Title:

Polymorphic minisatellites in the mitochondrial DNAs of *Oryza* and *Brassica*

Authors:

Yujiro Honma, Yu Yoshida, Toru Terachi, Kinya Toriyama, Tetsuo Mikami, Tomohiko Kubo

Addresses:

Y. H., Y. Y., T. M. and T. K.:

Laboratory of Genetic Engineering, Research Faculty of Agriculture, Hokkaido University, N-9, W-9, Kita-ku, Sapporo 060-8589, Japan

T. T.:

Department of Bioresource and Environmental Sciences, Faculty of Life Sciences, Kyoto Sangyo University, Kamigamo, Kita-ku, Kyoto, 603-8555, Japan

K. T.:

Laboratory of Environmental Plant Biotechnology, Graduate School of Agricultural Science, Tohoku University, 1-1 Tsutumidori-Amamiyamachi, Aoba-ku, Sendai 981-8555, Japan

Corresponding author:

Tomohiko Kubo

gelab@abs.agr.hokudai.ac.jp

tel/fax +81-11-706-2484

Abstract

Polymorphic analyses of angiosperm mitochondrial DNA are rare in comparison with chloroplast DNA because few target sequences in angiosperm mitochondrial DNA are known. Minisatellites, a tandem array of repeated sequences with a repeat unit of 10 to ~100 bp, are popular target sequences of animal mitochondria, but *Beta vulgaris* is the only known angiosperm species for which such an analysis has been conducted. From this lack of information, it was uncertain as to whether polymorphic minisatellites existed in other angiosperm species. Ten plant mitochondrial DNAs were found to contain minisatellite-like repeated sequences, most of which were located in intergenic regions but a few occurred in gene-coding and intronic regions. *Oryza* and *Brassica* accessions were selected as models for the investigation of minisatellite polymorphism because substantial systematic information existed. PCR analysis of 42 *Oryza* accessions revealed length polymorphisms in four of the five minisatellites. The mitochondrial haplotypes of the 16 *Oryza* accessions with chromosomal complement (genome) types of CC, BBCC and CCDD were identical but were clearly distinguished from BB-genome accessions, a result consistent with the notion that the cytoplasmic donor parent of the amphidiploid species might be the CC-genome species. Twenty-nine accessions of six major cultivated species of *Brassica* were classified into five mitochondrial haplotypes based on two polymorphic minisatellites out of six loci. The haplotypes of *Brassica juncea* and *Brassica carinata* accessions were identical to *Brassica rapa* and *Brassica nigra* accessions, respectively. The haplotypes of *Brassica napus* accessions were heterogeneous and unique, results that were consistent with previous studies.

Key words

Plant mitochondria, mitochondrial genome, variable number of tandem repeat loci, mitochondrial evolution, rice, rapeseed

Introduction

Minisatellites are a class of tandemly repeated sequences whose repeat unit ranges from 10 bp to ~100 bp, whereas microsatellites, or simple sequence repeats, are defined as tandem repeats with a repeat unit of less than 10 bp (usually 1-5 bp) (Richard and Paques 2000). Since the first report by Jeffreys et al. (1985), it has been widely recognized that minisatellites could be used as genetic markers if the number of repeat units varies among alleles.

Minisatellites occur not only in nuclear DNA but also in the organellar DNA found in chloroplasts and mitochondria (Bastien et al. 2003; Powell et al. 1995; Rand 1993). In a number of animal mitochondrial DNAs, minisatellites have been found in the D-loop region, an area that is noncoding but that contains promoters and a replication origin and is used extensively for polymorphic analyses (reviewed in Lunt et al. 1998); however, fewer such analyses have been conducted on angiosperm chloroplast or mitochondrial DNAs compared with the number of studies on animal mitochondrial DNA. Examples of chloroplast minisatellites include those found in the intergenic region between *trnM*-CAU and *rbcL* of *Sorbus* (Rosaceae) (King and Ferris 2002), in the polymorphic region of alfalfa (Skinner 2000), and in the intronic region of *trnL*-UAA of orchids (Cafasso et al. 2001).

To the best of our knowledge, polymorphic minisatellites in angiosperm mitochondria have been described only in *Beta vulgaris*, a species that includes sugar beet (Nishizawa et al. 2000). The four minisatellites in beet mitochondria, three of which occur in intergenic regions and one in the N-terminal extension of *ccmC* (Kitazaki et al. 2009), have become useful markers for genetic diagnosis in sugar beet breeding, evaluation of genetic resources and ecological study (Nishizawa et al. 2007; Fievet et al. 2007; Fenart et al. 2008; Cheng et al. 2009) because the number of mitochondrial haplotypes based on minisatellite polymorphism exceeds 20 (Cheng et al. 2011), which makes fine analysis possible. It is advantageous that the differences in electrophoretic mobility of PCR products targeting beet minisatellites are easy to distinguish on a gel because the size of minisatellite repeat units range from 30 to 66 bp (Nishizawa et al. 2000), and the difference of one repeat unit can be detected by a simple electrophoresis

apparatus. Minisatellite-like repeated sequences have also been found in *Arabidopsis* and wheat mitochondrial DNAs (Nishizawa et al. 2000), but it was not determined whether they were polymorphic.

Mitochondrial polymorphism has been used less often in studies on the systematics or ecology of angiosperms than chloroplast DNA polymorphorism (e. g. CBOL Plant Working Group 2009) partly because angiosperm mitochondrial DNAs are so large (200 to 2400 kbp) and divergent (Kubo and Newton 2008) that it was difficult to choose a region to be analyzed. Additionally, nucleotide sequences of mitochondrial DNA, which could serve as a reference for polymorphic analyses, were not available for many plants until recently. However, the number of angiosperm genomic sequences that are accessible is increasing (Kitazaki and Kubo 2010), and there is no doubt that the number will be more abundant in the future. The addition of another source of markers from angiosperm mitochondrial minisatellites would be invaluable for both fundamental and applied aspects of plant genetics.

In this study, we addressed the question as to whether mitochondrial minisatellites in angiosperms other than *Beta vulgaris* contain informative polymorphism. At first, we found novel minisatellites from 10 mitochondrial DNAs, representing 10 genera. To assess the polymorphism of these minisatellites, *Oryza* and *Brassica* plants were selected as models because intensive studies have been conducted to establish systematic relationships between the species. Our results show that some of the mitochondrial minisatellites in *Oryza* and *Brassica* are polymorphic in terms of the number of repeat units, and mitochondrial haplotypes based on the minisatellites were fairly consistent with previous classifications based on other parameters.

Materials and methods

Genomic sequence analysis

The nucleotide sequences of angiosperm mitochondrial DNAs were obtained from the NCBI web site

(http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=33090&opt=organelle). Minisatellites were found by using Tandem Repeats Finder (Benson 1999), a program available on the internet (http://tandem.bu.edu/trf/trf.html), using the following parameters: alignment parameters, (2, 5, 7); minimum alignment score to report repeat, 30; maximum period size, 500. Results consisting of repeat units less than 20 bp were discarded. Nucleotide sequences were analyzed by using Sequencher (Hitachi Software Engineering Co., Ltd., Tokyo, Japan) or Genetyx (Genetyx Corporation, Tokyo, Japan). A BLAST search was done at the NCBI web site (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with default parameters.

Plant materials

Plant materials used in this study are listed in Tables 1 and 2. The wild rice accessions were distributed from the National Institute of Genetics supported by the National Bioresource Project, MEXT, Japan (http://www.shigen.nig.ac.jp/rice/oryzabase/wild/coreCollection.jsp). The *Brassica* accessions used in this study were gifts from Professor Dr. Takeshi Nishio, Tohoku University, Japan.

DNA isolation

Total cellular DNA was isolated from fresh green leaves according to the standard CTAB-based method as described by Doyle and Doyle (1990).

Polymerase chain reaction (PCR) and nucleotide sequencing

Oligonucleotides for PCR primers are listed in Table 3. PCR was conducted in a 10μL reaction mixture containing 5μL of GoTaq (Promega, Madison, WI, U. S. A.), 0.2 μM of each primer and ~10ng of genomic DNA. The PCR program consisted of 1 cycle of 94˚C for 5min, 35 cycles of 94˚C for 30sec, 54-56˚C (depending on the combination of primers, see Table 3) for 30sec and 72˚C for 1min, and 1 cycle of 72˚C for 5min. PCR products were electrophoresed in 4% polyacrylamide gels and visualized by EtBr staining. PCR products were either cloned into the pBluescript vector (Stratagene, La Jolla, CA, U. S. A.) prior to sequencing or directly sequenced using an ABI3130 genetic analyzer (Applied Biosystems, Foster City, CA, U. S. A.). Nucleotide sequence data were deposited in the DDBJ/EMBL/GenBank database under accession numbers of AB627032-AB627044.

Results

Novel minisatellites in angiosperm mitochondrial DNAs

Tandem Repeats Finder was used to find mitochondrial minisatellites in angiosperm mitochondrial sequences. We first confirmed the program's ability to identify the four previously characterized minisatellites from sugar beet mitochondrial DNA (data not shown). After the analysis of 10 mitochondrial sequences shown in Table 4, minisatellite-like repeated sequences were selected by visual inspection. Two sequence blocks with more than 80% homology were considered repeated sequences, and a related block of similar length with 50-80% homology was considered a divergent repeat. We found a total of 84 minisatellites as detailed in Table S1. The locations of these minisatellites included gene-coding regions (4 minisatellites), intronic regions (8) and intergenic regions (72). Mitochondrial genes having minisatellites in their coding regions were sorghum- and wheat *rps4*, gamagrass *rps2A* and gamagrass *cox1* (Table S1). The two minisatellites in *rps4* were different in their sequences but both occurred in the less conserved internal ORF of angiosperm *rps4* (Itchoda et al. 2002). The minisatellite in gamagrass *rps2A* was located in

the C-terminal extension of the ORF, a region known to be not conserved among maize, rice and wheat (Perrotta et al. 2002). In maize, the C-terminal extension of *rps2A* is translated but cut away from the mature RPS2 polypeptide (Perrotta et al. 2002). The gamagrass *cox1* ORF has a unique C-terminal extension, where the minisatellite was located. Apparently independent minisatellite-like structures also occur in the carrot- and the sorghum C-terminal extensions (Bailey-Serres et al. 1986; Robison et al. 2006). Previously, the mature translation products of carrot and sorghum *cox1* were shown to exclude or include the C-terminal extension, respectively (Bailey-Serres et al. 1986; Robison et al. 2006). It is unclear whether gamagrass *cox1* is more similar to carrot or sorghum *cox1* because no gene expression analysis was conducted. The three intronic regions containing minisatellites were the *nad1*-intron 3 of seven species, rapeseed *nad2*-intron 4 and grapevine *nad7*-intron 3. The minisatellite in the *nad1*-intron 3 is conserved in five grass-family species, grapevine and onion but does not occur in papaya, tobacco, *Arabidopsis* or sugar beet (Fig. S1). The minisatellite in the rapeseed *nad2*-intron 4 is a non-duplicated form in *Arabidopsis* and the corresponding sequence appears to be missing in the other eight plants listed in Table 4 (data not shown). No corresponding sequence of the grapevine-*nad7*-intron 3 minisatellite was found in the other nine plants in Table 4. Of the remaining 72 minisatellites, two pairs, *Arabidopsis AtTR1* and rapeseed *BnTR1* and *Arabidopsis AtTR4* and rapeseed *BnTR4*, were identical in terms of the nucleotide sequence of the repeat unit but the number of iterations differed (Table S1). Papaya *CcTR2* and grapevine *VvTR10*, and gamagrass *TdTR4* and maize *ZmTR4* were the pairs of repeated sequences where the nucleotide sequence of the repeat unit and the number of iterations was identical (Table S1). Repeat units of another twenty-one minisatellites had their homologous sequence as a non-duplicated form in at least one other plant species. The nucleotide sequences of the other 43 minisatellites were unique to each plant species.

The most prevalent number of iterations was 2 (in 74 cases), followed by 3 (in 6), 4 (in 3) and 5 (in 1). The length of repeat units ranged from 20 to 72 bp, of which 67 units were in the range of 22 to 62 bp. Of the 84 minisatellites, 60 exhibited an organization with short direct repeats that were homologous to the beginning of the repeat unit immediately following the array (Table S1; see examples in Figs 1 and 2). Such an organization has been proposed to be involved in the initial duplication of the repeat unit in yeast and

humans (Haber and Louis 1998). The number of minisatellites per mitochondrial DNA ranged from 1 (tobacco) to 15 (*Arabidopsis*), and there seemed to be no correlation between the number of minisatellites and the size of mitochondrial DNA (Table 4).

Mitochondrial minisatellites in *Oryza*

As a model, we next examined the length polymorphism of the identified minisatellites in *Oryza*. There were six available mitochondrial sequences in *Oryza*: two from *O. sativa* Japonica cultivars, three were *O. sativa* Indica cultivars and one was an accession of *O. rufipogon*, which is a close relative of *O. sativa* (Fujii et al. 2010; Notsu et al. 2002; Tian et al. 2006). We searched the five additional *Oryza* mitochondrial genome sequences for counterparts of the five minisatellites identified in the cultivar Nipponbare (Table 4) and compared their nucleotide sequences. The result showed that all five minisatellites were preserved, and the number of repeat units was unchanged in each locus (data not shown).

In the genus *Oryza*, all the species are classified according to their chromosome complements (genomes) such as AA, BB, BBCC, CC, CCDD, EE, FF, GG and HHJJ (Vaughan and Morishima 1999; Vaughan et al. 2003). This classification system is well supported by cytogenetic- and molecular phylogenetic studies using nuclear and/or chloroplast DNAs. Because all six of the above-mentioned mitochondrial sequences belonged to an AA-genome species, another species with a different genome types remained to be examined. A core collection comprised of 42 accessions from 18 *Oryza* species, covering nine genome types, was evaluated (Table 1). PCR primers targeting the five minisatellites were designed based on the Nipponbare sequence and used for PCR analysis. When *OsTR3*, one of the five minisatellites (see Table S1 and Fig. S1), was targeted, PCR fragments of the expected size were obtained from all 42 plants and there was no apparent polymorphism (data not shown). On the other hand, PCR analysis of the other four loci exhibited length polymorphism of the amplicons (Fig. 1).

OsTR1 is a minisatellite composed of two repeats in the Nipponbare sequence (Table S1 and Fig. 1).

We obtained two kinds of amplicons, one was the same size as the Nipponbare sequence (21 plants) and the other was smaller (16 plants). No clear amplification was observed in 5 plants (Fig. 1a). To examine the organizational difference of the amplicons, nucleotide sequences of the PCR products of accessions W1020 and W0002 were determined. Comparison of the two sequences revealed that one repeat copy was missing in W0002 (Fig. 1b).

The repeat organization of *OsTR2* in the Nipponbare sequence is two copies of 23-bp repeats followed by a divergent repeat copy of 33 bp (Table S1 and Fig. 1). Two kinds of amplicons appeared on the polyacrylamide gel, one was expected from the Nipponbare sequence (15 plants) and the other was smaller (2 plants). No clear amplification was observed for 25 plants. We sequenced the PCR products from W2003 and W1401 and found that the 33-bp repeat copy was missing from W1401 (Fig. 1c).

All 42 plants yielded PCR products corresponding to *OsTR4*, but the PCR products of three plants were smaller than the expected size of the Nipponbare sequence (Fig. 1a). Nucleotide sequencing of the W0002 and W0003 amplicons revealed that a 32-bp repeat copy was missing in W0003 as well as a 4-bp deletion in the remaining repeat copy (Fig. 1d).

PCR analysis of *OsTR5* gave amplicons with the expected size of the Nipponbare sequence (16 plants), smaller amplicons (21 plants), and no amplicons were observed for 5 plants (Fig. 1a). The W0002 amplicon lacked a 20-bp repeat copy compared to the W1020 amplicon (Fig. 1e).

Mitochondrial minisatellites in *Brassica*

We selected *Brassica* as another model to examine the polymorphism of mitochondrial minisatellites because the relationships between six major cultivated species of *Brassica* have been established (U 1935). Total cellular DNAs were isolated from a total of 29 accessions, comprising the six major *Brassica* cultivated species (Table 2). Six sets of PCR primers were designed to target the mitochondrial minisatellites based on the mitochondrial DNA of *Brassica napus* cultivar Westar (Table S1). As a result of

PCR analysis of the 29 accessions, we obtained PCR products from all the plants with each of the primer sets, but no length polymorphism was apparent on the gel in the case of *BnTR2*, *BnTR3*, *BnTR5* or *BnTR6* (Fig. 2 and data not shown). The remaining two polymorphic minisatellites were investigated further.

BnTR1 consisted of 29-bp repeat copies with a 33-bp intervening sequence, and the number of repeat copies was four in Westar (Table S1). The PCR products of accession C-701, whose size was the same as expected from the Westar sequence, and O-141, whose amplicon was smaller (Fig. 2a), were sequenced and compared. As a result, the lack of a repeat copy from the O-141 sequence was revealed (Fig. 2b).

BnTR4 is a tandem array of 31-bp repeat units and a divergent copy, reiterated five times in Westar (Table S1). The sizes of the obtained PCR products were ~0.3, ~0.25 and ~0.22 kbp on the polyacrylamide gel (Fig. 2a), of which the 0.3-kbp amplicon was expected from the Westar sequence. The nucleotide sequence of the PCR products from N-118, O-141 and C-701 revealed a different number of repeat copies, which was five, three and two, respectively (Fig. 2c).

Discussion

Minisatellites occurred more frequently in regions with less functional significance in angiosperm mitochondrial DNAs. Identification of mitochondrial minisatellites is possible if a reference sequence is available. It should be noted that the number of informative mitochondrial minisatellites in a plant may be insufficient for use as markers because angiosperm minisatellites are not always polymorphic and, if they exist, the number of alleles may be small or some specimen may exhibit null alleles. However, mitochondrial minisatellites have the potential to increase the robustness of an analysis when combined with nuclear and chloroplast markers.

The number of repeat units in all the *Oryza* mitochondrial sequences and *Oryza* accessions with an AA genome examined in this study were 2, 3, 2 and 2 for *OsTR1*, *OsTR2*, *OsTR4* and *OsTR5*, respectively. Hereafter, haplotypes based on the mitochondrial minisatellites are coded in parenthesis. Therefore, the

haplotype of AA-genome accessions is coded as (2, 3, 2, 2) (see Table 1). The BB-genome accessions examined in this study possessed a null allele at *OsTR2*, and its haplotype code is (2, -, 2, 2). Accordingly, all the CC-, BBCC- and CCDD-genome accessions were coded as (1, -, 2, 1), which can be distinguished from the AA- or BB-genome accessions at *OsTR1* and *OsTR5*. This result is consistent with the notion that during the formation of these amphidiploid species the cytoplasmic donor parent of BBCC- and CCDD-genome species may have been a CC-genome species (Nishikawa et al. 2005). However, the code of EE-genome accessions cannot be distinguished from the HHJJ-genome accessions, a result that was not supported by previous studies (e. g. Nishikawa et al. 2005) where EE-genome accessions were grouped into an *O. officinalis* complex containing BB-, CC-, BBCC-, CCDD- and EE-genome accessions. It is possible that the two identical codes occurred by chance; however, further investigation of *Oryza* mitochondrial DNAs will be necessary. Because of the occurrence of more than two null alleles, the minisatellite set consisting of *OsTR1* to *OsTR5* was less informative for the FF- and GG-genome accessions.

In the case of *Brassica*, five distinctive haplotypes based on two mitochondrial minisatellites were apparent (Table 2). In *Brassica rapa*, for example, the numbers of repeat units in *BnTR1* and *BnTR4* were four and two, respectively, and they were coded as (4, 2). Of the six major cultivated *Brassica* species, the genomes of three diploid species, *Brassica rapa*, *Brassica nigra* and *Brassica oleracea*, have been designated as AA, BB and CC, respectively (U 1935). As shown in Table 2, the haplotype codes (4, 2), (3, 2) and (3, 3) corresponded to the AA-, BB- and CC-genome accessions, respectively. Of particular interest is that the haplotype codes of two amphidiploid accessions, *Brassica juncea* (AABB-genome species) and *Brassica carinata* (BBCC-genome species), were (4, 2) and (3, 2), respectively. These results are consistent with the notion that the cytoplasmic donor of the AABB-genome species was an AA-genome species and that the BBCC-genome species donor was a BB-genome species, a hypothesis well supported by chloroplast DNA analyses (Erickson et al. 1983; Palmer, et al. 1983). On the other hand, none of the two haplotype codes of the *Brassica napus* (AACC-genome species) accessions matched with the other five *Brassica* species. According to chloroplast DNA analyses, there were multiple chloroplast types in *Brassica napus* accessions and such chloroplast DNA types were very few in diploid *Brassica* accessions

(Song, et al. 1988; Flannery et al. 2006; Allender and King 2010).

It should be noted that two mitochondrial minisatellites of *Brassica*, *BnTR1* and *BnTR4*, exhibited their potential as mitochondrial markers (Table 2). Such mitochondrial markers can be applied to explore the mitochondrial diversity of *Brassica* genetic resources. Another potential application is the diagnosis of the heteroplasmic state of mitochondria, for example, in cybrids that were produced by fusion of two protoplasts with different mitochondrial types. Unlike animal mitochondria (Lunt et al. 1998), there was no clear evidence of heteroplasmy in our plant materials (Figs. 1 and 2) (Nishizawa et al. 2000), but heterogeneous PCR products targeting a mitochondrial minisatellite were seen when a *Brassica-Arabidopsis* cybrid was evaluated (our unpublished result).

Why was the level of polymorphism in *Oryza* or *Brassica* lower than that of *Beta vulgaris*? Intraspecific variation is rare in the former two but frequent in the latter genus: for example, a minisatellite locus of *Beta vulgaris* named TR1 has at least 6 alleles whose iteration numbers are 4, 5, 6, 7, 10 and 13 (Cheng et al. 2011). Preliminary tests on the intraspecific variation of radish (*Raphanus sativus*) using PCR primers for *BnTR1* to *BnTR6* revealed that, although PCR products were obtained with some primer sets, there were no length polymorphisms among nine accessions with various mitochondrial types (our unpublished result). It remains unknown whether the trends in mitochondrial minisatellite evolution differ among plant species or among loci. In gymnosperms, mitochondrial minisatellites showing intraspecific variation have been reported (Bastien et al. 2003; Godbout et al. 2005), but there is no information about other loci as far as we know. Mechanisms involved in the organizational polymorphism of minisatellites in *Drosophila* and yeast nuclear DNAs have been proposed such as DNA recombination including gene-conversion like activity rather than replication error, (Richard and Paques 2000), but mechanisms explaining plant mitochondrial minisatellites are unclear. The identity of factors accelerating variation in the number of repeat units is an open question.

Acknowledgements

The authors wish to thank Prof. Dr. Takeshi Nishio and the National Institute of Genetics for providing plant materials, and Prof. Dr. Hiroshi Yamagishi for valuable comments. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan; the Private University Strategic Research Foundation Support Program; and the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (BRAIN).

References

Allender CJ, King GJ (2010) Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. BMC Plant Biol 10: 54

Bailey-Serres J, Hanson DK, Fox TD, Leaver CJ (1986) Mitochondrial genome rearrangement leads to extension and relocation of the cytochrome c oxidase subunit I gene in sorghum. Cell 47: 567-576

Bastien D, Favre JM, Collignon AM, Sperisen C, Jeandroz S (2003) Characterization of a mosaic minisatellite locus in the mitochondrial DNA of Norway spruce *Picea abies* (L.) Karst. Theor Appl Genet 107: 574-580

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucl Acids Res 27: 573-580

Cafasso D, Pellegrino G, Musacchio A, Widmer A, Cozzolino S (2001) Characterization of a minisatellite repeat locus in the chloroplast genome of *Orchis palustris* (Orchidaceae). Curr Genet 39: 394-398

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797

Cheng D, Kitazaki K, Xu DC, 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, Xu D, Mikami T, Kubo T (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

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

Erickson LR, Straus NA, Beversdorf WD (1983) Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphiploids. Theor Appl Genet 65: 201-206

Fenart S, Arnaud JF, de Cauwer I, Cuguen J (2008) Nuclear and cytoplasmic genetic diversity in weed beet and sugar beet accessions compared to wild relatives: new insights into the genetic relationships within the *Beta vulgaris* complex species. Theor Appl Genet 116: 1063-1077

Fievet V, Touzet P, Arnaud JF, Cuguen J (2007) Spatial analysis of nuclear and cytoplasmic DNA diversity in wild sea beet (*Beta vulgaris* ssp *maritima*) populations: do marine currents shape the genetic structure? Mol Ecol 16: 1847-1864

Flannery ML, Mitchell FJG, Coyne S, Kavanagh TA, Burke JI, Salamin N, Dowding P, Hodkinson TR (2006) Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. Theor Appl Genet 113: 1221-1231

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: 15

Godbout J, Jaramillo-Correa JP, Beaulieu J, Bousquet J (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. Mol Ecol 14: 3497-3512

Haber JE, Louis EJ (1998) Minisatellite origins in yeast and humans. Genomics 48: 132-135

Itchoda N, Kubo T, Estiati A, Yamamoto MP, Handa H, Mikami T (2002) The *rps4* gene in sugar beet mitochondria: insertion/deletion mutations occur within the gene but do not disrupt the reading frame. J Plant Physiol 159: 211-215

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable minisatellite regions in human DNA. Nature 314: 67-73

King RA, Ferris C (2002) A variable minisatellite sequence in the chloroplast genome of *Sorbus* L. (Rosaceae : Maloideae). Genome 45: 570-576

Kitazaki K, Nomoto Y, Aoshima A, Mikami T, Kubo T (2009) A mitochondrial gee involved in cytochrome *c* maturation (*ccmC*) is expressed as a precursor with a long NH_2-terminal extension in sugar beet. J Plant Physiol 166: 775-780

Kitazaki K, Kubo T (2010) Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. J Botany 2010: Article ID: 620137

Kubo T, Newton KJ (2008) Angiosperm mitochondrial genomes and mutations. Mitochondrion 8: 5-14

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol Ecol 7: 1441-1455

Nishikawa T, Vaughan DA, Kadowaki K (2005) Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. Theor Appl Genet 110: 696-705

Nishizawa S, Kubo T, Mikami T (2000) Variable number of tandem repeat loci in the mitochondrial genomes of beets. Curr Genet 37: 34-38

Nishizawa S, Mikami T, Kubo T (2007) Mitochondrial DNA phylogeny of cultivated and wild beets: relationships among cytoplasmic male-sterility-inducing and nonsterilizing cytoplasms. Genetics 177: 1703-1712

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai 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 Gen Genomics 268: 434-445

Palmer JD, Shields CR, Cohen DB, Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65: 181-189

Perrotta G, Grienenberger JM, Gualberto JM (2002) Plant mitochondrial *rps2* genes code for proteins with a C-terminal extension that is processed. Plant Mol Biol 50: 523-533

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence

repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92: 7759-7763

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. J Mol Evol 37: 281-295

Richard GF, Paques F (2000) Mini- and microsatellite expansions: the recombination connection. EMBO Rep 1: 122-126

Robison MM, Wolyn DJ (2006) A 60 kDa COX1 protein in mitochondria of carrot irrespective of the presence of C-terminal extensions in the *cox1* reading frame. Mol Gen Genomics 275: 68-73

Skinner DZ (2000) Non random chloroplast DNA hypervariability in *Medicago sativa*. Theor Appl Genet 101: 1242-1249

Song KM, Osborn TC, Williams PH (1988) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPS) 1 Genome evolution of diploid and amphidiploid species. Theor Appl Genet 75: 784-794

Tian XJ, Zheng J, Hu SN, Yu J (2006) The rice mitochondrial genomes and their variations. Plant Physiol 140: 401-410

U N (1935) Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilisation. Jap J Botany 7: 389-452

Vaughan DA, Morishima H (1999) Biosystematics of the genus *Oryza*. In: Smith CW, Dilday RH (eds) Rice: origin, history, technology, and production. John Wiley and Sons, New York, pp. 27-65.

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol 6: 139-146

Table 1 *Oryza* accessions used in this study

Species[a]	**Genome type**	**Accession name**	**Number of repeat units**			
			OsTR1	***OsTR2***	***OsTR4***	***OsTR5***
O. barthii	AA	W0652	2	3	2	2
		W1588	2	3	2	2
O. glumaepatula	AA	W1169	2	3	2	2
		W2145	2	3	2	2
		W2199	2	3	2	2
O. longistaminata	AA	W1413	2	3	2	2
		W1508	2	3	2	2
O. meridionalis	AA	W1625	2	3	2	2
		W1635	2	3	2	2
O. rufipogon	AA	W0106	2	3	2	2
		W0120	2	3	2	2
		W1294	2	3	2	2
		W1866	2	3	2	2
		W1921	2	3	2	2
		W2003	2	3	2	2
O. punctata (2X)	BB	W1514	2	-[b]	2	2
O. minuta	BBCC	W1213	1	-	2	1
		W1331	1	-	2	1
O. punctata (4X)	BBCC	W1024	1	-	2	1
O. eichingeri	CC	W1527	1	-	2	1
O. eichingeri or *O. rhizomatis*	CC	W1805	1	-	2	1
O. officinalis	CC	W0002	1	-	2	1
		W1361	1	-	2	1
		W1830	1	-	2	1
O. alta	CCDD	W0017	1	-	2	1
O. alta or O. latifolia	CCDD	W1182	1	-	2	1
O. grandiglumis	CCDD	W0613	1	-	2	1
		W1194	1	-	2	1
		W2220	1	-	2	1
O. latifolia	CCDD	W1166	1	-	2	1
		W1197	1	-	2	1
		W2200	1	-	2	1
O. australiensis	EE	W0008	2	-	2	1
		W1628	2	-	2	1
O. longiglumis	HHJJ	W1220	2	-	2	1

O. ridleyi	HHJJ	W0001	2	-	2	1
		W0604	2	-	2	1
O. brachyantha	FF	W1401	-	2	2	-
		W1711	-	2	2	-
O. granulata	GG	W0003	-	-	1	-
		W0067(B)	-	-	1	-
O. meyeriana	GG	W1356	-	-	1	-

[a]Descriptions followed the web site of the National Institute of Genetics, Japan; [b]Null allele

Table 2 *Brassica* accessions used in this study

Species	**Genome type**	**Accession name**	**Number of repeat unit**	
			BnTR1	*BnTR4*
Brassica rapa	AA	C-701	4	2
		C-656	4	2
		C-504	4	2
		C-482	4	2
		C-256	4	2
		C-333	4	2
Brassica napus	AACC	N-106	4	5
		N-107	4	3
		N-118	4	5
		N-263	4	5
		N-348	4	5
Brassica oleracea	CC	O-141	3	3
		O-145	3	3
		O-146	3	3
		O-162	3	3
		O-204	3	3
Brassica carinata	BBCC	Ca-112	3	2
		Ca-114	3	2
		Ca-116	3	2
Brassica nigra	BB	Ni-111	3	2
		Ni-116	3	2
		Ni-130	3	2
		Ni-131	3	2
		Ni-138	3	2
Brassica juncea	AABB	J-112	4	2
		J-113	4	2
		J-139	4	2
		J-473	4	2
		J-474	4	2

Table 3 Nucleotide sequences of PCR primers used in this study

Targets	Primer name	Nucleotide sequences	Annealing temperature (˚C)
OsTR1	OsTR1-Fw	5'-CAACCGGTTCTCTTCGTCTATC-3'	58
	OsTR1-Rv	5'-TGTAACTAGAGGGCGAGACTAG-3'	
OsTR2	OsTR2-Fw	5'-TCACTTTCCGCTTATCGGTAGG-3'	58
	OsTR2-Rv	5'-ATATGCACACCGTACCGACTTG-3'	
OsTR3	OsTR3-Fw	5'-ACCCTACGGTGCTGGTAAGG-3'	55
	OsTR3-Rv	5'-CTGGGGAACCATCACAAGTA-3'	
OsTR4	OsTR4-Fw	5'-CAGTCACCCGTATAGTACCA-3'	54
	OsTR4-Rv	5'-CGGTGCTGAATTCTCTGGCT-3'	
OsTR5	OsTR5-Fw	5'-ATTGGATTACTAGTCTCGCCCTC-3'	58
	OsTR5-Rv	5'-GGCAACAGAGCAATAGGCATTG-3'	
BnTR1	BnTR1-Fw	5'-CCGTTAGGGGTATTTAGTAACTCG-3'	56
	BnTR1-Rv	5'-ACATAATGGCAATGTATCGGACTG-3'	
BnTR2	BnTR2-Fw	5'-TGATAACAGTTTCCTCCTAGTTTGC-3'	56
	BnTR2-Rv	5'-GAACCTGTTAATTAGCACGGAACTA-3'	
BnTR3	BnTR3-Fw	5'-GAGTGCTGGCTGTTATAGTATGGTT-3'	56
	BnTR3-Rv	5'-ATAGGAACAGAAAGCTACGCTAACAA-3'	
BnTR4	BnTR4-Fw	5'-GAAGTCCGAGGACCTTTAGTACC-3'	56
	BnTR4-Rv	5'-AGTAAGTTGTAGGTAGGGGCTTCAT-3'	
BnTR5	BnTR5-Fw	5'-CTTACAGTCGAGCTCCTTTGTCACT-3'	56
	BnTR5-Rv	5'-GTAACTCGACCCCTCATCAACTAAT-3'	
BnTR6	BnTR6-Fw	5'-AGTATATTGACAGTGCCCCAAGAC-3'	56
	BnTR6-Rv	5'-AGTTACTCGACTGAAAAGGAGAGGT-3'	

Table 4 Number of minisatellite-like repeated sequence loci identified in this study

Scientific name	Common name	Size of mitochondrial DNA	DDBJ/EMBL/GenBank accession no.	No. of minisatellites			
				Gene-coding region	Intron	Intergenic region	Total
Arabidopsis thaliana	Thale cress	366924 bp	Y08501	0	0	15	15
Brassica napus	Rapeseed	221853 bp	AP006444	0	1	5	6
Carica papaya	Papaya	476890 bp	EU431224	0	0	3	3
Nicotiana tabacum	Tobacco	430597 bp	BA000042	0	0	1	1
Vitis vinifera	Grapevine	773279 bp	FM179380	0	2	10	12
Oryza sativa Japonica Group	Rice	490520 bp	BA000029	0	1	4	5
Sorghum bicolor	Sorghum	468628 bp	DQ984518	1	1	9	11
Tripsacum dactyloides	Gamagrass	704100 bp	DQ984517	2	1	10	13
Triticum aestivum	Wheat	452528 bp	AP008982	1	1	6	8
Zea mays subsp. *mays*	Maize	569630 bp	AY506529	0	1	9	10

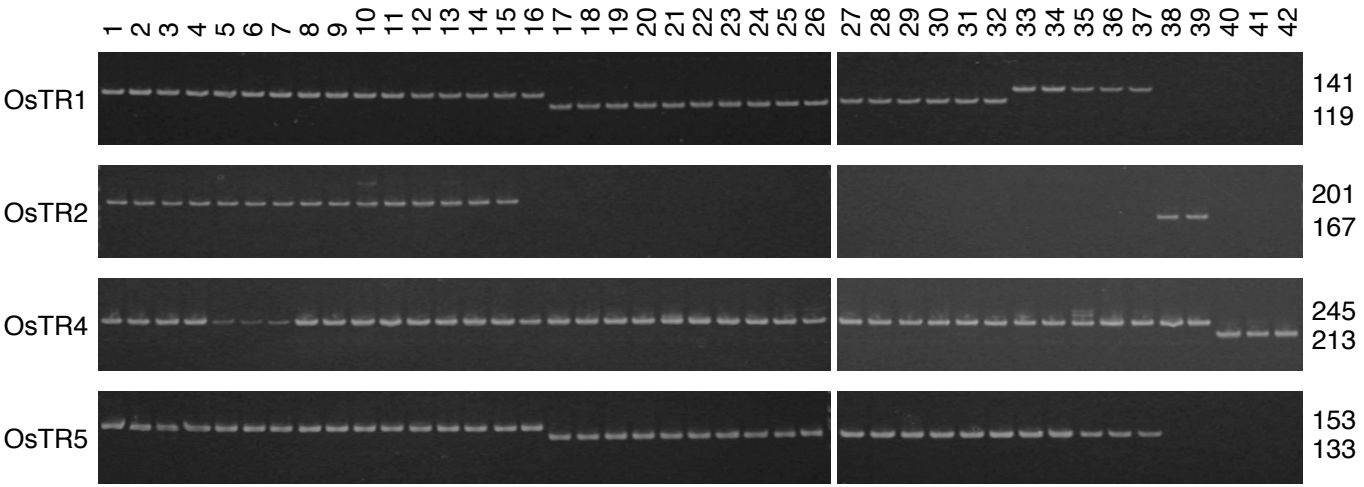
Figure legends

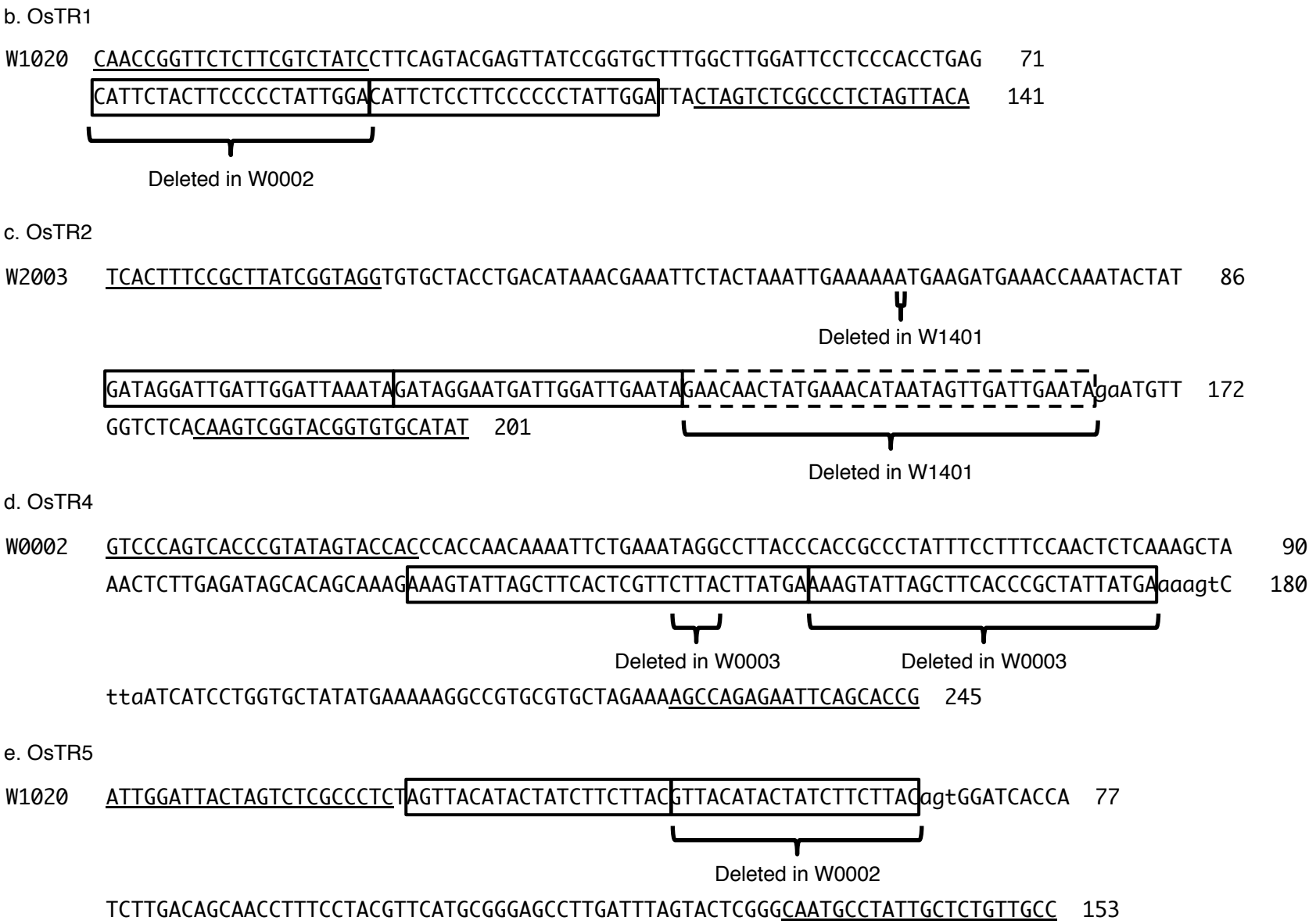
Fig. 1 **a.** Gel electrophoresis of PCR amplification products targeting *OsTR1*, *OsTR2*, *OsTR4* and *OsTR5* using 42 *Oryza* accessions. Sizes of the PCR products are shown on the right. Accession names are as follows: 1, W0652; 2, W1588; 3, W1169; 4, W2145; 5, W2199; 6, W1413; 7, W1508; 8, W1625; 9, W1635; 10, W0106; 11, W0120; 12, W1294; 13, W1866; 14, W1921; 15, W2003; 16, W1514; 17; W1213;18, W1331; 19, W1024; 20, W1527; 21, W1805; 22, W0002; 23, W1361; 24, W1830; 25, W0017; 26, W1182; 27, W0613; 28, W1194; 29, W2220; 30, W1166; 31, W1197; 32, W2200; 33, W0008; 34, W1628; 35, W1220; 36, W0001; 37, W0604; 38, W1401; 39, W1711; 40, W0003; 41, W0067(B); 42, W1356. **b-c.** Nucleotide sequences of representative PCR products in panel a. Positions of PCR primers are underlined. Repeated sequences are in boxes with solid lines. The divergent repeat copy is enclosed in a box with dashed lines. Short flanking repeats are indicated by lower case letters. Extents of deletion are shown by brackets. **b.** Nucleotide sequence comparison of *OsTR1* between W1020 and W0002. **c.** Nucleotide sequence comparison of *OsTR2* between W2003 and W1401. **d.** Summary of nucleotide sequence comparison of *OsTR4* between W0002 and W0003. **e.** Nucleotide sequence comparison of *OsTR5* between W1020 and W0002

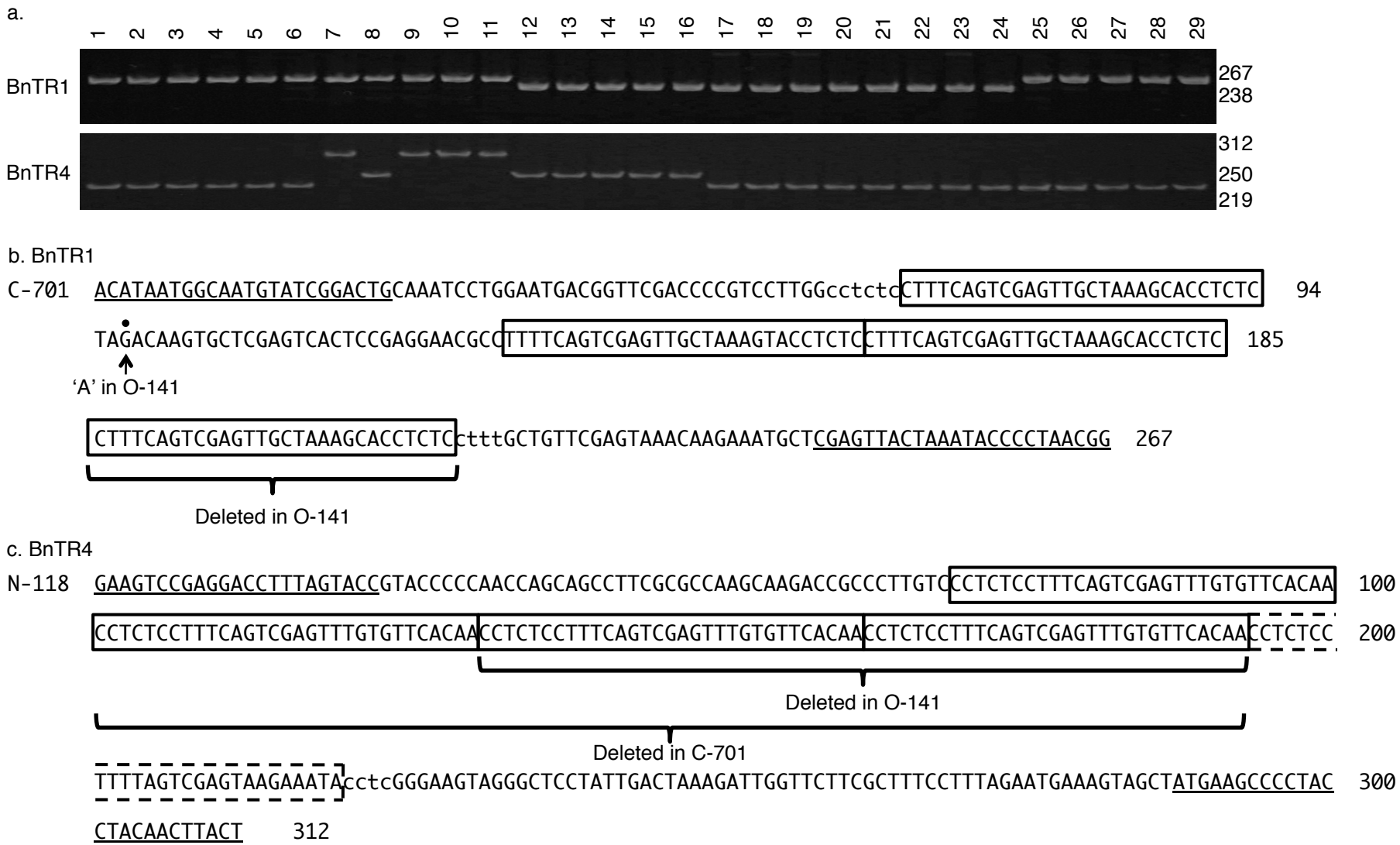
Fig. 2 Gel electrophoresis of PCR amplification products targeting *BnTR1* and *BnTR4* using 29 *Brassica* accessions. Sizes of the PCR products are shown on the right. Accession names are as follows: 1, C-701; 2, C-656; 3, C-504; 4, C-482; 5, C-256; 6, C-333; 7, N-106; 8, N-107; 9, N-118; 10, N-263; 11, N-348; 12, O-141; 13, O-145; 14, O-146; 15, O-162; 16, O-204; 17, Ca-112; 18, Ca-114; 19, Ca-116; 20, Ni-111; 21, Ni-116; 22, Ni-130; 23, Ni-131; 24, Ni-138; 25, J-112; 26, J-113; 27, J-139; 28, J-473; 29, J-474. **b-c.** Nucleotide sequence of representative PCR products in panel a. Positions of PCR primers are underlined. Repeated sequences are in boxes with solid lines. Divergent repeats are in boxes with dashed lines. Short flanking repeats are indicated by lower case letters. Extents of deletion are shown by brackets. **b.** Nucleotide sequence comparison of *BnTR1* between C-701 and O-141. A nucleotide substitution occurring in O-141 is indicated by a dot. **c.** Nucleotide sequence comparison of *BnTR4* among N-118,

O-141 and C-701

a.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42
OsTR1
141
119
OsTR2
201
167
OsTR4
245
213
OsTR5
153
133


b. OsTR1
W1020 CAACCGGTTCTCTTCGTCTATCCTTCAGTACGAGTTATCCGGTGCTTTGGCTTGGATTCCTCCCACCTGAG 71
CATTCTACTTCCCCCTATTGGACATTCTCCTTCCCCCCTATTGGATTACTAGTCTCGCCCTCTAGTTACA 141
Deleted in W0002
c. OsTR2
W2003 TCACTTTCCGCTTATCGGTAGGTGTGCTACCTGACATAAACGAAATTCTACTAAATTGAAAAAATGAAGATGAAACCAAATACTAT 86
Deleted in W1401
GATAGGATTGATTGGATTAAATAGATAGGAATGATTGGATTGAATAGAACAACTATGAAACATAATAGTTGATTGAATAgaATGTT 172
GGTCTCACAAGTCGGTACGGTGTGCATAT 201
Deleted in W1401
d. OsTR4
W0002 GTCCCAGTCACCCGTATAGTACCACCCACCAACAAAATTCTGAAATAGGCCTTACCCACCGCCCTATTTCCTTTCCAACTCTCAAAGCTA 90
AACTCTTGAGATAGCACAGCAAAGAAAGTATTAGCTTCACTCGTTCTTACTTATGAAAAGTATTAGCTTCACCCGCTATTATGAaaagtC 180
Deleted in W0003
Deleted in W0003
ttaATCATCCTGGTGCTATATGAAAAAGGCCGTGCGTGCTAGAAAAGCCAGAGAATTCAGCACCG 245
e. OsTR5
W1020 ATTGGATTACTAGTCTCGCCCTCTAGTTACATACTATCTTCTTACGTTACATACTATCTTCTTACagtGGATCACCA 77
Deleted in W0002
TCTTGACAGCAACCTTTCCTACGTTCATGCGGGAGCCTTGATTTAGTACTCGGGCAATGCCTATTGCTCTGTTGCC 153


a.
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
BnTR1
267
238
BnTR4
312
250
219
b. BnTR1
C-701 ACATAATGGCAATGTATCGGACTGCAAATCCTGGAATGACGGTTCGACCCCGTCCTTGGcctctcCTTTCAGTCGAGTTGCTAAAGCACCTCTC 94
TAGACAAGTGCTCGAGTCACTCCGAGGAACGCCTTTTCAGTCGAGTTGCTAAAGTACCTCTCCTTTCAGTCGAGTTGCTAAAGCACCTCTC 185
'A' in O-141
CTTTCAGTCGAGTTGCTAAAGCACCTCTCctttGCTGTTCGAGTAAACAAGAAATGCTCGAGTTACTAAATACCCCTAACGG 267
Deleted in O-141
c. BnTR4
N-118 GAAGTCCGAGGACCTTTAGTACCGTACCCCCAACCAGCAGCCTTCGCGCCAAGCAAGACCGCCCTTGTCCCTCTCCTTTCAGTCGAGTTTGTGTTCACAA 100
CCTCTCCTTTCAGTCGAGTTTGTGTTCACAACCTCTCCTTTCAGTCGAGTTTGTGTTCACAACCTCTCCTTTCAGTCGAGTTTGTGTTCACAACCTCTCC 200
Deleted in O-141
Deleted in C-701
TTTTAGTCGAGTAAGAAATAcctcGGGAAGTAGGGCTCCTATTGACTAAAGATTGGTTCTTCGCTTTCCTTTAGAATGAAAGTAGCTATGAAGCCCCTAC 300
CTACAACTTACT 312