Genetics and Molecular Research

Evolution of mustard (*Brassica juncea* Coss) subspecies in China: evidence from the chalcone synthase gene

F.B. Chen, H.F. Liu, Q.L. Yao and P. Fang

Department of Life Sciences, Yangtze Normal University, Fuling, Chongqing, China

Corresponding author: H.F. Liu
E-mail: chenfabo963@126.com

Genet. Mol. Res. 15 (2): gmr.15028045
Received November 11, 2015
Accepted January 18, 2016
Published May 6, 2016
DOI http://dx.doi.org/10.4238/gmr.15028045

ABSTRACT. To explore the phylogenetic relationship, genome donor, and evolutionary history of the polyploid mustard (*Brassica juncea*) from China, eighty-one sequences of the chalcone synthase gene (*Chs*) were analyzed in 43 individuals, including 34 *B. juncea*, 2 *B. rapa*, 1 *B. nigra*, 2 *B. oleracea*, 1 *B. napus*, 1 *B. carinata*, and 2 *Raphanus sativus*. A maximum likelihood analysis showed that sequences from *B. juncea* were separated into two well-supported groups in accordance with the A and B genomes, whereas the traditional phenotypic classification of *B. juncea* was not wholly supported by the molecular results. The SplitsTree analysis recognized four distinct groups of *Brassicaceae*, and the median-joining network analysis recognized four distinct haplotypes of *Chs*. The estimates of Tajima's *D*, Fu and Li's *D*, and Fu and Li's *F* statistic for the *Chs* gene in the B genome were negative, while those in the A genome were significant. The results indicated that 1) the *Chs* sequences revealed a high level of sequence variation in Chinese mustard, 2) both tree and reticulate evolutions existed, and artificial selection played an important role in the evolution of Chinese

mustard, 3) the original parental species of Chinese mustard are *B. rapa* var. *sinapis arvensis* and *B. nigra* (derived from China), 4) nucleotide variation in the B genome was higher than that in the A genome, and 5) cultivated mustard evolved from wild mustard, and China is one of the primary origins of *B. juncea*.

Key word: *Brassica juncea*; *Chs* genes; Evolution; Mustard; Polyploid

INTRODUCTION

Mustard (*Brassica juncea*), a species of the genus *Brassica* belonging to the family Brassicaceae, is an agriculturally and economically important crop widely cultivated in Asia and Europe (Warwick et al., 2006). All species of mustard are polyploids (AABB), with the chromosome number $2n = 36$. China possesses the richest mustard resources, which contain more than 1000 cultivars distributed all over the country, and are widely used as a vegetable as well as for producing cooking oil (Yao et al., 2012). Under the long evolutionary period imposed by nature and humans, Chinese mustard has evolved from the original dwarf variety into varieties possessing great variations in root, leaf, stem, and seed stalk forms (Yao et al., 2012). Morphologically, Yang et al. (1989) classified Chinese mustard into 16 varieties, which are widely accepted in the classification of Chinese mustards, including *B. juncea* var. *megarrhiza*, *B. juncea* var. *crassicaulis*, *B. juncea* var. *gemmifera*, *B. juncea* var. *tumida*, *B. juncea* var. *rugosa*, *B. juncea* var. *foliosa*, *B. juncea* var. *leucanthus*, *B. juncea* var. *multisecta*, *B. juncea* var. *longepetiolata*, *B. juncea* var. *linearifolia*, *B. juncea* var. *strumata*, *B. juncea* var. *latipa*, *B. juncea* var. *involuta*, *B. juncea* var. *capitata*, *B. juncea* var. *multiceps*, and *B. juncea* var. *utilis*. Meng et al. (2006) divided Chinese mustard into five different types - leaf mustard, stem mustard, root mustard, seed stalk mustard, and seed mustard - based on the basic characteristics of leaf, root, stem, flower, and seed, respectively. Cytologically, the karyotypes of three types of vegetable mustards were investigated by Xu et al. (2014) and were found to have the same chromosome number, $2n = 36$, although their karyotypes and symmetry exhibited some difference. Phytogeographically, Chen and Chen (1992) suggested that China, especially the northwest, is one of the primary origins of *B. juncea*, while the cultivated mustard appeared as early as the sixth century BC. Additionally, Sichuan is considered the secondary center of origin of *B. juncea*. Tong and Chen (1990) speculated that the original parental species of vegetable mustard are the wild black mustard and the primitive Chinese cabbage, derived from China, based on physiochemical evidence from acid phosphatase isozymes via zymography. Based on analysis using molecular markers, Qiao et al. (1998) divided the 16 vegetable mustard varieties into A, B, and C groups. Fu et al. (2006) classified nine typical accessions of Chinese mustard crops into two primary groups. Qi et al. (2008) reported that the vegetable mustard could be grouped into two main groups and some minor branches. Song et al. (2009) classified 28 accessions into three groups. Yao et al. (2012) separated 34 mustard landraces into seven clusters, which was not congruent with the classification based on phenotype. By sampling the DNA sequences of nuclear internal transcribed spacer (ITS) regions from the Chinese vegetable mustard and its putative parents, *B. rapa* and *B. nigra*, Qi et al. (2007) concluded as follows: two strongly supported clades were identified, one having a closer relationship with the B-genome species *B. nigra* lineage and the other with

the A-genome species *B. rapa* lineage, suggesting that *B. juncea* was closely related to the A-genome type, and that the traditional phenotypic classification of *B. juncea* was not wholly supported by the ITS results. Despite these studies, little is known about the evolutionary history of the Chinese mustard subspecies, especially the polyploid molecular evolution of the special AABB group. Hence, the phylogenetic relationships among these subspecies need to be reconsidered on a molecular level.

Recently, single- and low-copy genes have received increasing attention in plant evolution and are becoming ideal tools for studying the origin and evolution of polyploid taxa (Hochbach et al., 2015). The chalcone synthase gene (*Chs*), widespread in plants, is a single- and low-copy gene encoding the first enzyme in the flavonoid biosynthesis pathway (Bao et al., 2015). The chalcone synthase gene was recognized as an excellent marker for the analysis of the origin of polyploid species because 1) it is highly conserved in different species in plants (Abe and Morita, 2010), 2) it has provided highly robust phylogenetic reconstructions, particularly at deeper nodes (Yang et al., 2003), and 3) it is biparentally inherited and highly variable (Zhao et al., 2010). However, no evidence from single- and low-copy genes has been used to analyze the phylogenetic relationships of Chinese mustard. Therefore, in this study, the molecular phylogenetic relationships of Chinese mustard species and their closely related genera were analyzed using data from single-copy, nuclear *Chs* sequences. The aim of this study was to i) estimate the *Chs* nucleotide polymorphism in Chinese mustard, ii) elucidate the phylogenetic relationships among the subspecies of Chinese mustard, and iii) infer the maternal donors and relationships of the A and B genomes in polyploid mustard.

MATERIAL AND METHODS

Materials

Forty-three individuals were sampled, including 34 *B. juncea*, 2 *B. rapa*, 1 *B. nigra*, 2 *B. oleracea*, 1 *B. napus*, 1 *B. carinata*, and 2 *Raphanus sativus* (Table 1 and Figure 1). The individuals were chosen based on the genetic relationships between Chinese mustard and its relatives. Table 1 lists the names, origins, and GenBank accession Nos. of the individuals collected by the authors of this paper. The seeds and voucher specimens were deposited at the herbarium of Crop Genetics and Breeding Research Centre, Yangtze Normal University, China.

DNA amplification and sequencing

Total genomic DNA was extracted from fresh young leaves (Yao et al., 2012). The first and second exons, approximately 1200 bp in length, were amplified with *Chs*-specific primers. The sequences of Chs-specific primers were as follows: 5'-CTT CAT CTG CCC GTC CAT CAT ACC-3' (forward primer) and 5'-GGAACGCTGTGCAAGAC-3' (reverse primer). Primers were synthesized by Yinggen Bio-Tech, Shanghai, China. The polymerase chain reaction (PCR) was performed in a reaction mixture (25 µL) containing 12.5 µL 2X Taq Master Mix (Kangweishiji Biotech Co., Ltd., Beijing, China), 2 µL each primer (10 nmol/mL), 1 µL (50 ng) DNA template, and 7.5 µL RNase-free water. The thermocycler (Mastercycler Personal PCR System Eppendorf, Hamburg, Germany) program used for PCR was as follows: 5 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 55°C, and 2 min at 72°C, and a final extension phase of 10 min at 72°C.

Table 1. Materials used in the study and their source.

No.	Name	Chromosome	Latin name	Source	Accession No.
01	Qingcai	AA	*B. rapa* var. *chinensis* L.	Xichang, Sichuan, China	KP301150
02	Yili yueyoucai	AA	*B. rapa* var. *sinapis arvensis* Tsen et Lee	Yili, Xinjiang, China	KP301155
03	Heijie	BB	*B. nigra* L.	Yili, Xinjiang, China	KP301157
04	Shandong ganlan	CC	*B. oleracea* var. *capitata* L.	Shandong, Jinan, China	KP301158
05	Yaan cauliflower	CC	*B. oleracea* var. *botrytis* L.	Yaan, Sichuan, China	KP301162
06	Zhongyou1	AACC	*B. napus* L.	Beijing, China	KP301171,KP301172
07	Wild mustard	AABB	*B. juncea* var. *juncea* Tsen et Lee	Jiuquan, Ganshu, China	KP301175,KP301242
08	Zigong datoucai	AABB	*B. juncea* var. *megarrhiza* Tsen et Lee	Zigong, Sichuan, China	KP301176,KP301243
09	Wanyuan datoucai	AABB	*B. juncea* var. *megarrhiza* Tsen et Lee	Wanyuan, Sichuan, China	KP301177,KP301244
10	Neijiang bangcai	AABB	*B. juncea* var. *carassicaulis* Chen et Yang	Niejiang, Sichuan, China	KP301178,KP301179
11	Baijiacaitai	AABB	*B. juncea* var. *carassicaulis* Chen et Yang	Zigong, Sichuan, China	KP301180,KP301245
12	Chuannong1	AABB	*B. juncea* var. *gemmifera* Lee et Li	Yaan, Sichuan, China	KP301181,KP301247
13	Dianjiangbaoercai	AABB	*B. juncea* var. *gemmifera* Lee et Li	Dianjiang, Chongqing, China	KP301182,KP301246
14	Huangzhongzi	AABB	*B. juncea* var. *tumida* Tsen et Lee	Fuling, Chongqing, China	KP301190,KP301191
15	Zhetongyihao	AABB	*B. juncea* var. *tumida* Tsen et Lee	Yuyao, Zhejiang, China	KP301192,KP301249
16	Dongcai	AABB	*B. juncea* var. *rugose* Bailey	Dazhu, Chongqing, China	KP301193,KP301194
17	Midulvgan	AABB	*B. juncea* var. *rugose* Bailey	Midu, Yunnan, China	KP301195,KP301196
18	Baiganqingcai	AABB	*B. juncea* var. *foliosa* Bailey	Luzhou, Sichuan, China	KP301197,KP301198
19	Zhayetianqingcai	AABB	*B. juncea* var. *foliosa* Bailey	Mabian, Sichuan, China	KP301199,KP301200
20	Baihuacai	AABB	*B. juncea* var. *leucanthus* Chen et Yang	Luxian, Sichuan, China	KP301201,KP301202
21	Baihuaqingcai	AABB	*B. juncea* var. *leucanthus* Chen et Yang	Luxian, Sichuan, China	KP301203,KP301250
22	Huayejiecai	AABB	*B. juncea* var. *multisecta* Bailey	Ezhou, Hubei, China	KP301204,KP301252
23	Qinggenghuayejiecai	AABB	*B. juncea* var. *multisecta* Bailey	Nanchuang, Jiangxi, China	KP301205,KP301251
24	Liangpingxiangcai	AABB	*B. juncea* var. *longepetiolata* Yang et Chen	Liangping, Chongqing, China	KP301206,KP301253
25	Fengduxiangcai	AABB	*B. juncea* var. *longepetiolata* Yang et Chen	Fengdu, Chongqing, China	KP301207,KP301208
26	Yanjiweilacai	AABB	*B. juncea* var. *linearifolia* Sun	Xichang, Sichuan, China	KP301209,KP301210
27	Kuanyefengweicai	AABB	*B. juncea* var. *linearifolia* Sun	Zigong, Sichuan, China	KP301211,KP301212
28	Qingbangnainaicai	AABB	*B. juncea* var. *stromata* Tsen et Lee	Luxian, Sichuan, China	KP301213,KP301214
29	Daerduoqingcai	AABB	*B. juncea* var. *stromata* Tsen et Lee	Yuanjiang, Hunan, China	KP301215,KP301254
30	Dapianpianqingcai	AABB	*B. juncea* var. *latipa* Li	Meigu, Sichuan, China	KP301216,KP301217
31	Baiyeqingcai	AABB	*B. juncea* var. *latipa* Li	Zigong, Sichuan, China	KP301218,KP301255
32	Qingyebaobaocai	AABB	*B. juncea* var. *involute* Yang et Chen	Dianjiang, Chongqing, China	KP301219,KP301220
33	Baoxinqingcai	AABB	*B. juncea* var. *involute* Yang et Chen	Dazhou, Sichuan, China	KP301221,KP301256
34	Jixinjiecai	AABB	*B. juncea* var. *capitata* Hort	Chaozhou, Guangdong, China	KP301222,KP301223
35	Duanyejixinjiecai	AABB	*B. juncea* var. *capitata* Hort	Chenghai, Guangzhou, China	KP301224,KP301225
36	Dukexuelihong	AABB	*B. juncea* var. *multiceps* Tsen et Lee	Nantong, Jiangsu, China	KP301226,KP301227
37	Heiyexuelihong	AABB	*B. juncea* var. *multiceps* Tsen et Lee	Shanghai, China	KP301228,KP301257
38	Guizhoulacai	AABB	*B. juncea* var. *utilis* Li	Guiyang, Guizhou, China	KP301229,KP301230
39	Xiaoyechonglacai	AABB	*B. juncea* var. *utilis* Li	Banan, Chongqing, China	KP301231,KP301232
40	Maweisi	AABB	*B. juncea* Czern. et Coss	Suining, Sichuan, China	KP301233,KP301258
41	Aisaiebiyajie	BBCC	*B. carinata* Braun	Ethiopia	KP301234,KP301235
42	Mianyangluobo	RR	*R. Sativus* Linn	Mianyang, Sichuan, China	KP301238
43	Liupanshui luobo	RR	*R. Sativus* Linn	Liupanshui, Guizhou, China	KP301239

After electrophoresis of the PCR products on a 1.0% agarose gel, a single band of amplified product was excised and purified with an AxyPrep gel extraction kit (Axygen Biotechnology, Hangzhou, China). The purified DNA fragments were cloned into a pMD18-T vector (TaKaRa, Dalian, China). At least five positive clones for each species were randomly selected for sequencing. Positive clones were sequenced by Yinggen Bio-Tech.

The cloning of PCR amplicons from the single-copy nuclear genes of the allopolyploid species isolated homologous sequences from each nuclear genome. The DNAman 6.0 software (http://www.lynnon.com) was used to align and analyze the copy of the sequences from the A,

B, or C genome. Subsequently, primers specific to the A and B genomes were designed using the Primer 5 software (http://www.premierbiosoft.com). The primers specific to A and B were PA (R: 5'-GCA TTG ATC AAC CTC TTG TAA CT-3', F: 5'-GGA ACG CTG TGC AAG AC-3') and PB (R: 5'-TTG CAT AAA GTC ACA CAT CC-3', F: 5'-GGA ACG CTG TGC AAG AC-3'), respectively. After the DNA sequences of *Chs* genes were cloned and sequenced, they were submitted to GenBank. The homologous sequences EF408922 and GQ983033 representing *B. rapa* and *B. nigra* were downloaded from GenBank.

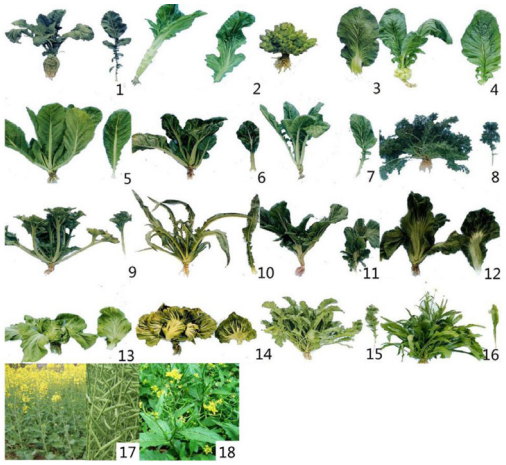

Figure 1. Phenotypes of 18 subspecies in Chinese mustard. The subspecies are arranged as follows: 1) *B. juncea* var. *megarrhiza*, 2) *B. juncea* var. *carassicaulis*, 3) *B. juncea* var. *gemmifera*, 4) *B. juncea* var. *tumida*, 5) *B. juncea* var. *rugosa*, 6) *B. juncea* var. *foliosa*, 7) *B. juncea* var. *leucanthus*, 8) *B. juncea* var. *multisecta*, 9) *B. juncea* var. *longepetiolata*, 10) *B. juncea* var. *linearifolia*, 11) *B. juncea* var. *strumata*, 12) *B. juncea* var. *latipa*, 13) *B. juncea* var. *involuta*, 14) *B. juncea* var. *capitata*, 15) *B. juncea* var. *multiceps*, 16) *B. juncea* var. *utilis*, 17) *B. juncea* Czern. et Coss (rape mustard), and (18) *B. juncea* var. *juncea* (wild mustard).

Phylogenetic analysis

Using maximum likelihood (ML) and Bayesian inference (BI), a phylogenetic analysis was performed using exon plus intron data matrices. The ML analysis of the exon plus intron data set was conducted using the PAUP* 4.0 software (Swofford, 2002). The outgroup was *R. sativus* and the evolutionary model used for the data set was determined using the Akaike information criterion (AIC) of the ModelTest v3.0 software (Darriba et al., 2012). The best-fit models for the data set were GTR + G + I. Maximum likelihood heuristic searches were performed with tree bisection-reconnection, branch-swapping algorithm and 1000 random-addition sequence replicates. Bootstrap support (BS) was used to estimate the topological robustness of the ML trees. The bootstrap analysis was carried out with 500 replications using simple taxon addition.

Bayesian interference was performed using the MrBayes v3.2 software (Ronquist et al., 2012). Using MrBayes default heating values (t = 0.2), with trees sampled every 100 generations, four chains of the Markov Chain Monte Carlo were simultaneously run for 4,000,000 total generations. The first 18,700 trees were “burned in” the chains and discarded. To ensure that log likelihoods were in the stationary “fury caterpillar” phase, the Tracer v1.4

program (Rambaut and Drummond, 2007; Fan et al., 2009) was used. The majority-rule consensus tree was established on the basis of the remaining trees. Two independent runs were conducted to examine the convergence on the same posterior distribution, and the statistical confidence of the nodes was estimated using the posterior probability (PP).

SplitsTree analysis

To detect reticulate evolution among *B. juncea*, the phylogenetic trees were inferred by SplitsTree 4.13 using the NeighborNet method (Huson and Bryant, 2006).

Network analysis

Relationships between haplotypes of the taxa sampled were analyzed by phylogenetic network reconstruction. The median-joining (MJ) network method was used in this study due to its robustness compared with other network methods for resolving gene phylogenies in a simulation study (Cassens et al., 2005). The MJ network was constructed using the Network 4.6.1.3 program (Fluxus Technology Ltd., Suffolk, UK). An analysis to detect recombination was performed using the *HyPhy* version 0.99 (Pond et al., 2005), since MJ networks are inferred from nonrecombinant DNA (Bandelt et al., 1999).

Estimate of nucleotide diversity

To evaluate the nucleotide diversity of the A and B genomes in Chinese mustard, sequence variations in *Chs* were estimated by Tajima's $\hat{\pi}$ (Tajima, 1983), Watterson's $\hat{\theta}$ (Watterson, 1975), and the number of shared polymorphisms and fixed differences. Tajima's $\hat{\pi}$ quantifies the mean pairwise differences between sequences, whereas Watterson's $\hat{\theta}$ refers to an index of the number of polymorphic sites. Both $\hat{\pi}$ and $\hat{\theta}$ have expected values of $4N\mu$, where N is the population size and μ the mutation per locus per generation. A fixed difference is a site where all sequences sampled in given a taxon have one base while those in another taxon have a different base, whereas in shared polymorphisms two taxa have the same two bases segregating at the same site (Hey, 1991). A test of the neutral evolution model (including Tajima's *D* and Fu and Li's *D* statistic, and HKA test) was carried out using the methods of Tajima (1989) and Fu and Li (1993). All parameters ($\hat{\pi}$, $\hat{\theta}$, fixed difference, shared polymorphisms, and Tajima's *D*, Fu and Li's *D* statistic, and HKA test) were computed using the DnaSP v5 software (Librado and Rozas, 2009).

RESULTS

Sequence analyses

Sequence comparison of all species showed that the average length of the *Chs* DNA sequence was 1455 bp, varying from 1295 to 1497 bp. The mean lengths of exons and introns were 1184 and 271 bp, respectively. Of the 1455 total sites of *Chs* sequence data, 334 were variable sites, 1116 were conserved, 236 were informative, and 98 were singleton. The nucleotide variation in the exon region (214 variable sites, 132 informative sites) was higher than that of the intron region (120 variable sites, 100 informative sites).

Phylogenetic analyses

Five positive clones were sequenced for each individual. In the case of multiple-identical sequences in an individual, only one sequence was used in the data set. In a tetraploid species, two copies of ancestral allelic types of the *Chs* gene were successfully cloned, while three copies of ancestral allelic types of the *Chs* gene were obtained in all tetraploids. To analyze the relationships among the *B. juncea* and its affinitive species, the phylogenetic analysis was conducted with 81 unique sequences of all the related diploid and polypoid species in *Brassica*.

Based on the nonrecombination signal in the alignment, only exon data were used to reconstruct the MJ network (-log likelihood = 3226.14; AIC = 2876.65). The ML analysis on the basis of the complete data set resulted in a single-phylogenetic tree with ML parameters as follows: the inferred nucleotide frequencies A: 0.2124, G: 0.2752, T: 0.2338, and C: 0.2786; the gamma distribution with shape parameter $k = 0.6372$; and the ratio of invariable sites = 0.1497. A similar topology was revealed in the BI analysis. The ML tree, with PP above and BS values below each branch, is displayed in Figure 2. Sequences from the family Brassicaceae, including *B. juncea*, *B. rapa*, *B. nigra*, *B. napus*, *B. carinata*, and *R. sativus*, separated into two well-supported groups (group I and group II), where sequences from *R. sativus* yielded a distinct group (group II). Group I, containing three clades (A, B, and C) with well-defined statistical support, is comprised of sequences from *B. juncea*, *B. rapa*, *B. nigra*, *B. napus*, and *B. carinata*. These findings are in accordance with the A, B, and C genomes revealed by *Chs*. Clade A (83% PP and 79% BS), which includes four subclades (A1, A2, A3, and A4), contains the A-genome sequences of *B. juncea* and the sequences *B. napus*, EF408922 (*B. rapa*), *B. rapa* var. *chinensis*, and *B. rapa* var. *sinapis arvensis*. Subclade A1 (86% PP and 81% BS) consists of eight *B. juncea*, including four leaf mustard, three stem mustard, and one seed stalk mustard. Subclade A2 (91% PP and 83% BS) consists of *B. rapa* var. *chinensis* and *B. napus*. Subclade A3 (83% PP and 59% BS) consists of 25 *B. juncea* sequences and one *B. rapa* var. *sinapis arvensis*, and includes wild mustard (*B. juncea* var. *juncea*), two seed mustards (wild mustard rape and mustard rape), two root mustards, 18 leaf mustards, and three stem mustards. Subclade A4 (98% PP and 88% BS) contains only one *B. juncea* var. *utilis* (seed stalk mustard) and one *B. rapa* (EF408922). Clade B (88% PP and 72% BS), which includes four subclades (B1, B2, B3, and B4), contains the B-genome sequences of *B. juncea* and the sequences *B. carinata*, *B. nigra*, and GQ983033.1. Subclade B1 (88% PP and 72% BS) consists of *B. juncea* var. *gemmifera* (stem mustard) and *B. juncea* var. *strumata* (leaf mustard). Subclade B2 (92% PP and 82% BS) consists of six *B. juncea*, including one seed stalk mustard, three leaf mustard, one stem mustard, and one mustard rape. Subclade B3 (85% PP and 59% BS) consists of two *B. juncea* var. *megarrhiza* (root mustard), *B. carinata*, and *B. nigra* (GQ983033.1). Subclade B4 (100% PP and 98% BS) consisted of 24 *B. juncea* and one *B. nigra*, including five stem mustard, 17 leaf mustard, one seed stalk mustard, and one wild mustard.

Splitstree analyses

The phylogenetic networks are mainly applied to display complicated reticulations above the species level, the relationships between intraspecific individuals and among populations, and the results of phylogenetic inference of contradicting data sets. In this study, the SplitsTree analysis was carried out to detect the reticulate evolution between *B. juncea*

and its closest relative taxa. The exon plus intron data were used to yield a Splitstree using the split-decomposition method. Figure 3 indicates that four distinct groups (I, II, III, and IV) of Brassicaceae were recognized. Group I includes four sequences from the C genome. Group II includes 37 sequences from the A genome that contains three clades (a, b, and c). Group III includes 36 sequences from the B genome that contains four clades (A, B, C, and D). Two sequences of *R. sativus* formed group IV as the outgroup. The classification results mostly overlapped with the results of ML phylogenetic tree. Moreover, as shown in Figure 3, both tree and reticulate evolutions occurred among Chinese mustards and related plants.

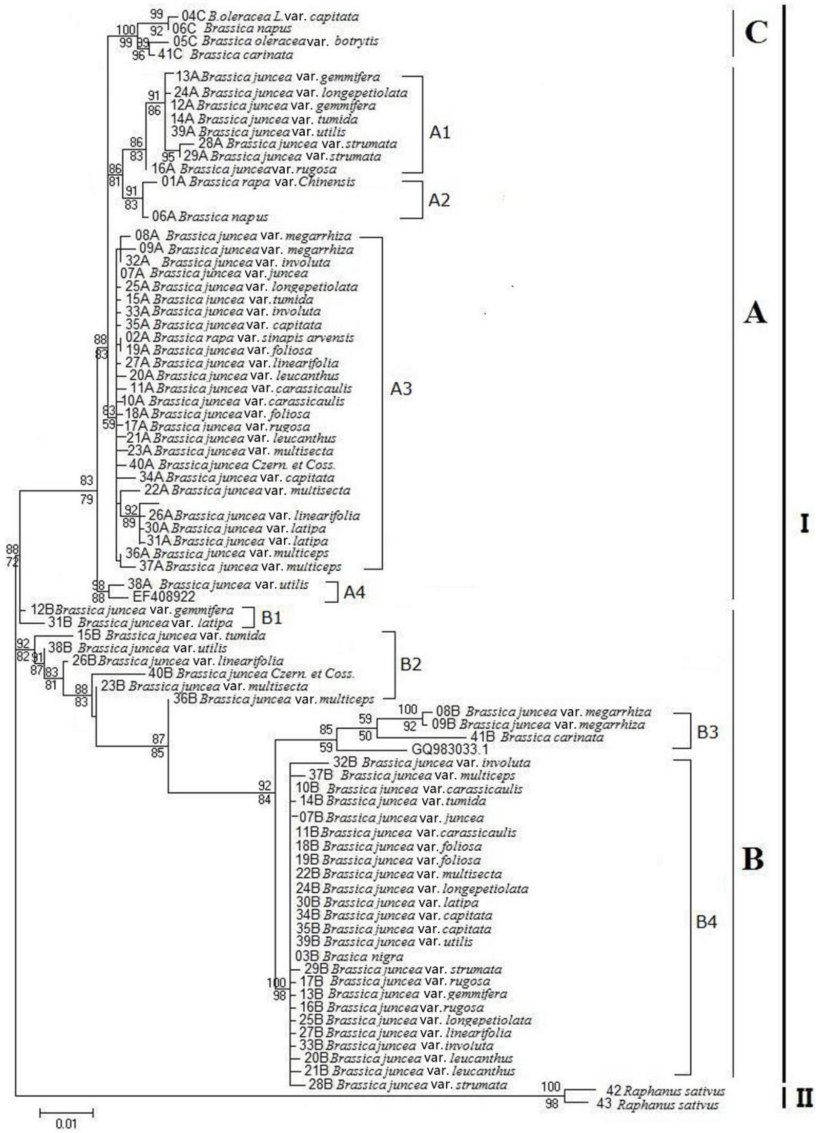

Figure 2. Maximum likelihood (ML) tree inferred from *Chs* sequences among *Brassica* species in China.

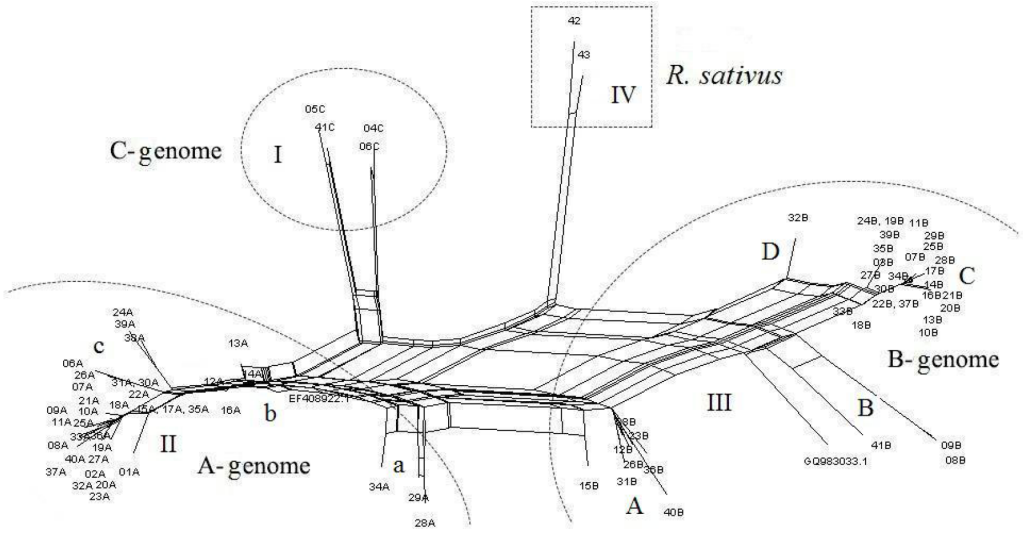

Figure 3. SplitsTree inferred from *Chs* sequences among *Brassica* species in China.

Network analyses

A network is used to reconstruct phylogenetic networks and trees, infer ancestral types and potential types, study evolutionary branching and variants, and estimate dating. The exon data set was used to infer the MJ network due to the absence of a recombination signal in its alignment. The MJ network illustrated the genealogical relationship between 38 haplotypes, derived from 81 sequences (Figure 4). In the MJ analysis, a circular network node represents a single haplotype, and the size of the node is proportional to the number of isolates with the haplotype. Median vectors refer to unsampled nodes assumed by the MJ network analysis, and the number along branches represents the mutation site. The MJ network illustrated in Figure 4 reveals a high level of haplotype diversity. Four distinct types of haplotypes within the taxa are recognized, which corresponds to the radish, A, B, and C genomes revealed by the *Chs* phylogeny. The A-genome haplotypes were two mutational steps (at positions 392 and 847) away from the B-genome haplotypes, and four mutational steps (at positions 300, 795, 1036, and 1175) away from the C-genome haplotypes. The relationship between the A and B genomes is closer than it is between the A and C, and B and C genomes. In the A-genome haplotypes, one diploid (*B. rapa* var. *sinapis arvensis*) and 19 allotetraploid are at the central branching points. This indicates that *B. rapa* may be the parental donor of the A genome in the allotetraploid *B. juncea*. In the B-genome haplotypes, one diploid (*B. nigra*) and 20 allotetraploid are at the central branching points. This indicates that *B. nigra* may be the parental donor of the A genome in the allotetraploid *B. juncea*.

Genetic relationships among the A and B genomes of *B. juncea*

Table 2 lists the *Chs* sequence data of the A and B genomes in *B. juncea*. The mean lengths of exons and introns in the A genome were 1186 and 259 bp, respectively, while the

mean lengths of exons and introns in the B genome were 1204 and 238 bp, respectively. The nucleotide variation in the B genome (350 variable sites, 161 informative sites, 26 synonymous polymorphisms in exons, and 109 base substitutions in introns) was higher than that in the A genome (98 variable sites, 36 informative sites, 13 synonymous polymorphisms in exons, and 33 base substitutions in introns).

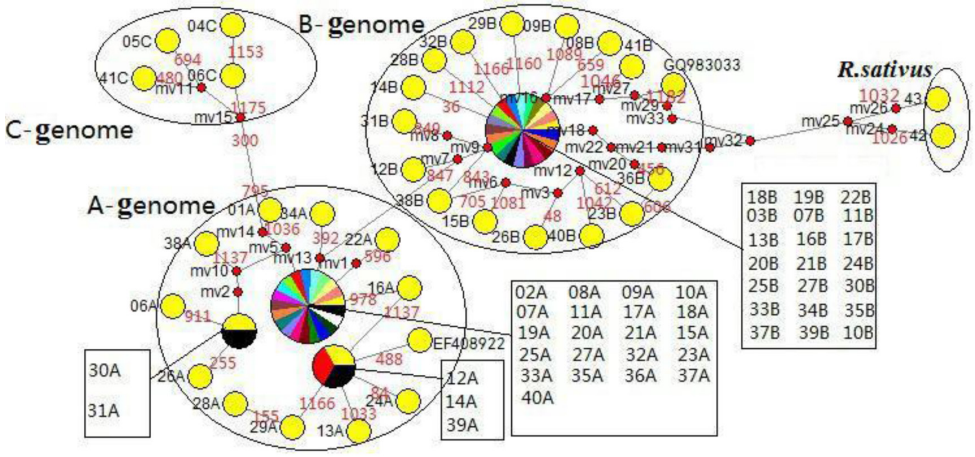

Figure 4. Median-joining (MJ) network derived from the *Chs* gene sequences among the *Brassica* species in China.

The B genome in *B. juncea* was also found to retain a high level of variation (0.192, 0.0261) in comparison to the A genome (0.0092, 0.0158) on the basis of estimates of $\hat{\pi}$ per bp and $\hat{\theta}$ per bp (Table 3). As estimated by the relative values of $\hat{\theta}$ per bp, the amount of variation detected in the A genome retained 57.47% of variation in the B genome. The estimates of Tajima's *D*, Fu and Li's *D*, and Fu and Li's *F* statistic for the B-genome *Chs* gene were negative (Table 3), suggesting that *Chs* of B genome is a neutral evolution gene. However, both the estimates of Fu and Li's *D* and Fu and Li's *F* statistic for the A-genome *Chs* gene were significant and large, indicating that there is evidence for the selection on sequence in the A genome. To further verify the factors affecting the evolution of the A genome in the *Chs* gene, the HKA test and misalignment analysis were carried out to detect the intergroup genetic evolutionary equilibrium. The HKA test of the A-genome sequence was significant ($\chi^2 = 6.086$, P = 0.0136), also suggesting that artificial selection plays an important role in the evolution of Chinese *B. juncea*. The results of the misalignment analysis of the *Chs* sequence from the A genome (Figure 5) showed that the genetic diversity of the A-genome sequences significantly decreased.

The genetic relationship between the A and B genomes was evaluated on the basis of the number of shared polymorphisms and fixed differences. A shared polymorphism exhibits a history of polymorphism that was not eliminated by genetic drift. By contrast, a fixed difference suggests that different taxa do not share genetic drift with independent evolution (Hey, 1991). The results show that many shared polymorphisms (32) and fixed differences (17) exist between the A and B genomes.

Table 2. Polymorphic sites of *Chs* sequences between the A and B genomes in *Brassica juncea*.

Population	Sample number	Exon					Intron			
		Syn	Rep	AL	VS	IS	Subs	AL	VS	IS
A genome	34	13	59	1186	65	22	33	259	33	14
B genome	34	26	114	1204	241	124	109	238	109	37

Syn = the number of synonymous polymorphisms in exon; Rep = the number of replacement or nonsynonymous polymorphisms in exons; Subs = the number of base substitutions in intron; AL = the average sequence length of the taxa; VS = variable sites; IS = informative sites.

Table 3. Sequence polymorphism and neutral evolution test in the A and B genomes of *Brassica juncea*.

Population	*N*	*S*	*H*	$\hat{\pi}$ (bp)	$\hat{\theta}$ (bp)	D	F_D	F_{fl}	R_m
A genome	34	81	30	0.0092	0.0158	-1.573 (P > 0.05)	-2.9161* (0.01 < P < 0.05)	-2.9138* (0.01 < P < 0.05)	3
B genome	34	131	25	0.0198	0.0261	-0.9035 (P > 0.05)	-0.99885 (P > 0.05)	1.14681 (P > 0.05)	13

N = sample numbers; *S* = the number of segregating sites; *H* = haploid numbers; *D* = Tajima's *D*; F_D = Fu and Li's *D*; F_{fl} = Fu and Li's *F*; R_m = the minimum recombination number.

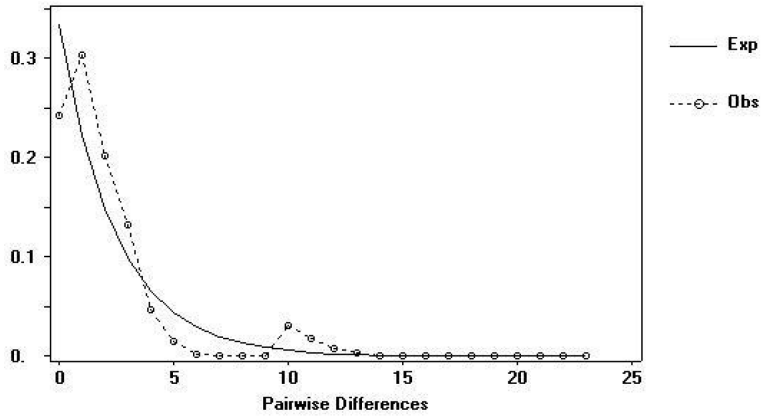

Figure 5. Misalignment analysis in the A gene of *Chs* sequence.

DISCUSSION

Sequence polymorphism of the *Chs* gene in Chinese mustard

In this study, a 1.2-kb domain of the *Chs* gene was detected from 43 individuals with 79 sequences representing the Chinese mustards and related species. Overall, 334 variable, 1116 conserved, 236 informative sites, and 98 singleton sites were found in the *Chs* sequences studied. Consistent with results reported by Qiao et al. (1998), Fu et al. (2006), Wu et al. (2009), Qi et al. (2008), and Yao et al. (2012), the highest level of sequence variation was also detected in Chinese mustards. In this study, the focus was on the polymorphisms between the A and B genomes of polyploid mustard. The nucleotide variation in the B genome was higher than that in the A genome. When investigating shared polymorphisms and fixed differences, the result showed that many shared polymorphisms (32) and fixed differences (17) existed between the A and B genomes, indicating no or very recent divergence between the two

genomes of Chinese mustard. The relationship between the A and B genomes is closer than between the A and C, and B and C genomes. According to the relative values of $\hat{\pi}$ and $\hat{\theta}$ per bp, the B genome in *B. juncea* retained a higher level of variation (0.192, 0.0261) than the A genome (0.0092, 0.0158). This result reinforced the data presented by Song et al. (2009) and Ge and Li (2007), which indicated that after the occurrence of polyploidy in mustard, the A and B genomes showed a different degree of variation and larger genetic variation within the B genome. The decreased variation in the A genome may be due to artificial selection, as evidenced by Fu and Li's *D*, Fu and Li's *F*, HKA test, and misalignment analysis. The genetic variation of Chinese mustard cultivars was affected by various factors throughout their evolutionary history. Outcrossing and fitness-relevant mutations generate intrapopulation diversity, while direct natural or human selection and bottleneck effects result in an increase in the diversity of the A genome (Fang et al., 2013).

Phylogenetic relationships of Chinese mustard

For many years, the relationships within the Chinese mustard subspecies were the subject of much controversy. Morphologically, Chinese mustard was divided into leaf mustard, stem mustard, root mustard, seed stalk mustard, and seed mustard, totaling 17 varieties (Yang et al., 1989; Meng et al., 2006). Using sequence-related amplified polymorphism markers, Li et al. (2014) classified 111 Chinese mustard accessions into four groups, which are in agreement with morphological classifications. The present *Chs* gene data show that Chinese mustard clustered into two distinct clades: clade A, containing all the *Chs* sequences from the A genome (including four subclades), while clade B is composed of four subclades from the B genome. The results support with studies by Qi et al. (2007), Qi et al. (2008), Song et al. (2009), and Yao et al. (2012), showing that the traditional phenotypic classification of *B. juncea* was not wholly supported by molecular results. The ML tree could not classify the Chinese mustard into leaf mustard, stem mustard, root mustard, seed stalk mustard, and seed mustard. For instance, with the exceptions of *B. juncea* var. *utilis* and *B. juncea* var. *gemmifera*, almost all subspecies were classified into subclade A3. However, with the exception of *B. juncea* var. *multiceps*, almost all subspecies were classified into subclade B4. Many subspecies with great original variations clustered into a zero-length branch (A3 and B4 clades). It is possible that natural hybridizations between mustards and the reticulate evolution events of Chinese mustard resulted in a mustard with a similar genetic background. The classification results based on the A- and B-genome sequences were also different, possibly as a result of asymmetric evolution in the A and B genomes.

Origin of *B. juncea* in China

According to a survey of agricultural heritage, mustard has been cultivated in China since ancient times, as early as the sixth century BC (Yao et al., 2012). Chinese vegetable mustards are highly diversified, with numerous variations of edible organs and morphologies. Evidence from chromosome number, chromosome pairing, artificial synthetic mustard, acid phosphatase isozymes via zymography, and DNA sequences (Qi et al., 2008) has demonstrated that mustard is an allotetraploid ($2n = 36$, AABB) synthesized by *B. rapa* ($2n = 20$, AA) and *B. nigra* ($2n = 16$, BB) via natural hybridization. In this study, two types of *Chs* sequences, A and B genomes, were cloned from all the *B. juncea* subspecies, allowing for the elucidation of the

phylogenetic relationships among *B. juncea* subspecies based on orthologous comparisons. The *Chs* data in the present study suggest that the *Chs* sequence of the A and B genomes are evolutionarily distinct. The nucleotide sequence diversity ($\hat{\pi}$) of the B genome was higher than that of the A genome, indicating that the *Chs* sequence of the B genome may evolve faster than that of the A genome. On the contrary, Liu et al. (2003) reported that the A genome evolved faster than the B genome in Chinese mustard. The results of network analyses further reinforced the conclusions of Chen et al. (1993) that the original parental species of Chinese mustard are *B. rapa* var. *sinapis arvensis* and *B. nigra* derived from China. As wild mustard distribution exists in China and wild mustards were at the central branching points of the haplotypes in both the A and B genomes, it was presumed that the cultivated mustard evolved from the wild mustard in China. The SplitsTree analyses indicated that many reticulate evolution events occurred during the evolutionary history of Chinese mustard and related species. Considering that no reproductive isolation exists among *B. juncea*, subspecies may incur many natural hybridization and reticulate evolution events. The results of the misalignment analysis in the A gene of *Chs* sequence, showing that the genetic diversity of A-genome sequences significantly decreased, also suggest that selection plays an important role in the evolution of Chinese mustard. It was presumed that a large number of natural mutations occurred in Chinese mustard during a certain historical period, after which different mustard varieties and cultivars were formed. Taken together, one can conclude that wild mustard is the progenitor of Chinese mustard and that China is one of the primary original locations of *B. juncea*, while the cultivated mustard has evolved from the wild mustard in China.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Projects of Team Innovation Plan of University in Chongqing (#KJTD201322), the Scientific and Technological Research Program of Chongqing Municipal Education Commission (#KJ121304), the Project of Chinese Ministry of Education (#Z2011143), and the Science Foundation Project of Fuling District (FLKJ, #2012ABB1085; FLKJ, #2013ABB2064).

REFERENCES

Abe I and Morita H (2010). ChemInform abstract: Structure and function of the chalcone synthase super family of plant type III polyketide synthases. *Nat. Prod. Rep.* 41: 809-838. http://dx.doi.org/10.1039/b909988n

Bandelt HJ, Forster P and Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16: 37-48. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026036

Bao Y, Guo CF, Chen SH and Liu M (2015). Molecular evolution of chalcone synthase gene superfamily in plants. *Chin. Bull. Bot.* 50: 55-71. http://dx.doi.org/10.3724/SP.J.1259.2015.00055

Cassens I, Mardulyn P and Milinkovitch MC (2005). Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Syst. Biol.* 54: 363-372. http://dx.doi.org/10.1080/10635150590945377

Chen CL and Chen XQ (1992). A study on origin of *Brassica juncea* Coss. in China. *J. Southwest Agri.* 5: 6-11.

Chen XQ, Zhou Y, Zhou GF, Fan YH, et al. (1993). Analysis of esterase isozyme in vegetable mustard (*Brassica juncea* Coss.). *Xi Nan Nong Ye Xue Bao* 6: 40-46.

Darriba D, Taboada GL, Doallo R and Posada D (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9: 772. http://dx.doi.org/10.1038/nmeth.2109

Fan X, Sha LN, Yang RW, Zhang HQ, et al. (2009). Phylogeny and evolutionary history of *Leymus* (Triticeae; Poaceae) based on a single-copy nuclear gene encoding plastid acetyl-CoA carboxylase. *BMC Evol. Biol.* 9: 247. http://dx.doi.org/10.1186/1471-2148-9-247

Fang P, Chen FB, Yao QL, Yang KC, et al. (2013). Analysis of genetic diversity in the tuber mustard (*Brassica juncea* var. *tumida* Tsenet Lee) in the Yangtze river basin of China. *Genet. Resour. Crop Evol.* 60: 129-143. http://dx.doi.org/10.1007/s10722-012-9821-y

Fu J, Zhang MF and Qi XH (2006). Genetic diversity of traditional Chinese mustard crops *Brassica juncea* as revealed by phenotypic differences and RAPD markers. *Genet. Resour. Crop Evol.* 53: 1513-1519. http://dx.doi.org/10.1007/s10722-005-7763-3

Fu YX and Li WH (1993). Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.

Ge XH and Li ZY (2007). Intra- and intergenomic homology of B-genome chromosomes in trigenomic combinations of the cultivated *Brassica* species revealed by GISH analysis. *Chromosome Res.* 15: 849-861. http://dx.doi.org/10.1007/s10577-007-1168-4

Hey J (1991). A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. *Theor. Popul. Biol.* 39: 30-48. http://dx.doi.org/10.1016/0040-5809(91)90039-I

Hochbach A, Schneider J and Röser M (2015). A multi-locus analysis of phylogenetic relationships within grass subfamily Pooideae (Poaceae) inferred from sequences of nuclear single copy gene regions compared with plastid DNA. *Mol. Phylogenet. Evol.* 87: 14-27. http://dx.doi.org/10.1016/j.ympev.2015.03.010

Huson DH and Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254-267. http://dx.doi.org/10.1093/molbev/msj030

Li N, Jun SJ, Ren XS, Song HY, et al. (2014). SRAP analysis of mustard germplasm genetic diversity and genetic relationship. *China Vegetables* 1: 26-33.

Librado P and Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452. http://dx.doi.org/10.1093/bioinformatics/btp187

Liu AH, Wang JB and Zhu YG (2003). RAPD analysis on the genome evolution of polyploids in *Brassica. Acta Phytotaxon. Sin.* 41: 520-530.

Meng QF, Wang YH, Huangpu WG, Xue XC, et al. (2006). Identification and cluster analysis of botanical traits in mustard vegetables. *Zhejiang Agri. Sci.* 3: 224-227.

Pond SL, Frost SD and Muse SV (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676-679. http://dx.doi.org/10.1093/bioinformatics/bti079

Qi XH, Zhang MF and Yang JH (2007). Molecular phylogeny of Chinese vegetable mustard (*Brassica juncea*) based on the internal transcribed spacers (ITS) of nuclear ribosomal DNA. *Genet. Resour. Crop Evol.* 54: 1709-1716. http://dx.doi.org/10.1007/s10722-006-9179-0

Qi XH, Yang JH and Zhang MF (2008). AFLP-based genetic diversity assessment among Chinese vegetable mustards (*Brassica juncea* L. Czern.). *Genet. Resour. Crop Evol.* 55: 705-711. http://dx.doi.org/10.1007/s10722-007-9278-6

Qiao AM, Liu PY and Lei JY (1998). RAPD analysis of sixteen varieties of mustard. *Acta Bot. Sin.* 40: 915-921.

Rambaut A and Drummond AJ (2007). Tracer v1.4. Available at: [http: //www. east. bio. ed. ac. Uk/Tracer].

Ronquist F, Teslenko M, van der Mark P, Ayres DL, et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61: 539-542. http://dx.doi.org/10.1093/sysbio/sys029

Song M, Liu T, Tang QL, Wang ZM, et al. (2009). Genetic diversity analysis of mustard germplasm based on RAPD and ISSR. *Acta Hortic. Sin.* 36: 835-842.

Swofford DL (2002). PAUP: phylogenetic analysis using parsimony (and other methods). Version 4.0b10. Sinauer Associates, Sunderland. MA.

Tajima F (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105: 437-460.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tong NK and Chen SR (1990). Analysis of acid phosphatase isozymes in *Brassica juncea* Coss. (L) and their original parental species. *Acta Hortic. Sin.* 17: 293-298.

Warwick SI, Guge RK, Mc-Donaland T and Falk KC (2006). Genetic variation of Ethiopian mustard (*Brassica carinata* A. Braun) germplasm in western Canada. *Genet. Resour. Crop Evol.* 53: 297-312. http://dx.doi.org/10.1007/s10722-004-6108-y

Watterson GA (1975). On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7: 256-276. http://dx.doi.org/10.1016/0040-5809(75)90020-9

Wu XM, Chen BY, Lu GY, Wang HZ, et al. (2009). Genetic diversity in oil and vegetable mustard (*Brassica juncea*) landraces revealed by SRAP markers. *Genet. Resour. Crop Evol.* 56: 1011-1022. http://dx.doi.org/10.1007/s10722-009-9420-8

Xu DM, Chen FB, Yao QL, Fang P, et al. (2014). Karyotypic analysis of three types of mustard (*Brassica juncea*). *J. Henan Agri. Sci.* 43: 111-115.

Yang JB, Tian X, Li DZ and Guo ZH (2003). Molecular composition and evolution of the chalcone synthase (CHS) gene family in five species of camellia (Theaceae). *Acta Bot. Sin.* 45: 659-666.

Yang YG, Liu NC, Chen XQ, Chen CL, et al. (1989). A study on classification of mustard. *Acta Hortic. Sin.* 16: 114-125.

Yao QL, Chen FB, Fang P, Zhou GF, et al. (2012). Genetic diversity of Chinese vegetable mustard (*Brassica juncea* Coss.) landraces based on SSR data. *Biochem. Syst. Ecol.* 45: 41-48. http://dx.doi.org/10.1016/j.bse.2012.05.006

Zhao B, Liu L, Tan DY and Wang JB (2010). Analysis of phylogenetic relationships of Brassicaceae species based on *Chs* sequences. *Biochem. Syst. Ecol.* 38: 731-739. http://dx.doi.org/10.1016/j.bse.2010.06.003