SUPPLEMENTAL DATA

Methods

Identification of 5’ and 3’ transcript termini

5’ transcript termini were determined using the GeneRacer^{TM} kit for RNA ligase-mediated rapid amplification of 5’ and 3’ cDNA ends (Invitrogen). To identify primary transcript 5’ termini, 4 µg of DNase I treated mtRNA from each plant line were treated with calf intestinal phosphatase (CIP). RNA was extracted with phenol/chloroform, precipitated with ethanol, and 5’ triphosphates were converted to monophospates by treatment with tobacco acid pyrophosphate (TAP). mtRNA was then extracted with phenol/chloroform, precipitated with ethanol and divided into two parts. For 5’ RACE analysis, one part was supplemented with 0.25µg GeneRacer^{TM} RNA Oligo and transcripts were attached to the adapter using RNA ligase. The other portion of the TAP treated mtRNA was circularized at 37°C for 1h with 50U of T4 RNA ligase. Following RNA ligation, the mtRNA was again extracted and precipitated and finally resuspended in water. The 5’adapter ligated mtRNAs were amplified by RT-PCR using the GeneRacer^{TM} 5’ primer and a gene specific primer. Circularized mtRNAs were amplified by RT-PCR gene specific primers. Amplification products were separated, cloned and sequenced. To analyze endonuclease-processed transcripts, mtRNA that had not been treated with either CIP or TAP was similarly analyzed.

Prediction of RNA secondary structure

Because the folding software M-fold (30; http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) only evaluates secondary structure in a single RNA molecule, to evaluate potential secondary structures that could contribute to mis-splicing it was necessary to first fuse, *in silico*, the downstream sequence (extending from the portion of nad5Ti2R upstream of exon c to the mapped 3’ terminus of nad5Ti3L) to the upstream sequence (extending from the mapped 5’ termini of the nad5a/b transcript to the mapped 3’ end of nad5Ti2L, but excluding the cis-spliced intron nad5i1). Potential secondary structures were evaluated by adjusting parameters such that only duplexes forming between the downstream and

upstream sequences were considered. This analysis indicated a number of potentially stable secondary could form between exon a and the 3’ end of nad5 Ti3L. The structure shown in Fig. 6A is that particular structure associated with the most negative free energy value.

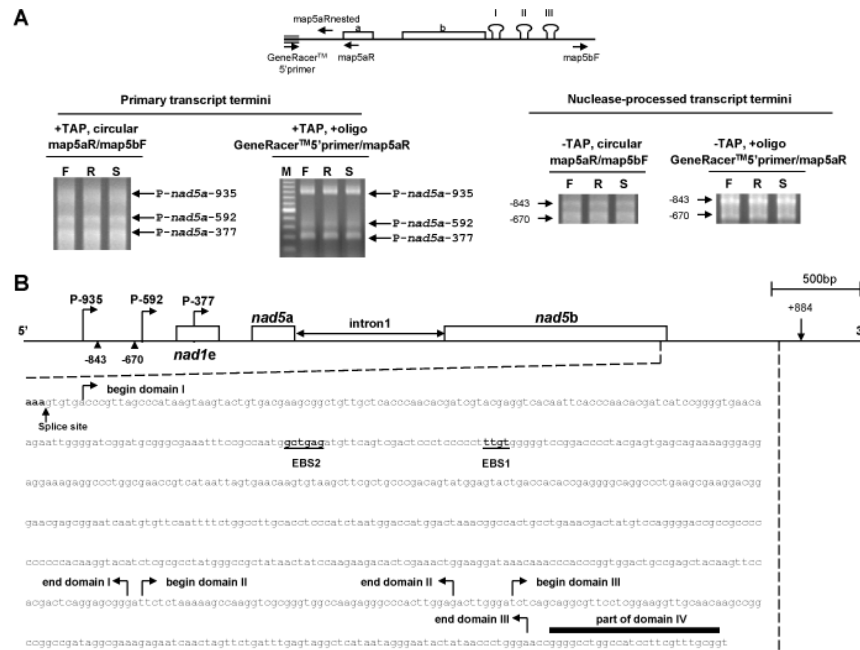
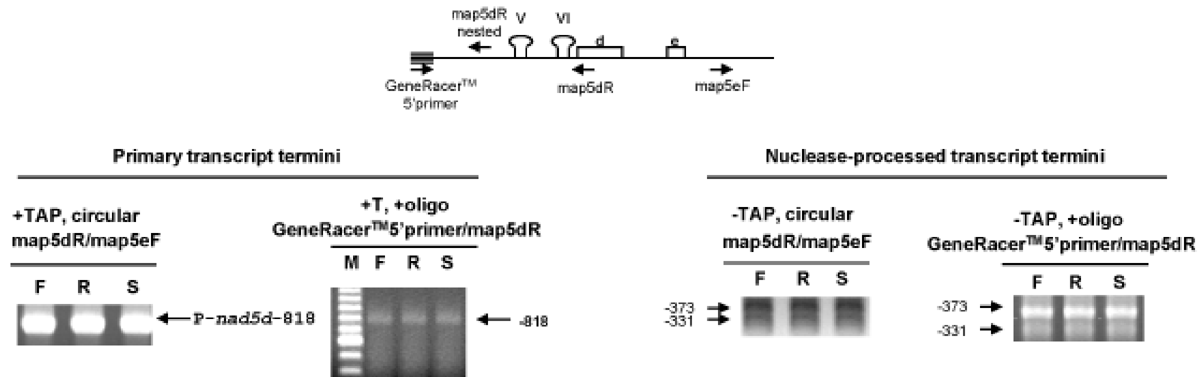


Figure 1 Mapping of primary and processed *nad5*a/*nad5*b transcript termini.

(A) Top: Schematic representation of the *nad5*a/*nad5*b region of *B. napus* mtDNA. Open boxes indicate the positions of *nad5* exons; arrows indicate the approximate positions and orientations of primers used in amplification reactions. The GeneRacer™ primer is a DNA oligonucleotide that primes within the GeneRacer™ RNA that is ligated to transcripts for 5' termini mapping by RACE. Stem loops are used to illustrate the presence of group II intron domains D1, D2 and D3 in the transcript.
Bottom: Gel photographs of RT-PCR products obtained during mapping of primary (left) and nuclease processed (right) RNA ends. Mapping of primary transcript termini was conducted by treating mtRNA with calf intestinal phosphatase (CIP), followed by tobacco acid pyrophosphatase (TAP). This RNA was either circularized with RNA ligase (circular), or joined to the GeneRacer™ RNA (+oligo) by RNA ligase; RT-PCR products were then generated by amplification using the indicated oligonucleotides separated on agarose gels and excised and sequenced to determine the precise location of 5' and 3' ends. RNAs with only a 5' terminal monophosphate were assumed to have been generated by nuclease processing, and were mapped by direct circularization or oligonucleoted ligation, then subjected to RT-PCR amplification using the indicated oligonucleotides (Nuclease Processed Transcripts).
(B) Schematic representation of the *nad1*e/*nad5*a/*nad5*b transcription unit. Products corresponding to primary transcript 5' ends are indicated by notched arrows and labeled according to the distance from the 5' end of *nad5*a. Solid triangles mark processing sites. The downward arrow points to the position of the mature 3' transcript terminus. Features of the 3' UTR are depicted in the sequence of the expanded region downstream of *nad5*b below; notched arrows illustrate the regions comprising the various group II intron structural domains of nad5Ti2L; EBS1 and EBS2 (bold underline) are sequences within domain I (D1) that pair with complementary exon sequences bordering the 5' splice site; the portion of domain IV (D4) that base pairs with a complementary sequence on nad5 Ti2R upstream of *nad5*c to form a splicing-competent structure is indicated by the horizontal bar.



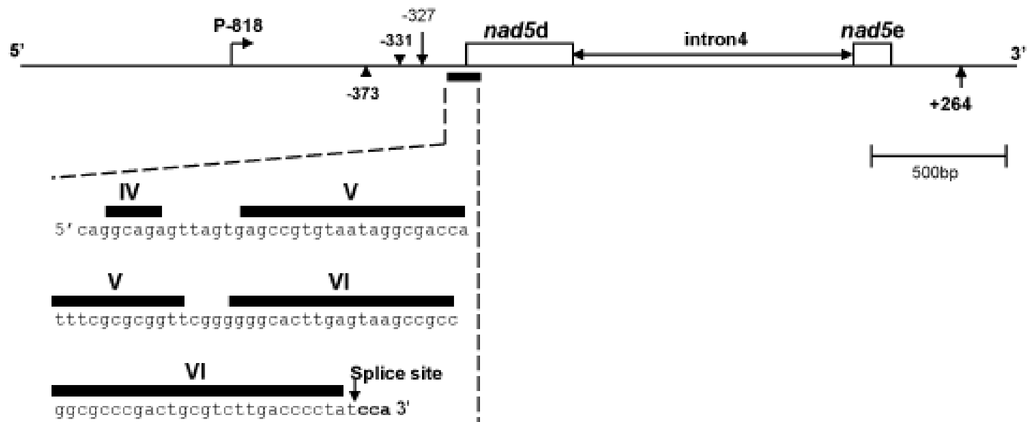


Figure 2. Mapping of primary and processed *nad5*d/*nad5*e transcript termini.

(A) Top: Schematic representation of the *nad5*a/*nad5*b region of *B. napus* mtDNA. Open boxes indicate the positions of *nad5* exons; arrows indicate the approximate positions and orientations of primers used in amplification reactions. Other symbols are the same as in Figure S1. Bottom: Gel photographs of RT-PCR products obtained during mapping of primary (left) and nuclease processed (right) RNA ends. Mapping of transcript termini was as described in Figure 1.

(B) Schematic representation of the *nad5*d/*nad5*e transcription unit. The site of the primary transcript 5' terminus is indicated by the notched arrow. Solid triangles and the downward arrow indicate processing sites. The upward arrow points to the position of the mature 3' transcript terminus. Features of the nad5Ti3 intron upstream of *nad5*d are depicted in the sequence of the expanded region; horizontal bars indicate the regions comprising the group II intron structural domains. The sequences under the bar indicating domain IV (D4) pair with complementary sequences on nad5Ti3L, located downstream of *nad5*c, to form a splicing-competent structure.

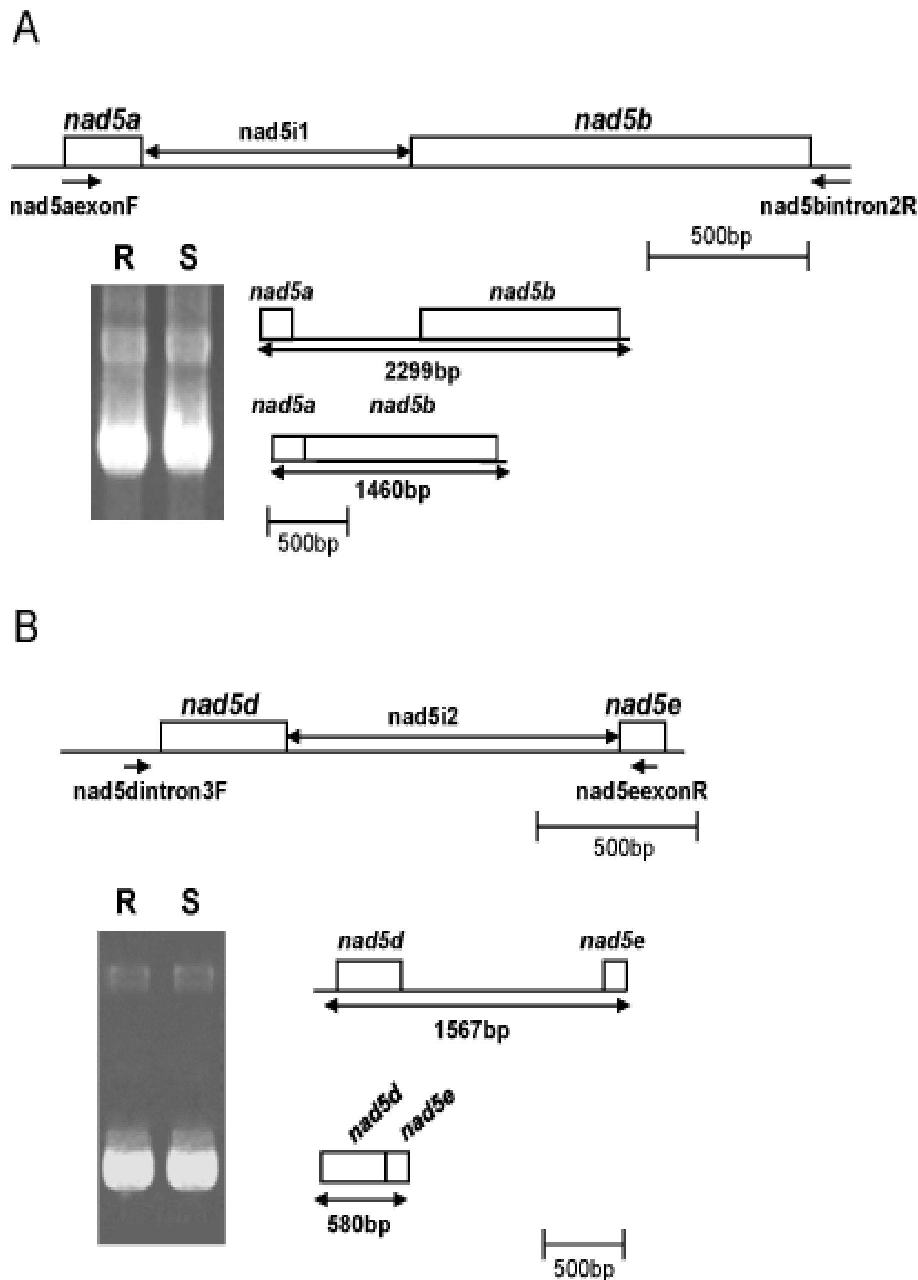


Figure 3. Cis-splicing of *Brassica nad5* transcripts. **(A)** *nad5* a/b transcripts; **(B)** *nad5* d/e transcripts (right). Locations and orientations of the primers employed in RT-PCR analysis are indicated in the upper diagrams. Gel images of the products obtained by RT-PCR, aligned with an illustration of the corresponding product and its size, are indicated below. The structures of the products were confirmed by sequencing individual clones of each gel-purified product, from both plant lines. R, fertility-restored line; S, CMS line.

A

	nt position	No. edited clones vs total no.cloned	
		Pre *cis*-splicing	Post *cis*-splicing
a	+154	5/5	5/5
b	+13	5/5	5/5
	+43	5/5	5/5
	+128	3/5	4/5
	+145	4/5	4/5
	+169	3/5	3/5
	+265	4/5	4/5
	+310	2/5	3/5
	+319	5/5	5/5
	+324	4/5	4/5
	+369	4/5	5/5
	+379	4/5	4/5
	+400	5/5	5/5
	+447	5/5	5/5
	+484	4/5	4/5
	+496	4/5	4/5
	+535	3/5	3/5
	+606	5/5	5/5
	+634	5/5	5/5
	+646	5/5	5/5
	+1171	4/5	4/5

B

	nt position	No. edited clones vs total no.cloned	
		Pre *cis*-splicing	Post *cis*-splicing
d	+22	8/8	8/8
	+82	7/8	8/8
	+112	3/8	6/8
	+142	6/8	8/8
e	+33	8/8	8/8
	+54	3/5	3/5
	+56	3/5	4/5
	+96	8/8	8/8

Figure 4. Editing sites and frequency in pre- and post- cis-spliced nad5a/b and nad5d/e transcripts.

The positions of the editing sites within each exon is indicated in the "nt position" column. The number of edited events at each site, as determined from the analysis of different cDNA clones obtained from pre and post cis-spliced transcripts is indicated under the "No. edited clones vs total no. cloned" column. The location of the edited cytosine is denoted as from its distance from the first nucleotide of the individual exon.

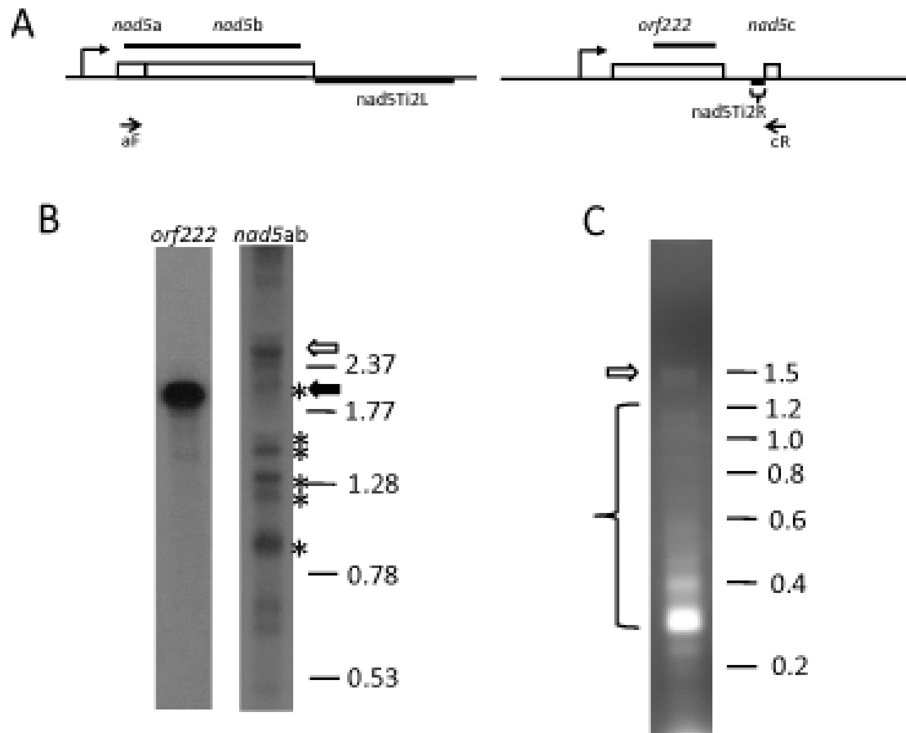


Figure 5. Relative levels of mis-spliced and properly trans-spliced *nad5*a-b-c transcripts.

(A) Schematic representation of *nad5* structural elements indicating locations of RNA gel blot probes (horizontal bars above diagrams) and RT-PCR primer sequences (arrows below diagrams. Notched arrows indicate the sites of transcription initiation closest to site of translation initiation.
(B) RNA gel blot showing RNAs detected using an exon a-b probe derived from an RT-PCR product corresponding to cis-spliced a-b transcripts. Open arrow indicates the position expected of the mature, fully-spliced nad5 transcript, filled arrow the position of a nad5a-b transcript that has undergone cis- but not trans-splicing. Sites with asterisks indicate RNAs of sizes expected of mis-spliced mis-spliced transcripts. Lane on the left shows same mtRNA preparation probed with orf222 sequences, demonstrating the integrity of RNA.
(C) RT-PCR analysis of mtRNA using a forward primer from exon a and reverse primer from exon c. The open arrow corresponds to the product obtained from correct b to c joining; the bracketed region indicates products derived from mis-spliced transcripts.

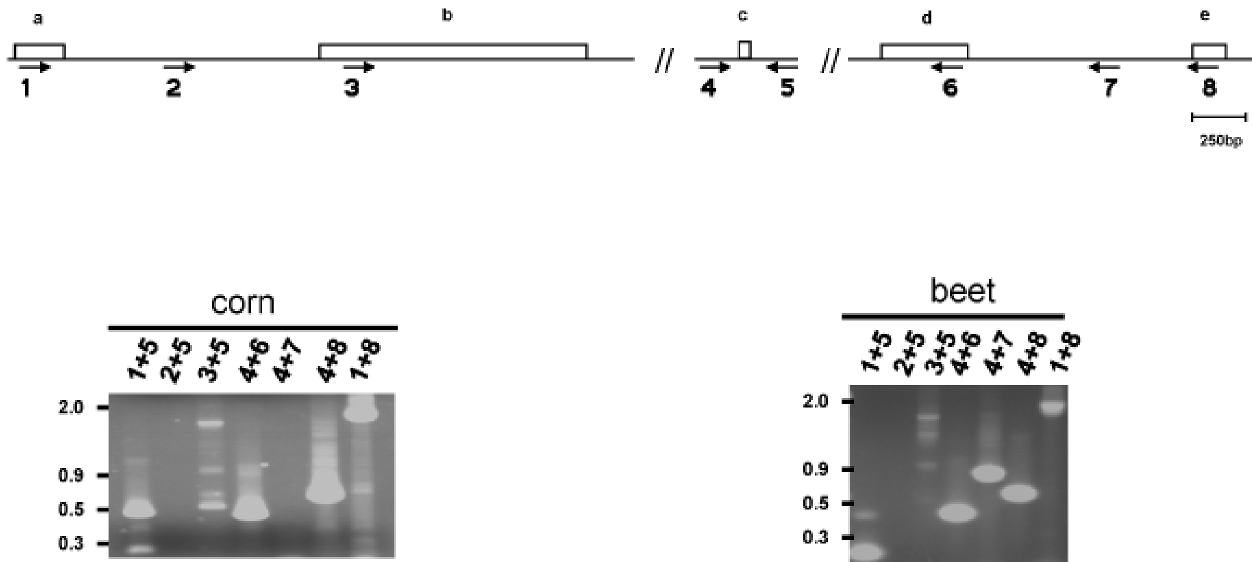


Figure 6. Splicing and mis-splicing of *nad5* exons in corn (*Zea mays*) and sugar beet (*Beta vulgaris*).

In the top diagram, open boxes indicate the positions of exons in mtDNA and arrows indicate the positions and orientation of primers (numbered 1 through 8) used in RT-PCR amplifications. The images below show RT-PCR products obtained using each primer combination. The small products obtained using primer pairs 1 + 5 and 3 + 5 correspond to mis-spliced RNAs.

Table 1. Promoter sequences at nad5 transcription initiation sites in *B. napus (nap)* mtDNA. Sequences corresponding to core motifs identified by Kuhn et al are underlined. Initiating nucleotides are capitalized.

Gene	Promoter site	Sequence
*nad5*a/b	-935	tt**atta**gtAaagcgcta
	-592	ctcataagAgaagaaat
	-377	ggcgtaagcAatgattc
*nad5*d/e	-818	TcgtaagatAagaaaga
orf222/*nad5*c	-424	tagataaattaAgtatt
	-150	aatctcataagAgaaga

Table 2. Primer sequences

Primer designation	**Primer sequence (5' to 3')**
Figure 2	
nad5aexonF	ttgattgctttttatgaagtcgcact
nad5bintron2R	tcacagtacttacttatgggctaa
nad5dintron3F-	gcacttgagtaagccggcgcc
nad5eexonR	tgcttaccaagccatcctggcaag
Figure 3	
aF	ttgattgctttttatgaagtcgcact
cintron2F	atgtatctacttatcgtattttg
cintron3R	cagggtttggtggccgaaag
222F	aacgaaatactcgatctgtttcga
eR	ttcttgatctgactttgtataaaa
Figure 5	
nad5bintron2R	tcacagtacttacttatgggctaa
nad5cRE	atgtatctacttatcgtattttg
nad5c3'intron	gcctctcacgcactaatccctaca
nad5bexon2R	tcatggcataggtcaaagggaagg
nad5dintron2R	gcacttgagtaagccgccggcgcc
Figure 7	
woF1	cttcaggcggataaagctgctata
woL1	gtcgactgaacatctcagccatt
woL2	gttaggactttgtctcccttttcgt
oeR1	cagtttgtggatcgcggtctcacg
whR1	cagtttgtggatcgcggtctcacg
oeL3	gaagccgagcctttccagcaa
whL3	gaagccgagcctttccagcaa
oeR2	ctactcactatccaaatgaaagacg
whR2	ctactcactatccaaatgaaagacg
Supplemental data Fig. 1	
map5aRnested	agcgattgagtgatagatttctcg
map5aR	ttcgagtgtgatcagaaacctaaa
map5bF	cattccttcaatacttctggctct
Supplemental data Fig. 2	
map5dRnested	cgccattataataatggcgagacaag
map5dR	tattacacggctcactaactctgc
Supplemental data Fig. 4	
aF	ttgattgctttttatgaagtcgcact
cR	tacctaaaccaatcatcatatc
Supplemental data Fig. 5	
1	atgtatctacttattgtctttttttgcct
2	gcctgactcgtaattcacttttga
3	cttcaggcggataaagctgctata
4	gaagccgagcctttccagcaa3
5	cagtttgtggatcgcggtctcacg3
6	tccactttgaagttgacttattcgctc
7	cactcactgcttcccctaatatcc3
8	ctcttgacttgacttattaataagaaaact