Mol Gen Genet (1990) 221:195–198

MGG

Sugarbeet mitochondria contain an open reading frame showing extensive sequence homology to the subunit 2 gene of the NADH: Ubiquinone reductase complex

Yongbiao Xue, D. Roy Davies, and Colwyn M. Thomas*

Department of Applied Genetics, John Innes Institute, AFRC Institute of Plant Science Research, Colney Lane, Norwich, NR4 7UH, UK

Summary. Sequence analysis of a transcribed region of mitochondrial DNA (mtDNA) from male fertile sugarbeet (*Beta vulgaris* L.) revealed an open reading frame showing extensive sequence homology to the subunit 2 gene of the NADH:ubiquinone reductase complex (*nad2*). Sugarbeet *nad2* in common with its proposed counterpart in animal mitochondria has no intron and thereby differs from the corresponding chloroplast gene. Northern RNA analysis of sugarbeet *nad2* suggested that transcription of this locus gives rise to at least three transcripts. No differences in transcript profile were detected between male fertile and cytoplasmic male sterile sugarbeet. This constitutes the first report of a mitochondrial *nad2* gene in higher plants.

Key words: *nad2* – NADH:ubiquinone reductase – Mitochondrial DNA – *Beta vulgaris* L.

Introduction

Despite large variations in the genome size of higher plant mitochondria, analysis of in vitro translation products in isolated mitochondria suggests they may encode 25–35 proteins (Leaver et al. 1983), consistent with the proposed number of transcriptional units (Makaroff and Palmer 1987). Several of the genes encoding these proteins have been isolated and characterized: these include *cox1*, *cox2* and *cox3* of the cytochrome *c* oxidase complex; apocytochrome *b* (*cyb*) of the *bc1* complex; *atp6*, *atp9* and *atpA* of the ATP synthase complex; and three genes encoding subunits of the NADH:ubiquinone reductase complex, *nad1* *nad3* and *nad5* (Lonsdale 1988).

Differential screening of cDNA libraries from cytoplasmic male sterile (CMS) and male fertile (MF) sugarbeet mitochondria has identified a uniquely transcribed open reading frame in CMS mitochondria co-transcribed with the gene encoding ATP6 (unpublished results). This transcriptional unit is flanked by a sequence repeated upstream of the CMS ATP9 gene (repeat II), consistent with results from cosmid mapping of the genome (Brears and Lonsdale 1988). Northern analysis demonstrated that mitochondrial DNA (mtDNA) upstream of a repeat II-homologous sequence in MF mtDNA is transcribed in both CMS and MF genotypes. Here we report on the DNA sequence analysis of this region and the identification of an open reading frame showing extensive homology to the subunit 2 gene of the NADH:ubiquinone reductase complex.

Materials and methods

Preparation of nucleic acids. Mitochondrial nucleic acids were isolated from lines of CMS and MF field grown sugarbeet as described previously (Thomas 1986).

Construction of mitochondrial genomic libraries. The detailed construction of mtDNA and cDNA libraries from MF and CMS sugarbeet is described elsewhere (Y. Xue et al. manuscript submitted).

Northern blot analysis. Mitochondrial RNAs were fractionated on 1.4% agarose gels containing formaldehyde/MOPS buffer, and blotted onto nylon membrane as recommended by the manufacturer (Amersham).

DNA sequence analysis. DNA fragments were subcloned and ordered deletions of the cloned inserts were created with exonuclease III. The sequence was determined by the dideoxy chain-terminator technique in both orientations (Sanger et al. 1977). The DNA sequence of the sugarbeet *nad2* gene can be accessed from the EMBL DNA database (Acc. No. X16828).

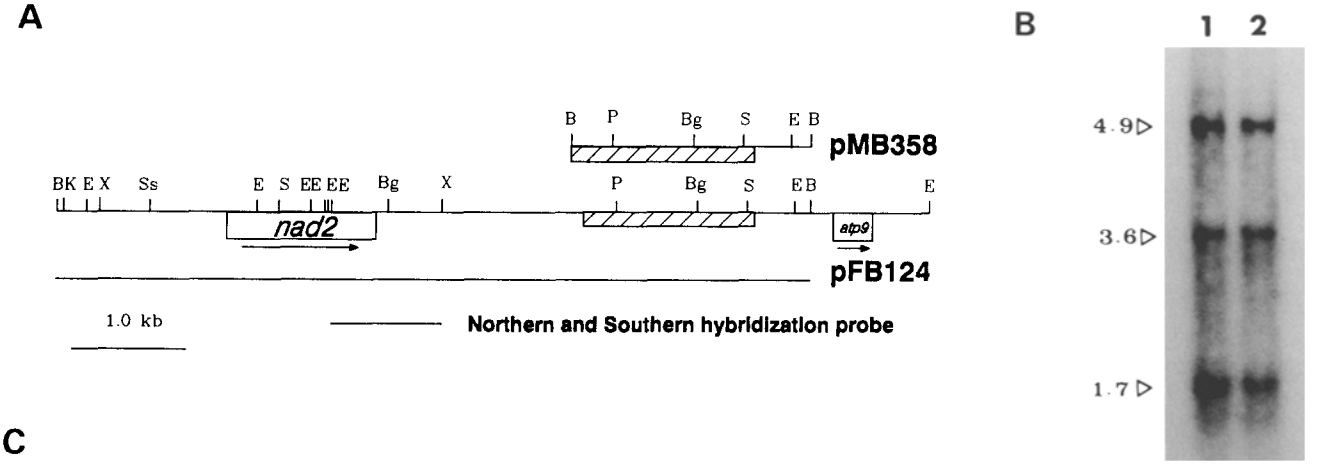
Results

Identification of mitochondrial nad2 *from sugarbeet by sequence analysis of a transcribed region*

We initially attempted to isolate the *atp6* locus from MF mitochondria using a clone isolated from CMS mtDNA containing part of repeat II (pMB358) and subsequently isolated a clone containing a 6.5 kb *Bam*HI fragment, pFB124 (Fig. 1A). Southern hybridization, restriction enzyme mapping and DNA sequence analysis revealed that this clone contained approximately 1.6 kb of repeat II-homologous sequence located upstream of MF *atp9* and not *atp6* (Fig. 1A). When used in Northern analysis pFB124 hybridized to three transcripts from MF and CMS mitochondria (Fig. 1B). These transcripts originate from se-

* *Present address:* The Sainsbury Laboratory, John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

Offprint requests to: C.M. Thomas



C

```
GGTCTGAAGAGCTCTGCCTCCTCTGTTATCAAGGCACGCCCCACTGCGTTAGGGTGGGTGCTTGTGAGCGCCTTGGTATGAGATCCTAAAAAAGCAAGTCCGGTATTGCGAAGATCTTCT
      10        20        30        40        50        60        70        80        90       100       110       120
TCATCAATAGATAGGAATGAGTGTTCGATATAGGGGATAGGATCTATCTGGTCTTTTTCTTTCTTTTTTTTGCATGGCATAGCTAGGACCCTTCAAATCATGTTTGAGCCTATGTTCAAA
     130       140       150       160       170       180       190       200       210       220       230       240
                                                [NAD2] M  L  S  L  T  I  K  G  K  A  R  R  R  K  E  R  A  F  G  D  R
CCAAACCCACTCCCTATTTGAGTAAGGCTGGAAGCCCGCCAAGTTCAGATAAGAGAATGCTTTCTCTAACCATTAAGGGAAAGGCTCGACGAAGGAAGGAAAGGGCTTTCGGAGATCGAG
     250       260       270       280       290       300       310       320       330       340       350       360
D  F  L  T  F  S  S  K  T  K  K  T  E  N  V  N  L  S  F  E  K  G  T  R  F  F  D  R  G  G  M  I  F  G  P  S  P  R  S  A
ATTTTCTTACTTTTTCATCGAAAACGAAGAAGACCGAGAATGTCAATCTTTCTTTCGAAAAGGGGAACACGCTTTTTCGACCGCGGTGGTATGATTTTCGGGCCTTCTCCTCGTTCCGCTC
     370       380       390       400       410       420       430       440       450       460       470       480
R  W  P  I  G  I  A  A  F  G  L  C  L  L  F  L  I  K  N  S  G  S  A  R  E  S  A  G  N  N  R  K  E  G  V  H  V  A  A  A
GCTGGCCTATTGGGATAGCAGCCTTTGGGCTTTGCCTGCTCTTTTTAATAAAGAATTCCGGCTCGGCCCGGGAAAGCGCTGGCAACAACAGAAAGGAAGGGGTCCATGTAGCTGCTGCGT
     490       500       510       520       530       540       550       560       570       580       590       600
S  A  P  F  L  V  N  R  A  A  G  S  A  T  T  T  K  E  R  I  H  F  K  I  T  N  A  S  A  M  A  A  C  G  M  A  G  S  D  L
CCGCCCCCTTCTTAGTCAATAGAGCAGCAGGTTCGGCTACTACTACAAAAGAGAGAATCCACTTCAAGATAACCAACGCCTCTGCAATGGCAGCGTGTGGAATGGCCGGGAGCGACCTTT
     610       620       630       640       650       660       670       680       690       700       710       720
F  G  Y  I  I  Q  V  E  S  G  V  T  G  T  A  G  L  M  E  N  N  F  H  G  S  V  Q  R  A  L  F  S  L  R  I  L  R  S  L  R
TTGGATATATAATCCAAGTCGAGAGTGGAGTTACGGGAACAGCCGGTCTGATGGAAAACAACTTTCACGGTTCGGTTCAGAGAGCACTTTTTTCGTTGAGAATTCTTCGTTCCCTTCGTG
     730       740       750       760       770       780       790       800       810       820       830       840
V  N  S  L  A  R  I  Q  N  F  W  G  P  S  I  P  S  S  S  P  A  K  T  P  L  P  F  G  L  N  I  F  F  D  S  Y  M  W  A  P
TGAATTCCCTAGCGCGAATTCAAAACTTTTGGGGCCCATCTATTCCATCATCGAGCCCCGCGAAAACCCCCCTCCCCTTCGGACTCAATATCTTTTTTGACTCATATATGTGGGCACCTG
     850       860       870       880       890       900       910       920       930       940       950       960
D  I  Y  E  G  S  P  T  P  V  T  A  F  F  S  I  A  P  E  R  S  I  S  A  N  I  L  R  V  F  I  Y  G  S  Y  G  A  T  L  Q
ACATCTATGAGGGTTCACCCACCCCGGTTACAGCATTCTTTTCTATTGCGCCTGAAAGATCTATTTCTGCTAATATTTTACGTGTTTTTATTTATGGTTCCTATGGAGCTACATTGCAAC
     970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
Q  I  F  F  F  C  S  I  A  L  R  L  R  S  T  G  A  M  A  N  E  G  K  A  S  S  S  I  G  Q  L  D  Y  G  G  L  Y  F  V  L
AAATCTTCTTTTTCTGCAGCATTGCTCTTAGACTTAGGAGCACTGGCGCCATGGCCAACGAAGGAAAAGCTTCTAGCTCCATAGGTCAATTGGACTATGGAGGGTTATATTTCGTACTGG
     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
V  L  M  W  N  R  E  G  I  Q  S  L  L  I  G  L  F  I  Y  A  S  M  D  D  R  C  F  A  I  V  S  A  L  R  Q  T  R  V  K  Y
TTCTCATGTGGAACCGAGAAGGAATTCAATCACTACTAATTGGTCTCTTTATTTATGCATCAATGGACGATAGATGCTTCGCTATAGTTTCAGCATTACGGCAAACACGTGTCAAATATA
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
I  A  D  L  G  A  L  A  K  T  N  P  I  S  A  I  T  F  S  I  T  M  F  S  Y  A  G  I  P  P  L  A  G  F  C  S  K  F  Y  L
TAGCGGATTTGGGCGCTCTAGCCAAAACGAATCCTATTTCGGCTATTACCTTCTCTATTACTATGTTCTCATACGCAGGAATACCCCCGTTAGCCGGCTTTTGTAGTAAATTCTATTTGT
     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
F  F  A  A  L  G  C  G  A  Y  F  L  A  P  V  G  V  V  T  S  V  I  G  C  W  A  A  G  R  L  P  R  V  S  Q  F  G  D  R  R
TCTTCGCCGCTTTGGGTTGTGGGGCTTACTTCCTAGCCCCAGTGGGAGTAGTGACTAGCGTTATAGGTTGTTGGGCGGCCGGAAGGTTGCCACGAGTAAGTCAGTTTGGGGACCGAAGGC
     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
Q  F  S  V  H  R  T  R  S  L  P  N  Q  L  R  H  G  W  E  C  M  L  R  K  I  G  S  S  L  I  H  Q  P  S  V  Y  S  I  S  L
AGTTCTCCGTGCACCGGACACGTAGCTTACCGAATCAGTTGCGACACGGATGGGAATGCATGCTACGAAAGATAGGGTCGAGTCTGATACATCAACCGTCTGTCTACTCAATATCCTTGT
     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
Y  E  S  T  I  T  T  R  D  E  P  W  F  G  E  F  E  L  A  L  G  V  I  G  L  P  V  T  A  H  D  R  I  L  R  C  S  P  P  V
ACGAGTCCACAATCACTACACGAGATGAACCCTGGTTTGGTGAATTTGAGTTGGCCTTAGGTGTAATAGGACTCCCAGTTACTGCGCACGATCGTATACTGAGGTGCTCCCCGCCGGTTG
     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
V  G  T  T  R  A  G  P  G  L  N  S  E  R  *
TTGGAACGACGCGAGCCGGGCCGGGCCTCAATTCAGAAAGATAAAGGGCCCAAAAGTCTTAATAAATAGGAGGTTCCAAATTCCCCATCTCATTGAGGGCGGAAAACGAATCGACATCTC
     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920
GATGTGATACAGCCTTTTCTATTTTAGTTGGGAAAGAACGGCGAAGTCCATCCAAACCGTCCAATGAAGAATAAGAAGAGAGCAAAGCGCATGCGGAACGGACACAGAAAAAAAGAAGTG
     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
TTGGAAGAGAAGCAGCCGAGCTCATTCCCTCGCTTCCTGGGCCCAAAGCAGTGCAGTCTTTCCTGGCCAAATAAAGGATTTGGGGCTTCTCTTGCTACGTAAAAAATTTCTTTTTTTTTT
     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
```

Fig. 1. A Organization of the *nad2* and *atp9* loci in male fertile sugarbeet mitochondria. The polarity of transcription is indicated by *arrows*. The location and extent of pFB124 is indicated as is the *Eco*RI-*Xho*I fragment used in Southern and Northern hybridizations. pMB358 is a clone isolated from CMS mitochondrial DNA (mtDNA) and contains part of repeat II (*hatched area*). The extent of repeat II-homologous sequence upstream of the MF *atp9* gene is also indicated. Sites for several restriction enzymes are also shown (E, *Eco*RI; B, *Bam*HI; K, *Kpn*I; P, *Pst*I; S, *Sma*I; X, *Xho*I; B, *Bgl*II; Ss, *Sst*I). **B** Northern analysis of the sugarbeet *nad2* locus using pFB124 as probe to MF mtRNA (lane 1) and CMS mtRNA (lane 2). Transcript sizes are indicated in kilobases. **C** DNA sequence of the sugarbeet *nad2* gene and its deduced amino acid sequence. The DNA sequence was translated according to the "universal" genetic code. The second in-frame ATG codon is denoted by an *arrow*, and putative ribosome-binding sites are *boxed*. DNA sequences downstream of the *nad2* gene capable of forming a hairpin loop structure which may be involved in transcript processing or termination are indicated by *horizontal arrows*

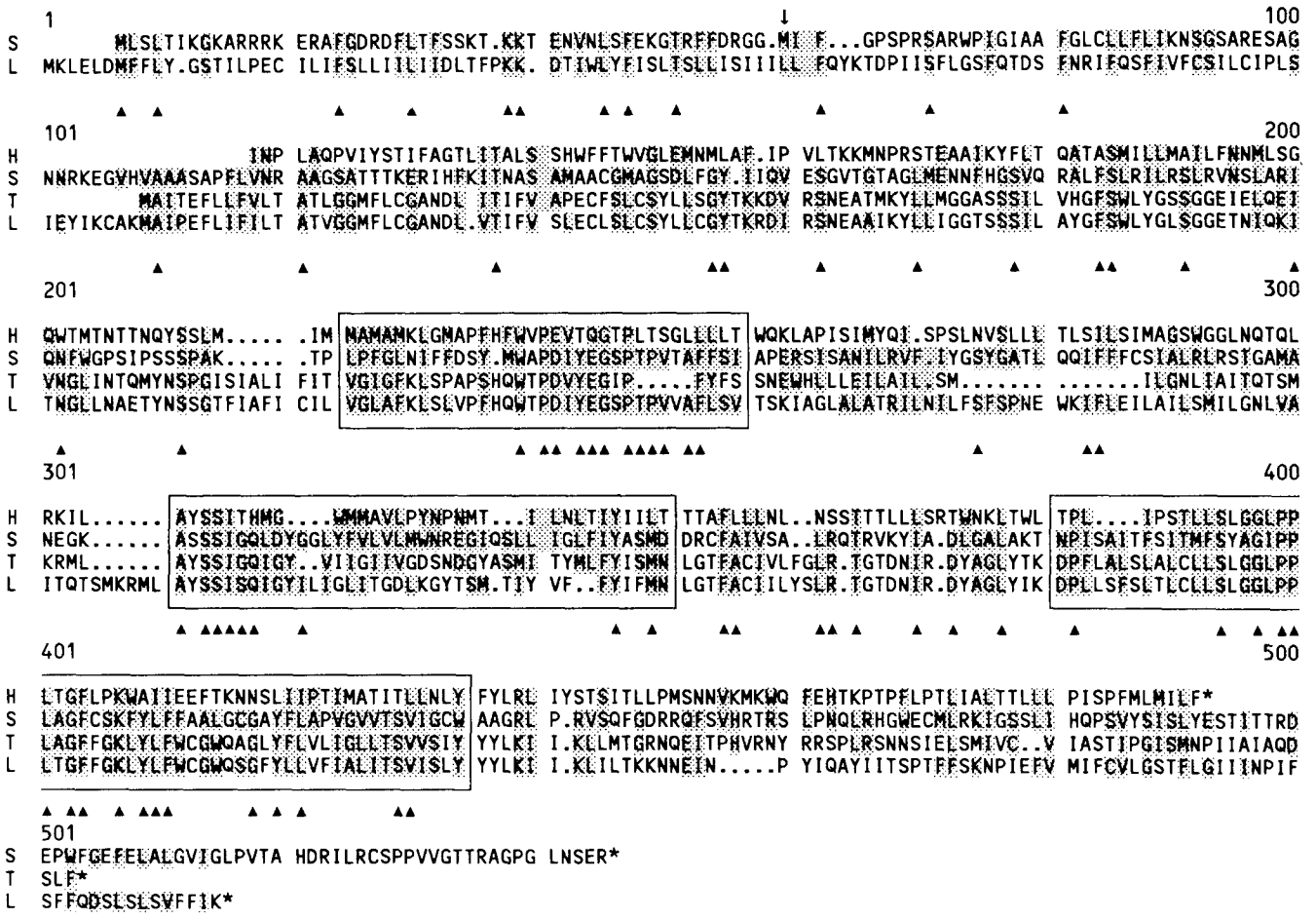


Fig. 2. Comparison of the deduced sugarbeet NAD2 polypeptide (S) with the corresponding human mitochondrial protein (H) and chloroplast proteins from tobacco and liverwort (T and L respectively). The second in-frame methionine of the sugarbeet sequence is indicated by an *arrow*. Amino acids which are conserved in at least the three plant proteins are indicated by *arrowheads*, and conservative amino acid replacements are *shaded*. The three most highly conserved sequence domains common to all four proteins are *boxed*. Gaps were introduced into the sequences in order to maximize the alignment

quences 5′ to the repeat II-homologous sequence since pMB358 detects different transcripts in both genotypes at much lower abundance (data not shown). The DNA sequence of pFB124 was determined. Analysis of the sequence revealed a large uninterrupted open reading frame of 515 amino acids (Fig. 1C). A bias towards the use of T in the third base position of codons is apparent, a feature of some plant mitochondrial genes (Isaac et al. 1985), and a consistent feature of sugarbeet mitochondrial genes including *atpA*, *atp6* and *atp9* (data not shown). The deduced amino acid sequence was found to show extensive homology to NAD2 proteins from liverwort and tobacco chloroplasts (Ohyama et al. 1986; Shinozaki et al. 1986), and mammalian mitochondria (Anderson et al. 1981). The predicted NAD2 polypeptides from sugarbeet mitochondria, liverwort and tobacco chloroplasts and from human mitochondria were compared and are shown in Fig. 2. The sugarbeet protein shows 46.8%, 43.4% and 34.3% similarity (including conservative amino acid substitutions) with the analogous tobacco, liverwort and human proteins respectively. Two CGG codons are present in the sugarbeet *nad2* gene which may encode tryptophan rather than arginine (Fox and Leaver 1981). However, one of these codons (at amino acid 335) occurs at a well conserved arginine residue in the corresponding chloroplast proteins (Fig. 2).

Transcription of the nad2 *locus in CMS and MF mitochondria*

At least three transcripts of 4.9, 3.6 and 1.7 kb are detected in both CMS and MF mitochondria when pFB124 is used as a probe (Fig. 1B). When the 1310 bp *Eco*RI-*Xho*I fragment encompassing the C-terminal region of the gene was used in Northern hybridizations (Fig. 1A), an identical set of transcripts was detected (data not shown). Since pMB358 detects different transcripts from *nad2*-specific probes the 3′ termini of these transcripts must be located between the end of the *nad2* gene and the repeat II-homologous sequence, and these transcripts may therefore be 3′ co-terminal. Southern analysis with the *Eco*RI-*Xho*I probe revealed that it hybridized to single fragments from *Eco*RI- or *Bam*HI-digested MF mtDNA of 4.1 and 6.5 kb respectively, and to *Eco*RI and *Bam*HI fragments of 2.4 and 5.7 kb respectively in CMS mtDNA. This indicates that sugarbeet mitochondria possess a single copy of the *nad2* gene and the three transcripts are not derived from multiple gene copies.

Discussion

Sugarbeet NAD2 is encoded by an uninterrupted DNA sequence in common with the mammalian and *Neurospora*

genes (Anderson et al. 1981; de Vries et al. 1986) and in contrast to its chloroplast counterpart (Ohyama et al. 1986; Shinozaki et al. 1986). Introns are not a consistent feature in genes for NADH: ubiquinone reductase subunits in higher plants. The *nad5* gene from another dicotyledonous species (*Oenothera*), contains two introns (Wissinger et al. 1988), and the *nad1* gene from watermelon has at least one (Stern et al. 1986), whereas the petunia, wheat and maize *nad3* genes lack introns (Gualberto et al. 1988; Rasmussen and Hanson 1989).

The precise translation initiation site of sugarbeet NAD2 is unknown. Two octanucleotide sequences showing 62.5% homology with a proposed ribosome binding site for plant mitochondrial genes (Dawson et al. 1984) are found close to the second in-frame ATG (Fig. 1C). Similar sequence motifs are observed in several other sugarbeet mitochondrial genes, in particular those encoding subunits of the mitochondrial ATPase (data not shown). In addition, the first 50 amino acids of the deduced sugarbeet polypeptide appear markedly more hydrophilic than the chloroplast and mammalian proteins, despite overall similarities in their hydrophobic properties (data not shown).

At least three *nad2*-specific transcripts are present in MF and CMS mitochondria. The size of the smallest transcript closely corresponds to that of the proposed coding region. The origin of the larger transcripts is unclear. Linkage and co-transcription of several plant mitochondrial genes has been observed, in particular the co-transcription of *nad5* with the 5 S rRNA gene in *Oenothera* (Wissinger et al. 1988), and *nad3* with the CMS-inducing gene (*S-pcf*) in petunia (Rasmussen and Hanson 1989) and the *rps12* gene in wheat and maize (Gualberto et al. 1989). No substantial open reading frames have been detected upstream or between *nad2* and the repeat II-homologous sequence. However, considering the large size of *nad2* transcripts, it is possible that sugarbeet *nad2* is part of a large transcriptional unit encompassing additional gene(s).

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.X. is grateful for a 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.

References

Anderson S, Bankier AT, Barrel BG, deBruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Brears T, Lonsdale DM (1988) The sugarbeet mitochondrial genome: a complex organization generated by homologous recombination. Mol Gen Genet 214:514–522

Dawson AJ, Jones VP, Leaver CJ (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 TD, Leaver CJ (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

Gualberto JM, Wintz H, Weil JH, Grienenberger JM (1988) The genes coding for subunit 3 of NADH dehydrogenase and for ribosomal protein S12 are present in the wheat and maize mitochondrial genomes and are co-transcribed. Mol Gen Genet 215:118–127

Isaac PG, Jones VP, Leaver CJ (1985) The maize cytochrome *c* oxidase subunit I gene: sequence, expression and rearrangements in cytoplasmic male sterile plants. EMBO J 4:1617–1623

Leaver CJ, Hack E, Forde BG (1983) Protein synthesis by isolated plant mitochondria. Methods Enzymol 97:476–484

Lonsdale DM (1988) Plant mitochondrial genes and sequences. Plant Mol Biol Rep 6:266–273

Makaroff CA, Palmer JD (1987) Extensive mitochondrial specific transcription of the *Brassica campestris* mitochondrial genome. Nucleic Acids Res 15:5141–5256

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Rasmussen J, Hanson M (1989) A NADH dehydrogenase subunit gene is co-transcribed with the abnormal gene associated with cytoplasmic male sterility. Mol Gen Genet 215:332–336

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci 74:5464–5467

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Stern DB, Bang AG, Thompson WF (1986) The watermelon URF-1 gene: evidence for complex structure. Curr Genet 16:857–869

Thomas CM (1986) The nucleotide sequence and transcription of minicircular mitochondrial DNAs associated with male fertile and cytoplasmic male sterile lines of sugarbeet. Nucleic Acids Res 14:9353–9370

deVries H, Alzner-DeWeerd B, Breitenberger CA, Chang DD, deJonge JC, RajBhandary UL (1986) The E35 stopper mutant of *Neurospora crassa*: precise localization of deletion end points in mitochondrial DNA and evidence that the deleted DNA codes for a subunit of NADH dehydrogenase. EMBO J 5:779–786

Wissinger B, Hiesel R, Schuster W, Brennicke A (1988) The NADH-dehydrogenase subunit 5 gene in *Oenothera* mitochondria contains two introns and is co-transcribed with the *5S* rRNA gene. Mol Gen Genet 212:56–65

Communicated by D.M. Lonsdale

Received September 21, 1989