Theor Appl Genet (2014) 127:2243–2252
DOI 10.1007/s00122-014-2376-x

ORIGINAL PAPER

A high-resolution linkage map of the *Rfd1*, a restorer-of-fertility locus for cytoplasmic male sterility in radish (*Raphanus sativus* L.) produced by a combination of bulked segregant analysis and RNA-Seq

Young-Pyo Lee · Youngcho Cho · Sunggil Kim

Received: 5 May 2014 / Accepted: 2 August 2014 / Published online: 14 August 2014

Abstract
Key message **We utilized a combination of BSA and RNA-Seq to identify SNPs linked to the *Rfd1* locus, a restorer-of-fertility gene in radish. A high-density linkage map was constructed using this approach.**
Abstract Male fertility of cytoplasmic male sterility conditioned by the Dongbu cytoplasmic and genic male-sterility cytoplasm can be restored by a restorer-of-fertility locus, *Rfd1*, in radish. To construct a high-density linkage map and to identify a candidate gene for the *Rfd1* locus, bulked segregant analysis and RNA-seq approaches were combined. A total of 26 and 28 million reads produced from male-fertile and male-sterile bulked RNA were mapped to the radish reference unigenes. After stringent screening of SNPs, 327 reliable SNPs of 109 unigenes were selected. *Arabidopsis* homologs for 101 of the 109 genes were clustered around the 4,000 kb region of *Arabidopsis* chromosome 3, which was syntenic to the *Rfd1* flanking region. Since the reference unigene set was incomplete, the contigs were de novo assembled to identify 134 contigs harboring SNPs. Most of SNP-containing contigs were also clustered on the same syntenic region in *Arabidopsis* chromosome. A total of 21 molecular markers positioned within a 2.1 cM interval including the *Rfd1* locus were developed, based on the selected unigenes and contigs. A segregating population consisting of 10,459 individuals was analyzed to identify recombinants containing crossovers within this interval. A total of 284 identified recombinants were then used to construct a high-density map, which delimited the *Rfd1* locus into an 83-kb syntenic interval of *Arabidopsis* chromosome 3. Since no candidate gene, such as a pentatricopeptide repeat (PPR)-coding gene, was found in this interval, 231 unigenes and 491 contigs containing putative PPR motifs were analyzed further, but no PPR gene in linkage disequilibrium with the *Rfd1* locus could be found.

Introduction

Male sterility, inability of producing viable pollen grains in plants, is naturally observed in many plant species (Laser and Lersten 1972), and can arise due to both environmental and genetic effects. In reference to genetic factors, causal genes for male sterility are positioned in either nuclear or mitochondrial genomes (Hanson 1991). Cytoplasmic male sterility (CMS), caused by aberrant mitochondrial genes, is a maternally inherited trait that has been widely used in F_1 hybrid seed production in many economically important crops (Budar et al. 2003; Hanson and Bentolila 2004; Knoop 2004; Kubo and Newton 2008). All CMS-inducing genes reported so far reside in mitochondrial genomes with no exception (Hu et al. 2014).

Mitochondrial genes responsible for CMS are mostly chimeric genes consisting of partial sequences of known mitochondrial genes, along with unidentified sequences. Partial sequences or promoters of genes coding for subunits of ATP synthase or cytochrome c oxidase have often been involved in the creation of chimeric CMS-inducing genes

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2376-x) contains supplementary material, which is available to authorized users.

Y.-P. Lee · Y. Cho
Biotech Research Center, Dongbu Advanced Research Institute, Dongbu Hannong Co., Ltd., Daejeon 305-708, Korea

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

(Hanson and Bentolila 2004). Dynamic mitochondrial genome rearrangements in plants mediated by frequent recombination are probably responsible for the production of such chimeric genes (Small et al. 1989; Kmiec et al. 2006). Frequent rearrangement of mitochondrial genomes is one the unique features of plant species.

Unlike stable, small-sized animal mitochondrial genomes, plant mitochondrial genomes are unstable due to frequent recombination (Backert et al. 1997; Oldenburg and Bendich 2001; Allen et al. 2007), and their genome sizes are highly variable among species, ranging from 208 kb in *Brassica hirta* (Palmer and Herbon 1987) to 11,319 kb in *Silene conica* (Sloan et al. 2012). Through recombination mediated by repeat sequences, plant mitochondrial genomes often form multipartite structures and multiple subgenomes (Palmer 1988; Small et al. 1989; Albert et al. 1998; Woloszynska and Trojanowski 2009). In addition, the stoichiometry of subgenomes is sometimes changed by a mechanism called substoichiometric shifting, in response to mutations in nuclear genes (Abdelnoor et al. 2006; Zaegel et al. 2006; Shedge et al. 2007), although specific stoichiometry is usually maintained throughout generations (Sakai and Imamura 1993; Bellaoui et al. 1998; Janska et al. 1998; Kim et al. 2007).

A CMS radish was first discovered by Ogura (1968), and has been widely used in commercial F_1 hybrid cultivar breeding. This CMS, normally called Ogura CMS, has been transferred to *Brassica* species, such as rapeseed and Chinese cabbage, through conventional backcross breeding (Bannerot et al. 1974; Dickson 1985) and protoplast fusion (Pelletier et al. 1983; Menczel et al. 1987; Jarl et al. 1989) for the establishment of stable CMS systems. Using segregating cybrids produced from protoplast fusion, the causal gene for Ogura CMS, *orf138*, was identified (Bonhomme et al. 1991; Grelon et al. 1994). Male fertility of Ogura CMS can be restored by a single restorer-of-fertility (Rf) locus, *Rfo*, which was isolated by map-based cloning approaches and was revealed to encode a pentatricopeptide repeat (PPR) protein (Brown et al. 2003; Desloire et al. 2003; Koizuka et al. 2003). Likewise, the Rf genes isolated from petunia (Bentolila et al. 2002), rice (Komori et al. 2004), and sorghum (Klein et al. 2005) also code for PPR proteins. However, *Rf2*, the first cloned Rf gene in maize, was shown to encode alcohol dehydrogenase (Cui et al. 1996), while two different rice Rf loci, *Rf2* and *Rf17*, encoded glycine-rich protein (Itabashi et al. 2011) and a protein with unknown function (Fujii and Toriyama 2009), respectively.

A different type of CMS called Dongbu cytoplasmic and genic male sterility (DCGMS) was also reported in radish (Lee et al. 2008), presenting a different phenotype from that of Ogura CMS. In Ogura CMS, the tapetal cell layer was degenerated at the tetrad stage of pollen development, thus no pollen grain was detected in mature flowers. On the other hand, maturation of the microspores released from the tetrad was disabled in DCGMS, and clumped non-viable pollen grain was detected in mature anthers (Lee et al. 2008). Molecular markers designed on the basis of variable mitochondrial (Lee et al. 2009) and chloroplast genomes (Kim et al. 2009) showed different patterns between the DCGMS and Ogura cytoplasms. Indeed, the organization of the mitochondrial genomes of DCGMS and Ogura CMS was proven to be highly variable, after the complete mitochondrial genomes of the two were released (Park et al. 2013). A novel chimeric gene designated as *orf463* was identified in the DCGMS-unique region, which may be responsible for CMS induction in DCGMS cytoplasm (Park et al. 2013).

In addition, the nuclear gene found to be responsible for fertility restoration of male sterility conferred by the DCMGS cytoplasm was different from that of Ogura CMS (Lee et al. 2008). A single Rf gene, designated as *Rfd1*, was identified in the previous study (Kim et al. 2010), and a high-density linkage map was constructed using synteny between radish and *Arabidopsis* genomes (Cho et al. 2012). To increase the resolution of the linkage map and to identify candidate genes for the *Rfd1* locus, a combination of bulked segregant analysis (BSA, Michelmore et al. 1991) and next-generation sequencing of transcript (RNA-Seq) was deployed in this study.

Materials and methods

Plant materials

The F_1 hybrid and F_2 populations originating from the cross between a male sterile line (MS19) containing the DCGMS cytoplasm and a male-fertile line (R219) were produced. The genotypes of the *Rfd1* locus of 224 F_2 plants were deduced from marker genotypes of five *Rfd1*-flanking markers (CAPS3700, ILP3585, ILP3510, CAPS3100, and RSS0818), which had been developed in the previous study (Cho et al. 2012). The *Rfd1* genotypes of the 224 F_2 plants showed a single-gene inheritance ($\chi^2 = 4.74$ and $P = 0.09$). Among the F_2 plants, 86 heterozygous plants were selected for random pollination in an isolated greenhouse to produce a large-sized segregating $F_{2:3}$ population. Pollination was carried out using honeybees as pollinators to produce a large amount of seeds. Seeds of the $F_{2:3}$ population were germinated in 128-cell plug trays measuring 54 × 28 × 8 cm, and total genomic DNA was extracted from the cotyledon of the seedlings using a CTAB (Cetyl trimethylammonium bromide) method (Doyle and Doyle

1987). Recombinants between two flanking markers (CAPS3100 and ILP3510) had been selected from 5,424 plants in 2012, and recombinants between CAPS3100 and CAPS3700 markers had been selected from 5,035 plants in 2013. Selected recombinants were transplanted in the greenhouse and vernalized through a winter season before male-fertility phenotypes were evaluated using multiple visual examinations.

RNA-Seq analysis and SNP identification between male-fertile and male-sterile bulked samples

To construct a high-density linkage map of the *Rfd1* locus, the BSA approach (Michelmore et al. 1991) and next-generation sequencing of the transcriptome were combined. Ten individuals of each group, male-fertile and male-sterile homozygous F_2 plants, were selected based on male-fertility phenotypes and the five previously mentioned flanking markers. Total RNA was extracted from the bulked samples of unopened flowers detached from the ten selected F_2 plants using an RNA extraction kit, following the manufacturer's instructions (RNeasy Plant Mini Kit, QIAGEN, Valencia, CA, USA).

The process of RNA-Seq, raw read mapping, and SNP calling was performed by a specialized company (Macrogen, Seoul, Republic of Korea). In brief, the cDNA sequencing library was produced using the Illumina TruSeq® RNA sample preparation v2 guide (Illumina, Hayward, CA, USA). The cDNA library was sequenced by the HiSeq 2000 (Illumina), and 101-bp paired-end reads were produced. Raw reads were mapped to the radish reference unigenes (Shen et al. 2013) using Bowtie software (Langmead et al. 2009), and quantification of transcripts was carried out using RSEM software (Li and Dewey 2011). The SNPs between reference unigenes and mapped reads were identified using SAMTools software (Li et al. 2009). Alternatively, raw reads produced from the male-fertile bulk were de novo assembled into contigs and isoforms using the Trinity software (Haas et al. 2013). Both raw reads from male-fertile and male-sterile bulks were mapped to the de novo assembled contigs. The SNPs between bulks were identified using SAMTools software (Li et al. 2009).

Development of molecular markers based on the SNPs discovered from transcriptome sequences

After the unigenes or contigs containing reliable SNPs were selected, putative intron positions were deduced by alignment of unigene/contig sequences and genomic sequences of corresponding *Arabidopsis* homologs. Primer pairs were designed based on the exon sequences separated by up to three introns to detect more polymorphic sequences. PCR amplifications were carried out using parental DNA as templates. PCR was performed in 25-μL reaction mixtures containing 0.05 μg template, 2.5 μL 10× PCR buffer, 0.2 μL forward primer (10 μM), 0.2 μL reverse primer (10 μM), 0.2 μL dNTPs (10 mM each), and 0.25 μL polymerase mix (Advantage 2 Polymerase Mix, Clontech, Palo Alto, CA, USA). PCR amplification was performed with an initial denaturation at 95 °C for 4 min, followed by 10 cycles of 95 °C for 30 s, 67 °C (with 0.8 °C decrements after each cycle) for 30 s, and 72 °C for 2 min, then 30 cycles of 95 °C for 30 s, 59 °C for 30 s, and 72 °C for 2 min, and a final 7-min extension at 72 °C. PCR products were visualized on 1.5 % agarose gels after ethidium bromide staining. If single clear PCR products were produced, the PCR products of both parental lines were directly sequenced. PCR products were purified 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).

If the PCR products showed length polymorphism in introns between the parental lines, the primer combination was used as an intron length polymorphism (ILP) marker. Otherwise, cleaved amplified polymorphic sequence (CAPS) markers were designed based on the SNPs between parental lines. For ILP and CAPS marker detection, PCR was performed in 10-μL reaction mixtures with the same conditions mentioned above, with substitution for an economical Taq polymerase (Prime Tag DNA polymerase; GeNet Bio, Nonsan, Korea). The PCR products of CAPS markers were digested with their respective restriction enzymes for 3 h at the recommended temperatures. The primer sequences for molecular markers developed in this study are presented in Supplementary Table 1.

Construction of a phylogenetic tree of PPR proteins isolated from radish and other species

The deduced amino acid sequences of PPR genes isolated from radish and other species were aligned using BioEdit software (Hall 1999). Five *Arabidopsis Rf-PPR-like* (*RFL*) genes which were evenly distributed in the phylogenetic tree of 26 *RFL* genes (Fujii et al. 2011) were included in the alignment. Five PPR-encoding Rf genes isolated from radish, maize, rice, and petunia were also included. The gaps in alignment were removed using Gblocks software (Castresana 2000), with options for less stringent selections. 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 1,000 bootstrap replicates.

Table 1 Filtering criteria for identification of reliable SNPs and differentially expressed genes between male-fertile and male-sterile bulks

Category	Filtering criteria	Number of SNPs or unigenes
SNPs in the unigenes	SNPs between reference unigenes and samples	437,474
	SNPs between bulks	135,068
	SNPs with quality scores higher than 200	1,679
	Homozygous SNPs[a]	1,288
	SNPs visually confirmed by the IGV viewer	327
SNPs in de novo assembled contigs	SNPs between bulks	28,320
	SNPs with quality scores higher than 200	974
	SNPs visually confirmed by the IGV viewer	752
Differentially expressed unigenes	Expression in male-fertile bulk (RPKM > 10)	16,664
	10-fold higher expression in male-fertile bulk	100
	Expression in male-sterile bulk (RPKM > 1)	63

[a] Homozygous SNPs were defined as the SNPs containing less than one read showing different SNP alleles

Table 2 Distribution of *Arabidopsis* homologs of radish unigenes or contigs containing SNPs or differential expression between male-fertile and male-sterile bulks in *Arabidopsis* chromosomes

Selected radish genes	*Arabidopsis* chromosomes					No hit[a]	Total
	1	2	3	4	5		
Unigenes containing SNPs	0	2	101	2	3	1	109
Differentially expressed unigenes	2	5	46	1	3	6	63
De novo assembled contigs containing SNPs	0	1	124	4	2	3	134

[a] No significantly homologous *Arabidopsis* genes were found using BLAST search

Results

Identification of SNPs from transcriptome constructed from male-fertile and male-sterile bulks

Bulked segregant analysis and next-generation sequencing were combined to develop molecular markers closely linked to the *Rfd1* locus, a restorer-of-fertility gene in radish. Transcriptome sequences were constructed from male-fertile and male-sterile bulked RNA. A total of 26 and 28 million reads were produced from the male-fertile and male-sterile bulks, respectively (Supplementary Table 2), and these raw reads were mapped to the radish reference unigene set consisting of 85,083 entries (Shen et al. 2013). Approximately, 30 % of the raw reads were mapped to the unigenes, and a total of 437,474 SNPs were identified between the unigene reference and the samples. Among them, reliable SNPs were screened using the criteria listed in Table 1. A total of 135,068 SNPs showed polymorphism between male-fertile and male-sterile bulks. After filtering low quality and heterozygous SNPs, 1,288 SNPs were selected for further analysis. Finally, the mapped reads were visually investigated using Integrative Genomics Viewer (IGV: Robinson et al. 2011) to confirm homozygous SNPs, and a final set of 327 reliable SNPs was screened (Supplementary Table 3). These SNPs were distributed over 109 unigenes.

In the previous study (Cho et al. 2012), we showed that the *Rfd1*-flanking regions were syntenic to the *Arabidopsis* chromosome 3. Therefore, if the selected SNPs were truly authentic, most of the *Arabidopsis* homologs to the 109 unigenes containing SNPs would be positioned in the *Arabidopsis* chromosome 3. To verify this hypothesis, *Arabidopsis* homologous genes showing the highest homology with each of the 109 unigenes were identified using a BLAST search. Most unigene homologs, excluding only eight of the unigenes, were positioned in the *Arabidopsis* chromosome 3 (Table 2). Furthermore, the positions of the *Arabidopsis* homologs were clustered around the 4,000 kb region, where the *Rfd1*-flanking regions showed synteny between radish and *Arabidopsis* genomes (Fig. 1a). This indicated reliable screening of SNPs between the male-fertile and male-sterile bulks.

Although multiple unigenes showing reliable SNPs were identified, the gene corresponding to the *Rfd1* locus could not be detected if the recessive mutant allele of the

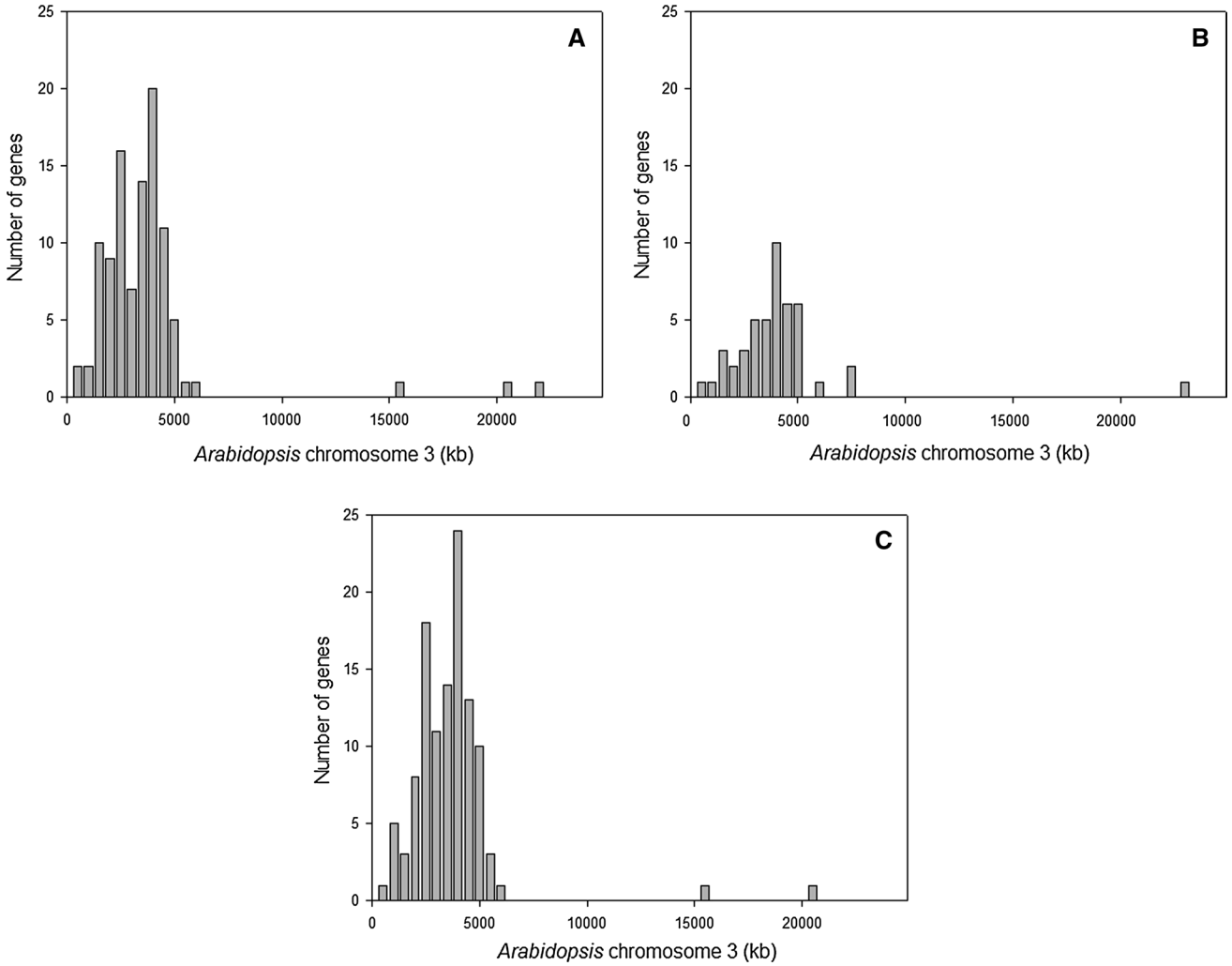

Fig. 1 Distribution of *Arabidopsis* genes homologous to the radish unigenes or contigs containing the SNPs between male-fertile and male-sterile bulks. **a** Unigenes containing SNPs. **b** Differentially expressed unigenes. **c** De novo assembled contigs containing SNPs

causal gene was not sufficiently expressed for detection in SNP calling. Thus, unigenes showing significant differential expression were screened using the criteria listed in Table 1. Overall, the expression levels of most unigenes were similar between the two bulks (Supplementary Fig. 1), and showed a high correlation coefficient of 0.88. First, the unigenes showing low expression in male-fertile bulk were first filtered to remove technical noise. A total of 94 unigenes showing 10-fold higher expression in the male-fertile bulk compared to the male-sterile bulk were identified, and a final set of 63 unigenes was screened after removal of the unigenes whose expression level in the male-sterile bulk was higher than one RPKM value, since these could have been detected in the preceding SNP screening. Similar to results of the previous SNP screening, *Arabidopsis* homologs of 46 of the 63 differentially expressed unigenes were positioned around the 4,000 kb region of the *Arabidopsis* chromosome 3 (Fig. 1b).

We mapped the raw reads to the radish unigene reference to identify SNPs, but it is unlikely that the unigene set included all radish genes, as it is possible that causal genes may not be included in the reference unigene list. Therefore, raw reads of the male-fertile bulk were de novo assembled into 96,748 contigs and isoforms to identify SNP-containing contigs which were not included in the reference unigenes. A total of 752 SNPs were selected after removal of low-quality SNPs and IGV investigation (Table 1). The selected SNPs were distributed in 134 contigs and isoforms. Like the SNP-containing unigenes, the majority of *Arabidopsis* homologs to the contigs (124 out of 134) were positioned around the 4,000 kb region in chromosome 3 (Table 2; Fig. 1c).

The selected SNP-containing and differentially expressed unigenes and contigs were annotated with BLAST search using the *Arabidopsis* database (Supplementary Tables 4, 5, and 6). Almost half of the

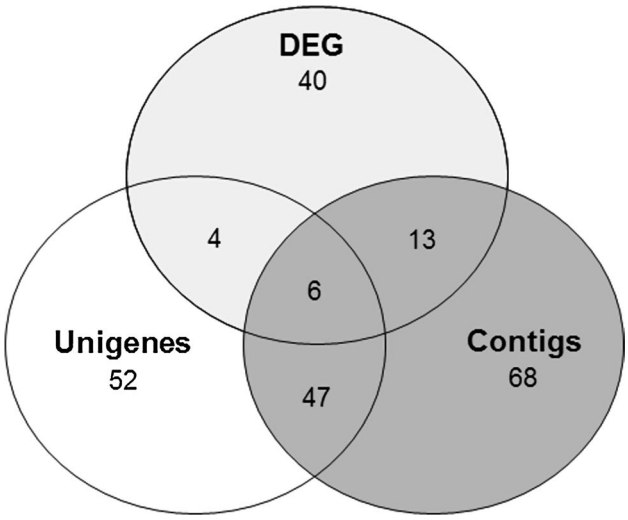

Fig. 2 Diagram showing the proportion of common unigenes or contigs screened by raw read mapping to the reference unigenes or de novo assembled contigs, and comparison of expression levels of unigenes. *DEG* Differentially expressed gene

SNP-containing unigenes and de novo assembled contigs were commonly selected by two different mapping methods. Among 53 common genes, six genes also showed differential expression between the two bulks (Fig. 2). However, in these six genes, no unigenes or contigs were found with similar functions encoded by previously isolated Rf genes, such as PPR proteins (Supplementary Table 7).

Construction of a high-density linkage map of the *Rfd1* locus

To construct a high-density linkage map of the *Rfd1* locus, a large-sized $F_{2:3}$ population was analyzed to identify recombinants between two previously developed molecular markers, CAPS3100 and CAPS3700 (Cho et al. 2012), which flanked the *Rfd1* locus within a 3.2 cM interval. A total of 284 recombinants were identified out of 10,459 individuals. The selected recombinants were further analyzed using molecular markers developed in this study, as follows, for fine mapping of the *Rfd1* locus.

For development of closely linked markers, SNP-containing unigenes or contigs for which the *Arabidopsis* homologs were positioned at the 3,098–3,702 kb interval in *Arabidopsis* chromosome 3 were selected. Two *Arabidopsis* genes, which were orthologous to the radish genes containing the CAPS3100 and CAPS3700 markers, were positioned at the boundary of this interval. In addition, SNP-containing unigenes or contigs whose *Arabidopsis* homologs were positioned in areas other than chromosome 3, or for which there were no *Arabidopsis* homologs,

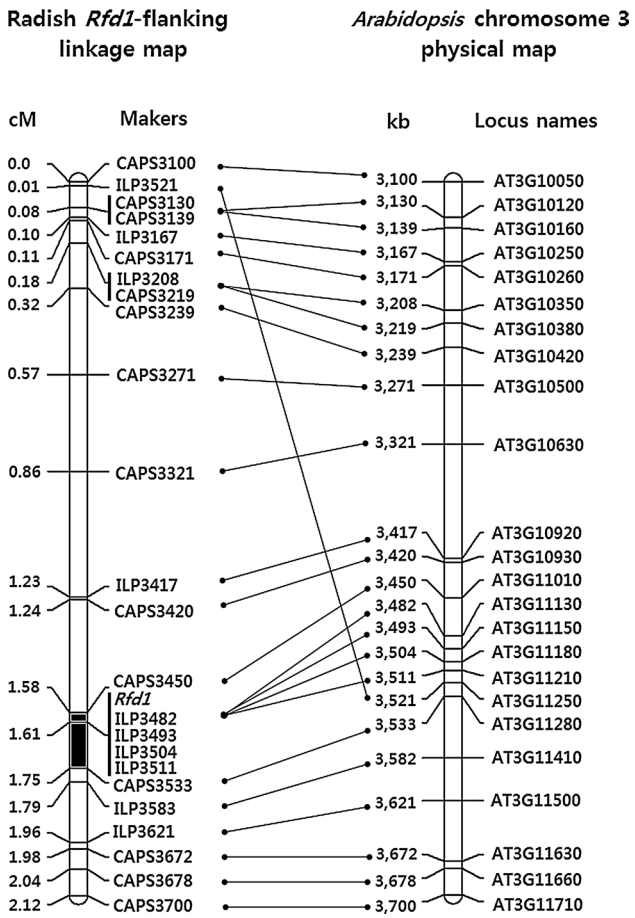

Fig. 3 A high-resolution linkage map of the *Rfd1* locus. The physical map of *Arabidopsis* chromosome 3 is shown on the *right*, and syntenic regions are connected with *horizontal lines*. The ILP3521 and ILP3585 markers were developed in the previous study (Cho et al. 2012). The interval delimited by CAPS3450 and CAPS3533 makers is indicated by a *black filled box*

were analyzed to confirm their linkage relationship with the *Rfd1* locus. ILP or CAPS markers were developed on the basis of polymorphic sequences between the male-fertile and male-sterile bulked DNAs. After confirming co-segregation of the markers and the *Rfd1* locus, 21 reliable and robust molecular markers were selected for analysis of the 284 identified recombinants. Among the 21 selected markers, one of the markers was designed based on SNPs in contigs which showed no *Arabidopsis* homologs, but this marker was positioned outside the interval delimited by the CAPS3100 and CAPS3700 markers (data not shown).

The 284 recombinants were analyzed using the newly developed markers. Radish genes containing marker sequences were perfectly collinear with the *Arabidopsis* orthologs except for one translocation of the radish gene harboring ILP3521 marker (Fig. 3). The CAPS3450 and

CAPS3533 markers further delimited the *Rfd1* locus to the interval of an 83 kb syntenic region on the *Arabidopsis* chromosome 3, and four markers showed no recombinants between the *Rfd1* locus and the markers (Fig. 3). A total of 25 genes were positioned in the delimited 83 kb region on the *Arabidopsis* chromosome 3. However, none of these 25 genes were PPR genes or Rf-like genes (Supplementary Table 8).

Isolation of radish PPR genes and their linkage relationship with the *Rfd1* locus

Since most restorer-of-fertility genes cloned so far encoded PPR genes, PPR genes may be the best candidate for the *Rfd1* locus. First, *Arabidopsis* PPR genes in the syntenic region to the *Rfd1* flanking region were searched for using the Gramene database (Monaco et al. 2013). A total of 492 *Arabidopsis* genes containing PPR domains were retrieved (data not shown). Among them, only one PPR gene (AT3G11380) was positioned in the 3,098–3,702 kb interval. However, no highly homologous radish unigene or contigs corresponding to AT3G11380 were identified, and radish genes showing some degree of homology did not show any SNPs between the male-fertile and male-sterile bulks (data not shown). Furthermore, the PPR protein encoded by AT3G11380 was distantly related with other *RFL* genes in *Arabidopsis* (Fujii et al. 2011) and other PPR-coding restorer-of-fertility genes isolated in petunia, rice, and maize (Fig. 4), suggesting involvement of AT3G11380 in biological processes other than fertility restoration.

To identify a candidate gene responsible for the *Rfd1* locus, radish genes showing homology with the radish *Rfo* gene, the previously isolated Rf gene encoding a PPR protein, were isolated from reference unigenes and de novo assembled contigs using a BLAST search. A total of 231 unigenes (Supplementary Table 9) and 491 contigs (Supplementary Table 10) were identified as PPR motif-containing genes (cut-off *E* value >E-03). Among them, seven SNP-containing PPR genes, designated as *RsPPR1-RsPPR7*, were selected by screening with SNP-calling quality and IGV investigation. (Supplementary Table 11). Primer pairs were designed for these PPR genes, and PCR products were sequenced. *RsPPR1* and *RsPPR4* showed close linkage with the *Rfd1* locus (Supplementary Table 11), but both were positioned outside the interval delimited by the CAPS3100 and CAPS3700 markers, suggesting that these genes were not responsible for fertility restoration. Alternatively, 12 radish PPR genes which were closely related to the RFL genes isolated from other species were selected for detection of SNPs (Fig. 4). These were designated as *RsRFL1-RsRFL12* (Supplementary Table 11). However, no SNPs between the male-fertile and male-sterile bulks were detected in these RFL genes.

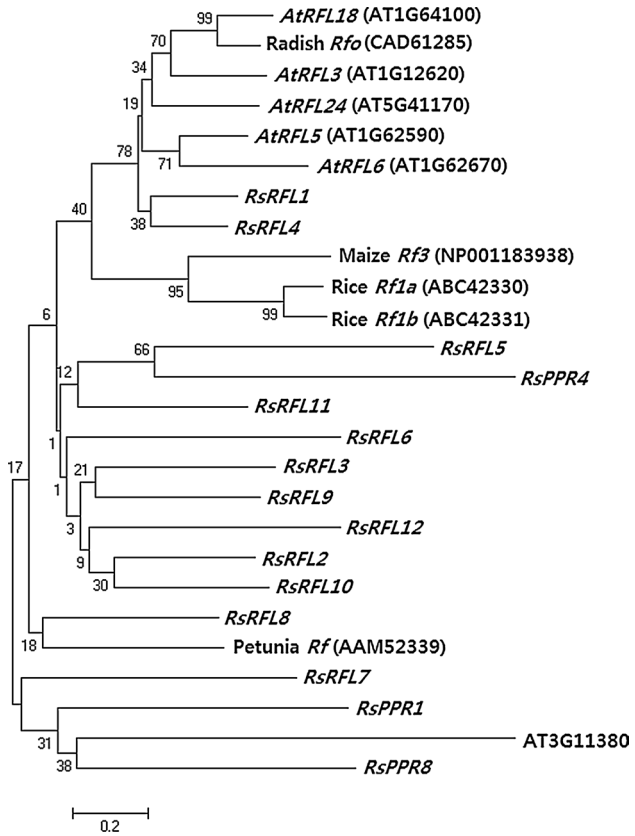

Fig. 4 Phylogenetic relationship of radish PPR genes with *Arabidopsis RFL* genes and PPR-coding Rf genes isolated from maize, petunia, rice, and radish. An *Arabidopsis* PPR gene (AT3G11380) positioned in the syntenic region containing the *Rfd1* locus was included to assess its relationship to other restorer-like PPR genes. The names and nucleotide sequences of radish PPR genes are shown in Supplementary Tables 11 and 10, respectively. The tree was constructed using the deduced amino acid sequences by a neighbor-joining method. The numbers at the nodes are the bootstrap probability (%) with 1,000 replicates

Discussion

Construction of a high-resolution linkage map using BSA and RNA-Seq

For construction of a high-resolution linkage map and identification of candidate genes for the *Rfd1* locus in radish, a combination of BSA and RNA-Seq technology was utilized. Since the BSA approach was first devised by Michelmore et al. (1991), numerous studies have been carried out to construct linkage maps encompassing genes of interest, and BSA has become the routine analysis used for mapping specific genomic regions. For map construction, random markers such as RAPD and AFLP markers have been used most commonly (Kim et al. 2011; Gonthier et al. 2013), but random markers have some limitations to practical application. First, RAPD markers need to be converted into SCAR

markers for reliable PCR amplification due to their inherent low reproducibility, and AFLP markers are not appropriate for large-scale analysis of breeding materials due to the complicated and time-consuming detection protocol required for analysis. Second, random markers usually tag non-coding regions, and tagging of regions tightly linked to the gene of interest solely depends on luck. The closer the molecular markers are positioned to the target genes, the more useful they are in marker-assisted selection in breeding programs.

In contrast, development of molecular markers based on coding regions used to be difficult because of inadequate availability of genomic sequence information for most crop species, except for a few model organisms. However, with the advent of next-generation sequencing technology, throughput of sequencing has sharply increased while the cost has steadily decreased. These days, a large amount of transcriptome sequence can be produced by RNA-Seq in any crop species regardless of the size and complexity of the genomes. Although RNA-Seq has been mainly used to estimate abundance of transcripts to identify differentially expressed genes in response to internal or external stimuli, this technique has also been used to detect a large number of SNPs for the development of molecular markers.

In most cases of SNP discovery, inbred parental lines have been used as materials for RNA-Seq to construct linkage maps. However, in this study, we used bulked samples of segregating populations for RNA-Seq to identify SNPs in the genes linked to the target locus. A few previous reports used similar approaches. Two near-isogenic lines showing polymorphisms in a grain protein content locus (*GPC-B1*) were used to identify SNPs around this locus in wheat (Trick et al. 2012). Liu et al. (2012) also used a similar approach to clone the *glossy3* gene in maize, and they coined the new term 'bulked segregant RNA-Seq (BSR-Seq)' to describe this kind of approach.

SNPs tightly linked to the radish *Rfd1* locus were successfully identified in this study. Unlike SNP discovery from two inbred parental lines, efficient removal of heterozygous SNPs in the bulked samples might be the key factor for successful selection of reliable SNPs tightly linked to the target locus. Most heterozygous SNPs were derived from unlinked genes. Initially, hundreds of thousands of SNPs were detected by the SAMtools software (Li et al. 2009). Reliable SNPs were selected using relatively stringent filtering criteria (Table 1). Particularly, visual investigation of read alignment using the IGV greatly improved the reliability of SNPs.

Since complete genomic sequences of radish are not yet available, 85,083 unigenes were used as the reference for raw read mapping and SNP discovery. Although the recently released radish unigene set seems to contain most radish genes, not all of the genes may be included, due to insufficient genomic sequence information of radish. Indeed, only 30 % of the raw reads were mapped to the reference unigenes. For this reason, we de novo assembled the contigs and isoforms using Trinity software (Haas et al. 2013), and identified ten novel genes which were tightly linked to the *Rfd1* locus, and not included in the unigene set. Therefore, SNP discovery using de novo assembled contigs can complement mapping with a unigene set and help find novel genes which might be responsible for the target phenotypes, in the case that reference data are either incomplete or unavailable.

Identification of candidate genes for the *Rfd1* locus in radish

We further delimited the interval containing the *Rfd1* locus down to 0.17 cM, using tightly linked markers and a large-sized segregation population in this study. This radish interval corresponded to an 83 kb interval in the syntenic *Arabidopsis* chromosome 3. Twenty-five genes were positioned in this *Arabidopsis* genome, but none of them contained PPR motifs. This result suggests that either the *Rfd1* gene may not be a PPR-coding gene, or the orthologous PPR gene may not be present in the *Arabidopsis* genome. Likewise, the *Rfo* gene responsible for fertility restoration of Ogura CMS in radish encoded a PPR protein, but has no corresponding gene in the *Arabidopsis* syntenic region (Brown et al. 2003). Most plant genomes have been found to contain approximately 400 PPR genes, but a subclass of 10–30 Rf-like PPR genes has been shown to be fast evolving compared to the other PPR genes (Fujii et al. 2011; Dahan and Mireau 2013). Therefore, an Rf-like PPR gene might have been created in the *Rfd1* locus after the divergence of *Arabidopsis* and radish.

Alternatively, as mentioned above, the *Rfd1* locus may not be a PPR gene, since other functions such as alcohol dehydrogenase (Cui et al. 1996), glycine-rich protein (Itabashi et al. 2011), and as of yet unknown functions (Fujii and Toriyama 2009) have also been reported for the Rf genes in other plant species. More than 500 unigenes or contigs containing putative PPR domains were isolated in this study, but we failed to identify a PPR gene which was in linkage disequilibrium with the *Rfd1* locus. However, we still cannot exclude the possibility of having missed the responsible PPR gene in the transcriptome analysis due to a low level of transcription. An upcoming radish genome project will be able to show the exact gene organization of the *Rfd1*-flanking regions. If the genomic sequences of the delimited region were revealed, closely linked markers and recombinants containing crossover events adjacent to the *Rfd1* locus would be useful for cloning the causal gene for fertility restoration of male-sterility conferred by DCGMS cytoplasm.

Author contributions Yong-Pyo Lee carried out sequence analysis, developed molecular markers, and wrote draft manuscript. Youngcho Cho analyzed plant materials and selected recombinants. Sunggil Kim organized and coordinated this research project, and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgments This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project title: Development of molecular breeding technology for major traits in radish and Chinese cabbage, Project No. PJ00799201)", Rural Development Administration and Golden Seed Project (Center for Horticultural Seed Development, No 213003-04-1-SB910). The authors thank Ji-wha Hur, Jeong-Ahn Yoo, and Sujung Kim for their dedicated technical assistance.

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

Ethical standard The authors declare that the experiments comply with the current laws of Republic of Korea.

References

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

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

Bannerot H, Loulidard L, Cauderon MY, Tempe J (1974) Transfer of cytoplasmic male sterility from *Raphanus sativus* to *Brassica oleracea*. In: Proceedings of Eucarpia meeting Cruciferae, pp 52–54

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 99:10887–10892

Bonhomme S, Budar F, Ferault M, Pelletier G (1991) A 2.5 kb *Nco* I fragment of Ogura radish mitochondrial DNA is correlated with cytoplasmic male-sterility in *Brassica* cybrids. Curr Genet 19:121–127

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

Cho Y, Lee Y, Park B, Han T, Kim S (2012) Construction of a high-resolution linkage map of Rfd1, a restorer-of-fertility locus for cytoplasmic male sterility conferred by DCGMS cytoplasm in radish (*Raphanus sativus* L.) using synteny between radish and *Arabidopsis* genomes. Theor Appl Genet 125:467–477

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

Dahan J, Mireau H (2013) The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. RNA Biol 10:1469–1476

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

Dickson MH (1985) Male sterile persistent white curd cauliflower NY 7642A and its maintainer NY 7642B. Hortscience 20:957

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

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 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 108:1723–1728

Gonthier L, Blassiau C, Mörchen M, Cadalen T, Poiret M, Hendriks T, Quillet MC (2013) High-density genetic maps for loci involved in nuclear male sterility (*NMS1*) and sporophytic self-incompatibility (*S*-locus) in chicory (*Cichorium intybus* L., Asteraceae). Theor Appl Genet 126:2103–2121

Grelon M, Budar F, Bonhomme S, Pelletier G (1994) Ogura cytoplasmic male-sterility (CMS)-associated *orf138* is translated into a mitochondrial membrane polypeptide in male-sterile *Brassica* cybrids. Mol Gen Genet 243:540–547

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 (1991) Plant mitochondrial mutations and male sterility. Annu Rev Genet 25:461–486

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:154–169

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, doi: 10.1016/j.mito.2014.02.008

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

Jarl CI, van Grinsven M, van den Mark F (1989) Correction of chlorophyll-defective male-sterile winter oilseed rape (*Brassica napus*) through organelle exchange: molecular analysis of the cytoplasm of parental lines and corrected progeny. Theor Appl Genet 77:135–141

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 Y, Lim H, Ahn Y, Sung S (2009) Identification of highly variable chloroplast sequences and development of cpDNA-based molecular markers that distinguish four cytoplasm types in radish (*Raphanus sativus* L.). Theor Appl Genet 119:189–198

Kim K, Lee Y, Lim H, Han T, Sung S, Kim S (2010) Identification of *Rfd1*, a novel restorer-of-fertility locus for cytoplasmic male-sterility caused by DCGMS cytoplasm and development of simple PCR markers linked to the *Rfd1* locus in radish (*Raphanus sativus* L.). Euphytica 175:79–90

Kim S, Song YH, Lee JY, Choi SR, Dhandapani V, Jang CS, Lim YP, Han T (2011) Identification of the *BrRHP1* locus that confers resistance to downy mildew in Chinese cabbage (*Brassica rapa* ssp. pekinensis) and development of linked molecular markers. Theor Appl Genet 123:1183–1192

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

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Curr Genet 50:149–159

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

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

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

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

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

Lee Y, Park S, Lim C, Kim H, Lim H, Ahn Y, Sung S, Yoon M, Kim S (2008) Discovery of a novel cytoplasmic male-sterility and its restorer lines in radish (*Raphanus sativus* L.). Theor Appl Genet 117:905–913

Lee Y, Kim S, Lim H, Ahn Y, Sung S (2009) Identification of mitochondrial genome rearrangements unique to novel cytoplasmic male sterility in radish (*Raphanus sativus* L.). Theor Appl Genet 118:718–728

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, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMTools. Bioinformatics 25:2078–2079

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

Menczel L, Morgan A, Brown S, Maliga P (1987) Fusion-mediated combination of Ogura-type cytoplasmic male sterility with *Brassica napus* plastids using X-irradiated CMS protoplasts. Plant Cell Rep 6:98–101

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 88:9828–9832

Monaco MK, Stein J, Naithani S, Wei S, Dharmawardhana P, Kumari S, Amarasinghe V, Youens-Clark K, Thomason J, Preece J, Pasternak S, Olson A, Jiao Y, Lu Z, Bolser D, Kerhornou A, Staines D, Walts B, Wu G, D'Eustachio P, Haw R, Croft D, Kersey PJ, Stein L, Jaiswal P, Ware D (2013) Gramene 2013: comparative plant genomics resources. Nucleic Acids Res 42:D1193–D1199

Ogura H (1968) Studies on the new male sterility in Japanese radish, with special reference to the utilization of this sterility towards the practical raising of hybrid seeds. Mem Fac Agr Kagoshima Univ 6:39–78

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

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

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

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

Pelletier G, Primard C, Vedel F, Chetrit P, Remy R, Rousselle P, Renard M (1983) Intergeneric cytoplasmic hybridization in Cruciferae by protoplast fusion. Mol Gen Genet 191:244–250

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

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

Shen D, Sun H, Huang M, Zheng Y, Li X, Fei Z (2013) Radishbase: a database for genomics and genetics of radish. Plant Cell Physiol 54:e3

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

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

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