ORIGINAL PAPER

S. Zanlungo · V. Quiñones · A. Moenne · L. Holuigue X. Jordana

Splicing and editing of rps10 transcripts in potato mitochondria

Received: 19 August 1994 / 17 November 1994

Abstract The structure and expression of the potato mitochondrial gene rps10, encoding ribosomal protein S10, has been characterized. The RPS10 polypeptide of 129 amino acids is encoded by two exons of 307 bp and 80 bp respectively, which are separated by a 774-bp class-II intron. Editing of the complete rps10 coding region was studied by sequence analysis of spliced cDNAs. Four C residues are edited into U, resulting in the creation of a putative translational initiation codon, a new stop codon which eliminated ten carboxy-terminal residues, and two additional amino-acid alterations. All these changes increase the similarity between the potato and liverwort polypeptides. One additional C-to-U RNA editing event, observed in the intron sequence of unspliced cDNAs, improves the stability of the secondary structure in stem I (i) of domain I and may thus be required for the splicing reaction. All spliced cDNAs, and most unspliced cDNAs, were completely edited, suggesting that editing is an early step of rps10 mRNA processing and precedes splicing. Earlier work on potato rps10 (Zanlungo et al. 1994) is now known to comprise only a partial analysis of the gene, since the short downstream exon was not identified.

Key words Group-II intron editing · rps10 transcript splicing · S10 ribosomal protein · Solanum tuberosum mitochondria

Introduction

The mitochondrial DNA (mtDNA) of flowering plants, with sizes between 200 and approximately 2500 kb, is

Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, P. Universidad Católica de Chile, P.O. Box 114-D, Santiago, Chile

much larger and more complex than that of animals and fungi (Levings and Brown 1989; Bonen 1991), and can thus potentially accomodate many more protein-coding genes. Contributing to the extended size of the genome are also large spacer sequences, promiscuous DNA of chloroplast and nuclear origin, numerous duplications, and sequences required for expression including introns. In flowering plants a number of mitochondrial genes are known to contain introns, all of which are group-II introns characterized by distinctive secondary structural features (Michel et al. 1989; Gray et al. 1992). In addition to splicing, RNA editing is required for maturation of mRNAs in plant mitochondria (Gray et al. 1992). In higher-plant mitochondria this modification occurs almost exclusively by C-to-U conversions and has so far been found in mRNAs of all conserved coding regions investigated. In most cases, editing leads to changes in the amino-acid sequence that result in a protein with a sequence better conserved when compared to the homologous protein in other organisms. Therefore, RNA editing seems to be essential to ensure the synthesis of functionally competent proteins.

Ribosomal protein genes comprise a significant fraction of the genes identified uniquely in plant mitochondria. While in vertebrates and many invertebrates these proteins are encoded by the nuclear genome, several ribosomal-protein genes have been identified in the mtDNA of higher plants, including rps3, rps7, rps12, rps13, rps14, rps19, rpl5 and rpl16 (Grohmann et al. 1993). Sixteen ribosomalprotein genes, including all of the above, have been identified in the complete sequence of the mitochondrial (mt) genome from the liverwort (Oda et al. 1992; Takemura et al. 1992).

Some of the ribosomal-protein genes that have been identified in the mtDNA of certain higher plants are incomplete in, or else absent from, the mt genome of other plants (Grohmann et al. 1993). Evolutionarily recent transfer events of functional genes to the nucleus, as shown for cox2 in legumes and rps12 in Oenothera (Nugent and Palmer 1991; Covello and Gray 1992; Grohmann et al. 1992), can explain these differences in the mt ribosomalprotein gene content between different plant lineages.

S. Zanlungo · V. Quiñones · A. Moenne · L. Holuigue X. Jordana (🖂)

We have recently reported the identification of an open reading frame (ORF) in the mitochondrial genome of *Solanum tuberosum* with significant similarity to ribosomal protein S10 genes (*rps10*) (Zanlungo et al. 1994). This *rps10* ORF is absent from wheat mtDNA. These results suggest a novel example of a rather recent gene-transfer event from the plant mitochondria to the nucleus. In this article we extend the results of our investigations on the organization and expression of the potato mitochondrial *rps10* gene and show that the *rps10* ORF is split by a group-II intron and undergoes RNA editing in both exons and the intron.

Materials and methods

Isolation of mitochondrial nucleic acids and cloning of the rps10 gene. Mitochondrial DNA and RNA were isolated from potato tubers (*S. tuberosum* cv Bintje) as previously described (Dell'Orto et al. 1993). A 5.1-kb *Eco*RI clone containing the potato *rps10* gene (located downstream from a truncated *cob* pseudogene) has been previously isolated from a potato mtDNA library (Zanlungo et al. 1994).

cDNA synthesis, PCR amplification and cDNA cloning. Potato mtRNA (50 μ g) was treated with 40 units of RNase-free DNase I (BRL) in the present of 40 units of RNasin (BRL), extracted with phenol, and precipitated with ethanol. First-strand cDNA synthesis and PCR amplification of the cDNA were performed as previously described (Dell'Orto et al. 1993), using the following primers (located as depicted in Fig. 1B):

1: 5'-CAAGGAGGATCCTATCCTAAAG-3' (BamHI site italicized) 2: 5'-AATTTGGATCCCCCTTTCCTCT-3' (created BamHI site italicized)

3: 5'-CCAGTCGGCAAGCT7TTTGGC-3' (created HindIII site italicized)

PCR products were analysed by electrophoresis through agarose gels, digested with BamHI or BamHI + HindIII, and ligated to BluescribeTM vectors (Stratagene).

DNA sequencing. DNA sequencing was performed on single- or double-stranded templates by the dideoxy chain-termination method, using a T7 DNA polymerase sequencing kit (Pharmacia LKB). Both strands of the genomic DNA were completely sequenced.

Northern-hybridization analysis. Total mtRNA (5–10 µg) was fractionated through agarose-formaldehyde gels, transferred to nylon membranes and hybridized using standard procedures (Sambrook et al. 1989). The oligonucleotide probes used were oligo 4 (exon-1 probe, 5'-ACTATGCCTATCTTGGTGGTC-3') and oligo 5 (intron probe, 5'-TACACTTCTTGAATTCTGTCCCAC-3'), and were labelled using T4 polynucleotide kinase and [γ -³²P]ATP as described by Sambrook et al. (1989).

Results and discussion

Identification of two potato rps10 exons

The potato rps10 gene has been previously identified on a 5.1-kb *Eco*RI restriction fragment of potato mtDNA, downstream from a truncated *cob* pseudogene and upstream of *cox1* (Fig. 1 A). Sequence analysis identified an open reading frame (positions 1–324 in Fig. 1B) whose

translation product displays similarity to bacterial, chloroplast, and liverwort mitochondrial RPS10 sequences (Zanlungo et al. 1994). However, comparison of liverwort and potato amino-acid sequences revealed a 18 amino-acid C-terminal extension in the liverwort polypeptide. This apparent divergence les us to determine additional downstream sequences. A stretch of 89 bp (positions 1088-1176 in Fig. 1 B) could be aligned with the sequence coding for the 20 C-terminal amino acids of the liverwort polypeptide (and the 3' flanking sequence). Analysis of the sequence immediately upstream of this 89-bp stretch identified conserved group-II-intron domains V and VI (discussed below), suggesting that a novel intron interrupts the rps10 coding sequence in potato mitochondria, while in liverwort this gene has no introns (Oda et al. 1992). Either gain of the intron in potato or its loss in Marchantia must have occurred after the divergence of mosses and angiosperms.

The presence of the intron was confirmed by sequence analysis of spliced *rps10* cDNAs. Mitochondrial RNA was reverse transcribed using primer 2, and amplified by PCR between primers 2 and 1 (Fig. 1 B). The main PCR product of 0.5–0.6 kb leads to an estimation of 0.7–0.8 kb for the intron length (data not shown). The position of the *rps10* 5' and 3' splice sites was precisely determined by sequence analysis of several cloned PCR-amplified cDNAs (Fig. 2). The two exons are separated by an intron of 774 bp and are predicted to encode 102 and 27 amino acids respectively (Fig. 1 B), taking into account RNA editing events (discussed below).

In our previous work the potato rps10 nucleotide sequence was aligned with a single-copy sequence located upstream of the pea cox1 gene (Kemmerer et al. 1989; Zanlungo et al. 1994). The published pea sequence does not contain a complete uninterrupted open reading frame, since deletions and insertions disturb the coding sequence and introduce frameshifts and stop codons. However, in light of the fact that the published homologous region in pea mitochondria contains rps10 exon-1/group-II intron/exon-2 sequences, these frameshifts might simply reflect sequencing errors in the early work by Kemmerer et al. (1989). This possibility has been considered by V. Knoop et al. (1995) who re-investigated this region in the pea mitochondrial genome and found a functional rps10 gene similar to the potato gene.

Intron structure

Several features of the intron are clearly indicative of a group-II intron (Michel et al. 1989): a conserved block at the 5' splice site (GUGCG), the helical domain-V structure, and the bulging A located 8 nt upstream of the 3' intron-exon junction on the 3' side of helix VI (this is the nucleotide which participates in "lariat" formation by covalent attachment to the 5' intron end during attack of the 5' splice site). The *rps10* intron can be folded into the conservative group-II structural model consisting of six domains (I through VI) radiating outward from a central core

Fig. 1A, B Nucleotide sequence of the potato rps10 gene. A restriction map of the potato mitochondrial pseudocob-rps10-cox1 locus. Black bars represent rps10 exons and the coxl coding region, and the hatched box the truncated cob coding region. The direction of transcription is indicated by arrows. A, AvaI; B, BamHI; E, EcoRI; H, HindIII; S, SalI. **B** nucleotide sequence of *rps10* (EMBL Accession number X74826). The genomic sequence is presented, with the derived amino-acid sequence given above the triplets in the single letter code. Sequence numbers are in relation to the rps10 genomic ATG codon at position 1. C-to-T conversions detected in rps10 cDNAs are shown below the genomic sequence and the altered codons are translated underneath. Sequences identical in the liverwort rrn26 intron are underlined. Arrows show the position of oligos (1-5) used in PCR amplifications and Northern analysis. The cox1 initiation codon is created by editing (unpublished results) and is located 61 nt downstream from the sequence shown



-123	TAC	[ACT]	AGA	АААТ(GGTG/	AG AG	CAT	TAGT	J AA	AGAA	GGAC	CAA	GAG	JAT	<u>ССТА:</u> 1	ICCT!	<u>AA A(</u>	<u></u> GAG	AAGGI	A
																м	R	Q	R	R
- 53	GGAO	GAG'	rag	GAGC?	TAGCI	T TZ	AGGGG	GCGG/	A ATO	GAA	FGAT	TACO	GAGA:	ГАА .	ACA	<u>ATG</u>	AGA	CAA	AGG	AGA
	A	L	R	R	v	S	Q	ĸ	Е	R	₽	P	ĸ	v	Т	т	т	к	I	G
16	GCA	CTT	AGA	CGA	GTC	AGC	CAA	AAA	GAA	AGA	CCA	CCA	ААА	GTA	ACG	ACC	ACC	AAG	ATA	GGC
															T				4	
	I	v	I	R	s	F	D	н	₽	F	L	Е	N	н	F	W	G	L	P	P
76	ATA	<u>GT</u> A	ATT	CGA	TCT	TTT	GAT	CAC	CCA	TTT	TTG	GAA	AAC	CAT	TTT	TGG	GGG	CTT	CCG	CCT
	Y	T	R	K	I	G	L	P	E	S	R	V	L	Y	T	V	L	R	S	P
136	TAC	ACA	CGG	AAG	ATT	GGA	116	CCT	GAA	TCA	CGA	GIC	TIA	TAT	ACT	GIG	TIA	ÇGA	ICA	CCI
	н	I	D	к	ĸ	s	R	Е	о	F	F	м	к	I	ĸ	к	E	F	L	v
196	CAT	- ATT	GAT	AAA	AAG	тсс	AGA	GAA	CÃA	TTT	TTT	ATG	AAA	ATA	AAG	AAA	GAA	TTT	CTG	GTC
															-					
	I	ĸ	т	Е	R	н	Е	г	R	ĸ	ĸ	F	F	R	L	ĸ	R			
256	ATA	àaa	ACA	GAA	AGG	CAT	GAA	TTG	CGC	AAG	AAG	TTC	TTT	CGG	TTA	AAA	CGC	С		
														T						
										,			3		J					
308	GTG	GAC	īde	GAGG	ACAT	A G	ACTT(CTTG	э тс	AAGC	CAAA	AAG	ATTG	CCG .	ACTG	- JATG(CT CO	CTAC	CCCA	2
			ЪГ.										-		-					
			岜																	
			Ë											۷			5			-
378	CAT	CCT	명 GG <u>C</u>	CCTT	TATC	TT A	CCTT	AAGA	A G <u>A</u>	AAGA	GGAG	GTA:	GAA		TGGG	ACAG	5	GCAA	JAAG	<u> </u>
378 448	CATO GTA	<u>SCCT</u> FGAT	GG <u>C</u> ACA	<u>CC</u> TT A <u>CA</u> G	TATCI ATAA(TT A	CCTTI FA <u>AG(</u>	AAGAI <u>GAGT</u> (A G <u>A</u> 3 <u>G</u> C	AAGA GG <u>CA</u>	GGAG ACCC	<u>GTA:</u> TCT:	<u>rgaa</u> rg <u>a</u> t'	<u>→</u> 3CG	TGGGJ TCAA	ACAG	AA TO	GCAAC GAAC	JAAG	
378 448 518 588	CATO G <u>TAC</u> T <u>T</u> AC	<u>BCCT</u> IGAT	GG <u>C</u> ACA CTT	<u>CC</u> TT A <u>CA</u> G CG <u>TA</u> GATC	TATCI ATAAC AAACC	TT AC	CCTT IA <u>AG</u> AC <u>CG</u> CGAT	AAGAI <u>GAGT</u> AT <u>CT</u> I GGGA	A G <u>A</u> G <u>G</u> C A CG G TT	<u>AAGA</u> 3G <u>CA</u> AGAG 1GCG	<u>GGAG</u> ACC <u>C</u> GAGC IGTC	<u>GTA:</u> TCT: TGA: TCG(<u>rgaa</u> rg <u>a</u> t' raga: stag	<u>JCG</u> T <u>GA</u> TGA JAA	TGGG TCAA GTAG AGAT	ACAG CGCG GCTT GCTT	AA TO AG TO CC CO	GCAAC GAAC CTTTC GAGTZ	JAAG IGTG GAT	
378 448 518 588 658	CATO G <u>TAC</u> T <u>TAC</u> CACO ACAD	GCCT <u>FG</u> AT GACG GAC GAT	GG <u>C</u> ACA CTT IGT CGC	<u>CC</u> TT A <u>CA</u> G CG <u>TA</u> GATC CTTT	TATC ATAAC AAACC IGGGI	TT AC C T' C C C AC AC AC	CCTTA TA <u>AG(</u> AC <u>CG)</u> CGAT(GCC <u>A</u>	AAGAI <u>GAGT</u> (<u>ATCT</u> I <u>GGG</u> A(<u>TGG</u> G(A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u>	AAGAG GG <u>CA</u> AGAG FGCG G <u>AC</u> C	<u>GGAG</u> ACC <u>C</u> GAGC <u>IGTC</u> I <u>GT</u> G	<u>GTA:</u> TCT: TGA: <u>TCG</u> (<u>AAC</u>)	IGAA IG <u>A</u> T IAGA IAGA AGG	JCG F <u>GA</u> FGA JAA	TGGG TCAA GTAG AGAT ACCC	ACAG CGCG GCTT GCTT ATCA	5 AA T(AG T(CC C(CC G(3G A)	GCAAC GAAC CTTTC GAGTZ A <u>CC</u> AZ	JAAG IGTG IGAT IGAT	
378 448 518 588 658 728	CATO GTAT TTAC CACO ACAJ CGGO	<u>CCT</u> <u>FGAT</u> <u>GAC</u> <u>GAA</u> <u>GAT</u> <u>GGG</u>	GGC ACA CTT IGT CGC IAC	<u>CC</u> TT A <u>CA</u> G CG <u>TA</u> GATC CTTT CGCA	TATCI ATAAC AAACC FGGGJ FCTTC T <u>TGG</u> G	$\begin{array}{c} \mathbf{T} & \mathbf{A} \\ \mathbf{C} & \mathbf{T} \\ \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \\ \mathbf{A} \\ \mathbf{C} & \mathbf{A} \\ \mathbf{C} & \mathbf{C} \\ \mathbf{C} \\ \mathbf{C} & \mathbf{C} \\ \mathbf{C} \\ \mathbf{C} \\ \mathbf{C} & \mathbf{C} \\ $	CCTT/ IA <u>AG(</u> AC <u>CG/</u> CGAT(GAAG/ GAAG/	AAGAI GAGT AT <u>CT</u> I GGGA GGGA TG <u>GG</u> A	A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u> F C <u>C</u>	AAGA GG <u>CA</u> AGAG IGCG GACC	GGAG ACC <u>C</u> GAGC IGTC IGTC A <u>GC</u> G	<u>GTA:</u> TCT: TGA: <u>TCG(</u> <u>AAC</u> <u>AAC</u>	rgaa rg <u>a</u> t raga <u>stag</u> agg agg	∠ JCG IGA IGA GAA FAA ACC	TGGG TCAA GTAG AGAT ACCC CAGC	ACAG CGCG CTTC ATCA ATCA TGAA	5 AA TO AG TO CC CO CC GO GG AJ	GCAAC GAAC CTTT GAGT A <u>CC</u> AJ AGG <u>G</u>	GAAG GAT ATAG AAAA GTG	
378 448 518 588 658 728 798	CATC GTA TTAC CACC ACA CGGC G <u>AT</u>	GCCT FGAT GACG GGAA GGGG GGGG	GGC ACA CTT IGT CGC IAC AAG	<u>CC</u> TT A <u>CAG</u> CGTA GATC CTTT CGCA AGGT	TATC: ATAA(AAAC(FGGGJ FCTT(FTGG(AGCT;	TT AC CC IT CG CI CG CI CC AC CC AC CG C	CCTTI IA <u>AG(</u> AC <u>CGJ</u> CGAT(GAAG) GAAGJ ITC <u>CJ</u>	AAGA GAGT AT <u>CT</u> GGGA GGGA ACGA ACGA	A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u> F C <u>C</u> C GG	AAGA GG <u>CA</u> AGAG IGCG G <u>ACC</u> AAAA	GGAG ACCC GAGC IGTC IGTC AGCG GCCG	GTA: TCT: TGA: <u>TCG(</u> <u>AAC</u> <u>AAG(</u> CGA(IGAA IG <u>A</u> T IAGA IAGG AAGG ICTC IGAA	JCG FGA FGA JAA FAA ACC FCA	TGGG TCAA GTAG AGAT ACCC CAGC	ACAG CGCG GCTT GATCA ATCA IGAA ATCT	5 AA TO AG TO CC CO CC GO GG AJ GG AJ TT GO	GC <u>AA(</u> GAAC CTTT(GAGT) ACCAJ AGG <u>G(</u> CCGG	JAAG IGTG CGAT ATAG AAAA JGTG IGCT(
378 448 518 588 658 728 798 868	CATC GTA TTAC CACC ACA CGGC GATZ ACT	GCCT GAT GAT GAA GAA GGGG AGGC IGGA	GGC ACA CTT IGT CGC IAC AAG ICT	<u>CCTT</u> A <u>CA</u> G CGTA GATC CTTT CGCA AGGTA C <u>GGG</u>	IATC ATAAC IGGG ICTTC ITGGC AGCT IGAC		CCTT/ TAAG ACCGJ CGAT CGAT CGAT GAAG TTCCJ ATAA	AAGAI GAGT(ATCTI GGGA(TG <u>GG</u> A ACGA: AAGC(GGCG(A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u> C GG G CC	AAGA GGCA AGAG IGCG GACC AAAA CCCG AGAA	GGAG ACCC GAGC IGTC IGTC GCCG GCCG GCCG	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CGA(IGAA IG <u>A</u> T IAGA <u>IAGG</u> <u>IAGG</u> IGAA IGAA	GA GA GA GA GAA GAA CC FCA FCA	TGGG TCAA GTAG AGAT ACCC CAGC AGAG	ACAG CGCG GCTTC ATCA ATCA IGAA ATCT CGAA	5 AA T(CC C) CC G(GC A) GG A) GG A) TT G(GC T)	GCAAC GAAC CTTTC GAGT ACCA AGG <u>G</u> CCGG TGGC	JAAG IGTG CGAT ATAG AAAA JGTG IGCT ICTA	
378 448 518 588 658 728 798 868 938	CATC GTA TTA CAC CAC CAC CAC CGG GAT ACT CCG	GCCT GAT GACG GGAA GGAT GGGC AGGC AGGC	GGC ACA CTT IGT CGC IAC AAG ICT GCT	CCTT ACAG CGTA GATC CTTT CGCA AGGT AGGT	TATCT ATAAC AAACC TGGGJ TCTTC T <u>TGGC</u> T <u>TGGC</u> GTAAC		CCTTI TAAG ACCGJ CGAT GAAG TCCJ ATAA FGGC	AAGAI <u>GAGT</u> I <u>GGG</u> A(<u>TGGGG</u> <u>ACGA</u> <u>A</u> AGC(GGCG(TACA(CGTG(A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u> C GG G CC G CC	AAGA GG <u>CA</u> AGAG TGCG GACC AAAA CCCG AGAA AACG	GGAG ACCC GAGC IGTC IGTC GCCG GCCG GCCG GCCG CTTA	GTA: TCT: TGA: <u>TCG(</u> <u>AACI</u> <u>AAG(</u> CGA(CCA)	rgaa rg <u>a</u> t raga <u>stag</u> <u>sct</u> c sgaa scag agcag	GA GA GA GA GA GA FCA FCA FCA FCA CGA	TGGG TCAA GTAG GTAG AGAT ACCC CAGC CAGC	ACAGI CGCG GCTTC ATCA AATGO IGAAC ATCT CGAAC AAGGO	5 AA T(CC C) CC G(JG A) JG A) JG A) JG A) JG A) JG C) JC C)	GCAAC GAAC CTTTC GAGT ACCAA AGG <u>GC</u> CCGG TGGC TTCGC	JAAG IGTG CGAT ATAG AAAA JGTG IGCT IGCTA CCAC	
378 448 518 588 658 728 798 868 938 1008	CATC GTA: TTAC CACC ACA CGGC GATZ ACT CCGZ TGAC	GCCT GAT: GAT: GAA: GGAT GGGC AGGC IGGA ATAG GCCG	GGC ACA CTT IGT CGC IAC AAG ICT GCT TAT	<u>CC</u> TT A <u>CAG</u> <u>CGTA</u> <u>GATC</u> CTTT <u>CGCA</u> AGG <u>T</u> AGG <u>T</u> AGCT <u>GCGG</u>	IATC: ATAA(AAAC(IGGG) ICTT(IGGG) AGCT: IGAC(GGGA)		CCTTI TAAGO AC <u>CGJ</u> CGATO GCC <u>A</u> GAAGJ FTC <u>CJ</u> ATAAO FGGC C <u>GCAO</u>	AAGAI <u>GAGT</u> I <u>GGGA(</u> <u>TGGG</u> A <u>ACGA:</u> <u>A</u> AGC(GGCG(TACA(<u>CGTG(</u>	A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> G <u>TC</u> G <u>CC</u> G <u>CC</u> G <u>CG</u> C <u>GG</u>	AAGA GG <u>CA</u> AGAG G <u>GCG</u> G <u>ACC</u> AAAA CCCG AAAA CCCG AAAA ACG ACCG	GGAG ACCC GAGC IGTC IGTC GCGG GCCG GCCG GCGA CTTA IAGG	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CGA(CCA) GGG(<u>rgaa</u> rg <u>a</u> t raga <u>stag</u> <u>scag</u> scag scag scag scag	GA FGA FGA FGA FCA FCA FCA FCG CGA AGC	TGGG TCAA GTAG GTAG AGAT ACCC CAGC CAGC	ACAGI CGCG/ GCTT(ATCA(AATG(TGAA(ATCT) CGAA(AAGGA(AGGA(AA TO AG TO CC CO CC GO 3G AJ 3G AJ 3G AJ 3G AJ 3G C 3C C	GCAAC GAAC CTTTC GAGT ACCAI AGG <u>GC</u> CCGG TGGC TTCG AT <u>CC</u>	JAAG IGTG CGAT ATAG ATAG AAAA JGTG IGCT IGCTA CCAC CATC	
378 448 518 588 658 728 798 868 938 1008	CATC GTA TTAC CACC ACA CGGC GATZ ACT CCGZ TGAC	GCCT GAT GAT GAT GGA GGA GGG GGA AGGC GGC GCC Q	GGC ACA CTT IGT CGC TAC AAG ICT GCT TAT R	<u>CC</u> TT A <u>CAG</u> <u>CGTA</u> <u>GATC</u> CTTT <u>CGCA</u> AGGT AGGT AGCT <u>GCGG</u> I	IATC: <u>ATAA(</u> <u>AAAC(</u> <u>IGGGJ</u> <u>IGGGJ</u> <u>IGAC(</u> <u>GGGAJ</u> F	$\begin{array}{c} \mathbf{T} \mathbf{A} \\ \mathbf{C} \mathbf{C} \\ \mathbf{C} \mathbf{C} \\ $	CCTTA TA <u>AG(</u> AC <u>CGJ</u> C <u>GAT(</u> 3CC <u>A</u> 3 <u>AAGJ</u> TTC <u>CJ</u> ATAA(C <u>GCA(</u> A	AAGAI GAGT(ATCTI GGGQA(TGGGQ ACGAC GGCGG GGCGG CGTGG Q	A G <u>A</u> G <u>G</u> G <u>TT</u> G <u>TC</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u>	AAGAG GG <u>CA</u> AGAGG <u>IGCG</u> <u>3ACC</u> <u>3ACC</u> AAAA CCCG AGAA AACG <u>ITCT</u> E	GGAG ACCC GAGC IGTC IGTC GCCG GCCG GCCG GCCG CTTA IAGG I	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CGA(CCA) GGG(L	r <u>GAA(</u> r <u>GA</u> T) rAGA <u>3T</u> A <u>G(</u> <u>3T</u> A <u>G(</u> <u>3C</u> AG) 3CAG 3CAG <u>3GA</u> G <u>3GA</u> G	JCG JCG IGA IGA JAA JACC ICA ICG CGA AGC S	TGGG TCAA GTAG GTAG AGAT AGAG CAGC AGAG AGGA TAGT C	ACAGI CGCGI GCTTC ATCAC ATCAC IGAAC ATCT: CGAAC AAGGA AGGAA K	AA T AG T CC C CC G G G G G G G G G C T T	GCAAC GAAC CTTTC GAGT ACCAA AGGG CCGG TGGC TGGC	JAAG IGTG CGAT ATAG AAAA JGTG IGCTG ICTAG CAC CAC	
378 448 518 588 658 728 798 868 938 1008	CATC GTA TTA CAC ACA CGG GAT ACT CCG TGA CAA	<u>3CCT</u> <u>FGAT</u> <u>3ACG</u> <u>3GGG</u> <u>3GGG</u> <u>3GGG</u> <u>1GGA</u> <u>4GGC</u> <u>1GGA</u> <u>4GGC</u> <u>1GGA</u> <u>4GGC</u> <u>1</u> GGA <u>4GGC</u> <u>1</u> GGA <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GGC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4GC</u> <u>4G</u>	GGC ACA CTT IGT CGC IGT CGC IAC AAG ICT GCT R CGT	CCTT ACAG CGTA CGTA CTTT CGCA AGGT AGGT	TATCT ATAAC AAACC IGGGJ IGGGG TTGGC GGGAC F TTT		CCTT IA <u>AG(</u> AC <u>CGJ</u> C <u>GAT(</u> GCC <u>A</u> ITC <u>CJ</u> ATAA(IGGC: C <u>GCA(</u> A GCT	AAGAA GAGT(AT <u>CT</u> A GGGA(TG <u>GGG</u> ACGAA GGCG(GGCG(CGTG(CGTG(CAA	A G <u>A</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> G <u>G</u> C <u>G</u> G <u>G</u> C <u>C</u> C <u>G</u> C <u>C</u> C <u>G</u> C <u>C</u> C	AAGAQ GG <u>CA</u> AGAGQ <u>IGCG</u> 3 <u>ACC</u> 3 <u>ACC</u> 4 AAAA CCCG AGAA ACCG ITCT E GAA	GGAG ACCC GAGC IGTC IGTG GCCG GCCG GCCG GCCG CTTA IAGG I ATC	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CGA(CCA) GGG(L CTA	rgaa rg <u>a</u> t raga <u>3Tag</u> <u>3Tag</u> 3 <u>Ga</u> 3 <u>Ga</u> 3 <u>Ga</u> 5 <u>3Ga</u> 5 <u>3Ga</u> 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	<u>JCG</u> F <u>GA</u> FGA FGA FCA FCA FCA FCA CGA S TCT	TGGG TCAA GTAG AGAT ACCC CAGC CAGC CAGC	ACAG2 CGCG2 GCTTC ATCA(AATGC IGAAC ATCT: CGAAC AAGGAC K AAGGAC K AAG	AA TO AG TO CC CO GG AJ GG AJ GG AJ GG AJ GG AJ GC C GC C ACC	GCAAC GAAC CTTTC GAGT AGGG CCGG TGGC TGGC	JAAG IGTG CGAT ATAG AAAA JGTG IGCTO ICTAO CATCO	
378 448 518 588 658 728 798 868 938 1008	CAT(G <u>TA</u>) T <u>TA</u> (<u>C</u> ACA) CGG0 G <u>AT</u>) ACT CCG1 TGA(C <u>A</u> A)	<u>3CCT</u> <u>FGAT</u> <u>3GAA</u> <u>3GAT</u> <u>3GGG</u> <u>3GGG</u> <u>1GGA</u> <u>ATAG</u> <u>3CCG</u> <u>Q</u> <u>F</u> AG	GGC ACA CTT IGT CGC IAC AAG ICT GCT IAT R CGT	CCTT ACAG CGTA GATC CTTT CGCA AGGT CGGG AGCT GCGG I I ATA	TATCT ATAAC AAACC TGGGJ TGGG TTGGC TGGC GGGA F TTT	TT AC CC II CC AC AC AC CC ACC	CCTT/ TA <u>AG(</u> AC <u>CGJ</u> CGAT(GCC <u>A</u> ; GAAGJ FCC <u>A</u> TCC <u>GCA</u> (A GCT	AAGAI GAGTO ATCTI GGGAO TGGGQ ACGAI AAGCO GGCGO TACAO CGTGO Q CAA	A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> <u>G</u> <u>T</u> <u>C</u> <u>G</u> <u>G</u> <u>G</u> <u>G</u> <u>C</u> <u>G</u> <u>G</u> <u>C</u> <u>G</u> <u>C</u> <u>G</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u>	AAGAG GGCA AGAGG IGCG GAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA CCCGG AAAA ACCGG AAAAA CCCGG AAAAA AAAAA CCCGG AAAAA CCCGG AAAAA CCCGG AAAAA CCCGG AAAAA AAAAA CCCGG AAAAA AAAAA CCCGG AAAAA AAAAA AAAAA CCCGG AAAAA AAAA AAAAA AAAAA AAAAA AAAAA AAAA	GGAG ACCC GAGC IGTC IGTG GCCG GCCG GCCG GCCG CTTA I ATC	GTA: TCT: TGA: <u>TCG(</u> <u>AAC</u> AAG(CGA(CGA(CCA) <u>GGG(</u> L CTA	r <u>Gaa</u> IG <u>a</u> T IAGA <u>3T</u> A <u>GG</u> <u>3C</u> AG 3CAG 3CAG 3CAG 3GAG F TTT	GA FGA FGA FGA FGA FCA FCA FCA FCA FCA FCA FCA FCA FCT	TGGG TCAA GTAG GTAG AGAT AGAG CAGC AGAG AGGA TAGT C TGC	ACAGI CGCGI GCTTCA ATCA TGAAC TGAAC TGAAC AATCT CGAAC AAGGAC K AAGGAC K AAG	AA T AG T CC C CC G G G G G G G G C T G C C C C C C C C C	GCAAC GAAC CTTTC GAGTI ACCAI AGG <u>GC</u> CCGG TGGC TGGC TGGC R R CGT	GAAG TGTG CGAT ATAG AAAA GGTG TGCTA CCAC CATC TCG T TCG T	
378 448 518 588 658 728 798 868 938 1008	CAT(G <u>TA</u>) T <u>TA</u> (CAC) ACA) CGG0 G <u>AT)</u> ACT CCGJ TGA(C <u>A</u> A)	<u>BCCT</u> <u>FGAT</u> <u>BACG</u> <u>GGA</u> <u>AGGC</u> <u>AGGC</u> <u>TGGA</u> <u>AGGC</u> <u>C</u> <u>Q</u> <u>F</u> AG	GGC ACA CTT TGT CGC CGC TAC GCT TAC CGT R CGT	CCTT ACAG CGTA GATC CTTT CGCA AGGT AGGT	TATCT ATAAC AAACC TGGGJ TCTTC TGGC TGAC GGGA F TTT	TT AC CC T CC C CC AC CC AC CC AC CC AC CC AC CC T CC T CC T CC T CC AC CC AC C	CCTTI TA <u>AG(</u> ACC <u>GJ</u> CGAT(3GCA(3GCA(ATAA(FGGC: C <u>GCA(</u> A GCT	AAGAI GAGT ATCTI GGGA GGCGA ACGAI AAGCC GGCGG TACAC CGTG Q CAA	A G <u>A</u> G <u>G</u> C A CG G <u>TT</u> <u>G</u> G <u>CC</u> G <u>CC</u> G <u>CC</u> G <u>CC</u> C <u>GG</u> Y TAT	AAGAG GGCA AGAG IGCG 3 <u>ACC</u> AAAA CCCG AGAA AACG ITCT E GAA	GGAG ACCC GAGC IGTC IGTG GCCG GCCG GCCG GCCG CTTA IAGG I ATC	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CGA(CCA) GGG(L CTA	<u>rgaa</u> r <u>ga</u> t raga <u>3Taga</u> <u>3CTC</u> <u>3CTC</u> 3GAA 3CAG 3GAG 5GAG	GA FGA FGA FGA FGA FCA FCA FCA FCA FCA FCA FCA FCA FCA FC	TGGG TCAA GTAG GTAG AGAT AGCC CAGC AGAG CGGT AGGA TAGT TGC	ACAGI CGCGI GCTT ATCA AATGO IGAAO AATGT CGAAO AAGGAO K AAGGAO K AAG	AA TO AG TO CC CC CC GG 3G AJ 3G AJ 3G AJ 3G AJ 3G C 3C CJ ACC	GCAAC GAAC CTTTC GAGT ACCA AGGG CCGG TGGC TGGC	JAAG TGTG GAT ATAG AAAA JGTG TCTA CATC CATC TCG TCG T L	
378 448 518 588 658 728 798 868 938 1008	CATC GTA: TTAS CACS CGGC GATZ CCGS TGA CCAS;	GCCT <u>GAT</u> <u>GAT</u> <u>GGAA</u> <u>GGGC</u> <u>GGA</u> <u>AGGC</u> <u>Q</u> <u>Q</u> <u>C</u> <u>G</u>	GGC ACA CTT IGT CGC IAC GCT IAC GCT R CGT R CGT	CCTTT ACAG CGTAG GATC CTTTT CGCA AGGT CGCGG AGCTC GCGGG I ATA	IATC: ATAAC AAACC IGGGI IGCTC IGAC GGGAI F TTT	TT A(CC T CG CA CG AC AC AC CG C CG A CG C CG A CG T G GGA	CCTTI IAAGO CGAC CGAT GGCA GCCA IGGCC A GCT	AAGAI GAGT(ATCT) GGGA(IGGGG ACGA: AAGCC GGCGC CGTGC CAA	A <u>GA</u> A <u>G</u> A <u>CG</u> 3 <u>TT</u> <u>TT</u> <u>T</u> <u>C</u> <u>G</u> <u>G</u> <u>G</u> <u>G</u> <u>C</u> <u>C</u> <u>G</u> <u>G</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u> <u>C</u>	AAGAG GG <u>CA</u> AGAGG <u>TGCCG</u> <u>3ACC</u> CCCGG AGAA CCCCG AGAA CCCCG AGAA CCCG AGAA	GGAG ACCC GAGC IGTC GGCG GCCG GCCG GCCG CTTA I ATC	GTA: TCT: TGA: TCG(AAC) AAG(CGA(CCA) GGG(L CTA	r <u>GAA(</u> IAGA <u>JTAGA</u> <u>JTAGG</u> <u>JTAGG</u> GGAG GGAG F TTT	GCG : GA : IGA	TGGGJ TCAA GTAGC AGATJ ACCCJ CAGC CAGC CAGCA TAGT TAGT TGC	ACAG2 CGCG2 GCTCA AATGA IGAAC ATCTC CGAAC AAGGAA K AAGGAA K AAGGAA	5 AA TC CC CC CC GC AI GG AI GG AI GG C TT ACC	GCAAC GAAC CTTTC GAGTI AGGGG CCGGC TGGC TTCGC ATCC CGT	JAAG GGAT ATAG GGAT GGTG IGCT ICTA CCAC CATC CATC ICTA ICCAC	
378 448 518 588 658 728 798 868 938 1008 1078	CATC GTA: TTAQ CACQ CGGC GATT CCGJ TGA(CAA; K AAG	GCCT GGAT JACGG GGAA GGGG GGGA GGGA GGGA GGGA	H GGC ACA CTT IGT IGC ICC GCT ICC GCT R CGT R CGT K AAA	CCTT ACAG CGTA CGTA CTTT CGCA AGGT CGGG AGCT GCGG I ATA L L CTC	IATC: ATAAC AAACO IGGGJ ITGGC ITGGC ITGAC GGGAJ F TTT TTT Q CAG	TT A(CC T CC A CC A A CC A A CC A A C C A C C A C C C A C C C C C A C C C C C C C C C C C C C	CCTTJ IA <u>AG</u> CGAT(GGAT(GGCA) ATAA(IGGC: CGCA(A GCT L TTG	AAGAJ GAGT(ATCT) GGGQ GGGQ GGCGG GGCGG GGCGG CAA L L CTT	A G <u>A</u> A CG A CG 3 <u>TT</u> <u>5</u> <u>TC</u> <u>5</u> <u>CC</u> <u>5</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u> <u>CC</u>	AAGA; GG <u>CA</u> AGAGG <u>IGCG</u> <u>IGCG</u> AAAA ACCG ITCT E GAA AACG AGAA	GAG ACCC GAGC IGTC IGTG AGCG AGCG GCGA CTTA IAGG I ATC K AAG	GTA: TGA: TGA: CGA: CGA: CGA: CGA: CCA: L CTA I ATC	IGAA IGAT IAGA JTAG JTAG JCTC JCAG JCAG GAA JCAG F TTT L L CTT	$\begin{array}{c} \\ \hline \\ GCG \\ \hline \\ GA \\ \hline \\ GA \\ \hline \\ GCA \\ \hline \\ GCA \\ \hline \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ $	TGGGI TCAAG GTAGG GTAGC CAGCC CAGCC CAGCC CAGCC CAGCC CGGTC CGGTC C TGC L TTGC	ACAGE CGCGC GCTTC ATCA(AATGC CGAA(AAGGA K AAGGA K AAG T ACC	5 AA T AG T CC C CC C GG AJ GG AJ GG AJ GG AJ GG C T ACC L CTT	GCAAC GAAC: CTTTC GAGTA AGGGC CCGGC TGGCC TGGCC R CGT R CGT S TCC	JAAG IGTG CGAT ATAG AAAAA GGTG IGCT ICTAA CCAC CATC T L STOO TGA	
378 448 518 588 658 728 728 738 868 938 1008 1078	CATC GTAT CACQ CGGC GATZ CCGJ TGAC CAA; K AAG	GCCT GAT GGAT GGA GGGG GGGG GGGA GGGA GG	GG <u>C</u> ACA CTT IGT IGT CGC IAC CGC IAC CGT R CGT K AAA	CCTT ACAG CGTA GATC: CTTT CGCA AGGT CGGG AGCT GCGGG I I ATA L L CTC	IATC' ATAAC TGGGJ FCTTU T <u>TGGG</u> AGCT <u>J</u> TGAC 3TAAC GGGA F TTT Q CAG	TT AC CC I CG C CG AC AC AC CG AC CG AC CC T GGA GGA R AGA	CCTTJ IA <u>AG</u> C <u>GAT</u> <u>GAT</u> <u>GCT</u> <u>GCT</u> A TAAO FGGC: A GCT L TTG	AAGAJ GAGT ATCTJ GGGQ GGCGG GGCGG GGCGG GGCGG CAA Q CAA L CTT	A GA GC GC GC GC GC GC GC GC GC GC C GC	AAGA; GG <u>CA</u> AGAG <u>IGCG</u> <u>JACC</u> ; J <u>ACC</u> ; CCCG AGAA AACG GAA N AAT	GGAG ACCC GAGC IGTC IGTG GCCG GCCG GCCG GCCG I ATC I ATC K AAG	GTA: TCT TGA: TCGA CGAC CGAC CCAA GGG L CTA I ATC	IGAA IGAT IAGA JTAG JTAG JGAA JCTC JGAA JCCG JGAG JGAG JCCG TTT L CTT	ACC GGA GA GA GGA CGA S TCT A GCG	TGGGJ TCAAG GTAGG GTAGG AGAT, ACCCC CAGC AGGA CCGGT CAGC TAGC TGC L TTG	ACAGE CGCGC GCTTC ATCAN TCAN AATGC TGAAN AAGGAN K AAGGAN K AAGG T ACC	5 AAA T CCC C CCC C CCC C AGG AJ AGG AJ AGG AJ AGG C T ACC L CTT	GCAAC GAACC CTTTC GAGT AGGG CGG CGG CGT R CGT S TCC	JAAG TGTG CGAT ATAG GGTG TCTAA CACC CATC TCG T L STO TGA	
378 448 518 588 658 728 738 938 1008 1078	CATC GTA TTA CAC CGGC GAT CCG TGA CCA CAA K AAG	GCCT GGAT GGAA GGGG GGGG GGGG GGGA GGGA	GG <u>C</u> ACA CTT TGGT CGC CGC CGC CGC CGT R CGT R CGT K AAAA	CCTT ACAG CGTA CGTA CTTT CGCA AGGT CGCG I AGCT CGCGG I ATA L L CTC	IATC' ATAAC TGGGJ TCTTC TGGG AGCT TGAC GGGA F TTT Q CAG	TT A(CC II CG C CA CG A CG A CG A CG A CG A CG A C	CCTTI IA <u>AG</u> C <u>GAT</u> GGCA <u>T</u> GGCA <u>T</u> ITCC <u>2</u> ATAA(IGGC C <u>GCA</u> A GCT L TTG	AAGAJ <u>GAGT</u> <u>GGG</u> <u>GGGG</u> <u>ACGA</u> <u>A</u> AGCC GGCGC GGCGC GGCGC CAA CAA L CTT	A GA G GC G GC G TT C CG G GG G CC C GG C GG C GG T T TAT	AAGA GG <u>CA</u> AGAG GG <u>CA</u> AAAA CCCCG AAAAA CCCCG AAAAA CCCCG AAAAA CCCCG AAAAA CCCCG AAAAA CCCCG AAAAA AACG GAA	GGAG ACC <u>C</u> GAGC I <u>GTC</u> GGCG GCCG GCCG CTTA I ATC K AAG	GTA: TCT TGA TCG CGA CGA CGA CCAA GGG L CTA I ATC	rgaaa rg <u>a</u> t itaga <u>itaga</u> <u>itag</u> <u>itag</u> itag itaga it	ACC GCG ACC CGA GCG ACC CGA S TCT A GCG	TGGGJ TCAAG GTAGG GTAGG AGAT, ACCCC CAGC AGGA CCGGT CAGC TAGC TGC L TTG	ACAG2 CGCG2 GCTTC ATCAC AATGC CGAAC AAGGA K AAGG K AAG T ACC	5 AA TC AG TCCC CC 2C Gd AJ 3G AJ 3G AJ 3G C T ACC L CTT	GCAAC GAAC CTTTT GAGTJ ACCAJ AGGG CCGGC TTCGC R CGT R CGT S TCC	GAAG TGTGG CGAT AAAAA GGTG TGCTA TCCAC TCCAC TCCAC TCCAC TCCG TCCG	
378 448 518 588 658 728 738 938 1008 1078	CATC GTA TTA CAC CGGC GAT CCG CAA CCAA K AAG	GCCT GAT GAT GGAA AGAT GGGA AGGC GGA GGA GGA	GGC ACA CTT TGT CGC TAC GCT TAC GCT TAT R CGT K AAA	CCTT ACAG CGTA GATC: CTTT CGCA AGGT CGGG AGCT AGCT AGCT ACCGG L CTC	IATC: ATAAC <u>AAACC</u> IGGG ICTTI TIGGG AGCTI GACC GGGA F TTT CAG CAG	TT AG <u>CC</u> <u>T</u> <u>CG</u> <u>C</u> <u>CG</u> <u>CG</u> <u>C</u> <u>CG</u> <u>CG</u> <u>C</u> <u>C</u> <u>CG</u> <u>CG</u> <u>C</u>	CCTTI IA <u>AG</u> C <u>GAT</u> GCC <u>A</u> GCC <u>A</u> TTC <u>C</u> A TAAA GCT L TTG	AAGAJ GAGT ATCTJ GGGA TGGGG GGGGG GGCGG GGCGG GGCGG CAA L CTT	A GA GC GC GC GC GC GC GC GC GC GC	AAGA GG <u>CA</u> AGAG GGAG GAA AACG GAA N AAT	GGAG ACCC GAGC IGTC GGCG GCCG GCCG GCCG CTTA I ATC K AAG	GTA: TCT TGA: TCGA CGA CGA CGA CCA GGG CCA L CTA I ATC	rgaa rgat Taga <u>Taga</u> <u>GCTC</u> GGAA GCAG GCAG GCAG F TTT L CTT	ACC GCG TGA TGA TGA TAA ACC TCA TCG CGA S TCT A GCG	TGGG TCAA GTAGG AGAT ACCCC CAGCC CGGT CGGT	ACAG2 CGCG2 GCTTC ATCAC AATGC TGAAC ATCT CGAAC AAGGAC K AAGG K AAG T ACC	5 AA T AG T CC C C C C C G G G G C C T ACC T C T C T C T C T	GCAAC GAAC CTTTC GAGT/ ACCAJ AGCGG CCCGG TTCGC R CGT R CGT S TCC	JAAG TGTGG CGAT AAAAA JGTG TGCTA CCAC CATC CATC CATC CATC TGA	
378 448 518 588 658 728 738 738 938 1008 1078	CATT GTA TTA CACA CGGG ACA CCGG TGA CAA CAA K AAG	GGGA	GGC CTT CGC CGC CGT CGC CGT R CGT R CGT R CGT R AAAA CCCG	CCTTT ACAGI CGTAI CTTTT CGGA AGGTI AGGTI AGGTI I ATA L L CTCC	IATC: ATAAC AAAC IGGG ICTTI TIGGC AGCT GAC GGGA F TTT CAG CAG AGGA:	TT AG CC II CC	CCTTI IA <u>AG(</u> CGAT(GCC <u>A</u> GCC <u>A</u> ATAA(C <u>GC</u> A GCT L TTG GAGG	AAGAJ GAGT ATCTJ GGGA GGGGG GGGGG GGCGG GGCGG GGCGG GGCGG CAA L CTT L CTT	A GA G G G G G C G G G C G G G C G G G C G G G C C G G G C G G G C C G G G C C C G G G G	AAGAG GG <u>CA</u> AGAGG <u>TGCG</u> <u>JACC</u> <u>AAAA</u> CCCCG AAACG TTCT E GAA N AAT	QGAQ ACCC GAGC IGTC IGTG GCG GCGG GCGG CTTA I ATC K AAG	GTAT TGAT TGAT TGGC CGAA CGAA CCAA GGGC L CTA I ATC	rgaaa rgat Taga <u>3Tag</u> <u>3Tag</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u> 3 <u>G</u>	ACC GCG TGA TGA TGA TCA S TCT A GCG A GCG TCA	TGGG TCAA GTAGG AGAT ACCCC CAGCC CGGT CGGT	ACAGI CGCGJ GCTTC ATCAC TGAAC ATCTC CGAAC AAGGA K AAGG T ACC AGTTC	5 AA TC AG TC CC CC CC G AA AG AA TT G GC CC ACC L CTT CC G G	GCAAA GAAC CTTTO GAGTJ ACCAA AGGGC TTCGG R CGGT R CGT S TCC GCTT	JAAG TGTG4 CGAT AAAAA JGTG IGCT TCTA CATC CATC CATC TCG T L STO TGA TCCA	

1265 TCTTCCTCCC CCCCCTCAC TCCCCCTTTT TC

(Fig. 3). The main features classifying this intron into subgroup IIB include the position of the bulging A at -8, a 4-nt C1(i)–C1(ii) internal loop with consensus sequence AARC, the intron 3' end with the consensus sequence RAY, and the conserved nucleotides circled in Fig. 3 (Michel et al. 1989).

Two rather close relatives of the potato rps10 intron were found by sequence comparisons in the liverwort mitochondrial genome (Oda et al. 1992). One of them is the intron in the rrn26 gene (Fig. 1B) and the other is intron 2 in the cox3 gene (data not shown). The fact that these introns are so similar in primary sequence implies a common evolutionary origin and suggests that they may have been mobile from gene to gene in a common ancestor of bryophytes and angiosperms. Horizontal spreading of group-II introns has been postulated in fungi (Hardy and Clark-Walker 1991), *Marchantia* (Oda et al. 1992; Ohta et al. 1993), and *Oenothera* (Lippok et al. 1994). Most of the other introns of *Marchantia* show no homology in sequence or integration site with higher-plant mitochondrial introns, suggesting different origins for these introns (Oda et al. 1992; Ohta et al. 1993).

```
Α
```







Fig. 2A-C Identification of splice sites between rps10 exons. A DNA sequence of the 3' end of exon 1 and the 5' end of the rps10 intron. B sequence of a spliced cDNA. C DNA sequence of the 3' intron end and the 5' end of exon 2. Exon and intron sequences are given in uppercase and lowercase letters, respectively. Horizontal arrows indicate exon-intron junctions in genomic sequences and the exon junction in the spliced cDNA. Vertical arrows indicate exon sequences

The rps10 stop codon and a putative initiation codon are created by editing

Comparison of cDNA and genomic sequences revealed RNA editing events in both exons. Six independently derived spliced cDNAs were completely sequenced. All had in common four C-to-U modifications as compared to the genomic sequence (Fig. 1B), with two editing sites in each exon. No partially edited sequences were observed, suggesting that these editing sites are altered efficiently in the mRNA. The first editing event in exon 1 changes a threonine ACG codon to an AUG methionine codon 20 amino acids downstream from a genomic-encoded in-frame AUG codon. The amino-acid sequence between the two potential initiation codons would extend the N-terminus of potato polypeptide in comparison to bacterial, chloroplast, and liverwort and pea mitochondrial RPS10 sequences, while the second AUG codon created by RNA editing coincides wih the liverwort and pea initiation codons (Fig. 4 and Knoop et al. 1995). These results suggest that translation initiates at the AUG codon created by editing in potato. C-to-U changes have been shown to produce an initiation codon in chloroplast rpl2 mRNA of maize (Hoch et al. 1991) and *psbL* transcripts of tobacco (Kudla et al. 1992), as well as in the mitochondrial nad1 mRNAs of wheat and probably also Oenothera (Chapdelaine and Bonen 1991; Wissinger et al. 1991). Isolation and direct sequencing of the N-terminus of potato RPS10 will answer this question. If the second methionine residue corresponds

to the first translated codon, this RNA editing event can be assumed essential for protein synthesis to proceed.

mRNA editing in plant mitochondria can also create stop codons from arg or gln triplets (Grav et al. 1992). The last editing event in exon 2 modifies a CGA arginine triplet to a UGA termination codon, shortening the deduced polypeptide by ten amino acids. After editing the predicted C-terminus is more similar to that of the liverwort mt RPS10. However, both plant-deduced amino-acid sequences are extended at the carboxy-terminal end in comparison with bacterial and algal chloroplast RPS10 polypeptide sequences (Fig. 4). A potential specific effect of these carboxy-terminal extensions in the function of the plant RPS10 remains to be established. The two other editing events (amino acids 99 and 120) also improve similarity to the liverwort polypeptide but not to the bacterial and algal chloroplast proteins (editing at position 99 actually decreases this similarity). The highest similarity between the predicted amino-acid sequences encoded by the rps10 genes of potato and other species is found with liverwort. After editing, potato and liverwort polypeptides show 60% identity in a 100 amino-acid overlap. The amino-acid sequence deduced for potato RPS10 corresponds to a basic polypeptide of 129 amino acids, with a predicted molecular weight of 15798 Da.

Unspliced rps10 transcripts are edited

The temporal relationship between RNA editing and intron splicing was examined by sequence analysis of cDNA clones of unspliced rps10 transcripts. Unspliced cDNAs were obtained by reverse transcription of total mtRNA with oligo 3 and PCR-amplification of exon-1 and intron sequences between oligos 1 and 3 (Fig. 1B). A RT-PCR product size with a length of 0.4–0.5 kb was as expected from the genomic sequence. Of the nine clones completely sequenced (Fig. 5), seven showed three C-to-U modifications with respect to the genomic sequence. These include Fig. 3 Secondary structural model of the rps10 intron. The secondary structure model of six domains (I to VI) is characteristic for group-II introns (Michel et al. 1989). The RNA editing event in domain I(i) is indicated by an arrow (C-to-U). The position of the bulging A to which the 5' end of the intron is expected to join during splicing is indicated by an asterisk. EBS and IBS are candidates for exon- and intron-binding sites, respectively. The γ tertiary interaction in the central hub is indicated. Arrows point to exon-intron borders



		1					
potato (mt) 1	MRQRRALRRV	SQKERPPKVM	TTKIGIVIRS	FDHPFLENHF	WGLPPYTRK-		
liverwort (mt)		M	:A::C:::K:	:::QR	S::LLN:::-		
Cryptomonas (c	cp)	MQ	KQ::R:RLKG	YETNL:LDSC	NKIIETATRT		
Cyanophora		MASNQ	QL::R:QL::	Y:SSL:::SC	EQIIEAAKRT		
E.coli		MQ	NQR:R:RLKA	:::RLIDQAT	AEIVETAKRT		
Mycoplasma		MA	EN:MR:KLKG	Y::AIVDQSI	TKIIQAAEGT		
			*	•••••	• • •		
potato 50	IGL	PESRVLYTVL	RSPHIDKKSR	EQFFMKIKKE	FLVIKTERHE		
liverwort	:::	: KKQT : : : : :		:::E:R:H:Q	L:::E::T:K		
Cryptomonas	NSVTAGP:P:	:TR:KI:C::	::::VN:D::	:H:ETR:H:R	VID:YEPNST		
Cyanophora	DATAVGP:P:	:TKKKI:C::	::::V::D::	:H:EIRVHRR	IIDLYLPSSK		
E.coli	GAQVRGP:P:	: TRKERF : : :	I:::VN:DA:	D:YEIRTHLR	LVD : VEPTEK		
Mycoplasma	GAKVRGP:P:	:TEKQVI:I:	:AV:KY:D::	:::E:RTH:R	L:E:LNPTAA		
	* *	**	* * .*		• •		
	ъ		e	ъ			
	R I DEVERSION ED	ODTHONOWHT			110 120)		
potato 93	LRKKFFWLKR	QRIFGAQIEI	LFSCKTRLDK	GRLQRLL~ (.	110 OF 129)		
liverwort	::E:LNW::L	HDLL:V:VK:	T:XXÖ::P::	VCKS* (.	102)		
Cryptomonas	TIDSLMQ:DL	PAGIDIDIK:	*	(102)			
Cyanophora	TIDTLTR:DL	PAGVDVEVKL	*	(:	105)		
E.coli	TVDALMR:DL	AAGVDV:ISL	G*	()	103)		
Mycoplasma	TMDVLKRVQL	PSGVDIEIKL	*	(:	102)		

Fig. 4 Comparison of RPS10 amino-acid sequences deduced from potato cDNA and genomic nucleotide sequences. Amino-acid sequence comparison between RPS10 polypeptides specified in potato mitochondria and by genomic sequences from liverwort mitochondria (Takemura et al. 1992). Cryptomonas phi chloroplast (EMBL SWISS-PROT Database Accession P19460), Cyanophora paradoxa cyanelle (Accession P17009), Escherichia coli (Accession P02364) and Mycoplasma capricolum (Accession P10129). Only amino ac-

ids altered by RNA editing are given in the top lane for the potato genomic-encoded sequence. Gaps, indicated by *dashes*, were introduced to maximize the alignment. A *colon* (:) indicates identity with the respective potato residue. *Asterisks* below the amino acids indicate positions conserved in all species compared, and *dots* indicate positions conserved in 5 out of 6 proteins or involving only conservative amino-acid changes. *Numbering* refers to the potato protein and total numbers of amino acids are given in *parentheses* 570



Fig. 5 RNA editing sites vary between individual unspliced rps10 cDNA clones. Shown on the first line is a representation of rps10 first exon (*open bar*) and intron (*thin line*). Line drawings below it represent the sequenced region of unspliced cDNAs. Edited or unedited nucleotides at the editing sites (nucleotides 59, 295 and 316) are indicated by T or C, respectively. Numbering is in relation to the rps10 genomic ATG codon at position 1

the two previously identified editing sites in exon 1 and an additional editing event in the intron sequence.

A low level of partially edited unspliced precursors was found. Two out of the nine cDNA clones were not fully edited, one clone lacking the first editing event in exon 1 and the other being edited only at this position. The presence of differentially edited cDNA clones confirms that the individual cDNA clones analyzed are derived from separate cloning events of the mRNA population. Fully edited atp9 mRNAs and low levels of partially edited cob mRNAs were previously observed in potato mitochondia (Dell'Orto et al. 1993; Zanlungo et al. 1993). Differentially edited mRNAs have also been observed in other plant mitochondrial genes and unselected translation of all these transcripts would direct the synthesis of variant polypeptides. However, partially edited transcripts appear to be, at least in some cases, intermediates in editing, since processed transcripts and polysomal mRNAs are more heavily edited than precursor or pooled total mtRNAs (Gualberto et al. 1991; Sutton et al. 1991; Yang et al. 1991). Consistent with this hypothesis, spliced rps10 transcripts are found fully edited while a low level of partially edited rps10 transcipts is detected in the unspliced fraction. On the other hand, RNA editing presumably precedes splicing in the maturation of rps10 transcripts, since most unspliced cDNA clones are already completely edited. Recent work has shown that chloroplast editing is also an early processing event (Freyer et al. 1993; Ruf et al. 1994).

The intron RNA editing event eliminates an A/C mismatch of the genomic sequence in stem I(i) of domain I, and allows an additional A-U base pairing presumably essential for stable folding of the stem structure (Fig. 3). This stem is usually well conserved in its secondary structure and thus presumably required for correct intron excision. This editing event may thus be necessary for splicing of the intervening sequence, an hypothesis further substantiated by the fact that it is altered with high efficiency, i.e. in 8 out of 9 cDNA clones. However, editing at the intron site detected here cannot be the sole rate-limiting step of splicing, since most of the immature molecules are already edited at this site.

Although editing takes place predominantly within the coding regions, editing sites have also been observed inside of intron sequences, in nad1 introns of Oenothera and petunia and in nad2 and nad5 introns of Oenothera (Knoop et al. 1991; Sutton et al. 1991; Wissinger et al. 1991; Binder et al. 1992). Some of these editing events improve base pairing in secondary structure models and may thus be required for correct processing of these introns. The editing event observed in the rps10 intron stem I(i) of domain I is similar to an editing event observed in the Oenothera nad2 intron 2 (Binder et al. 1992). A possible role of RNA editing in regulation splicing of rps10 transcripts needs to be investigated. Further editing within the intron may be required for splicing. A good candidate for such an editing site is found in the helical stem of domain IV which contains an A/C mismatch that could be eliminated by a C-to-U change, thus correcting the base pairing and stabilizing the structure (Fig. 3).

Transcription of the rps10 gene

Northern analysis previously showed that potato rps10 is expressed as multiple transcripts, ranging in size from 0.8 to 3.3 kb (Zanlungo et al. 1994). Although rps10 could be potentially co-transcribed with the upstream cob pseudogene, its complex set of transcripts differs from the hybridization pattern obtained with a cob probe. Intron and exon probes (probes 4 and 5 respectively, Fig. 1B) reveal similar transcript patterns of mRNAs larger than 1400 nt which thus probably represent unspliced precursors (Fig. 6). The relative abundance of these mRNAs is similar except for a 1.7 - 1.8 kb transcript detected by the intron probe that may represent a splicing intermediate (i.e. a intron-exon-2 lariat) or an anomalous migrating free intron. mRNA species with sizes of about 800–1000 nt, sufficient to encode the entire rps10 ORF, are specifically detected by the exon probe and thus might represent mature spliced rps10

Fig. 6 Northern analysis of transcripts from the *rps10* locus in potato mitochondria. Blots were hybridized with *A* an exon probe (oligo 4, Fig. 1 B), and *B* an intron probe (oligo 5, Fig. 1 B). Size markers are RNA length standards (BRL RNA ladders) co-fractionated in the agarose-formaldehyde gel



mRNAs. Accumulation of multiple unspliced intermediates that probably include immature molecules with different 5'- and 3'-extensions suggests that the splicing reaction may proceed more slowly than the other processing steps (i.e. the processing reactions at the 5' and 3' ends). On the other hand, different 5' ends may be generated by multiple promoters between the *cob* pseudogene and *rps10*. It will be necessary to analyse RNA transcription in greater detail to discover the nature and origin of these multiple mRNAs.

Potato *rps10* is likely to be a functional gene since its transcripts have been detected and shown to undergo processing in the form of editing and splicing. Evidence for the synthesis of ribosomal proteins in plant mitochondria includes the observation of Hunt and Newton (1991) that the NCS-3 phenotype in maize results from disruption of the *rps3/rpl16* locus, and the reaction of antibodies against RPS1A with proteins in wheat mitochondrial ribosomal fractions (González et al. 1993).

Acknowledgements The authors are greatly indebted to Dr. Simon Litvak for his continued support and encouragement. We thank Laura Tarragó-Litvak, Alejandro Araya and Alejandro Venegas for discussions and encouragement, and Dr. Jorge Garrido for improving the manuscript. We also thank V. Knoop and A. Brennicke (IGF, Berlin) for sharing data prior to publication, and F. Michel (Centre de Génétique Moléculaire du CNRS) for his help in analyzing the secondary structure of the intron. This work was supported by research grants 93/0584 from Fondecyt-Chile and Cl1*-CT93-0058 from the European Economic Community, and by The Ministère des Affaires Etrangères of France.

References

- Binder S, Marchfelder A, Brennicke A, Wissinger B (1992) RNA editing in trans-splicing intron sequences of *nad2* mRNAs in *Oenothera* mitochondria. J Biol Chem 267:7615–7623
- Bonen L (1991) The mitochondrial genome: so simple yet so complex. Curr Opin Genet Dev 1:515-522
- Chapdelaine Y, Bonen L (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. Cell 65:465–472
- Covello PS, Gray MW (1992) Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. EMBO J 11: 3815–3820
- Dell'Orto P, Moenne A, Graves PV, Jordana X (1993) The potato mitochondrial ATP synthase subunit 9: gene structure, RNA editing and partial protein sequence. Plant Sci 88:45–53
- Freyer R, Hoch B, Neckermann K, Maier RM, Kössel H (1993) RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocistronic mRNAs. Plant J 4:621– 629
- González DH, Bonnard G, Grienenberger J-M (1993) A gene involved in the biogenesis of cytochromes is co-transcribed with a ribosomal protein gene in wheat mitochondria. Curr Genet 24: 248–255
- Gray M, Hanic-Joyce PJ, Covello PS (1992) Transcription, processing and editing in plant mitochondria. Annu Rev Plant Physiol Plant Mol Biol 43:145–175
- Grohmann L, Brennicke A, Schuster W (1992) The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. Nucleic Acids Res 20:5641–5646
- Grohmann L, Brennicke A, Schuster W (1993) Genes for mitochondrial ribosomal proteins in plants. In: Nierhaus KH, Franceschi F, Subramanian AR, Erdmann VA, Wittmann-Liebold B (eds) The translational apparatus, structure, function, regulation, evolution. Plenum Press, New York London, p 599–607

- Gualberto JM, Bonnard G, Lamattina L, Grienenberger J-M (1991) Expression of the wheat mitochondrial *nad3-rps12* transcription unit: correlation between editing and mRNA maturation. Plant Cell 3:1109–1120
- Hardy C, Clark-Walker GD (1991) Nucleotide sequence of the *cox1* gene in *Kluyveromyces lactis* mitochondrial DNA: evidence for recent horizontal transfer of a group-II intron. Curr Genet 20: 99–114
- Hoch B, Maier RM, Appel K, Igloi GL, Kössel H (1991) Editing of a chloroplast mRNA by creation of an initiation codon. Nature 353:178–180
- Hunt MD, Newton KJ (1991) The NCS3 mutation: genetic evidence for the expression of ribosomal protein genes in Zea mays mitochondria. EMBO J 10:1045–1052
- Kemmerer EC, Kao T-H, Deng G-R, Wu R (1989) Isolation and nucleotide sequence of the pea cytochrome oxidase subunit I gene. Plant Mol Biol 13:121–124
- Knoop V, Schuster W, Wissinger B, Brennicke A (1991) Trans splicing integrates an exon of 22 nucleotides into the nad5 mRNA in higher plant mitochondria. EMBO J 10:3483–3493
- Knoop V, Ehrhardt T, Lättig K, Brennicke A (1995) The gene for ribosomal protein S10 is present in mitochondria of pea and potato but absent from those of *Arabidopsis* and *Oenothera*. Curr Genet 27:559–564
- Kudla J, Igloi GL, Metzlaff M, Hagemann R, Kössel H (1992) RNA editing in tobacco chloroplasts leads to the formation of a translatable *psbL* mRNA by a C to U substitution within the initiation codon. EMBO J 11:1099–1103
- Levings CS, Brown GG (1989) Molecular biology of plant mitochondria. Cell 56:171–179
- Lippok B, Brennicke A, Wissinger B (1994) Differential RNA editing in closely related introns in *Oenothera* mitochondria. Mol Gen Genet 243:39–46
- Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group-II catalytic introns – a review. Gene 82:5–30
- Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering-plant evolution. Cell 66:473–481
- Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Biol 223:1–7
- Ohta E, Oda K, Yamato K, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K, Michel F (1993) Group-I introns in the liverwort mitochondrial genome: the gene coding for subunit I of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res 21:1297-1305
- Ruf S, Zeltz P, Kössel H (1994) Complete RNA editing of unspliced and dicistronic transcripts of the intron-containing reading frame IRF170 from maize chloroplasts. Proc Natl Acad Sci USA 91:2295–2299
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Sutton CA, Conklin PL, Pruitt KD, Hanson MR (1991) Editing of pre-mRNAs can occur before *cis*- and *trans*-splicing in *Petunia* mitochondria. Mol Cell Biol 11:4274–4277
- Takemura M, Oda K, Yamato K, Ohta E, Nakamura Y, Nozato N, Akashi K, Ohyama K (1992) Gene clusters for ribosomal proteins in the mitochondrial genome of a liverwort, *Marchantia polymorpha*. Nucleic Acids Res 20:3199–3205
- Wissinger B, Schuster W, Brennicke A (1991) *Trans* splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon- and *trans*-splicing group-II intron-sequences. Cell 65:473-482
- Yang AJ, Mulligan RM (1991) RNA editing intermediates of cox2 transcripts in maize mitochondria. Mol Cell Biol 11:4278–4281
- Zanlungo S, Bégu D, Quiñones V, Araya A, Jordana X (1993) RNA editing of apocytochrome b (*cob*) transcripts in mitochondria from two genera of plants. Curr Genet 24:344–348
- Zanlungo S, Quiñones V, Moenne A, Holuigue L, Jordana X (1994) A ribosomal protein S10 gene is found in the mitochondrial genome in Solanum tuberosum. Plant Mol Biol 25:743–749