TÜBİTAK

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2015) 39: 403-412

doi:10.3906/tar-1405-50

Cytoplasmic-nuclear variation in a diversity-fixed foundation set of *Brassica juncea* (L.) Czern & Coss.

Javed AKHATAR, Nitin KUMAR, Chhaya ATRI, Surinder Singh BANGA*
Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India

Received: 14.05.2014 • **Accepted:** 26.09.2014 • **Published Online:** 12.06.2015 • **Printed:** 30.06.2015

Abstract: A core subset of genetically fixed lines was selected through stratified sampling to represent the allelic variations within *Brassica juncea* gene pools. It included germplasm accessions from India, China, and Australia, as well as *B. juncea* introgression lines carrying alien genetic variation and/or cytoplasm from wild crucifers, namely *Diplotaxis cardaminoides, Diplotaxis tenuisiliqua*, and *Erucastrum abyssinicum*. The aim of the study was to infer the influence of underlying cytoplasm on population structure and trait variation. Molecular characterization using nuclear SSR markers (158) and chloroplast SSR markers (9) revealed interesting germplasm patterns and trait variations. Diversity groups obtained using distance-based clustering were not fully compliant with structure analysis, especially for the nuclear genetic variation. Association between population structure and phenotypic variation was indicated by significant interactions between chlorotype and seed/biological yields. Thus, there may be a need to factor in the impact of background cytoplasm on population structure in association studies of crops where different cytoplasmic lineages are known.

Key words: Cytoplasmic lineage, gene pool, germplasm, Indian mustard, phylogenetic tree

1. Introduction

Brassica is an agriculturally important genus comprising species with a range of morphologies and food utilities. Oilseed forms include *Brassica napus*, *Brassica juncea*, *Brassica carinata*, and 3 ecotypes of *Brassica rapa*. These provide the third most important source of vegetable oil in the world after soybean and palm oil. Winter *B. napus* is widespread in Europe and parts of China, while spring forms of *B. napus* are cultivated in Canada and Australia. Spring *B. juncea* is the major oilseed crop of the Indian subcontinent. Both spring and winter types of *B. juncea* occupied large areas in China until their replacement by the more productive *B. napus* during the late 1950s. This crop also has areas of adoption in East Europe and Russia. The amphiploid *B. juncea* (2n = 36, AABB) arose several times from independent natural hybridization events between *B. rapa* (2n = 20, AA) and *B. nigra* (2n = 16, BB). Such events were earlier considered only unidirectional, with *B. rapa* as the cytoplasm donor parent (Banga et al., 1983; Erickson et al., 1983). However, it was recently established that most East European accessions of *B. juncea* carry *B. nigra* cytoplasm. Furthermore, 2 forms of *B. juncea* (with *B. rapa* or *B. nigra* cytoplasms) coexist in both India and China. For credible genetic research, diversity available in germplasm lines can be more useful if it is fixed as informative sets of genetically fixed lines representing a structured sampling of the variability in the form of fixed foundation sets (*Bj*DFFS). A core subset of 48 genotypes was selected to represent the majority of allelic variations within *B. juncea* gene pools for possible use as an association mapping panel. Population structure is generally confounded by the lack of random mating in largely self-pollinated crops like *B. juncea*. Factoring in type I and type II errors is crucial for removing the confounding effect of population structure in the association panel (Zhu et al., 2008). Genetic markers are mostly used to infer population structure. In plants, these markers can be located in either the nuclear or cytoplasmic genomes and can be used to decipher population structure. However, like most angiosperms, the cytoplasmic genomes in *B. juncea* are maternally inherited, whereas the nuclear genome is biparentally inherited. Thus, there is asymmetric gene flow due to contrasted pollen and seed dispersal capacities under natural conditions (Hamilton and Miller, 2002). Due to seed dispersal, cytoplasmic polymorphism is expected to be more pronounced than spatial patterns depicted by nuclear genetic markers with bimodal inheritance. Aside from the population genetic implications of this genetic asymmetry, there are adaptive repercussions of cytoplasm–nuclear interactions.

* Correspondence: nppbg@pau.edu

This becomes very important in crops where germplasm lines carrying different cytoplasm types coexist. This study incorporates information about the influence of underlying cytoplasm and nuclear genetic diversity on population structure and trait variation. Forty-eight fixed lines were genotyped using nuclear markers to quantify allelic variation within the diversity set. Chloroplast SSR (cpSSR) markers were also used to analyze the background natural or alien cytoplasm.

2. Materials and methods

2.1. Plant materials

Forty-eight pure-breeding genotypes were selected from a previously sorted *B. juncea* collection of diversity-fixed genotypes. Selection of genotypes was based on phenological data, geographic diversity, and possible genetic and cytoplasmic variability. This collection included representations from India, China, and Australia, as well as *B. juncea* lines carrying genetic introgression and/or cytoplasm from wild crucifers like *Diplotaxis cardaminoides*, *D. tenuisiliqua*, and *Erucastrum abyssinicum*. Procedures for synthesis of these wide hybrids and alien cytoplasmic substitution lines were described earlier (Banga et al., 2003; Chandra et al., 2004). Identified genotypes were advanced by single seed descent method through selfing over 7 generations. This set was considered representative with regard to the variation of morphophysiological characters and will be referred to in this study as the Diversity-Fixed Foundation Set (*Bj*DFFS). Each genotype was first tested for the expected euploid chromosome number for *B. juncea* (2n = 36).

2.2. DNA extraction

DNA was extracted from the first true leaf of a single plant from each accession by the modification of the cetyltrimethylammonium bromide (CTAB) extraction procedure of Doyle and Doyle (1990). Modified CTAB extraction buffer had a high NaCl concentration to remove polysaccharides (Lodhi et al., 1994) and polyvinylpyrrolidone to remove polyphenols (Maliyakal, 1992). DNA was suspended in TE buffer (pH 8) and digested with RNase A at 37 °C for 1 h. DNA quality and quantity were checked by spectrophotometric analysis and it was diluted to 5 ng/μL before conducting PCR.

2.3. Nuclear SSR and cpSSR selection and genotyping

A total of 158 polymorphic nuclear SSRs markers, representing all the *B. juncea* chromosomes, and 9 cpSSR markers were used. These included 73 SSRs for the A genome (Kim et al., 2009), 48 SSR markers for the B genome, and 64 SSRs constructed from genomic libraries of *B. nigra* (Lowe et al., 2004). In addition, 9 chloroplast-specific SSR primers were used to document plastid variability in the diverse *Brassica* gene pool (Flannery et al., 2006). The SSR polymorphism survey, genotyping, and scoring were conducted following Kim et al. (2009). DNA amplification for SSR markers was carried out in a final volume of 10.33 μL containing 1 U of Taq polymerase, 2 μM of primers, 400 μM of dNTPs, 1X PCR buffer with $MgCl_2$, and 15 ng of genomic DNA as templates. Reactions were performed in a 384-welled PCR plate from Applied Biosystems (model EN61328). The profile used was a hot start for 5 min at 94 °C and 35 amplification cycles of 30 s (DNA denaturation) at 94 °C, annealing at optimum Ta for 30 s, and 1 min at 72 °C, followed by a final extension step of 7 min at 72 °C. The success of the PCR reaction was verified by an automated high-throughput electrophoresis system (Caliper Lab Chip GX version 3.0.618.0).

2.4. Analysis of population structure

The genetic structure of the *Bj*DFFS was analyzed using STRUCTURE 2.1 software (Pritchard et al., 2000; Falush et al., 2003). Although STRUCTURE assumes Hardy–Weinberg equilibrium and linkage equilibrium within populations, it can also be applied to partially inbred genotypes by randomly choosing one allele. STRUCTURE sorts individuals into K clusters according to their genetic similarity. The best K is chosen based on the estimated membership coefficients (Q) for all individuals in each cluster. The number of subpopulations (K) was determined using the admixture model ranging from K = 1 to K = 10, a burn-in period of 1,000,000 and a Markov chain Monte Carlo model with 1,000,000 repetitions (Pritchard et al., 2010). This helped to identify subpopulations in both nuclear SSRs and cpSSRs. The admixture model was applied and no prior population information was used. STRUCTURE Harvester (Dent and Von Holdt, 2012) was employed to visualize ΔK, a modification of K suggested by Evanno et al. (2005). CLUMPP 1.1 (Jakobsson and Rosenberg, 2007) software was used to find optimal alignments of independent runs.

2.5. Phylogenetic tree

The phylogenetic tree was constructed separately by nuclear SSRs and chloroplast SSRs between 48 accessions (calculated from 158 SSRs and 9 cpSSRs) using PhyloDet. PhyloDet is a scalable phylogeny tree browser for visualizing multiple traits mapped to the tree (Lee et al., 2009).

2.6. Genetic diversity within groups of the germplasm collection

GenAlEx 6.501 (Peakall and Smouse, 2006) was used for analyses of molecular variance (AMOVA) and principal coordinate analysis (PCoA) via covariance matrix with data standardization. Population genetic parameters including Nei's unbiased genetic diversity (H) and genetic identity (I), genetic distance (D), and percentage of polymorphic loci (%P) were also calculated.

2.7. Field evaluation

The field assessments were carried out over 2 years (2011–2013). The test genotypes (48) were raised in 6 × 8 alpha lattice design with 2 replications. Each genotype was sown in a plot area of 2.4 m^2. Spacing of 30 cm was maintained between the rows. Standard cultural practices were followed throughout the growing season. Except for plant height, all data were recorded on a per-plot basis. The heights of 10 random plants were recorded at maturity and averaged. Drought tolerance index (DTI) was estimated using the following protocol: DTI = (Yc × Yrm) / $(Yc)^2$; Yc = yield of cultivar under control condition; Yrm = yield of cultivar under restricted moisture condition (Fernandez, 1992). For this, the crop was raised under both a standard irrigation schedule (3 irrigations) and a restricted moisture schedule (1 irrigation only). Field plot design remained the same, as described earlier. Two-way analysis of variance (ANOVA) was conducted in R when the sample sizes within each chloroplast groups were not the same.

3. Results

3.1. Nuclear diversity

The 158 nuclear SSR primers amplified 241 alleles among 48 *B. juncea* accessions. AMOVA analysis among populations showed that 70% of the variation was accounted for within individuals, and 30% of the variation was between individuals with zero population differentiation (ΦPT) (Table 1). Population structure analysis based on 158 nuclear SSR markers identified 3 groups at delta K-3. Percent polymorphism was very high in all 3 groups. Shannon's information index ranked the 3 populations similarly for their genetic diversity. Genetic distance analysis indicated that all 3 populations were dissimilar to each other with D = 0.000 to 0.002. Genetic identity calculation produced a similar result, with I ranging from 0.998 to 1.001 (Table 2). At the 0.60 membership probability threshold, almost half of the genotypes formed Nuclear Group 1 (NG 1), 8 genotypes (17%) were assigned to NG 2, and the remaining 16 genotypes formed NG 3 (Figure 1). The bulk of Indian low erucic acid and high erucic acid genotypes fell into NG 1. Four Indian genotypes and 1 Australian genotype were present as an admixture at the maximum threshold of 0.20. Based on their geographic origin, 1 Australian line and 7 Indian lines formed NG 2, while 1 Australian, 4 Chinese, and 11 Indian lines were included in NG 3. In the PCoA via distance matrix with data standardization, only 1 major coordinate explained 16.18%, 9.76%, and 8.27% of the molecular variance by the first 3 axes and confirmed the presence of all 3 populations (Figure 2). All the genotypes were dispersed evenly across the 2 coordinates with varying distances from the origin. The phylogenetic tree created 2 major clusters and 1 small cluster (Figure 3). Cluster I included 33 accessions with Indian, Chinese, Australian, and alloplasmic lines forming 1 clear lineage. The second clear lineage had 15 accessions conforming to all geographic and genetic formations.

3.2. Chloroplast diversity

Nine chloroplast-specific primer pairs amplified 34 cpSSR alleles following genotyping of 48 accessions forming the *Bj*DFFS set. Of these, 28 were polymorphic. The number of alleles amplified per primer pair varied from 2 to 4 with an average of 2.4 alleles per locus. AMOVA (Table 1) showed that 65% of the variation was accounted for by individual genotypes, 22% between groups, and the remaining 13% among genotypes ($P < 0.001$). Maximum population differentiation (ΦPT) was 0.22 ($P < 0.001$), suggesting significant plasmotype differences. The codominant nature of the cpSSR markers allowed analysis of the population structure. The estimated log probability of SSR marker data, calculated by structure analysis

Table 1. Analysis of molecular variance for *B. juncea* accessions based on chloroplast and nuclear genome SSRs. Grouping was based on population structure analysis.

Structure analysis	Source of valuation	df	Sum of squares	Estimated variance	Variation	%	ΦPT*	P-value
Chloroplast SSRs	Among groups	3	64.653	21.551	0.810	22%	0.22	<0.001
	Among individuals	44	147.201	3.345	0.464	13%		
	Within individuals	48	116.000	2.417	2.417	65%		
	Total	95	327.854		3.691	100%		
Nuclear SSRs	Among groups	2	173.083	86.542	0.000	0%	0.00	<0.001
	Among individuals	45	3905.750	86.794	19.981	30%		
	Within individuals	48	2248.000	46.833	46.833	70%		
	Total	95	6326.833		66.814	100%		

*ΦPT: population differentiation. Calculated as the proportion to the variance among the populations, relative to the total variance. Probability of obtaining equal or lower ΦPT value was determined by 9999 random permutations.

Table 2. Pairwise comparison of Nei's unbiased genetic identity* (above diagonal) and genetic distance* (below diagonal) among populations based on chloroplast SSR and nuclear SSR structure analysis in 48 *B. juncea* accessions.

On the basis of CpSSRs					On the basis of nuclear SSRs			
	Pop1	Pop2	Pop3	Pop4		Pop1	Pop2	Pop3
Pop1	-	0.660	0.652	0.740	Pop1	-	1.001	1.001
Pop2	0.415	-	0.968	0.914	Pop2	0.000	-	0.998
Pop3	0.428	0.033	-	0.957	Pop3	0.000	0.002	-
Pop4	0.301	0.090	0.044	-				

*Nei's unbiased genetic distance and genetic identity was calculated in GENALEX 6.501 (Peakall and Smouse, 2006) with 999 permutations (Nei, 1978).

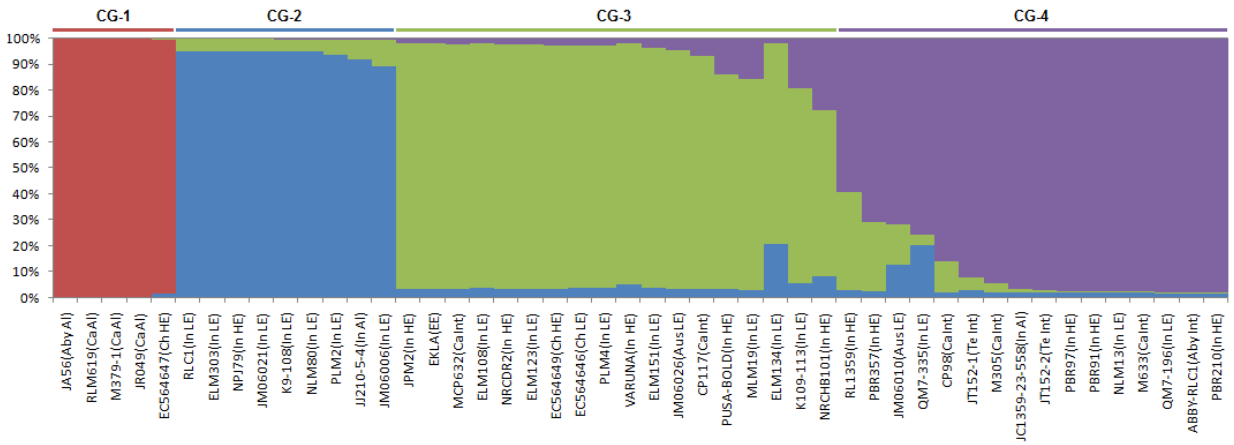

Figure 1. Probabilities of 48 *B. juncea* accessions belonging to 3 nuclear groups (NG 1–3) indicated by the representative bar's color.

(Pritchard et al., 2000), revealed a steep increase from K = 1 to K = 4, followed by relative stagnation from K = 5 to K = 10. The ad hoc measure of delta K described by Evanno et al. (2005) was applied to estimate the number of groups, and it showed a maximum at K = 4, indicating 4 major groups. Accessions were, thereafter, assigned to 4 chloroplast groups (CGs) based on the membership probability threshold at 0.60. Five (10%) accessions were assigned to CG 1, 9 (19%) to CG 2, 18 (38%) to CG 3, and 16 (33%) to CG 4 (Figure 4). Five accessions were identified as admixtures at the 0.20 maximum threshold. CG 1 included 3 alloplasmic lines with cytoplasm from *D. cardaminoides* and *E. abyssinicum* along with a Chinese genotype. CG 2 had 8 low erucic acid genotypes alongside 1 high erucic acid accession from India. CG 3 comprised 1 East European accession along with 5 Indian low erucic acid and 3 Indian high erucic acid genotypes and 1 high erucic acid accession from China. Some Indian high erucic acid lines occurred as admixtures. CG 4 comprised introgression lines along with some Indian low erucic acid and high erucic acid accessions. Shannon's information index ranked the genetic diversity somewhat differently. CG 4 (0.434) showed the highest genetic diversity, followed by CG 1 (0.384) and CG 3 (0.362), with CG 2 (0.332) having the lowest genetic diversity. The phylogenetic tree developed using PhyloDet showed 3 well-defined clusters (Figure 5), comprising 8, 21, and 19 genotypes. There were also subclusters. Assigning genotypes to various diversity clusters was mostly, though not exactly, supportive of population structure. Cluster 1 comprised 8 low erucic acid genotypes from India and 1 from Australia. Cluster 2 had 2 subclusters. Subcluster 1 included the bulk of low erucic acid Indian genotypes and 2 Australian genotypes. Subcluster 2 had East European genotypes, most of the high erucic acid Indian and Chinese genotypes, and alloplasmic lines based on *Erucastrum cardaminoides* and *E. abyssinicum*. Cluster 3 included another set of high erucic acid Indian genotypes and *B. juncea* carrying

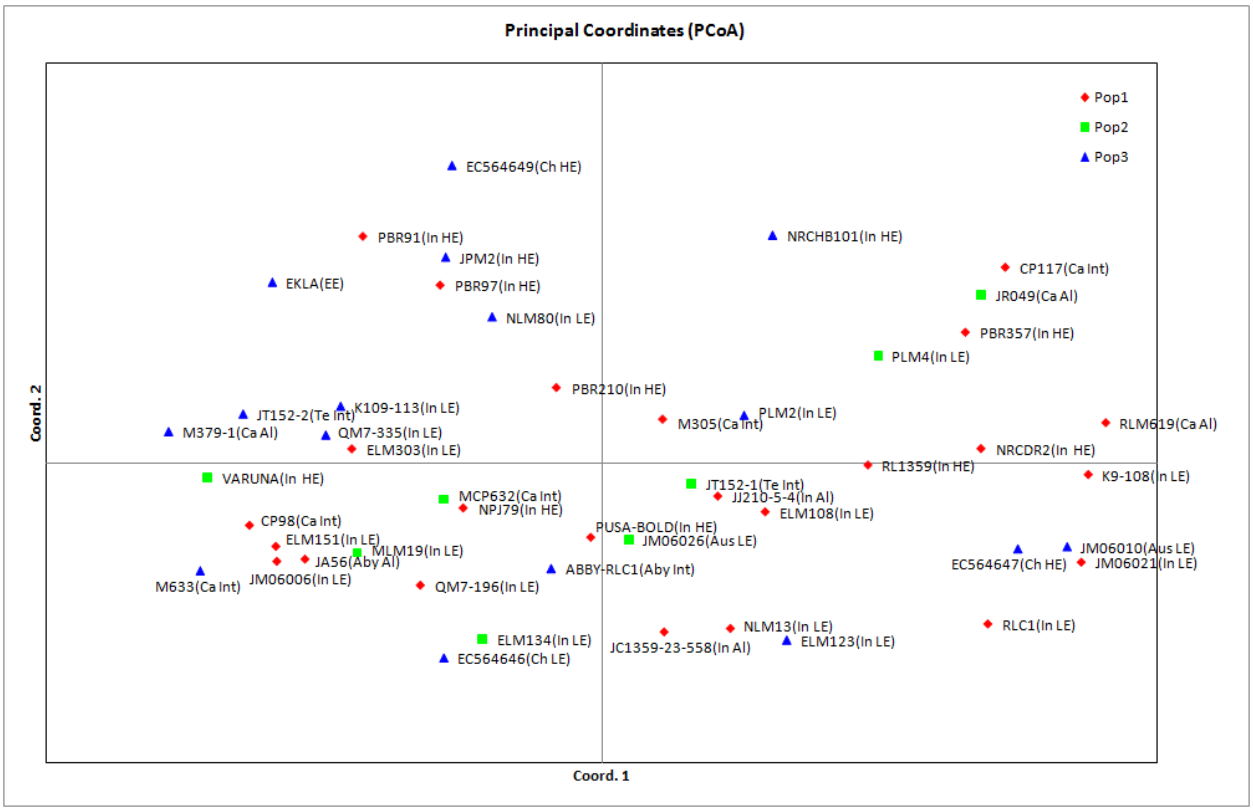

Figure 2. Plot of the 2 coordinates in the principal coordinate analysis of nuclear SSR marker polymorphism in 48 *B. juncea* accessions.

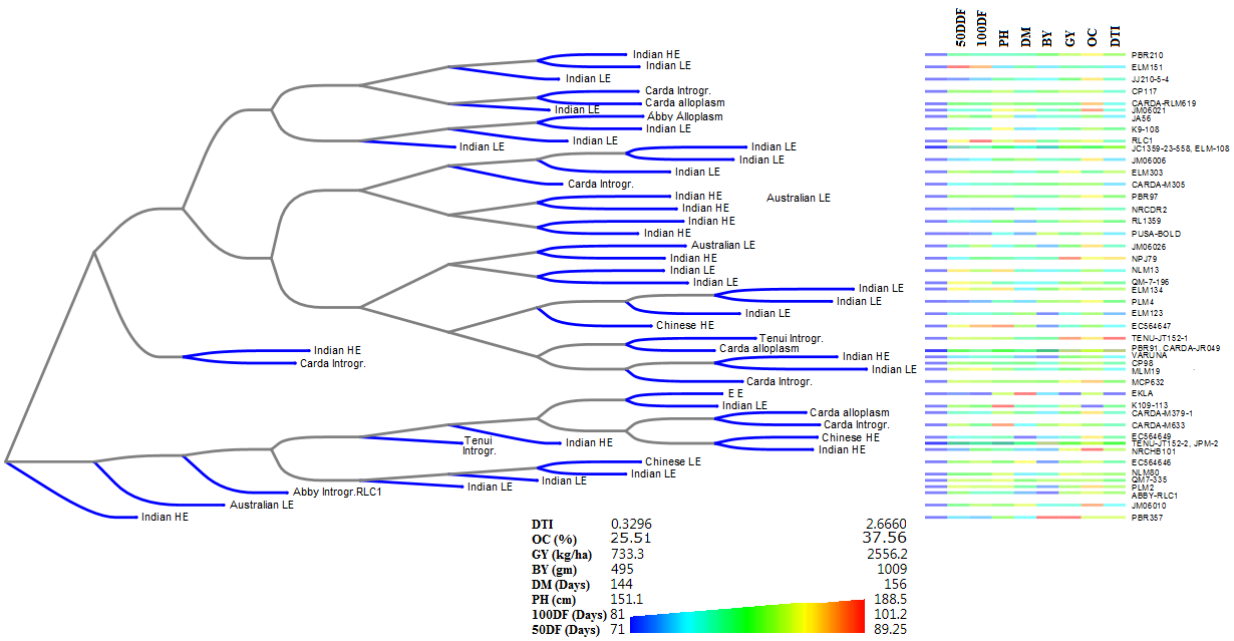

Figure 3. Phylogenetic tree of *B. juncea* accessions based upon the proportion of shared allele distance based on nuclear SSR markers.

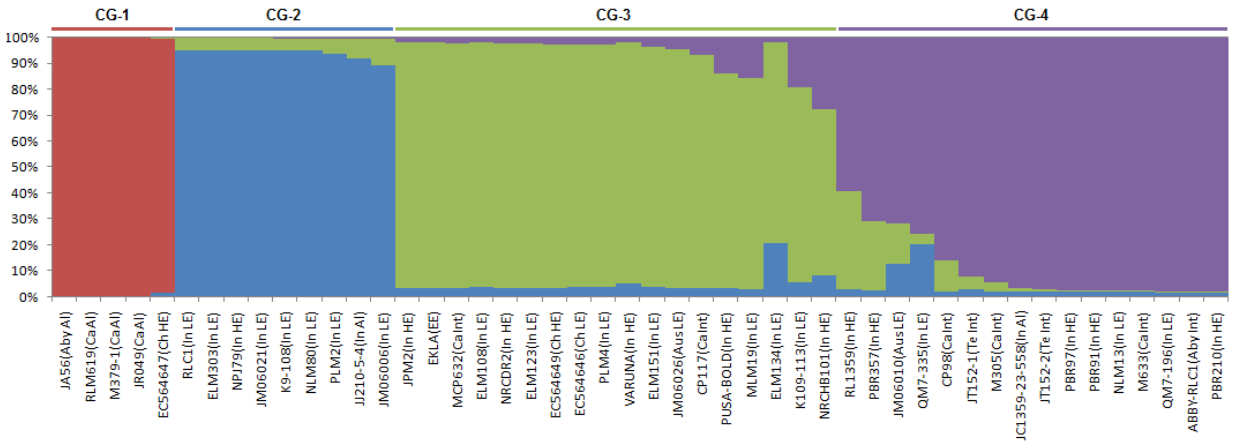

Figure 4. Probabilities of 48 *B. juncea* accessions belonging to 4 cytoplasmic groups (CG 1–4) indicated by the representative bar's color.

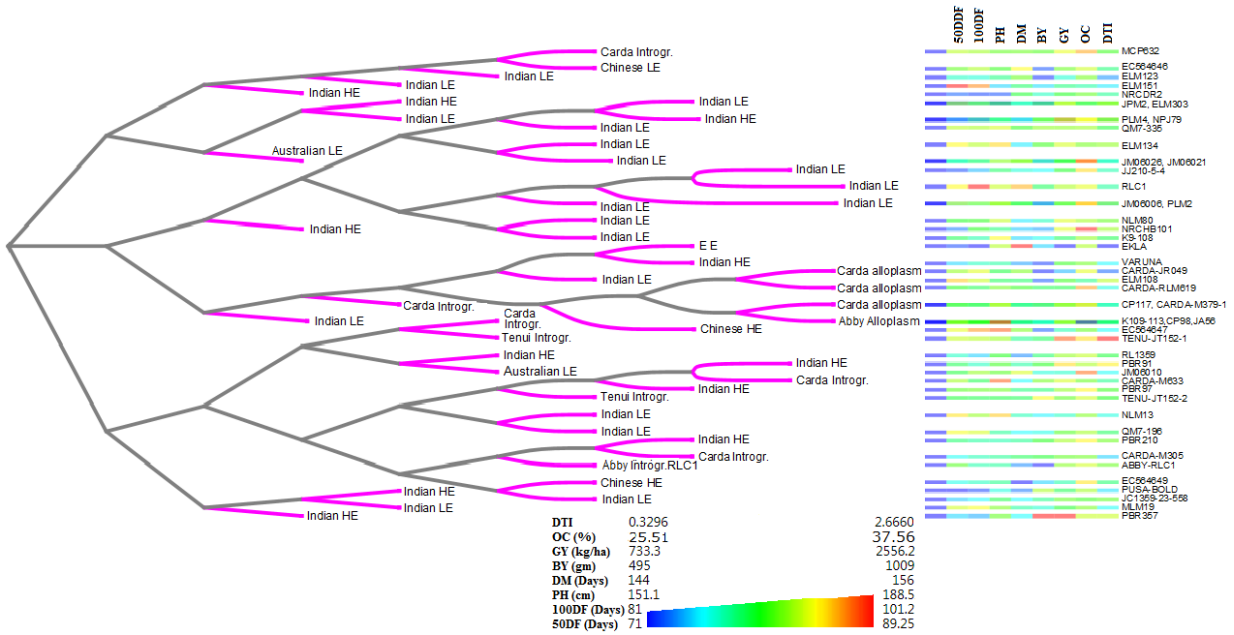

Figure 5. Phylogenetic tree of *B. juncea* accessions based on the proportion of shared allele distance based on chloroplast SSR markers.

introgressions from *Diplotaxis tenuisiliqua*, *E. abyssinicum*, and *E. cardaminoides*.

3.3. Phenotypic variation as associated with phylogenetic tree

Quantitative variation for key life history and productivity characteristics is depicted alongside the branches of the phylogenetic tree. Phenotypic variation in terms of a specific trait expression appeared not to be associated with diversity groups based on cytoplasmic or nuclear genetic SSR markers. Some differences were, however, apparent. We used cytoplasm-defined groups (CG 1–4) for further analysis. ANOVA for the linear model was estimated with weighted means of 8 morphological traits of the test accessions over 2 years of evaluation. Weighted means were used to overcome bias due to unequal sample size (genotypes) in the cytoplasm defined groups. ANOVA results are presented in Table 3. Variation over the years was indicated for days to maturity, plant height, yield, oil content, and drought tolerance index. Cytoplasm-defined groups varied significantly for plant height and biological yield. Variation within groups defined by year × genotype and year × cytoplasm interaction effects was nonsignificant. However, cytoplasm-defined grouping × genotype interactions were highly significant for grain

Table 3. Analysis of variance (ANOVA) for the linear model calculated with weighted means of 8 morphological traits of *B. juncea* accessions across 2 years with 2 replications.

Source	Df	50DF	100DF	DM	PH	BY	GY	OC	DTI
Year	1	89.38	49.01	867***	21,036.6***	-	8,968,251***	133.75***	14.192***
Replication	1	788.1***	656.4***	96.3*	1090.3*	102,345,218***	1,033,634	0.082	0.081
Chloroplast group	3	57.62	56.76	43.44	1126.8**	20,151,847*	278,287	1.949	0.706
Genotype	1	360.6**	518.6**	1	505.3	3,718,121	43,718	26.162	0.012
Year × chloroplast group	3	20.05	75.07	57.25*	164.7	-	350,248	9.366	0.447
Year × genotype	1	52.91	0.11	3.85	87.7	-	20,343	2.859	0.008
Chloroplast group × genotype	3	61.64	69.41	11.87	16.6	23,824,776*	1,164,251**	2.702	1.600
Residuals	160	42.02	48.21	16.79	228.8	6,814,880	282,293	7.495	0.627

*P < 0.05, **P < 0.01, and ***P < 0.001.
50DF: 50% days to flowering, 100DF: 100% days to flowering, PH: plant height, DM: days to maturity, BY: biological yield, GY: grain yield, DTI: drought tolerance index, OC: oil content.

yield and biological yield. Means and standard errors for individual traits are depicted in box plots (Figure 6). In terms of overall averages, CG 1 (78.1 days) and CG 4 (76.5 days) were earlier to flower and mature (149.7 days, 146.8 days). They also had relatively short stature. Average biological yield (10,339.57 kg/ha) was higher for CG 4, followed by CG 2 (7990.18 kg/ha). CG 3 (1801.67 kg/ha) and CG 4 (1904.51 kg/ha) showed higher mean yield(s) as compared to the remaining groups. Expression for oil content varied over 2 years. For year 1, CG 2 (34.27%) and CG 3 (34.95%) averaged higher oil content. In contrast, CG 3 (33.79%) had the highest mean oil content in the second year. Drought tolerance index appeared very variable across the years. However, genotypes included in CG 3 (1.4746) had better drought tolerance ability.

4. Discussion

Understanding population structure is crucial for any population panel developed for use in association genetic analysis. Divergent population structures result from alterations in allele frequencies between divergent structured populations due to genetic drift, mutation, and selection. If the differences in allele frequencies are correlated with morphological traits that differentiate 2 populations, statistical correlation between a gene and a trait may lead to spurious associations. The estimates of population structure are used in association with mapping studies to account for effects due to structure. Most such studies take into account population structure based on nuclear genetic variation. There are hardly any reports regarding the role of cytoplasmic background on population structure. This is in spite of the demonstrated influence of background cytoplasm on genome fractionation and, consequently, genetic diversity and trait variation (Banga et al., 1984, 2003). In this study, we generated polymorphisms with nuclear SSRs and cpSSRs through genotyping a set of sorted and true-breeding *B. juncea* germplasm accessions and breeding lines. These polymorphisms were used to draw the diversity of relationships between cytoplasmic and nuclear genomes. Population structures and familial relationships are implicit in diversity-fixed populations due to nonrandom mating and consequently Hardy–Weinberg disequilibrium.

Population structure analysis based on nuclear SSR markers identified 3 groups. Phylogenetic analysis suggested 3 lineages. These were reflective but not completely supportive of the population structure. The first lineage, which included 8 low/double low erucic acid genotypes from India and 1 from Australia, was expected to be associated with extant *B. rapa* cytoplasm, as most Indian low erucic acid genotypes in our study were developed using low erucic acid *B. rapa* as the female parent. The second lineage had 2 subclusters. The first carried the majority of low erucic acid lines, which were developed using the trait donors (Zem 1 and Zem 2) as the male parent. Apparently, these lines carried cytoplasm from ancient hybridization events. Subcluster 2 was distinct, and it included 1 East European accession, most of the high erucic acid Indian and Chinese genotypes, and alloplasmic *B. juncea* genotypes based on *Erucastrum cardaminoides* and *E. abyssinicum*. Convergence of the alloplasmic lines (carrying *Erucastrum cardaminoides* and *E. abyssinicum* cytoplasms) with natural germplasm lines from East Europe, known to carry *B. nigra* cytoplasm, and the traditional high erucic acid Indian and Chinese genotypes in subcluster 2 was considered significant, as both cytoplasm donors *Erucastrum cardaminoides* and *E. abyssinicum* belong to the *nigra* lineage, which is known to be distinct from the *rapa* lineage (Arias and Pires, 2012). Under this scenario, the entire lineage 2 (including

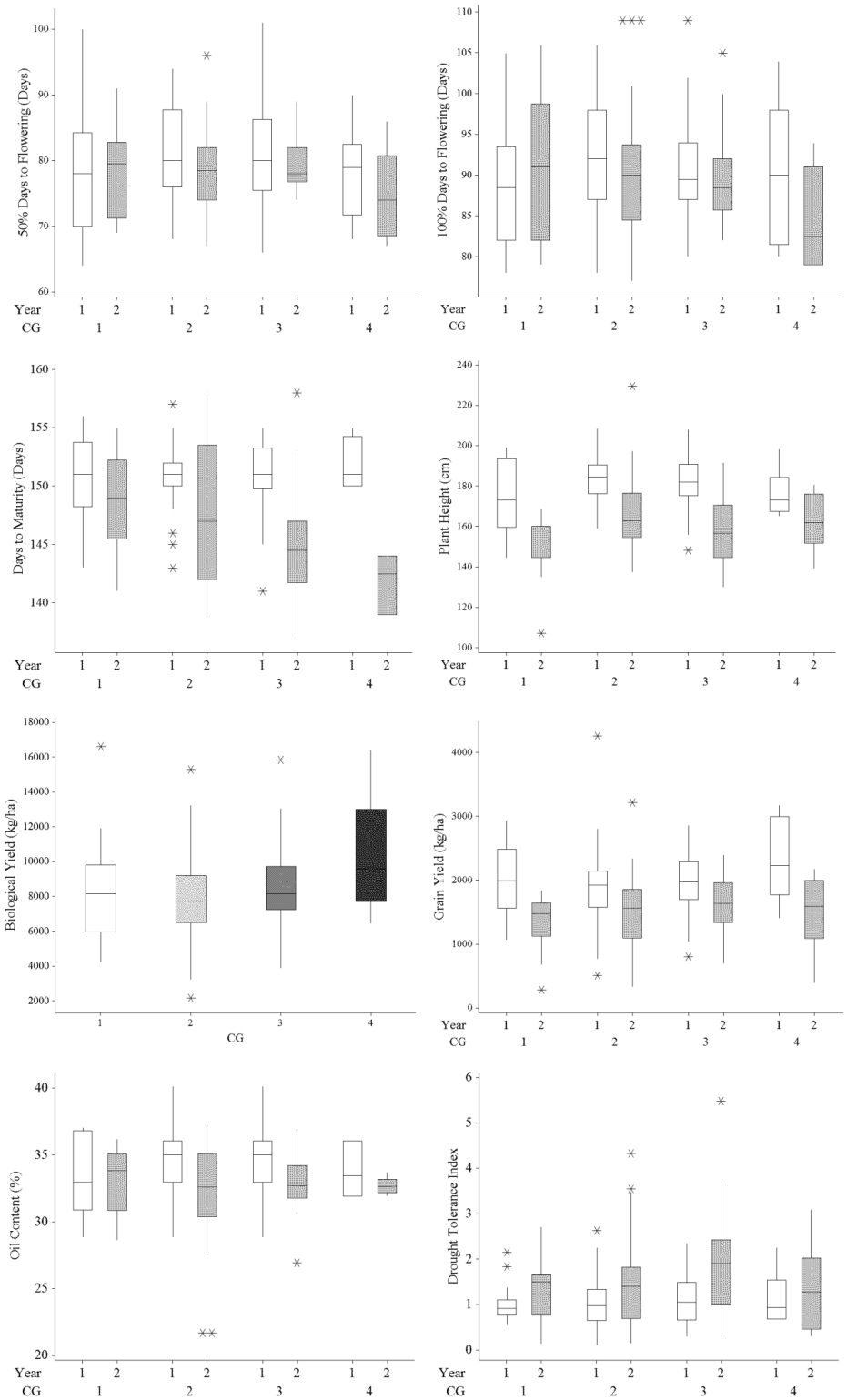

Figure 6. Box plots showing variation for morphological traits over 2 years in 4 groups (CG) based on plasmotype diversity. *P < 0.05, **P < 0.01, and ***P < 0.001.

subcluster 1), as developed through the phylogenetic tree, can be placed in the *nigra* lineage in terms of cytoplasm. The third lineage comprised high erucic acid Indian genotypes along with *B. juncea* accessions carrying genetic introgressions from *Diplotaxis tenuisiliqua*, *E. abyssinicum*, and *E. cardaminoides*. All these accessions were developed using natural *B. juncea* as a maternal parent. The cytoplasm in natural Indian and Chinese genotypes (both low erucic and high erucic acid genotypes) in lineage 3 were, therefore, considered distinct from the extant *B. rapa* cytoplasm known to be carried by genotypes in cluster 1. Mantel test statistics were nonsignificant, indicating no association between population structures developed based on nuclear and cytoplasmic diversities. At the micro level, however, some convergences did exist between the 2 groupings. Estimates of differentiations obtained using cpSSR markers were higher than those based on nuclear SSR markers. Cytoplasmic and nuclear differentiations are expected to contrast because organelle effective population size is generally half the size of the nuclear one, depending on the sex ratio and flower structure in terms of presence of pistil and stamens in the same flower or different flowers on the same plant or different plants (Laporte et al., 2001). There is also potential asymmetry in gene flow due to contrasted pollen and seed dispersal capacities (Hamilton and Miller, 2002). The spatial distribution of cytoplasmic polymorphism, established by seed dispersal during the settlement of new populations via range expansion or multiple colonization events, is expected to be more conserved and more obvious than the population patterns depicted with nuclear genetic markers with bimodal transmission through pollen and seed. The chloroplast genome is also known to have low mutation rates and to be more conserved than the nuclear genome. Germplasm patterns obtained using distance-based clustering were not in agreement with the structure analysis, especially for the nuclear genome-based diversity. Apparently, structure was more efficient for the detection of ancestry-based genetic differentiation than the subsequent genetic divergences resulting from geographic origin, natural selection, and crop breeding. This is, however, quite different from the situations where the habitats were considered the primary determinant of the population structure. The phylogenetic tree based on nuclear SSR markers was broadly reflective of the concept of separate *nigra* and *rapa* cytoplasm-based lineages for *B. juncea* as described earlier. Unlike the cpSSR-based phylogenetic tree, where alloplasmic lines clustered together, these dispersed across clusters in the phylogenetic tree based on nuclear SSR markers. Evidently, the influence of cytoplasm on nuclear diversity is a long-term evolutionary response. Besides the impact of population structure, cytoplasm was also found to influence trait variation in our studies. Genotypes carrying *B. nigra* cytoplasm (CG 3) averaged higher seed yield, oil content, and greater tolerance to drought. In contrast, genotypes carrying ancient *B. rapa* cytoplasm (CG 4) averaged higher biological yield.

Our studies showed that stratified sampling was effective in capturing the genome-wide SSR variation, as the core subset displayed a very high SSR variability. Maximum SSR variation resided among the accessions of different geographies. The accessions from China were the most genetically diverse. Further, the cytoplasmic background showed no correlation with the distribution of genetic variation based on neutral markers. However, association between population structure and phenotypic variation was indicated by significant interactions between chlorotype with seed/biological yields. There may thus be a need to factor in the impact of background cytoplasm on population structure in association studies of crops where different cytoplasmic lineages are known.

Acknowledgments

This study was supported by the Indian Council of Agricultural Research under the ICAR National Professor Project "Broadening the genetic base of Indian mustard (*Brassica juncea*) through alien introgressions and germplasm enhancement" awarded to SSB. Thanks are also due to Dr Isobel Parkin for providing access to chromosome-specific SSRs for the *Brassica* B genome.

References

Arias T, Pires JC (2012). A fully resolved chloroplast phylogeny of the *Brassica* crops and wild relatives (*Brassicaceae*: *Brassiceae*): novel clades and potential taxonomic implications. Taxon 61: 980–988.

Banga SS, Banga SK, Labana KS (1983). Nucleo-cytoplasmic interaction in *Brassica*. In: Proceedings of the 6th International Rapeseed Conference, 17–19 May 1983; Paris, France. pp. 602–606.

Banga SS, Deol JS, Banga SK (2003). Alloplasmic male sterile *Brassica juncea* with *Enarthrocarpus lyratus* cytoplasm and the introgression of gene(s) for fertility restoration from cytoplasm donor species. Theor Appl Genet 106: 1390–1395.

Banga SS, Labana KS, Medhi BN (1984). Alternaria incidence in some alloplasmic lines of Indian mustard (*Brassica juncea* (L.) Coss.). Theor Appl Genet 67: 195–197.

Chandra A, Gupta ML, Ahuja I, Kaur G, Banga SS (2004). Intergeneric hybridization between *Erucastrum cardaminoides* and two diploid crop *Brassica* species. Theor Appl Genet 108: 1620–1626.

Dent AE, von Holdt BM (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4: 359–361.

Doyle JJ, Doyle JL (1990). Isolation of plant DNA from fresh tissue. Focus 12: 13–15.

Erickson LR, Straus NA, Beversdorf WD (1983). Restriction patterns reveal origins of chloroplast genomes in *Brassica* amphidiploids. Theor Appl Genet 65: 201–206.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611–2620.

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Fernandez GCJ (1992). Effective selection criteria for assessing plant stress tolerance. In: Proceedings of the International Symposium on Adaptation of Vegetables and Other Food Crops in Temperature and Water Stress, 13–16 August 1992; Shanhua, Taiwan. pp. 257–270.

Flannery ML, Mitchell FJ, Coyne S (2006). Plastid genome characterisation in *Brassica* and *Brassicaceae* using a new set of nine SSRs. Theor Appl Genet 113: 1221–1231.

Hamilton MB, Miller JR (2002). Comparing relative rates of pollen and seed gene flow in the island model using nuclear and organelle measures of population structure. Genetics 162: 1897–1909.

Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806.

Kim H, Choi SR, Bae J (2009). Sequenced BAC anchored reference genetic map that reconciles the ten individual chromosomes of *Brassica rapa*. BMC Genomics 10: 432.

Laporte V, Viard F, Bena G, Valero M, Cuguen J (2001). The spatial structure of sexual and cytonuclear polymorphism in the gynodioecious *Beta vulgaris* ssp. *maritima*: I/ at a local scale. Genetics 157: 1699–1710.

Lee B, Nachmanson L, Robertson G, Carlson JM, Heckerman D (2009). PhyloDet: a scalable visualization tool for mapping multiple traits to large evolutionary trees. Bioinformatics 25: 2611–2612.

Lodhi MA, Ye GN, Weeden NF, Reisch BI (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol Biol Report 12: 6–13.

Lowe AJ, Moule C, Trick M, Edwards K (2004). Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108: 1103–1112.

Maliyakal EJ (1992). An efficient method for isolation of RNA and DNA from plants containing polyphenolics. Nucleic Acid Res 20: 23–81.

Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590.

Peakall R, Smouse PE (2006). GENALEX 6: Genetic Analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288–295.

Pritchard JK, Stephens M, Donnelly PJ (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Pritchard JK, Wena X, Falush D (2010). Documentation for STRUCTURE software: Version 2.3. Chicago, IL, USA: Department of Human Genetics, University of Chicago.

Zhu C, Gore M, Buckler ES, Yu J (2008). Status and prospects of association mapping in plants. Plant Gen 1: 5–20.