Evidence from Genome-wide Simple Sequence Repeat Markers for a Polyphyletic Origin and Secondary Centers of Genetic Diversity of *Brassica juncea* in China and India

Sheng Chen, Zhenjie Wan, Matthew N. Nelson, Jitendra S. Chauhan, Robert Redden, Wayne A. Burton, Ping Lin, Phillip A. Salisbury, Tingdong Fu, and Wallace A. Cowling

From the School of Plant Biology (M084) (Chen and Nelson) and The UWA Institute of Agriculture (Chen, Nelson, and Cowling), The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia; National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China (Wan and Fu); Directorate of Rapeseed-Mustard Research, Bharatpur, Rajasthan 321303, India (Chauhan); Department of Primary Industries, Horsham, VIC 3401, Australia (Redden, Burton, and Salisbury); Economic Crop Research Institute, Xinjiang Academy of Agricultural Sciences, Urumqi 830000, China (Lin); and the Faculty of Land and Food Resources, The University of Melbourne, Parkville, VIC 3010, Australia (Salisbury).

Address correspondence to Sheng Chen at the address above, or e-mail: sheng.chen@uwa.edu.au.

Abstract

The oilseed *Brassica juncea* is an important crop with a long history of cultivation in India and China. Previous studies have suggested a polyphyletic origin of *B. juncea* and more than one migration from the primary to secondary centers of diversity. We investigated molecular genetic diversity based on 99 simple sequence repeat markers in 119 oilseed *B. juncea* varieties from China, India, Europe, and Australia to test whether molecular differentiation follows Vavilov’s proposal of secondary centers of diversity in India and China. Two distinct groups were identified by markers in the A genome, and the same two groups were confirmed by markers in the B genome. Group 1 included accessions from central and western India, in addition to those from eastern China. Group 2 included accessions from central and western China, as well as those from northern and eastern India. European and Australian accessions were found only in Group 2. Chinese accessions had higher allelic diversity per accession (Group 1) and more private alleles per accession (Groups 1 and 2) than those from India. The marker data and geographic distribution of Groups 1 and 2 were consistent with two independent migrations of *B. juncea* from its center of origin in the Middle East and neighboring regions along trade routes to western China and northern India, followed by regional adaptation. Group 1 migrated further south and west in India, and further east in China, than Group 2. Group 2 showed diverse agroecological adaptation, with yellow-seeded spring-sown types in central and western China and brown-seeded autumn-sown types in India.

Key words: allelic diversity, crop evolution, Indian mustard, migration route, oriental mustard

*Brassica juncea*, commonly known as oriental mustard or Indian mustard, is an economically important oilseed crop widely cultivated in China, India, Europe, Canada, and Australia. Mustard has been cultivated in China for 6000 to 7000 years (Institute of Archaeology of Chinese Academy of Sciences 1963), where it is used as both an oilseed and vegetable crop. It is the predominant oilseed crop in India and has been an important component of Indian agriculture since the Indus Valley civilization flourished from 2300 to 1750 BC (Prakash and Hinata 1980). *Brassica juncea* has been used as the condiment mustard in Europe (Vaughan and Hemingway 1959) and was selected for canola quality recently in Canada and Australia (Woods et al. 1991; Burton et al. 2004).

Oilseed *B. juncea* types in India and China are morphologically distinct from one other and adapted to different environments. The Indian type is typically brown seeded and
adapted to autumn sowing in northern and central India, where winters are mild and dry. Chinese types include both yellow- and brown-seeded ones; some are adapted to spring sowing in elevated areas of central and western China, and some are winter hardy types with a strong vernalization requirement. Vavilov (1949) proposed that India and China are secondary centers of diversity of *B. juncea*, and this is supported by their distinct agroecological adaptation and seed morphology in these countries.

*Brassica juncea* (2n = 36, AABB) is an allotetraploid derived from interspecific hybridization between *B. rapa* (2n = 20, AA) and *B. nigra* (2n = 16, BB) (U 1935). Burkill (1930), Olsson (1960), Vaughan et al. (1963), and Sun (1970) hold the view that *B. juncea* originated from several independent hybridizations between wild *B. rapa* and *B. nigra* at multiple places in the Middle East and neighboring regions where their distributions overlapped. Prakash and Himata (1980) also considered that several independent polyploidization events occurred over the past several thousand years and that true wild types of *B. juncea* exist in Asia Minor. Its allotetraploid origins are well demonstrated by studies in taxonomy, artificial syntheses, molecular analysis, and chromosome mapping (Axelsson et al. 2009; Prakash et al. 2009).

Prain (1898) proposed that the oilseed mustard “Rai” migrated into India from China through a northeastern route because a wild form of Rai was far more common along this route than in the plains of India. Sinskaja (1928) supported Prain’s view and proposed that *B. juncea* var. *sareptana*, a wild type in the Kirgiz steppes and in China, was the “basic type” from which all other *B. juncea* varieties evolved and claimed that East European *B. juncea* was of Chinese origin, from where it migrated naturally through the Kirgiz steppes to Europe. However, Burkill (1930) disagreed and suggested that the Middle East was its place of origin. This was supported by Sun (1970) because the two parent species, *B. nigra* and *B. rapa*, were not found as wild species in China. Vavilov (1949) proposed that Afghanistan and its adjoining regions were the primary centers of origin of *B. juncea* and considered Asia Minor, central/western China, and eastern India as secondary centers of diversity. However, the cytogenetic and biochemical evidence points to a polyphyletic origin (Olsson 1960; Prakash 1973; Gómez-Campo and Prakash 1999). With strong biochemical evidence, Vaughan et al. (1963) and Vaughan (1977) proposed two geographic races of oilseed *B. juncea*—the Indian race and the Oriental race—which were closer to the *B. rapa* and *B. nigra* progenitors, respectively.

The genetic diversity of oilseed *B. juncea* in India and China has been studied using molecular markers. Random amplified polymorphic DNA (RAPD) assays of Indian mustard detected a low level of genetic variability among 12 Indian *B. juncea* accessions compared with 11 exotic accessions (Jain et al. 1994). Two groups of Indian mustard were identified by Khan et al. (2008) based on RAPD markers. Amplified fragment length polymorphism (AFLP) markers showed clustering of the Indian and Chinese lines, whereas other lines from Australia, Canada, Eastern Europe, and Russia formed another group (Srivastava et al. 2001). High diversity of *B. juncea* was discovered in southwestern and western China by RAPD analysis (Li et al. 1997; An et al. 1999). Grouping of 73 *B. juncea* landraces from southwestern China was related to agroecological adaptations (Pu et al. 2007). The molecular genetic diversity of 101 accessions from western China, analyzed by AFLP, sequence-related amplified polymorphism (SRAP), and simple sequence repeat (SSR) markers, was related to geological and biological conditions at collection sites (Xu et al. 2008). Genetic diversity in winter types of *B. juncea* was higher than that in spring types and genetic diversity of *B. juncea* in Shaanxi and Xinjiang provinces was more abundant than that among the genetically distinct types in Tibet (Xu et al. 2008). Using SRAP markers, Wu et al. (2009) assessed 67 oilseed *B. juncea* accessions from China and 10 accessions from other countries, together with 18 non-oilseed accessions, and found that the oilseed mustards were divided into different groups based on their growth habit (spring or winter type).

In this study, genetic diversity in a diverse, global collection of oilseed *B. juncea* was investigated with SSR markers distributed across the A and B genomes. The objective of the research was to assess molecular genetic differentiation in Vavilov’s proposed secondary centers of diversity in China and India. Clustering of molecular groups based on SSR alleles specific to the A or the B genome may shed light on variation in parental genomes at the time of polyploidization and the potential migrant groups of *B. juncea* from the primary to the secondary centers of diversity.

### Materials and Methods

#### Plant Materials

A total of 123 *B. juncea* accessions (landraces, cultivars, and breeding lines) formed the basis of this study, including 53 from China (CN), 46 from India (IN), 12 from Australia (AU), and 12 from Europe (EU; see Supplementary Table S1 online). Four of the *B. juncea* accessions were subsequently shown to be other *Brassica* species misclassified or mislabeled as *B. juncea* (see Results) and were removed from the analysis. Accessions from CN and IN were chosen to cover the widest possible geographic distribution and the longest possible breeding history. The samples originated from 13 Chinese Provinces (date of release or collection: from 1950 to 2005) and eight Indian States (date of release or collection: from 1948 to 2006). Five related *Brassica* species in U’s triangle (U 1935) were also used as genetic diversity outgroups and to identify accessions that were possibly misclassified as *B. juncea*.

#### DNA Extraction and SSR Analysis

Young leaves were collected from 2-week-old seedlings and the genomic DNA was extracted from a single individual of each accession using the method described previously (Chen et al. 2008). A total of 99 SSR primer pairs amplifying clear, reproducible, and polymorphic products were used for the genetic diversity study, of which 51 and 48 were originally designed to target the A and C genomes of *B. napus* (Wang et al. 2008).
et al. 2011) and the B genome of *B. juncea*, respectively (Axelsson et al. 2000). The primer sequences were kindly provided by the Agriculture and AgriFood Canada Saskatoon Research Centre, Saskatoon (Isobel Parkin, personal communication; for more information, see http://aafc-aac.usask.ca/BrassicaMAST/). Polymerase chain reaction was performed with a volume of 15 μL containing 7.5 μL of 2× AmpliTaq Gold Master Mix (Applied Biosystems, Foster City, USA), 1.0 μL of 5 μM of each SSR primer pair, and 25–50 ng of template DNA following the methods in Chen et al. (2010). Six to eight SSR markers were pooled together for analysis on an AB3730xl capillary sequencer (Applied Biosystems) according to complementary primer dye types and amplicon size range. The electrophoresis data were then processed using GeneMapper, version 3.7 (Applied Biosystems).

Determination of Allele Location on A or B Genome in *B. juncea*

In most cases, it was possible to allocate alleles to the locus position on the A genome, as indicated by Wang et al. (2011) and on the B-genome as provided by I. Parkin (personal communication). B-genome linkage group names followed the convention of Lagercrantz and Lydiate (1995). If a pair of SSR primers amplified two or more loci in the *B. juncea* genome, the location of alleles was judged by their presence or absence in other *Brassica* species, including the five control samples in this study, and by information in previous studies on 95 *B. napus* lines (Chen et al. 2010), 374 *B. rapa* lines (unpublished data of the authors), 44 *B. nigra* accessions (Pradhan et al. 2011), and 2 *B. carinata* lines (Nelson et al. 2009; Mason et al. 2011). For example, if an allele was present in *B. rapa* and/or *B. napus*, but not in *B. nigra* or *B. carinata*, this allele was designated as an A-genome allele. If an allele was present in *B. nigra* and/or *B. carinata*, but not in *B. rapa* or *B. napus*, this allele was designated as a B-genome allele.

Population Structure Analysis

The program STRUCTURE, version 2.3.3, was used to identify subpopulations in the population using a Bayesian approach (Pritchard et al. 2000). When applying STRUCTURE, we tested several numbers of subgroups (K) from 2–8, assuming that genotypes have independent allele frequencies between subpopulations. Each process was run using 100 000 iterations, with burn in of 100 000.

Multivariate Analysis

Allele frequencies were calculated using NTYSYSpc, version 2.21c (Rohlf 2006). Dissimilarity matrices were calculated as suggested by Nei and Li (1979) and were based on Euclidean distance. These matrices were subjected to hierarchical cluster analysis using the unweighted pair group method and arithmetic averages (UPGMA), as proposed by Sneath and Sokal (1973), and ordination by 2D multidimensional scaling (2D-MDS) in PRIMER 6 software (Clarke and Gorley 2006). The nonparametric analysis of molecular variance (AMOVA) was conducted using GenAlEx, version 6.3 (Peakall and Smouse 2006) to partition the total variation and test significance among and within populations. A set of null distributions was used to select 999 random samples from the raw distance matrix in order to generate significance values for the variance components.

Analysis of Common and Private SSR Alleles

*Brassica juncea* populations from AU, CN, EU, and IN were assessed for common and private SSR alleles as described previously (Chen et al. 2008; 2010). Saltkin and Takahata (1985) define common alleles as those that are shared between two or more populations or countries, whereas private alleles are unique to one population or country.

Geographic Distribution of Allelic Diversity

Geographic distribution of allelic diversity of Chinese and Indian *B. juncea* accessions was analyzed according to the longitude and latitude of the capital city of each State or Province from where the accession was collected. Many of the varieties were bred in State- or Province-based institutions in or close to these cities. Shannon diversity index was calculated using GenAlEx, version 6.3 (http://biology.anu.edu.au/GenAlEx/Welcome.html), based on the article by Sherwin et al. (2006).

Results

Allelic Diversity and Genetic Distance Analysis of *B. juncea*

A total of 650 alleles were amplified using 99 pairs of SSR primers from 128 DNA samples, including the original collection of 123 putative *B. juncea* accessions and 5 accessions from other *Brassica* species. Hierarchical clustering analysis of Nei's matrix distances among the 128 samples showed that three accessions (CN19, CN20, and CN22 [see Supplementary Table S1 online]) clustered with *B. carinata* at the dissimilarity level of 0.33 and that one accession (IN24) was grouped with *B. carinata* at the dissimilarity level of 0.39, whereas the other 119 accessions were grouped together at the genetic dissimilarity level of 0.44 and were distant from the 5 outgroup accessions (Figure 1). These results were confirmed using 2D-MDS analysis (see Supplementary Figure S1 online).IN24 was reclassified as *B. carinata*. Further cytogenetic analysis showed that CN19, CN20, and CN22 had 20 chromosomes (2n) and these were reclassified as *B. rapa*. On average, 6.6 alleles were amplified from each SSR primer pair, with an average of 36 alleles per *B. juncea* chromosome in the 128 DNA samples. Following removal of the four misclassified accessions, 119 accessions were used in the genetic diversity analysis of *B. juncea*.

A total of 500 alleles were recorded from 119 *B. juncea* accessions: 252 in the A genome (at 63 loci) and 248 in the B genome (at 49 loci), with an average of 4.0 alleles per locus on the A genome and 5.1 alleles per locus on the B genome (see Supplementary Table S2 online). The 2D-MDS analysis, based
Figure 1. Hierarchical clustering analysis of Nei's matrix distances among 128 samples of *B. juncea* and other *Brassica* species. In this study, 123 accessions were described as *B. juncea* from Australia (upright triangles), China (inverted triangles), Europe (squares), and India (rhombuses) and five accessions represented five other *Brassica* species (asterisks). The cluster analysis was based on 650 alleles from 99 pairs of SSR primers, including 51 from the A genome and 48 from the B genome. AA, *B. rapa*; BB, *B. nigra*; CC, *B. oleracea*; AACC, *B. napus*; and BBCC, *B. carinata*.
on 500 alleles from both the A and the B genomes (Figure 2), showed that at the genetic dissimilarity level of 0.50, all *B. juncea* accessions were grouped together, and at the genetic dissimilarity level of 0.47, they were divided into two major groups. Group 1 included 8 CN accessions and 23 IN accessions, and Group 2 contained 88 *B. juncea* accessions from IN, CN, EU, and AU. At the genetic dissimilarity level of 0.45, Group 2 was further divided into three subgroups: the first with 16 IN accessions; the second with 3 AU, 20 CN, and 6 IN accessions; and the third with 9 AU, 22 CN, and 12 EU accessions.

The 2D-MDS analysis was repeated, first based on 252 polymorphic A-genome alleles (see Supplementary Figure S2A online), and next based on 248 polymorphic B-genome alleles (see Supplementary Figure S2B online). In both analyses, the 119 *B. juncea* accessions were divided into the same two major groups. This increased confidence in assigning accessions to groups based on the combined analysis of A- and B-genome alleles (Figure 2), with one exception—accession AU09 was outside the two major groups (Group 1 and 2) when the analysis was based on the 252 polymorphic A-genome alleles (see Supplementary Figure S2A online). Based on the A genome, Group 2 was divided into five overlapping subgroups (see Supplementary Figure S2A online), but based on the B genome, Group 2 was divided into three separate subgroups (see Supplementary Figure S2B online).

The divisions revealed by the 2D-MDS analysis were supported by STRUCTURE analysis. In Bayesian assignment tests, the log probability reached a plateau around $k = 4-6$, suggesting that the most likely number of clusters was within that range (Figure 3). The first split ($k = 2$) was between Groups 1 and 2. This division remained for all values of $k$, suggesting very little admixture between Group 1 and Group 2. There was clearly more genetic structure within Group 2, with CN accessions separated from IN accessions (at $k = 3$) and the CN accessions were further subdivided (at $k = 4$) with little admixture. The remaining clusters (defined at $k = 5$ and 6) were less clear (Figure 3).

In order to evaluate the possible allelic differentiation in A and B genomes of *B. juncea*, common and private SSR alleles among the 119 *B. juncea* accessions from AU, CN, EU, and IN were analyzed in the A and B genomes separately (Figure 4). There were 148/500 (29.6%) common alleles in all of these four populations, of which 83 were in the A genome and 65 in the B genome. AU and EU accessions shared the same number of private alleles (seven) in both A and B genomes. CN and IN accessions shared the same number of private alleles (27) in the A genome, whereas in the B genome, CN had 42 private alleles but IN had 20 (Figure 4).

The allelic diversity per accession was higher in CN than in IN in both Groups 1 and 2, as well as in both the A and B genomes. Averaged across both genomes,
the frequency of private SSR alleles was 1.80, 1.53, and 1.99 times higher in CN than in IN in Group 1, Group 2, and the combined groups, respectively (Table 1).

Molecular Variation and Geographic Distribution of *B. juncea*

AMOVA of the 95 CN and IN accessions was based on 237 A-genome alleles and 232 B-genome alleles (Table 2). The proportions of the variance attributed to differences between CN and IN were only 13.6% in the A genome, 21.1% in the B genome, and 17.3% if considering both A and B genomes. Most of the variance occurred within CN or IN populations.

One-third of the variance was attributed to differences between Groups 1 and 2 whether based on the A/B genome or the combined data (Table 2). The variance explained by differences between accessions from Group 1 and Group 2 was significantly higher than that explained by differences between accessions from China and India. This supported the results of 2D-MDS analysis that grouping should occur on the basis of molecular genetic results rather than on country of origin.

Passport records on Province (CN) or State (IN) of origin were available for 90 *B. juncea* accessions (see Supplementary Table S1 online). This was used to calculate the Shannon diversity index, which averaged 0.194 over all locations. A genetic diversity hotspot was identified in central and western China, including Yunnan, Qinghai, and Gansu (25–36°N, 101–103°E), with a diversity index of 0.224 based on 17 accessions in this region (Figure 5). There was a clear geographic clustering of accessions in IN and CN within groups. Group 1 included accessions from central and western India (Rajasthan, Gujarat, Madhya Pradesh, Delhi, and Uttar Pradesh) and from eastern China (Guangdong, Beijing, Shanghai, Inner Mongolia, and Shanxi). Group 2 included accessions from central and western China (Inner Mongolia, Shaanxi, Guizhou, Shaanxi, Qinghai, Yunnan, Tibet, Xinjiang, Gansu, and Hubei) and from northern and eastern India (Delhi, Uttar Pradesh, West Bengal, Punjab, and Haryana; Figure 6).

Discussion

Allelic diversity of oilseed *B. juncea* was evaluated using SSR markers distributed across the A and B genomes of 119 landraces, cultivars, and breeding lines from India, China, Europe, and Australia. Two major SSR-defined groups of
B. juncea were formed in hierarchical clustering and multidimensional scaling analysis. Groups 1 and 2 both contained accessions from IN and CN, whereas EU and AU accessions were found only in Group 2 (Figure 2, see Supplementary Table S1 online). Allocation of accessions to groups was consistent across the A and B genomes. Allelic diversity per accession was higher in CN than in IN for Group 1 accessions, but the opposite trend occurred in Group 2 accessions. The richness (frequency) of private SSR alleles was higher in CN than in IN in both Groups 1 and 2, particularly in the B genome (Table 1). The Shannon diversity index, based on analysis of geographic distribution, also revealed a diversity hotspot in central and western China (Figure 5).

SSR allelic diversity in B. juncea in this study (an average of 4.0 alleles per locus on the A genome and 5.1 alleles per locus on the B genome, with an average of 4.5 alleles per locus across the whole genome) is similar to the level of SSR allelic diversity identified previously in B. napus, whereby there were 4.3 alleles per locus in 96 European genotypes (Hasan et al. 2006) and 3.0 alleles per locus in 72 accessions from IN, CN, EU, and AU (Chen et al. 2008).

Our results provide a unique insight into the evolution of this allotetraploid agricultural species. For the first time, we describe two distinct SSR-marker-defined groups of oilseed B. juncea in India and China—Group 1 includes accessions from 1) central and western India and 2) eastern China, and Group 2 includes accessions from 1) central and western China and 2) northern and eastern India. These groups were clearly distinct in separate analyses of the A and B genomes. One explanation for these observations was that B. juncea was formed in at least two independent polyploidization events outside of China and India and later migrated to these countries, as proposed by earlier authors (Burkill 1930; Vavilov 1949; Sun 1970; Prakash and Hinata 1980). Alternatively, isolation of two main pools (Group 1 and Group 2) occurred after a single polyploidization event. In either case, it appears that Groups 1 and 2 migrated independently and Group 1 went further south and west in India and further east in China than Group 2.

Our collection of landraces and cultivars from India and China was based on oilseed B. juncea varieties released by institutional breeding programs during the 20th century. The presence of two distinct SSR-based groups in each country indicates very low rates of interbreeding of B. juncea among regional breeding programs within countries, or between countries, before the release of these varieties. Local selection and isolation of breeding programs in India and China over many years explains why CN accessions are mostly yellow and IN accessions are all brown (see Supplementary Table S1 online; Vaughan 1977). Indian B. juncea is adapted to sowing...
in autumn and flowering under short days, whereas many CN accessions are spring sown or winter hardy types adapted to flowering in spring and summer. A range of seed colors and flowering types have been selected in both Groups 1 and 2 because flowering time and seed color are controlled by relatively few quantitative trait loci of large effect in Brassica species (Badani et al. 2006; Fu et al. 2007; Long et al. 2007).

Our results support a polyphyletic origin of B. juncea, as suggested by earlier authors (Burkill 1930; Olsson 1960; Vaughan et al. 1963; Sun 1970; Prakash and Hinata 1980), and are consistent with the proposition by Vavilov (1949) that the primary center of diversity of B. juncea lies in West Asia, such as Afghanistan and adjoining regions, and that India and China are secondary centers of diversity. The two SSR-marker-defined groups may have been separated from each other during thousands of years of cultivation and selection of B. juncea in India and China. Groups 1 and 2 may have originated as unique polyploidization events because they were differentiated equally from one another in both the A and B genomes. Investigation of cytoplasmic DNA markers (Pradhan et al. 1992; Warwick and Black 1997) would help resolve the question about the proposed polyphyletic origin of B. juncea.

### Table 1
Allelic diversity (number and average per accession) of Chinese (CN) and Indian (IN) oilseed B. juncea in Groups 1 and 2, based on total and private alleles

<table>
<thead>
<tr>
<th>Group</th>
<th>Country of origin</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Groups 1 + 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CN</td>
<td>IN</td>
<td>Subtotal</td>
<td>CN</td>
</tr>
<tr>
<td>I. Total alleles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A genome</td>
<td>Number</td>
<td>83</td>
<td>112</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>10.4</td>
<td>4.9</td>
<td>4.1</td>
</tr>
<tr>
<td>B genome</td>
<td>Number</td>
<td>92</td>
<td>117</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>11.5</td>
<td>5.1</td>
<td>3.9</td>
</tr>
<tr>
<td>A + B genomes</td>
<td>Number</td>
<td>175</td>
<td>229</td>
<td>248</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>21.9</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>II. Private alleles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A genome</td>
<td>Number</td>
<td>10</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>1.3</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>B genome</td>
<td>Number</td>
<td>5</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>A + B genomes</td>
<td>Number</td>
<td>15</td>
<td>24</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Average per accession</td>
<td>1.9</td>
<td>1</td>
<td>1.3</td>
</tr>
</tbody>
</table>

### Table 2
AMOVA for 95 Chinese and Indian oilseed B. juncea based on 237 A-genome alleles and 232 B-genome alleles with the populations defined by two countries of origin (I) and two groups according to 2D-MDS analysis results (II)

<table>
<thead>
<tr>
<th>Genome</th>
<th>Source of variation</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>Variance component</th>
<th>Total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Populations were defined as two countries of origin: CN and IN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Between CN and IN</td>
<td>1</td>
<td>209,225</td>
<td>209,225</td>
<td>3.893</td>
<td>13.6%</td>
</tr>
<tr>
<td></td>
<td>Within CN or IN</td>
<td>93</td>
<td>2306,838</td>
<td>24,805</td>
<td>24,805</td>
<td>86.4%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>2516,063</td>
<td>234,030</td>
<td>28,698</td>
<td>100.0%</td>
</tr>
<tr>
<td>B</td>
<td>Between CN and IN</td>
<td>1</td>
<td>303,318</td>
<td>303,318</td>
<td>5.935</td>
<td>21.1%</td>
</tr>
<tr>
<td></td>
<td>Within CN or IN</td>
<td>93</td>
<td>2063,398</td>
<td>22,187</td>
<td>22,187</td>
<td>78.9%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>2366,716</td>
<td>325,505</td>
<td>28,122</td>
<td>100.0%</td>
</tr>
<tr>
<td>A + B</td>
<td>Between CN and IN</td>
<td>1</td>
<td>512,543</td>
<td>512,543</td>
<td>9.828</td>
<td>17.3%</td>
</tr>
<tr>
<td></td>
<td>Within CN or IN</td>
<td>93</td>
<td>4370,236</td>
<td>46,992</td>
<td>46,992</td>
<td>82.7%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>4882,779</td>
<td>559,535</td>
<td>56,820</td>
<td>100.0%</td>
</tr>
<tr>
<td>II. Populations were defined as two groups: Group 1 and Group 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Between groups</td>
<td>1</td>
<td>495,922</td>
<td>495,922</td>
<td>11.353</td>
<td>34.3%</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>93</td>
<td>2020,141</td>
<td>21,722</td>
<td>21,722</td>
<td>65.7%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>2516,063</td>
<td>0,000</td>
<td>33,075</td>
<td>100.0%</td>
</tr>
<tr>
<td>B</td>
<td>Between groups</td>
<td>1</td>
<td>431,519</td>
<td>431,519</td>
<td>9.833</td>
<td>32.1%</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>93</td>
<td>1935,197</td>
<td>20,809</td>
<td>20,809</td>
<td>67.9%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>2366,716</td>
<td>0,000</td>
<td>30,642</td>
<td>100.0%</td>
</tr>
<tr>
<td>A + B</td>
<td>Between groups</td>
<td>1</td>
<td>927,441</td>
<td>927,441</td>
<td>21,186</td>
<td>33.3%</td>
</tr>
<tr>
<td></td>
<td>Within groups</td>
<td>93</td>
<td>3955,338</td>
<td>42,531</td>
<td>42,531</td>
<td>66.7%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>94</td>
<td>4882,779</td>
<td>0,000</td>
<td>63,717</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

df, degrees of freedom; SS, sum of squares; MS, mean square.
It was proposed that \textit{B. juncea} was transported to India from China in ancient times (Vaughan et al. 1963; Prakash and Hinata 1980). Our results support this proposition, based on the higher frequency (richness) of private alleles in China than in India in Group 1. Group 2 showed higher allelic diversity per accession in China than in India and included all samples from southwest China, where a diversity hotspot was found based on the Shannon diversity index (Figure 5).

The 12 AU and 12 EU accessions used in this study were only found in Group 2, where they were aligned more with CN than IN accessions (Figure 2), and they also shared the yellow seed color of \textit{CN} \textit{B. juncea} (see Supplementary Table S1 online). However, 31/500 (6.2\%) alleles were found in AU and/or EU accessions but not in CN or IN accessions based on SSR allele distinctness (Figure 4). This may be due to the low sample size or novel alleles may have been selected in AU and EU accessions after introduction to these countries.

In this study, alleles on the B genome of \textit{B. juncea} were detected at a total of 41 loci, with four to six loci on each chromosome (see Supplementary Figure S3 online). The frequency of private alleles on chromosomes varied from 15.7\% (J01) to 47.2\% (J10), with an average of 28.0\% on the \textit{A} genome and 24.3\% on the \textit{B} genome, indicating that there was no obvious clustering of private alleles on any chromosome. Apparently, there have been no recent introgressions of private alleles from related diploid species such as \textit{B. rapa} (\textit{A} genome) or \textit{B. nigra} (\textit{B} genome), which would be clustered if these introgressions were large.

SSR markers were useful to correct the misclassification of some accessions in this study. We received a total of 123 \textit{B. juncea} accessions from gene banks and other sources, but four accessions were proved not to be \textit{B. juncea} by SSR-based hierarchical clustering analysis. Three CN accessions (CN19, CN20, and CN22) were grouped with \textit{B. rapa} (AA), and one IN accession (IN24) was grouped with \textit{B. carinata} (BBCC). SSR markers were useful to verify the relationship of lines with ambiguous pedigree records. For example, Haoyou 11 (CN15) was reportedly derived from an introduced line and released after domestication and adaptation tests in Qinghai Province, in the Northwest of China (Zhongyuan Li, personal communication). However, there is no record of the geographic location from which this introduced line was obtained. Our molecular marker clustering result showed that Haoyou 11 is distant from all other 49 CN accessions, but it clustered with 8 out of 12 EU accessions tested in study. This molecular evidence suggested that Haoyou 11 was probably derived from a European introduction to China. Moreover, Haoyou 11 showed only 17.3\% genetic dissimilarity with Rk2 (EU08), the lowest dissimilarity among 119 \textit{B. juncea} accessions from among all four populations. We conclude that Haoyou 11 was most likely derived from the European line Rk2 or a close relative. Locus-specific SSR markers are valuable to confirm species identity in \textit{Brassica
germplasm collections and recently were used to identify extensive misclassification of *B. nigra* accessions (Pradhan et al. 2011).

These results are of great importance to oilseed *B. juncea* breeders who seek to increase genetic diversity in their breeding programs. Molecular markers are useful to assist the broadening of genetic diversity for sustainable crop genetic improvement (Chen and Zhang 2000). SSR markers were used to evaluate the genetic diversity and genetic distinctiveness of oilseed rape (*B. napus*) from different countries (Zhou et al. 2006; Chen et al. 2008). Restricted fragment length polymorphism (RFLP) markers were used to demonstrate the distinction between winter and spring breeding pools of oilseed rape (*B. napus*) and the discovery of new genetic diversity introduced from resynthesized oilseed rape (Becker et al. 1995). Similar benefits arise from SSR markers, which were used to identify novel genetic diversity in *B. napus* introduced through interspecific hybridization with closely related species (Chen et al. 2010; Zou et al. 2010). As revealed in this study, two SSR-marker-defined groups of oilseed *B. juncea* exist in both China and India, and there appears to have been little mixing of the groups within each country. Oilseed *B. juncea* breeders in China and India can make use of *B. juncea* genetic diversity by making crosses between the two different groups that exist inside each country.

**Conclusions**

We found two distinct SSR-defined groups in *B. juncea*, Group 1 and Group 2, which occurred in overlapping regions of India and China. Our data support the proposal of Vavilov that China and India are secondary centers of diversity of *B. juncea*, but these also provide new evidence that Groups 1 and 2 have been genetically isolated within China and India since they were introduced. Our analysis was based on breeding lines, cultivars, and landraces of *B. juncea*, many of which were bred and released by regional research stations in India.

**Figure 6.** Geographic clustering of oilseed *B. juncea* accessions in China and India within groups. Chinese accessions in Group 1 (circles) are mostly collected from a narrow range of eastern China along the coastal line, whereas Chinese accessions in Group 2 (triangles) are collected from a wide range of “western” China. Indian accessions in Group 1 are mostly collected from western and central India, whereas India accessions in Group 2 are mostly from northern and eastern India. The line of dashes indicates the separation of Groups 1 and 2 in China and India.
and China many years after Vavilov’s collections, but the patterns are consistent with Vavilov’s proposals. Within Groups 1 and 2, breeders have selected similar phenologies—for example, all Indian varieties are brown seeded and adapted to autumn sowing, whereas many Chinese varieties are yellow seeded and adapted to spring sowing. Groups 1 and 2 were equally differentiated by SSR markers in the A or B genome. This is consistent with the hypothesis that B. juncea had a polyphyletic origin prior to migrations to China and India from Vavilov’s proposed center of origin in middle Asia.

Supplementary Materials

Supplementary Tables S1 and S2 and Figures S1, S2, and S3 can be found at http://www.jhered.oxfordjournals.org/.

Funding

Australian Centre for International Agricultural Research (ACIAR) and Grains Research and Development Corporation (GRDC) (CIM/1999/072).

Acknowledgments

W.A.C. and M.N. acknowledge partial support of their salaries by Canola Breeders Western Australia Pty Ltd during the course of this research. We would like to thank Dr Jacqueline Batley at The University of Queensland for critical reading of this manuscript.

References


Received October 22, 2012; First decision November 15, 2012; Accepted February 15, 2013

**Corresponding Editor:** Brian Murray