Plant Molecular Biology **25**: 91–103, 1994.

Differential screening of mitochondrial cDNA libraries from male-fertile and cytoplasmic male-sterile sugar-beet reveals genome rearrangements at *atp6* and *atpA* loci

Yongbiao Xue, Sylvie Collin, D. Roy Davies and Colwyn M. Thomas [1,*]
AFRC Institute of Plant Science Research, John Innes Research Centre, Colney Lane, Norwich, NR4 7UH, UK; [1] *Present address: The Sainsbury Laboratory, John Innes Research Centre, Colney Lane, Norwich NR4 7UH, UK (*author for correspondence)*

Received 12 August 1993; accepted in revised form 18 February 1994

Key words: atpA, atp6, genome rearrangements, cytoplasmic male sterility

Abstract

As part of a strategy to define differences in genome organization and expression between cytoplasmic male-sterile (CMS) and male-fertile (MF) sugar-beet mitochondria, cDNA libraries from both mitochondrial genotypes were constructed. Preliminary screening with ribosomal RNA gene probes identified candidate cDNA clones corresponding to structural genes. In addition, reciprocal hybridization experiments were performed using labelled first-strand cDNA to identify uniquely transcribed sequences. One cDNA clone (pYC700) is unique to CMS mitochondria and is located upstream of the F_oF_1-ATPase subunit 6 gene (*atp6*). Another cDNA clone (pYC130), when used as a probe in northern hybridization analysis, revealed novel transcript profiles in CMS sugar-beet mitochondria. Sequence analysis of this cDNA showed strong homology with the F_oF_1-ATPase subunit α (*atpA*) coding sequences from several higher plants. The *atp6* and *atpA* loci from each genotype were cloned and the genomic organization, DNA sequence and transcription of each locus was studied. Differences in the transcript profiles of each gene are a consequence of genomic rearrangements 5′ to the coding sequence.

Introduction

Cytoplasmic male sterility (CMS) is a maternally inherited trait widely exhibited in higher plants and is routinely used in the large scale production of F_1 hybrid seed in several crops including sugar-beet. Molecular analysis of CMS from several plant species has suggested that genes responsible for CMS reside in mitochondrial DNA (mtDNA) [12]. The detailed analysis of CMS in *Zea mays* CMS-T and *Petunia hybrida* has led to the identification of CMS-inducing genes. A novel mitochondrial gene (*T-urf13*) in maize CMS-T [9] and a chimeric mitochondrial gene (S-*pcf*) in somatic hybrid plants of *Petunia* [38] are associated with CMS. These genes are either deleted [7, 26] or their transcription altered in plants restored to fertility [17, 25, 34]. In addition, chimeric sequences resulting from inter- and intramolecular recombinational events are associated with CMS

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X55076 (MF *atp6*, X54722 (CMS *atp6*), X68690 (MF *atpA*) and X68691 (CMS *atpA*).

in many other plant species, for example, sorghum, common bean, sunflower, 'Ogura' radish and rapeseed [12] and rice [15]. In most cases there is no definitive evidence to support the suggestion that these genes are involved in expression of the CMS phenotype.

All commercially used CMS lines in sugar-beet are thought to contain the Owen type CMS which arose spontaneously in a natural population [23]. Differences in the genomic organization and transcription of CMS and male-fertile (MF) mtDNA have been identified by restriction enzyme mapping [24], heterologous probing [10] and limited DNA sequence analysis [30, 31]. In addition, restriction enzyme mapping of overlapping cosmid clones of the mitochondrial genome of CMS sugar-beet has revealed a complex multicircular organization generated by homologous recombination across repeated DNA sequences derived from a putative 'master chromosome' of 386 kb [4]. In order to define a possible molecular basis for CMS in sugar-beet we have constructed cDNA libraries from CMS and MF lines. Labelled first-strand cDNAs have been used in reciprocal hybridization experiments to identify uniquely transcribed loci in each mitochondrial genotype. One cDNA clone (pYC700) was derived from a uniquely transcribed sequence in CMS mitochondria and is located upstream of the F_oF_1-ATPase complex subunit 6 gene (*atp6*). A second cDNA clone (pYC130) when used as a probe in northern hybridization analysis detected different transcription patterns between CMS and MF mitochondria. This clone is derived from the gene encoding the α-subunit of the F_oF_1-ATPase complex (*atpA*). Both loci were cloned from each mitochondrial genotype and subsequently characterized by northern and Southern hybridization analysis and DNA sequencing.

Materials and methods

Isolation of mitochondrial nucleic acids

MtDNA and mtRNA were isolated from the sugar-beet lines I 13M4 (MF) and 01 I 13M4 (CMS) as described previously [33]. Southern and northern hybridization analysis and S1 nuclease mapping were also performed as described [33].

Preparation of ^{32}P-labelled probes

Plasmid DNA and gel-purified DNA fragments were labelled by nick translation for Southern and northern hybridizations as described by Sambrook *et al.* [27]. First-strand cDNAs used for reciprocal screening of mt cDNA libraries were synthesized as follows. Five μg of mtRNA was incubated in a reaction volume of 50 μl containing 50 mM Tris-HCl pH 8.3, 45 mM KCl, 10 mM $MgCl_2$, 20 mM DTT, 7.5 μg d(N)6 oligomer primer, 1.5 μg actinomycin D, 1 mM each of dATP, dGTP and dTTP, 40 μCi α^{32}P-dCTP and 10 U AMV reverse transcriptase (Pharmacia) for 1 h at 42 °C. The unincorporated label was removed by chromatography through Sephadex G-50.

Preparation of mitochondrial cDNA and genomic libraries

Prior to cDNA synthesis purified mtRNAs were polyadenylated. Five μg of mtRNA was incubated in a reaction volume of 50 μl containing 50 mM Tris-HCl pH 7.9, 10 mM $MgCl_2$, 2.5 mM $MnCl_2$, 50 mM NaCl, 250 mM ATP, 500 μg/ml BSA, 20 μCi ^{3}H-ATP (40 Ci/mmol) and 4 U poly(A) polymerase (BRL) for 2 h at 37 °C. Polyadenylated mtRNA was purified by two phenol/chloroform extractions and ethanol-precipitated. The extent of polyadenylation was estimated by monitoring AMP incorporation into the mtRNA sample. Synthesis of cDNAs was performed using a kit essentially as described by the manufacturer (Amersham). The cDNA products were blunt-ended by treatment with T4 DNA polymerase and cloned into *Sma* I-digested pUC19 [27].

Mitochondrial genomic libraries were constructed by ligating restriction enzyme-digested

mtDNAs (*Eco* RI, *Bam* HI and *Hin*d III) with the appropriately digested pUC19 vector DNA which had been previously treated with calf alkaline phosphatase. All ligated DNAs were used to transform *Escherichia coli* JM109 [27]. Recombinant clones were screened by the *lac* complementation assay [27].

Reciprocal screening of cDNA libraries was carried out by *in situ* colony hybridization using Hybond N membrane filters. The filters were first hybridized with the CMS cDNAs and then reprobed with the MF cDNAs after removal of the CMS probe according to the manufacturer (Amersham). The cDNA libraries were also hybridized with nuclear, chloroplast and mitochondrial rRNA gene probes. Wheat mitochondrial 5S and 18S rRNA genes and 26S rRNA genes were kindly provided by Dr D. Falconet. Two clones containing pea nuclear rRNA genes (26S, 18S and 5.8S) or barley chloroplast rRNA genes were supplied by Dr T.H.N. Ellis.

DNA sequence analysis

DNA sequences were determined on both strands by the dideoxy terminator technique after subcloning into the phagemid pBluescript vector (Stratagene). Ordered deletions of large restriction fragments were created with exonuclease III [13]. Single-stranded templates were isolated after superinfection with the helper phage K07.

Results

Analysis of cDNA libraries from CMS and MF mitochondria

Higher-plant mtRNAs lack a poly(A) tail to facilitate oligo dT primed cDNA synthesis. Therefore, total mtRNAs from each genotype were first polyadenylated prior to cDNA synthesis and cloning as described in Materials and methods. The size of cDNA inserts ranged from 0.1 to 1.5 kb (results not shown). Screening of 500 randomly selected cDNA clones with rRNA gene probes of nuclear, plastid and mitochondrial origin indicated that about 50 of cDNA clones were non-ribosomal (results not shown). Clones which did not hybridize to rRNA gene probes were ^{32}P-labelled and used sequentially to probe Southern and northern blots of both mitochondrial genotypes to detect differences in genomic organization or transcript profiles. One cDNA clone, pYC130, containing a 143 bp insert, hybridized to *Hin*d III fragments which distinguished each genotype (results not shown). In northern hybridization analysis, genotypic differences in transcript profiles were observed (Fig. 1). Sequence analysis revealed that the cDNA clone corresponds to an internal region of the *atpA* gene (see Fig. 9).

Additionally, clones unique to each mitochondrial genotype were detected by reciprocal screening of 1500 randomly selected clones from each cDNA library. One clone hybridized specifically to CMS first-strand cDNAs but not MF cDNAs. The cDNA clone (pYC700) hybridized to two major transcripts of 3.3 kb and 3.1 kb from CMS mtRNAs (Fig. 2A). Southern hybridization analysis showed that pYC700 hybridized exclusively to CMS mtDNA (Fig. 2B). When pYC700 was used to probe *Sma* I digests of cosmid clones derived from CMS mtDNA it hybridized to a

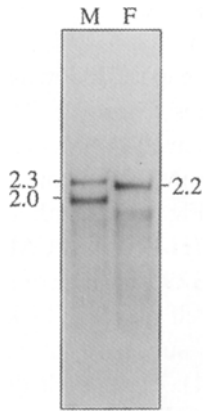


Fig. 1. Northern hybridization analysis of CMS and MF mtRNA using pYC130 as a probe. Four μg of mtRNA from MF (F) or CMS (M) genotype were loaded in each lane and electrophoresed on a 1.4% agarose gel containing 6.6% formaldehyde. The sizes of transcripts are indicated in kb.

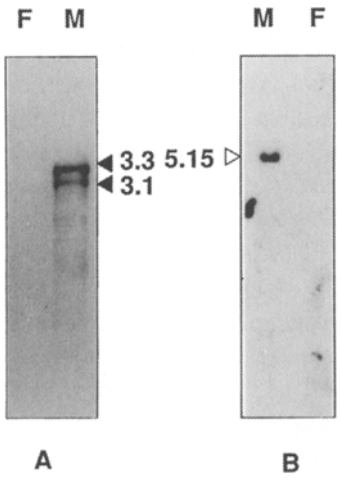


Fig. 2. Northern and Southern hybridization analysis of CMS and MF mitochondrial nucleic acids using pYC700 as a probe. A. 4 μg of mtRNA from CMS (M) or MF (F) mitochondria were loaded in each lane and electrophoresed on a 1.4% agarose gel containing 6.6% formaldehyde. B. Mitochondrial DNA (1 μg) from MF (F) and CMS (M) mitochondria were digested with *Eco* RI and electrophoresed on a 0.8% agarose gel. The size of transcripts and DNA fragments are indicated in kb and kbp respectively.

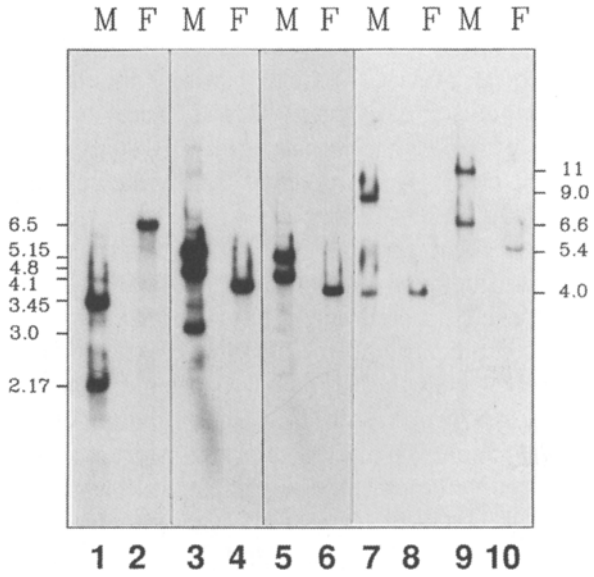


Fig. 3. Southern hybridization analysis showing that sequences on pMB796 are reiterated in CMS mitochondria. DNAs (1 μg each lane) from CMS (M) or MF (F) mitochondria were restriction enzyme digested and electrophoresed on 0.8% agarose gels. Lanes 1 and 2, *Bam* HI-digested DNAs probed with pMB796; Lanes 3 and 4, *Eco* RI-digested DNAs probed with pMB796; lanes 5 and 6, *Eco* RI-digested DNAs probed with pMB358; lanes 7 and 8, *Sma* I-digested DNAs probed with pMB796; lanes 9 and 10, *Xho* I-digested DNAs probed with pMB796. The sizes of DNA fragments are indicated in kb.

4 kb *Sma* I fragment encompassing the *atp6* gene (T. Brears and D.M. Lonsdale, personal communication).

Molecular cloning of the atp6 *locus from CMS and MF mitochondria*

The cDNA clone pYC700 was used to isolate overlapping clones from *Eco* RI and *Bam* HI libraries of CMS mtDNA. One clone from the *Bam* HI library (pMB796) contained an insert of 3.45 kb (see Fig. 4). Southern hybridization analysis of *Bam* HI-digested CMS and MF mtDNAs using pMB796 as a probe showed that it hybridized to itself and a 2.17 kb *Bam* HI fragment which was subsequently cloned (pMB358, see Figs. 3 and 4). This result demonstrates that sequences in pMB796 are repeated in CMS mtDNA. Only single *Bam* HI and *Eco* RI fragments of 6.5 kb and 4.1 kb respectively were detected in MF mtDNA (Fig. 3). Southern hybridization analysis of *Eco* RI digested CMS mtDNA showed that pMB796 hybridized to three fragments of 3.0, 4.8 and 5.15 kb (Fig. 3). The clone pMB358 also hybridized to two of these fragments (4.8 and 5.15 kb) due to the presence of the repeat sequence. However, only the 5.15 kb *Eco* RI fragment hybridized to pYC700 (results not shown) demonstrating that the 3.0 kb and 5.15 kb *Eco* RI fragments share homologous sequences with pMB796; these fragments were subsequently cloned (pMR115 and pMR397). Restriction enzyme mapping and Southern hybridization analysis were used to construct a locus map (Fig. 4).

Cosmid mapping of the mitochondrial genome of CMS sugar-beet has demonstrated that *atp6* and *atp9* loci are flanked by a repeat sequence, repeat II [4]. When pMB796 was used as a probe to *Sma* I- and *Xho* I-digested DNAs, fragments of a similar size to those previously reported at

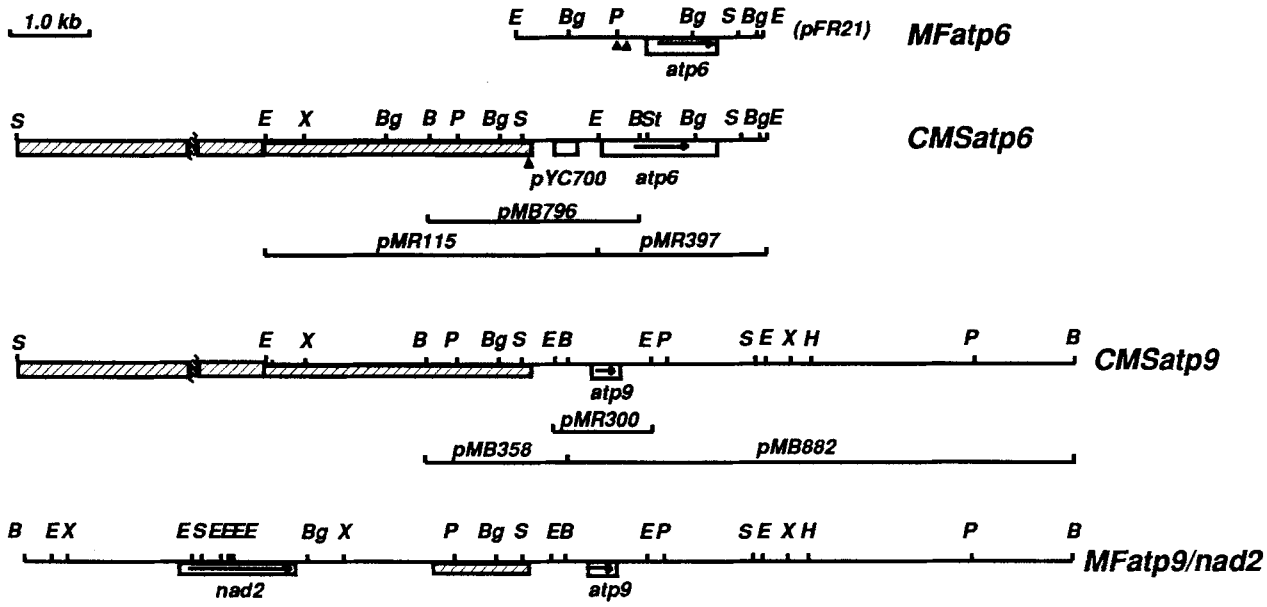


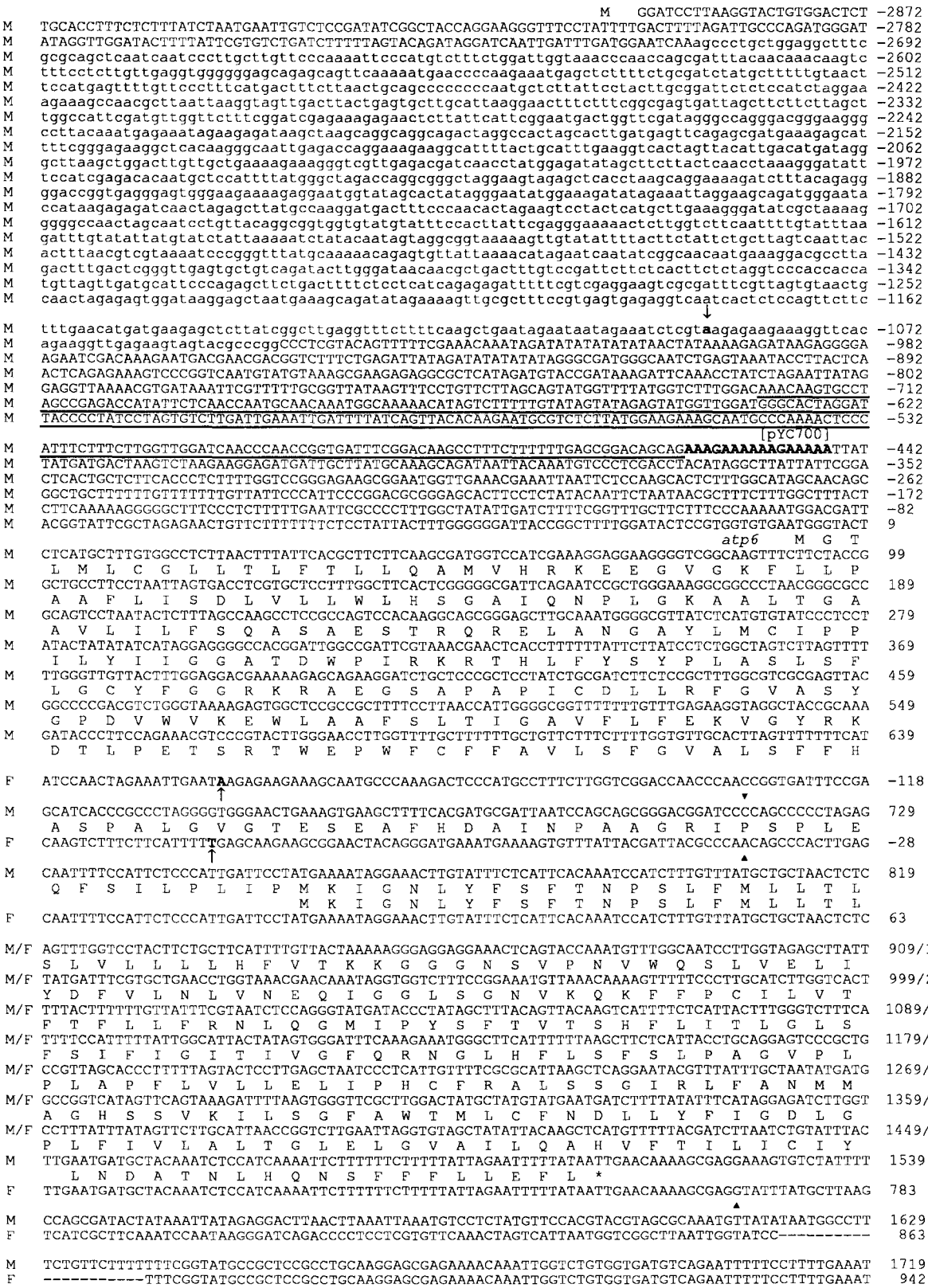
Fig. 4. Genomic organization of the *atp6* and *atp9* loci in CMS and MF sugar-beet mitochondria. The location and extent of repeat II [4] upstream of CMS *atp6* and CMS *atp9* is represented by a broken hatched box. The full extent of repeat II is not shown to scale here. The extent of repeat II homologous sequences upstream of the MF *atp9* locus is also shown by a hatched box. Clones isolated from CMS and MF mtDNA to construct maps are shown below the appropriate locus. The location of *atp6*, *atp9* and *nad2* genes relative to repeat II are shown and the direction of transcription indicated by arrows. The mapped 5′ termini of *atp6* are indicated by triangles (▲). The location of pYC700 homologous sequences upstream of the CMS *atp6* gene are indicated along with restriction enzyme sites employed in mapping; *B*, *Bam* HI; *Bg*, *Bgl* II; *E*, *Eco* RI; *H*, *Hin*d III; *P*, *Pst* I; *S*, *Sma* I; *X*, *Xho* I.

atp6 and *atp9* loci were observed (Fig. 3 [4]). In CMS mtDNA two *Sma* I fragments of 9.0 kb and 4.0 kb were detected. The latter fragment hybridizes with pYC700 and encompasses the *atp6* gene (see below). Analysis of cosmid clones suggested that repeat II extends for ca. 4.0 kb [4]; however, pMB796 hybridizes approximately twice as strongly to the *Sma* I fragment of 9.0 kb (Fig. 3), suggesting repeat II is larger than reported and extends for at least 9.0 kb upstream of *atp6* and *atp9* loci. Sequence analysis has confirmed the presence of repeat II upstream of CMS and MF *atp9* genes ([36], see below) and downstream of the MF *nad2* locus. The MF *atp6* locus was isolated on a 4.1 kb *Eco* RI fragment (pFR21) using an *Eco* RI clone from CMS mtDNA (pMR397) as a hybridization probe. The organization of *atp6* and *atp9* loci in CMS and MF mitochondria is summarized in Fig. 4.

DNA sequence analysis of the atp6 *locus in CMS and MF sugar-beet*

The DNA sequence of MF *nad2* and *atp9* loci have been reported previously [36, 37]. The DNA sequence of ca. 5.8 kb of the CMS *atp6* locus extending from the *Bam* HI site in repeat II to the *Eco* RI site 1.3 kb downstream of the *atp6* gene, was determined as described in Materials and methods. The DNA sequence of pFR21 (Fig. 4) encompassing ca. 4.0 kb of the MF *atp6* locus was also determined for comparison. The relevant sequence information is summarized in Fig. 5.

When both sequences were translated each showed extensive homology to ATP6 from the maize [8], soybean [11], tobacco [2], *Oenothera* [29], wheat [3], radish [18, and EMBL accession number M24672], sorghum [16] and rice [15] counterparts. The deduced MF ATP6 protein


M GGATCCTTAAGGTACTGTGGACTCT -2872
M TGCACCTTTCTCTTTATCTAATGAATTGTCTCCGATATCGGCTACCAGGAAGGGTTTCCTATTTTGACTTTTAGATTGCCCAGATGGGAT -2782
M ATAGGTTGGATACTTTTATTCGTGTCTGATCTTTTTAGTACAGATAGGATCAATTGATTTGATGGAATCAAagccctgctggaggctttc -2692
M gcgcagctcaatcaatcccttgcttgttcccaaaattcccatgtctttctggattggtaaacccaaccagcgatttacaacaaacaagtc -2602
M tttcctcttgttgaggtgggggagcagagcagttcaaaaatgaaccccaagaaatgagctctttttctgcgatctatgctttttgtaact -2512
M tccatgagttttgttccctttcatgacttttcttaactgcagccccccccaatgctcttattcctacttgcggattctctccatctaggaa -2422
M agaaagccaacgcttaattaaggtagttgacttactgagtgcttgcattaaggaactttctttcggcgagtgattagcttcttcttagct -2332
M tggccattcgatgttggttctttcggatcgagaaagagaactcttattcattcggaatgactggttcgatagggccagggacgggaaggg -2242
M ccttacaaatgagaaatagaagagatgaagcagcaggcaggcagactaggccactagcacttgatgagttcagagcgatgaaagagcat -2152
M tttcgggagaaggctcacaagggcaattgagaccaggaaagaaggcattttactgcatttgaaggtcactagttacattgacatgatagg -2062
M gcttaagctggacttgttgctgaaaagaaagggtcgttgagacgatcaacctatggagatatagcttcttactcaacctaaagggatatt -1972
M tccatcgagacacaatgctccatttttatgggctagaccaggcgggctaggaagtagagctcacctaagcaggaaaagatctttacagagg -1882
M ggaccggtgagggagtgggaagaaaagaggaatggtatagcactatagggaatatggaaagatatagaaattaggaagcagatgggaata -1792
M ccataagagagatcaactagagcttatgccaaggatgactttcccaacactagaagtcctactcatgcttgaaagggatatcgctaaaag -1702
M ggggccaactagcaatcctgttacaggcggtggtgtatgtatttccacttattcgagggaaaaactcttggtcttcaattttgtatttaa -1612
M gatttgtatattatgtatctattaaaaatctatacaatagtaggcggtaaaaagttgtatattttacttctattctgcttagtcaattac -1522
M actttaacgtcgtaaaatcccgggtttatgcaaaaacagagtgttattaaaacatagaatcaatatcggcaacaatgaaaggacgccttа -1432
M gactttgactcgggttgagtgctgtcagatacttgggataacaacgctgactttgtccgattcttctcacttctctagtgccaccacca -1342
M tgttagttgatgcattcccagagcttctgacttttctcctcatcagagagatttttcgtcgaggaagtcgcgatttcgttagtgtaactg -1252
M caactagagagtggataaggagctaatgaaagcagatatagaaaagttgcgctttccgtgagtgagaggtcaatcactctccagttcttc -1162
M tttgaacatgatgaagagctcttatcggcttgaggtttcttttcaagctgaatagaataatagaaatctcgtaagagaagaaaggttcac -1072
M agaaggttgagaagtagtacgcccggCCCTCGTACAGTTTTTCGAAACAAATAGATATATATATATAACTATAAAAGAGATAAGAGGGGA -982
M AGAATCGACAAAGAATGACGAACGACGGTCTTTCTGAGATTATAGATATATATATAGGGCGATGGGCAATCTGAGTAAATACCTTACTCA -892
M ACTCAGAGAAAGTCCCGGTCAATGTATGTAAAGCGAAGAGAGGCGCTCATAGATGTACCGATAAAGATTCAAACCTATCTAGAATTATAG -802
M GAGGTTAAAACGTGATAAATTCGTTTTTGCGGTTATAAGTTTCCTGTTCTTAGCAGTATGGTTTTATGGTCTTTGGACAAACAAGTGCCT -712
M AGCCGAGACCATATTCTCAACCAATGCAACAAATGGCAAAAACATAGTCTTTTTGTATAGTATAGAGTATGGTTGGATGGGCACTAGGAT -622
M TACCCCTATCCTAGTGTCTTGATTGAAATTGATTTTATCAGTTACACAAGAATGCGTCTCTTATGGAAGAAAGCAATGCCCAAAACTCCC -532
[pYC700]
M ATTTCTTTCTTGGTTGGATCAACCCAACCGGTGATTTCGGACAAGCCTTTCTTTTTTGAGCGGACAGCAGAAAGAAAAAGAAAAATTAT -442
M TATGATGACTAAGTCTAAGAAGGAGATGATTGCTTATGCAAAGCAGATAATTACAAATGTCCCTCGACCTACATAGGCTTATTATTCGGA -352
M CTCACTGCTCTTCACCCTCTTTTGGTCCGGGAGAAGCGGAATGGTTGAAACGAAATTAATTCTCCAAGCACTCTTTGGCATAGCAACAGC -262
M GGCTGCTTTTTTGTTTTTTTGTTATTCCCATTCCCGGACGCGGGAGCACTTCCTCTATACAATTCTAATAACGCTTTCTTTGGCTTTACT -172
M CTTCAAAAAGGGGGCTTTCCCTCTTTTTGAATTCGCCCCTTTGGCTATATTGATCTTTTCGGTTTGCTTCTTTCCCAAAAATGGACGATT -82
M ACGGTATTCGCTAGAGAACTGTTCTTTTTTTCTCCTATTACTTTGGGGGGATTACCGGCTTTTGGATACTCCGTGGTGTGAATGGGTACT 9
atp6 M G T
M CTCATGCTTTGTGGCCTCTTAACTTTATTCACGCTTCTTCAAGCGATGGTCCATCGAAAGGAGGAAGGGGTCGGCAAGTTTCTTCTACCG 99
L M L C G L L T L F T L L Q A M V H R K E E G V G K F L L P
M GCTGCCTTCCTAATTAGTGACCTCGTGCTCCTTTGGCTTCACTCGGGGGCGATTCAGAATCCGCTGGGAAAGGCGGCCCTAACGGGCGCC 189
A A F L I S D L V L L W L H S G A I Q N P L G K A A L T G A
M GCAGTCCTAATACTCTTTAGCCAAGCCTCCGCCAGTCCACAAGGCAGCGGGAGCTTGCAAATGGGGCGTTATCTCATGTGTATCCCTCCT 279
A V L I L F S Q A S A E S T R Q R E L A N G A Y L M C I P P
M ATACTATATATCATAGGAGGGGCCACGGATTGGCCGATTCGTAAACGAACTCACCTTTTTTATTCTTATCCTCTGGCTAGTCTTAGTTTT 369
I L Y I I G G A T D W P I R K R T H L F Y S Y P L A S L S F
M TTGGGTTGTTACTTTGGAGGACGAAAAAGAGCAGAAGGATCTGCTCCCGCTCCTATCTGCGATCTTCTCCGCTTTGGCGTCGCGAGTTAC 459
L G C Y F G G R K R A E G S A P A P I C D L L R F G V A S Y
M GGCCCCGACGTCTGGGTAAAAGAGTGGCTCCGCCGCTTTTCCTTAACCATTGGGGCGGTTTTTTTGTTTGAGAAGGTAGGCTACCGCAAA 549
G P D V W V K E W L A A F S L T I G A V F L F E K V G Y R K
M GATACCCTTCCAGAAACGTCCCGTACTTGGGAACCTTGGTTTTGCTTTTTTGCTGTTCTTTCTTTTGGTGTTGCACTTAGTTTTTTCAT 639
D T L P E T S R T W E P W F C F F A V L S F G V A L S F F H
F ATCCAACTAGAAATTGAATAAGAGAAGAAAGCAATGCCCAAAGACTCCCATGCCTTTCTTGGTCGGACCAACCCAACCGGTGATTTCCGA -118
M GCATCACCCGCCCTAGGGGTGGGAACTGAAAGTGAAGCTTTTCACGATGCGATTAATCCAGCAGCGGGACGGATCCCCAGCCCCCTAGAG 729
A S P A L G V G T E S E A F H D A I N P A A G R I P S P L E
F CAAGTCTTTCTTCATTTTTGAGCAAGAAGCGGAACTACAGGGATGAAATGAAAAGTGTTTATTACGATTACGCCCAACAGCCCACTTGAG -28
M CAATTTTCCATTCTCCCATTGATTCCTATGAAAATAGGAAACTTGTATTTCTCATTCACAAATCCATCTTTGTTTATGCTGCTAACTCTC 819
Q F S I L P L I P M K I G N L Y F S F T N P S L F M L L T L
M K I G N L Y F S F T N P S L F M L L T L
F CAATTTTCCATTCTCCCATTGATTCCTATGAAAATAGGAAACTTGTATTTCTCATTCACAAATCCATCTTTGTTTATGCTGCTAACTCTC 63
M/F AGTTTGGTCCTACTTCTGCTTCATTTTGTTACTAAAAAGGGAGGAGGAAACTCAGTACCAAATGTTTGGCAATCCTTGGTAGAGCTTATT 909/153
S L V L L L L H F V T K K G G G N S V P N V W Q S L V E L I
M/F TATGATTTCGTGCTGAACCTGGTAAACGAACAAATAGGTGGTCTTTCCGGAAATGTTAAACAAAAGTTTTTCCCTTGCATCTTGGTCACT 999/243
Y D F V L N L V N E Q I G G L S G N V K Q K F F P C I L V T
M/F TTTACTTTTTTGTTATTTCGTAATCTCCAGGGTATGATACCCTATAGCTTTACAGTTACAAGTCATTTTCTCATTACTTTGGGTCTTTCA 1089/333
F T F L L F R N L Q G M I P Y S F T V T S H F L I T L G L S
M/F TTTTCCATTTTTATTGGCATTACTATAGTGGGATTTCAAAGAAATGGGCTTCATTTTTTAAGCTTCTCATTACCTGCAGGAGTCCCGCTG 1179/423
F S I F I G I T I V G F Q R N G L H F L S F S L P A G V P L
M/F CCGTTAGCACCCTTTTTAGTACTCCTTGAGCTAATCCCTCATTGTTTTCGCGCATTAAGCTCAGGAATACGTTTATTTGCTAATATGATG 1269/513
P L A P F L V L L E L I P H C F R A L S S G I R L F A N M M
M/F GCCGGTCATAGTTCAGTAAAGATTTTAAGTGGGTTCGCTTGGACTATGCTATGTATGAATGATCTTTTATATTTCATAGGAGATCTTGGT 1359/603
A G H S S V K I L S G F A W T M L C F N D L L Y F I G D L G
M/F CCTTTATTTATAGTTCTTGCATTAACCGGTCTTGAATTAGGTGTAGCTATATTACAAGCTCATGTTTTTACGATCTTAATCTGTATTTAC 1449/693
P L F I V L A L T G L E L G V A I L Q A H V F T I L I C I Y
M TTGAATGATGCTACAAATCTCCATCAAAATTCTTTTTTCTTTTTATTAGAATTTTTATAATTGAACAAAAGCGAGGAAAGTGTCTATTTT 1539
L N D A T N L H Q N S F F F L L E F L *
F TTGAATGATGCTACAAATCTCCATCAAAATTCTTTTTTCTTTTTATTAGAATTTTTATAATTGAACAAAAGCGAGGTATTTATGCTTAAG 783
M CCAGCGATACTATAAATTATAGAGGACTTAACTTAAATTAAATGTCCTCTATGTTCCACGTACGTAGCGCAAATGTTATATAATGGCCTT 1629
F TCATCGCTTCAAATCCAATAAGGGATCAGACCCCTCCTCGTGTTCAAACTAGTCATTAATGGTCGGCTTAATTGGTATCC---------- 863
M TCTGTTCTTTTTTTTCGGTATGCCGCTCCGCCTGCAAGGAGCGAGAAAACAAATTGGTCTGTGGTGATGTCAGAATTTTTCCTTTTGAAAT 1719
F ----------TTTCGGTATGCCGCTCCGCCTGCAAGGAGCGAGAAAACAAATTGGTCTGTGGTGATGTCAGAATTTTTCCTTTTGAAAT 942

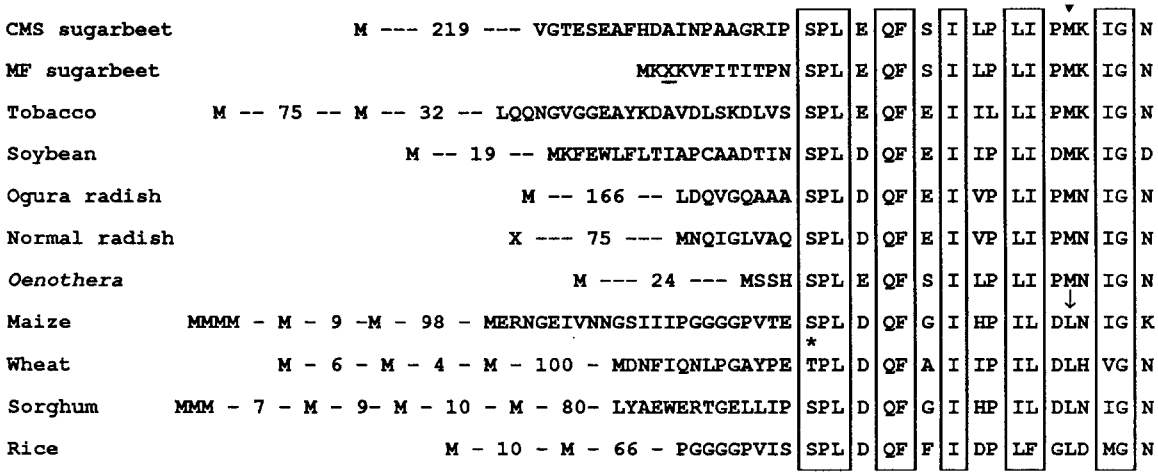


Fig. 6. Alignment of the predicted N-terminal amino acid sequence of ATP6 in higher plants. In-frame stop codons in sugar-beet MF ATP6 is represented by *X*. The position of the N-terminal serine residue of the mature wheat protein is indicated with an asterisk (*). A discrepancy between the deduced amino acid sequence of wheat ATP6 [3] and the N-terminal protein sequence data [1] is also denoted by the same asterisk. The methionine present in the ATP6 polypeptides of all dicotyledonous species is indicted by a triangle (▼) and the leucine at an equivalent position in that of monocotyledonous species by an arrow (↓). The identical amino acids between deduced ATP6 polypeptides are shown boxed.

sequence, translated according to the universal genetic code, appears to be the shortest reported to date if translation were initiated at the first in-frame AUG. The predicted ATP6 polypeptide from CMS sugar-beet (502 amino acids) appears to contain a large N-terminal extension compared to its MF counterpart. No homology has been detected between the nucleic acid or predicted polypeptide sequence of the N-terminal extension and database entries. The CMS and MF gene sequences diverge extensively 41 bp upstream of the first in-frame ATG codon of the MF sequence (Fig. 5), however, within this 41 bp region the predicted translation product shows extensive homology with the translation products of all higher-plant ATP6 polypeptides (Fig. 6) suggesting translation of the MF gene may be initiated further upstream.

Comparison of the restriction maps for the CMS and MF *atp6* locus suggests that sequences downstream of the coding sequence are very similar (Fig. 4). Sequence analysis revealed that a 21 bp sequence (positions 1620 to 1640 on the CMS sequence) 3′ to the coding sequence is absent in the corresponding MF sequence; sequences immediately downstream of the proposed *atp6* stop codon (nucleotides 1529 to 1619 on the CMS sequence) differ between the two genotypes. Downstream of this region both genes show extensive homology for a further 1.3 kb to the *Eco* RI site (data not shown).

Computer analysis revealed a region identical

Fig. 5. Comparison of the DNA sequences of the *atp6* loci from CMS (M) and MF (F) sugar-beet mitochondria. The amino acid sequence of *atp6* genes were derived by translation according to the universal genetic code. Identical nucleotides between CMS repeat II and its MF homologue are shown in lower case. The DNA sequence corresponding to the cDNA clone pYC700 is underlined. *atp6*, the gene for subunit 6 of the F_O-ATPase complex. The points of sequence divergence between CMS and MF *atp6* genes are marked by triangles (▲). The amino acid sequences are shown between the nucleotide sequences. Mapped transcript termini are indicated by vertical arrows (↑ and ↓ for MF and CMS transcripts respectively). Nucleotides deleted in one sequence relative to the other are indicated by (–). The DNA sequences are numbered from the first in-frame ATG codon of each *atp6* gene (position 1).

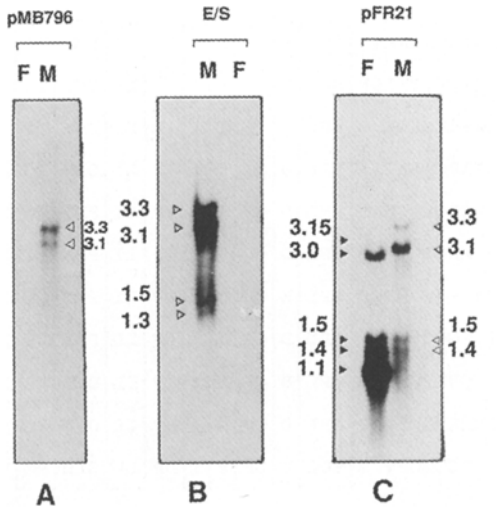


Fig. 7. Northern hybridization analysis of the *atp6* locus in CMS (M) and MF (F) sugar-beet mitochondria. Four μg of mtRNA was loaded in each lane and electrophoresed on a 1.4% agarose gel containing 6.6% formaldehyde. The blots were hybridized with pMB796 (A); an *Eco* RI/*Sty* I fragment of 724 bp containing the 5′ end of CMS *atp6* (B); and the probe pFR21 encompassing the MF *atp6* gene (C). Transcript sizes are shown in kb. The locations of all relevant probes are illustrated in Fig. 4. RNAs from two different preparations were used in panels B and C which accounts for the different proportions of 3.3 kb and 3.1 kb transcripts present in CMS mtRNA samples.

in sequence to pYC700, 462 bp upstream of the CMS *atp6* gene (Fig. 5). The DNA sequence immediately downstream of the 3′ terminal nucleotide in the cDNA clone contains an A-rich sequence which probably facilitated oligo-dT priming for cDNA synthesis (Fig. 5).

Transcription of the atp6 *locus in CMS and MF mitochondria*

Northern hybridization analysis using pMB796 as a probe detected transcripts exclusively in CMS mtRNA (Fig. 7), similar in size to those detected by pYC700 (Fig. 2A). When a probe containing the 5′ end of the CMS *atp6* was used as a probe (nucleotides −142 to 584, Fig. 5), it hybridized to the 3.3 kb and 3.1 kb transcripts, to a shorter transcript of 1.5 kb and to lesser extent a transcript of 1.3 kb (Fig. 7). No homologous transcripts were detected in MF mtRNA. A probe containing the MF *atp6* gene hybridized to similar sized transcripts in CMS mtRNA preparations as the probe derived above and to a major transcript of 1.1 kb in MF mitochondria, in addition to several transcripts of higher molecular weight (Fig. 7). Northern analysis with strand

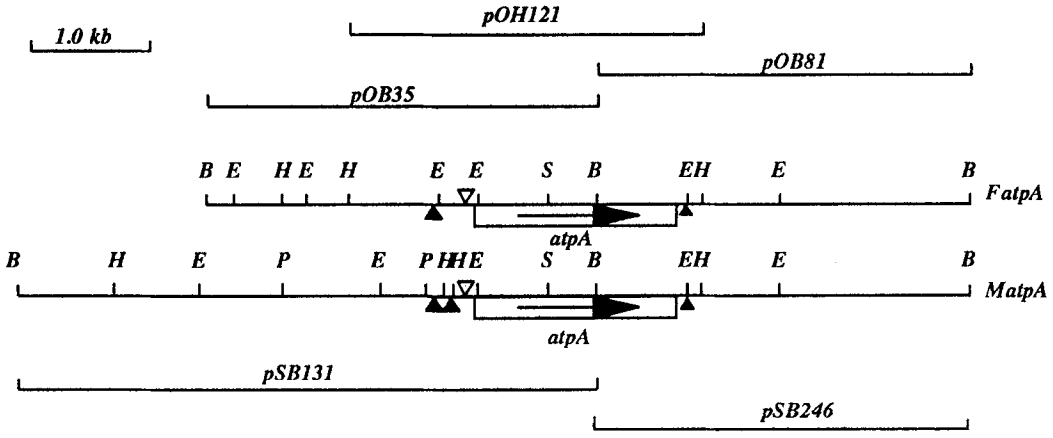


Fig. 8. Genomic organization of the *atpA* locus from CMS and MF mitochondria. Clones isolated from CMS and MF mtDNA used to construct maps are shown below or above the appropriate locus respectively. The points of sequence divergence between CMS and MF *atpA* genes are indicated by open triangles and mapped transcript termini by filled ones. The location of *atpA* is shown and the direction of transcription is indicated by arrows. Sites for the following restriction enzymes are also shown: *B*, *Bam* HI; *E*, *Eco* RI; *H*, *Hin*d III; *P*, *Pst* I; *S*, *Sal* GI.

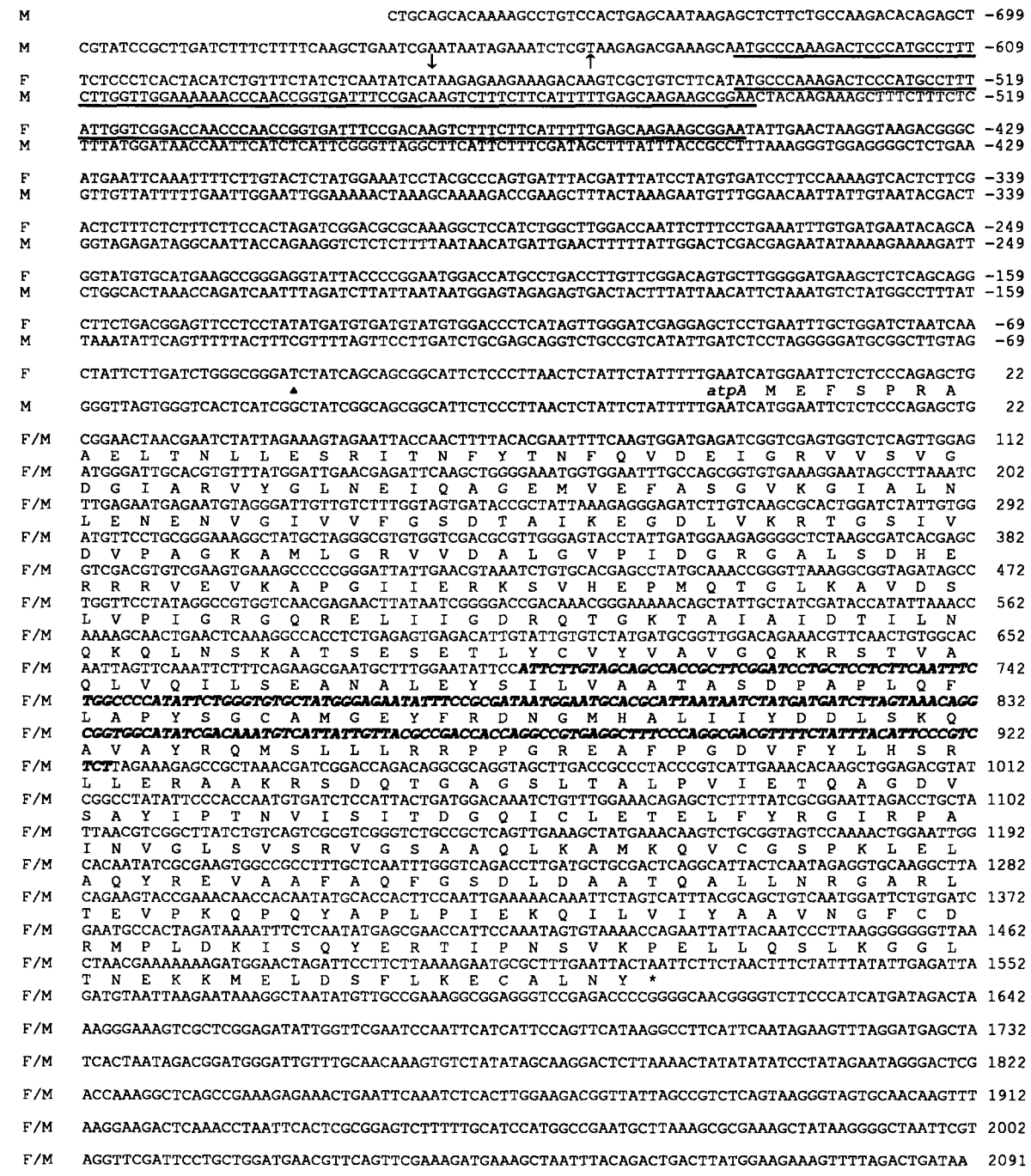


Fig. 9. DNA sequence comparison of the CMS (M) and MF (F) *atpA* gene. The points of sequence divergence between CMS and MF *atpA* genes are marked by a dark triangle. The amino acid sequences are shown between the nucleotide sequences which are numbered from the presumed translation initiation codon at position 1. The extent of DNA sequence corresponding to the cDNA clone pYC130 is shown in bold and italic. A homologous sequence of 92 bp upstream of the putative coding sequence in both *atpA* genes is underlined. Transcript termini determined by S1 nuclease protection analysis [33] are indicated by arrows.

specific probes confirmed that all transcripts from each genotype are of the same polarity (results not shown).

Molecular cloning and sequence analysis of the atpA *locus from CMS and MF mitochondria*

The cDNA clone pYC130 was used to screen a *Hind* III library of MF mtDNA. Restriction enzyme mapping of one positive clone (pOH121) identified *Eco* RI, *Sal* GI and *Bam* HI sites conserved in several higher plant *atpA* genes (Fig. 8). Subsequently, a 2.0 kb *Eco* RI fragment isolated from pOH121 was used to probe *Bam* HI libraries of CMS and MF mtDNA in order to obtain overlapping clones encompassing the *atpA* gene from each mitochondrial genotype (Fig. 8). The DNA sequence of c.2.9 kbp of CMS mtDNA and 2.7 kb of MF mtDNA at the *atpA* locus was determined as described in Materials and methods and found to be identical to sequences reported recently by Senda *et al.* [31]. The relevant sequence is summarized in Fig. 9.

Analysis of the sequence revealed a common large open reading frame of 506 amino acids (Fig. 9) potentially encoding a protein of 55 kDa, showing greater than 95% amino acid sequence homology with the corresponding sequences from pea [20], maize [14], tobacco [6] and *Oenothera* [28]. The coding and 3′-flanking regions of both genes appear identical. MF and CMS sequences diverge extensively 47 bp upstream of the proposed translation initiation site (Fig. 9). Comparison of the pYC130 cDNA sequence with the corresponding genomic sequence revealed one difference; in the genomic sequence a T was observed at position 863 and C in the cDNA sequence. The cDNA sequence would encode a different amino acid, alanine instead of valine. However, a conserved valine is found in the ATPA of other plant species. If this change is as a result of RNA editing [35] it appears to result in a conservative amino acid change. As observed in the sequence analysis of pYC700, the genomic sequence adjacent to the 3′-terminal nucleotide in pYC130 is composed of an oligopurine sequence which facilitated oligo-dT priming (Fig. 9).

Transcription of the atpA *locus from MF and CMS mitochondria*

Northern hybridization analysis of mtRNAs with a mtDNA fragment containing sugar-beet *atpA* showed that similar sized transcripts were detected as those by pYC130 (see Fig. 1). Two transcripts (2.3 kb and 2.0 kb) were found in CMS mitochondria and only one (2.2 kb) in MF mitochondria. S1 nuclease protection experiments were performed to determine the basis for differences in the *atpA* transcript profiles in CMS and MF mitochondria. When such experiments were

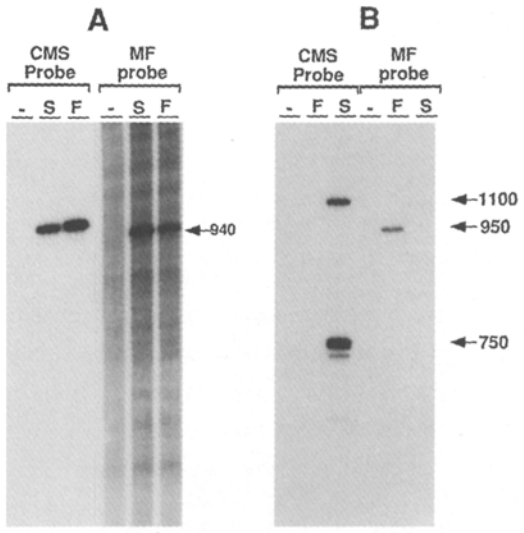


Fig. 10. S1 nuclease protection analysis of *atpA* transcripts from CMS and MF mitochondria. Probes for S1 analysis, asymmetrically labelled at the 5′ or 3′ terminus were prepared with polynucleotide kinase or DNA polymerase I (klenow fragment), followed by secondary restriction enzyme digestion as described by Thomas [33]. A. Mapping of 3′ termini; hybridization probes 3′ end-labelled at the internal *Bam* HI site (nucleotide 720, see Fig. 9) derived from CMS mtDNA (pSB246) or MF mtDNA (pOB81) were used in protection assays in the absence of RNA (–), in the presence of CMS mtRNA (S) or MF mtRNA (F) as indicated in the figure. B. Mapping of 5′ termini; probes 5′ end-labelled at the internal *Sal* GI site (nucleotide 383, see Fig. 9) derived from CMS mtDNA (pSB131) or MF mtDNA (pOB35) were hybridized in the absence of RNA, or mtRNA from each genotype as indicated above. S1 protection fragments were resolved on 4% polyacrylamide/8 M urea sequencing gels alongside labelled size markers (not shown). The sizes of S1 protection fragments are indicated in bases.

performed using 3′ end-labelled probes (Fig. 10) the results demonstrated that *atpA* transcripts from each genotype were 3′ co-terminal. Transcripts are terminated ca. 200 bp downstream of the proposed termination codon (Fig. 9). In nuclease protection experiments using 5′ end-labelled probes, two 5′ termini were observed in CMS mitochondria, but only one in MF mitochondria (Fig. 10), consistent with the data from northern hybridization analysis (Fig. 1). The precise location of 5′ termini have been determined in high resolution mapping experiments [31, 33] and are located upstream of a conserved 92 bp sequence in CMS and MF genotypes (Fig. 9); this sequence may play a role in the regulation of transcription and/or for translation of the *atpA* gene.

Discussion

We have described experiments to identity novel transcript patterns or rearranged loci between two different mitochondrial genotypes. A uniquely expressed cDNA from CMS mitochondria and a cDNA detecting different transcript patterns in CMS and MF mitochondria (Figs. 1 and 2) were used as molecular probes to isolate the corresponding genomic loci. One cDNA clone (pYC130) originated from the CMS *atpA* locus (Figs. 8 and 9). Another cDNA clone (pYC700) hybridizes exclusively to RNA from CMS mitochondria (Fig. 2); characterization of the corresponding genomic DNA revealed that pYC700 is part of the CMS *atp6* locus. The sugar-beet CMS and MF *atpA* genes appear to have identical coding sequences (Fig. 9, [31]), ruling out an obvious role in expression of the CMS phenotype. Genomic rearrangements were observed 5′ to the proposed coding sequences (Fig. 9), resulting in different transcript profiles in each genotype (Figs. 1 and 10).

Restriction enzyme mapping and Southern hybridization analysis have shown that a sequence of at least 9 kb is repeated upstream of *atp6* and *atp9* loci in CMS mitochondria (repeat II [4], Figs. 3 and 4). Only 1665 bp of this sequence is present as a single copy in MF mitochondria upstream of the *atp9* locus (Fig. 4). High resolution transcript mapping suggests that the 5′ termini of the largest transcripts identified with *atp6* probes (Fig. 7) are located within repeat II ([33], results not shown). Sequences around the termini of *atp6* and *atpA* transcripts from both CMS and MF genotypes are highly homologous to sequences at the 5′ termini of transcripts from sugar-beet minicircular DNAs which may act as promoter sequences [33]. Northern hybridization analysis also identified smaller transcripts which may be generated as a consequence of independent transcription initiation events, as described in maize [21, 22] or processing of larger primary transcripts (e.g. the 3.3 kb and 3.1 kb RNAs).

The predicted amino acid sequence of sugar-beet MF ATP6 is the shortest described so far when translated according to the universal genetic code (250 amino acids), whereas CMS ATP6 appears to be expressed as a larger protein of 502 amino acids due to the presence of a large N-terminal extension. It has been suggested that wheat ATP6 [1] is synthesized with presequences as in yeast [19] where post-translational cleavage of the amino terminal sequence is involved in its maturation. N-terminal sequencing of these proteins has revealed the sequence SPL at the mature N-terminus. The mature ATP6 polypeptide in CMS sugar-beet and other dicotyledonous plants may be generated by a similar mechanism since this sequence appears to be highly conserved (see Fig. 6). This latter sequence, however, appears to be upstream of the first in-frame initiation codon in the MF gene. A potential initiation codon is located 36 bp upstream of the SPL sequence but is closely followed by a termination codon. Therefore, it is possible that translation of MF ATP6 mRNA is initiated either from an alternative initiation codon, splicing of a RNA leader sequence or creation of an AUG initiation codon by RNA editing as recently demonstrated for the wheat *nad1* gene [5].

Chimeric genes are a common feature of plant mitochondria [12], and their relationship to the CMS phenotype is largely circumstantial, with the exception of S-*pcf* in *Petunia* [38], *T-urf13* in

maize [9] and *atp6* in rice [15]. In these latter examples, their role in expression of the CMS phenotype has been deduced from the analysis of somatic hybrid plants [7], analysis of cytoplasmic revertants to MF [26] and by determining the effect of dominant nuclear genes which restore fertility [15, 17, 25]. In sugar-beet, no well characterized lines containing restorer genes, or spontaneous revertants to MF are known. Therefore, the potential role of any genome rearrangements in expression of the CMS phenotype will remain unclear.

Acknowledgements

This work was funded by a grant from the European Economic Community (Contract No. GBI.4.116) and by the Agriculture and Food Research Council via a grant-in-aid to the John Innes Institute. Y. Xue is grateful for a postgraduate studentship from the University of East Anglia and an ORS award from the Committee of Vice-Chancellors and Principals of Universities of the United Kingdom. We would like to thank Dr T. Brears and Dr D.M. Lonsdale for screening the cosmid library of CMS sugar-beet mtDNAs and Dr D. Falconet and Dr T.H.N. Ellis for providing rRNA gene clones.

References

1. Begu D, Graves PV, Araya A, Litvak S: In: Quétier F, Lejeune B (eds) Proceedings of the 3rd International Workshop on the Mitochondrial Genome of Higher Plants, (28 August to 2 September 1988, Roscoff, France), p. 2, Centre National de la Recherche Scientifique, France (1988).
2. Bland MM, Levings CS III, Matzinger DF: The ATPase subunit 6 gene of tobacco mitochondria contains an unusual sequence. Curr Genet 12: 475–481 (1987).
3. Bonen L, Bird S: Sequence analysis of the wheat mitochondrial atp6 gene reveals a fused upstream reading frame and markedly divergent N termini among plant ATP6 proteins. Gene 73: 47–56 (1988).
4. Brears T, Lonsdale DM: The sugarbeet mitochondrial genome: a complex organization generated by homologous recombination. Mol Gen Genet 213: 514–522 (1988).
5. Chapedelaine Y, Bonen L: The mitochondrial gene for subunit I of the NADH dehydrogenase complex: A *trans*-splicing model for this gene-in-pieces. Cell 65: 465–472 (1991).
6. Chaumont F, Boutry M, Briquet M, Vassarotti A: Sequence of the gene encoding the mitochondrial F1-ATPase α subunit from *Nicotiana plumbaginifolia*. Nucl Acids Res 17: 2613–2631 (1988).
7. Clark E, Gafni Y, Izhar S: Loss of CMS-specific mitochondrial DNA arrangement in fertile segregants of *Petunia* hybrids. Plant Mol Biol 11: 249–253 (1988).
8. Dewey RE, Levings CS III,Timothy DH: Nucleotide sequence of the ATPase subunit 6 gene of maize mitochondria. Plant Physiol 79: 914–919 (1985).
9. Dewey RE, Levings CS III, Timothy DH: Novel recombination in the maize mitochondrial genome produces a unique transcriptional unit in the Texas male-sterile cytoplasm. Cell 44: 439–49 (1986).
10. Duchenne M, Lejeune B, Fouilard P, Quétier F: Comparison of the organization and expression of mtDNA of fertile and male-sterile sugar beet varieties (*Beta vulgaris* L.). Theor Appl Genet 78: 633–640 (1989).
11. Grabau E, Havlik M, Gesteland R: Chimeric organization of two genes for the soybean mitochondrial ATPase subunit 6. Curr Genet 13: 83–89 (1988).
12. Hanson MR: Plant mitochondrial mutations and male sterility. Annu Rev Genet 25: 461–486 (1991).
13. Henikoff S: Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359 (1984).
14. Isaac PG, Brennicke A, Dunbar SM, Leaver CJ: The mitochondrial genome of fertile maize (*Zea mays* L.) contains two copies of the gene encoding the α-subunit of the F_1-ATPase. Curr Genet 10: 321–328 (1985).
15. Iwabuchi M, Kyozuka J, Shimamoto K: Processing followed by complete editing of altered mitochondrial *atp6* RNA restores fertility of cytoplasmic male sterile rice. EMBO J 12: 1437–1446 (1993).
16. Kempken F, Mullen JA, Pring DR, Tang HV: RNA editing of sorghum mitochondrial *atp6* transcripts changes 15 amino acids and generates a carboxyl-terminus identical to yeast. Curr Genet 20: 417–422 (1991).
17. Kennell JC, Pring DR: Initiation and processing of *atp6*, T-*urf13* and *ORF221* transcripts from mitochondria of T cytoplasm of maize. Mol Gen Genet 216: 16–24 (1991).
18. Makaroff CA, Apel IJ, Palmer JD: The *atp6* coding region has been disrupted and a novel reading frame generated in the mitochondrial genome of cytoplasmic male-sterile radish. J Biol Chem 20: 11706–11713 (1989).
19. Michon T, Galante M, Velours J: NH_2-terminal sequence of the isolated yeast ATP synthase subunit 6 reveals post-translational cleavage. Eur J Biochem 172: 621–625 (1988).
20. Morikami A, Nakamura K: Structure and expression of pea mitochondrial F1 ATPase α-subunit gene and its

pseudogene involved in homologous recombination. J Biochem 101: 967–976 (1987).
21. Mulligan RM, Maloney AP, Walbot V: RNA processing and multiple transcription initiation sites result in transcript size heterogeneity in maize mitochondria. Mol Gen Genet 211: 373–380 (1988).
22. Mulligan RM, Lau GT, Walbot V: Numerous transcription initiation sites exist for the maize mitochondrial genes for subunit 9 of the ATP synthase and subunit 3 of cytochrome oxidase. Proc Natl Acad Sci USA 85: 7998–8002 (1988).
23. Owen FV: Cytoplasmically inherited male sterility in sugarbeets. J Agric Res 71: 423–440 (1945).
24. Powling A: Restriction endonuclease analysis of mitochondrial DNA from sugar beet with normal and male sterile cytoplasms. Heredity 49: 117–120 (1982).
25. Pruitt KD, Hanson MR: Transcription of the *Petunia* mitochondrial CMS-associated Pcf locus in male sterile and fertility restored lines. Mol Gen Genet 227: 348–355 (1991).
26. Rottman WH, Brears T, Hodge TP, Lonsdale DM: A mitochondrial gene is lost via homologous recombination during reversion of CMS-T to fertility. EMBO J 6: 1541–1546 (1987).
27. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
28. Schuster W, Brennicke A: Pseudocopies of the ATPase α-subunit gene in *Oenothera* mitochondria are present on different circular molecules. Mol Gen Genet 204: 29–35 (1986).
29. Schuster W, Brennicke A: Nucleotide sequence of the *Oenothera* ATPase subunit 6 gene. Nucl Acids Res 15: 9092 (1987).
30. Senda M, Harada T, Mikami T, Sugiura M, Kinoshita T: Genomic organization and sequence analysis of the cytochrome oxidase subunit II gene from normal and male-sterile mitochondria in sugarbeet. Curr Genet 19: 175–181 (1991).
31. Senda, M, Mikami, T, Kinoshita, T: The sugar beet mitochondrial gene for the ATPase alpha-subunit: sequence, transcription and rearrangements in cytoplasmic male-sterile plants. Curr Genet 24: 164–170 (1993).
32. Singh M, Brown GG: Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. Plant Cell 3: 1349–1362 (1991).
33. Thomas CM: Sugarbeet minicircular mitochondrial DNAs: high-resolution transcript mapping, transcript abundance and copy number determination. Mol Gen Genet 234: 457–465 (1992).
34. Wise RP, Pring DR, Gengenbach BG: Mutation to male fertility and toxin insensitivity in T-cytoplasm maize is associated with a frameshift in a mitochondrial open reading frame. Proc Natl Acad Sci USA 84: 2858–62 (1987).
35. Wissinger B, Brennicke A, Schuster W: Regenerating good sense. Trends Genet 8: 322–328 (1992).
36. Xue Y, Davies DR, Thomas CM: Sugarbeet mitochondria contain an open reading frame showing extensive sequence homology to the subunit 2 gene of the NADH:ubiquinone reductase complex. Mol Gen Genet 221: 195–198 (1990).
37. Xue Y, Thomas CM, Davies DR: Nucleotide sequence and transcription of the sugarbeet mitochondrial F_oF_1-ATPase subunit 9 gene. Nucl Acids Res 21: 8857 (1989).
38. Young EG, Hanson MR: A fused mitochondrial gene associated with cytoplasmic male sterility is developmentally regulated. Cell 50: 41–49 (1987).