Theor Appl Genet (2014) 127:1975–1989
DOI 10.1007/s00122-014-2354-3

ORIGINAL PAPER

Comparative analysis of the radish genome based on a conserved ortholog set (COS) of *Brassica*

Young-Min Jeong · Won-Hyong Chung · Hee Chung · Namshin Kim · Beom-Seok Park · Ki-Byung Lim · Hee-Ju Yu · Jeong-Hwan Mun

Received: 24 February 2014 / Accepted: 4 July 2014 / Published online: 24 July 2014

Abstract
Key message **This manuscript provides a *Brassica* conserved ortholog set (COS) that can be used as diagnostic cross-species markers as well as tools for genetic mapping and genome comparison of the Brassicaceae.**
Abstract A conserved ortholog set (COS) is a collection of genes that are conserved in both sequence and copy number between closely related genomes. COS is a useful resource for developing gene-based markers and is suitable for comparative genome mapping. We developed a COS for *Brassica* based on proteome comparisons of *Arabidopsis thaliana*, *B. rapa*, and *B. oleracea* to establish a basis for comparative genome analysis of crop species in the Brassicaceae. A total of 1,194 conserved orthologous single-copy genes were identified from the genomes based on whole-genome BLASTP analysis. Gene ontology analysis showed that most of them encoded proteins with unknown function and chloroplast-related genes were enriched. In addition, 152 *Brassica* COS primer sets were applied to 16 crop and wild species of the Brassicaceae and 57.9–92.8 % of them were successfully amplified across the species representing that a *Brassica* COS can be used as diagnostic cross-species markers of diverse *Brassica* species. We constructed a genetic map of *Raphanus sativus* by analyzing the segregation of 322 COS genes in an F_2 population (93 individuals) of Korean cultivars (*WK10039* × *WK10024*). Comparative genome analysis based on the COS genes showed conserved genome structures between *R. sativus* and *B. rapa* with lineage-specific rearrangement and fractionation of triplicated subgenome blocks indicating close evolutionary relationship and differentiation of the genomes. The *Brassica* COS developed in this study will play an important role in genetic, genomic, and breeding studies of crop Brassicaceae species.

Communicated by Isobel Parkin.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2354-3) contains supplementary material, which is available to authorized users.

Y.-M. Jeong · H. Chung · 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, Daejon 305-806, Korea

B.-S. Park
The Agricultural Genome Center, National Academy of Agricultural Science Rural Development Administration, Suwon 441-707, Korea

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

J.-H. Mun (✉)
Department of Bioscience and Bioinformatics, Myongji University, Yongin 449-728, Korea
e-mail: munjh@mju.ac.kr

Introduction

The Brassicaceae is a large plant family including approximately 3,700 species in 350 genera. The species of this family have diverse morphological characteristics, many of which are of agronomic importance as vegetables, oilseeds, condiments, fodders, and ornaments (Beilstein et al. 2006). Economically, the crop species of Brassicaceae contribute to 10 % of the world's vegetable crop produce and 12 % of the worldwide edible oil supplies (Economic Research Service

USDA 2008). The tribe Brassiceae, which is one of 13–19 tribes in the Brassicaceae, consists of approximately 240 species in monophyletic group and contains most crop species of *Brassica* and its relatives (Al-Shehbaz et al. 2006). Species of particular importance are *B. napus* (*Bna*) and *B. juncea* (*Bj*) as sources of canola oil, *B. rapa* (*Br*) and *B. oleracea* (*Bo*), as vegetable cole crops, *Raphanus sativus* (*Rs*) as root vegetable, and *B. nigra* (*Bn*) and *Sinapis alba* (*Sa*) as sources of mustard condiment. In addition to the crop species, many of the wild species in the tribe Brassiceae have potential as new crops, sources of condiments, industrial oil, and other diverse products and/or host systems for molecular farming. Moreover, wild relatives possess a number of useful agronomic traits, including nuclear and cytoplasmic male sterility, resistance to disease, insect, and nematode pests, tolerance of cold, salt, and drought stresses. For this reason, an understanding of the genetic potential and genomic information of Brassiceae is highly helpful for the establishment of long-term breeding programs of these crops.

Knowledge transfer from the sequenced genomes to the species of interest is critical for comparative and evolutionary genomic studies. As of 2013, six sequenced genomes, *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative 2000), *A. lyrata* (Hu et al. 2011), *Thellungiella parvula* (Dassanayake et al. 2011), *B. rapa* (The *Brassica rapa* Genome Sequencing Project Consortium 2011), *T. salsuginea* (Wu et al. 2012; Yang et al. 2013), and *Capsella rubella* (Slotte et al. 2013) were published in the Brassicaceae and genome sequencing of several species including *Bo*, *Bn*, *Bna*, and *Camelina sativa* is ongoing. Sequence level comparison between the sequenced genomes enabled the reconstruction of evolutionary relationships and the identification of lineage-specific events. For example, genome comparison of *At* and *Br* suggested that the triplicated *Br* genome might be arisen by hybridization between the less fractionated diploid genome (LF) and the more fractionated (MF1 and MF2) tetraploid genome (Cheng et al. 2013). Genome size change derived by deletions and/or insertions and recent transition to self-fertilization from outcross by genome-wide relaxation of purifying selection were found from *A. lyrata* (Hu et al. 2011) and *C. rubella* (Slotte et al. 2013), respectively.

Meanwhile, the extensive sequence and genetic information of the sequenced genomes can be easily applied to the closely related crop species based on comparative genome mapping. In the early stage, restriction fragment length polymorphism (RFLP) markers facilitated genetic linkage analyses and comparative mapping studies. Because RFLP markers use DNA–DNA hybridization, it became possible to examine homologous or similar genes between species, which enabled comparative mapping. With the advances in bioinformatics and molecular genetics, polymerase chain reaction (PCR) markers replaced the RFLP markers. Among them, gene-based markers targeting conserved genes had significant utility for comparative genome analysis as these were anticipated to have wider application between the closely related species. Furthermore, markers developed from single-copy genes are useful for studying polyploid genomes such as *Bj* and *Bna* because the orthologous (allelic) and paralogous (non-allelic) variants for single-copy genes do not need to be distinguished (Huang et al. 2013; Panjabi et al. 2008). In this regard, conserved ortholog set (COS) is an ideal candidate of gene-based marker development. COS is a collection of single- or low-copy genes that are evolutionarily conserved or orthologous between the species. Initially, COS was developed from the Solanaceae family and the COS markers were successfully applied in the comparative mapping of *Solanum* species by Fulton et al. (2002). Furthermore, Wu et al. (2006) reported COSII markers which were developed from a set of single-copy conserved orthologous genes in Euasterid species as an expanded version of original COS markers. So far, COS has been identified from diverse plant species such as cereals (Quraishi et al. 2009), Rosaceae (Bushakra et al. 2011; Cabrera et al. 2009), Asteraceae (Chapman et al. 2007; Timms et al. 2006), and Pinaceae (Krutovsky et al. 2004; Liewlaksaneeyanawin et al. 2008). Considering the fact that COS genes are highly conserved and low in copy number, PCR-based COS markers provide an important opportunity of rapid genetic map construction and genome comparison as shown in species of Solanaceae (Wu et al. 2009a, b, 2010).

Radish (*Rs*) is an edible root vegetable crop cultivated worldwide and is one of the key items of seed industry in East Asia. It is a member of tribe Brassiceae and a close relative of *Brassica* crops such as Kimchi cabbage (*Br*) and cabbage (*Bo*) that share mesopolyploidy before domestication. Despite of its importance in agriculture and evolutionary biology, little is known about the radish genome. Therefore systematic study of the radish genome can provide us detailed information on the genome as well as useful resources that can be used for breeding. Recently, a few studies of genetic map construction and comparative mapping showing chromosome relationship between *Rs* and *Br* based on EST-SSR (Shirasawa et al. 2011) or EST-SNP (Li et al. 2011) markers have been reported. Furthermore, a genome sequencing project for a Korean cultivar of radish, *WK10039* ($2n = 18$, 510 Mb) is currently underway (The Agricultural Genome Center, Rural Development Administration, Korea) using a next-generation sequencing strategy. As a part of genomic studies of the radish genome, here we report the development of COS markers on the basis of the sequenced genomes in the Brassicaceae and their application as diagnostic markers for diverse *Brassica* species. In addition, we applied these markers as genomic tools for genetic mapping of radish and identification of conserved genome blocks between the *Rs* and *Br* genomes.

Materials and methods

Sequence data and identification of COS

Sequence data were collected from TAIR (www.arabidopsis.org) for *At* and BRAD database (www.brassicadb.org) for *Br* in March, 2012. The predicted gene models of *Bo* were provided by Shengyi Liu at Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, China. Conserved orthologous genes between *At*, *Br*, and *Bo* were identified by a proteome comparison based on BLASTP analysis (Altschul et al. 1997). First we identified single-copy genes of each genome based on self-BLASTP (minimum of 50 % alignment coverage at a cutoff of $<1E^{-4}$) then compared the single-copy genes between the genomes. The top reciprocal BLASTP matches (minimum of 50 % alignment coverage at a cutoff of $<1E^{-10}$) of *At–Br*, *Br–Bo*, and *Bo–At* comparisons were selected and mutually connected with each other. Only the set of genes showing self-closed relationships among the genomes (i.e. a gene tracing the relation within *At–Br*, *Br–Bo*, and *Bo–At* comparisons comes back to itself) were selected as *Brassica* COS.

Alignment of COS and primer design

Sequences of selected COS were aligned using the CLUSTALW program (Larkin et al. 2007) and the intron-flanking regions showing more than 10 % size differences of introns between the genomes were selected for further analysis. Primers targeting the selected regions were designed based on the exon sequences of *Br* using Primer3 software (Rozen and Skaletsky 2000). The Primer3 software was configured to design five sets of oligonucleotide primers flanking each intron with a target amplicon size range of 250–1,000 bp. Primer specifications were melting temperature (T_m) 57–63 °C (target 60 °C) with GC content 40–60 % and a primer length of 18–27 nucleotides (target 24 nucleotides).

Plant materials and DNA extraction

Seeds of 11 wild *Brassica* species (*B. balearica*, *B. barrelieri*, *B. bivoniana*, *B. drepanensis*, *B. desnottesii*, *B. fruticulosa*, *B. macrocarpa*, *B. repanda*, *B. rupestris*, *B. spinescens*, *B. villosa*) were obtained from Leibniz Institute of Plant Genetics and Crop Plant Research, Germany. Seeds of *At*, *Brassica* species and *Rs* were surface sterilized in 12 % sodium hypochlorite and were 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 of 5 days after germination were transferred into soil and were grown in a culture room with the same growth condition. Genomic DNA from the leaves of each plant was isolated using an extraction buffer [2 % hexadecyltrimethylammonium bromide (CTAB), 100 mM TrisHCl pH 8.0, 20 mM EDTA, 1.4 M NaCl, 0.2 % β-mercaptoethanol] followed by ethanol precipitation and RNase treatment.

PCR amplification and sequencing

PCR amplification was performed using 2× TOPsimple Premix-Tenuto (Enzynomics, Daejeon, Korea) with a following condition of 3 min at 95 °C, 35 cycles of 30 s at 95 °C, 30 s at 55 °C, 3 min at 72 °C, and a final 5 min extension step at 72 °C. Primers used in this study are listed in Supplemental Table S1. PCR products were resolved on a 3 % agarose gel and bands were visualized by staining with ethidium bromide. Sequencing of PCR products was performed using an ABI3730xl automated sequencer and the ABI BigDye Terminator V3.1 Cycle Sequencing kit (Life Technologies, USA).

COS gene finding and construction of genetic map in *R. sativus*

To identify the COS genes in *Rs*, coding sequences of *At*, *Br*, and *Bo* COS genes were BLASTZ searched with cutoff of $1E^{-10}$ against the whole-genome contig sequences of *Rs* cv. *WK10039* covering 469 Mb of the genome (unpublished data). The *Rs* sequences showing the highest sequence similarity to the BLAST query were selected and the gene models were ab initio predicted using the FGENESH program with the parameter trained using the *Rs* matrix. The region of contig that contains the COS gene was also identified.

For genetic mapping, SNPs within the COS genes were searched between mapping parents, cv. *WK10039* and cv. *WK10024*, and SNP genotyping of F_2 mapping population composed of 93 individuals were performed by resequencing of each F_2 progeny for 10X genome coverage using Illumina HiSeq1000 (unpublished data). In brief, a total of 5,851,762,272 F_2 reads corresponding to 591 Gb were mapped to the reference contig sequences of parent lines using Bowtie (Langmead et al. 2009) and genotypes with more than 8 reads mapped on the SNP sites of the COS genes were recorded as follows: homozygous maternal (*WK10039*) as "A", homozygous paternal (*WK10024*) as "B", heterozygous as "H", and missing or ambiguous data as "–". For purposes of accurate genotype analysis, SNP genotypes from single COS gene were analyzed together. When the identity of genotypes for SNPs within the same COS gene was higher than 80 %, the genotype of major SNP was determined as representative for the COS gene. Otherwise, the genotype of COS was treated as unknown.

A genetic map was constructed by JoinMap 4 (van Ooijen 2006). A LOD threshold of 5.0 was used for the assignment of linkage groups and the genetic distance between the markers was calculated by the Kosambi map function.

Synteny block comparison between the genomes

Physical positions of the COS genes in the *At* and *Br* genomes were determined and compared with those in the genetic map of *Rs*. Syntenic region between the *Br* and *At* genomes were identified by a proteome comparison based on BLASTP analysis, as described in our previous report (Kim et al. 2012; Mun et al. 2009). The assignment of 24 genome building blocks of the ancestral karyotype (AK) of both genomes was carried out according to previous comparative genome mapping studies (Cheng et al. 2013; Schranz et al. 2006) and subgenome information of *Br* (Cheng et al. 2013). Meanwhile, *Bo* was excluded due to insufficient genome information. Syntenic regions between *Rs* and *At* or *Br* were determined based on collinearity of COS genes. The isolated non-collinear COS genes in the AK blocks were ignored. Synteny block pairs between the genomes were illustrated by Circos (Krzywinski et al. 2009).

K_s calculation of COS

To calculate the number of synonymous substitutions per synonymous site (K_s) between homologous genes of *At*, *Br*, and *Rs*, the coding sequences of the COS genes were aligned pairwise. In the case of *Rs* genes, coding sequences were deduced from the gene models predicted as described above. In addition, approximately 1.2 Gb expressed sequence tag (EST) reads generated from seedling (5 days after germination), root and leaf (1 month old), and flower tissues by 454 GS FLX were also analyzed (unpublished data). For the comparison of orthologous or paralogous gene pairs, annotated gene models for *At* and *Br* and predicted gene models or ESTs for *Rs* were used. After removing gaps, the K_s values from pairwise alignments of homologous sequences were determined using the maximum likelihood method implemented in the CODEML program (Goldman and Yang 1994) of the PAML package (Yang 2007).

Results

Identification and characterization of COS in the *A. thaliana*, *B. rapa*, and *B. oleracea* genomes

Considering the history of frequent polyploidy events in Brassicaceae, we created a list of single-copy genes in the *At*, *Br*, and *Bo* genomes before genome to genome comparison. Self-BLASTP analysis identified that 12.2–13.7 % of the genes were single copy in each genome. BLASTP comparison of the single-copy gene sets between the genomes identified a total of 1,194 conserved orthologous single-copy genes representing 2.6–3.4 % of total genes in each genome (Table 1). These genes were named as *Brassica* COS. The functional characteristics of COS were classified according to Gene Ontology (GO) analysis (Supplemental Fig. S1). We compared the results of GO-based classification of COS with those of whole *At* genes under statistical analysis of *Z* test. This revealed several categories for which the functional complement of COS is atypical ($P < 0.0001$). For example, it has higher proportions of genes pertaining to chloroplast and plastid of the GO cellular component category between the two data sets. In addition, there are differences in genes classified as related to 'DNA or RNA metabolism' under the GO biological process and unknown function and putative function of 'protein binding' of the GO molecular function category compared to the *At* genes.

Table 1 Conserved orthologous single-copy genes identified in the *A. thaliana*, *B. rapa*, and *B. oleracea* genomes

Species	Total gene	Single-copy gene (%)	COS gene (%)
A. thaliana	35,386	4,845 (13.7)	1,194 (3.4)
B. rapa	41,019	5,036 (12.2)	1,194 (2.9)
B. oleracea	45,758	5,581 (12.2)	1,194 (2.6)

To examine the distribution of COS genes in the *At*, *Br*, and *Bo* genomes, physical position was determined (Fig. 1). Overall, they were widely distributed across the chromosomes of each species with an average interval of 0.10 Mb (*At*), 0.22 Mb (*Br*), and 0.36 Mb (*Bo*) between the genes, respectively. There was no significant bias of COS distribution in the genome except the centromeric region. Centromeric regions of *At* chromosomes had few COS genes because gene density was low. As the heterochromatin sequences are unavailable, it is not clear whether the COS genes are unevenly distributed in the centromeric regions of *Br* and *Bo* genomes. Recently, Cheng et al. (2012) reported the comparative analysis of triplicated subgenome blocks (LF, MF1, and MF2) of *Br* with their syntenic counterpart of *At*. Using this information, the distribution patterns of COS in the triplicated subgenomes of *Br* were analyzed. A total of 994 COS genes (83.2 %) were identified in the subgenome blocks of *Br* that are syntenic to *At*. Most COS genes (56 %) were distributed in the LF blocks whereas MF1 and MF2 blocks contained 24 and 20 % of COS genes, respectively, representing more frequent distribution of COS genes in the LF blocks of the *Br* genome ($P < 0.001$; Table 2).

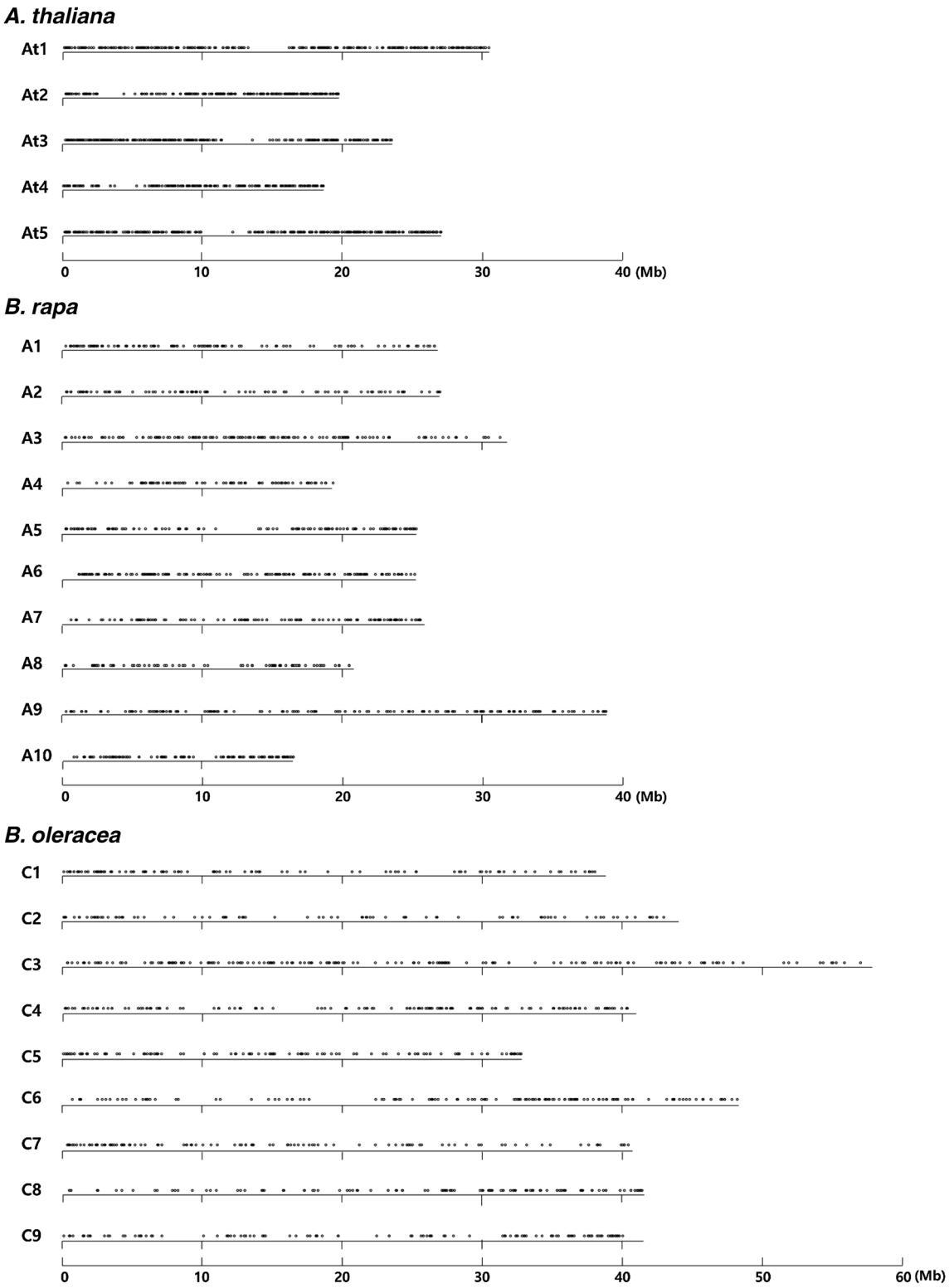

Fig. 1 Distribution of the COS genes in the *A. thaliana*, *B. rapa*, and *B. oleracea* genomes. *Circles* indicate position of COS genes on the chromosomes

Table 2 Distribution of the *Brassica* COS genes in the triplicated subgenomes of *B. rapa*

	Number			Total	Ratio (%)		
	LF	MF1	MF2		LF	MF1	MF2
B. rapa genes	16,270	11,633	9,653	37,556	43.3	31.0	25.7
COS genes	557	241	196	994	56.0	24.2	19.7

Number of genes in the triplicated subgenomes of *B. rapa* was counted based on the information of Cheng et al. (2012)

Table 3 Summary of PCR amplification of COS genes applied to crop and wild species

Species	Single-band amplification	Multiple-band amplification	Noisy or no amplification	Single-band amplification rate (%)	Overall success rate (%)
A. thaliana	117	23	12	77.0	92.1
B. rapa	125	16	11	82.2	92.8
B. nigra	100	36	16	65.8	89.5
B. oleracea	122	21	9	80.3	94.1
R. sativus	107	18	27	70.4	82.2
B. balearica	72	16	64	47.4	57.9
B. barrelieri	83	41	28	54.6	81.6
B. bivoniana	98	33	21	64.5	86.2
B. drepanensis	104	31	17	68.4	88.8
B. desnottesii	64	29	59	42.1	61.2
B. fruticulosa	75	35	42	49.3	72.4
B. macrocarpa	106	29	17	69.7	88.8
B. repanda	73	33	46	48	69.7
B. rupestris	108	28	16	71.1	89.5
B. spinescens	88	27	37	57.9	75.7
B. villosa	101	31	20	66.4	86.8
Average	96.4	27.9	27.6	63.4	81.8

PCR analysis of the COS in the Brassicaceae species

With the goal of developing cross-species diagnostic markers for *Brassica* species, we used the Primer3 software to design multiple sets of PCR primers flanking introns of the COS genes. Among 1,194 COS genes, 152 genes that have more than 10 % size differences between *At*, *Br*, and *Bo* were selected for PCR analysis. Primers were designed based on the *Br* sequences and amplification was performed using genomic DNA of *At*, *Br*, *Bo*, *Bn*, *Rs*, and wild *Brassica* species (Table 3). Approximately 82.2–94.1 % of the primer pairs yielded amplification products and the efficiency of single-band amplification was highest for *Br* (82.2 %) whereas less efficient for *Rs* (70.4 %) and *Bn* (65.8 %). To develop diagnostic markers for *Brassica* species, a total of 51 (34 %) of the primer pairs that amplified single amplicons in each genome of *At*, *Rs*, *Br* (A genome), *Bn* (B genome), and *Bo* (C genome) were selected. Among them, 7 primer pairs showed clear size difference between the genomes in agarose gel. Application of these 7 primer pairs successfully distinguished diploid genomes as well as allotetraploid species of *Brassica*, *Bj* (AB genome), *Bna* (AC genome), and *B. carinata* (*Bc*, BC genome) (Fig. 2a; Supplemental Fig. S2).

PCR amplification of the 152 primer pairs for the diverse wild *Brassica* species showed that 57.9–89.5 % of the genes were amplified and most of them were single-band amplifications (Table 3). However, there were variations in ratio of the single-banded amplicon per species. For example, *B. rupestris* showed the highest rate (71.1 %) of single-band amplicon whereas only 42.1 % of COS primer pairs amplified single-band amplicons in *B. desnottesii*. Of particular interest, 5 wild Mediterranean *Brassica* species (*B. bivoniana*, *B. drepanensis*, *B. macrocarpa*, *B. rupestris*, and *B. villosa*) showed higher rate of single-band amplification (64.5–71.1 %) than the other species (Supplemental Fig. S3). As shown in Fig. 2b, a total of 23 primer pairs were amplified as single amplicons across all the 11 wild *Brassica* species tested with size difference suggesting possible development of genome diagnostic markers for wide variety of wild *Brassica* species. To evaluate whether the COS primers correctly amplified the target genes,

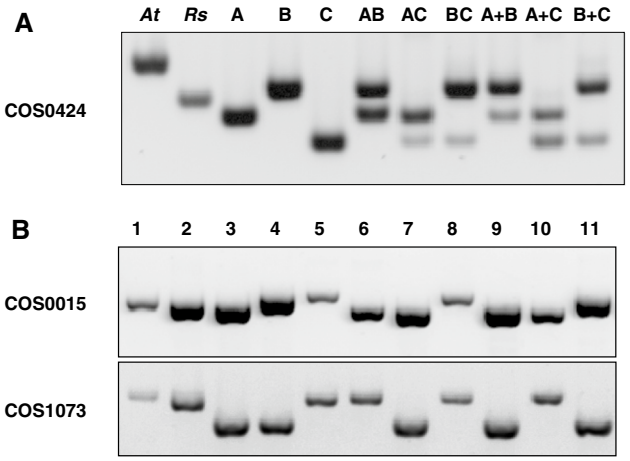

Fig. 2 PCR amplification of the COS genes in selected species of Brassicaceae. **a** An example of PCR amplification using COS0424 primer set. *At*, *A. thaliana*; *Rs*, *R. sativus*; *A*, *B. rapa* (*A* genome); *B*, *B. nigra* (*B* genome); *C*, *B. oleracea* (*C* genome); *AB*, *B. juncea* (*AB* genome); *AC*, *B. napus* (*AC* genome); *BC*, *B. carinata* (*BC* genome); *A* + *B*, genomic DNA mixture of *A* and *B*; *A* + *C*, genomic DNA mixture of *A* and *C*; *B* + *C*, genomic DNA mixture of *B* and *C*. **b** PCR amplification of COS genes in 11 wild *Brassica* species using COS0015 and COS1073 primer sets. Amplicons were examined in 3 % agarose gel. *1*, *B. balearica* (BRA2850); *2*, *B. barreilieri* (BRA1877); *3*, *B. bivoniana* (BRA2922); *4*, *B. drepanensis* (BRA2949); *5*, *B. desnottesii* (BRA2919); *6*, *B. fruticulosa* (BRA1810); *7*, *B. macrocarpa* (BRA2944); *8*, *B. repanda* (BRA1645); *9*, *B. rupestris* (BRA2992); *10*, *B. spinescens* (BRA2994); *11*, *B. villosa* (BRA1896)

sequences of the PCR products amplified by 6 primer pairs from 16 diploid species were analyzed. Among 96 amplicons, 93 sequences (96.9 %) were shown to be generated from the target genes and 32 sequences (33.3 %) had mismatches within intron. In this regard, the *Brassica* COS genes are conserved in the genomes of Brassicaceae with sequence or copy number variation.

Divergence of the COS genes between the genomes

To deduce the approximated divergence time point of the COS genes, we compared the distribution of synonymous substitution (K_s) of orthologous counterparts of COS genes between *At*, *Br*, and *Rs* genomes. As shown in Fig. 3a, *At* shares a single peaks with *Br* and *Rs* at K_s modes of 0.40 and 0.45, respectively, indicating a split of the tribe Brassiceae from the *Arabidopsis* lineage. Recent analyses based on the Bayesian approach suggest that diversification time between the two lineages occurred 43.2–47.5 million years ago (Mya) (Arias et al. 2014; Beilstein et al. 2010). The age distribution of *Br* and *Rs* yields a clear peak at $K_s = 0.20$ which is essentially identical to that obtained using the comparison of an orthologous gene set between *Br* gene models and *Rs* ESTs, which yields a peak at $K_s = 0.25$. A comparison of the K_s mode for the paralogs in *Rs* ESTs identified a peak for recent genome polyploidy event that is also evident in the *Br* genome at $K_s = 0.30$ as whole-genome triplication (Fig. 3b). Taken together, these findings suggest that the *Br* and *Rs* genomes share a recent polyploidy event, presumably a whole-genome triplication at 22.5 Mya (Beilstein et al. 2010), and separate soon after this polyploidy event.

Genetic mapping of the COS genes in *R. sativus*

For purposes of mapping the COS genes to the *Rs* genome, the genotypes of COS genes were scored in a reference mapping population consisted of 93 F_2 individuals from a cross between cultivars *WK10039* and *WK1002*4. In the initial analysis, we identified 11,198 SNPs from 632 COS genes in the whole-genome contig sequences of *Rs* and could score genotypes of 77.4 individuals of mapping population per SNP on average. In the case of COS genes mapped to multiple contigs were treated as paralogs. Genotyping data of all SNPs from a COS gene within a contig were compared and integrated into a major SNP only if the identity of genotypes is higher than 80 %. After removing ambiguous or insufficiently genotyped loci, SNP genotypes of 527 COS genes were selected for map construction. Linkage analysis identified COS genes showing identical genotypes or multiple mapping positions, and these genes were removed from map construction. A genetic map shown in Fig. 4 was derived from the analysis of 5,922 SNPs in 322 COS genes spanning 693.4 cM with an average distance between COS genes of 2.2 cM (Table 4). The biggest linkage group R4 contained 31 COS genes spanning 113.1 cM, while the smallest group R8 had 26 COS genes spanning 51.8 cM. Based on the estimated genome size (510 Mb) of *WK10039* based on K-mer analysis using 454 reads (unpublished data), 1 cM corresponds to 735.4 kb.

The reliability of linkage groups within the COS genetic map was tested with linkage comparisons, which were performed based on a putative ancestral karyotype (AK) comprising 24 genome building blocks named as A to X (Lysak et al. 2006; Schranz et al. 2006). Comparison of the COS genes of *Rs* with those of *At* and *Br* enabled comparative mapping of the inferred AK genome building blocks of the Brassicaceae. We defined conserved AK blocks on the *Rs* genome according to genome mapping of AK blocks on the *At* and *Br* genomes and pairwise information of *At–Rs* and *Br–Rs* collinearity blocks. A total of 45 AK blocks, excluding G and H blocks, were mapped to the *Rs* genome (Fig. 5a; Supplemental Table S2). The pattern of block boundaries on each linkage of the COS map was similar to those reported patterns for genetic maps based on EST markers (Li et al. 2011; Shirasawa et al. 2011). Overall, 42 of 45 blocks were present in the EST-based maps and 34 blocks were common in all three genetic maps with similar

block order (Supplemental Table S3). Comparison of the *Rs* genome with the *At* and *Br* genomes viewed through AK genome blocks revealed that three copies of the genomes, especially A, E, F, I, N, R, U, and W blocks, were evident (Supplemental Fig. S4); this suggests whole-genome triplication (WGT) of the *Rs* genome as predicted from a phylogeny of tribe Brassiceae. Other blocks occurred once or twice, presumably due to limited marker information or loss of blocks during genome evolution. Therefore, the COS genetic map defines nine linkage groups corresponding to those in previously defined maps.

Comparative analysis of subgenomic synteny between *Rs* and *Br*

Comparative mapping of COS genes between the *Rs* and *Br* genomes enabled characterization of the subgenomic identity of AK blocks on the triplicated *Rs* genome because the COS genes identified in this study are single-copy, syntenic orthologs. As shown in Fig. 5b and Supplemental Fig. S5, most *Rs* COS genes from the same block were mapped to a specific subgenome of the *Br* chromosome. For example, AK1 blocks (A, B, and C blocks) were located in R1, R5, and R8 of the *Rs* genome whereas they were dispersed in 6 *Br* chromosomes (A5, A6, A7, A8, A9, and A10).

Fig. 4 Genetic map of *R. sativus*. A total of 322 COS genes were positioned on nine linkage groups. Linkage groups were named as R1–R9 based on the comparative block models of Li et al. (2011). Maker positions (in cM) are shown on the *left margin* and corresponding COS genes are shown on the *right margin* of each linkage group. The position of AK genome blocks in the *R. sativus* chromosome was defined by a comparison of *Rs–At* and *Rs–Br* relationships of COS and the *At–AK* mapping results. *At*, *A. thaliana*; *Br*, *B. rapa*; *Rs*, *R. sativus*

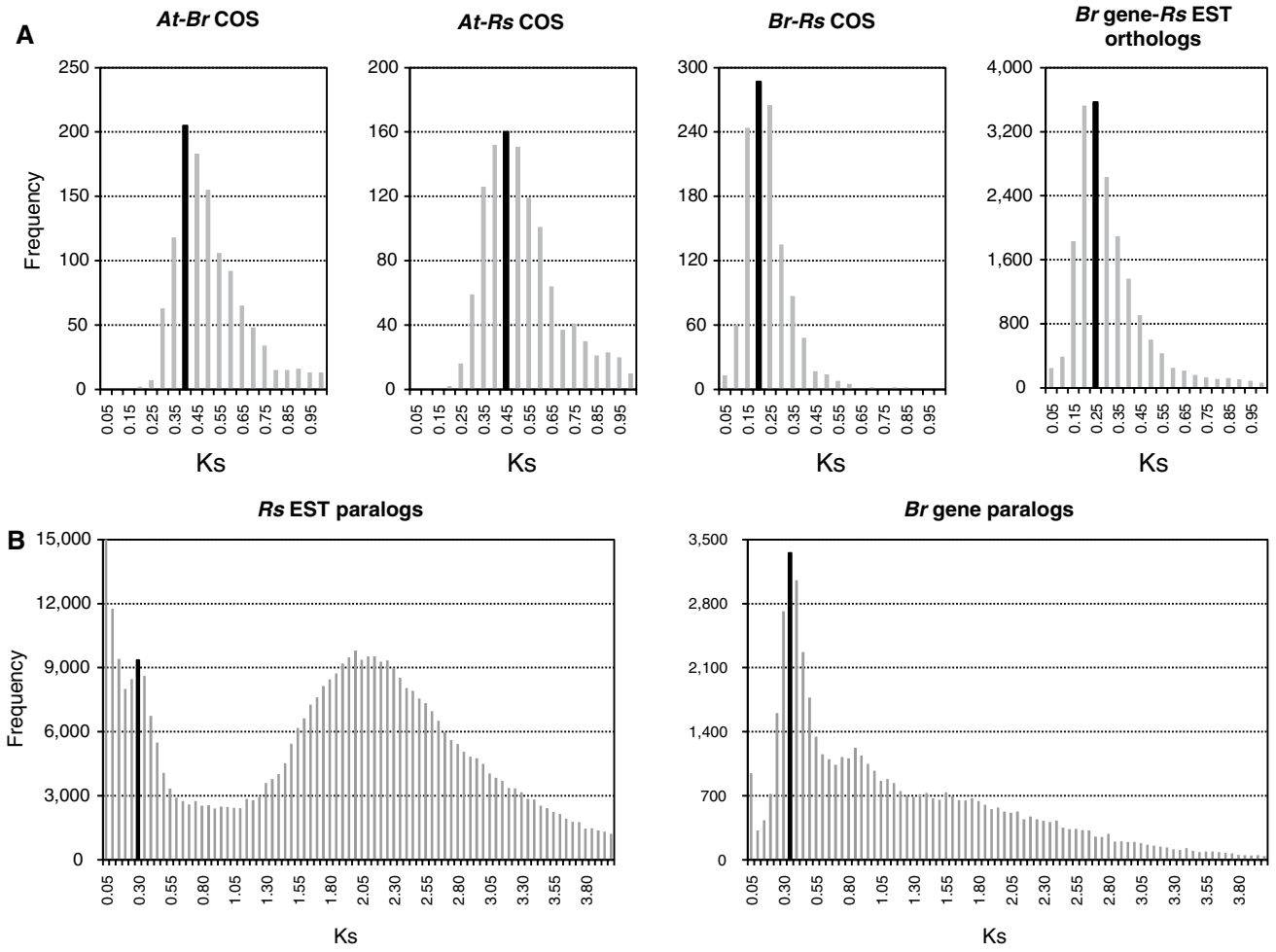

Fig. 3 Traces of polyploidy events in the *R. sativus* genome and divergence of COS genes. Distributions of K_s values were obtained from comparison of sets of orthologous COS genes (**a**) and paralogous genes (**b**) pairs. The *vertical axes* indicate the frequency of paired sequences and the *horizontal axes* denote K_s values at 0.05 intervals. The *black bars* indicate the position of the mode of K_s distributions. *At*, *A. thaliana*; *Br*, *B. rapa*; *Rs*, *R. sativus*

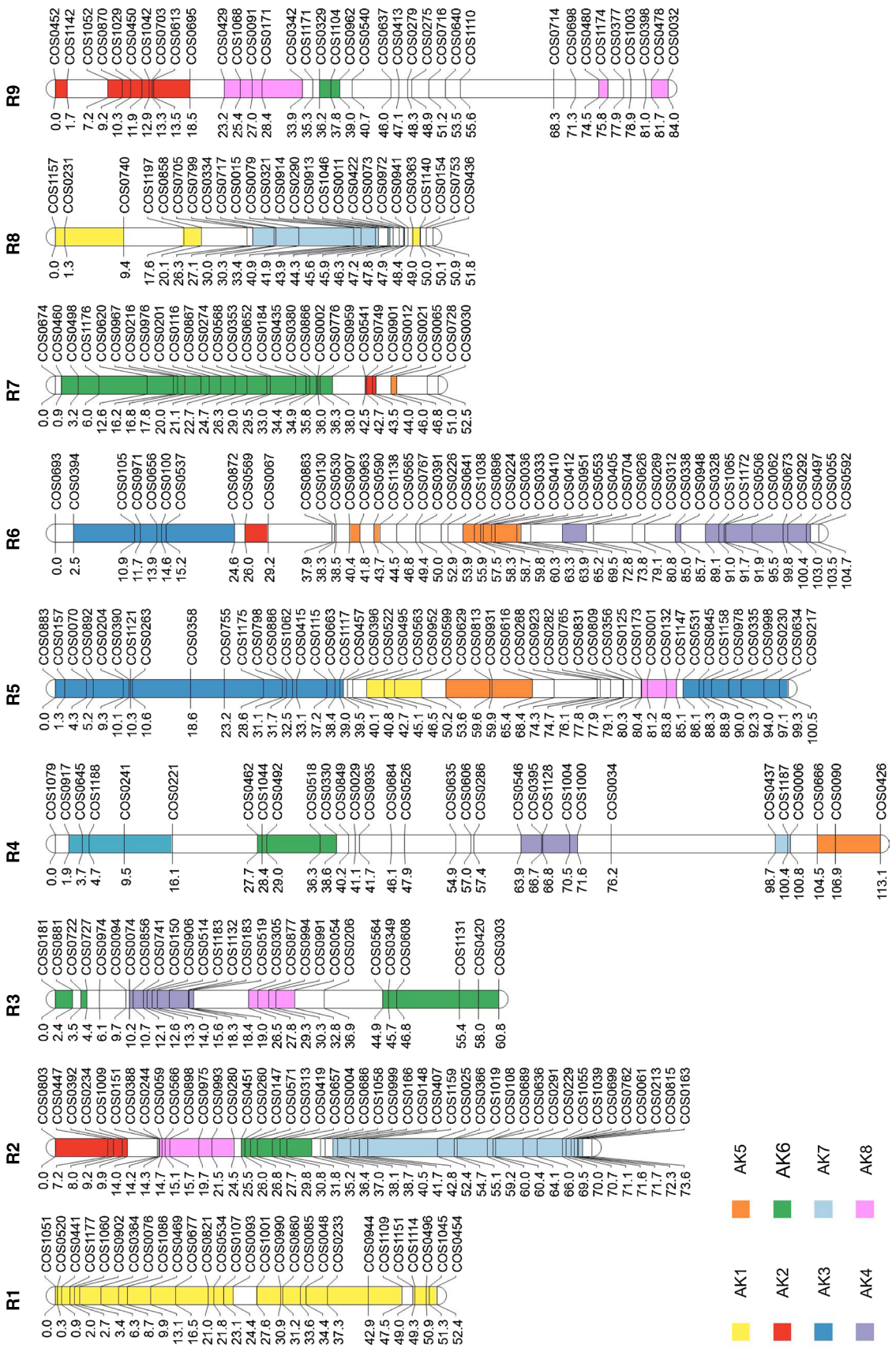
R1
R2
R3
R4
R5
R6
R7
R8
R9
AK1
AK2
AK3
AK4
AK5
AK6
AK7
AK8

Table 4 Summary of the genetic map constructed using the COS markers in *R. sativus*

Linkage group	Number of markers	Genetic length (cM)	Average distance per marker (cM)
R1	28	52.4	1.9
R2	45	73.6	1.6
R3	28	60.8	2.2
R4	31	113.1	3.6
R5	50	100.5	2.0
R6	48	104.7	2.2
R7	30	52.5	1.8
R8	26	51.8	2.0
R9	36	84.0	2.3
Total	322	693.4	2.2

Interestingly, except one gene, all *Rs* COS genes in R1 had counterparts in the LF subgenomes of *Br* A6, A9, and A10. Moreover, 7 of 8 *Rs* COS genes in R8 were mapped to the MF1 subgenomes in *Br* A8, whereas 5 of 6 *Rs* COS genes in R5 were orthologous to genes in the MF2 subgenomes of *Br* A5 and A9. The overall subgenomic structure of the *Rs* genome is presented in Fig. 6. These results collectively indicate that the *Brassica* COS is a useful resource for comparative genomic study of the Brassicaceae.

Discussion

Genome information of crop species provides fundamental resources for biology, phylogeny, and breeding studies. For genome analysis, investigation of structural and genetic features of the genome is crucial. In this regard, crop species in tribe Brassiceae can benefit directly from the sequenced genomes and gene-based markers which have utility for cross-species comparison are good targets of marker development. In this study, we developed conserved orthologous single-copy gene set of *Brassica* and positioned 322 COS genes on the genetic map of *Rs* for targeting gene-rich fractions of the genome.

The *Brassica* COS identified in this study was developed from 1,194 conserved single-copy genes in the genomes of *At*, *Br*, and *Bo* which correspond to small portion (2.6–3.4 %) of total genes in each genome. However, PCR analysis of the COS genes in 16 crop and wild species resulted in successful single-band amplification (42–82 %) suggesting a potential role as universal PCR primers for wide range of taxa in Brassicaceae. Wu et al. (2006) identified 2,869 COSII genes using EST sequences of 6 euasterid species (tomato, potato, pepper, coffee, sunflower, and lettuce) and one eurosid species *At*. Although not all the COSII were single copy in euasterid genomes, all COSII genes were single copy in *At*. Sequence comparison of the *Brassica* COS with the *At* COSII genes revealed that only 443 (37.1 %) genes are overlapped representing copy number variation of the *At* COSII genes in the *Brassica* genomes due to triplicated nature of the genomes. Enrichment of chloroplast and plastid-related genes of the *Brassica* COS based on GO in cellular component mapping (Supplemental Fig. S1) also suggests that evolutionary selection might involve in identity of single-copy genes in the *Brassica* genome. However, distribution of the COS genes on the *At*, *Br*, and *Bo* genomes was not uneven and they distributed throughout the genomes (Fig. 1). Taken together with successful cross-species amplification in the diverse crop and wild species, *Brassica* COS can serve as core gene-based markers for obtaining gene-rich fractions of *Brassica* and closely related species and investigating syntenic relationships between genomes.

We anticipated that application of the *Brassica* COS as well as the strategy of cross-species marker development will accelerate knowledge transfer from the sequenced genomes to closely related taxa and can make a significant contribution to structural and comparative genomic studies of *Brassica* and its relative crop species. For instance, we have constructed a core genetic map of *Rs* which has been serving a key resource for ongoing genome sequencing project using 322 COS genes (Fig. 4). We note that integration of sequence contigs or scaffolds into the genetic map is crucial for genome sequencing. Because the *Rs* genome has polyploidy nature, genetic markers originated from highly conserved multiplicated genome segments or recently evolved gene paralogs can result in inaccurate mapping of the assembled sequences. In this case, the resulting anchor of the sequence contigs or scaffolds would be flawed. In this regard, COS markers provide excellent tools to determine precise location and order of assembled sequences on the chromosomes because they represent non-redundant unique loci. Genetic markers continue to be added to the COS map, furthering the integration of genetic map and sequence assembly in this species and providing anchoring for the ongoing genome sequencing effort.

The utility of COS played an important role in understanding organization and evolution of the *Rs* genome.

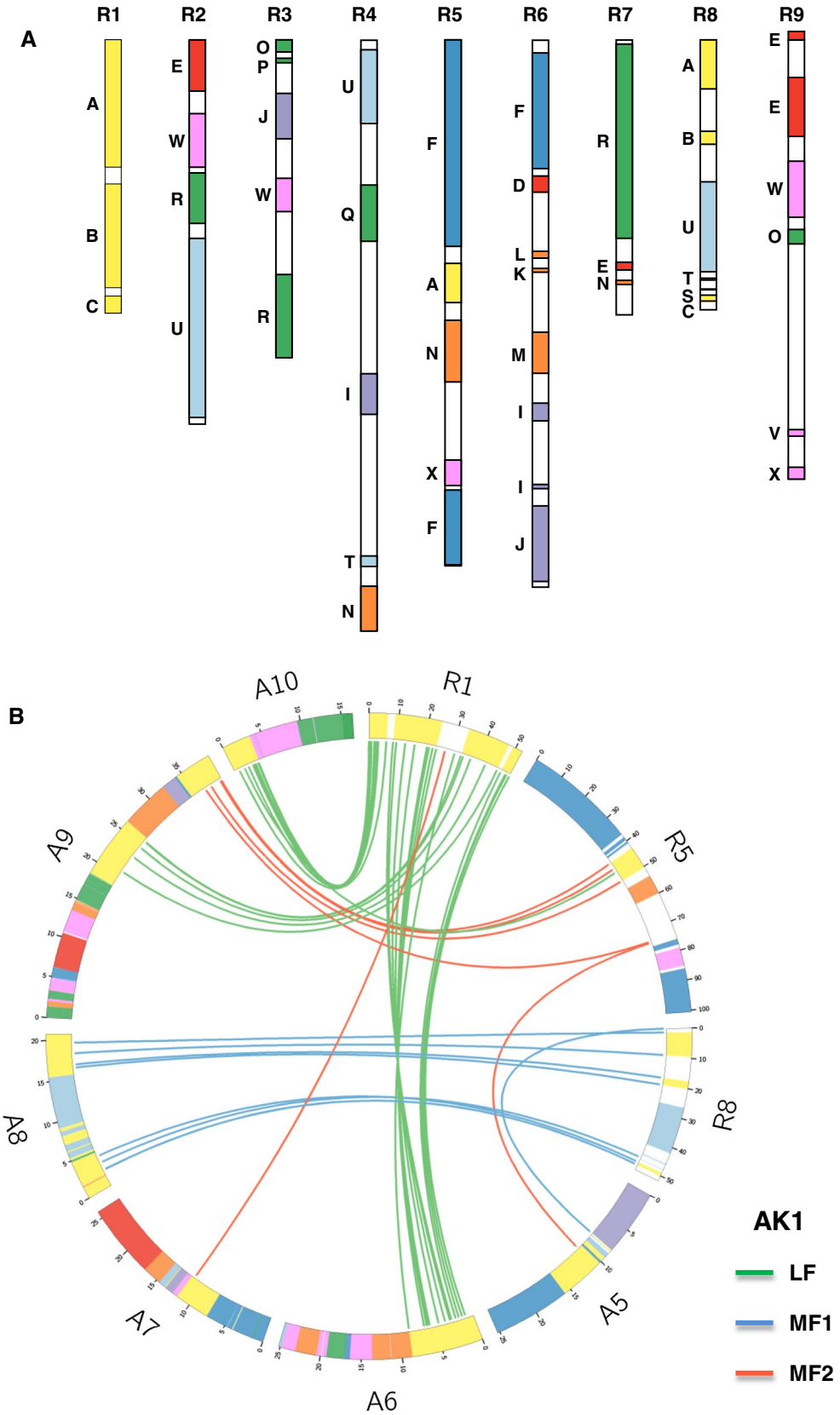

Fig. 5 Genomic structure of *R. sativus* based on 24 ancestral karyotype (AK) genome building blocks and syntenic relationships of AK1 between *R. sativus* and *B. rapa* genomes. **a** AK blocks were assigned based on the report of Cheng et al. (2013) and comparison of *Rs–Br* COS relationships. *Each color* indicates 8 ancient chromosomes and alphabets indicate AK blocks. Most of AK blocks in *Rs* genome are duplicated or triplicated and thoroughly rearranged compared to the ancient chromosomes. Block boundaries between syntenic blocks are shown in Supplemental Table S2. **b** Circos diagram of COS gene pairs in AK1 regions of *Rs* and *Br*. Only chromosomes or linkage groups with genes belonging to AK1 are displayed. Conserved orthologous genes in *Rs* are plotted against their syntenic counterparts in *Br*. The *numbers* in *Rs* linkage groups indicate genetic distance in cM whereas those in *Br* chromosomes indicate 5 Mb intervals. The syntenic counterparts of conserved gene pair between the genomes are interconnected by *colored lines*. The *line colors* indicate three differentially fractionated subgenome types in the *Br* genome. *Br*, *B. rapa*; *Rs*, *R. sativus*

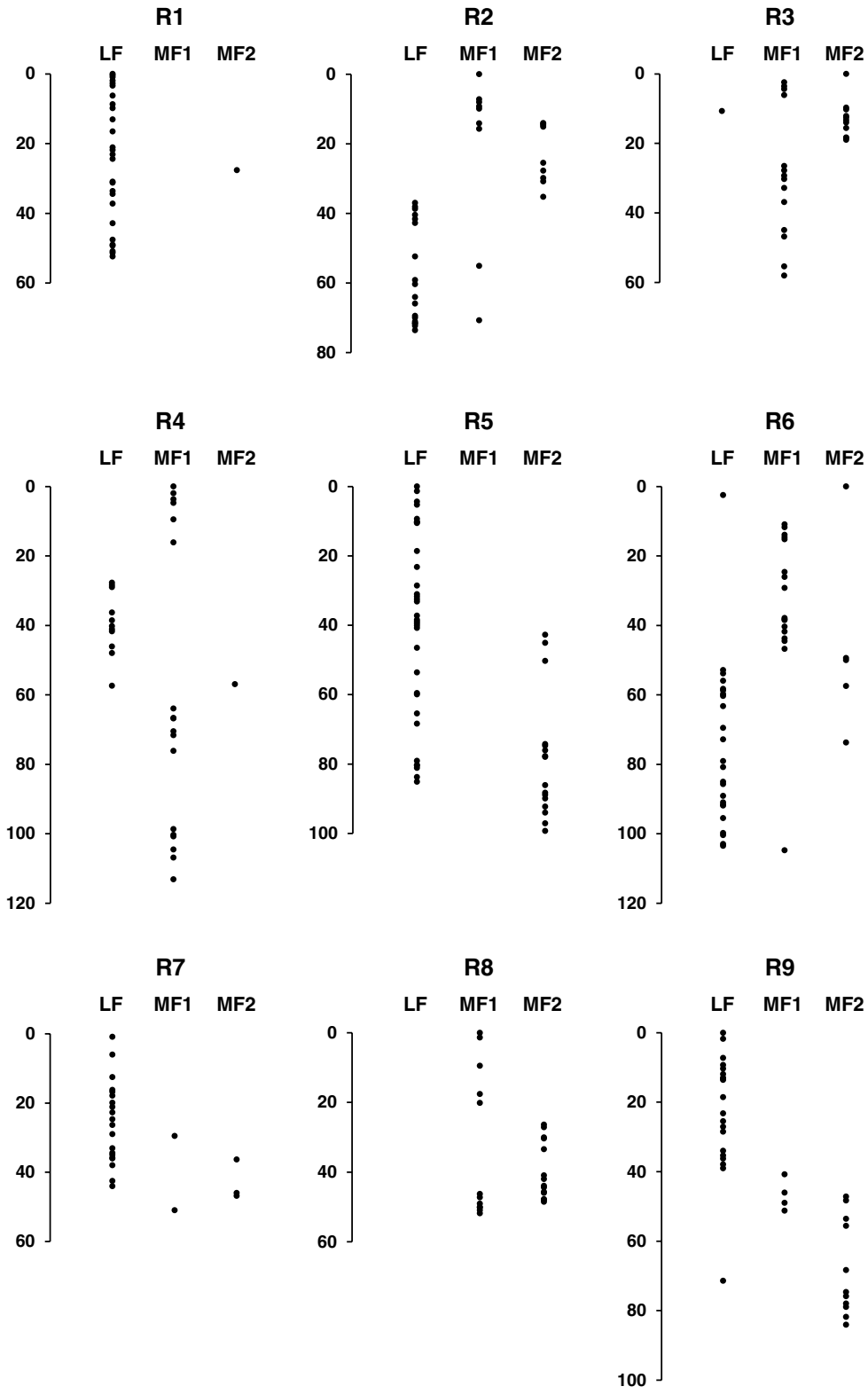

Fig. 6 Subgenomic characteristics of the triplicated *Rs* genome. The COS genes mapped to the *Rs* genetic map were plotted according to the position of their counterparts in the triplicated subgenomes of *Br*. LF, MF1, and MF2 indicate three differentially fractionated *Br* subgenomes. The *vertical axes* indicate genetic position in cM

Multiple rounds of polyploidy are evident during *Brassica* evolution (Mun et al. 2009). K_s estimates suggest that the *Rs* genome shared similar polyploidy events with *Br* (Fig. 3) and the most recent polyploidy event might be whole-genome triplication based on the mapping of AK blocks (Fig. 5a). The *Rs* genomes might be evolved from a pre-triplicated common ancestral genome of *Brassica* and *Raphanus* by chromosomal rearrangement considering

the triplicated nature of *Brassica* genomes and the similar genome divergence time point of *Rs–Br* in this study and *Br–Bo* in our previous report (Mun et al. 2009). This is in agreement with the taxonomic studies that genus *Raphanus* is a close relative of genus *Brassica* (Lü et al. 2008; Warwick and Black 1991, 1997) in the tribe Brassiceae.

In addition to the polyploidy and speciation events, a chromosome synteny map based on *Rs* and *Br* COS genes revealed subgenomic organization of the *Rs* genome (Fig. 5b; Supplementary Fig. S5). Previous genetic maps (Li et al. 2011; Shirasawa et al. 2011) indicated a syntenic relationship between *Rs* and *Br* genomes along with AK mapping but differentially fractionated subgenome blocks were not defined presumably due to non-allelic relationships between paralogs. In contrast, we could identify the subgenomic block structure of the *Rs* genome based on the orthologous syntenic relationship of COS genes. One notable characteristics of the *Rs* genome is R1, which retained all three blocks of AK1 as a LF subgenome. The similar structure of retained AK1 was also evident in the S05 chromosome of *Sa*, which has a triplicated genome. Comparative mapping of the single copy RFLP marker, pO105, suggested that S05 is syntenic to A6 of *Br*, which corresponds to the LF subgenome in *Br* (Nelson et al. 2011). BLAST search of the pO105 sequence in *Br* also identified Bra020001 in A6 (data not shown). Thus, it is likely that S05 is a LF subgenome in *Sa*. Phylogenic study based on chloroplast genes suggested that *Rs* belongs to the *Br*/*Bo* lineage whereas *Sa* is related to the *Bn* lineage, and the two lineages diverged approximately 20 Mya (Arias et al. 2014). In contrast, studies of nuclear DNA markers suggested that *Rs* is closely related to *Bn* than *Br* and *Bo* (Song et al. 1990, 1988; Thormann et al. 1994; Yang et al. 1999). Considering the fact that the LF subgenome of AK1 has been rearranged in the genomes of *Brassica* species including *Br*, *Bo*, and *Bna*, R1 and S05 chromosomes might be inherited from the common ancestor of *Rs* and *Sa*. This result supports previous suggestions on the origin of *Rs* as a hybrid between *Br*/*Bo* and *Bn* lineages (Song et al. 1990; Yang et al. 2002).

Further analysis of the complete genome sequences of both *Rs* and *Brassica* species will allow determination of the precise location of AK blocks on the *Rs* genome as well as the genomic relationships between *Rs–Br*/*Bo* and *Rs–Bn*. We anticipate that the COS markers developed in this study will have applications for taxonomic classification, comparative mapping, and genomic analysis of *Brassica* species and their relatives on the basis of orthology.

Acknowledgments This work was supported by grants from the Next-Generation Biogreen21 program (PJ008019), Rural Development Administration, Korea to HJY and 2013 Research Fund of Myongji University to JHM. We thank Dr. Shengyi Liu (Oil Crops Research Institute of CAAS, China) for kindly providing sequence information of *Brassica oleracea*, Sin-Gi Park (National Academy of Agricultural Science of RDA, Korea) for bioinformatics support, and Dr. Suhyoung Park (National Institute of Horticultural and Herbal Science of RDA, Korea) for providing plant material.

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

Ethical standards 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

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Arias T, Beilstein M, Tang M, McKain M, Pires J (2014) Diversification times among *Brassica* (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. Am J Bot 101:86–91

Beilstein M, Al-Shehbaz I, Kellogg E (2006) Brassicaceae phylogeny and trichome evolution. Am J Bot 93:607–619

Beilstein M, Nagalingum N, Clements M, Manchester S, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 107:18724–18728

Bushakra JM, Sargent DJ, Cabrera A, Crowhurst R, Girona EL, Velasco R, Symonds VV, Knaap E, Troggio M, Gardiner SE, Chagné D (2011) Rosaceae conserved orthologous set (RosCOS) markers as a tool to assess genome synteny between *Malus* and *Fragaria*. Tree Genet Genomes 8:643–658

Cabrera A, Kozik A, Howad W, Arus P, Iezzoni AF, van der Knaap E (2009) Development and bin mapping of a Rosaceae conserved ortholog set (COS) of markers. BMC Genom 10:562

Chapman MA, Chang J, Weisman D, Kesseli RV, Burke JM (2007) Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). Theor Appl Genet 115:747–755

Cheng F, Wu J, Fang L, Wang X (2012) Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. Front Plant Sci 3:198

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

Dassanayake M, Oh DH, Haas JS, Hernandez A, Hong H, Ali S, Yun DJ, Bressan RA, Zhu JK, Bohnert HJ, Cheeseman JM (2011) The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet 43:913–918

Economic Research Service USDA (2008) Vegetables and melons outlook. http://www.ers.usda.gov/Publications/VGS/Tables/World.pdf

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Yang L, Nasrallah ME, Bergelson J, Carrington JC, Gaut BS, Schmutz J, Mayer KF, Van de Peer Y, Grigoriev IV, Nordborg M, Weigel D, Guo YL (2011) The *Arabidopsis lyrata*

genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

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 Genom 14:717

Kim B, Yu H, Park S, Shin J, Oh M, Kim N, Mun J (2012) Identification and profiling of novel microRNAs in the *Brassica rapa* genome based on small RNA deep sequencing. BMC Plant Biol 12:218

Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the Pinaceae. Genetics 168:447–461

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

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

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

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

Liewlaksaneeyanawin C, Zhuang J, Tang M, Farzaneh N, Lueng G, Cullis C, Findlay S, Ritland CE, Bohlmann J, Ritland K (2008) Identification of COS markers in the Pinaceae. Tree Genet Genomes 5:247–255

Lü N, Yamane K, Ohnishi O (2008) Genetic diversity of cultivated and wild radish and phylogenetic relationships among *Raphanus* and *Brassica* species revealed by the analysis of *trnK/matK* sequence. Breed Sci 58:15–22

Lysak M, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA 103:5224–5229

Mun J-H, Kwon S, Yang T, Seol Y, Jin M, Kim J, Lim M, Kim J, Baek S, Choi B, Yu H, Kim D, Kim N, Lim K, Lee S, Hahn J, Lim Y, Bancroft I, Park B (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

Nelson MN, Parkin IA, Lydiate DJ (2011) The mosaic of ancestral karyotype blocks in the *Sinapis alba* L. genome. Genome 54:33–41

Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using intron polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. BMC Genom 9:113

Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genomics 9:473–484

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542

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

Slotte T, Hazzouri K, Ågren J, Koenig D, Maumus F, Guo Y, Steige K, Platts A, Escobar J, Newman L, Wang W, Mandáková T, Vello E, Smith L, Henz S, Steffen J, Takuno S, Brandvain Y, Coop G, Andolfatto P, Hu T, Blanchette M, Clark R, Quesneville H, Nordborg M, Gaut B, Lysak M, Jenkins J, Grimwood J, Chapman J, Prochnik S, Shu S, Rokhsar D, Schmutz J, Weigel D, Wright S (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat Genet 45:831–835

Song KM, Osborn TC, Williams PH (1988) *Brassica* taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs): 2. Preliminary analysis of subspecies within *B. rapa* (syn. campestris) and *B. oleracea*. Theor Appl Genet 76:593–600

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

The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

The *Brassica rapa* Genome Sequencing Project Consortium (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43:1035–1039

Thormann CE, Ferreira ME, Camargo LE, Tivang JG, Osborn TC (1994) Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor Appl Genet 88:973–980

Timms L, Jimenez R, Chase M, Lavelle D, McHale L, Kozik A, Lai Z, Heesacker A, Knapp S, Rieseberg L, Michelmore R, Kesseli R (2006) Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. Genetics 173:2227–2235

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

Warwick S, Black L (1991) Molecular systematics of *Brassica* and allied genera (Subtribe Brassicinae, Brassiceae)—chloroplast genome and cytodeme congruence. Theor Appl Genet 82:81–92

Warwick S, Black L (1997) Phylogenetic implications of chloroplast DNA restriction site variation in subtribes Raphaninae and Cakilinae (Brassicaceae, tribe Brassiceae). Canadian J Bot 75:960–973

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407–1420

Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD (2009a) A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus Capsicum. Theor Appl Genet 118:1279–1293

Wu F, Eannetta NT, Xu Y, Tanksley SD (2009b) A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. Theor Appl Genet 118:927–935

Wu F, Eannetta NT, Xu Y, Plieske J, Ganal M, Pozzi C, Bakaher N, Tanksley SD (2010) COSII genetic maps of two diploid Nicotiana species provide a detailed picture of synteny with tomato and insights into chromosome evolution in tetraploid *N. tabacum*. Theor Appl Genet 120:809–827

Wu HJ, Zhang Z, Wang JY, Oh DH, Dassanayake M, Liu B, Huang Q, Sun HX, Xia R, Wu Y, Wang YN, Yang Z, Liu Y, Zhang W, Zhang H, Chu J, Yan C, Fang S, Zhang J, Wang Y, Zhang F, Wang

G, Lee SY, Cheeseman JM, Yang B, Li B, Min J, Yang L, Wang J, Chu C, Chen SY, Bohnert HJ, Zhu JK, Wang XJ, Xie Q (2012) Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Natl Acad Sci USA 109:12219–12224

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang YW, Lai KN, Tai PY, Ma DP, Li WH (1999) Molecular phylogenetic studies of *Brassica*, *Rrippa*, *Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S–25S rDNA. Mol Phylogenet Evol 13:455–462

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

Yang R, Jarvis DE, Chen H, Beilstein MA, Grimwood J, Jenkins J, Shu S, Prochnik S, Xin M, Ma C, Schmutz J, Wing RA, Mitchell-Olds T, Schumaker KS, Wang X (2013) The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci 4:46