analysis, discussed in §4, suggests that this is a minimal estimate of genome utilization, because stable transcripts corresponding to these genes or gene segments virtually saturate the mtDNA.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

BIOLOGICAL SCIENCES THE ROYAL SOCIETY PHILOSOPHICAL TRANSACTIONS OF

4. Transcription of the *C. reinhardtii* mitochondrial genome

As figure 1 indicates, both strands of *C. reinhardtii* mtDNA are used to encode genetic information. A 'leftward transcriptional unit' (LTU) contains the *nd*5 and *orf*x genes, the latter continuing into the leftward uncloned terminal region; this LTU possibly includes the *cyb* gene, as well. All genes to the right of and including the *cox*I gene are transcribed in the same direction, and could conceivably constitute a single 'rightward transcriptional unit' (RTU) if they are transcriptionally as well as physically linked. Because many of the transcripts do abut one another as they map to the genome, they must be produced by precise endonucleolytic cleavage of larger precursors. We have previously documented transcriptional linkage of the *nd*5 and *orf*x genes (Boer & Gray 1986*a*), and have more recently obtained evidence for cotranscription of the following clusters: *cox*I–*nd*2–*urf*A–tRNA^{Trp}; L_6–S_1–L_5; L_7–S_2–*nd*1–L_3–*orf*y–L_8–S_3; and tRNA^{Met}–L_4–S_4–L_1–L_2 (P. H. Boer, unpublished work). Whether the cotranscripts corresponding to these clusters are in turn derived from a larger precursor covering the entire RTU is not known. Apparent short spacers (25–200 bp) are situated as follows: between the tRNA^{Trp} and L_6 coding regions; upstream and downstream of the tRNA^{Gln} gene; between the S_3 and tRNA^{Met} coding regions; and downstream of the L_2 module. In total, these spacers make up less than 6% of the sequenced 12.4 kbp portion of *C. reinhardtii* mtDNA, 94% being represented in stable transcripts coming from one or the other strand.

Primer extension experiments suggest that *C. reinhardtii* mitochondrial mRNAs contain few, if any, 5′-non-coding nucleotides; indeed, the *nd*5, *cox*I, and *orf*x (minor species; Boer & Gray 1986*a*) mRNAs all appear to begin at the AUG initiation codon. This is a feature shared with animal mitochondrial mRNAs (Montoya *et al.* 1981). However, the latter also lack an encoded 3′-untranslated region (Ojala *et al.* 1981), whereas *C. reinhardtii* mitochondrial mRNAs contain a genome-specified termination codon and a substantial 3′-non-coding region.

From the data currently available, it is not possible to tell how many promoters or transcriptional initiation sites there may be in the *C. reinhardtii* mitochondrial genome, or where such regulatory sequences may be located. One obvious possibility is the short intergenic region between the *nd*5 and *cox*I cistrons, which might nicely accommodate a bidirectional promoter. However, transcript-mapping studies have identified low levels of a putative *nd*5 precursor RNA whose 5′-end overlaps the beginning of the *cox*I gene (Boer & Gray 1986) (see figure 1), likewise, we have found rare *cox*I transcripts whose 5′-ends overlap the beginning of the *nd*5 gene (P. H. Boer, unpublished work). Hence, transcription of both the LTU and the RTU appears to initiate outside the *nd*5–*cox*I intergenic region.

Northern hybridization analysis has provided an assessment of transcript size and abundance. Putative mRNAs corresponding to each of the protein-coding genes (including ORFs x and y) have been identified; these are discrete, low-abundance, high-molecular-mass (0.55–1.9 kb) transcripts whose sizes closely approximate those predicted from mapping of transcript termini. Transcripts corresponding to each of the S and L modules have also been detected; these are high-abundance, low-molecular-mass (100–700 bp) RNAs, again with sizes that closely approximate the sizes of the rRNA modules, as defined by primary sequence comparisons and secondary structure modelling (see §7). In view of the evidence for cotranscription of protein-coding genes and rRNA modules, the very different steady-state levels of mature mRNAs and rRNAs indicates that mitochondrial RNA abundance must be regulated primarily by differential transcript stability.

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

5. Processing of *C. reinhardtii* mitochondrial transcripts

What mechanism(s) operate(s) in *C. reinhardtii* mitochondria to produce the specificity required for the precise endonucleolytic cleavage of precursor RNAs? In animal (Battey & Clayton 1980; Ojala *et al.* 1980) and some fungal (Lang *et al.* 1983; Breitenberger *et al.* 1985; Burger *et al.* 1985) mitochondrial systems, where processing also involves precise endonucleolytic cleavage of precursors, tRNA sequences flank and abut protein-coding and rRNA sequences and are implicated as processing signals. In *C. reinhardtii* mitochondria, the paucity of tRNA genes precludes a major role for tRNA sequences in the production of stable RNAs. However, the 3′-terminal regions of *C. reinhardtii* mRNAs do contain short direct and inverted repeats that have the potential to form characteristic secondary structure motifs (Pratje *et al.* 1984; Vahrenholz *et al.* 1985; Boer & Gray 1986*a*). The location of these motifs close to transcript termini suggests that they may play a role in directing the precise endonucleolytic cleavage of long primary transcripts (Boer & Gray 1986*a*). An intriguing feature of these putative processing signals is their ability to assume alternate secondary structures; thus, a conformational change from one secondary structure form to another, either during or after transcription, could serve to trigger an endonucleolytic cleavage in the vicinity of the motif (Boer & Gray 1986*a*).

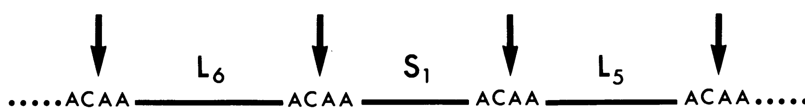

Figure 2. Primary sequence processing motifs in *C. reinhardtii* mitochondrial transcripts. Endonucleolytic cleavages (vertical arrows) map to ACAA sequences flanking the L_6, S_1 and L_5 rRNA modules, mature transcripts of which abut one another as they map to the genome.

Primary sequence motifs also are implicated in processing, especially in rRNA transcripts. Here, scissions occur primarily within short A+C-rich sequences, with the sequence ... ACAA ... predominating. A good example is the L_6–S_1–L_5 cluster (figure 2). It is difficult, however, to propose a simple, all-inclusive model of transcript processing in *C. reinhardtii* mitochondria, because several cleavages occur in regions lacking the motifs described above. Other motifs (as, for example, the tertiary structure of $tRNA^{Trp}$ and $tRNA^{Met}$) may play a role in processing at these sites.

6. The genetic code usage in *C. reinhardtii* mitochondria

Codon usage in the five sequenced and assigned protein-coding genes (*cox*I, *nd*1, *nd*2, *nd*5 and *urf*A) (Boer & Gray 1986*a*, *b*; P. H. Boer, unpublished work) is summarized in figure 3. The distribution of codons is highly biased, with 12 codons not used at all and a number of others used rarely. Among the 'absent' codons is CGG, which in plant mitochondria codes for Trp rather than Arg (Fox & Leaver 1981). The TGA termination codon, which is used as a Trp codon in animal and fungal mitochondrial genes (see Gray (1982) for references), is not used in this capacity in *C. reinhardtii* mitochondria (or in plant mitochondria (Fox & Leaver 1981)); neither is it used as a termination codon in any of the *C. reinhardtii* mitochondrial protein-coding genes we have sequenced, which end either in TAA or TAG.

Downloaded from rstb.royalsocietypublishing.org

TTT Phe 59 (3.1)	TCT Ser 43 (2.3)	TAT Tyr 24 (1.3)	TGT Cys 12 (0.6)
TTC Phe 87 (4.6)	TCC Ser 13 (0.7)	TAC Tyr 38 (2.0)	TGC Cys 18 (1.0)
TTA Leu 0 (0.0)	TCA Ser 0 (0.0)	TAA End 0 (0.0)	TGA End 0 (0.0)
TTG Leu 179 (9.5)	TCG Ser 0 (0.0)	TAG End 0 (0.0)	TGG Trp 43 (2.3)
CTT Leu 15 (0.8)	CCT Pro 9 (0.5)	CAT His 14 (0.7)	CGT Arg 34 (1.8)
CTC Leu 0 (0.0)	CCC Pro 3 (0.2)	CAC His 32 (1.7)	CGC Arg 15 (0.8)
CTA Leu 86 (4.6)	CCA Pro 50 (2.6)	CAA Gln 26 (1.4)	CGA Arg 3 (0.2)
CTG Leu 13 (0.7)	CCG Pro 4 (0.2)	CAG Gln 11 (0.6)	CGG Trp 0 (0.0)
ATT Ile 80 (4.2)	ACT Thr 66 (3.5)	AAT Asn 8 (0.4)	AGT Ser 53 (2.8)
ATC Ile 34 (1.8)	ACC Thr 38 (2.0)	AAC Asn 35 (1.9)	AGC Ser 35 (1.9)
ATA Ile 0 (0.0)	ACA Thr 0 (0.0)	AAA Lys 25 (1.3)	AGA Arg 0 (0.0)
ATG Met 87 (4.6)	ACG Thr 0 (0.0)	AAG Lys 3 (0.2)	AGG Arg 0 (0.0)
GTT Val 58 (3.1)	GCT Ala 124 (6.6)	GAT Asp 24 (1.3)	GGT Gly 124 (6.6)
GTC Val 27 (1.4)	GCC Ala 60 (3.2)	GAC Asp 19 (1.0)	GGC Gly 30 (1.6)
GTA Val 72 (3.8)	GCA Ala 3 (0.2)	GAA Glu 0 (0.0)	GGA Gly 5 (0.3)
GTG Val 4 (0.2)	GCG Ala 3 (0.2)	GAG Glu 39 (2.1)	GGG Gly 0 (0.0)

FIGURE 3. Codon usage in *C. reinhardtii* mitochondrial protein-coding genes, translated according to the plant mitochondrial genetic code (in which CGG encodes Trp rather than Arg). The table includes a total of 1887 codons from the five protein-coding genes (*nd*1, *nd*2, *nd*5, *cox*I and *urf*A) whose identities are known. Codons that are not used at all in these five genes have been boxed.

The extreme codon bias shown in figure 3 is not seen with plant mitochondrial protein genes, where codon usage is much more balanced. However, plant mitochondrial protein-coding genes do show a bias in the sense that 40% of the codons have a third position T (Fox & Leaver 1981; Dawson *et al.* 1984). With the exception of *orfy* (where 31.2% of codons have T in the third position), this particular bias is also displayed by *C. reinhardtii* mitochondrial protein-coding genes (37.0–43.2%; average 39.6%).

It is interesting that a biased codon usage has been noted in the case of protein-coding genes in *C. reinhardtii* chloroplast DNA (Rochaix *et al.* 1984). However, the nature of the bias is quite different in the two organelle compartments. For instance, TTA is the major Leu codon in a number of *C. reinhardtii* chloroplast genes but is not used at all in *C. reinhardtii* mitochondrial genes; on the other hand, TTG is the major Leu codon in *C. reinhardtii* mitochondria but is a rarely used Leu codon in *C. reinhardtii* chloroplasts.

The first 75 codons of the *orfx* coding region (Boer & Gray 1986*a*) do not contain any of the 12 codons that are absent in the collection of assigned protein-coding genes. In contrast, the longest *orfy* sequence (368 codons) contains 2 CTC (Leu), 1 ATA (Ile), 2 ACA (Thr) and 3 GGG (Gly) codons. These codons are distributed throughout the *orfy* sequence, so that even the shortest open reading frame (19 codons) contains an ACA codon. Thus, if the *orfy* gene is included in a compilation of codon usage, the number of unused codons drops to 8.

To decode the remaining 53 codons, even assuming an expanded codon recognition pattern ('two out of three' reading), a minimum of 22 tRNAs is required (23 if separate initiator and elongator $tRNAs^{Met}$ are used). As discussed in §3, this minimum complement of tRNAs does not appear to be encoded by *C. reinhardtii* mtDNA. This raises the possibility that extramitochondrial tRNAs are imported into *C. reinhardtii* mitochondria to function in translation there, as has been proposed for *Tetrahymena pyriformis* mitochondria (Suyama 1986).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

7. Scrambled ribosomal RNA coding segments in *C. reinhardtii* mitochondrial DNA

The pattern of segmented and dispersed rRNA coding segments (figure 1) raises the question of whether these modules encode functional LS and SS rRNAs. Secondary structure modelling strongly suggests that they do, because the high-abundance, small RNAs transcribed from these modules have the potential to reconstitute the conserved, functionally important core regions of LS and SS rRNA (Brimacombe 1982; Noller 1984). Figure 4 indicates those portions of the *E. coli* 16*S* rRNA secondary structure that can be reconstructed with the *C. reinhardtii* S modules. It can be seen that the four S segments are held together by *inter*molecular, long-range hydrogen-bonding interactions (between S_1 and S_2, S_1 and S_3, and S_3 and S_4) that reproduce *intra*molecular interactions present in *E. coli* 16*S* rRNA. The *C. reinhardtii* S segments can form an exact replica of about 56% of the proposed secondary structure of *E. coli* 16*S* rRNA. This conserved portion is virtually identical to a minimal core recently identified by Gutell *et al.* (1985) in a broad phylogenetic comparison of SS rRNA structure (figure 4, inset).

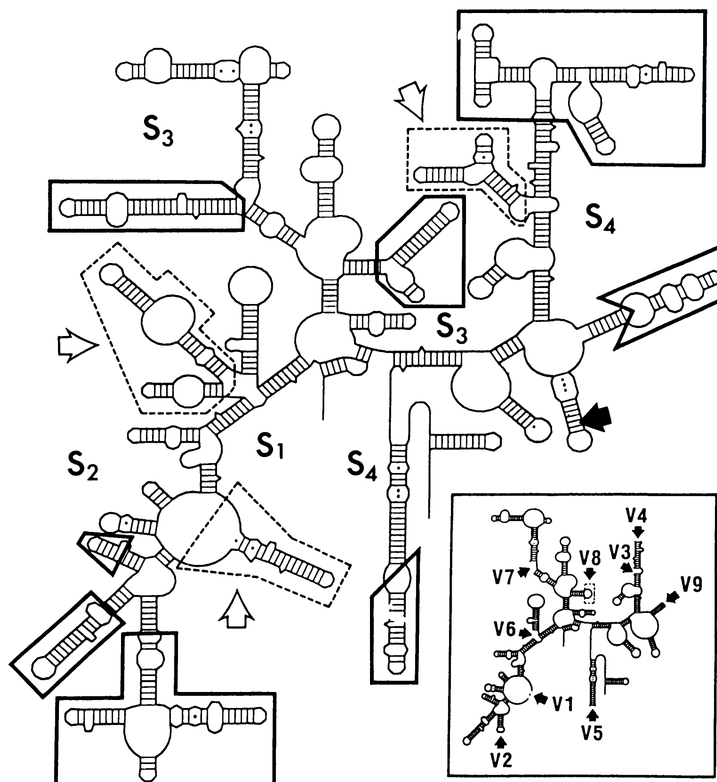

Figure 4. Secondary structure model of *C. reinhardtii* mitochondrial SS rRNA. The framework corresponds to the most recent secondary structure proposed for *E. coli* 16*S* rRNA (R. R. Gutell & C. R. Woese, unpublished work). *C. reinhardtii* segments S_1–S_4 can form an exact replica of the *E. coli* model except in those variable regions contained within the solid black lines. The three regions outlined by the dashed lines are effectively absent in the *C. reinhardtii* mitochondrial SS rRNA because they represent the discontinuities between different rRNA segments (open arrows). The solid arrow indicates the position of a unique 46 nucleotide insertion in the *C. reinhardtii* sequence. Inset: minimal structure, or core, showing those secondary structure elements universally conserved in SS rRNA (Gutell *et al.* 1985). Arrows denote the positions of the nine variable regions (V1, V2, etc.) defined by a comparison of potential secondary structure among eubacterial, chloroplast and plant mitochondrial SS rRNAs (Gray *et al.* 1984).

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Outside the conserved core, there is no primary sequence or secondary structure correspondence between the *C. reinhardtii* S segments and the analogous regions in *E. coli* 16*S* rRNA. These variable regions are designated V1 to V9 (Gray *et al.* 1984). It is noteworthy that the discontinuities in the *C. reinhardtii* mitochondrial SS rRNA map to regions that are normally variable (and therefore presumably of little functional significance) in covalently continuous SS rRNAs. Thus, the break between S_1 and S_2 coincides with V1; the break between S_2 and S_3 corresponds to V6; and the break between S_3 and S_4 is localized to V3.

The same picture emerges when the secondary structure of the *C. reinhardtii* LS rRNA is modelled. Here, eight L segments are held together by conserved, long-range interactions that form the basis of the secondary structure core in *E. coli* 23*S* rRNA. These intermolecular interactions tie together L_1 and L_{2a}; L_{2a} and L_{3b}; L_{3b} and L_4; L_4 and L_5; L_5 and L_6; L_6 and L_{2a}; L_6 and L_7; and L_7 and L_8. About 50% of the *E. coli* 23*S* secondary structure (Noller 1984) can be reproduced precisely by the *C. reinhardtii* L segments, and again the discontinuities between L segments are strictly localized to regions that are normally variable in covalently continuous LS rRNAs.

Recent studies from our laboratory have reinforced the idea that a ribosomal RNA molecule need not be a single, covalently continuous polynucleotide chain to be functional (Schnare *et al.* 1986; Spencer *et al.* 1987). Moreover, we have recently reported one other example of rearranged rRNA-coding segments: in *Tetrahymena pyriformis* mtDNA, a module encoding the 5′-terminal 280 nucleotides of the LS rRNA is located *downstream* of the remainder of the LS rRNA gene, and separated from it by a $tRNA^{Leu}$ gene (Heinonen *et al.* 1987). What we see in *C. reinhardtii* mtDNA is a particularly dramatic example of dispersed as well as discontinuous rRNA-coding segments. Among other things, our observations imply that ribosome biogenesis must be largely uncoupled from transcription and processing, with ribosomal subunits being assembled from mature S and L rRNA segments (perhaps already complexed with ribosomal proteins).

8. Contrasting patterns of mitochondrial genome organization and expression in *C. reinhardtii* and higher plants

As summarized in table 2, the *C. reinhardtii* mitochondrial genome has few features in common with the mitochondrial genome of higher plants (e.g. wheat). Instead, it displays a number of properties reminiscent of the small, economically organized animal mitochondrial

Table 2. Contrasting patterns of mitochondrial genome organization and expression

	human	*Chlamydomonas*	wheat
size (kbp)	16.6	15.8	430
gene linkage			
physical	extensive	extensive	minimal
transcriptional	extensive	extensive	minimal
endonucleolytic processing of transcripts	+	+	
introns	−	−	+
mRNAs			
5′-non-coding sequence	−	−	+
3′-non-coding sequence	−	+	+
TGA = Trp	+	−	−

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

genome. *C. reinhardtii* and human mtDNA are of a similar size (although different in form), and each contains intronless genes that are closely and extensively linked, both physically and transcriptionally. In both cases, processing of long cotranscripts occurs by way of precise endonucleolytic cleavages, giving rise to mRNAs having little or no 5′-non-coding sequence. In contrast, plant mtDNA is large, and its genes (some of which contain introns) are widely dispersed within a virtual 'sea' of non-coding DNA, with little evidence of physical or transcriptional linkage.

Although 'animal-like' in some respects, *C. reinhardtii* mtDNA does display major differences in organization and expression when compared with animal mtDNA. These differences include: (i) *C. reinhardtii* mtDNA lacks a number of the genes specified by animal mtDNA, yet contains two apparently novel protein-coding genes (ORFs x and y); (ii) endonucleolytic processing of transcripts is signalled differently in animal and *C. reinhardtii* mitochondria; (iii) *C. reinhardtii* mtDNA probably does not encode a full set of tRNA genes, but those three it does contain show none of the structural peculiarities displayed by their animal mitochondrial counterparts (Anderson *et al.* 1982); and (iv) *C. reinhardtii* mitochondria use the universal genetic code (albeit dispensing with a number of codons altogether). Perhaps the most dramatic contrast is provided by the rRNA genes. In animal mtDNA, the condensed LS (16*S*) and SS (12*S*) rRNA genes together occupy less than 3 kbp; each is uninterrupted, and the two are separated by only a tRNA gene. In *C. reinhardtii* mtDNA, the rRNA genes are each divided into a number of segments that are scattered throughout more than 6 kbp of the mtDNA, interspersed with one another and with protein-coding and tRNA genes. These and other novel features of *C. reinhardtii* mtDNA serve to emphasize once again the wide diversity that exists in the patterns of mitochondrial gene arrangement and expression.

Although comparative studies of a number of animal and plant mtDNAs have defined 'prototypical' (and quite different) animal and plant mitochondrial genomes, one may question whether the results reported here are representative of mtDNA in unicellular algae. Boynton *et al.* (1987) have recently reported that *C. smithii*, which is interfertile with *C. reinhardtii*, contains mtDNA that is collinear with *C. reinhardtii* mtDNA except in the region of the *cyb* gene, where *C. smithii* mtDNA has a 1 kbp insertion. We have previously presented evidence that two other *Chlamydomonas* species, *C. eugametos* and *C. moewusii* (which are interfertile with each other but not with *C. reinhardtii*), have mitochondrial genomes that are somewhat larger than *C. reinhardtii* mtDNA, and may be circular rather than linear (Boer *et al.* 1985*a*). These observations are supported by recent findings that *C. moewusii* mtDNA has a circular map of 21 kbp (C. Lemieux, personal communication). On the basis of heterologous hybridization experiments with gene-specific mtDNA probes from *C. reinhardtii*, it appears that the *C. moewusii* mitochondrial genome is completely rearranged relative to *C. reinhardtii* mtDNA (R. W. Lee, C. Lemieux and M. Turmel, unpublished work). From these results, it is already clear that additional unicellular and multicellular algae will have to be examined before it is possible to tell whether a 'prototypical' algal mitochondrial genome exists.

10. Is there a specific evolutionary connection between green algal and plant mitochondrial genomes?

At the beginning of this paper, we posed the question of whether there is a specific evolutionary link between the mitochondrial genomes of green algae and higher plants. Because of the wide variation in the size, organization and mode of expression of mitochondrial

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

DNA, these parameters are of limited value in attempts to trace the evolutionary history of the mitochondrial lineage (Gray 1983). Ribosomal RNA sequence information, on the other hand, has proved exceptionally useful in this regard, providing a basis for the construction of global phylogenetic trees connecting all organisms and organelles (discussed in detail in Gray *et al.* (1984)). Because rRNA genes are encoded by all three genomes (nuclear, chloroplast, mitochondrial) present in all algal or plant cells, we may ask whether phylogenies (i.e. tree topologies) based on nuclear, chloroplast and mitochondrial SS and/or LS rRNA sequences are congruent, i.e. whether the three genomes or organelles appear to have evolved in parallel. Such is the expectation if a single ancestral cell was the source of the nuclear, chloroplast and mitochondrial genomes in unicellular and multicellular green algae and higher plants.

Phylogenetic trees based on nuclear 5*S* (Küntzel *et al.* 1983; Vandenberghe *et al.* 1984) and SS (Gunderson *et al.* 1987) rRNA sequences do group *C. reinhardtii* and other unicellular green algae together specifically with higher plants, indicating a recent common origin of the nuclear genome of chlorophytes and angiosperms. The same holds true for the chloroplast genome of *C. reinhardtii* and higher plants (Gray *et al.* 1984). However, when the *C. reinhardtii* mitochondrial SS or LS rRNA sequences (P. H. Boer, unpublished results) are analysed in concert with the corresponding rRNA sequences from other mitochondria, the algal and higher plant mitochondria do not branch together in the resulting phylogenetic trees (unpublished results with Y. Abel, R. Cedergren and D. Sankoff). Instead, the *C. reinhardtii* mitochondrial branch is positioned as an early offshoot of the branch leading to animal mitochondria, and is quite separate from the plant mitochondrial branch. These results suggest a separate evolutionary origin of at least the ribosomal RNA genes in *C. reinhardtii* and plant mitochondria, and raise the fascinating possibility that mitochondria may have originated as eubacteria-like endosymbionts (Gray 1983) in at least two separate evolutionary events. These observations provide an additional incentive for further investigations of mitochondrial genome organization, expression and evolution in other chlorophytes, and in members of the Kingdom Plantae other than angiosperms.

We thank L. Bonen, C. Lemieux and U. L. RajBhandary for communicating unpublished results; R. Gutell for supplying a line drawing of the most recent secondary structure of *E. coli* 16*S* rRNA (from which figure 4 was prepared); and L. Bonen, C. Lemieux and R. W. Lee for helpful discussions. This work was supported by a grant (MT-4124) from the Medical Research Council of Canada to M.W.G., who is a Fellow of the Canadian Institute for Advanced Research.

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. 1982 Comparison of the human and bovine mitochondrial genomes. In *Mitochondrial genes* (ed. P. Slonimski, P. Borst & G. Attardi), pp. 5–43. New York: Cold Spring Harbor Laboratory.

Attardi, G. 1985 Animal mitochondrial DNA: an extreme example of genetic economy. *Int. Rev. Cytol.* **93**, 93–145.

Battey, J. & Clayton, D. A. 1980 The transcription map of human mitochondrial DNA implicates transfer RNA excision as a major processing event. *J. biol. Chem.* **255**, 11599–11606.

Benne, R. 1985 Mitochondrial genes in trypanosomes. *Trends Genet.* **1**, 117–121.

Boer, P. H., Bonen, L., Lee, R. W. & Gray, M. W. 1985*a* Genes for respiratory chain proteins and ribosomal RNAs are present on a 16-kilobase-pair DNA species from *Chlamydomonas reinhardtii* mitochondria. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3340–3344.

Boer, P. H. & Gray, M. W. 1986*a* The URF 5 gene of *Chlamydomonas reinhardtii* mitochondria: DNA sequence and mode of transcription. *EMBO J.* **5**, 21–28.

Downloaded from rstb.royalsocietypublishing.org

Boer, P. H. & Gray, M. W. 1986*b* Nucleotide sequence of a protein coding region in *Chlamydomonas reinhardtii* mitochondrial DNA. *Nucl. Acids Res.* **14**, 7506–7507.

Boer, P. H., McIntosh, J. E., Gray, M. W. & Bonen, L. 1985*b* The wheat mitochondrial gene for apocytochrome *b*: absence of a prokaryotic ribosome binding site. *Nucl. Acids Res.* **13**, 2281–2292.

Bonen, L., Boer, P. H. & Gray, M. W. 1984 The wheat cytochrome oxidase subunit II gene has an intron insert and three radical amino acid changes relative to maize. *EMBO J.* **3**, 2531–2536.

Boynton, J. E., Harris, E. H., Burkhart, B. D., Lamerson, P. M. & Gillham, N. W. 1987 Transmission of mitochondrial and chloroplast genomes in crosses of *Chlamydomonas*. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2391–2395.

Breitenberger, C. A., Browning, K. S., Alzner-DeWeerd, B. & RajBhandary, U. L. 1985 RNA processing in *Neurospora crassa* mitochondria: use of transfer RNA sequences as signals. *EMBO J.* **4**, 185–195.

Breitenberger, C. A. & RajBhandary, U. L. 1985 Some highlights of mitochondrial research based on analyses of *Neurospora crassa* mitochondrial DNA. *Trends biochem. Sci.* **10**, 478–483.

Brimacombe, R. 1982 The secondary structure of ribosomal RNA, and its organization within the ribosomal subunits. *Biochem. Soc. Symp.* **47**, 49–60.

Burger, G., Citterich, M. H., Nelson, M. A., Werner, S. & Macino, G. 1985 RNA processing in *Neurospora crassa* mitochondria: transfer RNAs punctuate a large precursor transcript. *EMBO J.* **4**, 197–204.

Chomyn, A., Mariottini, P., Cleeter, M. W. J., Ragan, C. I., Doolittle, R. F., Yagi, A. M., Hatefi, Y. & Attardi, G. 1985 Functional assignment of the products of the unidentified reading frames of human mitochondrial DNA. In *Achievements and perspectives of mitochondrial research* (ed. E. Quagliariello, E. C. Slater, F. Palmieri, C. Saccone & A. M. Kroon), vol. 2 (*Biogenesis*), pp. 259–275. Amsterdam, New York and Oxford: Elsevier Science Publishers.

Chomyn, A., Cleeter, M. W. J., Ragan, C. I., Riley, M., Doolittle, R. F. & Attardi, G. 1986 URF6, last unidentified reading frame of human mtDNA, codes for an NADH dehydrogenase subunit. *Science, Wash.* **234**, 614–618.

Clayton, D. A., 1984 Transcription of the mammalian mitochondrial genome. *A. Rev. Biochem.* **53**, 573–594.

Dawson, A. J., Jones, V. P. & Leaver, C. J. 1984 The apocytochrome *b* gene in maize mitochondria does not contain introns and is preceded by a potential ribosome binding site. *EMBO J.* **3**, 2107–2113.

Fox, T. D. & Leaver, C. J. 1981 The *Zea mays* mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. *Cell* **26**, 315–323.

Grant, D. & Chiang, K.-S. 1980 Physical mapping and characterization of *Chlamydomonas* mitochondrial DNA molecules: their unique ends, sequence homogeneity, and conservation. *Plasmid* **4**, 82–96.

Gray, M. W. 1982 Mitochondrial genome diversity and the evolution of mitochondrial DNA. *Can. J. Biochem.* **60**, 157–171.

Gray, M. W. 1983 The bacterial ancestry of plastids and mitochondria. *BioScience* **33**, 693–699.

Gray, M. W., Sankoff, D. & Cedergren, R. J. 1984 On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. *Nucl. Acids Res.* **12**, 5837–5852.

Gunderson, J. H., Elwood, H., Ingold, A., Kindle, K. & Sogin, M. L. 1987 Phylogenetic relationships between chlorophytes, chromophytes and oomyctes. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5823–5827.

Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. 1985 Comparative anatomy of 16-S-like ribosomal RNA. *Prog. Nucl. Acid Res. molec. Biol.* **32**, 155–216.

Heinonen, T. Y. K., Schnare, M. N., Young, P. G. & Gray, M. W. 1987 Rearranged coding segments, separated by a transfer RNA gene, specify the two parts of a discontinuous large subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. biol. Chem.* **262**, 2879–2887.

Kück, U. & Neuhaus, H. 1986 Universal genetic code evidenced in mitochondria of *Chlamydomonas reinhardtii*. *Appl. Microbiol. Biotechnol.* **23**, 462–469.

Küntzel, H., Piechulla, B. & Hahn, U. 1983 Consensus structure and evolution of 5S rRNA. *Nucl. Acids Res.* **11**, 893–900.

Lang, B. F., Ahne, F., Distler, S., Trinkl, H., Kaudewitz, F. & Wolf, K. 1983 Sequence of the mitochondrial DNA, arrangement of genes and processing of their transcripts in *Schizosaccharomyces pombe*. In *Mitochondria 1983: nucleo-mitochondrial interactions* (ed. R. J. Schweyen, K. Wolf & F. Kaudewitz), pp. 313–329. Berlin, New York: Walter de Gruyter.

Montoya, J., Ojala, D. & Attardi, G. 1981 Distinctive features of the 5′-terminal sequences of the human mitochondrial mRNAs. *Nature, Lond.* **290**, 465–470.

Netzker, R., Köchel, H. G., Basak, N. & Küntzel, H. 1982 Nucleotide sequence of *Aspergillus nidulans* mitochondrial genes coding for ATPase subunit 6, cytochrome oxidase subunit 3, seven unidentified proteins, four tRNAs and L-rRNA. *Nucl. Acids Res.* **10**, 4783–4794.

Noller, H. F. 1984 Structure of ribosomal RNA. *A. Rev. Biochem.* **53**, 119–162.

Ojala, D., Merkel, C., Gelfand, R. & Attardi, G. 1980 The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* **22**, 393–403.

Ojala, D., Montoya, J. & Attardi, G. 1981 tRNA punctuation model of RNA processing in human mitochondria. *Nature, Lond.* **290**, 470–474.

Downloaded from rstb.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Pickett-Heaps, J. D. 1975 *Green algae: structure, reproduction and evolution in selected genera.* Sunderland, Massachusetts: Sinauer.

Pratje, E., Schnierer, S. & Dujon, B. 1984 Mitochondrial DNA of *Chlamydomonas reinhardtii*: the DNA sequence of a region showing homology with mammalian URF2. *Curr. Genet.* **9**, 75–82.

Pring, D. R. & Lonsdale, D. M. 1985 Molecular biology of higher plant mitochondrial DNA. *Int. Rev. Cytol.* **97**, 1–46.

Ragan, M. A. & Chapman, D. J. 1978 *A biochemical phylogeny of the protists.* New York, San Francisco and London: Academic Press.

Rochaix, J.-D., Dron, M., Rahire, M. & Malnoe, P. 1984 Sequence homology between the 32 K dalton and the D2 chloroplast membrane polypeptides of *Chlamydomonas reinhardtii. Pl. molec. Biol.* **3**, 363–370.

Ryan, R., Grant, D., Chiang, K.-S. & Swift, H. 1978 Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii. Proc. natn. Acad. Sci. U.S.A.* **75**, 3268–3272.

Schnare, M. N., Heinonen, T. Y. K., Young, P. G. & Gray, M. W. 1986 A discontinuous small subunit ribosomal RNA in *Tetrahymena pyriformis* mitochondria. *J. biol. Chem.* **261**, 5187–5193.

Sederoff, R. R. 1984 Structural variation in mitochondrial DNA. *Adv. Genet.* **22**, 1–108.

Spencer, D. F., Collings, J. C., Schnare, M. N. & Gray, M. W. 1987 Multiple spacer sequences in the nuclear large subunit ribosomal RNA gene of *Crithidia fasciculata. EMBO J.* **6**, 1063–1071.

Suyama, Y. 1986 Two dimensional polyacrylamide gel electrophoresis analysis of *Tetrahymena* mitochondrial tRNA. *Curr. Genet.* **10**, 411–420.

Vahrenholz, C., Pratje, E., Michaelis, G. & Dujon, B. 1985 Mitochondrial DNA of *Chlamydomonas reinhardtii*: sequence and arrangement of *URF5* and the gene for cytochrome oxidase subunit I. *Molec. gen. Genet.* **201**, 213–224.

Vandenberghe, A., Chen, M.-W., Dams, E., de Baere, R., de Roeck, E., Huysmans, E. & de Wachter, R. 1984 The corrected nucleotide sequences of 5S RNAs from six angiosperms. *FEBS Lett.* **171**, 17–23.

Wolstenholme, D. R., Macfarlane, J. L., Okimoto, R., Clary, D. O. & Wahleithner, J. A. 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1324–1328.