Theor Appl Genet (2015) 128:259–272
DOI 10.1007/s00122-014-2426-4

ORIGINAL PAPER

Construction of a reference genetic map of *Raphanus sativus* based on genotyping by whole-genome resequencing

Jeong-Hwan Mun · Hee Chung · Won-Hyong Chung · Mijin Oh · Young-Min Jeong · Namshin Kim · Byung Ohg Ahn · Beom-Seok Park · Suhyoung Park · Ki-Byung Lim · Yoon-Jung Hwang · Hee-Ju Yu

Received: 11 June 2014 / Accepted: 2 November 2014 / Published online: 18 November 2014

Abstract

Key message **This manuscript provides a genetic map of *Raphanus sativus* that has been used as a reference genetic map for an ongoing genome sequencing project. The map was constructed based on genotyping by whole-genome resequencing of mapping parents and F_2 population.**

Abstract *Raphanus sativus* is an annual vegetable crop species of the Brassicaceae family and is one of the key plants in the seed industry, especially in East Asia. Assessment of the *R. sativus* genome provides fundamental resources for crop improvement as well as the study of crop genome structure and evolution. With the goal of anchoring genome sequence assemblies of *R. sativus* cv. *WK10039* whose genome has been sequenced onto the chromosomes, we developed a reference genetic map based on genotyping of two parents (maternal *WK10039* and paternal *WK10024*) and 93 individuals of the F_2 mapping population by whole-genome resequencing. To develop high-confidence genetic markers, ~83 Gb of parental lines and ~591 Gb of mapping population data were generated as Illumina 100 bp paired-end reads. High stringent sequence analysis of the reads mapped to the 344 Mb of genome sequence scaffolds identified a total of 16,282 SNPs and 150 PCR-based markers. Using a subset of the markers, a high-density genetic map was constructed from the analysis of 2,637 markers spanning 1,538 cM with 1,000 unique framework loci. The genetic markers integrated 295 Mb of genome sequences to the cytogenetically defined chromosome arms. Comparative analysis of the chromosome-anchored sequences with *Arabidopsis thaliana* and *Brassica rapa* revealed that the *R. sativus* genome has evident triplicated sub-genome blocks and the structure of gene space is highly similar to that of *B. rapa*. The genetic map developed in this study

Communicated by Isobel Parkin.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2426-4) contains supplementary material, which is available to authorized users.

J.-H. Mun
Department of Bioscience and Bioinformatics, Myongji University, Yongin 449-728, Korea

H. Chung · Y.-M. Jeong · H.-J. Yu (✉)
Department of Life Science, The Catholic University of Korea, Bucheon 420-743, Korea
e-mail: yuheeju@catholic.ac.kr

W.-H. Chung · N. Kim
Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 305-806, Korea

M. Oh · B. O. Ahn
Department of Genomics, National Academy of Agricultural Science, Rural Development Administration, Wanju 565-851, Korea

B.-S. Park
The Agricultural Genome Center, National Academy of Agricultural Science, Rural Development Administration, Wanju 565-851, Korea

S. Park
Department of Horticultural Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Suwon 440-706, Korea

K.-B. Lim
Department of Horticultural Science, Kyungpook National University, Daegu 702-701, Korea

Y.-J. Hwang
Department of Life Science, Sahmyook University, Seoul 139-800, Korea

will serve as fundamental genomic resources for the study of *R. sativus*.

Introduction

Raphanus sativus L. includes important root vegetable crop, radish, that is cultivated worldwide and various related wild species having potential of germplasms. The most marketable part of radish is its swollen taproot which provides a source of carbohydrates, phytochemicals, minerals and organic nutrients, and dietary fiber. The radish taproot varies in color (white, red, and black) as well as in size and shape from short and globular type to spindle-shaped, conical, and oblong-elongated type. In addition, there are several cultivars used as oil crops, silique vegetables, and leaf vegetables. *R. sativus* belongs to the tribe Brassiceae of the Brassicaceae family. Systematic and comparative genetic studies have shown that *R. sativus* is a close relative of *Brassica* species (Al-Shehbaz et al. 2006; Hall et al. 2002). Successful development of intergeneric allopolyploid hybrids between *R. sativus* and *B. rapa* (Lee et al. 2011) or *B. oleracea* (Prakash et al. 2009) also supports their close relationship. Phylogenetic studies have suggested that *Raphanus* species may have arisen from hybridization between *B. rapa* (A genome)/*oleracea* (C genome) and *B. nigra* (B genome) lineages (Song et al. 1990; Yang et al. 2002). However, the origin of *Raphanus* is still controversial.

The importance of *R. sativus* in agriculture and plant evolution has motivated efforts to develop tools for genetic and genomic studies on the *R. sativus* genome. One aspect of this effort has been the development of genetic maps. In the early stage of genetic mapping, restriction fragment length polymorphism (RFLP) or amplified fragment length polymorphism (AFLP) markers have played an important role in genetic studies (Bett and Lydiate 2003; Budahn et al. 2009; Kamei et al. 2010; Tsuro et al. 2005). A major focus of recent genetic mapping effort is the simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) of expressed sequence tags (ESTs) including EST-SSR (Shirasawa et al. 2011) and EST-SNP markers (Li et al. 2011). EST-based genetic maps have enabled comparative genome mapping between *R. sativus* and the sequenced genomes of *Arabidopsis thaliana* and *B. rapa*. In parallel to these activities, transcriptome analysis using the Illumina platform (Wang et al. 2012) and development of databases for ESTs, markers, and genetic maps such as RadishBase (Shen et al. 2013) have been reported. In addition, 402 Mb genome sequences of the Japanese cultivar 'Aokubi' have recently been published (Kitashiba et al. 2014). However, the sequences are draft scaffolds with a small size of N50 (46.3 kb) and only 116 Mb of them has been assigned to chromosomes showing incomplete genome assembly.

R. sativus has a diploid genome (RR, $2n = 2x = 18$), and the genome size of *R. sativus* is estimated from 539 Mb (Hall et al. 2002) to 573 Mb (Johnston et al. 2005). Our previous cytogenetic data predicted that the chromosome size and structure of *R. sativus* are similar to those of *B. rapa*. Moreover, the heterochromatin regions are positioned on the centromeric region of whole chromosomes with an estimation of 36.3 % of the total genome (Hwang et al. 2012). These results underlie a strategy for sequencing the *R. sativus* genome based on whole-genome shotgun sequencing. To make a reference genome sequence of *R. sativus*, we have been sequencing the Korean cultivar cv. *WK10039* (510 Mb) which is a typical oriental radish with white skin and large elongated root. As of September 2014, whole-genome scaffold sequences covering 344 Mb with an N50 of 1.22 Mb (67.5 % of the total genome and almost all of the predicted euchromatin) were assembled by next-generation sequencing (NGS) technologies (our unpublished data). A crucial component of successful genome sequencing is the availability of a high-density genetic map of the reference species. Sites representing potential sequence polymorphisms within the assembly as well as previously reported genetic markers can be used to facilitate chromosome assignment of the sequence scaffolds.

Recent advances in NGS technologies have led to the development of rapid genome-wide SNP detection at low costs where sequences are directly used to detect and score SNPs. A new approach, namely genotyping by sequencing (GBS), uses sequence reads from the mapping population being genotyped. In brief, two parental lines are deeply sequenced using NGS sequencers and then SNPs or other variations between them are identified. The mapping population of the two parental lines is sequenced through a reduced-representation or a whole-genome resequencing using NGS technologies. The choice of whether to sequence a reduced fraction or the entire genome depends on several factors including ploidy, genome size, repetitive sequence content, or presence of homeologs. The resulting sequences are used to determine allelic diversity for each individual (reviewed in Deschamps et al. 2012). So far, GBS by whole-genome resequencing has been performed on small genome species including *Arabidopsis* (Cao et al. 2011; Ossowski et al. 2008) and rice (Huang et al. 2009; Subbaiyan et al. 2012; Xu et al. 2011), whereas reduced-representation resequencing has been applied to complex genomes such as maize (Elshire et al. 2011), wheat (Trick et al. 2012), potato (Uitdewilligen et al. 2013), and soybean (Sonah et al. 2013).

In this study, we aimed to develop a reference genetic map of *R. sativus* to provide a framework for anchoring and ordering the sequence scaffolds enabling their assembly into pseudomolecules of the genome. Here, we report on the construction of a high-density SNP map of *R. sativus* by whole-genome

resequencing of the mapping parents (*WK10039* and *WK10024*) and 93 individuals of the F_2 population. Relatively small genome size and availability of the genome sequences of *R. sativus* enabled us to construct a genetic map based on whole-genome resequencing. The resulting genetic map facilitated anchoring of sequence assemblies as well as sequenced BAC clones for the whole genome of *R. sativus*. In addition, the characteristics of the gene space and overall genomic structure of *R. sativus* are described.

Materials and methods

Plant materials

Two parental inbred lines of *R. sativus*, cv. *WK10039* (maternal line and the reference cultivar investigated in our genome sequencing project) and cv. *WK10024* (paternal line), were chosen to develop a mapping population. Cultivar *WK10039* is a typical winter type with white cylindrical root and green siliques, whereas cv. *WK10024* has red round root and pink siliques. Both lines were obtained from the National Institute of Horticultural & Herbal Science (RDA, Korea) and were self-pollinated for eight generations before crossing. A single plant of cv. *WK10039* was used as the female partner and a single plant of cv. *WK10024* was used as the male partner to produce F_1 hybrids whose genotypes were confirmed by co-dominant gene-based markers. A single F_1 plant was selected and self-pollinated to produce the segregating F_2 population. A total of 93 individuals were randomly selected as the F_2 mapping population for segregation analysis and genetic mapping. Seeds of parental lines and F_2 individuals were surface sterilized in 12 % sodium hypochlorite and germinated on 0.5× Murashige and Skoog (MS) agar plates (0.7 %) in a growth chamber at 22 °C with a 16-h light/8-h dark cycle and 60 % humidity. Seedlings harvested 5 days after germination were transferred into soil and grown in a culture room under the same growth conditions.

Genomic DNA extraction and Illumina sequencing

Genomic DNA (gDNA) was extracted from the leaves of each plant using the DNeasy Plant MaxiKit (Qiagen, USA) and standard paired-end (PE) libraries with 500 bp inserts were constructed for Illumina sequencing. In brief, extracted gDNA was sheared by Covaris and the fragmented DNA was purified with the QIAquick PCR Purification Kit (Qiagen, USA) followed by end repair and index adapter ligation using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA). The resulting libraries were size selected and then pooled with up to six samples per flow cell lane for cluster generation on the Illumina flow cell using the cBot. The library clusters were sequenced in the Illumina HiSeq1000 sequencer to generate paired-end sequences (2 × 100 bp) using the TruSeq™ SBS Kit v3-HS. Reads from each library were collected at a minimum Illumina quality score of 31 and then filtered for adapter contamination, ambiguous residues (N residues), PCR duplicates, and low-quality regions.

Mapping reads, filtration, and polymorphism detection

All PE reads generated from the parental lines were aligned to the *WK10039* draft genome sequence and BAC sequences (Table 1; NCBI BioProject Accession PRJNA239785) using Burrows Wheeler Aligner (BWA, version 0.5.9) with default options (Li and Durbin 2009). After initial mapping of the raw reads, the reads that mapped to multiple locations or unmapped reads were excluded. In addition, aligned reads considered to be PCR duplicates were also removed using the MarkDuplicates in the Picard software package 1.48 (McKenna et al. 2010). Base quality scores were recalibrated with CountCovariates and TableRecalibration, and regions near short Insertion or Deletion (InDel) were realigned with IndelRealigner in the Genome Analysis Toolkit (GATK, version 1.0.5974). The realigned and recalibrated Sequence Alignment/Map (SAM) files produced by these processing steps were used

Table 1 Summary of reference sequences for *WK10039* and NGS reads used in this study

Class	Plant	Mapping population	Sequence type	NGS	Number	Length (Mb)	Coverage (X)[a]
Reference genome	*WK10039*	Maternal	Genome scaffold	GS-FLX Plus	1,414	344.0	0.7
			BAC sequence[b]	GS-FLX Plus	188	23.6	0.05
WGS	*WK10039*	Maternal	PE read	HiSeq1000	427,289,276	43,156.2	84.6
	WK10024	Paternal	PE read	HiSeq1000	394,451,376	39,839.6	78.1
	93 progeny[c]	F_2	PE read	HiSeq1000	62,922,175	6,355.1	12.5

WGS whole-genome shotgun, *PE* paired-end

[a] Genome coverage was calculated with the genome size of *WK10039* as 510 Mb

[b] Phase 2 sequence with ordered contigs per BAC clone

[c] Average values are presented

for SNP, InDel, and SSR detection. We used UnifiedGenotyper in GATK for calling variants such as SNP, InDel, and SSR. Polymorphisms between the parental lines were called with a standard call confidence (-stand_call_conf) set to 30.0 and standard emit confidence (-stand_emit_conf) set to 10.0. To reduce the false discovery rate of SNP, InDel, and SSR, raw variant calls were filtered out using VariantFiltration in GATK under the following conditions: base quality score of ≤50.0, mapping quality (MQ) ≤30, strand bias (SB) ≥−1.0, MQzero reads (MQ0) ≥4, and MQ0 divided by depth (MQ0/DP) ≥0.1. In addition, InDels with at least 10 bp size difference and SSRs of di- and tri-nucleotide repeats were considered for further analysis. The SNP information was saved asvcf files, whereas InDel and SSR information was converted into SAMtoolsmpileup format (http://samtools.sourceforge.net).

SNP marker selection and genotyping of the F_2 population

To obtain high-confidence SNP markers for genetic map construction, all SNPs identified between the parental lines were filtered as follows: length of the reads mapped to the SNP sites >75 bp, depth at each SNP in both parental lines >7, mapping quality of reads >20, and base quality >30. We further selected SNPs which were 100 %homozygous alleles in each parental line. The final candidate SNPs were used for genotyping of the F_2 population. For genotype calling, reads of F_2 individuals were filtered as follows: read depth ≥8, mapping quality of reads >20, base quality >30, and default values for the other options. SNPs that had >30 % (28 plants) missing or ambiguous data were excluded. In addition, a position was considered to be homozygous if more than 90 % of the nucleotides at the SNP position were the same. A heterozygous position had two different alleles with a frequency between 0.3 and 0.7. For multi-sample genotyping, every 93 individuals were grouped into 4 groups and genotypes were determined as follows: homozygous maternal (*WK10039*) as "A", homozygous paternal (*WK10024*) as "B", heterozygous as "H", and missing or ambiguous data as "-."

InDel and SSR marker development

To discover InDel and SSR regions for PCR-based marker development, we used BWA, Picard, and GATK tools. The polymorphic loci between *WK10039* and *WK10024* were filtered as follows: read depth of >7, length difference between the parental lines ≥10 bp, flanking length of >150 bp, and SNP variation within the flanking region <4. Size differences of more than 10 nucleotides between the parental lines were considered as InDels. For SSRs, di- and tri-repeats with ≥12 nucleotide differences were selected for further analysis. Sequences from the candidate regions were extracted by SAM tools. Primers targeting the selected regions were designed using Primer3 software (Rozen and Skaletsky 1999) with a target amplicon size range of 100–250 bp. Primer specifications were as follows: melting temperature (T_m) 55–60 °C, GC content 40–60 %, and primer length 18–27 nucleotides. PCR amplification was performed using Tenuto 2X premix (Enzynomics, Korea), 0.5 μM of each primer, and 20 ng genomic DNA. The PCR cycle conditions were 5 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C, and a final 5 min extension step at 72 °C. The PCR products were separated on a 3 % agarose gel in 1× sodium borate buffer and bands were visualized by ethidium bromide.

Construction of genetic map

To construct a genetic map, the JoinMap 4 (van Ooijen 2006) program was used. Linkage analysis and marker grouping were performed at a logarithm of odds (LOD) threshold of 10 and genetic distance between the markers was calculated by the Kosambi map function (Kosambi 1944). SNP (RsSNP), InDel (RsInD andRsIBP), and SSR (RsSSR) markers as well as previously reported markers we developed based on *B. rapa* sequences (BrEST and RsSSR) (Chung et al. 2014; Park et al. 2010) were integrated to construct a genetic map, and the markers which were positioned at the same loci were manually curated. The genetic position of each BAC clone was determined based on the selected BAC SNP (RsH) markers. To evaluate co-linearity of markers in the linkage groups with their physical positions, each linkage group on the genetic map was aligned with the genome assembly of *WK10039*. Consensus sequences of SNPs were searched with BLASTN at an *E* value cutoff of $1E^{-10}$ and sequence coverage greater than 85 % against the genome sequence.

Florescent in situ hybridization of BAC clones

Mitotic chromosomes of *WK10039* were prepared as described previously (Hwang et al. 2012; Lim et al. 2005). DNA (1–2 μg) extracted from BAC clones, 45S rDNA, and 5S rDNA was labeled with either biotin-16-dUTP or digoxigenin-11-dUTP by nick translation (Roche, Germany), and FISH was performed as described by Lim et al. (2005). The hybridization mixture consisted of 50 % deionized formamide, 10 % dextran sulfate, 2× SSC, and salmon sperm DNA. The probes were mixed to a final concentration of 20 ng/mL and then denatured at 70 °C for 10 min. The slides were subsequently washed in 2× SSC at room temperature for 5 min and 0.1× SSC at 42 °C for 30 min. The fluorescence signal was detected using FITC-conjugated anti-digoxigenin antibodies (Roche, Germany) and streptavidin Cy3 (Zymed Lab., USA). The slides were

then mounted and counterstained in Vectashield containing 2 μL/mL 4′,6-diamidino-2-phenylindole (DAPI). The images were observed using an Olympus BX 61 fluorescent microscope equipped with a CCD camera and analyzed with the Genus Image Analysis Workstation (Genus version 3.8, Applied Imaging Corporation, UK).

DNA sequence analysis and genome comparison

BAC sequences were masked using Repeat-Masker (http://www.repeatmasker.org) and gene-coding regions in the BAC sequences were predicted by the Fgenesh+ (http://www.softberry.com) program with parameters trained using the *R. sativus* matrix. Predicted gene models were searched against a database of plant transposon-encoded proteins (Nussbaumer et al. 2013). Predicted proteins with a top match to transposon-encoded proteins were excluded from the annotation and gene counts. For comparative genomics, scaffold sequences containing SNPs and sequence-tagged site (STS) genetic markers were aligned to the genome of *A. thaliana*. Alignment was performed using LASTZ (version 1.03.02) with the gapped alignment option. Only LASTZ matches longer than or equal to 300 bp with identity >80 %, and LASTZ score >20,000 were used. Analysis of chromosomes based on a syntenic dot matrix plot was performed using an internally developed visualization program.

Results

Mapping population and reference sequence assembly

To facilitate genetic mapping, we developed an F_2 population composed of 93 individuals. The F_2 mapping population was derived from a single F_1 plant based on a cross between cv. *WK10039*, the reference species of our genome sequencing project, and cv. *WK10024*. Because the alleles of the F_1 plant are assumed to segregate according to Mendelian rules in the F_2 population, co-dominant markers are expected to segregate at 1:2:1 ratio of homozygous *WK10039*:heterozygous *WK10039* and *WK10024*:homozygous *WK10024*. We applied 109 co-dominant intron-target markers and identified 89 markers (82 %) showing a goodness-of-fit ratio (1:2:1, $P < 0.005$). This result suggested that the segregation of the markers in the F_2 population was not distorted significantly from the expected ratio.

As a reference sequence to identify polymorphic loci between the mapping parents, we used a draft genome assembly of cv. *WK10039*. The reference genome sequences of cv. *WK10039* were previously generated from the ongoing genome sequencing project. Whole-genome scaffold sequences of cv. *WK10039* were assembled using single and mate-paired reads of 454 sequences, PacBio RSII sequences, and end sequences of ~23,000 BAC clones. In brief, approximately 50× raw sequences of single (1.5 kb) and mate-paired (3, 8, and 20 kb) DNA sequences were generated using Roche/454 GS-FLX Plus (average read length of ~600 bp). Approximately 24X PacBio RSII sequences and ~23,000 BAC-end ABI sequences were also generated. The sequence reads were assembled into scaffolds using Newbler assembler, and homopolymers in the scaffolds were corrected using 85× paired-end short-read sequences from Illumina HiSeq1000. The resulting genome sequence assembly included 1,414 scaffolds with an N50 of 1.22 Mb and spanned 344 Mb, representing 67.5 % of the genome (our unpublished data). In addition, sequences of 188 gene-rich BAC clones were also determined as phase 2 sequences with ordered contigs per clone. All the sequence data used in this study are summarized in Table 1.

Sequence variations between the parental lines and PCR-based marker development

For identification of polymorphic loci between the mapping parents, we generated >394 million Illumina PE reads from the paternal line *WK10024*. These reads corresponded to 39 Gb sequence data representing approximately 78× genome coverage. In addition, >427 million Illumina PE reads of *WK10039* covering 85× of the genome were available from our genome sequencing activity. Using the bioinformatics pipeline that allowed read mapping to the reference genome and identification of polymorphic alleles between the parental lines (Fig. 1), we aligned 60.8 Gb of high-quality sequence reads from the parental lines to 326.6 Mb (86 %) of the reference scaffold sequences, providing an average read depth of 99× when all the reads were considered together. As a consequence of the read mapping and identification of polymorphic loci, we found a total of 14,741 length variants (1,039 InDels and 13,702 SSRs; Table 2) and 1,258,476 SNPs between the parental lines. The average density of polymorphic loci in the genome was 3.9/kb which is almost two times higher than *A. thaliana* (2/kb; Cao et al. 2011).

With the goal of establishing genetic map positions for sequence assembly scaffolds, we designed PCR primers flanking InDels and SSR motifs. Among the 14,741 length variants, 265 InDel (RsInD) and 206 SSRs (RsSSR) were randomly selected to test PCR amplification. Additional PCR primers that were developed based on BAC-end sequence information from *WK10039* (RsIBP; our unpublished data) and tissue-specific ESTs from *B. rapa* (BrEST; Yu et al. 2011) were applied. Moreover, we also screened previously reported EST-SSR markers from *R. sativus* including 610 RSS (Shirasawa et al. 2011). However, most

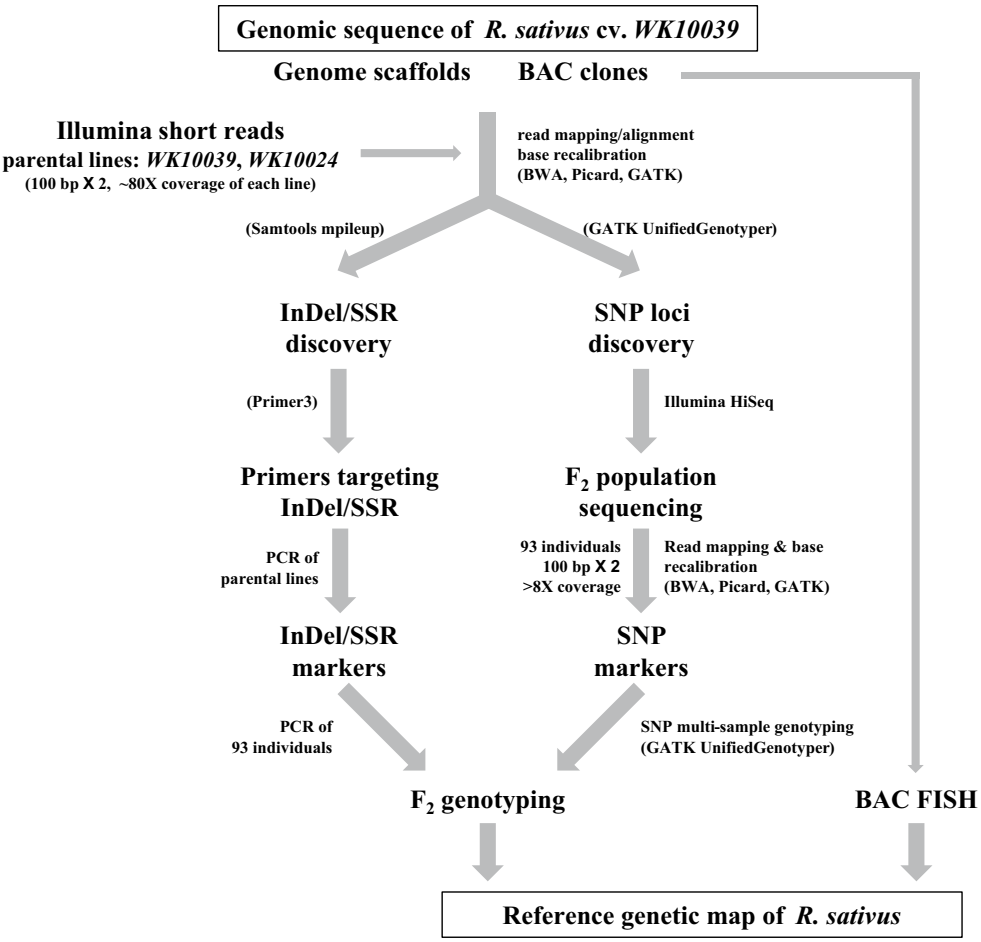

Fig. 1 Workflow to develop genome-wide InDel/SSR and SNP markers based on whole-genome resequencing. More than 80× Illumina reads generated from each mapping parent (*WK10039* and *WK10024*) were mapped to the reference genome sequences of *WK10039* (genomic scaffolds and BAC sequences) and polymorphic loci were identified. For PCR-based markers, primers targeting InDels and SSRs were designed. For SNP markers and genotyping, resequencing of the F_2 population was performed with 8× genome coverage and high-confidence SNP loci were selected. Candidate polymorphic markers were genotyped using 93 individuals of the F_2 population followed by linkage analysis and genetic map construction. For integration of cytogenetic and genetic linkage maps, FISH analysis of BAC clones was performed. The bioinformatics programs used in this study are presented in *parentheses*

of the RSS primers failed PCR amplification or did not show polymorphisms between *WK10039* and *WK10024*. All primer sets showing polymorphisms between the parent lines were selected as PCR-based markers. In total, 150 primer pairs including 92 RsInD, 20 RsSSR, 11 RSS, 22 RsIBP, and 5 BrEST markers were selected to construct a genetic map.

SNP genotyping by whole-genome resequencing of the F_2 population

To generate a substantial number of reads for increased target coverage, at least 41 million PE reads for the F_2 progeny were produced by Illumina sequencing. A total of 5,851,762,272 reads corresponding to 591 Gb were generated from 93 individuals. The number of reads varied from 41,207,730 to 112,117,888 (average of 62,922,175) covering 8× to 22× of the genome. All reads passing quality filter were mapped to the reference sequences. A total of 71 % of the reads were mapped to the reference scaffolds. Although the number of mapped reads varied among individuals, genome coverage of the reads was not significantly varied representing an average of 87.5 % coverage with an average read depth of 13.6 (Supplemental Figure S1 and Supplemental Table S1).

According to the genotyping criteria of the SNP position, we considered all SNPs having genotypes for at least 65 individuals with more than 8 reads. Consequently, we obtained a total of 772,678 SNP loci from the reference sequences. To develop high-confidence SNP markers from the candidate SNPs, we further filtered out the SNPs that were unable to define the plant genotype as homozygous or heterozygous at the nucleotide frequency cutoff. As a result, a total of 16,069 genomic SNP markers from 1,353 genome scaffolds (258.9 Mb) along with 188 BAC SNP and 25 conserved ortholog set (COS) SNP markers were selected for genetic mapping (Supplemental Table S2 and Supplemental Fig. S2). The transition/transversion ratio of genomic SNP markers was 1.39, thus, showing a high rate of transition similar to other plant species (Table 3).

Genetic map construction

For purposes of genetic mapping of polymorphic markers and anchoring the sequence scaffolds of *R. sativus*, the genotypes of 16,282 SNP and 150 PCR-based markers in the

Table 2 Summary of polymorphic InDels and SSRs identified between the mapping parents, *WK10039* and *WK10024*

	Characteristics	Number
InDel[a] size difference (bp)	10	264
	11	199
	12	180
	13	88
	14	98
	15	58
	16	40
	17	25
	18	17
	19	19
	≥20	51
SSR[b] motif type	AT/TA	5,775
	AG/GA/CT/TC	2,867
	AC/CA/TG/GT	501
	AAG/AGA/GAA/CTT/TTC/TCT	1,393
	ATG/TGA/GAT/CAT/ATC/TCA	746
	AAT/ATA/TAA/ATT/TTA/TAT	606
	AGG/GGA/GAG/CCT/CTC/TCC	542
	AAC/ACA/CAA/GTT/TTG/TGT	484
	ACC/CCA/CAC/GGT/GTG/TGG	336
	AGC/GCA/CAG/GCT/CTG/TGC	175
	AGT/GTA/TAG/ACT/CTA/TAC	168
	ACG/CGA/GAC/CGT/GTC/TCG	62
	GGC/GCG/CGG/GCC/CCG/CGC	47
Total		14,741

[a] InDels of ≥10 bp size difference were considered

[b] SSRs of di- and tri-repeats with ≥12 nucleotides were considered

Table 3 SNP characteristics of transition or transversion identified between the mapping parents, *WK10039* and *WK10024*

DNA substitution types [*WK10039*/*WK10024*]	Number (%)
Transition	
[G/A]	4,556 (28.4)
[C/T]	4,776 (29.7)
Sub-total	9,332 (58.1)
Transversion	
[C/G]	1,018 (6.3)
[T/A]	2,274 (14.2)
[A/C]	1,691 (10.5)
[G/T]	1,754 (10.9)
Sub-total	6,737 (41.9)
Total	16,069 (100)

F_2 mapping population were scored. SNP genotypes from the same scaffolds were analyzed together and the redundant SNPs showing same genotypes were collapsed into a major SNP. As a result, we selected 2–3 of the best SNPs per scaffold for linkage analysis. Genotypes for all markers were integrated into a color-coded genotype matrix using Excel (Kiss et al. 1998). An initial linkage analysis was performed with a total of 2,678 markers including 2,319 genomic SNP, 188 BAC SNP, 25 COS SNP, and 146 PCR-based markers using the JoinMap 4 program. For further analysis, the redundant markers showing identical genotypes were collapsed to a single representative framework marker (Supplemental Table S3), and the markers locating on multiple mapping positions were excluded. As a result, 2,637 out of 2,678 markers (99 %) were grouped into 9 linkage groups at LOD 10. As shown in Fig. 2, a genetic map was derived from the 1,000 non-redundant framework genetic markers including 689 genomic SNPs, 168 SNPs from BAC clones, 25 COS SNPs, and 118 PCR-based markers. The genetic map spanned 1,538 cM with an average distance of 1.5 cM between markers (Table 4). We assigned the name of each linkage group according to the R groups of previously reported genetic maps (Jeong et al. 2014; Shirasawa et al. 2011). The biggest linkage group R9 contained 136 markers spanning 263.5 cM, while the smallest group R8 had 75 markers covering 99.4 cM. A detailed list of marker attributes along with the genetic positions is given in Supplemental Table S4. The sequence defined genetic markers (framework markers and their equivalent markers) anchored 390 assembly scaffolds accounting for ~275 Mb and 168 BAC clones covering ~20 Mb on 9 chromosomes. Therefore, the extent of the genome sequence associated with genetic loci was ~295 Mb or 85.8 % of the genome assembly. Considering the estimated genome size of *WK10039* as 510 Mb by K-mer analysis (our unpublished data), 1 cM of the current genetic map corresponded to 331.6 kb. Figure 3 shows the correlation between the genetic map and the draft genome sequence of cv. *WK10039*. The physical position of the markers was determined by a BLASTN search. Most of the markers in the linkage groups showed co-linearity with their physical position, especially with R1, R3, R4, R5, and R6. However, uneven distributions of markers were found in R2, R8, and R9, suggesting that there may be a biased distribution of makers presumably due to centromeres or heterochromatin.

Anchoring of BAC clones on the genetic map enabled us to align the linkage groups and chromosomes of *R. sativus*. Although the chromosomes of cv. *WK10039* are small and compact, each chromosome can be distinguishable based on 5S and 45S rDNA hybridization signals and morphological characteristics including chromosome size and the chromosome arm length ratio (Hwang et al. 2012). To establish correspondence between each linkage group and chromosome, we performed FISH analysis of BAC clones that were anchored on the genetic map. Consequently, nine BAC clones (one BAC clone per linkage group) were

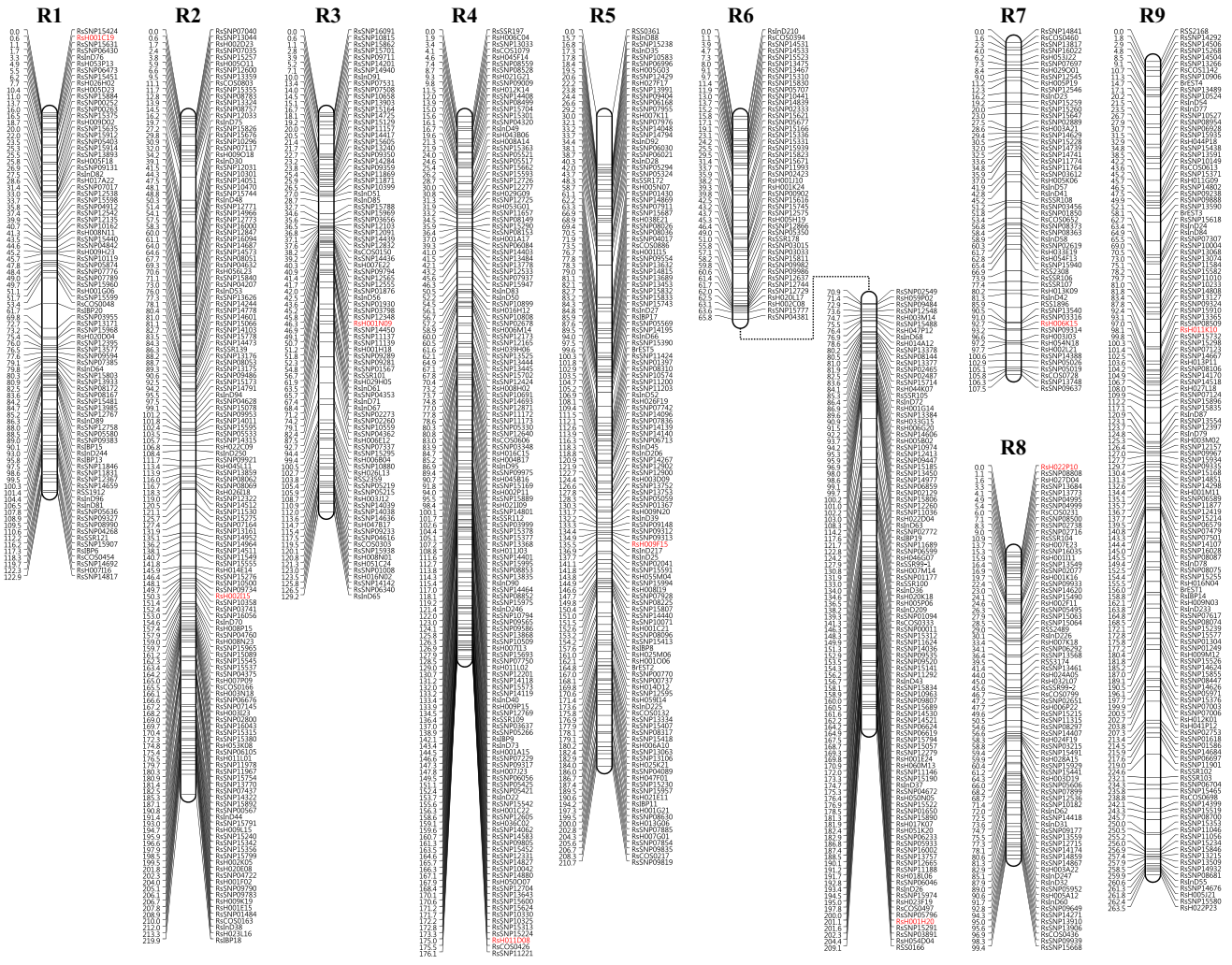

Fig. 2 Genetic map of *R. sativus.* Nine linkage groups were assigned according to previously reported RsCOS (Jeong et al. 2014) and RSS markers (Shirasawa et al. 2011). Map distances calculated by Kosambi map units (cM) are shown on the *left* margin and the marker names are shown on the *right* margin of each linkage group. Marker types are indicated with the following prefixes: RsInD, InDel; RsSSR, SSR; RsIBP and RsCOS, gene-based; BrEST, *B. rapa* EST; RSS, EST-SSR. Markers developed from BAC clones that were mapped to chromosomes by FISH analysis are *highlighted in red*

Table 4 Summary of the reference genetic map of *R. sativus* constructed using SNP, InDel/SSR, and COS markers

Group	No. linked framework markers				Framework markers per group	Length (cM)	Average length/marker (cM)
	Genomic SNP	BAC SNP	COS SNP	PCR-based markers			
R1	58	12	2	13	85	122.9	1.4
R2	104	23	3	11	141	219.9	1.6
R3	64	12	2	10	88	129.2	1.5
R4	99	28	3	13	143	176.1	1.2
R5	82	24	3	20	129	210.7	1.6
R6	101	28	3	15	147	209.1	1.4
R7	31	12	3	10	56	107.5	1.9
R8	47	15	3	10	75	99.4	1.3
R9	103	14	3	16	136	263.5	1.9
Total	689	168	25	118	1,000	1,538.2	1.5

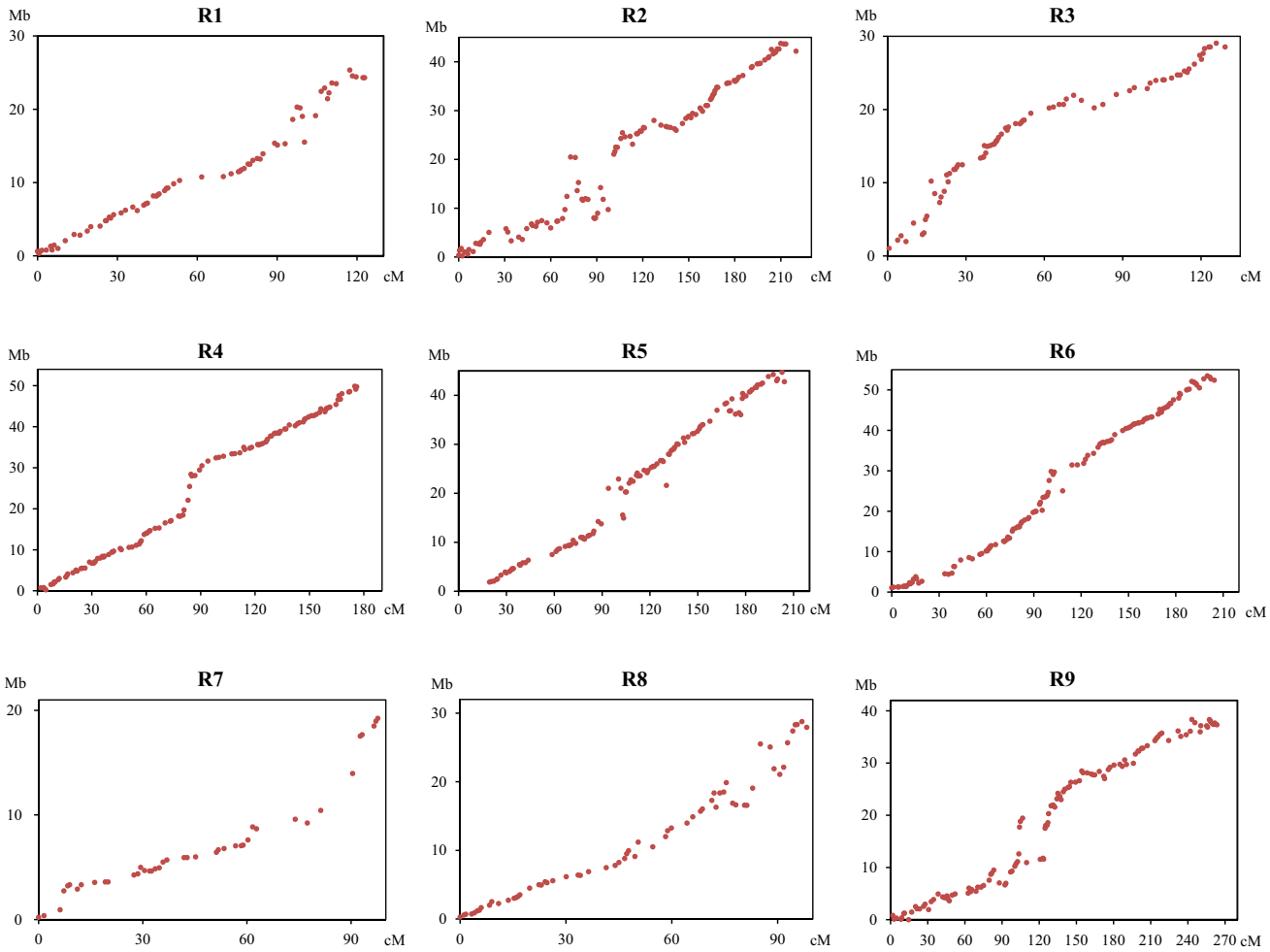

Fig. 3 Correlation between the genetic map and the draft genome sequence of *R. sativus* cv. *WK10039*. The physical position of each marker was identified in the reference genome based on a BLASTN search. *X* axis denotes genetic distance (cM) and *Y* axis indicates physical distance

hybridized on metaphase chromosomes (Fig. 4). Using the fluorescent signals of BAC clones detected, the destination of each BAC clone was determined based on the size and morphological characteristic of the chromosome, which were previously numbered from the largest (a) to the smallest (i) chromosomes. In addition, the orientation of each linkage group could be revised based on the orientation of the chromosome deduced from the FISH signal. These results confirmed that analysis of BAC clones by FISH is highly useful for distinguishing chromosomes as well as integrating genetic and cytogenetic maps.

Characteristics of the gene space in the *R. sativus* genome and comparative analysis

To overview the characteristics of the gene space in the *R. sativus* genome, gene prediction was carried out for the sequenced BAC clones. *Ab initio* gene prediction of 188 BAC clones by the Fgenesh+ program resulted in the construction of 4,957 gene models. The gene structure and density statistics are shown in Table 5. The exon–intron composition and average length of the *R. sativus* genes were highly similar to those of *B. rapa* genes. Moreover, the average gene density of 1 per 4.8 kb of *R. sativus* BAC clones was also similar to that of the gene-rich BAC clones of *B. rapa* (1 per 4.2 kb) (Mun et al. 2009) indicating that the gene-containing regions of *R. sativus* occupy almost the same space as *B. rapa*. Considering the overall gene density of the *B. rapa* genome (1 per 6.9 kb) and the coverage of euchromatin in *R. sativus*, the overall number of protein-coding genes in the *R. sativus* genome was estimated up to ~47,000.

A large-scale sequence level comparison between chromosome-assigned *R. sativus* sequence scaffolds and the whole genome of *A. thaliana* demonstrated that there was a three-to-one syntenic match between five chromosomes of *A. thaliana* and nine chromosomes of *R. sativus*, indicating genome-wide triplication of the *R. sativus* genome (Fig. 5). Similar to the

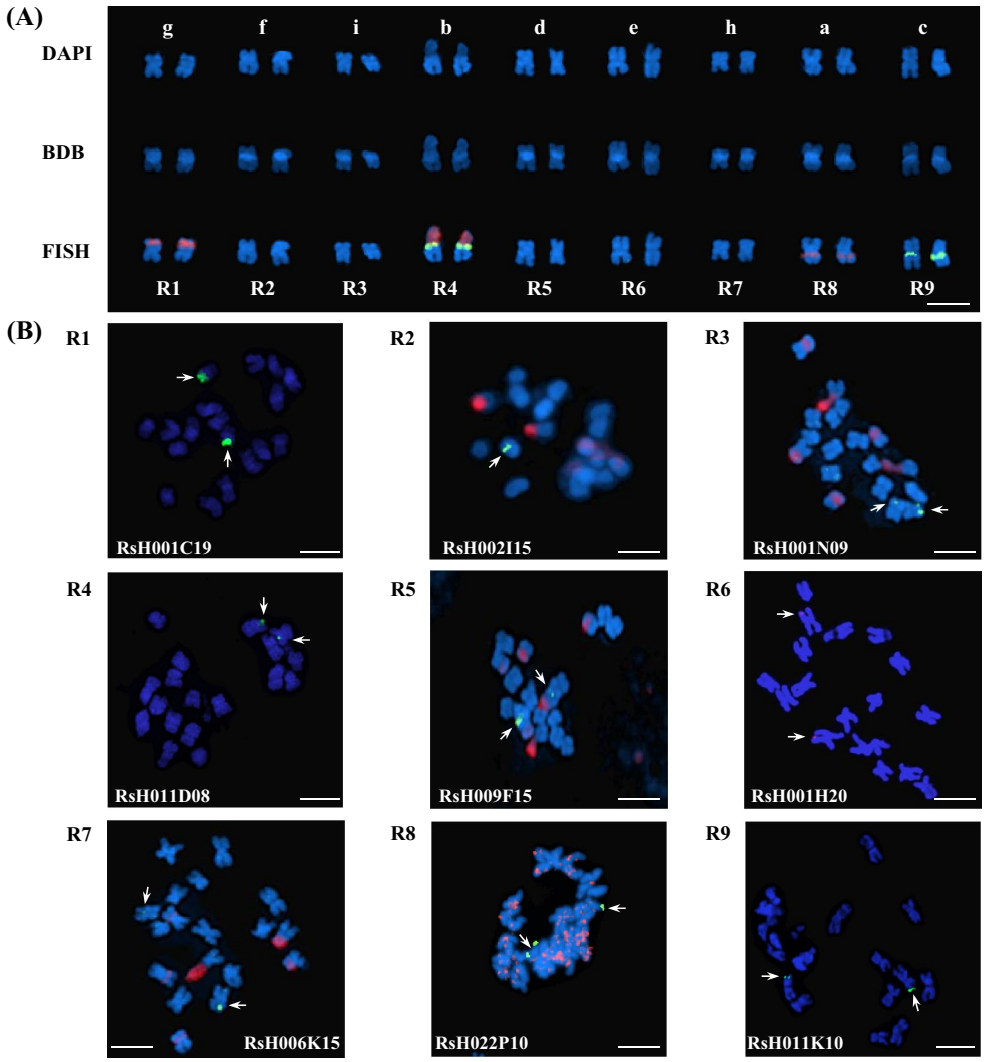

Fig. 4 Chromosome karyotype of *R. sativus* cv. *WK10039* and FISH mapping of BAC clones on mitotic metaphase chromosomes. **a** Mitotic metaphase chromosomes stained with DAPI, bleached DAPI (BDB) technique, and 5S (*green*) and 45S (*red*) rDNA. Chromosomes are arranged according to linkage group name (*R1–R9*) and previous IDs (a toi) by Hwang et al. (2012) are also provided on the *top* margin. *Bar* 3 μm. **b** FISH analysis of chromosomes using BAC clones that were mapped to linkage groups of the genetic map. *Arrows* to *green* or *red* areas point to the position of the hybridization signal of BAC clones. *Red* signals (*R2*, *R3*, *R5*, and *R7*) are 45S rDNA. *Bars* 5 μm

results from the *B. rapa* and *A. thaliana* genome comparison, *R. sativus* and *A. thaliana* genomes share a minimum of seven large-scale triplicated syntenic block sets. Within the individual syntenic block set, the coverage of syntenic region (sum of aligned sequences divided by the total length of sequences in the blocks) was different among the triplicated blocks; one has higher coverage than the others (Supplemental Table S5). Taken together, it is evident that the diploid *R. sativus* genome had a triplicated genome structure that consisted of three sub-genome blocks with different fractionation. Additional sequence information will present a more detailed view of synteny between the genomes.

SNP markers continue to be added to the genetic map, furthering the integration of genetic and physical scaffolds in this species and providing additional anchoring and assembly validation for the ongoing genome sequencing effort. Together with further scaffolding of the current assembly using clone-end sequences such as fosmid-end and BAC-end sequences, the resulting assembly will allow construction of more accurate chromosome pseudomolecules.

Discussion

Sequence-based genetic maps provide the framework for sequencing projects because they indicate the position of sequences and guide assembled sequences to their specific locations in linkage groups. Of particular importance,

the positioning of sequence assemblies to cytogenetically defined chromosomes greatly aids the production of chromosome pseudomolecules (Mun et al. 2006). The ultimate goal of our genome sequencing project in *R. sativus* is to provide pseudochromosome arms that cover the entire gene space of the whole genome. As of May 2014, a draft sequence of 'Aokubi', a Japanese cultivar of *R. sativus*, has been published (Kitashiba et al. 2014). It is noteworthy that a total of 402 Mb scaffold sequences were generated; however, only a part of them (116 Mb) has been assigned to the linkage map showing insufficient anchoring of assembled sequences to produce chromosome pseudomolecules. In this regard, determination of the order and orientation of sequence scaffolds on each chromosome are crucial to the final assembly of the ordered scaffolds. Therefore, the primary objective of this study was to develop a reference genetic map to anchor the scaffolds generated from whole-genome shotgun sequencing directly onto specific chromosomes and chromosome positions.

Genetic mapping of sequenced genome regions has been performed using STS markers. Genomic STS markers identified from genome sequences provide immediate links between sequence assembly and the genetic map. Previously reported STS markers can be applied for genetic mapping; however, PCR amplification of STS markers in different species or cultivar often fails resulting in limited cross-species/cultivar utility. For instance, we applied 610 previously reported RSS markers to the mapping parents but only 20 markers (3.3 %) showed polymorphism. Therefore, an effective development of STS markers should include direct analysis of genomic sequence data. One of the co-dominant STS markers of frequent choice has been genomic SSRs. Automated identification of SSRs in DNA sequence databases using public software programs has become

Table 5 Overall composition of predicted protein-coding genes in the gene-rich BAC sequences of *R. sativus* cv. *WK10039*, *B. rapa*, and *A. thaliana* genomes

Features	*R. sativus*	*B. rapa*[a]	*A. thaliana*[b]
Number of BACs	188	–	–
Total sequence length (Mb)	23.6	283.8	119.1
Transposon (%)	20.1	39.5	20.9
Number of protein-coding genes	4,957	41,090	27,411
Number of exons per gene	5.1	5.0	5.9
Average intron size (bp)	202	209	165
Average exon size (bp)	228	233	296
Average gene size (kb)[c]	2.0	2.0	2.6
Average gene density (kb/gene)	4.8	6.9	4.3

[a] *B. rapa* genome annotation v1.0 was used

[b] TAIR10 genome statistics was used

[c] Average gene size was defined as average length of coding region including introns

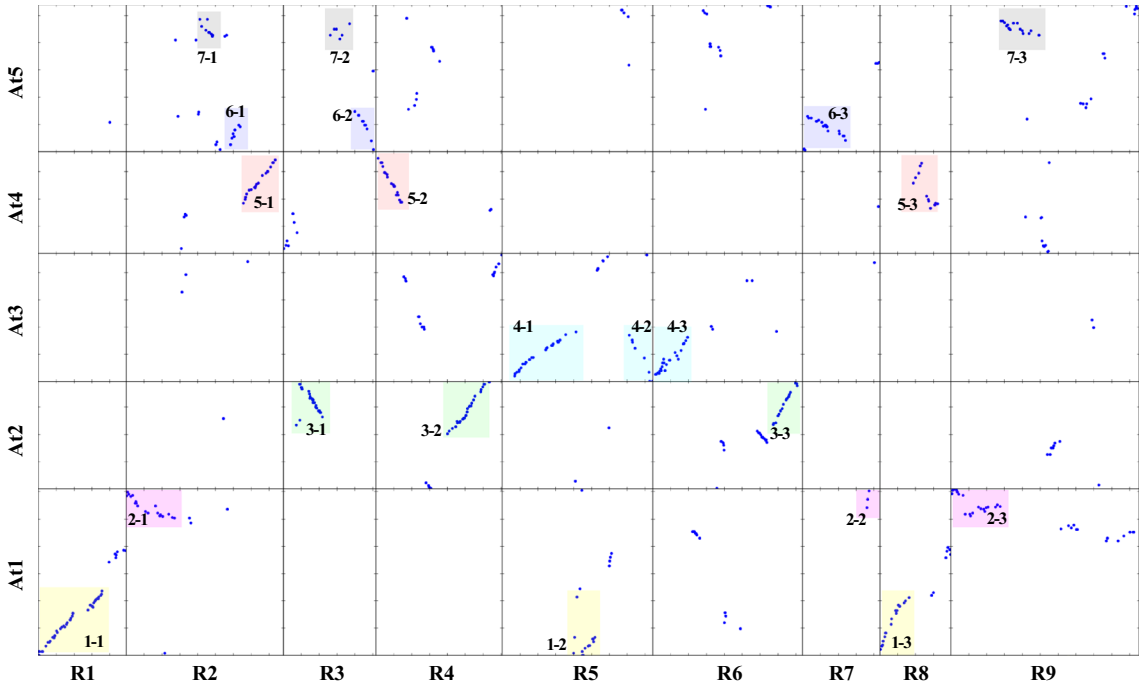

Fig. 5 Chromosome correspondence between *R. sativus* and *A. thaliana* represented as a *dot* plot. Each dot demonstrates the best LASTZ match between sequence pairs. Scale unit of *X* and *Y* axes is 5 Mb. The *color-shaded boxes* in the *dot* plots represent long-range synteny blocks along chromosome pairs. *Boxes* with the *same color* are putative triplicated sub-genomic blocks

commonplace in a wide variety of plant species. Polymorphic SSR alleles can be scored in agarose or polyacrylamide gels or capillary sequencers; however, genotyping of the mapping population still has limited throughput, high cost, and is time consuming due to PCR amplification of the target alleles. Recently, a major focus of genome mapping efforts has been SNPs because there are vast numbers of SNPs in every genome enabling highly detailed genome maps to be constructed. Furthermore, SNPs have a lower error rate compared with SSRs (Jones et al. 2007) and rapid, high throughput detection and genotyping methods have been developed. These methods include the Illumina GoldenGate assay, oligonucleotide hybridization analysis using high-density SNP chips, and NGS genotyping (Edenberg and Liu 2009). In particular, whole-genome resequencing of mapping populations using NGS technologies with a known reference has created a powerful way to construct a high-density SNP map covering the entire genome (Sonah et al. 2013). Although it is still expensive to carry out on a large set of genotypes, resequencing efforts are suitable for SNP genotyping not only for small genome species including *Arabidopsis* (Cao et al. 2011) and rice (Subbaiyan et al. 2012; Xu et al. 2011) but also for large genome polyploidy species such as *B. napus* (Huang et al. 2013).

The current work has contributed to a successful application of NGS genotyping by resequencing. It is noteworthy that the genome of *R. sativus* has been indicated as a polyploid genome (Jeong et al. 2014; Li et al. 2011; Shirasawa et al. 2011). Polyploid genomes present challenges to genome-wide SNP discovery due to the presence of multiple homoelogous sequences (Bancroft et al. 2011; Huang et al. 2013; Lai et al. 2012). To reduce false SNPs from non-allelic paralogous variations, we applied three stringent conditions and identified polymorphic alleles with high confidence. First, we selected only single-hit reads that uniquely mapped to the reference sequences for further analysis. Second, because we aimed to genotype individuals of the F_2 population (homozygous or heterozygous at a single genomic locus), SNPs derived from 100 % homozygous alleles in each parental line were considered as allelic differences for genotyping of the mapping population. Third, to remove sequencing errors, only alleles with depth $\geq$8 reads were used for SNP genotyping. This read depth is more than two times higher than the minimum recommended read depth of $\geq$3 per genotype for genotyping by resequencing (Santosh et al. 2012). As a result, a total of 16,282 SNP markers along with 150 length polymorphic PCR markers were identified and 2,637 unique markers among them were mapped to 9 linkage groups. The resulting genetic map of *R. sativus* had more than three times as many markers than any of the genetic maps reported so far (Li et al. 2011; Shirasawa et al. 2011). Inclusion of the PCR-based markers ensured direct comparison with map positions of previously reported maps of *R. sativus* and *B. rapa*. Comparison of the PCR-based markers showed collinearity of linkage groups between our map and previously reported maps. Moreover, by means of inferred marker sequence scaffold relationships, 390 sequence scaffolds and 168 BAC clones were incorporated into the reference genetic map, anchoring ~295 Mb of the genome sequences to genetic loci, which covers approximately 86 % of the current genome assembly and is 2.5 times more than that of the Japanese 'Aokubi' assembly (116 Mb; Kitashiba et al. 2014). Because the current map has restricted resolution of markers due to limited number of F_2 population (93 plants), additional sequencing of F_2 plants and reanalysis of the SNP data such as bin mapping (Celton et al. 2010) will be helpful for further improvement of genetic map and additional anchor of the genome assembly.

The sequence scaffolds anchored onto chromosomes provided an overall view of the euchromatin structure in *R. sativus*. A genome-scale comparison of *R. sativus* genomic sequences with the *A. thaliana* genome showed chromosome level synteny between the two genomes and presented clear evidence of a triplicated genome structure in the *Raphanus* lineage. There have been several reports that the *R. sativus* genome has triplicated sub-genome blocks based on the comparative mapping (Jeong et al. 2014; Li et al. 2011; Shirasawa et al. 2011) or a close phylogenetic relationship of *Raphanus* with *Brassica* in tribe Brassiceae (Lysak et al. 2005). Our results provide direct chromosome to chromosome relationships at the sequence level. In addition, sequence analysis of gene-rich BAC clones showed that the gene structure and density of *R. sativus* are expected to be highly similar to those of *B. rapa*. The total number of genes in the *R. sativus* genome was estimated to be approximately 47,000, which is similar to *B. rapa* (~41,000) and *B. oleracea* (~45,000, personal communication with Dr. Sheingyi Liu) and only 1.7-fold more compared with that of *A. thaliana* (~27,000). It is noteworthy that the 61,572 gene models predicted from the 'Aokubi' assembly appears to be an over-estimation because the gene models were predicted by Augustus with a training set using the *A. thaliana* gene model and partial genes as well as very small genes (3–300 bp) and transposons were also counted (Kitashiba et al. 2014). Despite whole-genome triplication, a reduced number of genes in the *R. sativus* genome indicate a genome-wide loss of redundant components. Even though the *R. sativus* sequencing project is still underway and the sequence scaffolds used in this study are incomplete, the scale of synteny between *R. sativus* and *A. thaliana* is enough to estimate the degree of fractionation in each syntenic sub-genome block. We identified differential coverage of syntenic region per sub-genome block in the *R. sativus* genome. Taken together, it appears that a similar mode of whole-genome triplication proposed as a two-step merging of three ancestral genomes

in tribe Brassiceae (Cheng et al. 2013) might be involved in the organization of the *R. sativus* genome.

The map presented here successfully incorporates sequence scaffolds which can be assembled into super-scaffolds or chromosome pseudomolecules. In addition, it helps us to identify potential errors in order and orientation of sequence assembly. We evaluated the reliability of scaffolds based on SNP genotypes. Contigs showing conflict of genotyping data compared to neighbor contigs in the same scaffold were considered as possible assembly errors. Among 33,319 contigs investigated, 2,685 (8.1 %) contigs fell into this category. By using multiple evidences including SNP genotypes and sequence matches of mate-paired reads, 206 scaffolds were split into separate contigs. In addition, the genetic map identified several inversion errors in orientation of scaffolds. For example, the region at 34.5 Mb to 38.4 Mb in R2 was inversed based on the genetic position. The possible origins of such errors include highly conserved triplicated genome segments, misassignment of highly similar repetitive sequence blocks, and experimental error. Increasing the marker density of the genetic map will further enable analyses of genomic sequences associated with high recombination rates. Furthermore, the sequence data generated in this study increase the genetic and genomic resources available for *R. sativus* by adding a new category of genetic markers and a pipeline of resequencing data, thus, greatly enriching genomic tools that can be used to study the *R. sativus* genome and facilitate the genetic improvement of this crop.

Authors contributions JHM and HJY conceived the projects, designed research, and wrote the manuscript. HC performed the experiments and analyzed data. YMJ and SP contributed to genetic mapping. WHC and NK performed bioinformatics analysis. KBL and YJH contributed to FISH. MO, BOA, and BSP participated in NGS sequencing.

Acknowledgments This work was supported by grants from the National Academy of Agricultural Science (PJ009795), RDA to JHM and the Next-Generation Biogreen21 program (PJ008019), RDA to HJY.

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

Ethical standard The authors declare that the experiments complied with current laws of the country in which they were performed.

References

Al-Shehbaz I, Beilstein M, Kellogg E (2006) Systematics and phylogeny of the Brassicaceae (Cruciferae): an overview. Plant Syst Evol 259:89–120

Bancroft I, Morgan C, Fraser F, Higgins J, Wells R, Clissold L, Baker D, Long Y, Meng J, Wang X, Liu S, Trick M (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. Nat Biotechnol 29:762–766

Bett K, Lydiate D (2003) Genetic analysis and genome mapping in *Raphanus*. Genome 46:423–430

Budahn H, Peterka H, Mousa M, Ding Y, Zhang S, Li J (2009) Molecular mapping in oil radish (*Raphanus sativus* L.) and QTL analysis of resistance against beet cyst nematode (*Heterodera schachtii*). Theor Appl Genet 118:775–782

Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid K, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956–963

Celton J-M, Christoffels A, Sargent D, Xu X, Rees D (2010) Genome-wide SNP identification by high-throughput sequencing and selective mapping allows sequence assembly positioning using a framework genetic linkage map. BMC Biol 8:155

Cheng F, Mandáková T, Wu J, Xie Q, Lysak M, Wang X (2013) Deciphering the diploid ancestral genome of the Mesohexaploid *Brassica rapa*. Plant Cell 25:1541–1554

Chung H, Jeong Y-M, Mun J-H, Lee S-S, Chung W-H, Yu H-J (2014) Construction of a genetic map based on high throughput SNP genotyping and genetic mapping of a TuMV resistance locus in *Brassica rapa*. Mol Genet Genomics 289:149–160

Deschamps S, Llaca V, May G (2012) Genotyping-by-sequencing in plants. Biology 1:460–483

Edenberg H, Liu Y (2009) Laboratory methods for high-throughput genotyping. Cold Spring Harb Protoc 2009:pdb-top62

Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K, Buckler E, Mitchell S (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Hall A, Fiebig A, Preuss D (2002) Beyond the *Arabidopsis* genome: opportunities for comparative genomics. Plant Physiol 129:1439–1447

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. Genome Res 19:1068–1076

Huang S, Deng L, Guan M, Li J, Lu K, Wang H, Fu D, Mason A, Liu S, Hua W (2013) Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop *Brassica napus*. BMC Genomics 14:717

Hwang Y-J, Yu H-J, Mun J-H, Ryu K, Park B-S, Lim K-B (2012) Centromere repeat DNA originated from *Brassica rapa* is detected in the centromere region of *Raphanus sativus* chromosomes. Korean J Hortic Sci Technol 30:751–756

Jeong Y-M, Chung W-H, Chung H, Kim N, Park B-S, Lim K-B, Yu H-J, Mun J-H (2014) Comparative analysis of the radish genome based on a conserved ortholog set (COS) of *Brassica*. Theor Appl Genet 127:1975–1989

Johnston J, Pepper A, Hall A, Chen Z, Hodnett G, Drabek J, Lopez R, Price H (2005) Evolution of genome size in Brassicaceae. Ann Bot 95:229–253

Jones E, Sullivan H, Bhattramakki D, Smith J (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371

Kamei A, Tsuro M, Kobo N, Hayashi T, Wang N, Fujimura T, Hirai M (2010) QTL mapping of clubroot resistance in radish (*Raphanus sativus* L.). Theor Appl Genet 120:1021–1027

Kiss G, Kereszt A, Kiss P, Endre G (1998) Colormapping: a non-mathematical procedure for genetic mapping. Acta Biol Hung 49:125–142

Kitashiba H, Li F, Hirakawa H, Kawanabe T, Zou Z, Hasegawa Y, Tonosaki K, Shirasawa S, Fukushima A, Yokoi S, Takahata Y, Kakizaki

T, Ishida M, Okamoto S, Sakamoto K, Shirasawa K, Tabata S, Nishio T (2014) Draft sequences of the radish (*Raphanus sativus* L.) genome. DNA Res. doi:10.1093/dnares/dsu014

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Lai K, Duran C, Berkman P, Lorenc M, Stiller J, Manoli S, Hayden M, Forrest K, Fleury D, Baumann U, Zander M, Mason A, Batley J, Edwards D (2012) Single nucleotide polymorphism discovery from wheat next-generation sequence data. Plant Biotechnol J 10:743–749

Lee S-S, Lee S-A, Yang J, Kim J (2011) Developing stable progenies of x*Brassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. Theor Appl Genet 122:885–891

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler Transform. Bioinformatics 25:1754–1760

Li F, Hasegawa Y, Saito M, Shirasawa S, Fukushima A, Ito T, Fujii H, Kishitani S, Kitashiba H, Nishio T (2011) Extensive chromosome homoeology among Brassiceae species were revealed by comparative genetic mapping with high-density EST-based SNP markers in radish (*Raphanus sativus* L.). DNA Res 18:401–411

Lim K, de Jong H, Yang T, Park J, Kwon S, Kim J, Lim M, Kim J, Jin M, Jin Y, Kim S, Lim Y, Bang J, Kim H, Park B (2005) Characterization of rDNA and tandem repeats in the heterochromatin of *Brassica rapa*. Mol Cell 19:436–444

Lysak M, Koch M, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res 15:516–525

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303

Mun J-H, Kim D-J, Choi H-K, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudez A, Kiss G, Roe B, Young N, Cook D (2006) Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. Genetics 172:2541–2555

Mun J-H, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Lee SI, Baek S, Choi BS, Kim DS, Kim N, Yu HJ, Lim KB, Lim YP, Bancroft I, Hahn JH, Park BS (2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol 10:R111

Nussbaumer T, Martis M, Roessner S, Pfeifer M, Bader K, Sharma S, Gundlach H, Spannagl M (2013) MIPS PlantsDB: a database framework for comparative plant genome research. Nucleic Acids Res 41:D1144–D1151

Ossowski S, Schneeberger K, Clark R, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. Genome Res 18:2024–2033

Park S, Yu HJ, Mun JH, Lee SC (2010) Genome-wide discovery of DNA polymorphism in *Brassica rapa*. Mol Genet Genomics 283:135–145

Prakash S, Bhat S, Quiros C, Kirti P, Chopra V (2009) *Brassica* and its close allies: cytogenetics and evolution. In: Jules J (ed) Plant Breed Reviews, vol 31. Wiley, London, pp 21–187

Rozen S, Skaletsky H (1999) Primer3 on the WWW for General Users and for Biologist Programmers. In: Misener S, Krawetz S (eds) Bioinformatics Methods and Protocols. Humana Press, Totowa, pp 365–386

Santosh K, Travis W, Sylvie C (2012) SNP Discovery through next-generation sequencing and its applications. Int J Plant Genomics 2012:ID 831460

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

Shirasawa K, Oyama M, Hirakawa H, Sato S, Tabata S, Fujioka T, Kimizuka-Takagi C, Sasamoto S, Watanabe A, Kato M, Kishida Y, Kohara M, Takahashi C, Tsuruoka H, Wada T, Sakai T, Isobe S (2011) An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the Brassicaceae. DNA Res 18:221–232

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau É, Laroche J, Larose S, Jean M, Belzile F (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8:e54603

Song K, Osborn T, Williams P (1990) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs): 3. Genomic relationships in *Brassica* and related genera and the origin of *B. oleracea* and *B. rapa* (syn. *campestris*). Theor Appl Genet 79:497–506

Subbaiyan G, Waters D, Katiyar S, Sadananda A, Vaddadi S, Henry R (2012) Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. Plant Biotechnol J 10:623–634

Trick M, Adamski N, Mugford S, Jiang C-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

Tsuro M, Suwabe K, Kobo N, Matsumoto S, Hirai M (2005) Construction of a molecular linkage map of radish (*Raphanus sativus* L.), based on AFLP and *Brassica*-SSR markers. Breed Sci 55:107–111

Uitdewilligen J, Wolters A, D'hoop B, Borm T, Visser R, van Eck H (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS ONE 8:e62355

van Ooijen JW (2006) JoinMap® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B. V, Wageningen

Wang S, Wang X, He Q, Liu X, Xu W, Li L, Gao J, Wang F (2012) Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. Plant Cell Rep 31:1437–1447

Xu X, Liu X, Ge S, Jensen J, Hu F, Li X, Dong Y, Gutenkunst R, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2011) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Yang Y, Tai P, Chen Y, Li W (2002) A study of the phylogeny of *Brassica rapa*, *B. nigra*, *Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. Mol Phylogenet Evol 23:268–275

Yu H-J, Park S-G, Oh M, Hwang H-J, Kim N, Chung H, Sohn S-H, Park B-S, Mun J-H (2011) The *Brassica rapa* tissue-specific EST database. Korean J Hortic Sci Technol 29:633–640