

Comparison of mitochondrial and chloroplast genome segments from three onion (*Allium cepa* L.) cytoplasm types and identification of a *trans*-splicing intron of *cox2*

Sunggil Kim · Moo-Kyoung Yoon

Received: 2 December 2009 / Revised: 11 January 2010 / Accepted: 15 January 2010 / Published online: 2 February 2010
© Springer-Verlag 2010

Abstract To study genetic relatedness of two male sterility-inducing cytotypes, the phylogenetic relationship among three cytotypes of onions (*Allium cepa* L.) was assessed by analyzing polymorphisms of the mitochondrial DNA organization and chloroplast sequences. The *atp6* gene and a small open reading frame, *orf22*, did not differ between the normal and CMS-T cytotypes, but two SNPs and one 4-bp insertion were identified in CMS-S cytotype. Partial sequences of the chloroplast *ycf2* gene were integrated in the upstream sequence of the *cob* gene via short repeat sequence-mediated recombination. However, this chloroplast DNA-integrated organization was detected only in CMS-S. Interestingly, disruption of a group II intron of *cox2* was identified for the first time in this study. Like other *trans*-splicing group II introns in mitochondrial genomes, fragmentation of the intron occurred in domain IV. Two variants of each exon1 and exon2 flanking sequences were identified. The predominant types of four variants were identical in both the normal and the CMS-T cytotypes. These predominant types existed as sublimons in CMS-S cytotypes. Altogether, no differences were identified between normal and CMS-T, but significant differences in gene organization and nucleotide sequences were

identified in CMS-S, suggesting recent origin of CMS-T male-sterility from the normal cytotype.

Keywords Onion (*Allium cepa* L.) · Cytoplasmic male sterility (CMS) · Substoichiometric shifting · *cox2* · *Trans*-splicing

Introduction

Plant mitochondrial genomes possess several unusual features in contrast to their animal mitochondrial counterparts, which are generally small (15–18 kb) and circular (Budar et al. 2003; Hanson and Bentolila 2004; Knoop 2004; Kubo and Newton 2008). First, plant mitochondrial genomes are comparatively large and variable, ranging from 208 kb in *Brassica hirta* (Palmer and Herbon 1987) to 2,400 kb in Cucurbitaceae (Ward et al. 1981). Second, the precise configurations of plant mitochondrial genomes are still elusive in spite of complete sequencing of mitochondrial genomes from many higher plant species (Unsel et al. 1997; Kubo et al. 2000; Notsu et al. 2002; Handa 2003; Sugiyama et al. 2005). Observation of plant mitochondrial DNA (mtDNA) using electron microscopy and pulsed-field gel electrophoresis showed mostly linear genome structures with some circular and branched forms (Backert et al. 1997; Oldenburg and Bendich 2001).

The complexity of plant mitochondrial genomes is complicated by multipartite structures and subgenomic mtDNA, termed sublimons (Palmer 1988; Small et al. 1989; Albert et al. 1998; Woloszynska and Trojanowski 2009). Sublimons exist at low copy numbers, sometimes less than one copy per every 100 cells (Arrieta-Montiel et al. 2001). Homologous recombination mediated by repeat sequences present throughout mitochondrial genomes is responsible

Communicated by R. Bock.

S. Kim (✉)
Department of Plant Biotechnology, Biotechnology Research
Institute, Chonnam National University, Gwangju 500-757, Korea
e-mail: dronion@jnu.ac.kr

M.-K. Yoon
National Institute of Horticultural and Herbal Science,
RDA, Suwon 441-440, Korea

for the mtDNA multipartite structure and sublimons (Palmer 1988; Albert et al. 1998). Repeats larger than 1–10 kb are involved in formation of the interconvertible multipartite structures. MtDNA recombination through short repeats of less than several hundred base pairs is responsible for dynamic mtDNA rearrangement and production of multiple sublimons (Small et al. 1989; Kmiec et al. 2006).

Specific stoichiometry of subgenomic mtDNAs is maintained throughout generations during reproduction (Sakai and Imamura 1993; Bellaoui et al. 1998; Janska et al. 1998; Kim et al. 2007). However, the stoichiometry can change due to substoichiometric shifting triggered by events, such as tissue culture (Kanazawa et al. 1994), or nuclear genes, such as *Fr* in *Phaseolus vulgaris* (Mackenzie and Chase 1990; Janska et al. 1998; Abdelnoor et al. 2003). In addition, nuclear genes such as *Msh1* (Abdelnoor et al. 2006), *RecA* (Shedge et al. 2007), and *OSB1* (Zaegel et al. 2006) can suppress mtDNA recombination and maintain sublimon transmission. For example, transgenic tobacco and tomato in which *Msh1* genes were inactivated using RNAi displayed rearranged mtDNA molecules and male-sterility induction (Sandhu et al. 2007).

Dynamic mitochondrial genome rearrangements produced by short repeat sequence-mediated recombination is a driving force of plant mtDNA evolution and is responsible for the creation of chimeric open reading frames (ORFs). Frequent mtDNA rearrangements result in highly variable genome organizations, even within a single species (Satoh et al. 2004; Allen et al. 2007). Chimeric ORFs can sometimes cause phenotypic changes, such as cytoplasmic male-sterility (CMS). CMS plants whose pollen grains are non-viable have been commercially utilized in F_1 hybrid cultivar development for many crop plants.

In onions (*Allium cepa* L.), two types of CMS (CMS-S and CMS-T) have been utilized in F_1 hybrid development. The CMS-S cytoplasm type (cytotype) was discovered first (Jones and Emsweller 1936) and is more widely used. In fact, the scheme of F_1 hybrid seed production using CMS systems, which is globally used in many crops, was developed using onion CMS-S by Jones and Clarke (1943). The fertility of CMS-S male sterility is restored by a single restorer-of-fertility locus (*Ms*). Fertility restoration of CMS-T male sterility is controlled by three independent loci (Berninger 1965; Schweisguth 1973). Therefore, the CMS-S system is preferable in commercial F_1 cultivar breeding because of its simple inheritance of restoration of fertility and stability of male-sterility in diverse environmental conditions (Havey 2000).

However, the male-sterility phenotypes of CMS-S and CMS-T are indistinguishable by visual examination. Furthermore, it takes 4–8 years for breeders to differentiate between the three onion cytotypes (normal, CMS-S, and

CMS-T) by progeny tests since onion is a biennial crop. For these reasons, several molecular markers for differentiation between the normal and CMS-S cytotypes have been developed based on polymorphic sequences of mitochondrial (Sato 1998) and chloroplast genomes (Havey 1995). Meanwhile, a molecular marker for differentiating CMS-S and CMS-T was first reported by Engelke et al. (2003). We previously reported a molecular marker for differentiating between the three onion cytotypes by one simple PCR based on different copy numbers of the chimeric gene *orf725* (Kim et al. 2009a).

Despite several reports of molecular markers, few studies on the identification of male sterility-inducing genes and the phylogenetic relationship among the three onion cytotypes have been performed. To evaluate the mtDNA sequence variation among the three onion cytotypes and to identify additional candidate genes responsible for male-sterility, several key genes frequently involved in creation of chimeric ORFs causing male-sterility (Hanson and Bentolila 2004) were selected for genome walking to isolate their flanking sequences and non-coding sequences of chloroplast genomes were analyzed in this study.

Materials and methods

Plant materials

Two breeding lines of each cytotype whose male-fertility phenotypes had been confirmed in a previous study (Kim et al. 2009a) were used as representatives of the three onion cytotypes. The cytotypes were further confirmed by multiple molecular markers as previously described (Havey 1995; Kim et al. 2009a). Three-leaf stage seedlings were used for total genomic DNA extraction using a commercial DNA extraction kit (DNeasy Plant Mini Kit, QIAGEN, Valencia, CA, USA) according to the manufacturer's protocol.

PCR amplification and genotyping of the cleaved amplified polymorphic sequence (CAPS) marker for *atp6*-flanking regions

PCR was performed in a 10 μ L reaction mixture containing 0.05 μ g template, 0.2 μ L forward primer (10 μ M), 0.2 μ L reverse primer (10 μ M), 0.2 μ L dNTPs (10 mM each), 1 μ L 10 \times PCR buffer, and 0.1 μ L polymerase mix (Advantage 2 Polymerase Mix, Clontech, Palo Alto, CA, USA). PCR amplifications were carried out with an initial denaturation step at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 57–65°C for 30 s, and 72°C for 90 s, and a final 10 min extension at 72°C. The primer sequences used for amplification of different mitochondrial gene organizations

Table 1 Primer sequences used in PCR and RT-PCR analysis

Primer names	Primer sequences (5′–3′)
atp6-F1	CCCAAACCTCTTCCCAGCCCTAACCTCA
atp6-F2	GGATCCGACAACGAACAAGGGAGGTTT
atp6-R1	TGGCTATCGAAAGAATGAGTCCGCAA
Atp1-F1	CCCCCTTCTAGTTTCACACCTTGCCTTG
Atp1-F2	CGGGATGCGATCCTTAATAG
Atp1-F3	GTCTCAGTTGAGATGGGATTGCACGA
Atp1-R1	CGAAAGCCAGGCGACTAGTGTTTCCAT
Atp1-R2	TGGTGAATATTGTGGTTGTTTGGGCACT
Atp1-R2	TGAAAGGGTTGTTTCTGAGGGGAACGA
cob-F1	GGCGCTGATGGGGTGAATCTTCTTTGT
cob-F2	CGGAGCGAAAAGGGTTTTCCATGAGAT
cob-R1	GCATTACGGTATGTCTGGTGGAGATCA
cob-R2	TTGTATGTATGCCCGATCCA
cox2-F1	TCTGATTTTGGTTTTCTGATCACGGATCT
cox2-F2	GCACCTCCGCTGCTTACCAAATTCTT
cox2-R1	GGATGATCAACTCCGCAAAT
cox2-R2	CTCCTCATTGGCAGGAAATC
cox2-R3	GGCGTAAAGGCATGATTAGTTCCACGA
cox2-R4	CCTTCAGTGCGGGATTCAAGATGTTCC
cox2-R5	TCCCATTCACGACCCATGTAACAAGC
Tubulin-F	GGAAGCATGTGCCCGTGCTATATTTG
Tubulin-R	ACAATCTGGATCGTGCGCTTCGTCTTT

The positions of primers are depicted in figures

are presented in Table 1. For genotyping the CAPS markers for *atp6*-flanking regions, PCR products were digested with restriction enzyme, *ApoI*, for 3 h at 37°C. The digested PCR products were electrophoresed on a 1%-agarose gel.

RT-PCR and rapid amplification of cDNA ends (RACE)

Total RNA was extracted from unopened flowers using an RNA extraction kit (RNeasy Plant Mini Kit, QIAGEN) as per the manufacturer's protocol. The extracted RNAs were treated with DNase to remove residual DNA. cDNA was synthesized from total RNA using a commercial cDNA synthesis kit (SuperScript™ III First-Strand Synthesis System for RT-PCR, Invitrogen, Carlsbad, CA, USA) as per the manufacturer's protocol. RT-PCR amplification was performed with an initial denaturation step at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 2 min, and a final 10 min extension at 72°C. The control reaction without RT was carried out at the same time. Primer sequences used for RT-PCR are presented in Table 1. The sequence of onion tubulin that was used as a control was obtained from EST sequences (TC125) from the DFCI *Allium cepa* Gene Index (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi>).

RACE was carried out with a commercial RACE kit (SMART RACE cDNA Amplification Kit; Clontech) as per the manufacturer's protocol.

Genome walking and sequencing of PCR products

Total genomic DNA from normal, CMS-S, and CMS-T cytotypes were extracted from leaf tissues of three-leaf-stage seedlings using a commercial DNA extraction kit (DNeasy Plant Mini Kit, QIAGEN) as per the manufacturer's protocol. Genome walking libraries of the three cytotypes were constructed using the Universal Genome-Walker Kit (Clontech) as per the manufacturer's protocol.

PCR products were purified using the QIAquick PCR Purification kit (QIAGEN). The purified PCR products were either sequenced directly or after cloning into TOPO TA cloning vector for sequencing (Invitrogen). Sequencing reactions were carried out using Big Dye (Applied Biosystems, Foster City, CA, USA) and analyzed using an ABI 3700 Genetic Analyzer (Applied Biosystems) as per the manufacturer's protocol. Nucleotide sequences obtained in this study were deposited into GenBank/EMBL data libraries under accession numbers from GU253298 to GU253309.

Construction of cladogram

Genomic sequences, including introns of *cox2* genes from other plant species, were obtained from GenBank. Alignments were performed by BioEdit (Hall 1999), and gaps were removed by Gblocks software (Castresana 2000). The cladogram was constructed using *MEGA* version 4 (Tamura et al. 2007).

Results

Identification of candidate ORFs in the flanking sequences of *atp6* and *Atp1* genes

Partial sequences of the onion *atp6* gene were isolated using primers designed based on conserved sequences from other plant species since few onion mtDNA sequences were available. Approximately 3.5 kb of mtDNA sequences, including full-length *atp6* and its 5′- and 3′-flanking regions were obtained by genome walking (Fig. 1a). Two types of *atp6*-flanking sequences (*atp6*-type1 and *atp6*-type2) were isolated from the three onion cytotypes (Fig. 1a).

There were two single nucleotide polymorphisms (SNPs) and one 4-bp insertion between the two variants. A CAPS marker was developed to distinguish between the two variants. *Atp6*-type1 was exclusively present in both

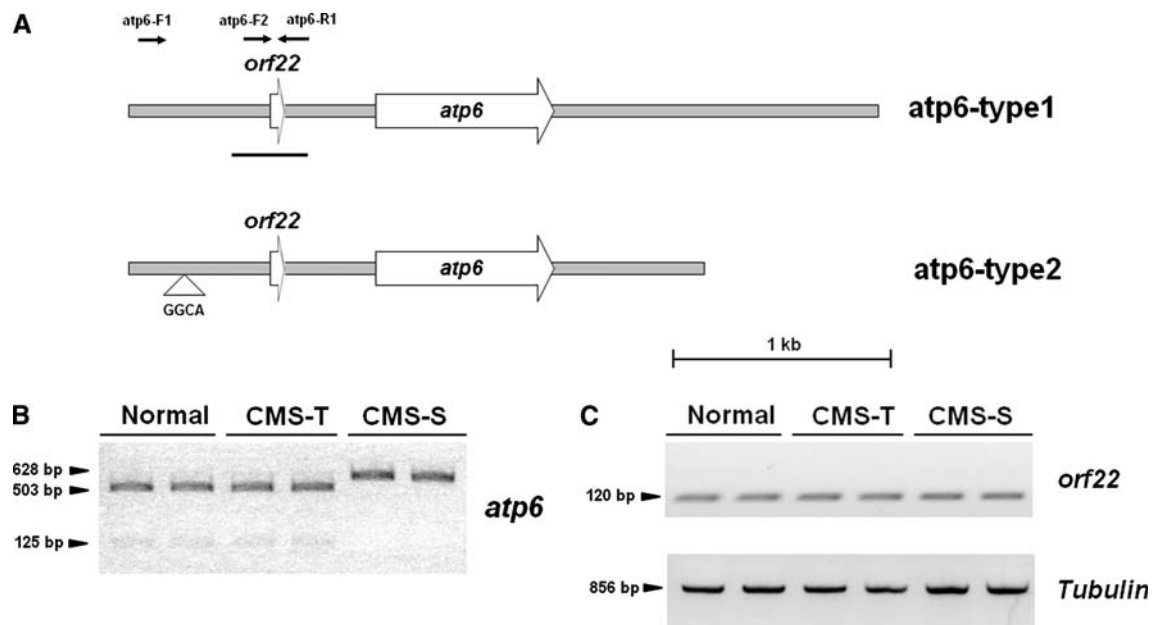


Fig. 1 Comparison of onion *atp6* and its flanking sequences among the three onion cytotypes. **a** Organizations of *atp6* and its flanking sequences. Arrow-shaped boxes the 5'–3' direction. Horizontal arrows primer-binding sites. The horizontal line under the *orf22* indicates the position of the full-length cDNA revealed by RACE. Different colored boxes indicate dissimilar sequences. The recognition site of

ApoI restriction enzyme is broken by the 'GGCA' insertion in *atp6*-type2. Sequences were deposited into GenBank under the accession numbers of GU253298 (*atp6*-type1) and GU253299 (*atp6*-type2). **b** PCR products digested with the *ApoI* restriction enzyme. PCR products were amplified using *atp6*-F1 and *atp6*-R1 primers. **c** RT-PCR products amplified using *atp6*-F2 and *atp6*-R1 primers

the normal and CMS-T cytotypes, whereas *atp6*-type2 was detected only in the CMS-S cytotype (Fig. 1b).

A small ORF (*orf22*) was identified in the upstream region of the *atp6* gene. Initially, this ORF was thought to be co-transcribed with the *atp6* gene due to its close proximity to the *atp6* promoter region. However, *orf22* was transcribed independently, as confirmed by RACE analysis. The position of the full-length transcript of the *orf22* is depicted in Fig. 1a. However, the transcription level of *orf22* was relatively low and was not significantly different among the three cytotypes (Fig. 1c).

The publicly available EST sequence of the partial *Atp1* gene was used as template for genome walking. Three variants of the *Atp1*-flanking sequences were obtained (Fig. 2a). Similar to the *atp6*-flanking sequences, the variant containing the entire *Atp1* and *cob* genes at its 5' portion (*Atp1*-type1) was predominantly present in both normal and CMS-T cytotypes. The variant consisting of the entire *Atp1* gene and exon2 of the *nad1* gene at its 5' portion (*Atp1*-type2) was detected only in CMS-S cytotype (Fig. 2b). Interestingly, the third variant was a chimeric ORF (*orf435*), which consisted of the majority of the *Atp1* coding sequences and upstream sequences of the *cox3* gene. However, this chimeric ORF appeared to exist at a substoichiometric level in all three cytotypes because only small amounts of PCR products were observed in comparison to those of normal *Atp1* genes (Fig. 2b).

Integration of a partial sequence of the chloroplast *ycf2* gene in the flanking sequence of the *cob* gene

The partial coding sequence of the *cob* gene was obtained using primers designed from conserved sequences of *cob* genes from other plant species. From this sequence, complete coding and flanking sequences of the *cob* genes were isolated from onion genome walking libraries. Two variant organizations of *cob*-flanking sequences were obtained (Fig. 3a). The PCR results indicated that *cob*-type1 is present on master chromosomes of normal and CMS-T cytotypes, but it existed as a sublimon in CMS-S cytotype (Fig. 3b).

The second organization (*cob*-type2) was shown to possess partial sequences of the chloroplast *ycf2* gene at the immediate upstream sequence of the *cob* gene (Fig. 3a). PCR analysis indicated that the second variant was predominant only in the CMS-S cytotype (Fig. 3b). This *ycf2* gene integration might correspond to the chloroplast sequence integration reported by Sato (1998), though detail comparison was not possible due to unavailability of his sequence information.

Comparison of the integrated *ycf2* gene with complete coding sequences of *ycf2* genes from other plant species indicated that an approximately 2.2 kb coding region within the 3' end was integrated (Fig. 4a). Partial sequences of the onion chloroplast *ycf2* gene were isolated using primers

Fig. 2 Onion *Atp1* and its flanking sequences from the three cytotypes. **a** Organizations of *Atp1* and its flanking sequences. Arrow-shaped boxes the 5'–3' direction. Horizontal arrows primer-binding sites. Different colored boxes dissimilar sequences. Sequences were deposited into GenBank under the accession numbers of GU253300 (*Atp1*-type1), GU253301 (*Atp1*-type2), and GU253302 (*orf435*). **b** PCR results from three different variants. The primer pairs of *Atp1*-F1 + *Atp1*-R1, *Atp1*-F2 + *Atp1*-R1, and *Atp1*-F3 + *Atp1*-R2 were used for amplification of *Atp1*-type1, *Atp1*-type2, and *orf435*, respectively

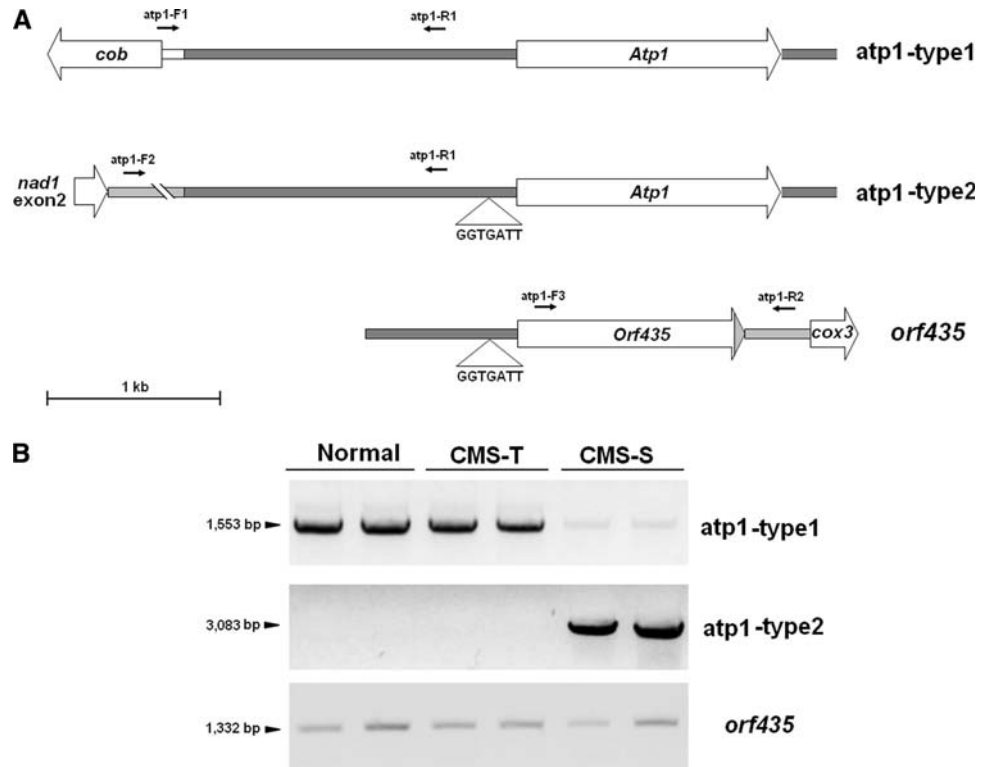
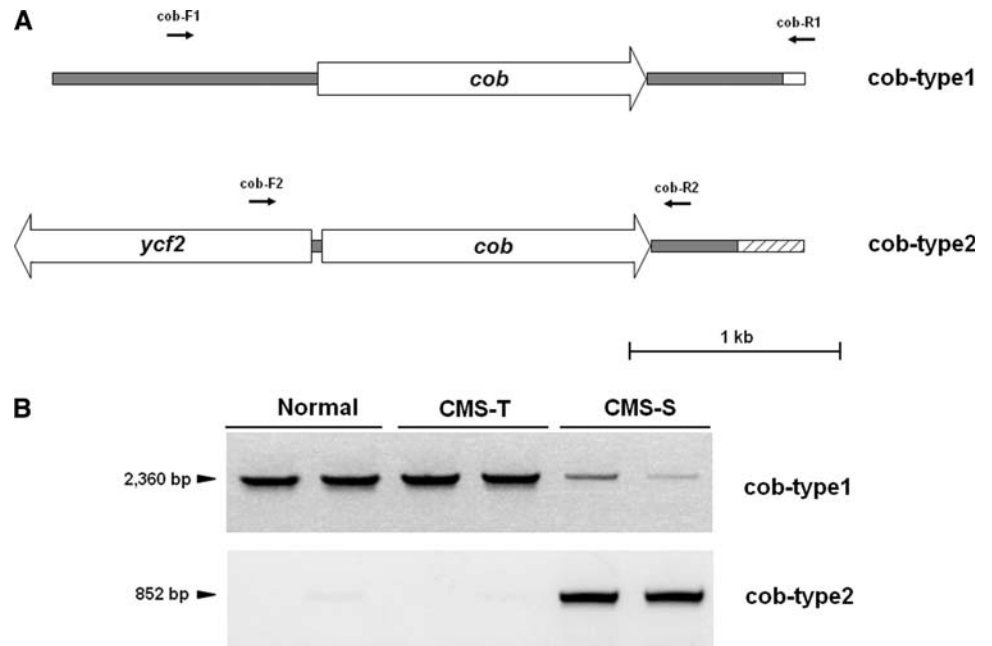


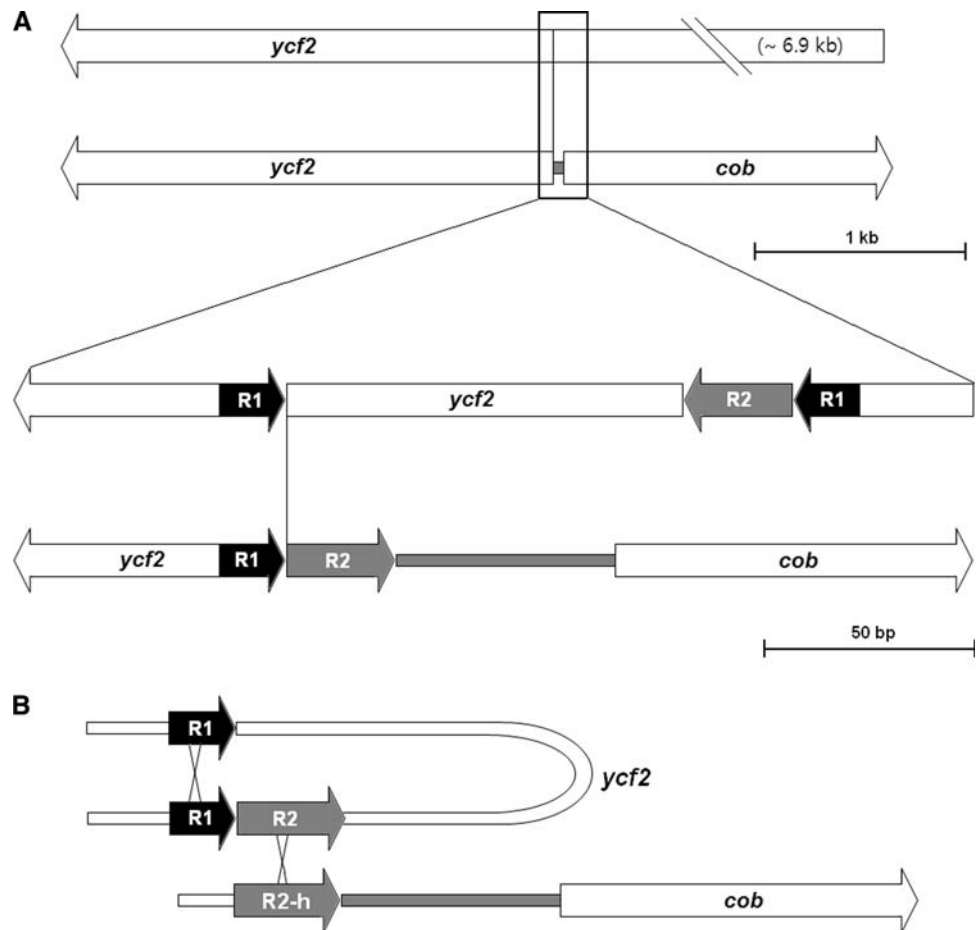
Fig. 3 Onion *cob* and its flanking sequences from the three cytotypes. **a** Organizations of *cob* and its flanking sequences. Arrow-shaped boxes the 5'–3' direction. Horizontal arrows primer-binding sites. Different colored boxes dissimilar sequences. Sequences were deposited into GenBank under the accession numbers of GU253303 (*cob*-type1) and GU253304 (*cob*-type2). **b** PCR products specific to two different variants. The primer pairs of *cob*-F1 + *cob*-R1 and *cob*-F2 + *cob*-R2 were used for amplification of *cob*-type1 and *cob*-type2, respectively



based on the conserved *ycf2* coding sequences which were not present in the integrated *ycf2* region. This partial sequence of chloroplast *ycf2* gene was deposited into GenBank under the accession number of GU253309. There was only one SNP and one 4-bp insertion within the integrated 2,189 bp and chloroplast *ycf2* coding sequences, suggesting relatively recent integration of chloroplast DNA (cpDNA) into mitochondrial genomes. Interestingly, there were two

short repeat sequences (R1 and R2) at the breakpoint of rearrangement ((Fig. 4a). The R1 and R2 repeats were also identified in the chloroplast *ycf2* gene, and were 94 bp away from the breakpoint with inverted orientation (Fig. 4a). This structural feature suggests that the integration of chloroplast *ycf2* might have occurred by multiple homologous recombinations mediated by short repeat sequences, as shown in Fig. 4b. The homologous R2 repeat sequence

Fig. 4 Integrated partial sequence of the chloroplast *ycf2* gene in the 5' flanking region of the *cob* gene. **a** Alignment of the chloroplast and integrated *ycf2* genes. Enlargement of sequences in the rectangular box are shown below. R1 and R2 represent repeat sequences. **b** Model showing integration of the chloroplast *ycf2* gene via repeat sequence-mediated recombination. R2-h: putative R2 repeat homologous sequence in mtDNA



(R2-h) showing 65% nucleotide identity with the chloroplast R2 repeat was identified at the rearrangement breakpoint of the *cob*-type1 sequence.

Identification of a *trans*-splicing intron of the *cox2* gene in onion mitochondrial genomes

Except for a few species, such as pea in which the *cox2* gene is intronless, *cox2* genes in plant mitochondrial genomes generally contain a single group II intron (Bonen 2008). These *cox2* group II introns in most plant species are less than 1.5 kb. Therefore, primers based on the conserved exon1 and exon2 sequences of *cox2* genes of other plant species were used to isolate the onion *cox2* gene. However, we failed to obtain any PCR products using multiple pairs of primers. Thus, genome walking was carried out separately from exon1 and exon2 conserved regions. Two variants of each exon1 and exon2 flanking sequences were isolated.

Surprisingly, the group II intron was disrupted by rearrangements in domain IV in all of the organizations (Fig. 5a). Since the group II intron sequences of the *cox2* genes of other plant species were sufficiently conserved around the breakpoints of the rearrangements, the exact

positions of rearrangements in both exon1 and exon2 could be easily identified, as shown in Fig. 6a. In addition, it is unlikely that the onion *cox2* gene contained an unusually large group II intron, since both closely related monocots and distantly related dicotyledonous species were shown to have relatively conserved intron sequences of which sizes range from 794 bp in maize to 1,463 bp in sugar beet (Fig. 6b). Therefore, these results indicate that the onion *cox2* group II intron is probably *trans*-spliced. RT-PCR and RACE analyses indicated that the *cox2* transcript sequences were identical to the genomic sequences of exon1 and exon2 (Fig. 5b), except for 17 RNA editing sites. In addition, 3' RACE analysis identified intermediary primary transcripts, which included exon1 and intron domain I–IV, and terminated at the breakpoint. Another primary transcript whose transcription start site was positioned at 95 bp upstream of the breakpoint of exon2 was identified during 5' RACE (Fig. 5a).

PCR analysis indicated that Exon1-1 and Exon2-1 organizations were present on the master chromosomes of the normal and CMS-T cytotypes. Exon1-2 and Exon2-2 were the predominant structures in the CMS-S cytotype, and were present as sublimons in both normal and CMS-T cytotypes (Fig. 5c).

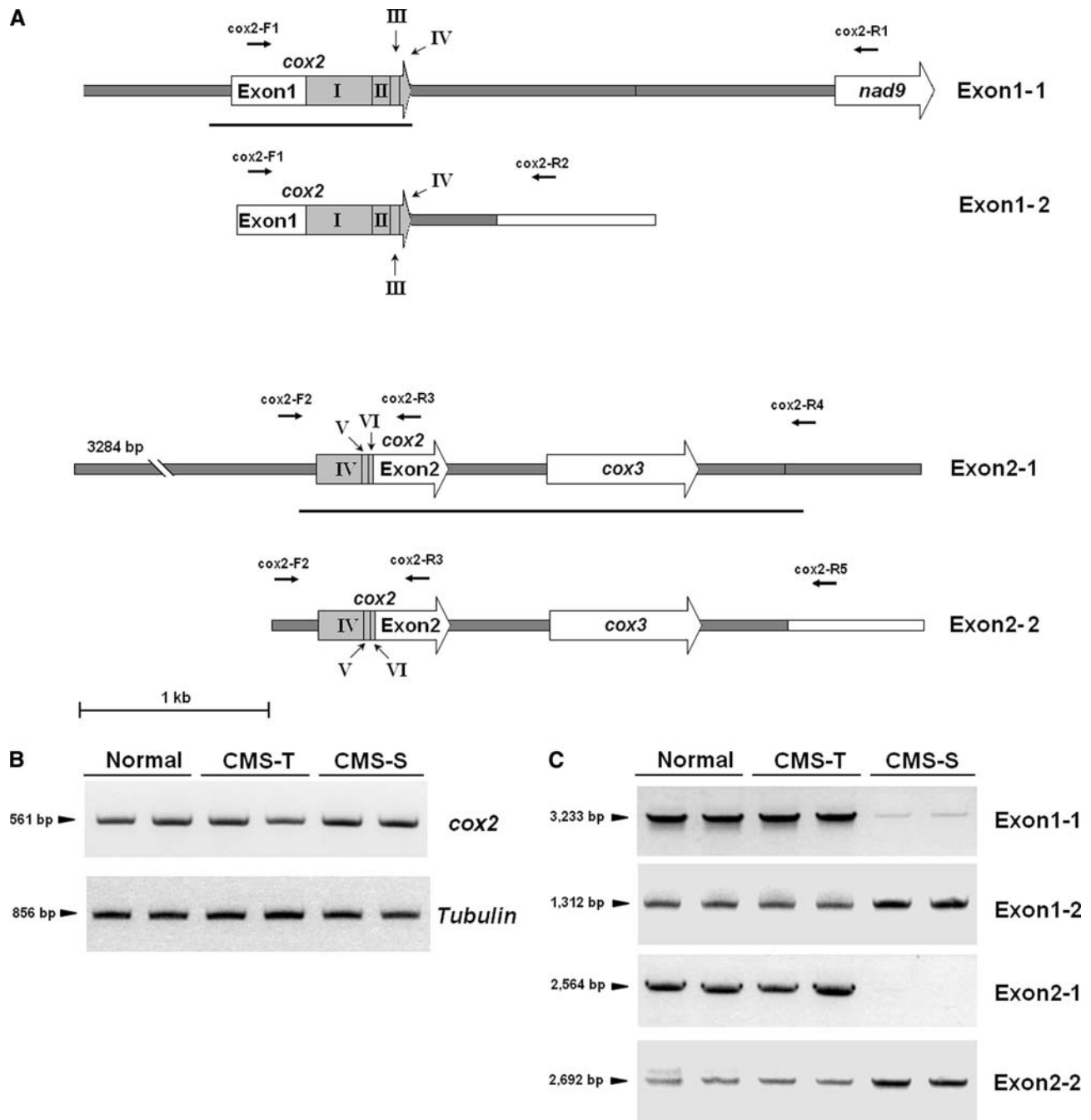


Fig. 5 Onion *cox2* gene and its flanking sequences from the three cytotypes. **a** Organizations of *cox2* and its flanking sequences. Arrow-shaped boxes the 5'–3' direction. Horizontal arrows primer-binding sites. Different colored boxes dissimilar sequences. The horizontal lines under the *cox2* gene indicate the position of the full-length primary transcripts revealed by RACE. I–VI six conserved helical domains of a group II intron. Sequences were deposited into GenBank

under the accession numbers of GU253305 (Exon1-1), GU253306 (Exon1-2), GU253307 (Exon2-1), and GU253308 (Exon2-2). **b** RT-PCR products of the *cox2* gene amplified with *cox2*-F1 and *cox2*-R3 primers. **c** PCR results from four different organizations. The primer pairs of *cox2*-F1 + *cox2*-R1, *cox2*-F1 + *cox2*-R2, *cox2*-F2 + *cox2*-R4, and *cox2*-F2 + *cox2*-R5 were used for amplification of Exon1-1, Exon1-2, Exon2-1, and Exon2-2, respectively

Comparison of non-coding chloroplast sequences among the three cytotypes

Except for a previously reported chimeric gene, *orf725* which is present in the master chromosomes of both the CMS-T and

the CMS-S cytotypes, but as a sublimon in the normal cytotype (Kim et al. 2009a), no differences between the nucleotide sequences and stoichiometry of the variant organizations between normal and CMS-T cytotypes were identified in this study. To assess genetic relatedness among the three cytotypes,

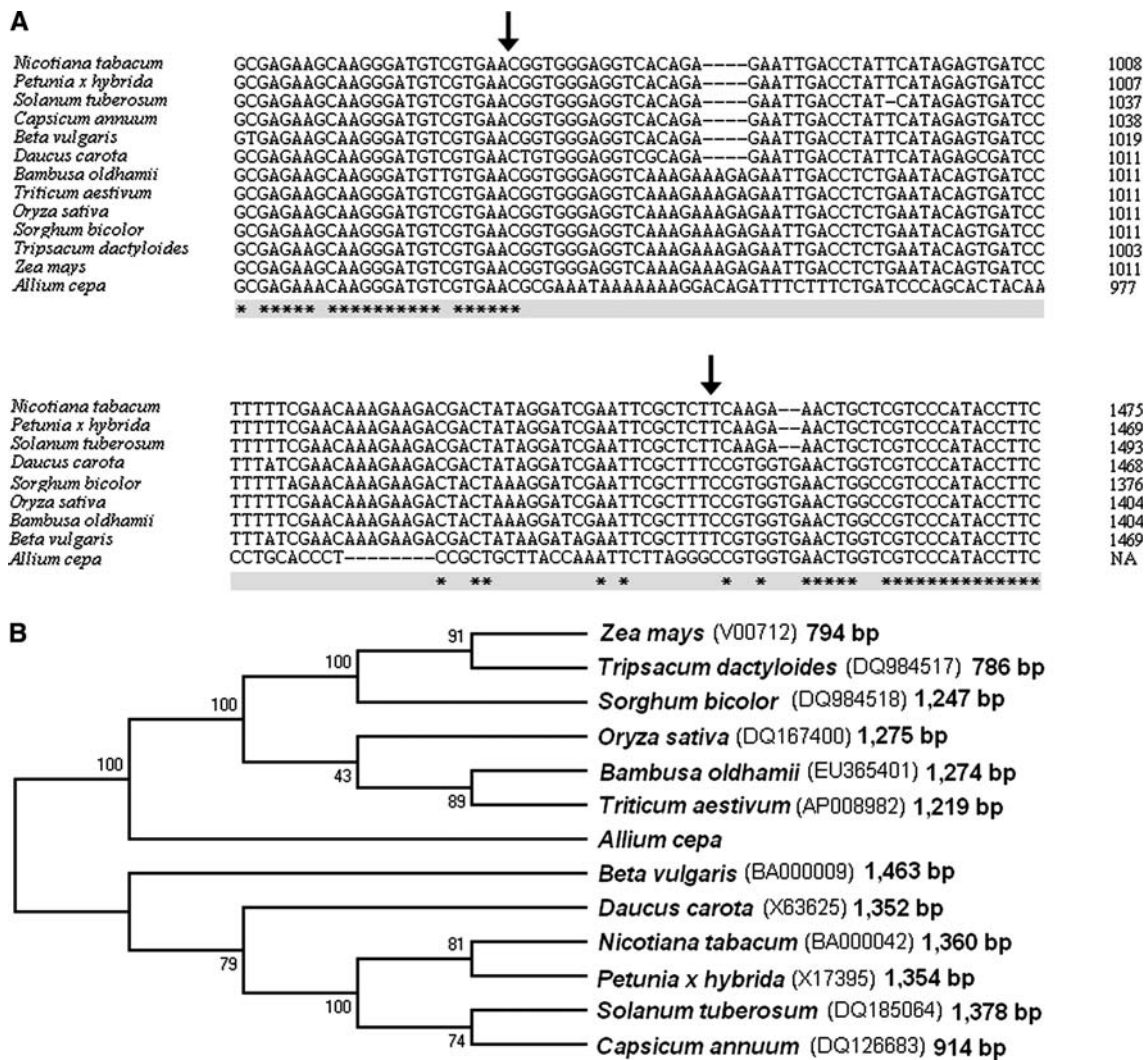


Fig. 6 Discontinuous group II intron of the onion *cox2* gene. **a** Alignment of domain IV sequences from the onion *cox2* gene around breakpoints created by onion mtDNA rearrangement with those from other species. In case of breakpoint regions for the 3' side, only nine plant species were aligned due to lack of those regions in four plant species. **b** Cladogram showing the genetic relationship of the onion *cox2* gene

with those from other species. Both exon and intron sequences were used to construct the cladogram. The numbers at the nodes are the bootstrap probability (%) with 1,000 replicates. GenBank accession numbers are shown in parentheses. Lengths of group II introns of the *cox2* genes are shown next to the GenBank accession numbers

a total of 4.6 kb of non-coding sequences of chloroplast genomes from the three cytotypes were obtained (Table 2). Interestingly, there were no polymorphisms between normal and CMS-T cytotypes, while many SNPs and indels were identified in the CMS-S cytotype, suggesting a very recent divergence of CMS-T male-sterility from the normal cytotype.

Discussion

Identification of a *trans*-splicing group II intron in the onion *cox2* gene

Introns identified in all kingdoms are classified into five categories: group I and group II introns, nuclear tRNA

introns, archaeal introns, and spliceosomal mRNA introns based on splicing mechanisms and conserved RNA-folding patterns. Group II introns are present in chloroplast and mitochondrial genomes of plants and some lower eukaryotes and prokaryotes (Glanz and Kück 2009). Group II introns are large self-splicing ribozymes whose secondary structure generally contains six distinct conserved helical domains radiating from a central hub (Pyle et al. 2007; Toor et al. 2008). Some group II introns are discontinuous, having exons dispersed in genomes. Therefore, at least two primary transcripts are ligated by *trans*-splicing. In chloroplasts, *trans*-splicing group II introns have been identified in six genes (*pbsA*, *petD*, *psaA*, *psaC*, *rbcL*, and *rps12*) of some algae and higher plants (Glanz and Kück 2009). Meanwhile, most group II introns in plant mitochondrial

Table 2 Polymorphisms among non-coding sequences of chloroplast genomes from the three onion cytotypes

Positions	Length (bp) ^a	Numbers of polymorphisms among the three cytotypes ^b		
		Normal versus CMS-T	Normal versus CMS-S	CMS-T versus CMS-S
<i>trnS-trnG</i>	621	0	4	4
<i>rbcL-accD</i>	230	0	4	4
<i>atpB-rbcL</i>	325	0	2	2
<i>trnQ-psbK</i>	325	0	2	2
<i>psbK-psbI</i>	341	0	2	2
<i>rps16</i> intron	358	0	3	3
<i>rps16-trnQ</i>	819	0	6	6
<i>trnQ-trnS</i>	668	0	3	3
<i>trnK-rps16</i>	890	0	6	6
Total	4,577	0	32	32

^a Fragment lengths of the normal cytotype are given. Fragment lengths of other mitotypes are slightly different due to indels

^b Insertions or deletions are counted as one polymorphic nucleotide, regardless of the length

genomes are *cis*-spliced. However, *trans*-splicing group II introns have been exclusively identified in genes (*nad1*, *nad2*, *nad3*, and *nad5*) encoding subunits of the NADH dehydrogenase complex (Bonen 2008; Glanz and Kück 2009). To date, no other mitochondrial genes containing *trans*-splicing group II introns have been reported. However, here we identified a putative *trans*-splice of the onion *cox2* gene, making it the first mitochondrial gene containing *trans*-splicing group II introns not coding for the NADH dehydrogenase complex.

The first evidence suggesting *trans*-splicing of the onion *cox2* gene was that no expected PCR products could be obtained when multiple sets of primers encompassing exon1 and exon2 were used for PCR. More than ten combinations of primer pairs were used for amplification of the expected *cox2* intron without successful PCR amplification. Kudla et al. (2002) used PCR to identify the presence of a *cox2* intron in 36 monocotyledonous species using similar primers designed from the conserved *cox2* exons. *Cis*-splicing group II introns from 32 species have been identified. Interestingly, they also found *cis*-splicing *cox2* introns in two onion relatives: *Allium ramosum* and *A. sativum*. This suggests that disruption of the onion *cox2* group II intron occurred very recently in onions. All *trans*-splicing introns of plant mitochondrial genes are assumed to be originally *cis*-splicing, but disruption of introns has occurred multiple times during evolution (Bonen 2008). Qiu and Palmer (2004) showed that fracture of *nad1* group II intron has occurred 15 times during angiosperm evolution. Therefore, recent dynamic rearrangement of onion mitochondrial genomes probably led to fragmentation of the onion *cox2* intron.

The genomic sequences of onion *cox2* exons were identical to the mRNA sequence of onion *cox2*, except for RNA editing sites. RNA editing also proved that genomic

sequences of onion *cox2* identified in this study were present in mitochondrial genomes. In addition, PCR analysis showed that the *cox2* organizations containing the disrupted *cox2* intron were the only identified *cox2* mtDNA molecules (Fig. 5c). Furthermore, the possibility of an exceptionally long intron can be eliminated because the entire *nad9* gene was positioned at the 3' end of the disrupted intron connected to exon1. Likewise, no expected exon1 sequence was identified in the 3,284 bp upstream sequence of the disrupted intron connected to exon2. Lengths of most *cox2* introns are less than 1.5 kb, and the largest *cox2* intron identified to date is 2,659 bp from *Ginkgo biloba* (Bonen 2008). Therefore, it is unlikely that the onion *cox2* gene contains an intron larger than 6 kb.

Fragmentation of group II introns in plant mitochondrial genes has occurred many times, as described above. However, the positions of fragmentations of all reported *trans*-splicing mitochondrial group II introns are in the loop of domain IV. The consensus positions of fragmentation might be related to functional constraints of the tertiary structure of group II introns (Bonen 2008; Glanz and Kück 2009). As expected, the rearrangements disrupting the onion *cox2* intron were positioned in the domain IV (Fig. 5a). Furthermore, identification of intermediary primary transcripts starting and terminating around the rearrangement breakpoint strongly support *trans*-splicing of the onion *cox2* gene.

Comparison of partial sequences of mitochondrial and chloroplast genomes among the three onion cytotypes

To identify candidate male sterility-inducing genes and to assess the variability of mitochondrial genomes of the three onion cytotypes, some of the genes known to be frequently involved in the creation of chimeric ORFs responsible for

male-sterility in many plant species (Hanson and Bentolila 2004) were analyzed in this study. Ideally, it would be better to compare complete mitochondrial genome sequences from multiple cytotypes for the identification of candidate ORFs as shown in maize (Allen et al. 2007). However, it is difficult to obtain complete mitochondrial genome sequences of higher plants because existence of master circles and even in vivo structures of plant mitochondrial genomes are still controversial (Backert et al. 1997; Oldenburg and Bendich 2001). In addition, most complete plant mitochondrial genome sequences lack information about substoichiometric mtDNA molecules which play a crucial role in evolution of plant mitochondrial genomes (Small et al. 1989).

Here, we reported various organizations of mtDNA molecules and differential stoichiometry of these variants among the three onion cytotypes. No highly promising candidate ORFs causing male-sterility were identified. However, an interesting relationship among the three cytotypes was illustrated. No differences in the nucleotide sequences, gene organizations, and relative copy numbers of the isolated mtDNA were identified between normal and CMS-T cytotypes. However, there were many polymorphisms in the CMS-S cytotype. These findings strongly suggest that CMS-T male-sterility is recently derived from the normal cytotype. This hypothesis was further supported by comparison of highly variable non-coding sequences of chloroplast genomes of the three cytotypes. Likewise, no polymorphism was detected in non-coding sequences between normal and CMS-T cytotypes. Since mitochondria and chloroplasts are strictly co-transmitted to the next generation in most angiosperms (Reboud and Zeyl 1993), non-coding chloroplast genomes have been utilized in phylogenetic studies at the intra-specific level (Ohsako et al. 1996; Gao et al. 2007; Meng et al. 2007; Kim et al. 2009b).

Though differences were identified between the CMS-S cytotype and the other cytotypes, no additional differences were uncovered between the normal and CMS-T cytotypes. We previously reported that *orf725* was not detected in the normal cytotype, but abundant copy numbers were detected in both CMS-T and CMS-S cytotypes (Kim et al. 2009a). In total, these results imply that recent substoichiometric shifting of a few substoichiometric chimeric ORFs, including *orf725*, in the normal cytotype might have induced CMS-T male-sterility. Additionally, the phylogenetic relationship between the normal and CMS-T cytotypes was too close to produce any *de novo* mtDNA rearrangements or nucleotide changes. Therefore, it would be worth studying the function of *orf725* as a male sterility-inducing gene in CMS-T. More extensive sequencing of complete mitochondrial genomes is required for identification of additional candidate chimeric ORFs.

In conclusion, dynamic onion mtDNA rearrangement was shown by identification of *trans*-splicing intron of *cox2* gene, but the fact that few polymorphisms on the mtDNA organizations and non-coding chloroplast sequences were identified between normal and CMS-T cytotypes implies that CMS-T male-sterility might be induced by very recent substoichiometric shifting of a few mtDNA molecules from the normal cytotype. These results are informative to understand the mechanism of male-sterility induction and fertility restoration in two male-sterility inducing onion cytotypes, and to utilize male-sterility in onion F₁ hybrid breeding. In addition, several molecular markers for identifying polymorphic mtDNA organizations and cpDNA sequences were developed in this study. Previously, two novel cytotypes were found using such mtDNA-based molecular markers from radish germplasm (Kim et al. 2007; Lee et al. 2008). Similarly, these molecular markers can be useful for identifying novel onion cytotypes including new forms of CMS from diverse onion germplasms.

Acknowledgments This study was supported by the Specific Joint Agricultural Research-Promoting Projects (Project No. 200803A0 1081219), RDA, Republic of Korea, and by the Biotechnology Research Institute, Chonnam National University, Republic of Korea. We thank Seon Chong and Mi-Young Kim for their dedicated technical assistance.

References

- Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proc Natl Acad Sci* 100:5968–5973
- Abdelnoor RV, Christensen AC, Mohammed S, Munoz-Castillo B, Moriyama H, Mackenzie SA (2006) Mitochondrial genome dynamics in plants and animals: convergent gene fusions of a *MutS* homologue. *J Mol Evol* 63:165–173
- Albert B, Godelle B, Gouyon PH (1998) Evolution of the plant mitochondrial genome: dynamics of duplication and deletion of sequences. *J Mol Evol* 46:155–158
- Allen JO, Fauron CM, Mink P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177:1173–1192
- Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics* 158:851–864
- Backert S, Neilsen BL, Börner T (1997) The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trend Plant Sci* 2:477–483
- Bellaoui M, Martin-Canadell A, Pelletier G, Budar F (1998) Low-copy-number molecules are produced by recombination, actively maintained and can be amplified in the mitochondrial genome of Brassicaceae: relationship to reversion of the male sterile phenotype in some cybrids. *Mol Gen Genet* 257:177–185
- Berninger E (1965) Contribution à l'étude de la stérilité mâle de l'oignon (*Allium cepa* L.). *Ann Amélior Plant* 15:183–199

- Bonen L (2008) *Cis-* and *trans-*splicing of group II introns in plant mitochondria. *Mitochondrion* 8:26–34
- Budar F, Touzet P, De Paepe R (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revised. *Genetica* 117:3–16
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
- Engelke T, Terefe D, Tatlioglu T (2003) A PCR-based marker system monitoring CMS-(S), CMS-(T) and (N)-cytoplasm in the onion (*Allium cepa* L.). *Theor Appl Genet* 107:162–167
- Gao LM, Möller M, Zhang XM, Hollingsworth ML, Liu J, Mill RR, Gibby M, Li DZ (2007) High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Mol Ecol* 16:4684–4698
- Glanz S, Kück U (2009) Trans-splicing of organelle introns—a detour to continuous RNAs. *Bioessays* 31:921–934
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res* 31:5907–5916
- Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16:S154–S169
- Havey MJ (1995) Identification of cytoplasmic using the polymerase chain reaction to aid in the extraction of maintainer lines from open-pollinated populations of onion. *Theor Appl Genet* 90:263–268
- Havey MJ (2000) Diversity among male-sterility-inducing and male-fertile cytoplasmic of onion. *Theor Appl Genet* 101:778–782
- Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA (1998) Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. *Plant Cell* 10:1163–1180
- Jones HA, Clarke A (1943) Inheritance of male sterility in the onion and the production of hybrid seed. *Proc Am Soc Hortic Sci* 43:189–194
- Jones HA, Emsweller SL (1936) A male-sterile onion. *Proc Am Soc Hortic Sci* 34:582–585
- Kanazawa A, Tsutsumi N, Hirai A (1994) Reversible changes in the composition of the population of mtDNAs during dedifferentiation and regeneration in tobacco. *Genetics* 138:865–870
- Kim S, Lim H, Park S, Cho K, Sung S, Oh D, Kim K (2007) Identification of a novel mitochondrial genome type and development of molecular markers for cytoplasm classification in radish (*Raphanus sativus* L.). *Theor Appl Genet* 115:1137–1145
- Kim S, Lee E, Cho DY, Han T, Bang H, Patil BS, Ahn YK, Yoon M (2009a) Identification of a novel chimeric gene, *orf725*, and its use in development of a molecular marker for distinguishing among three cytoplasm types in onion (*Allium cepa* L.). *Theor Appl Genet* 118:433–441
- Kim S, Lee Y, Lim H, Ahn Y, Sung S (2009b) Identification of highly variable chloroplast sequences and development of cpDNA-based molecular markers that distinguish four cytoplasm types in radish (*Raphanus sativus* L.). *Theor Appl Genet* 119:189–198
- Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr Genet* 50:149–159
- Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet* 46:123–139
- Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8:5–14
- Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA^{cys}(GCA). *Nucleic Acids Res* 28:2571–2576
- Kudla J, Albertazzi FJ, Blazevic D, Hermann M, Bock R (2002) Loss of the mitochondrial *cox2* intron 1 in a family of monocotyledonous plants and utilization of mitochondrial intron sequences for the construction of a nuclear intron. *Mol Genet Genomics* 267:223–230
- Lee Y, Park S, Lim C, Kim H, Lim H, Ahn Y, Sung S, Yoon M, Kim S (2008) Discovery of a novel cytoplasmic male-sterility and its restorer lines in radish (*Raphanus sativus* L.). *Theor Appl Genet* 117:905–913
- Mackenzie SA, Chase CD (1990) Fertility restoration is associated with loss of a portion of the mitochondrial genome in cytoplasmic male-sterile common bean. *Plant Cell* 2:905–912
- Meng L, Yang R, Abbott RJ, Miede G, Hu T, Liu J (2007) Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan plateau and adjacent highlands. *Mol Ecol* 16:4128–4137
- Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Kirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268:434–445
- Ohsako T, Wang G, Miyashita NT (1996) Polymerase chain reaction-single strand conformational polymorphism analysis of intra- and interspecific variations in organellar DNA regions of *Aegilops mutica* and related species. *Genes Genet Syst* 71:281–292
- Oldenburg DJ, Bendich AJ (2001) Mitochondrial DNA from the Liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. *J Mol Biol* 310:549–562
- Palmer JD (1988) Intraspecific variation and multicircularity in *Brassica* mitochondrial DNAs. *Genetics* 118:341–351
- Palmer JD, Herbon LA (1987) Unicircular structure of the *Brassica hirta* mitochondrial genome. *Curr Genet* 11:565–570
- Pyle AM, Fedorova O, Waldsich C (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trend Biochem Sci* 32:138–145
- Qiu Y, Palmer JD (2004) Many independent origins of *trans* splicing of a plant mitochondrial group II intron. *J Mol Evol* 59:80–89
- Reboud X, Zeyl C (1993) Organelle inheritance in plants. *Heredity* 72:132–140
- Sakai T, Imamura J (1993) Evidence for a mitochondrial sub-genome containing radish *Atp1* in a *Brassica napus* cybrid. *Plant Sci* 90:95–103
- Sandhu AS, Abdelnoor RV, Mackenzie SA (2007) Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. *Proc Natl Acad Sci* 104:1766–1770
- Sato Y (1998) PCR amplification of CMS-specific mitochondrial nucleotide sequences to identify cytoplasmic genotypes of onion (*Allium cepa* L.). *Theor Appl Genet* 96:367–370
- Satoh M, Kubo T, Nishizawa S, Estiati A, Itchoda N, Mikami T (2004) The cytoplasmic male-sterile type and normal type mitochondrial genomes of sugarbeet share the same complement of genes of known function but differ in the content of expressed ORFs. *Mol Gen Genomics* 272:247–256
- Schweigsuth B (1973) Étude d'un nouveau type de stérilité male chez l'oignon, *Allium cepa* L. *Ann Amélior Plant* 23:221–233
- Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA (2007) Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. *Plant Cell* 19:1251–1264
- Small I, Suffolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. *Cell* 58:69–76
- Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome:

- comparative analysis of mitochondrial genomes in higher plants. *Mol Gen Genomics* 272:603–615
- Tamura K, Dudley J, Nei M, Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Toor N, Keating KS, Taylor SD, Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* 320:77–82
- Unsold M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature genet* 15:57–61
- Ward BL, Anderson RS, Bendich AJ (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell* 25:793–803
- Woloszynska M, Trojanowski D (2009) Counting mtDNA molecules in *Phaseolus vulgaris*: sublimons are constantly produced by recombination via short repeats and undergo rigorous selection during substoichiometric shifting. *Plant Mol Biol* 70:511–521
- Zaegel V, Guermann B, Le Ret M, Andrés C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell* 18:3548–3563