The EMBO Journal vol.4 no.7 pp.1617–1623, 1985

The maize cytochrome *c* oxidase subunit I gene: sequence, expression and rearrangement in cytoplasmic male sterile plants

Peter G. Isaac, Valerie P. Jones[1] and Christopher J. Leaver

Department of Botany, The King's Buildings, University of Edinburgh, Mayfield Road, Edinburgh, EH9 3JH, UK

[1]Present address: Amersham International plc, White Lion Road, Amersham, Bucks, HP7 9LL, UK

Communicated by C.J. Leaver

The single copy of the gene for cytochrome *c* oxidase subunit I (*COX* I) present in the mitochondrial genome of fertile maize (*Zea mays* L.) is encoded by a continuous open reading frame of 1584 nucleotides. The predicted polypeptide encoded by the gene has a mol. wt. of 58 219 daltons and shows >60% amino acid sequence homology with the corresponding fungal and animal polypeptides. Two major transcripts of 2400 and 2300 nucleotides can be detected and the 5′ end of the larger transcript maps to a sequence from −161 to −153 (relative to the initiator codon) which shows high homology to the yeast mitochondrial promoter. In mitochondrial DNA from the S male-sterile cytoplasm of maize, which also characteristically contain two low mol. wt. linear DNAs (S1 and S2), rearrangements just 5′ (at −175) to the *COX* I gene, generate additional DNA restriction fragments containing entire copies of the gene. These rearrangements involve a sequence identical to the terminal 186 bp of the 208-bp inverted repeat sequence found at either end of the S1 and S2 DNAs.

Key words: cytochrome *c* oxidase/plant mtDNA/cytoplasmic male sterility/terminal inverted repeats/DNA recombination

Introduction

The physical map and organisation of the mitochondrial genomes from two higher plants have recently been described (Palmer and Shields, 1984; Lonsdale *et al.*, 1984). Despite a 2-fold difference in size, the mitochondrial genomes of *Brassica campestris* (218 kb) and maize (*Zea mays*, 570 kb) have been shown to have a similar complex multipartite structure. While the entire genome can be represented as a single master circle, recombination between inverted and/or directly repeated sequences leads to the generation of smaller subgenomic circles. The number and relative recombination frequency of the homologous repeated sequences and the efficiency of different replication origins are thought to determine the genome organisation *in vivo* (Lonsdale, 1984).

In maize, cytoplasmic male sterility (CMS), a phenotype recognised by the failure to produce functional pollen, is associated with sequence reorganisation of the mitochondrial genome, as characterised by restriction endonuclease analysis of the normal (fertile) and the three CMS types, CMS-T, C and S (Pring and Levings, 1978). Such genome reorganisation could be mediated by recombination between the types of repeated sequences mentioned above. In the S-male sterile cytoplasm of maize, mitochondria contain two small, double-stranded linear DNAs (sometimes referred to as episomal or 'plasmid-like' DNAs) in addition to the main mitochondrial DNA (mtDNA) (Pring *et al.*, 1977). These 'linear' DNAs, termed S1 and S2, are 6397 and 5453 bp long, contain terminal inverted repeats of 208 bp with a protein covalently attached to their 5′ termini (Kemble and Thompson, 1982), and share sequence homology with the main mitochondrial genome of all maize cytoplasms examined to date (Lonsdale *et al.*, 1981; Levings and Sederoff, 1983). They are normally present in equimolar amounts and are at least 5-fold more abundant than the main band mtDNA. We have previously shown that mitochondria containing the S1 and S2 DNAs synthesise unique, high mol. wt., variant polypeptides which are not synthesised by mitochondria from the other cytoplasms (Forde and Leaver, 1980).

Because the S1 and S2 DNAs share common structural features with transposable elements, viruses and phages with terminal inverted repeat sequences, it has been suggested that in some cases genome rearrangement may be related to the presence of specific sequences within them (Levings *et al.*, 1980; Levings and Sederoff, 1983).

To extend our knowledge of the information content and structure of the large higher plant mitochondrial genome, as well as to investigate reorganisation of the genome associated with the male sterile phenotype, we have isolated and sequenced the gene for subunit I of cytochrome *c* oxidase (*COX* I) from maize. In contrast to the homologous gene in yeast, the plant gene does not contain introns and gives rise to two principal transcripts of 2.4 and 2.3 kb. S1 nuclease analysis reveals that the 5′ end of the larger transcript lies within a sequence which shows strong homology to the conserved nonanucleotide sequences thought to be involved in transcription initiation of yeast mitochondrial genes (Osinga and Tabak, 1982).

Using the cloned maize *COX* I gene as a hybridisation probe we have shown that the N (fertile) and CMS-T and C maize mitochondrial genomes each contain single copies of the gene, whereas there are apparently several distinct copies associated with DNA rearrangement in the flanking sequences in CMS-S mtDNA. Sequence analysis 5′ (upstream) to one copy of the gene in CMS-S mtDNA reveals the presence of an almost complete copy of the inverted repeat (TIR) sequence found at the termini of the S1 and S2 DNAs.

Results and Discussion

Identification and cloning of the maize COX I gene

The plasmid pKLD (a gift from L.A.M. Hensgens, University of Amsterdam, Hensgens *et al.*, 1983) containing part of the yeast gene for subunit I of cytochrome *c* oxidase was labelled with ^{32}P by nick translation (Rigby *et al.*, 1977) and hybridised under non-stringent conditions to restriction fragments of maize mtDNA which had been separated by gel electrophoresis and transferred to nitrocellulose. The probe hybridised to a single 12-kb *Bam*HI fragment (data not shown). Fragments of this size were eluted from a preparative gel (Thuring *et al.*, 1975) and cloned into the *Bam*HI site of pAT153. One clone, designated pBN1266, hybridised strongly to the yeast probe under non-stringent con-

© IRL Press Limited, Oxford, England.

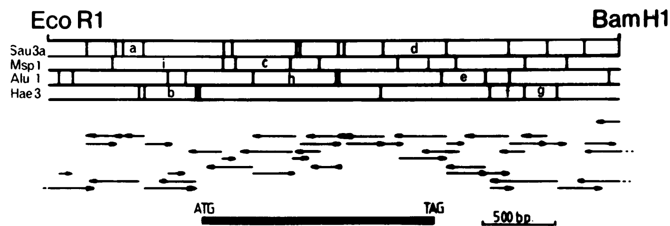

Fig. 1. Restriction map of maize mtDNA containing the cytochrome *c* oxidase subunit I gene and flanking sequences. A 3.9-kb *Eco*RI to *Bam*HI restriction fragment showing homology to the *oxi*-3 gene from yeast was cloned into pAT153 to generate the clone pBN 6601 shown above. Letters a – i indicate the M13mp8 clones referred to in the text and the arrows indicate the origin, direction and extent of sequence analysis from each restriction site. The box at the bottom of the figure represents the position of the *COX* I gene.

ditions. The region of homology to the probe was subsequently found to be restricted to a 3.9-kb *Bam*HI-*Eco*RI mtDNA fragment which was subsequently sub-cloned into pAT153, to produce the plasmid designated pBN6601 (Figure 1).

Nucleotide sequence of the maize COX I gene

The mtDNA inserted in pBN6601 was sequenced using the strategy summarised in Figure 1. The sequence contains a continuous open reading frame of 1584 bp (Figure 2) which shows 58% homology with the corresponding sequences of the cytochrome *c* oxidase subunit I genes from yeast (Bonitz *et al.*, 1980) and man (Anderson *et al.*, 1981). In common with the homologous gene in man (Anderson *et al.*, 1981), *Drosophila* (de Bruijn, 1983) and *Neurospora* (Burger *et al.*, 1982; de Jonge and de Vries, 1983) the maize *COX* I gene does not appear to contain introns. In contrast, the *COX* I gene in yeast contains several introns; in *Saccharomyces cerevisiae* for example, the split *COX* I gene contains up to 10 exons, the number of exons depending on the strain (Bonitz *et al.*, 1980; Hensgens *et al.*, 1983). The only maize mitochondrial gene shown to contain an intron to date is the gene for subunit II of cytochrome *c* oxidase (Fox and Leaver, 1981).

The *COX* I gene is preceded by the sequence 5′-GGTTTTCA-3′ (Figure 2) which may function as a ribosome binding site in maize mitochondrial mRNAs (Dawson *et al.*, 1984). Eleven of the 13 bases separating the initiator methionine codon from the putative ribosome binding site are A residues. This arrangement is similar to that described for the homologous yeast gene, where 13 of the 15 bases immediately preceding the initiator codon are A residues (Bonitz *et al.*, 1980).

Amino acid sequence

The maize polypeptide, predicted from the DNA sequence, is 528 amino acids long and has a mol. wt. of 58 219 (Figure 2). It shares extensive overall amino acid sequence homology with the *Neurospora* (62%), yeast (60%), man (68%) and *Drosophila* (65%) proteins, although homology at the carboxy terminus is low (Figure 3) and the proteins vary in length. The non-conserved region extends from the last hydrophobic domain to the relatively hydrophilic carboxy terminus.

Examination of codon usage in the maize *COX* I gene reveals that, in common with other plant mitochondrial protein codon genes (*COX* II, Fox and Leaver, 1981; *COB*, Dawson *et al.*, 1984), there is a strong bias towards the use of T (35%) in the third position of the codon. The maize *COX* I gene does not contain any TGA codons which encode tryptophan in other mitochondria, but does contain one CGG codon. In view of the findings of Fox and Leaver (1981), we predict that this triplet encodes tryptophan.

A hydropathy profile of the predicted maize cytochrome *c* oxidase subunit I protein sequence reveals the presence of 12 hydrophobic regions (data not shown) which correspond exactly with the 12 transmembrane hydrophobic regions described by Saraste and Wikstrom (1983) in the human, ox, mouse, yeast and *Neurospora* protein. This overall conservation of hydrophobic and hydrophilic domains extends to the carboxy terminus, despite a lack of conserved amino acids. In addition, histidine residues occur in the maize protein at positions 64 and 154 and these align precisely with the histidine residues which Saraste and Wikstrom (1983) proposed to be important in haem binding by the protein.

Transcription of the COX I gene

To determine whether the maize *COX* I gene is transcribed, specific gene probes were hybridised to total mitochondrial (mt) RNA which had been separated by electrophoresis on agarose-formaldehyde gels and transferred to nitrocellulose. When an M13 clone containing sequences lying completely within the *COX* I coding region is used as a probe, two major transcripts of ~2.4 and 2.3 kb are detected (Figure 4, lane c). To identify in more detail the 5′ and 3′ ends of the two *COX* I transcripts with respect to the protein coding sequence, identical RNA blots were hybridised with a series of ^{32}P-labelled M13 clones of the *COX* I region (Figure 4, a – g). Probe b hybridises only to the higher mol. wt. transcript, suggesting that the two principal transcripts differ at their 5′ termini. Probe f hybridises weakly to both transcripts.

S1 nuclease mapping experiments were performed in order to characterise the 5′ termini of the *COX* I transcripts in more detail. An M13 clone containing a 759-bp *Msp*I insert which spans the 5′ ends of the two principal transcripts (fragment i, in Figure 1) was uniformly labelled, hybridised with total mtRNA, treated with S1 nuclease and the protected hybrids analysed by electrophoresis on 6% (w/v) DNA sequencing gels. A major S1 nuclease-protected fragment of 286 bp and several less abundant fragments (of which the most abundant is ~191 bp) were detected (see Figure 5). This indicates that the two major transcripts visible in Figure 4 begin at approximately −152 bp and −57 bp from the putative ATG initiation codon (see Figure 2). We do not know whether the two transcripts originate from separate initiation events or from processing of the larger molecule. Guanylyl transferase 'capping' experiments in progress may resolve these two alternatives.

Examination of the DNA sequence around the 5′ end of the larger of these two transcripts reveals the presence of a sequence 5′-TCATAAGTA-3′ (Figure 2) which exhibits a seven out of nine homology to the conserved nonanucleotide 5′-ATATAAGTA-3′ thought to be involved in transcription initiation in the mitochondrial genome of *S. cerevisiae* (Osinga and Tabak, 1982; Osinga *et al.*, 1984). No other sequences with a seven out of nine match to the nonanucleotide were found on the same DNA strand in the entire region 5′ to the gene which has been sequenced (1070 bp).

These results show that the transcripts of the maize *COX* I gene (~2300 – 2400 nucleotides) are considerably larger than the coding region of the gene (1584 nucleotides), most of the additional sequence being at the 3′ end of the mRNA. The transcript pattern of the *COX* I gene is, however, much simpler than the complex pattern of transcripts described for the maize *COX* II (Fox and Leaver, 1981) and *COB* (Dawson *et al.*, 1984) genes.

```
-180
ACTTTTGCACCGAAGAAACTCATAAGTAATCCAATTTCCGAGGACACGAACGGAAGTAAGCTGTTTTTGGTATTCGTATCGGAATTAGCC
                                      COX I mRNA
-90
AGAAGAAGAAACCATTTTCACTGCCCTTCATTCTTTGGCCTGCCTGCCTTAGTGGCCCCTCTCTGATAAGGTTTTCAAAACGAAAAAAAA

1
ATGACAAATCTGGTCCGATGGCTCTTCTCCACTAACCACAAGGATATCGGTACTCTATATTTCATCTTCGGTGCCATTGCAGGAGTGATG
 M  T  N  L  V  R  W  L  F  S  T  N  H  K  D  I  G  T  L  Y  F  I  F  G  A  I  A  G  V  M
91
GGCACATGCTTCTCCGTACTGATTCGTATGGAATTAGCCCGACCCGGCGATCAAATTCTTGGTGGGAATCATCAACTTTATAATGTTTTA
 G  T  C  F  S  V  L  I  R  M  E  L  A  R  P  G  D  Q  I  L  G  G  N  H  Q  L  Y  N  V  L
181
ATAACGGCTCACGCTTTTTTAATGATCTTTTTTATGGTTATGCCGGCGATGATAGGTGGATTTGGGAATTGGTTTGTTCCGATTCTGATA
 I  T  A  H  A  F  L  M  I  F  F  M  V  M  P  A  M  I  G  G  F  G  N  W  F  V  P  I  L  I
271
GGTGCACCTGACATGGCATTTCCACGATTAAATAATATATCATTCTGGTTGTTGCCACCAAGTCTCTTGCTCCTATTAAGCTCAGCCTTA
 G  A  P  D  M  A  F  P  R  L  N  N  I  S  F  W  L  L  P  P  S  L  L  L  L  L  S  S  A  L
361
GTAGAAGTGGGCAGCGGCACTGGGTGGACGGTCTATCCGCCCTTAAGTGGTATTACCAGCCATTCTGGAGGAGCAGTTGATTTAGCAATT
 V  E  V  G  S  G  T  G  W  T  V  Y  P  P  L  S  G  I  T  S  H  S  G  G  A  V  D  L  A  I
451
TTTAGTCTTCATCTATCAGGTGTTTCATCAATTTTAGGTTCTATCAATTTTATAACTACTATCTTCAACATGCGTGGACCTGGAATGACT
 F  S  L  H  L  S  G  V  S  S  I  L  G  S  I  N  F  I  T  T  I  F  N  M  R  G  P  G  M  T
541
ATGCATAGATTACCACTTTTTGTGTGGTCCGTTTTAGTGACAGCATTCCTACTTTTATTATCACTTCCGGTACTGGCAGGGGCAATTACA
 M  H  R  L  P  L  F  V  W  S  V  L  V  T  A  F  L  L  L  L  S  L  P  V  L  A  G  A  I  T
631
ATGTTATTAACCGATCGAAACTTTAATACAACCTTTTTTGATCCAGCAGGAGGGGGAGACCCAATATTATACCAGCATCTCTTTTGGTTC
 M  L  L  T  D  R  N  F  N  T  T  F  F  D  P  A  G  G  G  D  P  I  L  Y  Q  H  L  F  W  F
721
TTCGGTCATCCAGAGGTGTATATTCTCATTCTGCCTGGATTCGGTATTATTAGTCATATCGTATCGACCTTTTCAAGAAAACCGGTCTTC
 F  G  H  P  E  V  Y  I  L  I  L  P  G  F  G  I  I  S  H  I  V  S  T  F  S  R  K  P  V  F
811
GGGTATCTAGGCATGGTTTATGCCATGATAAGTATAGGTGTTCTTGGATTTCTAGTTTGGGCTCATCATATGTTTACTGTGGGCTTAGAC
 G  Y  L  G  M  V  Y  A  M  I  S  I  G  V  L  G  F  L  V  W  A  H  H  M  F  T  V  G  L  D
901
GTTGATACGCGTGCCTACTTCACCGCAGCTACCATGATCATAGCTGTGCCCACTGGAATCAAAATCTTTAGTTGGATCGCTACCATGTGG
 V  D  T  R  A  Y  F  T  A  A  T  M  I  I  A  V  P  T  G  I  K  I  F  S  W  I  A  T  M  W
991
GGAGGTTCGATACAATACAAAACACCCATGTTATTTGCTGTAGGGTTCATCTTTTTGTTCACCATAGGAGGGCTCACTGGAATAGTTCTA
 G  G  S  I  Q  Y  K  T  P  M  L  F  A  V  G  F  I  F  L  F  T  I  G  G  L  T  G  I  V  L
1081
GCAAACTCTGGGCTAGACATTGCTCTACATGATACTTATTATGTGGTTGCACATTTCCATTATGTACTTTCTATGGGAGCCGTTTTTGCT
 A  N  S  G  L  D  I  A  L  H  D  T  Y  Y  V  V  A  H  F  H  Y  V  L  S  M  G  A  V  F  A
1171
TTATTTGCTGGATTTTACTATTGGGTGGGTAAAATCTTTGGTCGGACATATCCTGAAACTTTAGGCCAAATCCATTTTTGGATCACTTTT
 L  F  A  G  F  Y  Y  W  V  G  K  I  F  G  W  T  Y  P  E  T  L  G  Q  I  H  F  W  I  T  F
1261
TTCGGGGTTAATCTGACCTTCTTTCCCATGCATTTCTTAGGGCTTTCGGGTATGCCACGTCGCATTCCAGATTATCCAGATGCTTACGCC
 F  G  V  N  L  T  F  F  P  M  H  F  L  G  L  S  G  M  P  R  R  I  P  D  Y  P  D  A  Y  A
1351
GGATGGAATGCTCTGAGCAGTTTCGGTTCTTATATATCCGTAGTTGGGATTCGTCGTTTCTTCGTAGTTGTCGCAATCACTTCAAGCAGT
 G  W  N  A  L  S  S  F  G  S  Y  I  S  V  V  G  I  R  R  F  F  V  V  V  A  I  T  S  S  S
1441
GGAAAGAACAAAAGATGTGCGGAAAGTCCTTGGGCTGTTGAGCAGAATCCAACCACACTAGAATGGTTGGTACAAAGCCCTCCAGCCTTT
 G  K  N  K  R  C  A  E  S  P  W  A  V  E  Q  N  P  T  T  L  E  W  L  V  Q  S  P  P  A  F
1531
CATACTTTTGGAGAACTTCCAACTATCAAAGAAACTAGAAACCAATCCAGTTGCTAGAAATCCATCATTAGCGAGACAATTATTTTTCTT
 H  T  F  G  E  L  P  T  I  K  E  T  R  N  Q  S  S  C
1621
TTCTTTTGATTTGTTTGGCTTTGCCTTAACCGAAGCTATTGCATTGTTTGCACTTATGAGGGATCTTATTTGTGTTCTAAAATATTCGAG

1711
TAAAGTATGAAAGGCTGGAGCAACACTTCTCAATGCCACTTCCGGAAACAAATTCGT
```

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the mitochondrial cytochrome *c* oxidase subunit I gene from fertile maize. The sequence is numbered relative to the A of the ATG initiation codon. The boxed sequences from −161 to −153 and −65 to −58 represent the DNA sequences to which the 5′ end of the *COX* I transcripts have been mapped (see Figure 5). The putative ribosome binding site at −21 to −14 is underlined (▮▮▮▮▮▮). The predicted amino acid sequence was translated according to the universal genetic code except that the CGG codon at position 1213 was translated as tryptophan as proposed by Fox and Leaver (1981).

A DNA sequence rearrangement 5′ to the COX I gene in the CMS-S maize mitochondrial genome

To examine the genome organisation around the *COX* I gene in cytoplasmic male sterile lines of maize, a labelled M13 clone (fragment h in Figure 1) containing DNA from the coding sequence was hybridised to Southern blots of mtDNA from CMS-T, C and S lines cut with various restriction endonucleases (Figure 6). While the same sized restriction fragments were identified in the fertile (N) and CMS-C mtDNA, rearrangements near to the *COX* I gene were apparent in the CMS-T and S mtDNA. The rearrangement in the CMS-T genome is >1 kb from the 3′ terminus of the gene and has not been characterised further.

The rearrangement around the *COX* I region in the CMS-S mitochondrial genome is apparently more complicated (Figure 6) since several DNA fragments hybridise to the *COX* I-specific probe. Two fragments show strong hybridisation with the labelled

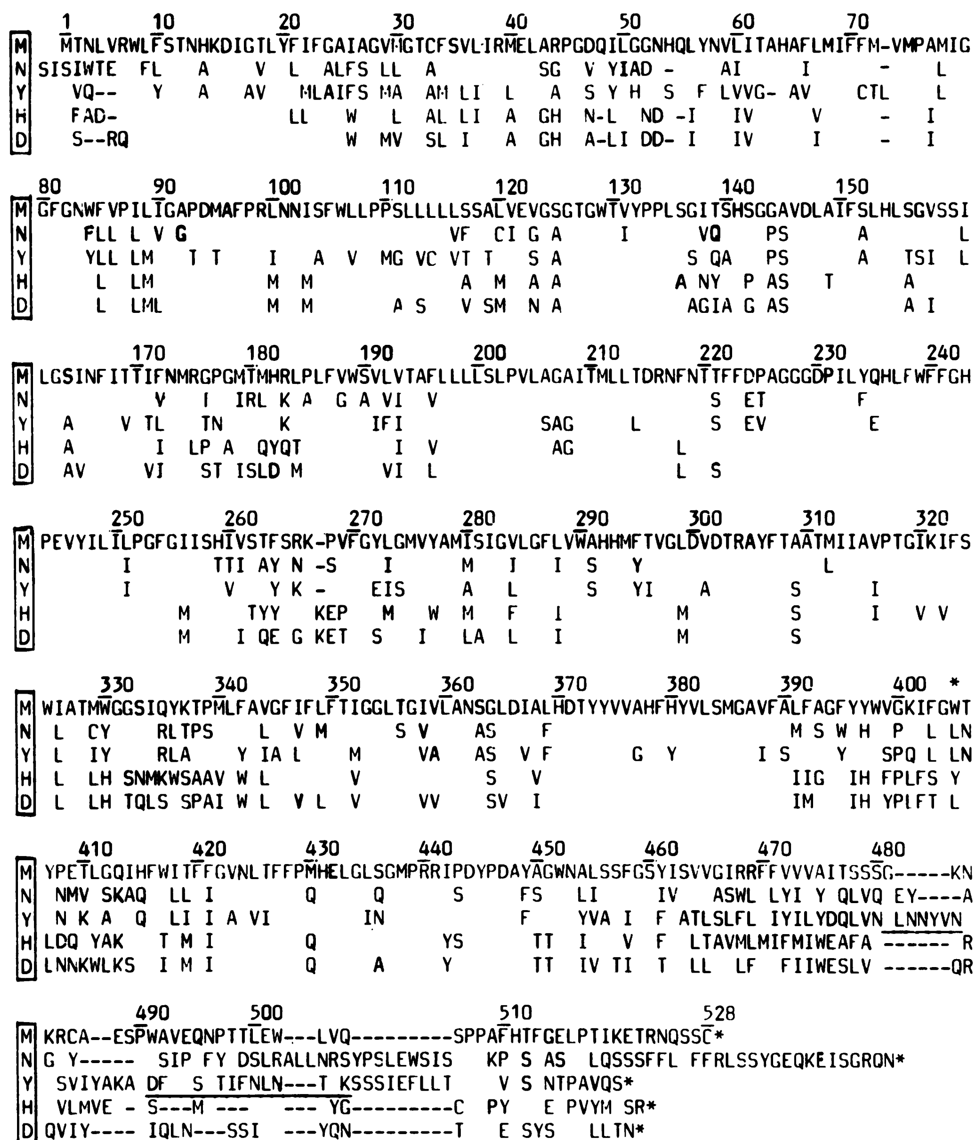

Fig. 3. Comparison of the predicted amino acid sequence of the maize cytochrome *c* oxidase subunit I with the predicted amino acid sequences of the homologous proteins from *N. crassa* (N), *S. cerevisiae* (Y), *H. sapiens* (H) and *D. melanogaster* (D). The numbering refers to the maize sequence and only this sequence is shown in full. The amino acids are only shown in the other sequences where they differ from maize. The two underlined portions of the yeast sequence represent the sequences originally thought to be small introns. The non-conserved carboxy ends of the sequences were aligned using the two conserved proline residues (positions 507 and 508) and the conserved histidine residue (position 511).

probe when CMS-S mtDNA is digested with either *Eco*RI or *Hind*III, together, in each case, with four fragments showing weaker hybridisation. When digested with *Bam*HI, however, a single fragment of 4.8 kb shows strong hybridisation, together with four less intensely hybridising fragments. The 4.8-kb *Bam*HI fragment is presumably contained within both the two major hybridising fragments generated by the other two enzymes.

An essentially similar pattern of hybridisation was obtained using either DNA probes encoding the amino terminus (fragment i in Figure 1) or carboxy terminus (fragment j in Figure 1) of the gene (results not shown) suggesting that the S genome contains at least six fragments of DNA containing the entire *COX* I gene in two different stoichiometries. The two fragments in CMS-S mtDNA showing strong hybridisation are apparently present in a similar abundance to the single fragment containing the gene in the fertile cytoplasm, while the other four fragments are present at much lower abundance.

The results presented here were obtained from plants with a B37 nuclear genotype. We have found that S type cytoplasms in other nuclear backgrounds, e.g., M825, 38-11 and WF9, the relative abundance of the restriction endonuclease-generated DNA fragments containing the *COX* I gene show characteristically different stoichiometries (Isaac *et al.*, in preparation). This suggests that the nucleus has a major role in determining the structure of the mitochondrial genome in CMS-S lines, perhaps by promoting specific recombination in sequences flanking the *COX* I gene. It also emphasises the importance of the interaction between nuclear and mitochondrial genomes during mitochondrial biogenesis. The CMS phenotype apparently results from the introduction of a cytoplasm into a foreign nuclear background and may be due to changes in mitochondrial genome organisation and/or expression.

These results show that sequences around the *COX* I gene in CMS-S mtDNA are arranged in at least six different permutations. To investigate the nature of these rearrangements a *Hind*III library of CMS-S maize mtDNA (isolated from the line B37S)

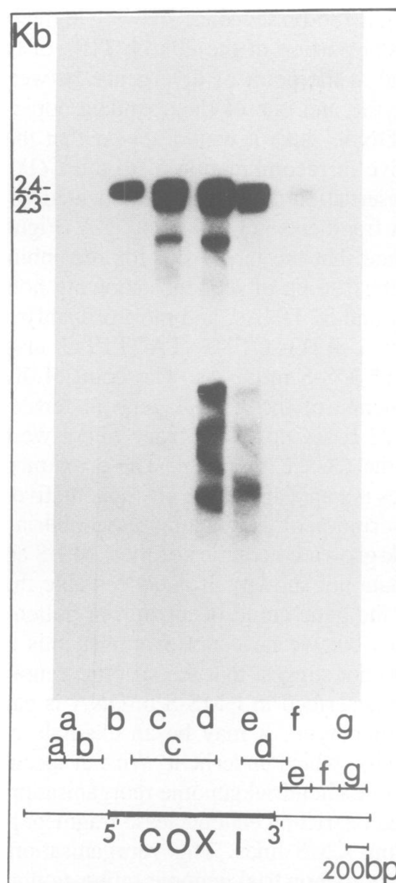

Fig. 4. Transcripts of the maize cytochrome *c* oxidase subunit I gene. Total maize mtRNA was fractionated by electrophoresis through 1.3% (w/v) agarose-formaldehyde gels, transferred to nitrocellulose and parallel tracks were hybridised with individual ^{32}P-labelled M13 DNA subclones from the *COX* I coding and flanking sequences. The origins of the M13 clones a–g are indicated in the lower part of this Figure and in Figure 1. (The darkening of the film in the middle of the Figure represents non-specific hybridisation.)

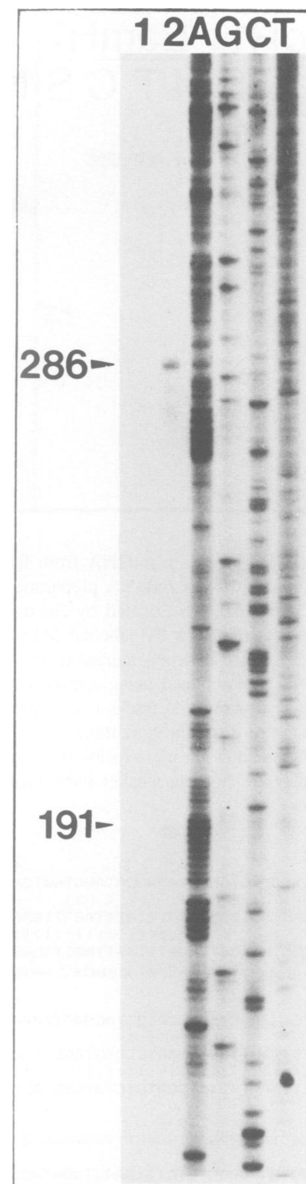

Fig. 5. Determination of the 5′ ends of the *COX* I RNA transcripts by S1 nuclease mapping. Hybrids were formed with mtRNA and a ^{32}P-labelled M13 mtDNA probe (*Msp* fragment; in Figure 1) which contained the start of both transcripts. S1 nuclease-resistant hybrids were fractionated in a 6% (w/v) sequencing gel. **Lanes (1)** input DNA plus tRNA, **(2)** hybrids treated with 400 U/ml S1 nuclease. **Lanes A, G, C** and **T**, M13 sequencing ladder. The apparent size of the protected fragments are marked.

was probed with a *COX* I internal gene probe (fragment h in Figure 1). Plasmid DNA was isolated from a single colony identified by the probe and shown to contain a DNA insert of ~9.7 kb. This corresponded to the larger of the two major *Hind*III fragments identified by hybridisation in Figure 6, and contained the 4.8-kb *Bam*HI fragment identified in CMS-S mtDNA. Sequence analysis of the 4.8-kb *Bam*HI fragment revealed that it contained the entire sequence of the *COX* I gene and that the nucleotide sequence was identical to that in N mtDNA from the 3′ *Bam*HI site to a point 174 bp 5′ to the initiation codon. The sequences 5′ to nucleotide −174 diverge in N and CMS-S mtDNAs and following a single inserted 'A' residue (relative to the common sequence in N mtDNA) the next 186 nucleotides in CMS-S mtDNA are identical to a portion of the free end of the 208-bp TIR of the S1 and S2 DNA elements (Levings and Sederoff, 1983) (Figure 7). The sequence 5′ to the S1 and S2 DNA homolgy in CMS-S mtDNA, shows no homology to other regions of the S1 and S2 DNAs, nor to any other known maize mtDNA sequence (data not shown). At the point where rearrangement has occurred in CMS-S mtDNA there is a short sequence, 5′-ATG--CTTG--TACTTTT-3′ (see Figure 7), in N mtDNA, showing partial sequence homology to the distal end of the TIR of S1 and S2. As noted by Levings and Sederoff (1983) the terminal 5′-TACTTTT-3′ of the TIRs shows a high degree of homology with the short TIRs found in adenovirus and in a range of linear *Bacillus* phages.

Schardl *et al.* (1984) have demonstrated that a high proportion of the mitochondrial DNA in CMS-S maize is present as linear molecules. They postulate that linearisation occurs as a result of recombination between the usual circular mitochondrial chromosome and the S1 and S2 linear DNAs. The recombination occurring between sequences in the CMS-S mitochondrial genome and homologous sequences in the TIRs of the S1 and S2 DNAs, results in the production of linear molecules with S1 and S2 covalently linked at the end. The sequence data presented in Figure 7 probably represent such a target site for homologous recombination. Those DNA fragments which hybridise faintly with a *COX* I gene probe may well represent such linearised molecules present in lower abundance than those fragments which hybridise more strongly. It should be noted that neither the free

EcoR1 | BamH1 | HindIII
Kb N T C S | N T C S | N T C S
9.42
6.56
4.38
2.32

Fig. 6. Location of the *COX* I gene in mtDNA from fertile (N), and CMS-T, C and S maize cytoplasms. Total mtDNA preparations were cleaved with either *Eco*RI, *Bam*HI or *Hind*III, fractionated by electrophoresis, blotted onto nitrocellulose and probed with a ^{32}P-labelled M13 clone (fragment h in Figure 1) containing a sequence entirely internal to the *COX* I gene. An essentially similar pattern was obtained using amino- and carboxy-terminal-specific gene probes (not shown). The tracks containing CMS-S mtDNA contain labelled fragments of different stoichiometries. One (*Bam*HI) or two (*Eco*RI, *Hind*III) fragments show strong hybridisation to the probe, and four fragments from each enzyme showing weaker hybridisation (indicated by dots) can be seen.

```
      *-380             *-360             *-340
(1) GAATCACAGCCTGACTTCCTCTAGATGAAAGCATAGGTAGTGACGACCTATTGTTTTCTT
(2) TTTTGTTTCAGTTTCACTATTGGTTGGATTTATTTAGTTTGAGACATGTAAGGACTTCTC
(3) TTTTTTGTGTGTCTTCTTGCAGGTTGGATTTATTTAGTTTGAGACATGTAAGGACTTCTC
    Proximal end S episome terminal inverted repeat

      *-320             *-300             *-280
(1) CTCAGCGGACCCAATGAGTAACTACGCCATGTTACGAATAAAAGCGAGGGAAAATCGGAT
(2) CGGAGAGGTATTCTCTGCAGTTCTGTCCATCTCTGTCGGCACCATTGTTATCTGTCTCTT
(3) CGGAGAGGTATTCTCTGCAGTTCTGTCCATCTCTGTCGGCACCATTGTTATCTGTCTCTT

      *-260             *-240             *-220
(1) GAGCAATCTTTTGCAGCTATATGAAATGCACTTTGGAAAATCAGGTCTTGAGGACGAAAA
(2) AATAGTAGAGGATCGCCGACGCCCTGTCTTTGAAGTAGATACCTGGTTGGTATCTTTATG
(3) AATAGTAGAGGATCGCCGACGCCCTGTCTTTGAAGTAGATACCTGGTTGGTATCTTTATG

      *-200             *-180
(1) ACACGATAAAATGGACTTGCCTACTTTTGCACC<-N-fertile
                                     <---->170 bases 5'to COX I
(2) TAGATTGGACATGTGCTTGTATACTTTAGCACC<-CMS-S
(3) TAGATTGGACATGTGCTTGTATACTTTT<--end S episome
```

Fig. 7. Comparison of DNA sequence homology 5′ to the *COX* I gene around the point of divergence between mtDNAs from N-fertile **(1)** and CMS-S **(2)**, maize cytoplasms and the 208-bp TIR of S1 and S2 DNAs **(3)**. The numbering refers to the number of bases 5′ to the ATG initiation codon of the *COX* I gene in mtDNA from the N-fertile maize line. Homology between a sequence in N mtDNA and the distal end of the S TIR around the point of divergence is underlined.

S1 nor S2 molecules contain sequences which hybridise with any region of the *COX* I gene (data not shown).

The biological significance of the sequence rearrangement 5′ to the *COX* I gene and increased copy number of the gene in the CMS-S maize mitochondrial genome is unknown. These changes are however associated with the presence of the S1 and S2 linear DNA molecules characteristically found in mitochondria from CMS-S cytoplasms. The involvement of sequences related to these linear DNAs in the process leading to the rearrangement of the mitochondrial genome is supported by our observation that a 186-bp sequence showing uninterrupted homology to the distal portion of the 208-bp TIRs of the S1 and S2 DNAs is found at the point of divergence between the *COX* I gene of the fertile and one of the abundant copies of the gene in CMS-S mtDNA. Thus it would appear that the presence of sequences active in recombination 5′ to the *COX* I gene could explain our observation that the gene is located on a number of different DNA fragments in CMS-S mtDNA (Figure 6). We can further speculate that the target site for recombination is contained within the 186 bp of sequence showing homology to the TIRs of the S1 and S2 DNAs, and most probably in the vicinity of the sequence 5′-ATG--CTTG--TACTTT-3′ also found in the fertile (N) and CMS-S mtDNAs. The point of divergence between the sequence of the *COX* I gene in fertile and CMS-S mtDNA is ~23 bases upstream from where we have mapped the 5′ end of the *COX* I transcript. The proximity of this rearrangement does not appear to have any quantitative or qualitative effect on transcription of the gene in mitochondria isolated from 4-day-old dark-grown coleoptiles of the CMS-S compared with fertile lines (data not shown). It is conceivable that differential expression of the gene could occur during pollen development in CMS-S lines but we have not examined this as yet.

While we do not suggest that the specific genome rearrangement we have described in CMS-S mtDNA is causally related to the CMS phenotype, it may be an example of the type of molecular event(s) which underlie it. We can speculate that these events lead to mitochondrial genome reorganisation, as indicated by the observed restriction endonuclease fragment polymorphism in mtDNA from CMS lines. This reorganisation in turn may directly affect mitochondrial genome replication and/or expression. Support for this hypothesis is provided by our earlier observations that characteristic variant translation products are synthesised by mitochondria from each of the different male sterile lines (Forde and Leaver, 1980; Leaver and Gray, 1982). Rearrangements of the kind we have described could affect mitochondrial genome replication or expression resulting in a failure or deficiency in mitochondrial function or biogenesis during the critical period of mitochondrial biogenesis which occurs in both the tapetum and microspores during pollen formation.

Materials and methods

Isolation of mitochondrial DNA and RNA

Mitochondria were isolated from 4-day-old dark-grown maize coleoptiles (lines B37N, T, C and S, Pioneer Hi-Bred International, Des Moines, IA) and DNA and RNA extracted as described previously (Fox and Leaver, 1981).

Restriction, electrophoresis, blotting and hybridisation of DNA

Restriction enzymes were obtained from Bethesda Research Laboratories and used as specified by the supplier. Electrophoresis was carried out in 0.8% (w/v) agarose gels containing 40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, adjusted to pH 8 with acetic acid. The separated fragments were transferred to nitrocellulose (Southern, 1975) after depurination (Wahl *et al.*, 1979).

The plasmid pKLD which contains exons 3 and 4, intron 3 and part of intron 2 of *S. cerevisiae oxi*-3 gene (Hensgens *et al.*, 1983), was used to identify the maize mitochondrial DNA fragments containing the *COX* I gene by heterologous hybridisation. These heterologous hybridisations were carried out in 0.6 M NaCl, 60 mM Na-citrate (4 x SSC), 0.1% (w/v) SDS, 0.2% (w/v) bovine serum albumin, 0.2% (w/v) Ficoll (type 400, Sigma), 0.2% (w/v) polyvinyl pyrrolidone (mol. wt. 40 000) and 200 μg/ml sonicated, denatured herring sperm DNA (Serva Fine Biochemicals). Filters were pre-hybridised at 45°C in the above buffer for at least 30 min before pKLD DNA labelled with ^{32}P was added (Rigby *et al.*, 1977). Filters were hybridised for 18 h at 45°C and then washed in three changes of 0.6 M NaCl, 60 mM Na-citrate, 0.1% (w/v) SDS at 52°C. Homolgous hybridisations were carried out at 37°C in 50% (v/v) formamide, 4 x SSC 0.02% (w/v) bovine serum albumin, 0.02% (w/v) Ficoll, 0.02% (w/v) polyvinyl pyrrolidone, 50 μg/ml, sonicated denatured herring sperm DNA.

Cloning and sequencing of the COX I gene

10 μg of maize mtDNA was digested with the restriction enzyme *Bam*HI and fractionated by agarose gel electrophoresis. The area of the gel containing fragments of ~12 kb (i.e., which contained fragments hybridising to the yeast *oxi*-3 DNA probe) was excised, and the DNA eluted using the 'freeze-squeeze' method of Thuring *et al.* (1975). The DNA fragments were ethanol precipitated and one-third of the preparation was ligated into the *Bam*HI site of pAT153 (Twigg and Sherratt, 1980). The ligated DNA was transformed into the *Escherichia coli* strain HB101, and ampicillin-resistant colonies were selected. DNA was prepared (Birnboim and Doly, 1979) from those colonies which were tetracycline sensitive.

An initial restriction map was prepared of the 3.9-kb *Eco*RI to *Bam*HI subfragment of the original 12-kb fragment using the method of Smith and Birnstiel (1976) (Figure 1). Subclones of the fragment were made in the replicative form of M13mp8 (Messing and Vieira, 1982). Single-stranded isolates of the phages were sequenced by the extension dideoxyribonucleotide termination procedure (Sanger *et al.*, 1977,1980) using [^{32}P]dCTP (410 Ci/mM, Amersham International), except that a synthetic pentadecamer primer (New England Biolabs) was used. The sequencing strategy used is shown in Figure 1.

RNA transcript analysis and S1 nuclease mapping

Northern blot analysis was performed as described previously (Dawson *et al.*, 1984). The 5′ ends of transcripts were mapped using S1 nuclease (Berk and Sharp, 1977). The single-stranded M13 clone used as a probe was labelled by a modification of the sequencing reaction (Hu and Messing, 1982), using the pentadecamer sequencing primer and the strands were separated prior to hybridisation. The labelled DNA probe was hybridised with 20 μg total mitochondrial RNA at 43°C for 3 h in 80% (v/v) formamide after which S1 nuclease was added to 400 U/ml and digestion allowed to proceed for 30 min at 37°C. The protected hybrid fragments were fractionated by electrophoresis on 6% (w/v) polyacrylamide-urea sequencing gels in parallel with m13 sequence ladders to provide size markers.

Acknowledgements

We are grateful to Professor L.A. Grivell and Dr. L.A.M. Hensgens of the University of Amsterdam for advice and the gift of the yeast gene probe for *oxi*-3 and to Pioneer Hi-Bred International Inc. for supplying maize seed. This work was supported by a grant from the Agricultural and Food Research Council to C.J. Leaver and an S.E.R.C. post-graduate studentship to V.P. Jones.

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. and Young,I.G. (1981) *Nature*, **290**, 457-464.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721-732.

Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.

Bonitz,S.C., Coruzzi,G., Thalenfeld,B.E., Tzagoloff,A. and Macino,G. (1980) *J. Biol. Chem.*, **255**, 11927-11941.

Burger,G., Scriven,C., Machliedt,W. and Werner,S. (1982) *EMBO J.*, **1**, 1385-1391.

Dawson,A.J., Jones,V.P. and Leaver,C.J. (1984) *EMBO J.*, **3**, 2107-2113.

de Bruijn,M.M.L. (1983) *Nature*, **304**, 234-241.

de Jonge,J.C. and de Vries,H. (1983) *Curr. Genet.*, **7**, 21-28.

Forde,B.G. and Leaver,C.J. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 418-422.

Fox,T.D. and Leaver,C.J. (1981) *Cell*, **26**, 315-323.

Hensgens,L.A.M., Arnberg,A.C., Roosendal,E., Van der Horst,G., Van der Veen,R., Van Ommen,G.-J.B. and Grivell,L.A. (1983) *J. Mol. Biol.*, **164**, 35-38.

Hu,N.-T. and Messing,J.C. (1982) *Gene*, **17**, 271-277.

Kemble,R.J. and Thompson,R.D. (1982) *Nucleic Acids Res.*, **10**, 8181-8190.

Leaver,C.J. and Gray,M.W. (1982) *Annu. Rev. Plant Physiol.*, **33**, 373-402.

Levings,C.S.,III and Sederoff,R.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4055-4059.

Levings,C.S.,III, Kim,B.D., Pring,D.R., Conde,M.F., Mans,R.J., Laughnan,J.R. and Gabay-Laughnan,S.J. (1980) *Science (Wash.)*, **209**, 1021-1023.

Lonsdale,D.M. (1984) *Plant Mol. Biol.*, **3**, 201-206.

Lonsdale,D.M., Thompson,R.D. and Hodge,T.P. (1981) *Nucleic Acids Res.*, **9**, 3657-3669.

Lonsdale,D.M., Hodge,T.P. and Fauron,C.M.-R. (1984) *Nucleic Acids Res.*, **12**, 9249-9261.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.

Osinga,K.A. and Tabak,H.F. (1982) *Nucleic Acids Res.*, **10**, 3617-3626.

Osinga,K.A., De Vries,E., Van der Horst,G.T.J. and Tabak,H.F. (1984) *Nucleic Acids Res.*, **12**, 1889-1900.

Palmer,J.D. and Shields,C.R. (1984) *Nature*, **307**, 437-440.

Pring,D.R. and Levings,C.S.,III (1978) *Genetics*, **89**, 121-136.

Pring,D.R., Levings,C.S.,III, Hu,W.W.L. and Timothy,D.H. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2904-2908.

Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **113**, 237-251.

Sanger,F., Air,G.M., Barrell,B.G., Brown,N.L., Coulson,A.R., Fiddes,J.C., Hutchinson,C.A., Slocombe,P.M. and Smith,M. (1977) *Nature*, **265**, 687-695.

Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.

Saraste,M. and Wikstrom,M. (1983) in Quagliariello,E. and Palmieri,F. (eds.) *Structure and Function of Membrane Proteins*, Elsevier Science Publications, B.V. Amsterdam, pp. 139-144.

Schardl,C.L., Lonsdale,D.M., Pring,D.R. and Rose,K.R. (1984) *Nature*, **310**, 292-296.

Smith,H.D. and Birnsteil,M.L. (1976) *Nucleic Acids Res.*, **3**, 2387-2398.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.

Thuring,R.W.J., Sanders,J.P.M. and Borst,P. (1975) *Anal. Biochem.*, **66**, 213-220.

Twigg,A. and Sherratt,D. (1980) *Nature*, **283**, 216-218.

Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683-3687.

Received on 9 April 1985; revised on 29 April 1985