Title: **Two male sterility-inducing cytoplasms of beet (*Beta vulgaris*) are genetically distinct but have closely related mitochondrial genomes: implication of a substoichiometric mitochondrial DNA molecule in their evolution**

Authors: Yasuyuki Onodera[1], Takumi Arakawa[1], Rika Yui-Kurino[1], Masayuki P. Yamamoto[1, *], Kazuyoshi Kitazaki[1, **], Shigehiko Ebe[2, ***], Muneyuki Matsunaga[1], Kazunori Taguchi[3], Yosuke Kuroda[3], Shiko Yamashita[1], Tomoyuki Sakai[1], Toshiro Kinoshita[2], Tetsuo Mikami[1], Tomohiko Kubo[1]

Affiliations: [1]Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

[2]Plant Breeding Institute, Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

[3]Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Memuro, Hokkaido 082-0081, Japan

Corresponding author:

Tomohiko Kubo

Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

Tel/Fax +81-11-706-2484

E-mail gelab@abs.agr.hokudai.ac.jp

Conflict of interest: The authors declare that they have no conflict of interest.

Present addresses:

[*]Department of Biology, Faculty of Science, University of Toyama, Toyama, 930-8555, Japan

[**]Lowland Farming and Horticulture Research Division, Kyushu Okinawa Agricultural Research Center, National Agriculture and Food Research Organization, Kurume, Fukuoka, 839–8503, Japan

[***]Tokachi Agricultural Experiment Station, Agriculture Research Department, Hokkaido Research Organization, Memuro, Hokkaido, 082-0081, Japan

Abstract

I-12CMS(2) and I-12CMS(3) are sugar beet lines with different sources of cytoplasmic male sterility (CMS) derived from wild beets in Turkey and Pakistan, respectively. We established that I-12CMS(2) has a genetically distinct cytoplasm, but its mitochondrial genome is very similar to I-12CMS(3). Male fertility was assessed in F_1 hybrids produced with a common pollen parent. Fertility in the F_1's carrying the I-12CMS(3) cytoplasm exceeded that of the F_1's with the I-12CMS(2) cytoplasm. Organization of the I-12CMS(2) and I-12CMS(3) mitochondrial genomes were compared based on their physical maps. Mitochondrial genomes of the two strains were largely collinear, except for a large deletion in the noncoding region of I-12CMS(2). Because a mitochondrial *orf129* in the I-12CMS(3) cytoplasm is associated with a male sterility phenotype and preservation of *orf129* was evident in I-12CMS(2), I-12CMS(2) *orf129* was investigated in detail. I-12CMS(2) plants contained three to five times more ORF129 protein than did I-12CMS(3) plants. A single nucleotide substitution, present in the putative promoter region of *orf129,* appeared to be responsible for the differential accumulation of *orf129* transcript. A long N-terminal extension of *atp6* is a common feature of some beet CMSs and is found in I-12CMS(2), but the amino acid sequence is unique. I-12CMS(3) mitochondria, but not I-12CMS(2) mitochondria, were found to be heteroplasmic. This heteroplasmy is characterized by a substoichiometric DNA molecule(s) that has at least two I-12CMS(2)-type mitochondrial loci, suggesting the possibility that the I-12CMS(2) mitochondrial genome might have evolved from such a substoichiometric DNA molecule in I-12CMS(3) mitochondria.

Key words: F_1 hybrid, mitochondrial evolution, mitochondrial gene, nuclear-cytoplasmic interaction, plant mitochondria, substoichiometric DNA

Author contribution: TKi, TM, and TKu designed this study; YO, TA, RY-K, MPY, KK, SE, MM, KT, YK, SY, TS and TKu did the experiments; YO, MPY, KK, SE, TKi, TM, and TKu analyzed the data; TM and TKu wrote the manuscript.

Key message

We characterized a novel male sterility-inducing cytoplasm by genetic and molecular biological approaches. The mitochondrial genome of this novel cytoplasm appears to have evolved from another male sterility-inducing cytoplasm.

Introduction

Cytoplasmic male sterility (CMS) is a maternally inherited character causing pollen abortion (Chen and Liu 2014). CMS occurs in more than 140 plant species (Laser and Lersten 1972). Affecting only male reproductive organs, CMS is used in hybrid seed production in which self pollination of the seed parent must be prevented (Budar et al. 2006). These CMS-based hybrid seed production systems, however, result in the cytoplasmic uniformity of the hybrids if the system relies on a single CMS source. This genetic vulnerability is of special concern (Laughnan and Gabay-Laughnan 1983; Levings 1993). Therefore, exploring novel cytoplasms is an important task for securing CMS-based hybrid seed production systems.

The expression of CMS is genetically explained by male-sterility (MS) inducing cytoplasm and one or several recessive nuclear genes termed restorer of fertility (*rf*) genes (Hanson and Bentolila 2004). Otherwise, CMS is not expressed. For example, the combination of MS-inducing cytoplasm and dominant alleles of restorer-of-fertility genes (*Rf*) results in the fertility restoration of a CMS plant. The *Rf* is, however, unable to restore non-cognate MS-inducing cytoplasm. Hence, one of the methodologies for distinguishing CMS resources uses test crosses to monitor responses to a specific *Rf* (Duvick 1965).

The test cross for cytoplasmic discrimination is, however, laborious. An alternative may be to use a molecular approach to find a characteristic of the MS-inducing cytoplasm. Because CMS is encoded by mitochondria (Schnable and Wise 1998), attempts to find molecular differences between multiple MS-inducing mitochondria have been made. An organizational comparison of multiple mitochondrial genomes of the same species is one such example (L'Homme et al. 1997; Allen et al. 2007; Kawanishi et al. 2010; Fujii et al. 2010).

Mitochondrial genomes of different MS-inducing cytoplasms, as well as of non-MS-inducing cytoplasms exhibit varying degrees of genomic rearrangements (inversions, translocations, and/or insertions/deletions) (Kubo et al. 2011). The mechanisms for how these rearrangements occurred are still a subject of debate (Kubo and Newton 2008; Christensen 2014). One of the unique features in plant mitochondrial genome evolution is the involvement of substoichiometric DNA molecules that are maintained for successive generations. Such DNA molecules are thought to accumulate mutations and may become predominant mitochondrial DNA molecules either spontaneously or from the effect of a particular nuclear gene (Janska et al. 1998).

In sugar beet (*Beta vulgaris* ssp. *vulgaris*), current commercial varieties are hybrids using a single source of CMS, termed CMS-Owen (Owen 1942; Bosemark 2006). This situation represents a significant genetic vulnerability (Bosemark 1979), and other sources of cytoplasm have been explored. I-12CMS(3) is a sugar beet CMS line developed by R. K. Oldemeyer, who introduced MS-inducing cytoplasm from wild beets (*Beta vulgaris* L. ssp. *maritima*) to sugar beets by recurrent backcrossing (Mikami et al. 1985). Molecular studies showed that the I-12CMS(3) cytoplasm differs from CMS-Owen in the electrophoretic

pattern of restriction endonuclease-digested mitochondrial (mt) DNA and the identity of polypeptides synthesized in mitochondria (Mikami et al. 1985; Hallden et al. 1992). An ORF termed *orf129* was identified as the CMS-associated ORF (Yamamoto et al. 2008) in I-12CMS(3) mitochondria. A large part of the primary sequence of *orf129* consists of a sequence of unknown origin that is absent from non-MS-inducing cytoplasms or CMS-Owen (Yamamoto et al. 2008). Translation products of *orf129* (ORF129) were detected in vegetative and reproductive organs as proteins loosely associated with the mitochondrial inner membrane. Transgenic tobacco plants expressing a transgene consisting of *orf129* fused with a mitochondrial transit peptide sequence exhibited male sterility (Yamamoto et al. 2008).

I-12CMS(3) and CMS-Owen are characterized by a long (389 and 387 amino-acid residues, respectively) N-terminal extension of *atp6* (Yamamoto et al. 2005; Yamamoto et al. 2008), whereas the length of that region in non-MS-inducing cytoplasm is only five amino acids. Although the detailed mechanism is unclear about how the *atp6* N-terminal extension is processed, it is likely that once the precursor protein containing both the N-terminal extension and mature ATP6 is translated, then the precursor is cut into two separate proteins (Yamamoto et al. 2005). The translation product of the N-terminal extension of CMS-Owen was detected in vegetative organs as well as in anthers, whereas that of I-12CMS(3) was detected in vegetative organs only (Yamamoto et al. 2005; Yamamoto et al. 2008).

Although the cytoplasm of I-12CMS(3) is derived from a wild Pakistani beet (Mikami et al. 1985), the same cytoplasm, known as CMS-E, is involved in the most common CMS found in European wild beets (Touzet 2012). Additionally, plants with *orf129* have been identified in leaf beet and garden beet accessions (Cheng et al. 2011). Therefore, I-12CMS(3)/CMS-E appears to be widely distributed in *B. vulgaris* in terms of taxonomy and geography.

At one time, I-12CMS(3)/CMS-E and CMS-Owen were thought to have emerged independently from non-MS-inducing cytoplasm, a notion based on the genealogy constructed from nucleotide sequence polymorphisms of plastid DNA fragments or mitochondrial DNA fragments (Fenart et al. 2006; Nishizawa et al. 2007). More recently, however, a mitochondrial genome-wide comparison produced a phylogenetic tree in which I-12CMS(3)/CMS-E and CMS-Owen were grouped into the cluster that contained other beet CMS but none of the non-MS-inducing cytoplasms. This result provided the basis for another hypothesis about the monophyletic origin of beet CMSs (Touzet 2012). The monophyletic origin of beet CMSs would be supported if an example could be found of MS-inducing cytoplasm evolving from one type to another.

We report here that I-12CMS(2), another Oldemeyer's CMS line whose cytoplasmic donor is a wild beet from Turkey (Mikami et al. 1985), has a novel cytoplasm that may have evolved from I-12CMS(3)/CMS-E. The objective of this study was to identify difference(s) between I-12CMS(3) and I-12CMS(2) at the phenotypic level by a test cross and at molecular level by analyzing mitochondrial genomes. These data are important not only for sugar beet breeding but also for obtaining insight into the evolution of CMS in *B. vulgaris*. Interestingly, a substoichiometric DNA molecule containing two

I-12CMS(2)-type loci exists in I-12CMS(3) mitochondria, suggesting that such a DNA molecule might play an important role in the evolution of beet CMS.

Materials and methods

Plant materials and fertility scoring

I-12CMS(R), I-12CMS(2) and I-12CMS(3) possess different sources of cytoplasm, but the nuclear genotype is equivalent to that of a sugar beet inbred maintainer line, I-12 61L (Mikami et al. 1985). I-12CMS(R) carries the CMS-Owen, whereas I-12CMS(2) and I-12CMS(3) have MS cytoplasms derived from wild beets collected in Turkey and Pakistan, respectively. NK-183mm-O (developed by the Hokkaido Agricultural Research Center, National Agriculture and Food Research Organization, Japan) is an inbred maintainer line of CMS-Owen (Moritani et al. 2013), i.e. with normal fertile cytoplasm and is devoid of *Rf* for CMS-Owen. TK-81mm-CMS and TK-81mm-O are a CMS-Owen line and its maintainer line, respectively (Yamamoto et al. 2005). Fertility was determined by the following three criteria: anther color, pollen shedding and pollen stainability. Pollen stainability estimates involved squashing anthers from open flowers in a drop of cotton-blue stain (Hagihara et al. 2005). At least three flowers per plant were examined. Fertile plants bore well-dehiscing, yellow anthers with 20% and more stainable pollen grains. Sterile plants had translucent or brown shrunken anthers. No dehiscence was observed in sterile plants. Partially fertile plants produced less than 20% stainable pollen. Anthers of partially fertile plants were light-yellow or orange in color and usually failed to dehisce. The statistical significance of data was calculated using Fisher's exact test posted at http://aoki2.si.gunma-u.ac.jp/exact/fisher/getpar.html (accessed on 10/Feb/2015).

Nucleic acids isolation

Mitochondrial (mt) DNA and mtRNA were isolated according to the methods of Mikami et al. (1985) and Kubo et al. (2000), respectively. Total cellular DNA was isolated according to the protocol of Doyle and Doyle (1990). Total cellular RNA was isolated from green leaves using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA).

Library construction and chromosome walking

MtDNAs from I-12CMS(2) and I-12CMS(3) were partially digested with *Mbo*I (Takara Bio, Ohtsu, Japan) and fractionated on continuous sucrose density gradients (Sambrook et al. 1989). Fragments of 15 to 20 kbp were collected and ligated into the *Bam*HI-digested lambda DASH vector (Stratagene, La Jolla, CA). The

175 DNA ligation mixture was packaged *in vitro* using Gigapack Gold (Stratagene). A total of 384 recombinant
176 phages were randomly chosen from each of the resulting mtDNA libraries and used for a physical mapping
177 study. DNA fragments located every 5 kbp on the TK-81mm-O (normal cytoplasm) mtDNA map (Kubo et
178 al. 1995), and several known mitochondrial gene sequences were used as probes for plaque hybridization.
179 DNA fragments at the extremities of the clone contigs were used to 'walk' the mitochondrial genome. All
180 overlapping clones selected were tested by hybridization with several mitochondrial gene probes as well as
181 with DNA fragments on the TK-81mm-O mtDNA map (Kubo et al. 1995), to ensure that each clone was a
182 true representation of the mtDNA sequences.
183
184 RNA gel blot hybridization
185
186 MtRNA was fractionated by electrophoresis in a 1.4% agarose gel containing 0.66 M formaldehyde and
187 then transferred to Hybond N+ membrane (GE Healthcare, UK, Amersham Place, England). Specific
188 probes for *orf129*- and *atp9* mRNA detection were PCR amplified from cloned DNA with the primers
189 described in Cheng et al. (2011) and Matsunaga et al. (2011), respectively. Hybridization probes were
190 labeled with ^{32}P using a Megaprime DNA labeling kit (GE Healthcare) or with alkaline phosphatase using
191 the AlkPhos Direct DNA labeling system (GE Healthcare).
192
193 DNA sequencing
194
195 The DNA fragments of interest were cloned into the pBluescript vector (Stratagene) and then sequenced
196 using a Li-COR4000L DNA sequencer (Li-COR, Lincoln, NE) or an ABI3130 genetic analyzer (Applied
197 Biosystems, Foster City, CA). Nucleotide sequences were assembled using Sequencher 4.0 (Gene Codes
198 Corporation, Ann Arbor, MI). A homology search (BLAST) was conducted on the NCBI web site
199 (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequence data were deposited with DDBJ/EMBL/GenBank under
200 accession nos. LC032133 (I-12CMS(2) *atp1-rrn26*), AB355937 (I-12CMS(3) *orf129*), LC032134
201 (I-12CMS(2) *atp6*), and AB490412 (I-12CMS(2) *orf129*).
202
203 Polymerase chain reaction (PCR)
204
205 Primers used in this study are listed in Table S1. For quantitative reverse-transcription (qRT) PCR, RNA
206 samples (4 μg) were treated with RNase-free DNase I and reverse transcribed with SuperScript III
207 First-Strand Synthesis System for RT-PCR (Life Technologies, Carlsbad, CA) using random hexamer
208 primers. Reagents for qRT-PCR contain SYBR GreenER qPCR SuperMix Universal (Life Technologies)
209 and primers (0.2 μM each). PCR was monitored using a Chromo 4, Opticon Monitor (ver 3.1) with a

DNA-Engine PTC-200 (Bio-Rad Laboratories, Hercules, CA). The PCR protocol was 95˚C for 10 min, then 40 cycles of 95˚C for 15 sec and 60.8˚C for 1min. A series of dilutions of I-12CMS(2) cDNA were used as the standard samples for quantification. Using this series, qRT-PCR generated a standard curve (X axis, factor of dilution; Y axis, threshold cycle number) for mRNAs of *orf129* (primers orf129-Fw and orf129-Rv) and an actin gene (*Act*) (primers actin-Fw and actin-Rv). The ratio of *orf129*:*Act* in each sample was calculated based on the results of qRT-PCR analysis (repeated three times for each sample). Student's *t*-test was done using Microsoft Excel (Microsoft Office 2008 for Mac; Microsoft Japan, Tokyo, Japan). For circularized RNA RT-PCR (Kuhn and Binder 2002), DNase I-treated RNA samples were treated with tobacco acid pyrophosphatase as described in Muller and Storchova (2013). The RNA was circularized as described previously (Matsunaga et al. 2011). Complementary DNA was synthesized with SuperScript III First-Strand Synthesis System for RT-PCR and primer orf129-162R. PCR amplification of *orf129* cDNA used primers orf129-54R and orf129-339F (protocol: 94˚C, 3min; 35 cycles of 94˚C, 30sec, 55˚C, 30sec, and 72˚C, 1min; 72˚C, 4min). PCR products were cloned using a TOPO TA Cloning kit (Life Technologies). For PCR amplification of genomic DNA, about 10 ng of total cellular DNA was mixed with 10 μl of 1 x Quick Taq HS DyeMix (Toyobo, Osaka, Japan) or 1 x GoTaq Green Master Mix (Promega, Madison, WI). For PCR amplification of PCR products, the template PCR products were treated with Exonuclease I (Takara Bio) (1 unit, 37˚C, 1h) to degrade residual primers before the PCR amplification. PCR protocols are described in the figure legends. PCR products were electrophoresed in 2% agarose gels. Gels were stained with ethidium bromide solution (Sambrook et al. 1989).

Immunoblotting

Anti-ORF129 antiserum was raised against the carboxyl half of ORF129 (D50-Q129) as described previously (Yamamoto et al. 2008). Anti-preSATP6 was raised against an oligopeptide corresponding to the C-terminal sequence of preSATP6 (Yamamoto et al. 2005). Total cellular proteins were isolated from green leaves as described in Cheng et al. (2009). Proteins were separated by SDS-PAGE and electroblotted onto Hybond ECL membrane (GE Healthcare). The concentration of primary antisera used for immunoblots was 42.5 ng/mL. Signal bands were detected as described previously (Yamamoto et al. 2005, 2008).

Results

I-12CMS(2), I-12CMS(3) and I-12CMS(R) differ in their ability to restore fertility

Three CMS lines, I-12CMS(2), I-12CMS(3) and I-12CMS(R) (CMS-Owen), all of which bear the same

nuclear background, were crossed individually with three plants (#1 to #3) randomly chosen from NK-183mm-O, a line that is devoid of the *Rf* for CMS-Owen but male fertile owing to a normal fertile cytoplasm (i.e. a maintainer for CMS-Owen). As shown in Table 1, NK-183mm-O served as a perfect sterility-maintainer when used as the pollen parent on I-12CMS(R). This same pollen parent line, however, failed to maintain male sterility in most of its hybrids with either I-12CMS(2) or I-12CMS(3) (Table 1). Interestingly, fertility in the F_1 hybrids carrying the I-12CMS(3) cytoplasm apparently exceeded that of the F_1s with the I-12CMS(2) cytoplasm (Fisher's exact test: p=1.5 x 10^{-23} for NK-183mm-O#1, p=5.1 x 10^{-12} for NK-183mm-O#2, and p=2.2 x 10^{-17} for NK-183mm-O#3). The cross of I-12CMS(3) x NK-183mm-O yielded fully fertile and partially fertile progeny, whereas the F_1 progeny of I-12CMS(2) with NK-183mm-O tended to group around the partially fertile class (0-20% stainable pollen with cotton-blue stain).

A fraction of the F_1 hybrids (e.g. five F_1s of I-12CMS(2) x NK-183mm-O#2 and four F_1s of I-12CMS(3) x NK-183mm-O#2) had white shriveled anthers and were classified as fully sterile. Because the frequency of the fully sterile plants in F_1 progeny appeared to be different depending on the pollen parents (i.e. NK-183mm-O#1 to NK-183mm-O#3), we tested a null hypothesis that fully sterile F_1 plants occurred at the same frequency irrespective of the pollen parents for each of the I-12CMS(2) and I-12CMS(3) by Fisher's exact test. The results rejected the null hypothesis [p=0.0007 for I-12CMS(2) and p=5.6 x 10^{-5} for I-12CMS(3)]. A possible explanation for these results is that, although NK-183mm-O has *Rf*s for I-12CMS(2), as well as for I-12CMS(3), the genetic composition of these *Rf*s differs among the pollen parents. For example, some of the *Rf*s were heterozygous in one pollen parent but homozygous in the others. Also, it is possible that fertility restoration conditioned by NK-183mm-O is highly sensitive to the environment, hence unstable, although the restoration was more efficient for I-12CMS(3).

Organizational differences between the I-12CMS(2) and I-12CMS(3) mitochondrial genomes

To gain insight into the molecular basis of cytoplasmic differences, we focused our analysis on the organizational differences between the I-12CMS(2) and I-12CMS(3) mitochondrial genomes. Chromosome walking using phage-cloned ~18 kbp DNA fragments that were generated by random shearing of I-12CMS(3) mitochondrial (mt) DNA predicted the existence of two circular chromosomes of 274 kbp and 234 kbp, sharing a 103 kbp sequence (Figs. 1 and S1). The I-12CMS(3) genome can also be interpreted as a 508 kbp master chromosome with 103 kbp repeated sequences (Figs. 1 and S1). We also identified two families of repeated sequences that are involved in inter- and/or intra-molecular recombination (shown as R1 [6.4 kbp] and R2 [5.7 kbp] in Figs. 1 and S1). Hence, the molecular organization of the mitochondrial genome of I-12CMS(3) appears to be very complex.

During our study, Darracq et al. (2011) reported the shotgun-sequencing-based prediction of the mitochondrial genome organization of CMS-E that is probably equivalent to the I-12CMS(3) cytoplasm. Although their model, one circular molecule (269 kbp) and one linear molecule (110 kbp), is different from ours, a comparison of gene order and restriction site arrangement indicated that the net sequence complexity is nearly the same between the two models. Differences in the configuration are due to the introduction of repeated sequences and the arrangement of unique sequences between repeats. It is likely that differences in the mapping strategy (chromosome walking using ~18 kbp-clone library vs. shotgun sequencing using 5 kbp-clone library) are associated with the different results.

An I-12CMS(2) mtDNA physical map was also constructed by identifying a set of overlapping I-12CMS(2) phage clones as was done for the I-12CMS(3) genome. The resultant map was identical to that of I-12CMS(3) except for missing a 13-kbp segment from a region between *atp1* and *rrn26* of I-12CMS(2) (shown as the stippled region in Fig. 1). In our physical maps, one of the repeated sequences named R1 partially overlapped with the 13-kbp region in I-12CMS(3), and this R1 copy is truncated in I-12CMS(2) (Fig. 1); however, another intact R1 copy is preserved in I-12CMS(2).

Inspection of the restriction site map confirmed the conserved organization of the *atp1-rrn26* region between I-12CMS(2) and TK-81mm-O. The latter is a normal, fertile sugar beet line whose entire mitochondrial DNA sequence is known (Kubo et al. 2000). Comparison of the *atp1-rrn26* regions between CMS-E (DDBJ/EMBL/GenBank accession number FQ014226) and TK-81mm-O (BA000009) revealed a 12860-bp sequence that is absent from the TK-81mm-O *atp1-rrn26* region (Fig. 2). In TK-81mm-O, the 12860-bp sequence is replaced with a 561-bp sequence that has little homology to the 12860-bp sequence. We confirmed that the 561-bp sequence is conserved in I-12CMS(2) by nucleotide sequencing (see below). No known mitochondrial genes were found in the 12860-bp sequence or the 561-bp sequence. Eight ORFs that can specify 105-265 amino acid residues are annotated in the 12860-bp sequence (FQ014226), but there was not an ORF identified in the 561-bp sequence (BA000009).

A larger amount of ORF129 protein is present in I-12CMS(2) than in I-12CMS(3)

We sought other molecular differences potentially associated with the cytoplasmic difference. We focused our analysis on *orf129*, an MS-associated mitochondrial gene in I-12CMS(3) (Yamamoto et al. 2008). Our physical mapping data indicated the preservation of *orf129* in I-12CMS(2) (Fig. 1). In order to confirm the presence of ORF129 polypeptide, immunoblot analysis using ORF129-specific antiserum (Yamamoto et al. 2008) was done. Total cellular proteins were isolated from leaves of I-12CMS(R), I-12CMS(2), and I-12CMS(3). ORF129 was detected from the preparations of I-12CMS(2) and I-12CMS(3), whereas no detectable signal was seen in I-12CMS(R) (Fig. 3a). What does merit attention is that the signal was more

intense in I-12CMS(2) than in I-12CMS(3) (Fig. 3a). Dilutions of the total protein preparations from I-12CMS(2) and I-12CMS(3) were used to compare the relative amounts of ORF129. Relative signal intensities obtained by immunoblot analysis indicated that I-12CMS(2) contains approximately three to five times more ORF129 than does I-12CMS(3) (Fig. 3b).

RNA gel blot analysis of total mitochondrial RNA isolated from leaves showed a stronger *orf129* mRNA signal for a 1.1-kb band in I-12CMS(2) than that of I-12CMS(3) (Fig. 3c). On the other hand, the smaller size transcripts are fairly more intense in I-12CMS(3). Accumulation of *orf129* transcripts was compared by quantitative real time RT-PCR (qRT-PCR). Total cellular RNA of leaves was isolated from two I-12CMS(3) plants and two I-12CMS(2) plants. Each of the four samples was subjected to qRT-PCR in which a nuclear actin gene (*Act*) was used as the reference gene. The results showed that accumulation of I-12CMS(2) *orf129* mRNA is about 2.7 times that of I-12CMS(3) *orf129* mRNA, a statistically significant difference (Table 2).

No apparent difference at the *orf129* locus was seen on the physical maps between I-12CMS(3) and I-12CMS(2), but it is possible that the difference in *orf129*-mRNA accumulation is associated with a genomic alteration that is difficult to detect by physical mapping. We determined the nucleotide sequence of a 2.2-kbp segment (1.1-kbp *Bam*HI-*Xho*I and 1.1-kbp *Xho*I sub-fragments of a phage clone) of the I-12CMS(2) mtDNA containing *orf129* (Fig. S2). The nucleotide sequence of the corresponding region in I-12CMS(3) was also determined from subclones of a phage clone. The I-12CMS(3) sequence was identical to that of CMS-E.

The sequences of the *orf129* coding region (387 bp in length) from the two beet lines, I-12CMS(2) and I-12CMS(3), were perfectly identical and, thus, can encode identical proteins. The sequence identity extended downstream for at least 714 bp, from nucleotide +388 (the first nucleotide residue of the *orf129* initiation codon is defined as +1) to the 3'-end of the sequenced region (nucleotide +1104). The 5' flanking regions (from the 5'-end nucleotide -1138 of the sequenced region to nucleotide -1) of the two genes were virtually identical. The only difference was a single nucleotide substitution identified at nucleotide -490; a G residue in the I-12CMS(3) *orf129* versus a T in the I-12CMS(2) *orf129* (Fig. S2).

To find an association between the G-T polymorphism and *orf129* expression, cDNAs containing transcript termini of *orf129* mRNA were PCR amplified by circularized RNA reverse transcription PCR (CR-RT-PCR), in which RNAs were reverse transcribed after their 5' and 3' ends were ligated by T4 RNA ligase (Kuhn and Binder 2002). The resultant cDNAs were PCR amplified, cloned into plasmid vectors and sequenced. Some of our cDNA clones have additional nucleotides that were not encoded in genomic DNA (Table S2). These additional nucleotides appeared to be inserted between transcript termini in the cDNA. It is likely that post-transcriptional modification, such as polyadenylation (Lange et al. 2009), occurred in some *orf129* mRNAs. The mapped termini are detailed in Fig. S2 and Table S2, and are summarized in Fig. 4a.

The mapped 3' termini were almost always confined within the same ~12 bp region in I-12CMS(3) and I-12CMS(2) (Fig. 4a and Table S2). On the other hand, the 5' termini were mapped to the region from -479 to -114. In I-12CMS(2), eight of the eleven cDNA clones were derived from transcripts whose 5' termini were mapped within a region of -479 to -472 (Table S2). Considering the size of these transcripts deduced from mapped transcript termini (~1050 bases), they likely represent the 1.1 kb-signal band on the RNA gel blot (Fig. 3c). In I-12CMS(3), the number of cDNA clones representing 1.1-kb signal band is three out of eleven (Table S2). The other cDNA clones enabled us to map the 5' termini from -278 to -153 (Table S2 and Fig. 4a), suggesting that they represent the smaller size transcripts on the RNA gel blot (Fig. 3c).

The G-T polymorphism at -490 was close to the 5' termini of the 1.1-kb transcripts (Fig. 4b). The DNA sequences surrounding these termini are similar to the promoter consensus sequence (5'-CRTAAGAGA-3' or 5'-CGTATATAA-3') that has been proposed for mitochondrial genes in dicot plants (Gagliardi and Binder, 2007) (Fig. 4b). This observation suggests that the 1.1-kb *orf129* mRNA is a primary transcript. The -490 polymorphic site is located only 5 bp upstream of this promoter consensus sequence.

Differences in the amino acid sequence of the N-terminal extension of *atp6*

In I-12CMS(3)/CMS-E mitochondria, there is another expressed ORF that is absent from normal fertile mitochondria. This ORF is fused with *atp6* as an N-terminal extension (Yamamoto et al. 2008; Darracq et al. 2011). On our physical map, no apparent structural alteration is seen between the *atp6* loci of I-12CMS(2) and I-12CMS(3) (Fig. 1), indicating that I-12CMS(2) *atp6* also has an homologous N-terminal extension. To compare the nucleotide sequence of *atp6*, we determined the sequence of I-12CMS(2) *atp6* that is on a subclone of a phage clone (Fig. S3).

I-12CMS(2) *atp6* is 1956 bp in length with the potential to encode a 652 amino acid residue polypeptide. We identified the N-terminal extension of I-12CMS(2) ATP6 polypeptide by comparison with I-12CMS(3) ATP6 (Fig. S4). In both I-12CMS(2) and I-12CMS(3) ATP6, the length of the N-terminal extension is 389 amino acid residues. Sequence homology between the two N-terminal extensions is 94.1%. Besides the N-terminal extension, I-12CMS(2) *atp6* and I-12CMS(3) *atp6* are identical at the nucleotide and amino acid sequence levels.

To see the expression of the N-terminal extension of I-12CMS(2) *atp6*, immunoblot analysis was done using antiserum against preSATP6, an N-terminal extension of *atp6* in CMS-Owen mitochondria (Yamamoto et al. 2005). Because the amino acid sequence of the antigen used for raising anti-preSATP6 antiserum is conserved in I-12CMS(2) ATP6 (Fig. S4), and because anti-preSATP6 antiserum reacted with I-12CMS(3) protein (Yamamoto et al. 2008), this antiserum is expected to react with the translation product of I-12CMS(2) N-terminal extension, if any. As shown in Fig. 5, we detected signal bands of low intensity

385 from total cellular proteins of I-12CMS(2) and I-12CMS(3) leaves. The apparent molecular mass of the two
386 signal bands was indistinguishable.
387
388
389 A substoichiometric mitochondrial DNA molecule in I-12CMS(3)
390
391 Our next task was to develop a molecular marker that enables one to discriminate I-12CMS(2) and
392 I-12CMS(3) cytoplasms. We thought that the *atp1-rrn26* region was a suitable target for designing
393 oligonucleotide primers that could be used for specific PCR amplification of I-12CMS(2) or I-12CMS(3)
394 mtDNA. Two primers were designed that correspond to specific regions of I-12CMS(2) or I-12CMS(3)
395 (Fig. 6a). The I-12CMS(2)-specific primer has no significant homology to nucleotide sequences of
396 FQ014226 or FQ014231, entries that constitute the complete mtDNA sequence of CMS-E. A common
397 primer was designed outside of the polymorphic region (Fig. 6a).
398 PCR amplification targeting the I-12CMS(3)-specific region gave us amplicons of the expected size of
399 0.6 kbp from two I-12CMS(3) plants (Fig. 6b). Nucleotide sequences of these amplicons matched with
400 those from CMS-E mitochondria. No amplification was observed from four I-12CMS(2) plants (Fig. 6b),
401 indicating that the I-12CMS(3)-specific sequence is absent from I-12CMS(2) mitochondria.
402 We tested another primer combination that is expected to PCR amplify the I-12CMS(2)-specific region.
403 From four I-12CMS(2) plants, PCR products of the expected size were obtained (Fig. 6b). Nucleotide
404 sequences of these amplicons were matched with TK-81mm-O mitochondria, confirming the
405 abovementioned notion that the I-12CMS(2)-specific region is conserved in TK-81mm-O. On the other
406 hand, PCR products of the same size were amplified from two I-12CMS(3) plants (Fig. 6b). The nucleotide
407 sequences of these PCR products matched perfectly with that of the I-12CMS(2) PCR products. We think
408 that contamination of reagents or utensils with I-12CMS(2) genomic DNA [or the I-12CMS(2) DNA clone]
409 is unlikely as indicated by the control experiment (lane 7 in Fig. 6b). Each of the DNA samples shown in
410 Fig. 6b was isolated from a piece of green leaf excised with a clean razor, making contamination of DNA
411 samples unlikely. Because of the absence of the I-12CMS(2)-specific region from the physical map of
412 I-12CMS(3) (or from FQ014226 or FQ014231), from the draft sequence of the sugar beet nuclear genome
413 (Dohm et al. 2014) or the plastid genome sequence (DDBJ/EMBL/GenBank accession number EF534108),
414 a possible explanation is heteroplasmy regarding the *atp1-rrn26* region in I-12CMS(3) in which a
415 I-12CMS(2)-like DNA molecule exists in substoichiometric levels.
416 The heteroplasmic state of the *orf129* locus was tested. We designed other primer sets for PCR that can
417 detect single nucleotide polymorphisms (SNPs) in the upstream region of *orf129* (Fig. 7a). Specific primers
418 have G or T residues at their 3' ends that correspond to the -490 SNP of *orf129*, by which PCR amplification
419 proceeds only when the 3' end of the primer is complementary to the template DNA (mutant allele-specific

PCR amplification, MASA). A common primer corresponds to the upstream region of *orf129* that is conserved between I-12CMS(2) and I-12CMS(3). As shown in Fig. 7b, PCR analysis using the common and I-12CMS(3) specific primers detected the expected signal bands from I-12CMS(3) samples but not from I-12CMS(2). On the other hand, when the I-12CMS(2) specific primer was used, we observed signal bands from I-12CMS(2) samples as well as I-12CMS(3) samples, suggesting heteroplasmy in I-12CMS(3), but PCR amplification from I-12CMS(3) *orf129* with mismatch primers could also be possible. If the observed amplification from I-12CMS(3) samples was due to the mismatch-primer PCR, the I-12CMS(2) specific primer could generate the same PCR product from I-12CMS(3) *orf129*. To test this possibility, PCR products of the upstream region of I-12CMS(3) *orf129* were subjected to PCR amplification with the I-12CMS(2)-specific primers, and we observed no amplification (lane 7 in Fig. 7b).

Finally, we tested the heteroplasmic state of the *atp6* locus. Of the several nucleotide substitutions and indels, SNP in +177 was selected as a target for MASA (Fig. 8a). Specific primers have C or A residues at their 3' ends that correspond to the +177 SNP of *atp6*. In this MASA, we saw no trace of heteroplasmy in the *atp6* loci of I-12CMS(2) or I-12CMS(3) (Fig. 8b).

Discussion

In our test cross presented here, fully restored F_1 plants from I-12CMS(3) and partially restored F_1 plants from I-12CMS(2) seed parents emerged. Because the pollen parent, NK-183mm-O, is devoid of the *Rf* for CMS-Owen, the I-12CMS(3) and I-12CMS(2) CMSs are considered as different cytoplasms from CMS-Owen at the genetic level. Darracq et al. (2011) reported the differences in mitochondrial genome organization between CMS-E and CMS-Owen, a result that included our data. We also observed differential fertility restoration reactions between I-12CMS(3) and I-12CMS(2) CMSs. I-12CMS(2) apparently tended to resist fertility restoration more than I-12CMS(3) when crossed with the pollen parent NK-183mm-O. We assume that I-12CMS(3) and I-12CMS(2) cytoplasms may differentially respond to the same *Rf*(s) even if the molecular data indicate a close relationship between these two cytoplasms. Mitochondrial genome organization is very similar between I-12CMS(3) and I-12CMS(2); both I-12CMS(3) and I-12CMS(2) possess *orf129*, whereas *orf129* is absent from CMS-Owen (Yamamoto et al. 2008); and both I-12CMS(3) and I-12CMS(2) possess a long N-terminal extension of *atp6*. An *Rf* can differentially restore pollen fertility to plants with different sources of CMS that share some features at the molecular level. For example, in rice, CMS-BT and CMS-LD are considered to be different CMSs, but both are associated with mitochondrial gene *orf79* (Chen and Liu 2014). Plants with these two CMSs can restore male fertility by *Rf2*, but the degree of restored fertility is different between CMS-BT and CMS-LD plants (Itabashi et al. 2009).

The observed difference in the degree of fertility restoration between I-12CMS(3) and I-12CMS(2) can be discussed in relationship to the molecular data. One of the molecular differences is seen in the *atp1-rrn26* region; however, we think it is unlikely that this difference affects MS expression for the following reasons: the altered region in *atp1-rrn26* contains no known mitochondrial gene in I-12CMS(3) or I-12CMS(2) and no function is known for the eight ORFs in the 12860-bp region in I-12CMS(3); and mitochondrial genes close to the altered region are *atp1* and *rrn26*, but they are 6.3 kbp and 2.0 kbp away from the altered region, respectively, which appears to be too remote to affect plant mitochondrial gene expression. Moreover, genetic association of the I-12CMS(2)- or the I-12CMS(3)-type *atp1-rrn26* with MS is improbable, because the former is seen in the normal, fertile line TK-81mm-O, and we found that the I-12CMS(3)-type *atp1-rrn26* is very similar to that of the mitochondrial genome organization of wild beet accession BGRC56777 that has another non-MS-inducing cytoplasm (Nishizawa et al. 2007).

On the other hand, the differential fertility restoration between I-12CMS(3) and I-12CMS(2) may be associated with quantitative differences in the accumulation of ORF129 polypeptide. ORF129 polypeptide is more abundant in I-12CMS(2), whose fertility restoration by NK-183mm-O is insufficient. This is an interesting correlation that should be analyzed in the future.

The relative abundance of ORF129 protein correlates with the difference in steady-state *orf129* transcript levels. In this connection, of particular interest is the single nucleotide substitution near the 5' terminus of the *orf129* transcript. We hypothesize that this substitution may affect the expression of *orf129*. Mitochondrial promoters have been analyzed in several plant species by identifying transcription initiation sites and aligning the surrounding sequences, thereby revealing conserved promoter motifs (Gagliardi and Binder 2007). In *Arabidopsis thaliana,* for example, half of the transcriptional starts identified are located within the conserved nonanucleotide core motif (CRTAAGAGA or CGTATATAA; transcription initiation nucleotide underlined) (Gagliardi and Binder 2007). Purines appear immediately downstream from the transcription initiation site. Preceding the core motif is usually a sequence rich in A and T nucleotides (AT-box) that has been proven essential for the full function of some mitochondrial promoters *in vitro* (Gagliardi and Binder 2007). DNA sequences around the 5' terminus of the 1.1-kb *orf129* transcripts were found to be very similar to the consensus promoter motif. The single base change mentioned above occurred within the putative AT-box. The promoter of the I-12CMS(2) *orf129* contained a conserved T nucleotide at the position in question, whereas in the I-12CMS(3) *orf129* promoter, a T was replaced by a G. It is probable that this T to G transversion exerted a substantial negative influence on the activity of the I-12CMS(3) *orf129* promoter, given that the AT-box is actually required for full-promoter function.

In both I-12CMS(3) and I-12CMS(2), a long N-terminal extension of *atp6* is expressed and the protein product is detected in leaves. However, because the protein product is undetectable in anthers of I-12CMS(3) (Yamamoto et al. 2008), there are no data associating the *atp6* loci with MS expression of I-12CMS(3). On the other hand, the N-terminal extension of CMS-Owen accumulated in anthers as a

membrane protein; hence, it may be associated with MS expression (Yamamoto et al. 2005). If very small amounts of the N-terminal extension of *atp6* can reduce the male fertility of I-12CMS(3) and I-12CMS(2), differences in the amino acid sequences of the N-terminal extension may be associated with differential fertility restoration, but this proposal needs further study.

The long N-terminal extension is a characteristic of I-12CMS(2), I-12CMS(3) and CMS-Owen (Onodera et al. 1999; Yamamoto et al. 2005; this study). The amino acid sequences of the N-terminal extensions are similar among I-12CMS(2), I-12CMS(3) and CMS-Owen, suggesting a monophyletic origin of these *atp6* genes. This result is consistent with the notion of Darracq et al. (2011), who proposed a monophyletic origin of MS-inducing mitochondria in beet.

We found that I-12CMS(3) is heteroplasmic at two mitochondrial loci (*atp1-rrn26* and *orf129*), a finding that provides insight into the evolutionary relationship between I-12CMS(3) and I-12CMS(2) mitochondria. In I-12CMS(3) mitochondria, we postulate the presence of a substoichiometric DNA molecule that has the 561-bp sequence at *atp1-rrn26* and a T residue at -490 in *orf129* (Fig. 9). This substoichiometric molecule resembles the I-12CMS(2) mitochondrial genome but differs at least at nucleotide +177 in *atp6*. On the other hand, the genomes of I-12CMS(2) mitochondria are homoplasmic having the 561-bp sequence at *atp1-rrn26*, a T residue at nucleotide -490 of *orf129* and an A residue at nucleotide +177 of *atp6*. Assuming an intermediate that has heteroplasmic mitochondria similar to I-12CMS(3) but whose substoichiometric DNA molecule encodes A at nucleotide +177 in *atp6* (Fig. 9), we propose a model to explain the evolution of I-12CMS(3) mitochondria to I-12CMS(2) mitochondria. First, the substoichiometric DNA molecule in the intermediate mitochondria evolved to have A residues at nucleotide +177 in *atp6*. In this regard, Small et al. (1987) proposed that mtDNA molecules accumulating mutations could be maintained as substoichiometric DNA molecules that play an important role in mitochondrial genome evolution. In the next step, the resultant substoichiometric DNA molecule predominated in mitochondria in a manner similar to that shown in common bean (Janska et al. 1998), whereas the other DNA molecule was lost, resulting in I-12CMS(2) in Fig. 9. We cannot exclude the possibility that evolution occurred in the opposite direction (i.e. from the right to the left in Fig. 9). However, evidence would be needed for a preexisting substoichiometric molecule that has the 12860-bp sequence at *atp1-rrn26*, and none was detected in this study. Such a substoichiometric DNA molecule may be found in other beet accessions.

We should point out the possibility that there might be an unidentified substoichiometric DNA molecule in I-12CMS(3) or other CMS beets, and such DNA molecules could serve as materials to evolve other MS-inducing mitochondria. Hence, variation in beet CMS might be wider than ever thought.

Acknowledgements

Part of this work was conducted at the Center for Genome Dynamics, Faculty of Science, Hokkaido

University. This work was supported in part by JSPS KAKENHI Grant Number 25292001.

References

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, et al. (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177: 1173-1192

Bosemark NO (1979) Genetic poverty of the sugarbeet in Europe. In: A.C. Zeven & A.M. Van Harten (Eds.). Proc Conf Broadening Genet Base Crops, Wageningen, the Netherlands. Pudoc Wageningen, pp. 29-35

Bosemark NO (2006) Genetics and Breeding. In: Draycott AP (ed) Sugar Beet. Blackwell Publishing, Oxford, UK, pp 50-88

Budar F, Touzet P, Pelletier G (2006) Cytoplasmic male sterility. In: Ainsworth C (ed) Flowering and its manipulation. Blackwell, Oxford, pp 147-180

Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65: 579-606

Cheng D, Kitazaki K, Xu D, Mikami T, Kubo T (2009) The distribution of normal and male-sterile cytoplasms in Chinese sugar-beet germplasm. Euphytica 165: 345-351

Cheng D, Yoshida Y, Kitazaki K, Negoro S, Takahashi H, et al. (2011) Mitochondrial genome diversity in *Beta vulgaris* L. ssp. *vulgaris* (Leaf and Garden Beet Groups) and its implications concerning the dissemination of the crop. Genet Res Crop Evol 58: 553-560

Christensen CA (2014) Genes and junk in plant mitochondria--repair mechanisms and selection. Genome Biol Evol 6: 1448-1453

Darracq A, Varre JS, Marechal-Drouard L, Courseaux A, Castric V, et al. (2011) Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. Genome Biol Evol 3: 723-736

Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, et al. (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). Nature 505: 546-549

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12: 13-15

Duvick DN (1965) Cytoplasmic pollen sterility in corn. Adv Genet 13: 1-56

Fenart S, Touzet P, Armaud J-F, Cuguen J (2006) Emergence of gynodioecy in wild beet (*Beta vulgaris maritima* L.): a genealogical approach using chloroplastic nucleotide sequences. Proc Royal Soc London B 273: 1391-1398

Fujii S, Kazama T, Yamada M, Toriyama K (2010) Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. BMC Genomics 11: 209

Gagliardi D, Binder S (2007) Expression of the plant mitochondrial genome. In: Logan DC (ed) Plant Mitochondria, Blackwell, Oxford, pp 50-96

560 Hagihara E, Itchoda N, Habu Y, Iida S, Mikami T, et al. (2005) Molecular mapping of a fertility restorer
561 gene for Owen cytoplasmic male sterility in sugar beet. Theor Appl Genet 111: 250-255
562 Hallden C, Lind C, Moller IM (1992) Variation in mitochondrial translation products in fertile and
563 cytoplasmic male-sterile sugar beets. Theor Appl Genet 85: 139-145
564 Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male
565 gametophyte development. Plant Cell 16: S154-S169
566 Itabashi E, Kazama T, Toriyama K (2009) Characterization of cytoplasmic male sterility of rice with Lead
567 Rice cytoplasm in comparison with that with Chinsurah Boro II cytoplasm. Plant Cell Rep 28:
568 233-239
569 Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA (1998) Stoichiometric shifts in the
570 common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility.
571 Plant Cell 10: 1163-1180
572 Kawanishi Y, Shinada H, Matsunaga M, Masaki Y, Mikami T, et al. (2010) A new source of cytoplasmic
573 male sterility found in wild beet and its relationship to other CMS types. Genome 53: 251-256
574 Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T (2011) Male sterility-inducing mitochondrial
575 genomes: how do they differ? Crit Rev Plant Sci 30: 378-400
576 Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion 8: 5-14
577 Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, et al. (2000) The complete nucleotide sequence of
578 the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for $tRNA^{Cys}$(GCA).
579 Nucleic Acids Res 28: 2571-2576
580 Kubo T, Satoh Y, Muro T, Kinoshita T, Mikami T (1995) Physical and gene organization of mitochondrial
581 DNA from the fertile cytoplasm of sugarbeet (*Beta vulgaris* L.). Curr Genet 28: 235-241
582 Kuhn J, Binder S (2002) RT-PCR analysis of 5' and 3'-end-ligated mRNAs identifies the extremities of *cox2*
583 transcripts in pea mitochondria. Nucleic Acids Res 30: 439-446
584 Lange H, Sement FM, Canaday J, Gagliardi D (2009) Polyadenylation-assisted RNA degradation processes
585 in plants. Trends Plant Sci 14: 497-504
586 Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile
587 angiosperms. Bot Rev 3: 425-454
588 Laughnan JR, Gabay-Laughnan S (1983) Cytoplasmic male sterility in maize. Ann Rev Genet 17: 27-48
589 Levings CS III (1993) Thoughts on cytoplasmic male sterility in *cms-T* maize. Plant Cell 5: 1285-1290
590 L'Homme Y, Stahl RJ, Li X-Q, Hameed A, Brown GG (1997) *Brassica nap* cytoplasmic male sterility is
591 associated with expression of a mtDNA region containing a chimeric gene similar to the *pol*
592 CMS-associated *orf224* gene. Curr Genet 31: 325-335
593 Matsunaga M, Nagano H, Mikami T, Kubo T (2011) Large 3' UTR of sugar beet *rps3* is truncated in
594 cytoplasmic male-sterile mitochondria. Plant Cell Rep 30: 231-238

595 Mikami T, Kishima Y, Sugiura M, Kinoshita T (1985) Organelle genome diversity in sugar-beet with
596 normal and different sources of male sterile cytoplasms. Theor Appl Genet 71:166-171
597 Moritani M, Taguchi K, Kitazaki K, Matsuhira H, Katsuyama K, et al. (2013) Identification of the
598 predominant nonrestoring allele for Owen-type cytoplasmic male sterility in sugar beet (*Beta vulgaris*
599 L.): development of molecular markers for the maintainer genotype. Mol Breed 32: 91-100
600 Muller K, Storchova H (2013) Transcription of *atp1* is influenced by both genomic configuration and
601 nuclear background in the highly rearranged mitochondrial genomes of *Silene vulgaris*. Plant Mol
602 Biol 81: 495-505
603 Nishizawa S, Mikami T, Kubo T (2007) Mitochondrial DNA phylogeny of cultivated and wild beets:
604 relationships among cytoplasmic male-sterility-inducing and nonsterilizing cytoplasms. Genetics
605 177: 1703-1712
606 Onodera Y, Yamamoto MP, Kubo T, Mikami T (1999) Heterogeneity of the *atp6* presequences in normal
607 and different sources of male-sterile cytoplasms of sugar beet. J Plant Physiol 155: 656-660
608 Owen FV (1942) Male sterility in sugar beets produced by complementary effects of cytoplasmic and
609 mendelian inheritance. Am J Bot 29:692
610 Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, Second edition. Cold
611 Spring Harbor Laboratory Press, Cold Spring Harbor, New York
612 Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration.
613 Trends Plant Sci 3: 175-180
614 Small ID, Isaac PG, Leaver CJ (1987) Stoichiometric differences in DNA molecules containing the *atpA*
615 gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. EMBO J 6:
616 865-869
617 Touzet P (2012) Mitochondrial genome evolution and gynodioecy. In: L. Marechal-Drouard (ed.).
618 Mitochondrial Genome Evolution, Academic Press, Oxford, UK, pp. 71-98
619 Yamamoto MP, Kubo T, Mikami T (2005) The 5'-leader sequence of sugar beet mitochondrial *atp6*
620 encodes a novel polypeptide that is characteristic of Owen cytoplasmic male sterility. Mol Genet
621 Genomics 273: 342-349
622 Yamamoto MP, Shinada H, Onodera Y, Komaki C, Mikami T, et al. (2008) A male sterility-associated
623 mitochondrial protein in wild beets causes pollen disruption in transgenic plants. Plant J 54:
624 1027-1036
625
626

627 **Table 1** Male fertility in the F_1 progeny of three CMS sources with a common pollen parent

Cross combination	Number of plants			
	Fertile	Partial fertile	Sterile	Total
I-12CMS(R) x NK-183mm-O#1	0	0	34	34
I-12CMS(2) x NK-183mm-O#1	0	30	8	38
I-12CMS(3) x NK-183mm-O#1	49	1	1	51
I-12CMS(R) x NK-183mm-O#2	0	0	89	89
I-12CMS(2) x NK-183mm-O#2	0	57	5	62
I-12CMS(3) x NK-183mm-O#2	22	15	4	41
I-12CMS(R) x NK-183mm-O#3	0	0	19	19
I-12CMS(2) x NK-183mm-O#3	0	55	0	55
I-12CMS(3) x NK-183mm-O#3	19	4	6	29

628

629

630 **Table 2** Relative accumulation of *orf129* mRNA in I-12CMS(2) and I-12CMS(3) measured by qRT-PCR

Reference gene	Sugar beet line	Relative accumulation of *orf129* mRNA (±SD)	p^1
Act	I-12CMS(2)	1.061±0.273	0.00029
	I-12CMS(3)	0.390±0.130	

631 [1]Student's *t*-test.

632

633

Figure legends

Fig 1 Physical map of the master chromosome of the I-12CMS(3) mitochondrial genome. The 103 kbp-repeated sequences are the regions marked from 1 to 2 and from 3 to 4. Regions from 2 to 3 and from 4 to 1 do not have features in common. Two recombinogenic sequence-families are indicated by solid lines (R1) and dashed horizontal lines (R2), respectively. Two of the R1 copies share additional regions of homology shown by brackets, and another copy of the additional region exists. The stippled region is deleted in the I-12CMS(2) mitochondrial genome (see Fig. 2). Locations of some mitochondrial genes are shown. Restriction enzymes are *Xho*I (Xh), *Sal*I (Sa), *Xba*I (Xb), and *Sma*I (Sm). The scale bar is shown below (kbp)

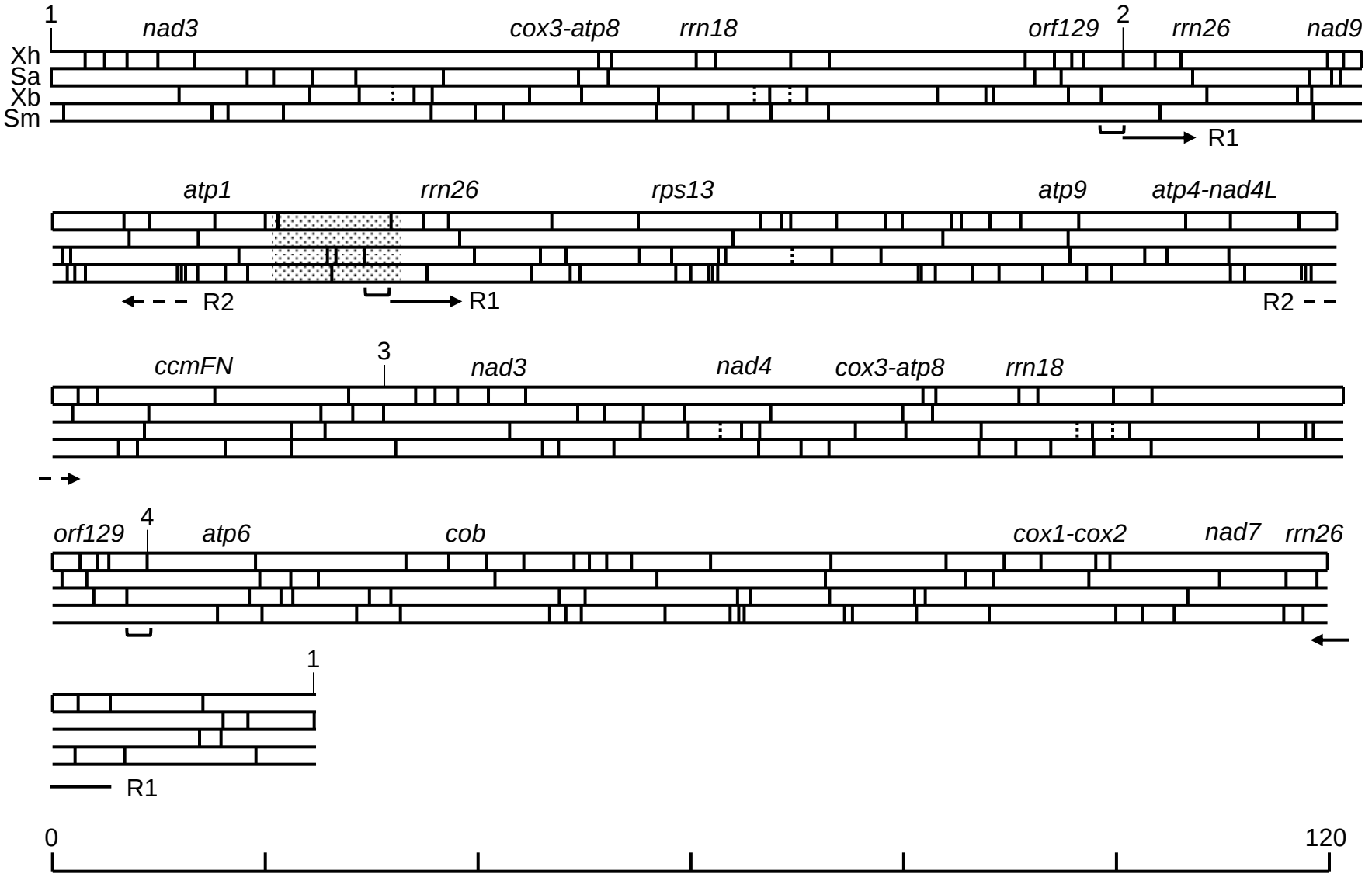
Fig. 2 Comparison of the mitochondrial *atp1-rrn26* intergenic region between CMS-E and TK-81mm-O. Nucleotide residues are numbered according to the DDBJ/EMBL/GenBank accession numbers of FQ014226 (CMS-E) and BA000009 (TK-81mm-O). Boxed regions with lower case letters indicate conserved sequences between CMS-E and TK-81mm-O. Regions with upper case letters indicate regions unique to CMS-E or TK-81mm-O. Internal sequences are omitted, and the length of the unique region is shown. The repeated sequences in CMS-E (Darracq et al. 2011) are underlined

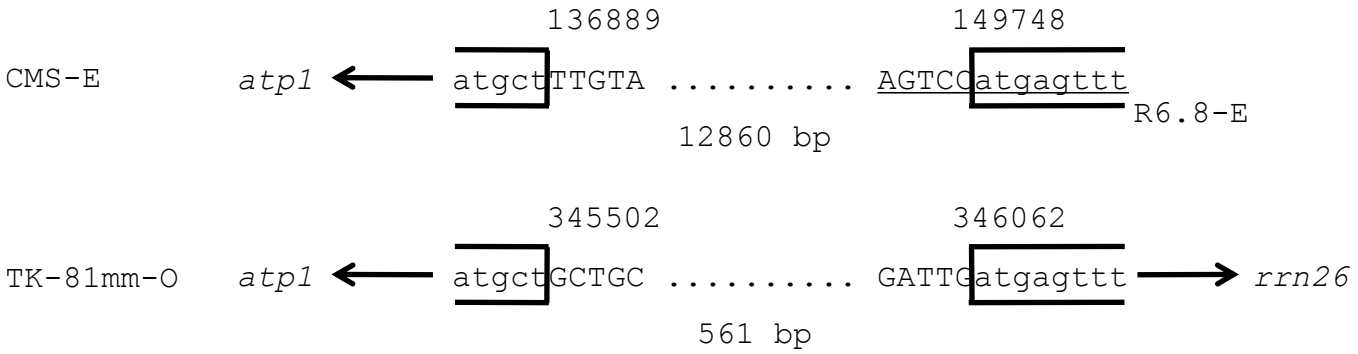
Fig. 3 Expression analysis of *orf129*. Panel a. Immunoblot analysis of total cellular proteins from green leaves of I-12CMS(R) (R), I-12CMS(2) (2), and I-12CMS(3) (3). Protein samples were electrophoresed in a 15% SDS-polyacrylamide gel. The blot was probed with anti-ORF129 antiserum or stained with Ponceau S. The size of the signal bands is shown on the left. Panel b. Immunoblot analysis of total cellular proteins from I-12CMS(2) and I-12CMS(3) green leaves. The total protein from I-12CMS(2) was diluted as indicated. The blot was probed with anti-ORF129 antiserum. Panel c. RNA gel blot analysis of total mitochondrial RNA from I-12CMS(2) and I-12CMS(3) leaves. The blots were probed with *orf129*- or *atp9*-specific probes. Sizes of the signal bands are shown on the left

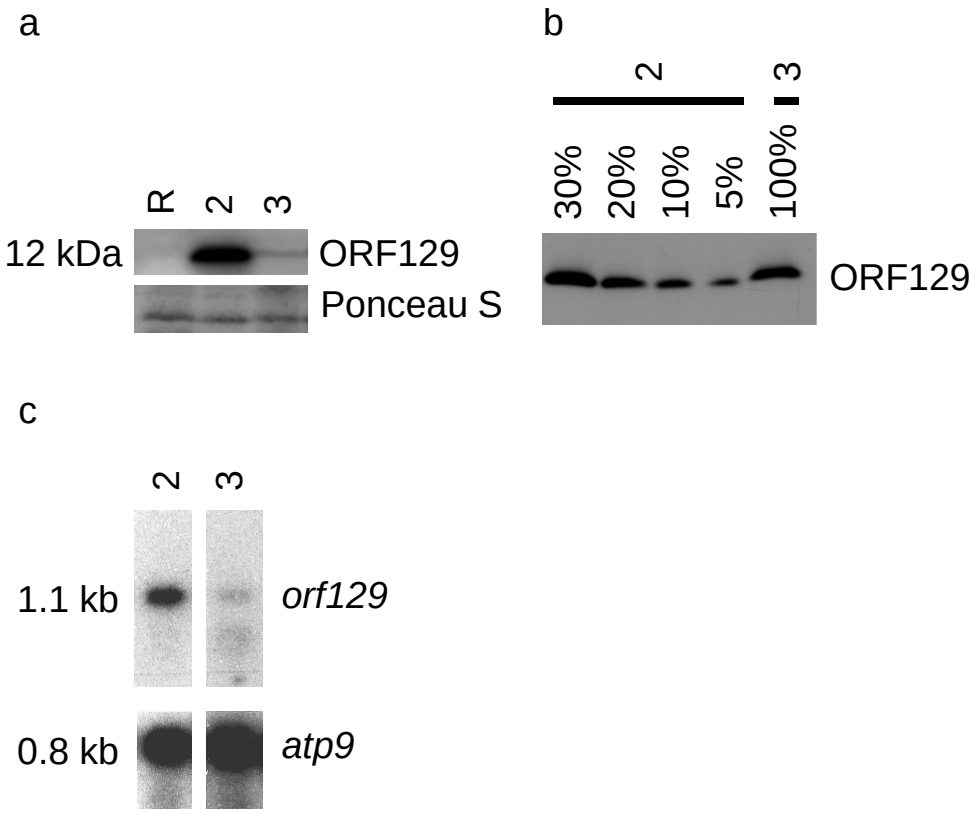
Fig. 4 Mapping transcript termini onto the *orf129* locus. Panel a. Organization of the *orf129* locus. The protein coding region is indicated by a box. The direction of transcription is from left to right. Vertical arrows indicate mapped transcript termini. Those above the horizontal line are from I-12CMS(3) and below are from I-12CMS(2). Numbers in parentheses indicate the number of cDNA clones representing the termini if the number of clones is more than one (see Table S2). Panel b. Alignment of nucleotide sequences in the region from -490 to -472 in I-12CMS(3) and I-12CMS(2). Lower case letters indicate alterations in nucleotide residues. Nonanucleotide motifs that resemble typical mitochondrial promoters are underlined. Transcript termini are indicated as in panel a

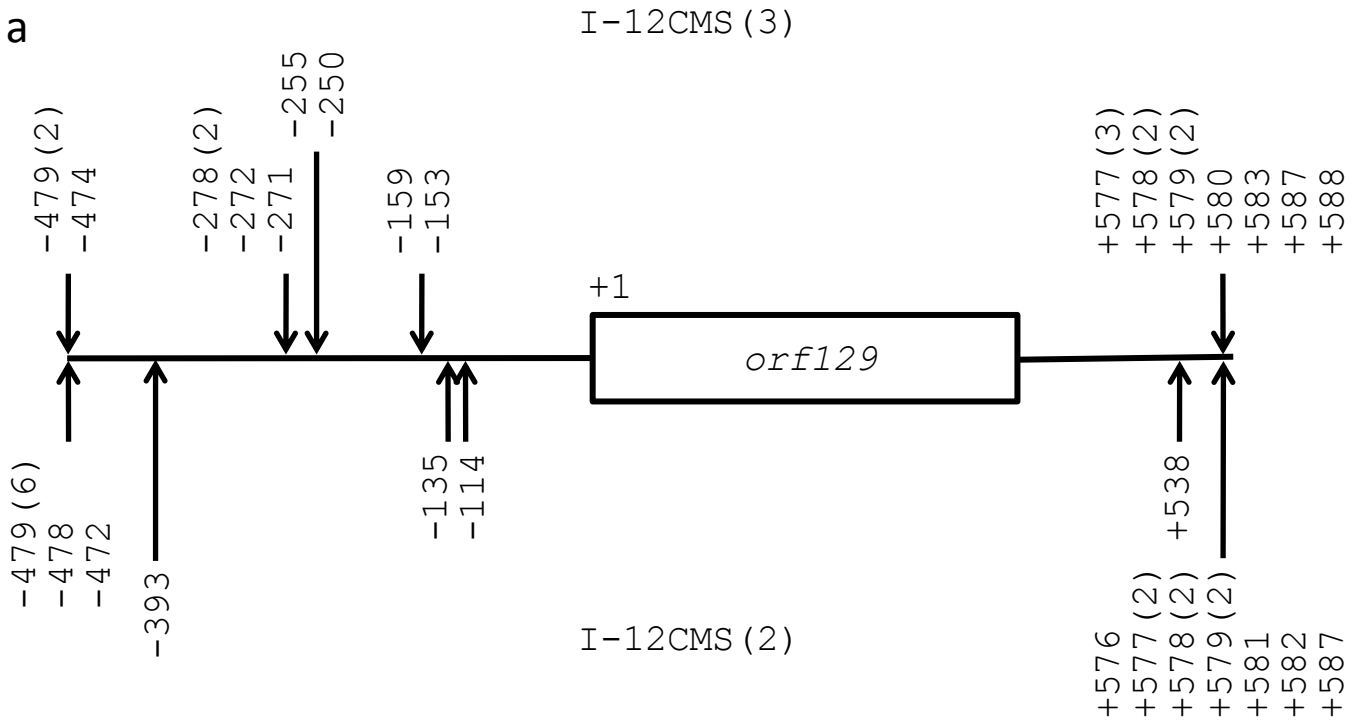
669 **Fig. 5** Expression analysis of N-terminal extension of *atp6*. Total cellular proteins from TK-81mm-O (lane
670 1), TK-81mm-CMS (lane 2), I-12CMS(2) (lane 3), I-12CMS(3) (lane 4) and I-12CMS(R) (lane 5) green
671 leaves were electrophoresed in a 12% SDS polyacrylamide gel. The blot was probed with anti-preSATP6
672 antiserum (a and b) or stained with Ponceau S (c). Exposure time to X-ray film was 1min (a) and 30min (b).
673 Size markers are shown on the left (kDa)
674
675 **Fig. 6** Heteroplasmy in the mitochondrial *atp1-rrn26* intergenic region of I-12CMS(3). Panel a.
676 Organization of the *atp1-rrn26* intergenic region in I-12CMS(3) and I-12CMS(2). Boxes indicate
677 conserved regions between the two sugar beet lines. Solid and dashed lines indicate regions unique to
678 I-12CMS(3) and I-12CMS(2), respectively. See also Fig. 2. Positions of PCR primers are shown by
679 triangles. Panel b. Negative images of gels after electrophoresis of PCR products. Total cellular DNAs of
680 four I-12CMS(2) plants (lanes 1 to 4) and two I-12CMS(3) plants (lanes 5 and 6) were subjected to PCR
681 analysis using two primer combinations (6-1 and 6-3, and 6-2 and 6-3). Lane 7 is a mock reaction in which
682 template DNA was omitted. The PCR protocol was: 94˚C for 2min; and 35 cycles of 94˚C for 30sec, 55˚C
683 for 30sec, and 68˚C for 30sec. The sizes of signal bands are shown on the left (kbp)
684
685 **Fig. 7** Heteroplasmy at the mitochondrial *orf129* locus of I-12CMS(3). Panel a. Organization of the *orf129*
686 locus. The open box indicates the *orf129*-coding region. The direction of transcription is shown by an arrow.
687 A single nucleotide polymorphism in the upstream of *orf129* is guanine (G) or thymine (T) in I-12CMS(3)
688 or I-12CMS(2), respectively. Positions of PCR primers are shown by triangles. Note that the 3' termini of
689 primer 7-1 and primer 7-2 are T and G, respectively. Panel b. Negative images of gels after electrophoresis
690 of PCR products. Total cellular DNAs of four I-12CMS(2) plants (lanes 1 to 4) and two I-12CMS(3) plants
691 (lanes 5 and 6) were subjected to PCR analysis using two primer combinations (7-1 and 7-3, and 7-2 and
692 7-3). Lane 7 is a control experiment in which PCR product from I-12CMS(3) total cellular DNA with
693 primers 7-2 and 7-4 was used as template. The PCR protocol was: 95˚C for 3min; and 30 cycles of 95˚C for
694 30sec, 64˚C for 30sec, and 72˚C for 30sec. The sizes of signal bands are shown on the left (kbp)
695
696 **Fig. 8** Mutant allele-specific PCR amplification of the *atp6* locus. Panel a. Organization of the *atp6* locus.
697 The open box indicates the *atp6*-coding region. The direction of transcription is shown by an arrow. A
698 single nucleotide polymorphism at nucleotide +177 (the first coding nucleotide is defined as +1) is cytosine
699 (C) or adenine (A) in I-12CMS(3) or I-12CMS(2), respectively. Positions of PCR primers are shown by
700 triangles. Note that the 3' termini of primer 8-1 and primer 8-2 are C and A, respectively. Panel b. Negative
701 images of gels after electrophoresis of PCR products. Total cellular DNAs of four I-12CMS(2) plants (lanes
702 1 to 4) and two I-12CMS(3) plants (lanes 5 and 6) were subjected to PCR analysis using two primer
703 combinations (8-1 and 8-3, and 8-2 and 8-3). Lane 7 is a mock reaction in which template DNA was omitted.

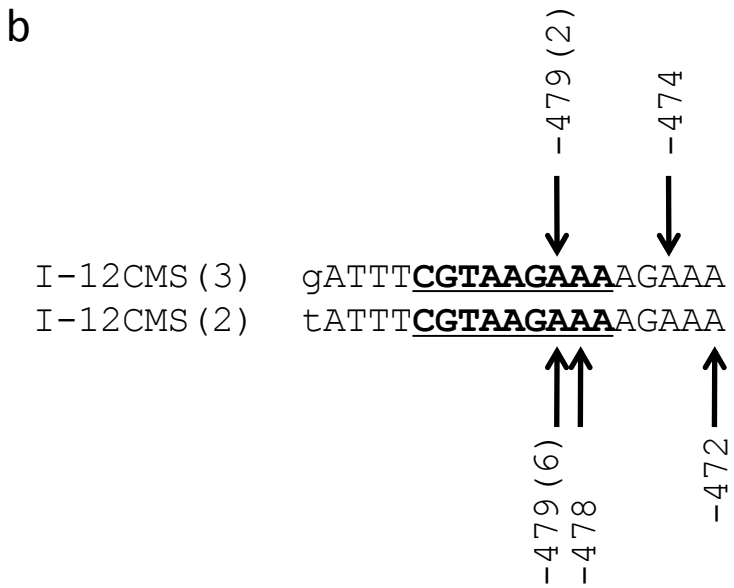
704 The PCR protocol was: 95˚C for 3min; and 30 cycles of 95˚C for 30sec, 60˚C for 30sec, and 72˚C for 30sec.
705 The sizes of signal bands are shown on the left (kbp)
706
707 Fig. 9 Evolutionary model for the emergence of I-12CMS(2) mitochondria from I-12CMS(3) mitochondria
708 through a possible intermediate. Circles represent mitochondrial genomes: those present as predominant
709 DNA molecules and those present as substoichiometric DNA molecules are shown by solid lines and dotted
710 lines, respectively. Three loci are shown on the circles but the position and the copy number of the loci does
711 not exactly correspond to the organization of the mitochondrial master chromosome shown in Fig. 1. At the
712 *atp6-rrn26* loci, large and small rectangles denote the 12860-bp sequence and the 561-bp sequence,
713 respectively. At the *orf129* loci, the nucleotide sequence polymorphism at nucleotide -490 (G or T) is
714 shown. At the *atp6* loci, the nucleotide sequence polymorphism at nucleotide +177 (C or A) is shown. In
715 I-12CMS(3) mitochondria, two types of mitochondrial DNA molecules coexist, one being at
716 substoichiometric levels. Note that the *atp6* allele of this substoichiometric DNA molecule is unknown
717 (shown by '?'). This substoichiometric DNA molecule evolved into an intermediate molecule with an A
718 residue at nucleotide +177 in *atp6* (this intermediate molecule has not been found and is still an assumption).
719 The evolved intermediate (substoichiometric) DNA molecule eventually predominated in I-12CMS(2).
720 Because no substoichiometric DNA molecule with the 12860-bp sequence at *atp1-rrn26*, G at nucleotide
721 -490 of *orf129*, or C at nucleotide +177 of *atp6* was found in I-12CMS(2), such substoichiometric DNA
722 molecules are missing from I-12CMS(2) mitochondria in this model

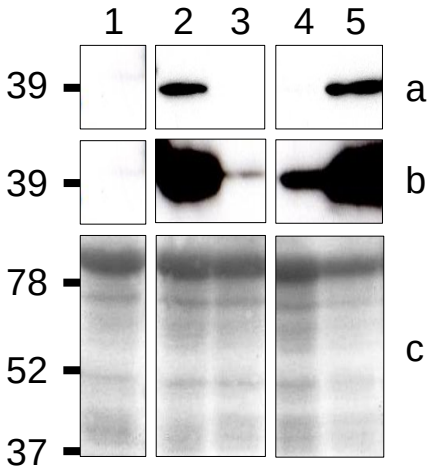

1
nad3
cox3-atp8
rrn18
orf129
2
rrn26
nad9
Xh
Sa
Xb
Sm
R1
atp1
rrn26
rps13
atp9
atp4-nad4L
R2
R1
R2
ccmFN
3
nad3
nad4
cox3-atp8
rrn18
orf129
4
atp6
cob
cox1-cox2
nad7
rrn26
1
R1
0
120

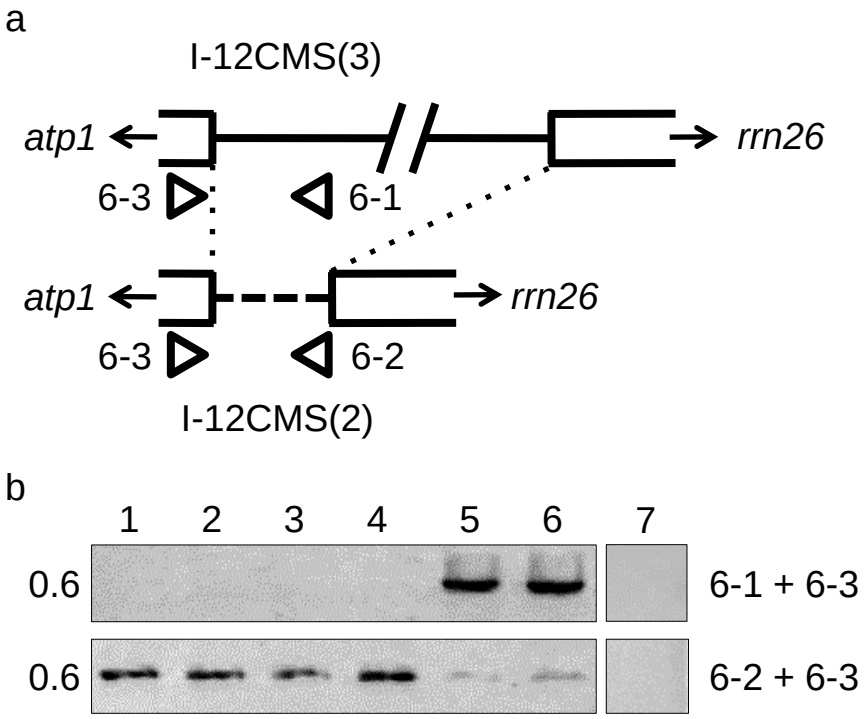

136889
149748
CMS-E
atp1
atgctTTGTA .....
AGTCCatgagttt
R6.8-E
12860 bp
345502
346062
TK-81mm-O
atp1
atgctGCTGC .....
GATTGatgagttt
rrn26
561 bp

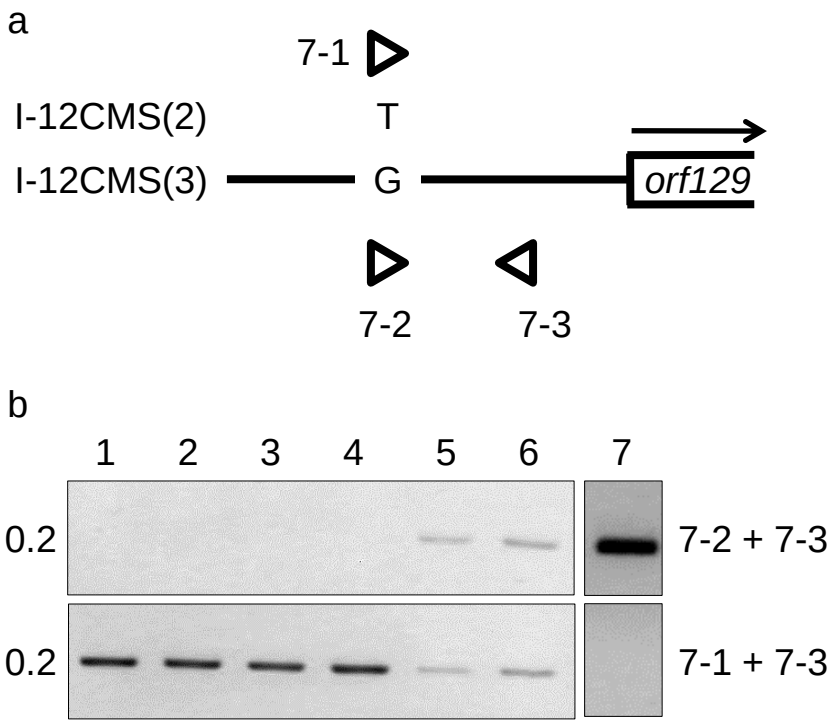

a
R
2
3
12 kDa
ORF129
Ponceau S
b
2
3
30%
20%
10%
5%
100%
ORF129
c
2
3
1.1 kb
orf129
0.8 kb
atp9

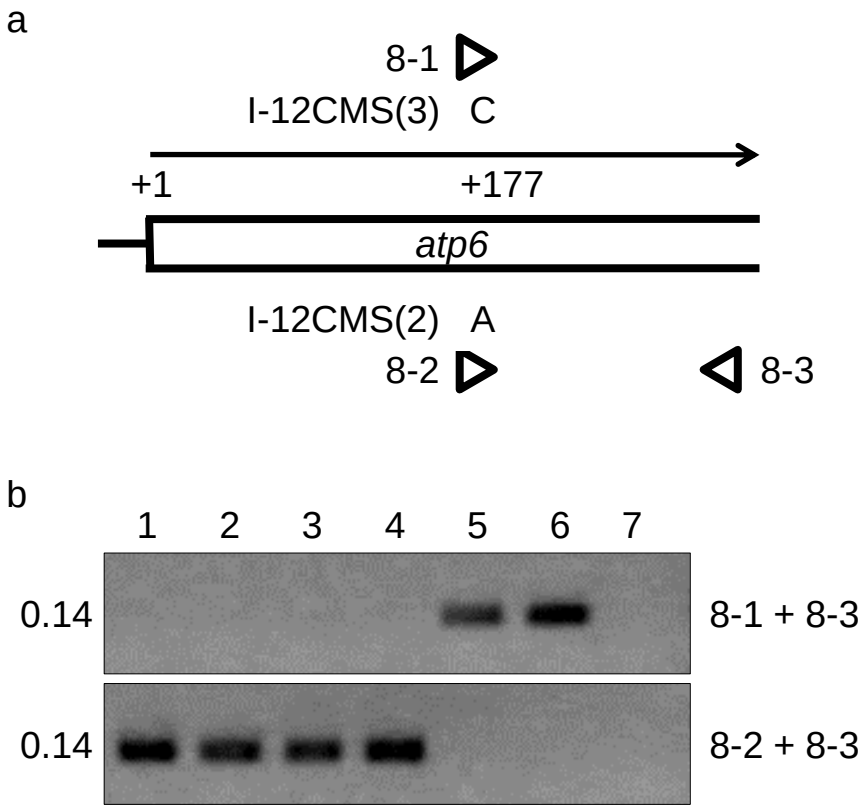

a
I-12CMS(3)
-479(2)
-474
-278(2)
-272
-271
-255
-250
-159
-153
+1
orf129
+577(3)
+578(2)
+579(2)
+580
+583
+587
+588
-479(6)
-478
-472
-393
-135
-114
+538
I-12CMS(2)
+576
+577(2)
+578(2)
+579(2)
+581
+582
+587

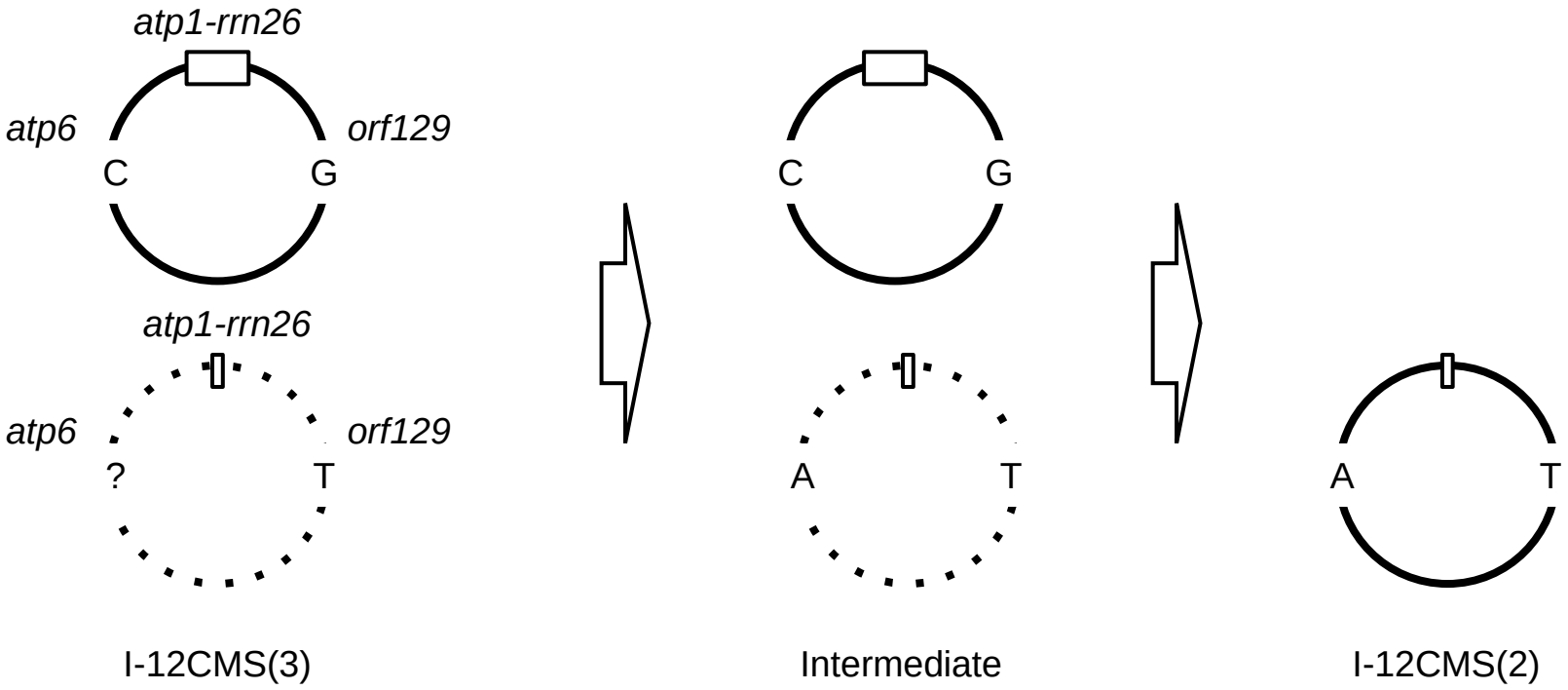

b
-479(2)
-474
I-12CMS(3) gATTTCGTAAGAAAAGAAA
I-12CMS(2) tATTTCGTAAGAAAAGAAA
-479(6)
-478
-472


1
2
3
4
5
39
39
78
52
37
a
b
c


a
I-12CMS(3)
atp1
rrn26
6-3
6-1
atp1
rrn26
6-3
6-2
I-12CMS(2)
b
1
2
3
4
5
6
7
0.6
6-1 + 6-3
0.6
6-2 + 6-3


a
7-1
I-12CMS(2)
T
I-12CMS(3)
G
orf129
7-2
7-3
b
1
2
3
4
5
6
7
0.2
7-2 + 7-3
0.2
7-1 + 7-3


a
8-1
I-12CMS(3) C
+1
+177
atp6
I-12CMS(2) A
8-2
8-3
b
1 2 3 4 5 6 7
0.14
8-1 + 8-3
0.14
8-2 + 8-3



I-12CMS(3) Intermediate I-12CMS(2)

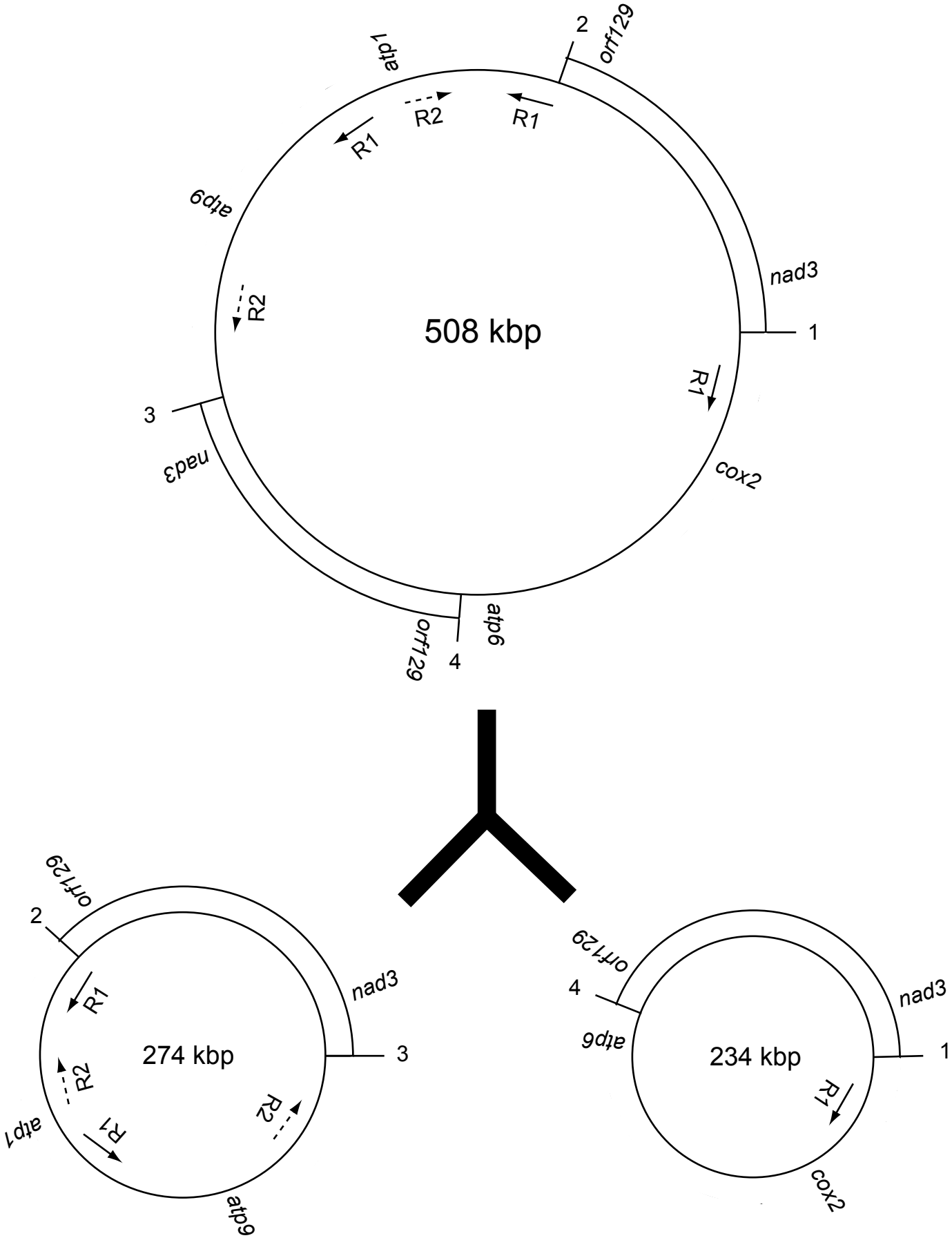


Fig. S1 Interconversion of the 508-kbp master chromosome of I-12CMS(3) mitochondria into two circular molecules of 274- and 234-kbp. Points indicated by 1 to 4 correspond to those in Fig. 1. The 103-kbp repeated sequences are boxed. Two recombinogenic sequence families are indicated by solid lines (R1) and dashed lines (R2), respectively. The locations of some mitochondrial genes are indicated.

```
GGATCCCGCAAACGGAGACATGAACCAAGCAGGAAGAGGAATTGGAGCAGCTCTTTGCCT  -1079
TATGTTATGATCTCATAACACCGAAAGATTAGACCCTTTTTTTTCCTTTCTTTCAGGATG  -1019
AAAACACAGGGCAGGAGGGTATAGAAAGTCCTCTATATGCCGTCCATAGTCCATCGACGC  -959
CTCGTAGTACCACTTACAGTACTCGCAGAGTGATTGAGTCCAAGATTTTAACTCAATCTT  -899
CTCAACGAGGAAAGAAGCATCGTCGTCGAGCTGGCGAGAGAAATCCTCTATAAGTCTCTC  -839
TACGTAACCCCAATCGGGCTTACAGCTCTCAACTAACATCATACCAATCTTATAACAGCT  -779
TACCAGAAAACCCTCTGGAAAACCTAACCAGACTGAGAAAGGCATAGGACGGATACCACC  -719
TGGAGAGAGGGCGACAAGAATGTGCCGAAACCCGTGACGATTGTAATTTGGGTTAATGTT  -659
ACAAGGCTTACAGGAAGACCTCCTAACCCCTCTTAATCTCAGAGAAGTCTGAATGTTATT  -599
ACACCCTATTTGCTTTGCTTCATCACCGGCATTTTAAAGCCACAGCGAACCAAGCTGACC  -539
                                                      -490
                                                      G
TACCTCTATCACTAAACTCTACCCTGGGCTCTTTTACCTCTAAAAAGATATTTCGTAAGA  -479
AAAGAAAAGTGCTTTTTCCTAGGGTAGATGAAAGGAATGCGGAAAGGCTATCCGAGTTCAC  -419
AGGCGTAGTTGCGTGTACTTCATTTCATTGAAATTGAGTTCGTTACCGCACCGTGGGTAG  -359
AAGATCCTTCTCTTTCTCTGTGAAGATAGGCTGTGCTTTGCATGTCAAGTGCGAAATTGG  -299
ATCGAAAAATTGCGTATGATATGAAAGGTTTATGGTCTATCTAAGACTCTCTTTGATATC  -239
AATCACTAAACTAAAGAAAGTAAAATTTGAAAAATCATAAGATAAGAGAAGAAAGAAGAG  -179
TATCTTTCGTTGTAGCCATAGAAGACTCGACCCAAGCAATTCTTATGCCCAAAACTCCCA  -119
TGTTTTCCGGATCATTGGTTAAGAACCAACCGGCAATTTATGTCTTCCTGAATTGGGAGA  -59
GCAAGAATCAGTCTTTCTTCATTTCTCGAGAGAGCAGAGCAGTCAAAGAATGAACCAAAT  +2
                                                          M
GAATCCTTATATTCTGCAGTTTTTGGCCGACATGGCCACAGCAATTTTAACGATTGCAGG  +62
  N  P  Y  I  L  Q  F  L  A  D  M  A  T  A  I  L  T  I  A  G
AGTTGCATATCTACCACTTATTGTGTTAATTGTATTTAGGGCCGGCGGCCTTCGAGGGTT  +122
  V  A  Y  L  P  L  I  V  L  I  V  F  R  A  G  G  L  R  G  L
AAATGAAGAGAATGCAAGCGAAAGACTCTTAGACCTTTGTTGTGATACGCTGAAGGAACA  +182
  N  E  E  N  A  S  E  R  L  L  D  L  C  C  D  T  L  K  E  Q
GATCAAAAACAAAATTGAAGAGTTACTCCAGGTTTATTATAATAATTCTGTACCTTTGCC  +242
  I  K  N  K  I  E  E  L  L  Q  V  Y  Y  N  N  S  V  P  L  P
GTCTGGTCGGAGGATTCAAGACGCAGCAGCCTTCCTCCATCAGGATTCAGAATCCTTGGA  +302
  S  G  R  R  I  Q  D  A  A  A  F  L  H  Q  D  S  E  S  L  E
ACAACTGCTAATGATTTTGAAAAATATGACCGAATTGGGCGTTCAAAGTCAAGAATTTTT  +362
  Q  L  L  M  I  L  K  N  M  T  E  L  G  V  Q  S  Q  E  F  L
ACAAGTTTTACTCTATCTCTCACAGTGAAGATAGATGGAGTCTTTGCACTACAATGGATT  +422
  Q  V  L  L  Y  L  S  Q  *
TGGCGGGAGTACTTCTTTATTATAAGCGAGGACCGCCTTTTATGTTACTCGTCGTTTGGC  +482
TTGAATCTCGTTCCTTAGGAGGCTCGCAAGTTTTTAAGAAGTTACATTGGAAAAAAGGCT  +542
CCTCCCCGAACATGCCTTCGGTGAAGGTGGGGGTCGTCAATAAGAACGAGGCCCGCTCTT  +602
TGGGCGGGGAAAGGGAAGTGGGTCCCTTTCGCTTTATTTCAGTTTAGGTCGATCAGAAGG  +662
CGCTGGAACTGCTTGTTGCACAGTTGATAATCGGATATAAGCGGATAGCTAATCCCTCTT  +722
TCAACGAGAGTTGCGGACCCTCTCTTACTTACTCTTAGTGCTGTGCGAGCCTCGGTTTCA  +782
TTTCTTCTTCTAAGGCAGCTAGCTAGGCCTCTTGAGAGCTCAGGATTGGACGGAACCCTT  +842
ACGGTTGGACTATCGAACAGCTGTTGGTGGGGAATAATCTCCTTTCCCTTGATTGCTGCG  +902
GGCCTTCATTGCCGAAACTGGACTGGCTAGGATCACTAGAGGGGACAGACTACCAGGACT  +962
AGTCCGCTTCCCAATAGGAAGAGCAAACAGCAACTATTAGTTCAGCTTTCCGGGCTCTCA  +1022
TAACAGCCCTGGTTTGAGCATCTCCCTAGCCTACCTACTTTCAGAAGGTCAAGTCTAATA  +1082
GAGTAGAGGCAATCCACTCGAG                                         +1104
```

Fig. S2 Nucleotide sequence of *orf129* from I-12CMS(2). Nucleotide residues are numbered from the first residue of the putative initiation codon of *orf129*. The deduced amino acid sequence of the translation product is shown below the nucleotide sequence. The only nucleotide alteration in the corresponding region of I-12CMS(3) is shown above the nucleotide sequence. Mapped transcript termini are highlighted in red.

```
I-12CMS-2 AAGCTTGTTTCAACTTTCTATAAATTGCTGCTTTTGCTTCTAGCTAGTTC -282
I-12CMS-3 AAGCTTGTTTCAACTTTCTATAAATTGCTGCTTTTGCTTCTAGCTAGTTC -290
          **************************************************

I-12CMS-2 AGTGGATAAGATGGCTGCGCTCCAGCTCACTCCTGACTCTTAAAAAGGGA -232
I-12CMS-3 AGTGGATAAGATGGCTGCGCTCCAGCTCACTCCTGACTCTTAAAAAGGGA -240
          **************************************************

I-12CMS-2 GGATTCACGGGTTTCAATAGATAATATCGTAAGTAAGAAAGAACGTTTTC -182
I-12CMS-3 GGATTCACGGGTTTCAATAGATAATATCGTAAGTAAGAAAGAACGTTTTC -190
          **************************************************

I-12CMS-2 ACTAATTACTTTCATAGAGAGAGACCCTCCCCACCCCAGTATAATGTCCT -132
I-12CMS-3 ACTAATTACTTTCATAGAGAGAGACCCTCCCCACCCCAGTATAATGTCCT -140
          **************************************************

I-12CMS-2 CTCGAAAATCGGATACCTTTTCGTTTTCATTTCTCCCATTTCTTTCTTGG -82
I-12CMS-3 CTCGAAAATCGGATACCTTTTCGTTTTCATTTCTCCCATTTCTTTCTTGG -90
          **************************************************

I-12CMS-2 TTGGATCAACCCAACCGGTGATTTCCGACA--------AGTCTTTCTTCA -40
I-12CMS-3 TTGGATCAACCCAACCGGTGATTTCCGACATTCCGACAAGTCTTTCTTCA -40
          ******************************        ************
                                                  >atp6
I-12CMS-2 TTTTTGAGCGGACAGCAGAAAGAAAAAAGAAAAAATTTGATGATGACTCA 11
I-12CMS-3 TTTTTGAGCGGACAGCAGAAAGAAAAAAGATAAATTTTGATGATGACTCA 11
          ****************************** *** ***************

I-12CMS-2 TTGGAAGAAGGAGATAACTGCTTTGGCACAACAGATAATTACTAATGTCC 61
I-12CMS-3 TTGGAAGAAGGAGATAACTGCTTTGGCAAACCAGATAATTACTAATGTCC 61
          **************************** * ******************

I-12CMS-2 CTCGACCTACACGGCGAGTCTTATTCGGACTCGGCGTGCTGTTCGCCCTC 111
I-12CMS-3 CTCGACCTACACGGCGAGTCTTATTCGGACTCGGCGTGCTGTTCACCCTC 111
          ********************************************* *****

I-12CMS-2 TTTTGGGTCGGGAGACGCGGGCTGTTGTTGTTTGAAACGAAAATTATTCT 161
I-12CMS-3 TTTGAGGTCGGGAGACACGGGCTGTTGTT---TGAAACGGAAATTATTCT 158
          ***  *********** ************   ******* **********

I-12CMS-2 ACAAGCCCTCTTTGGAATAGGAACAGCGGCTTCTATTTTGTTTTTTTGTT 211
I-12CMS-3 ACAAGCCCTCTTTGGCATAGGAACAGCGGCTTCTATTTTGTTTTTTTGTA 208
          *************** ********************************

I-12CMS-2 ATTTCCGCACCCGAGAGCACTTTCTCTATACATTTCTACTAACGATTTAT 261
I-12CMS-3 ATTTCCGCACCCGAGAGCACTTTCTCTATACATTTCTACTAACGATTTAT 258
          **************************************************

I-12CMS-2 TTGGCTTTATTCTTCCTAAAGAGGGCTTTCCCTCTTTTGGAATTCGCCTT 311
I-12CMS-3 TTGGCTTTATTCTTCCTAAAGAGGGCTTTCCCCCTTTTGGAATTCGCCTT 308
          ********************************* *****************

I-12CMS-2 TTTGGCTATATTGATCTTTTCGGTTTGCTTCTTTCCGAAAAATGGACGAT 361
I-12CMS-3 TTTGGCTATATTGATCTTTTCGGTTTGCTTCTTTCCTAAAAATGGACGAT 358
          ************************************ *************

I-12CMS-2 TACGGCATTCGCTAGAGAACTGTTCTTTTATTTTACTACTCCTTTGGGGG 411
I-12CMS-3 TACGGCATTCGCTAGAGAACTGTTCTTTTATTTTCCTACTCCTTTGGGGG 408
          ********************************** ***************

I-12CMS-2 AGTTACCGGCTTTTGGAGACTCCGTGGCACGAATGGCCACTCTATGTTTG 461
I-12CMS-3 AGTTACCGGCTTTTGGAGACTCCGTGGCACGAATGGCCACTCTATGTTTG 458
          **************************************************


I-12CMS-2 TGGTCTCTTAATGTTATTCACGCTTCTTCAAGCGCTGGTCCATCGAAAGG 511
I-12CMS-3 TGGTCTCTTAATGTTATTCACGCTTCTTCAAGCGCTGGTCCATCGAAAGG 508
          **************************************************
```

```
I-12CMS-2 ACGAAGGGGTCGGCAAGTTTATTCTGCCGGCTACCTTCCTAATTTTAATG 561
I-12CMS-3 ACGAAGGGGTCGGCAAGTTTATTCTGCCGGCTACCTTCCTAATTTTAATG 558
          **************************************************

I-12CMS-2 AGTGCCCTCGTGCTCCTTGGGCTTCACTCGGGGGCGATTCATAATCTGCC 611
I-12CMS-3 AGTGCCCTCGTGCTCCTTGGGCTTCACTCAGGGGCGATTCATAATCTACC 608
          ***************************** ***************** **

I-12CMS-2 GGAAAAGGCGGCCCTAACGGGCGCCGCAGTCTTAATACTCTTTAGCCTAG 661
I-12CMS-3 GGAAAAGGCGGCCCTAACGGGCGCCGCAGTCTTAATAATCTTTAGCCAAG 658
          ************************************ ********* **

I-12CMS-2 CGGCCGCCGAGTCCACACGGGAACGGTGGCTTGCATATGGGGCGTATCTC 711
I-12CMS-3 CGGCCGCCGAGTCCACAAGGGAGCGGTGGCTTGCAATTGGGGCGTATCTC 708
          ***************** **** ************  *************

I-12CMS-2 ATGTGTATCCCGTGTATACTATATCTCCTAGGAGGGGCCACGGATTGGCC 761
I-12CMS-3 ATGTGTATCCCGTGTATACTATATATCCTAGGAGGGGCCACGGATTGGCC 758
          ************************ ************************

I-12CMS-2 CCTATTCGTACAACAAATCGGCTTTTTTTTTATTATCCTCGTCCTAGTAT 811
I-12CMS-3 CCTATTCGTACAACAAATCCGCCTTTTTTTTTTTATCCTCGGCCTAGTAT 808
          ******************* ** ******** ********* ********

I-12CMS-2 TCG---TTTTTTGGGTTGTTACTTATGAGGACGTAAAAGAGCAGAAGAAT 858
I-12CMS-3 TCGGTTTTTTTTGGGTTGTTACTTATGAGGACGTAAAAGAGCAGAAGGAT 858
          ***   ************************************** **

I-12CMS-2 CTGCTCCCGATCCTATCTGCGATCTTCTCCGCTTTGGCGTCGGCGAGGTG 908
I-12CMS-3 CTGCTCCCGATCCTATCTGCGATCTTCTCCACTTTGGCGTCGGCGAGGTG 908
          ****************************** ******************

I-12CMS-2 TGGCCCCGGCGTCTGGGTAAAAGAGTGGCTCGCCGCCTTTTCCTTCACCA 958
I-12CMS-3 TGGCCCCGGCGTCTGGGTAAAAGAGTGGCTCGGCGCCTTTTCCTTCACCA 958
          ******************************** *****************

I-12CMS-2 TTGCGGCGTATTTGTTATTTGAGAAGGTAGGCTACCGCAAAGATACCCTT 1008
I-12CMS-3 TTGGGGCGTATTTTTTATTTGAGAAGGTAGGCTACCGCAAAGATACCCTT 1008
          *** ********* ************************************

I-12CMS-2 CAAGAAACGTCCCGTACTTGGGAAGCTTGGTTTTGCTTTTTTGCTGTTCT 1058
I-12CMS-3 CAAGAAACGTCCCGTACTTGGGAAGCTTGGTTTTGCTTTTTTGCTGTTCT 1058
          **************************************************

I-12CMS-2 TTCTTTTGGTGTTGCACTTAGTTTTTTTCATGCATCACCCGCCCCAGGGG 1108
I-12CMS-3 TTCTTTTGGTGTTGCACTTAGTTTTTTTCATGCATCACCCGCCCCAGGGG 1108
          **************************************************

I-12CMS-2 TCGGAACTGAAAGTGAAGCTTTTCACGATGCGATTAATCCAGCAGCGGGA 1158
I-12CMS-3 TCGGAACTGAAAGTGAAGCTTTTCACGATGCGATTAATCCAGCAGCGGGA 1158
          **************************************************

I-12CMS-2 CGGATCCCCAGCCCGCTAGAGCAATTTTCCATTCTCCCATTGATTCCTAT 1208
I-12CMS-3 CGGATCCCCAGCCCGCTAGAGCAATTTTCCATTCTCCCATTGATTCCTAT 1208
          **************************************************

I-12CMS-2 GAAAATAGGAAACTTGTATTTCTCATTCACAAATCCATCTTTGTTTATGC 1258
I-12CMS-3 GAAAATAGGAAACTTGTATTTCTCATTCACAAATCCATCTTTGTTTATGC 1258
          **************************************************


I-12CMS-2 TGCTAACTCTCAGTTTGGTCCTACTTCTGCTTCATTTTGTTACTAAAAAG 1308
I-12CMS-3 TGCTAACTCTCAGTTTGGTCCTACTTCTGCTTCATTTTGTTACTAAAAAG 1308
          **************************************************

I-12CMS-2 GGAGGAGGAAACTCAGTACCAAATGTTTGGCAATCCTTGGTAGAGCTTAT 1358
I-12CMS-3 GGAGGAGGAAACTCAGTACCAAATGTTTGGCAATCCTTGGTAGAGCTTAT 1358
          **************************************************
```

```
I-12CMS-2 TTATGATTTCGTGCTGAACCTGGTAAACGAACAAATAGGTGGTCTTTCCG 1408
I-12CMS-3 TTATGATTTCGTGCTGAACCTGGTAAACGAACAAATAGGTGGTCTTTCCG 1408
          **************************************************

I-12CMS-2 GAAATGTGAAACAAAAGTTTTTCCCTTGCATCTTGGTCACTTTTACTTTT 1458
I-12CMS-3 GAAATGTGAAACAAAAGTTTTTCCCTTGCATCTTGGTCACTTTTACTTTT 1458
          **************************************************

I-12CMS-2 TTGTTATTTCGTAATCTCCAGGGTATGATACCCTATAGCTTTACAGTTAC 1508
I-12CMS-3 TTGTTATTTCGTAATCTCCAGGGTATGATACCCTATAGCTTTACAGTTAC 1508
          **************************************************

I-12CMS-2 AAGTCATTTTCTCATTACTTTGGGTCTTTCATTTTCCATTTTTATTGGCA 1558
I-12CMS-3 AAGTCATTTTCTCATTACTTTGGGTCTTTCATTTTCCATTTTTATTGGCA 1558
          **************************************************

I-12CMS-2 TTACTATAGTGGGATTTCAAAGAAATGGGCTTCATTTTTTAAGCTTCTCA 1608
I-12CMS-3 TTACTATAGTGGGATTTCAAAGAAATGGGCTTCATTTTTTAAGCTTCTCA 1608
          **************************************************

I-12CMS-2 TTACCTGCAGGAGTCCCGCTGCCGTTAGCACCTTTTTTAGTACTCCTTGA 1658
I-12CMS-3 TTACCTGCAGGAGTCCCGCTGCCGTTAGCACCTTTTTTAGTACTCCTTGA 1658
          **************************************************

I-12CMS-2 GCTAATCCCTCATTGTTTTCGCGCATTAAGCTCAGGAATACGTTTATTTG 1708
I-12CMS-3 GCTAATCCCTCATTGTTTTCGCGCATTAAGCTCAGGAATACGTTTATTTG 1708
          **************************************************

I-12CMS-2 CTAATATGATGGCCGGTCATAGTTCAGTAAAGATTTTAAGTGGGTTCGCT 1758
I-12CMS-3 CTAATATGATGGCCGGTCATAGTTCAGTAAAGATTTTAAGTGGGTTCGCT 1758
          **************************************************

I-12CMS-2 TGGACTATGCTATGTATGAATGATCTTTTATATTTCATAGGAGATCTTGG 1808
I-12CMS-3 TGGACTATGCTATGTATGAATGATCTTTTATATTTCATAGGAGATCTTGG 1808
          **************************************************

I-12CMS-2 TCCTTTATTTATAGTTCTTGCATTAACCGGTCTTGAATTAGGTGTAGCTA 1858
I-12CMS-3 TCCTTTATTTATAGTTCTTGCATTAACCGGTCTTGAATTAGGTGTAGCTA 1858
          **************************************************

I-12CMS-2 TATTACAAGCTCATGTTTTTACGATCTTAATCTGTATTTACTTGAATGAT 1908
I-12CMS-3 TATTACAAGCTCATGTTTTTACGATCTTAATCTGTATTTACTTGAATGAT 1908
          **************************************************
                                                    atp6/
I-12CMS-2 GCTACAAATCTCCATCAAAATTCTTTTTTCTTTTTATTAGAATTTTTATA 1958
I-12CMS-3 GCTACAAATCTCCATCAAAATTCTTTTTTCTTTTTATTAGAATTTTTATA 1958
          **************************************************

I-12CMS-2 ATT 1961
I-12CMS-3 ATT 1961
          ***
```

Fig. S3 Comparison of nucleotide sequences derived from I-12CMS(2) and I-12CMS(3) *atp6* loci. Dashes are introduced for maximum matching. Matched residues are marked by asterisks. Nucleotide residues are numbered from the underlined translation initiation codons. The termination codons are double-underlined. Sources of sequence data are DDBJ/EMBL/GenBank accession number AB015177 and this study. Nucleotide sequences were aligned using ClustalW (http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja).

```
I-12CMS-2 MMTHWKKEITALAQQIITNVPRPTRRVLFGLGVLFALFWVGRRGLLLFETKIILQALFGI 60
I-12CMS-3 MMTHWKKEITALANQIITNVPRPTRRVLFGLGVLFTLFEVGRHG-LLFETEIILQALFGI 59
          ************* ******************** ** *** * ***** *********

I-12CMS-2 GTAASILFFCYFRTREHFLYTFLLTIYLALFFLKRAFPLLEFAFLAILIFSVCFFPKNGR 120
I-12CMS-3 GTAASILFFCNFRTREHFLYTFLLTIYLALFFLKRAFPLLEFAFLAILIFSVCFFPKNGR 119
          ********** ************************************************

I-12CMS-2 LRHSLENCSFILLLLWGSYRLLETPWHEWPLYVCGLLMLFTLLQALVHRKDEGVGKFILP 180
I-12CMS-3 LRHSLENCSFIFLLLWGSYRLLETPWHEWPLYVCGLLMLFTLLQALVHRKDEGVGKFILP 179
          *********** ************************************************

I-12CMS-2 ATFLILMSALVLLGLHSGAIHNLPEKAALTGAAVLILFSLAAAESTRERWLAYGAYLMCI 240
I-12CMS-3 ATFLILMSALVLLGLHSGAIHNLPEKAALTGAAVLIIFSQAAAESTRERWLAIGAYLMCI 239
          ********************************** ** ************ *******

I-12CMS-2 PCILYLLGGATDWPLFVQQIGFFFIILVLVF-VFWVVTYEDVKEQKNLLPILSAIFSALA 299
I-12CMS-3 PCILYILGGATDWPLFVQQIRLFFFILGLVFGFFWVVTYEDVKEQKDLLPILSAIFSTLA 299
          ***** **************  ** ** ***  ************* ********** **

I-12CMS-2 SARCGPGVWVKEWLAAFSFTIAAYLLFEKVGYRKDTLQETSRTWEAWFCFFAVLSFGVAL 359
I-12CMS-3 SARCGPGVWVKEWLGAFSFTIGAYFLFEKVGYRKDTLQETSRTWEAWFCFFAVLSFGVAL 359
          ************** ****** ** **********************************

                              presequence/core region
I-12CMS-2 SFFHASPAPGVGTESEAFHDAINPAAGRIPSPLEQFSILPLIPMKIGNLYFSFTNPSLFM 419
I-12CMS-3 SFFHASPAPGVGTESEAFHDAINPAAGRIPSPLEQFSILPLIPMKIGNLYFSFTNPSLFM 419
          ************************************************************

I-12CMS-2 LLTLSLVLLLLHFVTKKGGGNSVPNVWQSLVELIYDFVLNLVNEQIGGLSGNVKQKFFPC 479
I-12CMS-3 LLTLSLVLLLLHFVTKKGGGNSVPNVWQSLVELIYDFVLNLVNEQIGGLSGNVKQKFFPC 479
          ************************************************************

I-12CMS-2 ILVTFTFLLFRNLQGMIPYSFTVTSHFLITLGLSFSIFIGITIVGFQRNGLHFLSFSLPA 539
I-12CMS-3 ILVTFTFLLFRNLQGMIPYSFTVTSHFLITLGLSFSIFIGITIVGFQRNGLHFLSFSLPA 539
          ************************************************************

I-12CMS-2 GVPLPLAPFLVLLELIPHCFRALSSGIRLFANMMAGHSSVKILSGFAWTMLCMNDLLYFI 599
I-12CMS-3 GVPLPLAPFLVLLELIPHCFRALSSGIRLFANMMAGHSSVKILSGFAWTMLCMNDLLYFI 599
          ************************************************************

I-12CMS-2 GDLGPLFIVLALTGLELGVAILQAHVFTILICIYLNDATNLHQNSFFFLLEFL 652
I-12CMS-3 GDLGPLFIVLALTGLELGVAILQAHVFTILICIYLNDATNLHQNSFFFLLEFL 652
          ****************************************************
```

Fig. S4 Alignment of deduced amino acid sequences of I-12CMS(2) and I-12CMS(3) ATP6. Dashes are introduced for maximum matching. Matched residues are marked by asterisks. The junction between the N-terminal extension and the core region is shown by a slashed line. Antigenic regions of αpreSATP6 that are conserved in I-12CMS(2) and I-12CMS(3) are underlined.

Table S1 Nucleotide sequences of primers used in this study

Name of primer	Nucleotide sequence
orf129-Fw	5'-TTGTTGTGATACGCTGAAGGAA-3'
orf129-Rv	5'-CCAGACGGCAAAGGTACAGA-3'
actin-Fw	5'-AGACCTTCAATGTGCCTGCT-3'
actin-Rv	5'-ACGACCAGCAAGATCCAAAC-3'
orf129-162R	5'-ACA AAGGTCTAAGAGTCTTTCGCTTGCA-3'
orf129-54R	5'-CGTTAAAATTGCTGTGGCCATGTC-3'
orf129-339F	5'-GGGCGTTCAAAGTCAAGAATTTTTACAAG-3'
6-1	5'-TATCACCACCAACAAAGTATAAGTCAT-3'
6-2	5'-CAATCACATCTAGCTTCTTCTGAGATAT-3'
6-3	5'-GACTTGAAAGCCTTCTTCCCGAAA-3'
7-1	5'-GGCTCTTTTACCTCTAAAAAGAT-3'
7-2	5'-TCTACCCTGGGCTCTTTTACCTCTAAAAAGAG-3'
7-3	5'-GACATGCAAAGCACAGCCTA-3'
8-1	5'-TCTACAAGCCCTCTTTGGC-3'
8-2	5'-TCTACAAGCCCTCTTTGGA-3'
8-3	5'-CAAAAAGGCGAATTCCAAAA-3'

Table S2 Summary of *orf129*-transcript termini in I-12CMS(2) and I-12CMS(3) by cDNA sequencing.

Sugar beet line	Plasmid clone ID	5' terminus[1]	3' terminus[1]	Additional nucleotides[2]
I-12CMS(2)	2-1	-114	+579	
	2-2	-135	+577	
	2-3	-393	+577	
	2-4	-472	+576	
	2-5	-478	+578	AAAAAAAAA
	2-6	-479	+538	
	2-7	-479	+582	TAAA
	2-8	-479	+579	AA
	2-9	-479	+578	CTAA
	2-10	-479	+587	
	2-11	-479	+581	
I-12CMS(3)	3-1	-153	+577	
	3-2	-159	+580	
	3-3	-250	+587	
	3-4	-255	+583	
	3-5	-271	+579	
	3-6	-272	+579	
	3-7	-278	+588	
	3-8	-278	+577	
	3-9	-474	+578	A
	3-10	-479	+578	AA
	3-11	-479	+577	AAA

[1]Number corresponds to Fig. S2.

[2]Nucleotides not encoded by genomic DNA.