Theor Appl Genet (2015) 128:2289–2299
DOI 10.1007/s00122-015-2584-z

ORIGINAL ARTICLE

Identification of candidate genes associated with fertility restoration of cytoplasmic male-sterility in onion (*Allium cepa* L.) using a combination of bulked segregant analysis and RNA-seq

Sunggil Kim[1] · Cheol-Woo Kim[2] · Minkyu Park[3] · Doil Choi[3]

Received: 17 April 2015 / Accepted: 16 July 2015 / Published online: 28 July 2015

Abstract

Key message **A combination of BSA and RNA-seq was performed to identify candidates for the restorer-of-fertility gene in onion. The *AcPMS1* involved in DNA mismatch repair was identified as the best candidate.**

Abstract To identify candidate genes of the restorer-of-fertility gene (*Ms*) responsible for fertility restoration of onion cytoplasmic male-sterility, a combined approach of bulked segregant analysis and RNA-seq was employed. From 32,674 de novo assembled contigs, 430 perfectly homozygous SNPs between male-fertile (MF) and male-sterile (MS) bulks were identified in 141 contigs. After verifying the homozygosity of the SNPs by PCR amplification and sequencing, the SNPs on 139 of the contigs were genotypes for the two recombinants which contained crossover events between the *Ms* locus and two tightly linked molecular markers. As a result, 30 contigs showing perfect linkage with the *Ms* locus in the large-sized segregating population were identified. Among them, 14 showed perfect linkage disequilibrium (LD) with the *Ms* locus, as determined by genotyping 251 domestic breeding lines. Furthermore, molecular markers tagging the 14 contigs also showed almost perfect LD with each other in 124 exotic accessions introduced from 21 countries, except for one accession which contained a crossover event by which the 14 markers were divided into two groups. After sequencing of the full-length cDNA of the 14 contigs showing perfect LD, the deduced amino acids sequences of the MF and MS alleles were compared. Four genes were shown to harbor putative critical amino acid changes in the known domains. Among them, the gene encoding PMS1, involved in the DNA mismatch repair pathway, was assumed to be the best candidate gene responsible for fertility restoration of male-sterility in onion.

Communicated by M. Xu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2584-z) contains supplementary material, which is available to authorized users.

✉ Sunggil Kim
dronion@jnu.ac.kr

[1] Department of Plant Biotechnology, Biotechnology Research Institute, Chonnam National University, Kwangju 500-757, Korea

[2] National Institute of Horticultural & Herbal Science, RDA, Muan 534-833, Korea

[3] Department of Plant Science, Seoul National University, Seoul, Korea

Introduction

Cytoplasmic male-sterility (CMS) is a maternally inherited trait conferring failure in production of viable pollen without having a significant effect on female fertility. Although the selective advantage of CMS in evolution remains unclear (Hanson and Bentolila 2004), CMS has been adopted by more than 140 plant species as a successful reproductive strategy in natural populations (Laser and Lersten 1972). To date, all known naturally occurring cases of CMS have arisen by the creation of aberrant genes in the mitochondrial genomes (Hu et al. 2014).

Several peculiar features of the plant mitochondrial genomes were assumed to be responsible for the occurrence of CMS induction (Schnable and Wise 1998; Budar et al. 2003; Knoop 2004; Kubo and Newton 2008). In contrast to the small and stable animal mitochondrial genomes, their plant counterparts contain relatively large

genomes of variable size, ranging from approximately 200 kb in *Brassica* species (Palmer and Herbon 1987) to 11,319 kb in *Silene conica* (Sloan et al. 2012). In addition to the variable size, the genome organization is also highly variable due to dynamic rearrangements (Palmer 1988; Park et al. 2013b). Even the structure of plant mitochondrial genomes has not yet been unanimously defined (Backert et al. 1997; Oldenburg and Bendich 2001; Allen et al. 2007).

Frequent repeat sequence-mediated recombination leads to dynamic rearrangements (Palmer 1988; Small et al. 1989; Albert et al. 1998; Woloszynska and Trojanowski 2009). Relatively short (<1 kb) repeats can cause genome rearrangements, while large repeats are responsible for multipartite formation, through which diverse subgenomes are created. The copy numbers and stoichiometry of subgenomes are also variable, even at the intraspecific level. A specific stoichiometry of the subgenomes is normally maintained throughout generations, but the stoichiometry can sometimes be changed by a mechanism called substoichiometric shifting (Sakai and Imamura 1993; Bellaoui et al. 1998; Janska et al. 1998; Arrieta-Montiel et al. 2001; Kim et al. 2007). The copy number increase of a low-copy-number subgenome containing a CMS-inducing gene by substoichiometric shifting could result in the induction of CMS for a short period of time (Janska et al. 1998). Tissue culture (Kanazawa et al. 1994) and the mutation of certain nuclear genes are known to induce substoichiometric shifting in plants. The *Msh1* (Abdelnoor et al. 2006) and *RecA* (Shedge et al. 2007) genes involved in the DNA repair pathway, as well as OSB1 (Zaegel et al. 2006), one of the single-stranded DNA binding proteins, were all reported to be involved in suppression of mitochondrial genome rearrangements.

Male-sterility induced by aberrant mitochondrial genes is often reverted by nuclear restorer-of-fertility (Rf) genes in many plant species. Since the first isolation of an Rf gene in maize (Cui et al. 1996), several Rf genes have been cloned in other plant species. Four Rf genes were revealed to encode pentatricopeptide repeat (PPR) proteins (Bentolila et al. 2002; Brown et al. 2003; Desloire et al. 2003; Koizuka et al. 2003; Komori et al. 2004; Klein et al. 2005), but other functions, such as an aldehyde dehydrogenase (Cui et al. 1996), glycine-rich protein (Itabashi et al. 2011), and uncharacterized protein (Fujii and Toriyama 2009), have been reported as well.

Two types of CMS systems were previously reported in onions: CMS-S (Jones and Emsweller 1936) and CMS-T (Berninger 1965). These have been broadly used in F_1 hybrid seed production. Although the male-sterility phenotypes of these two CMS are indistinguishable upon visual observation, the inheritance mechanisms of fertility restoration were reported to differ between the two. One Rf gene, the *Ms* locus, was found to control fertility restoration in CMS-S (Jones and Clarke 1943), but more than three independent Rf loci were reported to be involved in the fertility restoration of CMS-T (Schweisguth 1973). However, Kim (2014) recently reported that the male fertility of both CMS-S and CMS-T might be commonly restored by the same *Ms* locus, or by tightly linked genes.

Since onion is a biennial crop and economic F_1 hybrid seed production is only possible using CMS, the rapid and reliable identification of cytoplasm types and Rf genotypes is fundamental for onion F_1 hybrid breeding. Accordingly, many molecular markers based on mitochondrial and chloroplast genome polymorphisms have been developed for identification of cytoplasm types (Havey 1995; Sato 1998; Engelke et al. 2003; Kim et al. 2009). In addition, several molecular markers linked to the *Ms* locus have also been developed for marker-assisted selection of the *Ms* genotypes (Gökçe et al. 2002; Bang et al. 2013; Park et al. 2013a; Yang et al. 2013; Kim 2014). Recently, Havey (2013) reported three SNP markers showing strong linkage disequilibrium (LD) with the *Ms* locus. In the present study, candidate genes responsible for the restoration of male fertility of CMS onions were identified through the combined approaches of bulked segregant analysis (BSA, Michelmore et al. 1991) and RNA-seq, and a reliable molecular marker for *Ms* genotyping was developed on the basis of polymorphism in the candidate gene.

Materials and methods

Plant materials

$F_{2:5}$ segregating populations originating from the cross between a male-sterile (MS) maternal line (506L) and a male-fertile (MF) paternal line (H6) were used for BSA and RNA-seq analysis. To screen tightly linked contigs, two recombinants (12-510 and 11-248) which contained crossover events between the *Ms* locus and the closely linked jnurf05 (Park et al. 2013a) and OPT (Bang et al. 2013) markers, respectively, were selected from among 4273 segregating plants produced in the previous study (Park et al. 2013a). The OPT marker was a simple PCR marker which had been converted from a RFLP marker, AOB272 (Gökçe et al. 2002). The contigs which were in LD with the *Ms* locus were identified using 251 breeding lines maintained by six different breeding institutes in Korea (Supplementary Table 1). The male fertility phenotypes and the cytoplasm types of these breeding lines were identified in the previous study (Kim 2014).

In addition, 124 exotic accessions introduced from 21 countries were also analyzed to estimate the LD levels of the contigs which showed perfect LD among the breeding

lines (Supplementary Table 2). Among them, 83 accessions were introduced from the National Plant Germplasm System, Agricultural Research Service, Baltimore, MD, USA. The other 41 were commercial cultivars bred in diverse countries. The origins of these cultivars are shown in Supplementary Table 2.

RNA-seq analysis and identification of homozygous SNPs between MF and MS bulks

Ten individuals from each group of MF and MS homozygous plants were selected by male fertility phenotypes, as determined by visual examination, and genotyping of the closely linked molecular markers (jnurf05 and OPT) flanking the *Ms* locus. These two markers were developed in previous studies (Bang et al. 2013; Park et al. 2013a). RNA was extracted from the unopened flowers of ten bulked samples of MF and MS plants using an RNA extraction kit following the manufacturer's instructions (RNeasy Plant Mini Kit, QIAGEN, Valencia, CA, USA). The ensuing RNA-seq and SNP discovery were carried out by a specialized company (Phyzen, Seoul, Republic of Korea). In brief, cDNA libraries were produced from the RNA samples using the Illumina TruSeq® RNA sample preparation v2 guide (Illumina, Hayward, CA, USA). The transcriptomes were sequenced by the HiSeq2000 (Illumina), and 101 bp paired-end reads were produced. The transcriptome sequences were then deposited into the SRA database under the accession numbers SRP056991. Trimmed raw reads from the MF and MS bulks were combined and de novo assembled into contigs using the Trinity software (Haas et al. 2013). Raw reads from the MF and MS bulks were separately mapped to the assembled contigs to find SNPs occurring between the bulks using SAMTools software (Li et al. 2009a). Quantification of each contig in the MF and MS bulks was carried out using RSEM software (Li and Dewey 2011).

To identify floral tissue-specific contigs showing significant increase of expression in the MF bulk, 30,004 contigs which were previously assembled from leaf tissues were employed (Kim et al. 2015). These contigs were produced using the same parental line (H6) as used in the present study. Differentially expressed contigs which were not detected among the 30,004 contigs were considered to be floral tissue specific.

DNA extraction, PCR amplification, and sequencing of PCR products

The total genomic DNA of all samples was extracted from the leaf or flower peduncle tissues using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1987).

PCR amplification of contigs was performed in a 25 μL reaction mixture containing a 0.1 μg template, 2.5 μL 10× PCR buffer, 0.5 μL forward primer (10 μM), 0.5 μL reverse primer (10 μM), 0.5 μL dNTPs (10 mM each), and 0.25 μL polymerase mix (Advantage 2 Polymerase Mix, Clontech, Palo Alto, CA, USA). PCR amplification consisted of an initial denaturation step at 95 °C for 4 min, 10 cycles at 95 °C for 30 s, 65 °C (0.8 °C decrements in each cycle) for 30 s, and 72 °C for 1 min, 35 cycles at 95 °C for 30 s, 57 °C for 30 s, and 72 °C for 1 min, and a final 10 min extension at 72 °C. The PCR products were visualized on 1.5 % agarose gels after ethidium bromide staining. If single clear PCR products were observed, the PCR products were purified for sequencing using a QIAquick PCR Purification kit (QIAGEN). Sequencing reactions were carried out using Big Dye (Applied Biosystems, Foster City, CA, USA), according to the manufacturer's protocol, and the sequences were obtained using an ABI PRISM 3730XL Analyzer (Applied Biosystems).

As for the PCR amplification of cleaved amplified polymorphic sequence (CAPS) and intron length polymorphism (ILP) markers, a 10 μL reaction mixture and 0.25 U Taq polymerase (Prime Tag DNA polymerase; GeNet Bio, Nonsan, Korea) were used for the analysis of a large number of samples. The same PCR amplification conditions were employed as described above. The PCR products of the CAPS markers were digested with their respective restriction enzymes for 3 h at the recommended temperatures. The PCR products or digested products were then visualized on 1.5 % agarose gels after ethidium bromide staining. The primer sequences and restriction enzymes are described in Supplementary Table 3.

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

To obtain MF and MS full-length cDNA sequences of the contigs which were in perfect LD with the *Ms* locus, RT-PCR was carried out if full-length cDNA sequences were available from the assembled contigs. If only partial sequences were assembled, RACE was performed to obtain the full-length cDNA sequences of both the MF and MS alleles. Regarding RT-PCR, the same MF and MS bulk RNAs used for the RNA-seq analysis were used to produce cDNA with a commercial cDNA synthesis kit (SuperScript™ III first-strand synthesis system for RT-PCR, Invitrogen, Carlsbad, CA, USA). RT-PCR amplification was performed with an initial denaturation step at 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 3 min, and a final 10-min extension at 72 °C. RACE was performed with a commercial RACE kit (SMART RACE cDNA Amplification Kit; Clontech), according to the

Table 1 Screening process for the selection of contigs in LD with the *Ms* locus

Step	Screening criterion	Number of contigs
1	De novo assembly of contigs from raw reads produced by RNA-seq	32,674
2	Contigs containing homozygous SNPs between MF and MS bulks	141
3	Contigs containing reliable SNPs verified by the IGV viewer	139
4	Contigs containing perfectly homozygous SNPs verified by PCR amplification and sequencing of bulked DNAs of MF and MS individuals	128
5	Contigs containing the SNPs showing perfect linkage with the *Ms* locus in the large-sized segregating population consisting of 4273 plants	30
6	Contigs containing the SNPs showing perfect LD with the *Ms* locus in diverse breeding lines	14

manufacturer's instructions. The PCR products were purified and sequenced as described above.

Identification of onion PPR genes and construction of a phylogenetic tree of PPR proteins isolated from onion and other plant species

Onion contigs containing the PPR motifs were identified through a local BLAST search using BioEdit software (Hall 1999). The rice Rf1a protein (GenBank accession number ABC42330) was used as the query. The deduced amino acid sequences of the onion PPR genes which were relatively close to the PPR-containing Rf genes were aligned using BioEdit software (Hall 1999), together with PPR-containing Rf genes isolated from maize, rice, radish, and petunia. Five *Arabidopsis Rf-PPR-like* (*RFL*) genes (Fujii et al. 2011) were also included in the alignment. The gaps in the alignment were removed using Gblocks software (Castresana 2000), with options for less stringent selection. A phylogenetic tree was constructed using MEGA version 4 (Tamura et al. 2007) using a neighbor-joining method. Node support of the phylogenetic tree was assessed by 1000 bootstrap replicates.

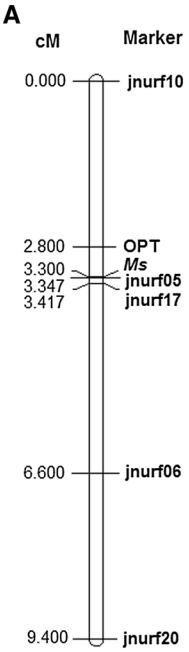
Results

Selection of contigs tightly linked to the *Ms* locus using BSA and RNA-seq analysis

To identify the candidate genes responsible for restoration of male fertility in onion, RNAs extracted from the floral tissues of MF and MS bulked samples were used for transcriptome sequencing. A total of 4.9 and 5.3 Gb of raw sequences were obtained from the MF and MS bulked samples, respectively. After trimming the low quality sequences and pooling the MF and MS sequences, 32,674 contigs were de novo assembled (Supplementary Table 4). The raw reads from the MF and MS bulks were separately mapped to the assembled contigs for identification of the SNPs between bulks. A total of 430 perfectly homozygous SNPs between bulks were identified, which were distributed in 141 contigs. The SNPs were further confirmed by visual investigation of the aligned reads using the IGV viewer (Robinson et al. 2011), and two contigs showing heterozygous SNPs were removed (Table 1).

For validation of the homozygous SNPs between bulks, primer pairs flanking the SNPs were designed for each of the 139 contigs (Supplementary Table 5). Putative exon–intron junctions were predicted using orthologous rice genes, if available, and the primers were designed based on the exon sequences flanking one or two introns. The rice orthologous sequences were retrieved from the Rice Genome Annotation Project (Ouyang et al. 2007). Initially, MF and MS bulk DNAs, each consisting of 10 homozygous individuals selected from the segregating population, were used as templates for the PCR amplification. If the PCR amplification failed, alternative pairs of primers were designed. In the case of successful PCR amplification, the PCR products were directly sequenced to verify the homozygous SNPs. Except for 11 contigs, all contigs were found to contain perfectly homozygous SNPs between bulks, as expected. In addition, the sequencing results showed that the intron positions of 81.3 % of the contigs (61 out of 75 contigs containing at least one intron) matched perfectly with those of rice orthologs (Supplementary Table 6). However, the chromosomal positions of the rice orthologous genes were not linked with each other in the rice genome, although the positions were not completely random (Supplementary Fig. 1).

As the next screening step, the recombinant (12-510) harboring a crossover event between the *Ms* locus and the jnurf05 marker was genotyped by sequencing for the SNPs on 128 selected contigs (Fig. 1). The jnurf05 marker was previously shown to be tightly linked to the *Ms* locus with a distance of 0.047 cM (Fig. 1a, Park et al. 2013a). The male fertility phenotype of the '12-510' plant was male-sterile; therefore, its *Ms* genotype should be



B

	Recombinants		Breeding lines			
Marker	12-510	11-248	90377	90143	90904	90027
jnurf10	B	H	A	B	B	H
OPT	B	H	A	A	A	H
Ms	B	B	H	B	B	B
jnurf05	H	B	A	A	H	H
jnurf17	H	B	A	H	H	B
jnurf06	H	B	B	B	B	H
jnurf20	D	B	B	B	B	B

Fig. 1 Linkage map of the *Ms* locus and marker genotypes of selected recombinants and breeding lines. **a** Linkage map flanking the *Ms* locus developed in the previous study (Park et al. 2013a). **b** Marker genotypes of two recombinants selected from a large-sized segregating population and four randomly selected breeding lines. Homozygous recessive genotypes are shown in *gray boxes*. *A* homozygous dominant, *H* heterozygous, *B* homozygous recessive, *D* homozygous dominant or heterozygous

homozygous recessive. However, the SNP genotypes of 85 contigs in this recombinant were heterozygous, indicating that these 85 genes were linked to the *Ms* locus at positions in the same direction as the jnurf05 marker (Fig. 1b). Next, another recombinant (11-248) containing a crossover between the *Ms* locus and the OPT marker (Fig. 1) was analyzed for the SNPs of the 43 contigs which had shown homozygous recessive SNP genotypes in the 12-510 recombinant. Among them, 30 contigs showed identical genotypes with that of the *Ms* locus, indicating perfect linkage with the *Ms* locus in this large-sized segregating population (Table 1).

Next, the SNPs of the 30 contigs showing perfect linkage with the *Ms* locus were genotyped in four randomly selected breeding lines (90377, 90143, 90904, and 90027) whose genotypes of the *Ms* and linked markers had been identified (Fig. 1b). Among the 30 contigs, 16 showed identical genotypes with those of the *Ms* locus. For analysis of a large number of samples, co-dominant CAPS or ILP markers were developed on the basis of SNPs and indel polymorphisms present in these 16 contigs (Supplementary Table 3). The developed markers were then used to analyze a total of 251 breeding lines bred by six different institutes, and the genotypes of all markers were found to be perfectly matched with the phenotypes of the *Ms* locus, except for the RF24000 and RF24437 markers (Supplementary Table 1), indicating that 14 of the contigs were in almost perfect LD with the *Ms* locus.

Additionally, 124 exotic accessions introduced from 21 countries were analyzed with the 14 markers (Supplementary Table 2) to estimate the LD level of these markers in diverse germplasm. One accession (PI233186) was found to have a crossover which separated the 14 markers into two groups (Fig. 2). The nine markers comprising group A in Fig. 2 showed perfect LD with each other, but one (RF27528) out of five markers of group B showed several recombinants. The jnurf13 marker developed in the previous study (Kim 2014) was also included in group B. However, we were unable to determine the position of the *Ms* locus since this accession (PI233186) contained normal cytoplasm (Supplementary Table 2).

Analysis of differentially expressed genes between MF and MS bulks and linkage relationship between the *Ms* locus and onion PPR gene family

If the recessive *Ms* allele were not transcribed in the MS plants, candidate genes could not be detected in the preceding SNP analysis. Therefore, contigs showing more than 10-fold increase of expression in the MF bulk compared with the MS bulks were screened. Using the stepwise screening process presented in Supplementary Table 7, a total of 145 contigs were finally selected (Supplementary Table 8). Transcription of 97 of the 145 contigs appeared to be floral tissue specific, since they were not detected in the transcriptome assembled from leaf tissues. However, none of the contigs showed perfect LD with the *Ms* locus after analysis of the PCR products amplified from the MF and MS bulks, recombinants, and four breeding lines, as shown in the SNP analysis, though five of the contigs (RF02974, RF06810, RF13742, RF15612, and RF27928) showed tight linkage with the *Ms* locus.

Since the majority of Rf genes cloned so far in other plant species were found to encode PPR proteins, the

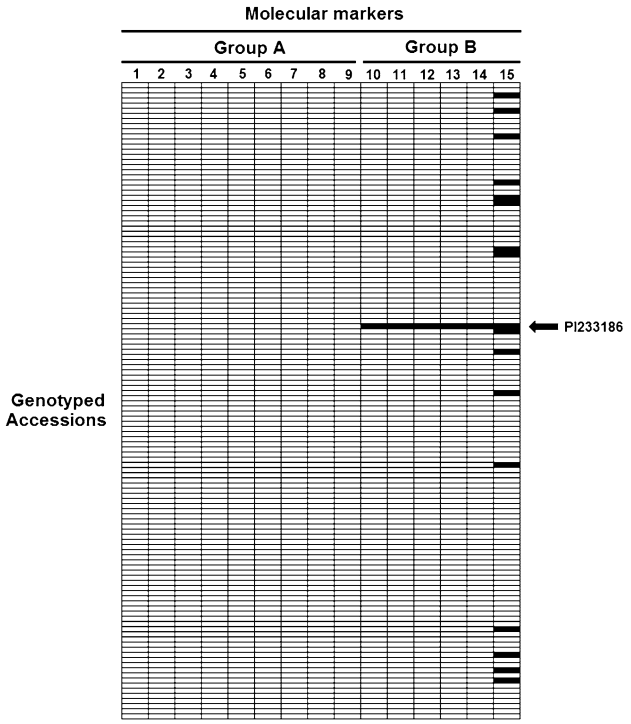


Fig. 2 Linkage relationship of 14 molecular markers tagging the candidate genes for the *Ms* locus in exotic accessions. The previously reported jnurf13 marker (Kim 2014) was also included. The marker genotypes which were inconsistent with those of the majority of markers are shown in *filled boxes*. The *horizontal arrow* indicates PI233186, which contains a crossover event separating the markers into two groups (Group A and B). The order of accessions is the same as that in Supplementary Table 2. The detailed marker genotypes of all tested markers and accession names are shown in Supplementary Table 2. *1* RF15334, *2* RF23881, *3* RF24123, *4* RF24998, *5* RF25191, *6* RF28184, *7* RF28314, *8* RF31446, *9* RF31869, *10* RF24501, *11* RF26780, *12* RF27463, *13* RF28839, *14* jnurf13, and *15* RF27528

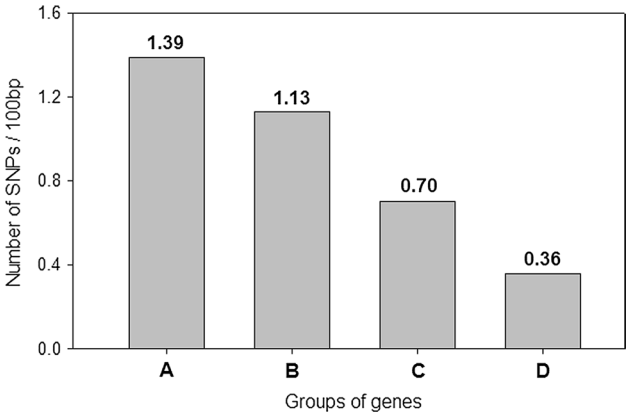


Fig. 3 Comparison of the SNP frequencies in four groups of genes which were linked or showed perfect LD with the *Ms* locus. *a* Nine genes showing perfect LD in the breeding lines and belonging to Group A in Fig. 2. *b* Five genes showing perfect LD in the breeding lines and belonging to Group B in Fig. 2. *c* 15 genes showing perfect linkage with the *Ms* locus in the large-sized segregating population, but not showing perfect LD in the breeding lines. *d* 98 Genes linked to the *Ms* locus, but not showing perfect linkage with the *Ms* locus in the segregating population

contigs containing PPR domains were selected from the onion transcriptome as candidate genes. A total of 483 contigs harboring at least one PPR motif were retrieved from a local BLAST search using the rice RF1a protein as the query (Supplementary Table 9). Two contigs (RF02436 and RF14756) were included from among the list of contigs containing homozygous SNPs between MF and MS bulks, but they did not show perfect linkage with the *Ms* locus (Supplementary Table 5). In addition, 14 PPR genes showing at least threefold increase of expression in the MF bulk were analyzed, but none of them showed any linkage with the *Ms* locus. Furthermore, 41 PPR genes which were closely related to the Rf and Rf-like PPR genes isolated from other species were also analyzed (Supplementary Fig. 2), but none contained homozygous SNPs or showed more than twofold increase of expression in the MF bulk (data not shown).

Annotation of full-length cDNA sequences of contigs showing LD with the *Ms* locus and identification of candidate genes for the *Ms* locus

The full-length cDNA sequences of 14 contigs which showed perfect LD with the *Ms* locus in the breeding lines were obtained by RT-PCR and RACE from both homozygous MF and MS individuals. The SNP frequencies in the coding regions of the 14 genes were significantly higher compared with the other linked genes (Fig. 3). In particular, the nine genes belonging to group A in Fig. 2 showed the highest SNP frequency. As the location of genes became closer to the *Ms* locus, the SNP frequency tended to become higher.

Next, the putative functions of the 14 genes were annotated by the BLAST-p search using their deduced amino acid sequences (Table 2). No previously known functions of Rf genes, such as PPR proteins, were identified. The deduced amino acid sequences of the 14 genes of the MF and MS alleles were compared to identify any putative critical mutations. One of the genes (RF15334) showed no polymorphic amino acid sequences between the MF and MS alleles, and was thus excluded from further analysis. As for the remaining 13, the amino acid sequences of the top 50 homologous proteins isolated from other plant species were aligned with the onion sequences to identify conserved regions. The amino acid changes which were unique to the MS allele and positioned in the known domains were searched. Four genes (RF31446, RF28839, RF26780, and RF27528) were found to contain

Table 2 Annotation of genes showing perfect LD with the *Ms* locus, and the number of amino acid changes between MF and MS alleles

Gene	Annotation	Length of coding sequences (bp)	Number of amino acid changes
RF15334	60S ribosomal protein L37a	276	0
RF23881	Peroxisome biogenesis protein 22	765	6
RF24123	Inactive leucine-rich repeat receptor-like serine/threonine protein kinase	2001	8
RF24501	66 kDa stress protein	1766	4
RF24998	Beta-galactosidase	2487	6
RF25191	Uncharacterized protein	744	11
RF26780	Protein AAR2 homolog	1224	10
RF27463	KH domain-containing protein	1920	6
RF27528	2-hydroxyacyl-CoA lyase	1701	5
RF28184	Uncharacterized protein	942	10
RF28314	Kinesin-13A	2163	7
RF28839	Mannose-6-phosphate isomerase	1275	13
RF31446	DNA mismatch repair protein PMS1	2640	26
RF31869	Serine/threonine protein kinase EDR1	2724	17

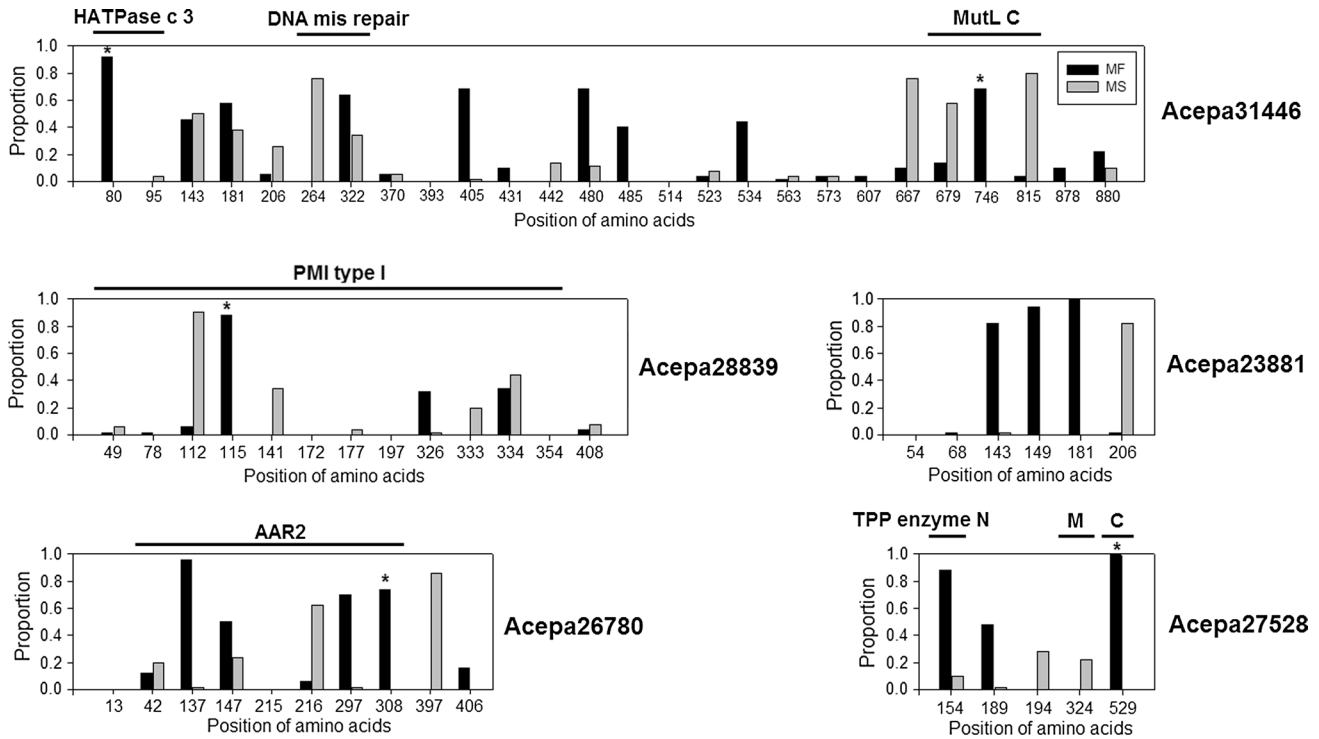


Fig. 4 Comparison of proportions of MF- and MS-specific amino acid changes in the top 50 homologous proteins isolated from other plant species. The *black* and *gray bars* indicate the proportion of the MF- and MS-specific amino acids, respectively. The known domains were identified using Pfam search (Finn et al. 2014), and the positions are shown with *horizontal lines* above the bar charts. The amino acid changes which were unique to the MS allele and positioned in the known domains are indicated with *asterisks* above the *bars*

putative critical amino acid changes in the known domains (Fig. 4). In the case of RF23881, two amino acid changes were observed in the conserved regions, though no known domain was found (Fig. 4). On the basis of annotated functions, LD analysis, and the presence of critical amino acid changes, the gene (RF31446) encoding DNA mismatch repair protein PMS1 might be the most plausible candidate for the restorer-of-fertility gene in onion. The *PMS1* gene identified in this study is hereafter designated as *AcPMS1*.

Discussion

Identification of genes in perfect LD with the *Ms* locus using a combination of BSA and RNA-seq approaches

Fourteen genes which showed perfect LD with the *Ms* locus were successfully identified herein through a combination of BSA and RNA-seq approaches. With the advent of next-generation sequencing, RNA-seq has become a powerful tool for massive SNP discovery in many species (Schneeberger and Weigel 2011; Mutz et al. 2013; Wolf 2013). RNA-seq analysis is especially useful for the examination of crop species containing complex or huge genomes, such as onion (16,400 Mb/1C), since no reliable reference genomic sequence is required for this analysis (Edwards et al. 2013). A few previous studies demonstrated the efficiency of combined BSA and RNA-seq analysis in cloning the causal genes for important traits in maize (Liu et al. 2012) and wheat (Trick et al. 2012).

To validate the SNPs identified in silico and to develop PCR-based markers, primers were designed based on the exon sequences flanking one or two introns. Since rice is the most closely related model plant with onion, the exon–intron boundaries of each contig were predicted using information from the rice orthologous genes. Approximately 81 % of the intron positions of onion were matched with rice orthologous genes. Similarly, Martin et al. (2005) showed that the positions of 83 % of onion introns were identical to those of rice introns, but there was no collinearity between the onion and rice genomes. We also confirmed the lack of microsynteny between two monocotyledonous crops. The positions of the rice genes which were orthologous to the onion genes that were tightly linked to the *Ms* locus were almost random (Supplementary Fig. 1). If the synteny between the rice and onion genomes were well conserved, the rice orthologous genes should form a cluster in the rice genome, since the onion genes analyzed herein were all linked to the *Ms* locus. Therefore, independent efforts are required in the future for establishment of onion genomic resources.

Identification of candidate genes responsible for restoration of male fertility in onion CMS

Two recombinants (12-510 and 11-248) selected from a large-sized population consisting of 4273 individuals were effectively used for the screening of candidate genes. Using these recombinants, 30 genes which were positioned within the approximately 0.55 cM interval between the OPT and jnurf05 markers could be successfully identified. Analysis of 251 breeding lines further eliminated 16 of the genes. Although these breeding lines were collected from six different institutes, their genetic diversity seemed to be narrow, since the six institutes were all located in Korea.

Extended analysis of 124 exotic accessions introduced from 21 counties enabled separation of 14 genes into two groups (group A and B in Fig. 2). In particular, one accession (PI233186) contained a crossover event between the two groups. The marker genotypes of group A were heterozygous, but genotypes of markers belonging to group B were homozygous recessive. However, we failed to observe male fertility phenotypes of this exotic accession due to disagreeable growth conditions. In addition, it will take at least 4 years to genotype the *Ms* locus of PI233186, because this accession contained normal cytoplasm. Despite the unidentified *Ms* genotypes, the *Ms* locus is likely located within group A, since one marker (RF27528) in group B contained several recombinants among the 124 accessions. In contrast, the markers belonging to group A showed perfect LD in all tested accessions. In addition, the SNP frequency in the coding sequences of the genes in group A was slightly higher than those of group B, and was 3.8-fold higher than the linked genes (group D in Fig. 3). The high SNP frequency in the LD block harboring the group A genes was considered to be a signature of selection during the long breeding history, as suggested by Gupta et al. (2005). The SNP frequencies gradually decreased as the distance between the genes and the *Ms* locus became farther due to decay of LD (Fig. 3).

Based on the annotated function, high SNP frequency (1.97/100 bp), and presence of putative critical amino acid changes, *AcPMS1* seems to be the most plausible candidate gene responsible for restoration of male fertility in onion CMS. This gene encodes the PMS1 protein, which is involved in DNA mismatch repair (MMR). The MMR system is one of the DNA repair pathways, and is highly conserved from bacteria to animals and plants. Three homodimeric proteins (MutS, MutL, and MutH) are members of the MMR system in *Escherichia coli*. MutS first recognizes mismatches and recruits MutL, which subsequently recruits MutH, possessing endonuclease activity, to the regions harboring mismatch (Kolodner and Marsischky 1999; Bray and West 2005; Kimura and Sakaguchi 2006). The PMS1 protein was known to form a heterodimer with the MLH1 protein, playing a similar role as MutL in yeast and plants (Bray and West 2005).

Two independent mutants of the *PMS1* gene in *Arabidopsis* showed more than 50 % collapsed pollen grains and increased homeologous recombination (Li et al. 2009b). A high proportion of collapsed pollen was also observed in the *Arabidopsis* with a mutant *MLH1* gene, a partner forming a heterodimer with *PMS1* (Dion et al. 2007). In addition, the *Msh1* gene product, which is homologous to MutS, is known to be involved in the MMR and control of plant mitochondrial recombination (Abdelnoor

et al. 2003). The *Msh1* mutant in *Arabidopsis* produced extensive mitochondrial genome rearrangement (Arrieta-Montiel et al. 2009). Furthermore, inactivation of the *Msh1* gene via RNAi created heritable cytoplasmic male-sterility in tobacco and tomato (Sandhu et al. 2007). Although the relationship between the MMR system and CMS induction is not yet clearly understood, the *AcPMS1* gene identified herein may be a good material for studying their relationship if this gene is truly responsible for the fertility restoration of onion CMS. It should be noted that although the *AcPMS1* gene was proposed as the best candidate for the *Ms* locus in this study, we could not exclude the possibility that the real causal gene might have escaped our screening process for identification of homozygous SNPs and differentially expressed genes.

Application of molecular markers in perfect LD with the *Ms* locus in onion breeding

Transition from open-pollinated cultivars into F_1 hybrid varieties has become a worldwide trend in the vegetable seed market, including onion. Therefore, CMS has become an essential tool in onion F_1 hybrid breeding, since CMS is the sole method by which F_1 hybrid seed production can be achieved. However, the development of diverse maintainer lines has been limited due to the requirement for time-consuming and labor-intensive progeny tests. This limitation can be overcome by using reliable molecular markers for genotyping of the *Ms* locus.

We previously reported a molecular marker (jnurf13) which was in LD with the *Ms* locus (Kim 2014). However, this marker was designed on the basis of indel polymorphisms in the intergenic region, and the size of the indel (12 bp) was so small that marker genotyping was only possible on acrylamide gels. In addition, jnurf13 was found to belong to the group B markers in Fig. 2. Therefore, if the *AcPMS1* gene proves to be responsible for fertility restoration in onion, the simple PCR marker (RF31446) based on polymorphism in the *AcPMS1* gene would be the most reliable marker for genotyping of the *Ms* locus. This marker belonged to group A, which showed perfect LD in all tested accessions collected from 21 different countries. In addition, this marker can easily be genotyped using agarose gels, since a relatively large 34-bp indel was utilized. Therefore, it is likely that the RF31446 marker developed in this study can be used universally, and will greatly enhance the efficiency of onion F_1 hybrid breeding. Besides practical application, *AcPMS1* and the other genes in perfect LD with the *Ms* locus will ultimately be invaluable materials for cloning of the *Ms* locus in the future.

Author contribution statement Sunggil Kim organized and coordinated this research project, and wrote the manuscript. Cheol-Woo Kim carried out analysis of the male fertility phenotypes of the breeding lines. Minkyu Park and Doil Choi produced and analyzed the transcriptome data. All authors read and approved the final manuscript.

Acknowledgments This research was supported by the Agriculture Research Center program, Golden Seed Project (Center for Horticultural Seed Development, No 213003-04-2-SB910), and a grant from the Next-Generation BioGreen 21 Program (Plant Molecular Breeding Center No. PJ011034). The authors thank Ji-wha Hur, Jeong-Ahn Yoo, and Su-jung Kim for their dedicated technical assistance.

Compliance with ethical standards

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

Ethical standards The authors declare that the experiments complied with the current laws of the Republic of Korea.

References

Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proc Natl Acad Sci USA 100:5968–5973

Abdelnoor RV, Christensen AC, Mohammed S, Munoz-Castillo B, Moriyama H, Mackenzie SA (2006) Mitochondrial genome dynamics in plants and animals: convergent gene fusions of a *MutS* homologue. J Mol Evol 63:165–173

Albert B, Godelle B, Gouyon PH (1998) Evolution of the plant mitochondrial genome: dynamics of duplication and deletion of sequences. J Mol Evol 46:155–158

Allen JO, Fauron CM, Mink P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177:1173–1192

Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie SA (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. Genetics 158:851–864

Arrieta-Montiel M, Shedge V, Davila J, Christensen AC, Mackenzie SA (2009) Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. Genetics 183:1261–1268

Backert S, Neilsen BL, Börner T (1997) The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trend Plant Sci 2:477–483

Bang H, Kim S, Park SO, Yoo K, Patil BS (2013) Development of a codominant CAPS marker linked to the *Ms* locus controlling fertility restoration in onion (*Allium cepa* L.). Sci Hortic 153:42–49

Bellaoui M, Martin-Canadell A, Pelletier G, Budar F (1998) Low-copy-number molecules are produced by recombination, actively maintained and can be amplified in the mitochondrial genome of Brassicaceae: relationship to reversion of the male sterile phenotype in some cybrids. Mol Gen Genet 257:177–185

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci USA 99:10887–10892

Berninger E (1965) Contribution à l'étude de la sterilité mâle de l'oignon (*Allium cepa* L.). Ann Amélior Plant 15:183–199

Bray CM, West CE (2005) DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity. New Phytol 168:511–528

Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS (2003) The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J 35:262–272

Budar F, Touzet P, De Paepe R (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revised. Genetica 117:3–16

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

Cui X, Wise RP, Schnable PS (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272:1334–1336

Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Rep 4:588–594

Dion E, Li L, Jean M, Belzile F (2007) An *Arabidopsis MLH1* mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. Plant J 51:431–440

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126:1–11

Engelke T, Terefe D, Tatlioglu T (2003) A PCR-based marker system monitoring CMS-(S), CMS-(T) and (N)-cytoplasm in the onion (*Allium cepa* L.). Theor Appl Genet 107:162–167

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–D230

Fujii S, Toriyama K (2009) Suppressed expression of retrograde-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. Proc Natl Acad Sci USA 106:9513–9518

Fujii S, Bond CS, Small ID (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci USA 108:1723–1728

Gökçe AF, Havey MJ (2002) Linkage equilibrium among tightly linked RFLPs and the *Ms* locus in open-pollinated onion populations. J Am Soc Hortic Sci 127:944–946

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Window 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:S154–S169

Havey MJ (1995) Identification of cytoplasms using the polymerase chain reaction to aid in the extraction of maintainer lines from open-pollinated populations of onion. Theor Appl Genet 90:263–268

Havey MJ (2013) Single nucleotide polymorphisms in linkage disequilibrium with the male-sterility restoration (*Ms*) locus in open-pollinated and inbred populations of onion. J Am Soc Hortic Sci 138:306–309

Hu J, Huang W, Huang Q, Qin X, Yu C, Wang L, Li S, Zhu R, Zhu Y (2014) Mitochondria and cytoplasmic male sterility in plants. Mitochondrion 19:282–288

Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K (2011) The fertility restorer gene, *Rf2*, for lead rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rice protein. Plant J 65:359–367

Janska H, Sarria R, Woloszynska M, Arrieta-Montiel M, Mackenzie SA (1998) Stoichiometric shifts in the common bean mitochondrial genome leading to male sterility and spontaneous reversion to fertility. Plant Cell 10:1163–1180

Jones HA, Clarke A (1943) Inheritance of male sterility in the onion and the production of hybrid seed. Proc Am Soc Hortic Sci 43:189–194

Jones HA, Emsweller SL (1936) A male-sterile onion. Proc Am Soc Hortic Sci 34:582–585

Kanazawa A, Tsutsumi N, Hirai A (1994) Reversible changes in the composition of the population of mtDNAs during dedifferentiation and regeneration in tobacco. Genetics 138:865–870

Kim S (2014) A codominant molecular marker in linkage disequilibrium with a restorer-of-fertility gene (*Ms*) and its application in reevaluation of inheritance of fertility restoration in onions. Mol Breed 34:769–778

Kim S, Lim H, Park S, Cho K, Sung S, Oh D, Kim K (2007) Identification of a novel mitochondrial genome type and development of molecular makers for cytoplasm classification in radish (*Raphanus sativus* L.). Theor Appl Genet 115:1137–1145

Kim S, Lee E, Cho DY, Han T, Bang H, Patil BS, Ahn YK, Yoon M (2009) Identification of a novel chimeric gene, *orf725*, and its use in development of a molecular marker for distinguishing three cytoplasm types in onion (*Allium cepa* L.). Theor Appl Genet 118:433–441

Kim S, Kim M, Kim Y, Yeom S, Cheong K, Kim K, Jeon J, Kim S, Kim D, Sohn S, Lee Y, Choi D (2015) Integrative structural annotation of *de novo* RNA-seq provides an accurate reference gene set of the enormous genome of the onion (*Allium cepa* L.). DNA Res 22:19–27

Kimura S, Sakaguchi K (2006) DNA repair in plants. Chem Rev 106:753–766

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the collinear region of rice chromosome 12. Theor Appl Genet 111:994–1012

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46:123–139

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J 34:407–415

Kolodner RD, Marsischky GT (1999) Eukayrotic DNA mismatch repair. Curr Opin Genet Dev 9:89–96

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). Plant J 37:315–325

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

Laser KD, Lersten NR (1972) Anatomy and cytology of microsporogenesis in cytoplasmic male sterile angiosperms. Bot Rev 38:425–454

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup (2009a) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079

Li L, Dion E, Richard G, Domingue O, Jean M, Belzile F (2009b) The Arabidopsis DNA mismatch repair gene *PMS1* restricts somatic recombination between homeologous sequences. Plant Mol Biol 69:675–684

Liu S, Yeh C, Tang HM, Nettleton D, Schnable PS (2012) Gene mapping via bulked segregant RNA-seq (BSR-Seq). PLoS One 7:e36406

Martin WJ, McCallum J, Shigyo M, Jakse J, Kuhl JC, Yamane N, Pither-Joyce M, Gokce AF, Sink KC, Town CD, Havey MJ (2005) Genetic mapping of expressed sequences in onion and in silico comparisons with rice show scant colinearity. Mol Gen Genomics 274:197–204

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Mutz K, Heilkenbrinker A, Lönne M, Walter J, Stahl F (2013) Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol 24:22–30

Oldenburg DJ, Bendich AJ (2001) Mitochondrial DNA from the Liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5′ protein. J Mol Biol 310:549–562

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR rice genome annotation resource: improvements and new features. Nucl Acids Res 35:D883–D887

Palmer JD (1988) Intraspecific variation and multicircularity in *Brassica* mitochondrial DNAs. Genetics 118:341–351

Palmer JD, Herbon LA (1987) Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr Genet 11:565–570

Park J, Bang H, Cho DY, Yoon M, Patil BS, Kim S (2013a) Construction of high-resolution linkage map of the *Ms* locus, a restorer-of-fertility gene in onion (*Allium cepa* L.). Euphytica 192:267–278

Park J, Lee Y, Lee J, Choi B, Kim S, Yang T (2013b) Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in radish (*Raphanus sativus* L.) containing DNCGMS cytoplasm. Theor Appl Genet 126:1763–1774

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26

Sakai T, Imamura J (1993) Evidence for a mitochondrial sub-genome containing radish *AtpA* in a *Brassica napus* cybrid. Plant Sci 90:95–103

Sandhu AP, Abdelnoor RV, Mackenzie SA (2007) Transgenic induction of mitochondrial rearrangements for cytoplasmic male sterility in crop plants. Proc Natl Acad Sci USA 104:1766–1770

Sato Y (1998) PCR amplification of CMS-specific mitochondrial nucleotide sequences to identify cytoplasmic genotypes of onion (*Allium cepa* L.). Theor Appl Genet 96:367–370

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3:175–180

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trend Plant Sci 16:282–288

Schweisguth B (1973) Étude d'un nouveau type de stérilité male chez l'oignon, *Allium cepa* L. Ann Amélior Plant 23:221–233

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA (2007) Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. Plant Cell 19:1251–1264

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10:e1001241

Small I, Suffolk R, Leaver CJ (1989) Evolution of plant mitochondrial genomes via substoichiometric intermediates. Cell 58:69–76

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Trick M, Adamski NM, Mugford SG, Jiang C, Febrer M, Uauy C (2012) Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploidy wheat. BMC Plant Biol 12:14

Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. Mol Ecol Resour 13:559–572

Woloszynska M, Trojanowski D (2009) Counting mtDNA molecules in *Phaseolus vulgaris*: sublimons are constantly produced by recombination via short repeats and undergo rigorous selection during substoichiometric shifting. Plant Mol Biol 70:511–521

Yang YY, Huo YM, Miao J, Liu BJ, Kong SP, Gao LM, Liu C, Wang ZB, Tahara Y, Kitano H, Wu X (2013) Identification of two SCAR markers co-segregated with the dominant *Ms* and recessive *ms* alleles in onion (*Allium cepa* L.). Euphytica 190:267–277

Zaegel V, Guermann B, Le Ret M, Andrés C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. Plant Cell 18:3548–3563