

# Genetic characterization and divergence studies of maize (*Zea mays* L.) lines developed from landraces indigenous to North Eastern Hill Region (NEHR) of India

K. L. Naveenkumar, Devyani Sen\* , Shimreiso Vashum and Miranda Sanjenbam

*School of Crop Improvement, College of Post-Graduate Studies in Agricultural Sciences, Central Agricultural University (Imphal), Umroi Road, Umiam 793103, Meghalaya, India*

Received 23 December 2019; Accepted 27 July 2020 – First published online 9 October 2020

## Abstract

The current study focused on characterization of the underlying genetic divergence in inbred lines developed from local landraces of North Eastern Hill Region of India – a designated Asiatic maize diversity centre – following six generations of inbreeding. Substantial genetic differentiation was indicated based on very high to moderate  $F_{st}$  values for 22 of the 38 simple sequence repeat markers studied. STRUCTURE analysis partitioned the subset into two distinct and one admixture subgroup (Populations I, II and III respectively) accompanied by a significant reduction in heterozygosity. Population II was further subdivided into subpopulations Pop-M9 and Pop-T9. Nei's pairwise genetic distance and population  $F_{st}$  values indicated that Populations I and II were more divergent with neighbour joining clustering analysis clearly defining landraces originating from the states of Tripura (Population II) and Sikkim (Population I) as most divergent. Principal coordinates analysis could explain 31.26% of the variation present in the subgroups wherein Population I was more variable. Analysis of molecular variance and  $F_{st}$  coefficients ( $P < 0.001$ ) indicated 17% population structuring with 55% variation detected for individuals within populations. These results combined with the presence of phenotypic variability in the subgroups for yield traits supported by results of a preliminary partial diallel analysis strongly suggest the existence of distinct heterotic groups. Divergence studies are crucial for exploiting heterosis, and the current study would go a long way to help establish a germplasm base for developing varieties with improved agronomic performance and surer commercial prospects no reports of which are available thus far.

**Keywords:** Asiatic maize landraces, genetic divergence, maize, North East India, population structure, SSR (simple sequence repeat) markers

## Introduction

Maize is the most cultivated cereal crop worldwide (Yang *et al.*, 2017; FAO, 2018). Being an extremely versatile crop, it originated in Mexico and diffused to the rest of the world through trade routes to diversify into distinct landraces (Rebourg, *et al.*, 2003; Bedoya *et al.*, 2017; Wang *et al.*, 2017). It was introduced in India both in pre- and post-Columbian era (Singh, 1977; Kumar and Sachan, 1994), and the Eastern Himalayan region reports a very

high genetic variability for maize (Sharma *et al.*, 2010; Prasanna, 2010, 2012; Hossain *et al.*, 2016) with North East India designated as an Asiatic maize diversity centre (Sharma and Brahma, 2011).

The genetic diversity of these landraces endemic particularly to the North Eastern Hill Region (NEHR) of India has been maintained for generations by tribal farmers as part of their socio-cultural practices and comprise a distinct gene pool of useful alleles (Singh, 1977; Prasanna, 2012). Very little of the novel alleles found in landraces however get utilized for cultivar development because of problems of adaptation in regions beyond their cultivation (Yao *et al.*, 2007; Lia *et al.*, 2009; Romay *et al.*, 2012). Nonetheless, in

\*Corresponding author. E-mail: [devyani.sen@gmail.com](mailto:devyani.sen@gmail.com)

the face of plateauing yields and climate change, the allelic richness of crop genetic resources such as landraces and/or close wild relatives that help to conserve heterogeneity in cropping systems will require to be harnessed for widening the genetic base of hybrid breeding programmes to ensure food security (FAO, 2019).

The commercial success of hybrid maize has been largely driven by the escalated demands as animal feed owing to the shift by farmers to cultivating high-yielding single cross hybrids (Kumar *et al.*, 2013; Pavithra *et al.*, 2018). With a fast growing livestock sector, the NEHR would stand to greatly benefit from locally available quality animal feed (Feroze *et al.*, 2010; Bhagat *et al.*, 2015), production of which is impeded by practices of shifting cultivation over reduced fallow periods and use of low-yielding varieties, leading to pronounced yield gaps compared to the national scenario (Tripathi *et al.*, 2003; Grogan *et al.*, 2012). One course of action to bridge this yield gap would be to emulate the commercial success of single cross hybrids as found elsewhere in the world and the country (Duvick, 2001, 2005; Dass *et al.*, 2009; Hallauer *et al.*, 2010; Dass *et al.*, 2012) using genetically divergent landraces of the region which are already fortified with the advantage of local adaptation.

Since knowledge of maize diversity is critical for exploitation of heterosis in single cross breeding programmes, the objective of the current study was to investigate genetic divergence in a subset of 111 inbred lines developed from landraces originating at varying altitudes across the NEHR of India. The process of inbreeding phenotypically favourable individuals can be further reinforced with genetic analysis (Liu *et al.*, 2003; Choukan *et al.*, 2006; Charlesworth and Willis, 2009; Lai *et al.*, 2010; Morrell *et al.*, 2012; Nyaligwa *et al.*, 2015) and therefore the subset was subjected to population subgrouping based on maximum likelihood and distance-based analysis using simple sequence repeat (SSR) markers. This would allow us to understand the extent of differentiation in the subgroups, the ultimate goal being establishment of a germplasm base for heterotic grouping of the genetically diverse maize indigenous to this centre of diversity.

## Materials and methods

### Study material

A total of 111 lines developed from seven different landraces collected from five different North Eastern states of Meghalaya (M9, M22), Manipur (Ma5), Nagaland (N11, N25), Sikkim (S16) and Tripura (T9) were studied. While some of these landraces had been identified as tolerant/resistant to biotic stresses predominant in the NEHR from previous studies (Sanjenbam *et al.*, 2018) certain other were reported to be tolerant to abiotic stresses (unpublished

data). The passport data of the landraces are provided in online Supplementary Table S1. The selection and inbreeding programme was initiated in 2015 at the Experimental Farm, College of Post-Graduate Studies in Agricultural Sciences, Central Agricultural University (Imphal), Umiam, Meghalaya, India (25°40'52.9"N, 91°54'40.7"E) employing full sib-mating from generations one to four and selfing in generations five and six respectively. Morphological data were recorded as per the standard descriptors outlined by the Protection of Plant Varieties and Farmers' Rights authority, New Delhi (Anonymous, 2007). Eight yield contributing quantitative traits viz. anthesis silking interval (ASI, days), plant height (PH, cm), ear height (EH, cm), ear with husk (EWH, g), ear without husk (EWWH, g), ear length (EL, cm), number of kernels per row (NK) and kernel yield per plant (GYP, g) were subjected to principal component analysis (PCA).

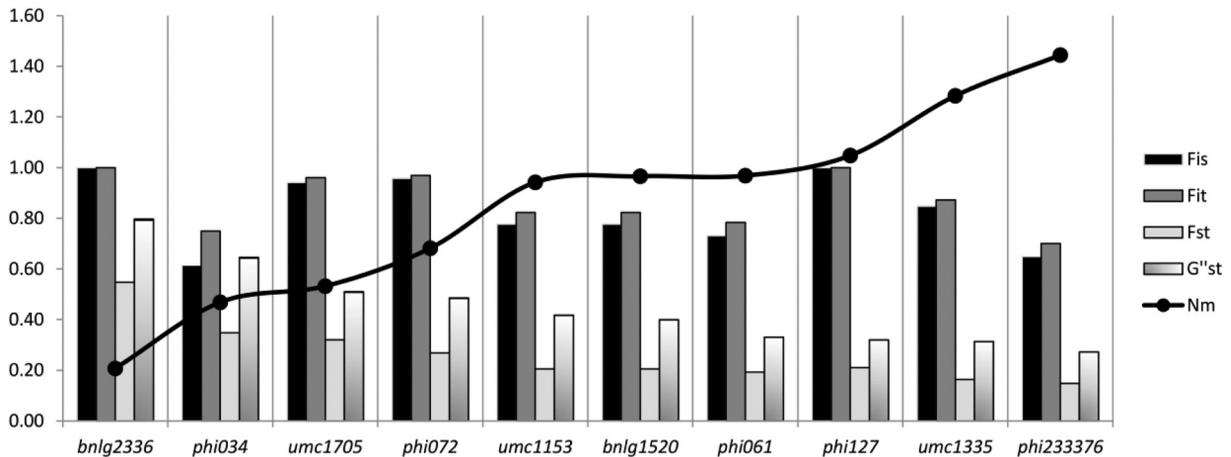
### DNA extraction, polymerase chain reaction and SSR genotyping

For genotyping studies, fresh leaf samples were collected at seedling stage for extraction of genomic DNA using the modified CTAB extraction protocol (Doyle and Doyle, 1990) from each of the individual lines under study. A total of 48 reported SSR markers consistent with high phenotypic variation and high polymorphic information content (PIC) values obtained from the maize GDB database (Portwood *et al.*, 2018) were used to genotype the panel. DNA quantity was assessed in 0.8% agarose gel (Sigma) and individual samples were uniformly diluted to a final concentration of 10 ng/μl.

Polymerase chain reactions (PCRs) performed in 10 μl reaction volume in a programmable thermal cycler included an initial denaturation step at 94°C for 4 min followed by 35 cycles of denaturation at 94°C for 30 s, annealing for 30 s adjusted to temperatures depending on length of the primer (online Supplementary Table S2) and extension at 72°C for 30 s. The final extension was carried out at 72°C for 4 min. Depending on the size of the amplicon, 7 μl mixture of the amplified PCR product was resolved in either 1.5 or 2% agarose gel and stained with ethidium bromide. The DNA amplicons separated on the basis of size as observed under a gel documentation unit (Alpha Imager Mini) were scored relative to the standard 100 bp ladder (Gene Ruler, Fermentas).

### Data analysis

SSR data were analysed using software package STRUCTURE (Version 2.3.4) developed by Pritchard *et al.* (2000). A total of 20 independent runs for each *K*, with *K* values ranging from 1 to 10 and a Markov chain Monte Carlo replication burn length of 10,000–100,000 was



**Fig. 1.**  $F$  coefficient values of SSR markers *bnlg2336*, *phi034*, *umc1705*, *phi072*, *umc1153*, *bnlg1520*, *phi061*, *phi127* and *umc1335* with highest contribution to genetic differentiation positively correlated with Hedrick's standardized  $G''_{st}$  values. The  $F$  values were negatively correlated with gene flow ( $N_m$ ).

performed as per Evanno *et al.* (2005), the results of which were corroborated with those of STRUCTURE HARVESTER (Fig. 2(a)) to calculate optimal  $K$  value. Lines with membership probabilities  $Q \geq 0.98$  were assigned to the two subgroups while lines with membership probabilities below 0.98 were clubbed into a single admixture subgroup. The larger subgroup was further subdivided similarly. These subgroups defined by STRUCTURE hereafter also designated as populations were then studied for frequency and distance-based analysis with respect to  $F$ -statistics,  $G_{st}$  adjusted for  $F_{st}$  bias, Hedrick's standardized  $G_{st}$  ( $G''_{st}$ ) further corrected for bias when the number of populations are small, analysis of molecular variance (AMOVA), principal coordinates analysis (PCoA), allele frequencies and expected and observed heterozygosity using GenAlEx software version 6.5 (Peakall and Smouse, 2012).

Allelic diversity as explained by PIC values was calculated as per Nei (1973) using the following formula:

$$H_j = 1 - \sum p_i^2,$$

where  $p_i$  is the frequency of the  $i$ th allele.

The phylogenetic tree was constructed using the unweighted neighbour joining clustering method based on dissimilarity index computed from simple matching coefficient in DARwin 6.0.21. Descriptive statistics for the eight yield contributing morphological traits, correlation studies and PCA was performed in MS excel using XLSTAT (Version 2014.5.03).

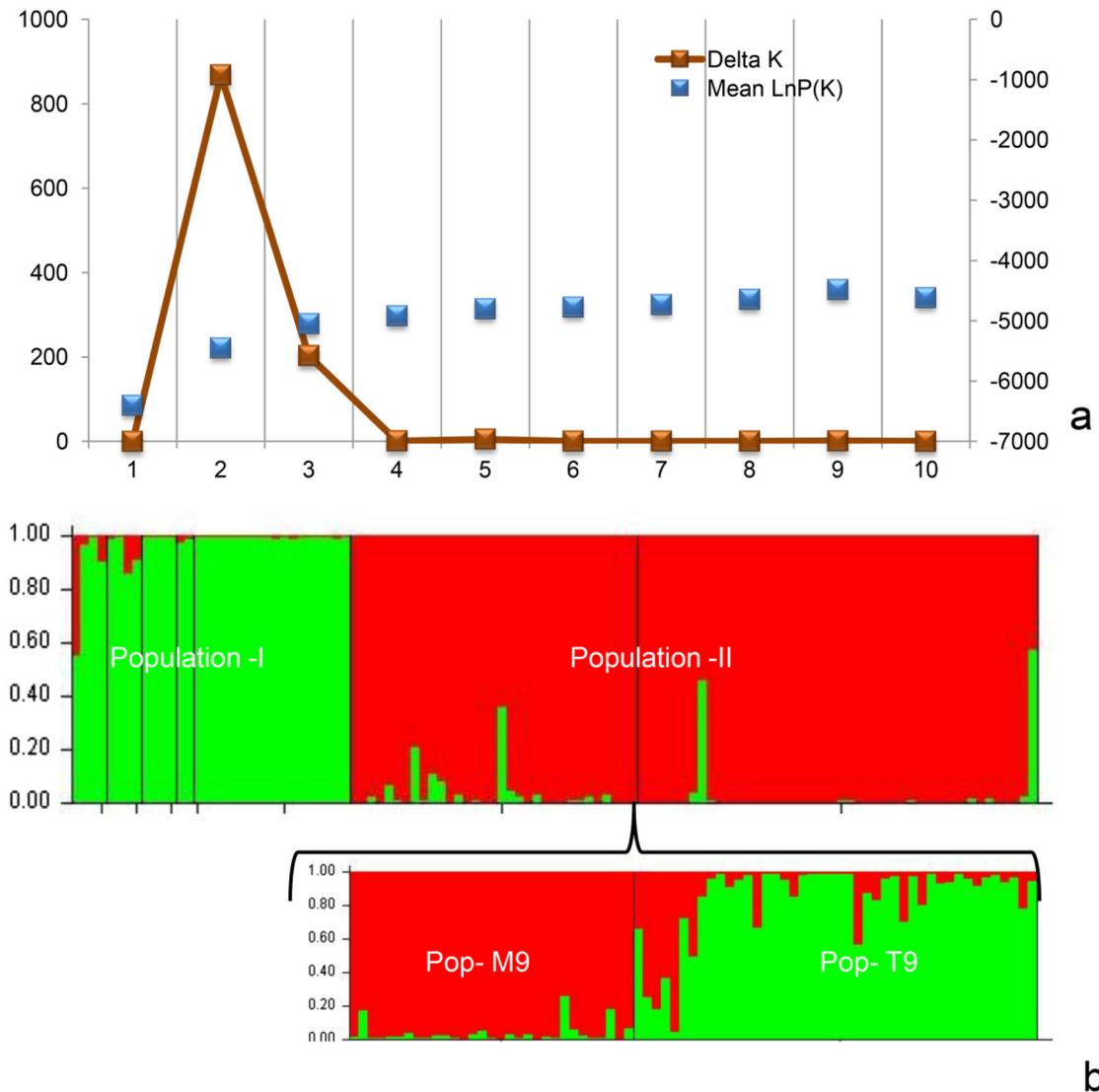
## Results

### SSR diversity studies

Of the 48 SSR markers studied, a total of 38 were polymorphic with missing values within permissible limits as

determined by Genalex software and were used for further analysis. With a mean of 2.32 alleles per locus, the average PIC value for the informative markers was 0.38 with 25 markers recording PIC values higher than the average of which five markers viz. *umc1277*, *umc2059*, *bnlg439*, *bnlg1484* and *umc1149* of bins 9.08, 6.08, 1.03, 1.03 and 8.06 respectively recorded PIC values ranging from 0.57 to 0.50 (online Supplementary Table S3).

The least informative marker with the lowest PIC value of 0.07 was reported for *phi053* located in bin 3.05. A total of 88 alleles were detected with observed heterozygosity (%Het) for the SSR markers studied ranging from 0 to 29.1%. Studies with respect to  $F$ -coefficients viz.  $F_{st}$  the coancestry coefficient,  $F_{is}$  the consanguinity coefficient and  $F_{it}$  the inbreeding coefficient, revealed that 22 SSR markers recorded  $F_{st}$  values in the range of 0.59 to 0.06 implicating their contribution in high to moderate population substructuring. With the exception of SSR markers *mmc0241* and *umc2101* of bins 6.05 and 3 which recorded low/negative  $F_{is}$  but high  $F_{st}$  values as a result of harbouring excess heterozygotes due to negative assortative mating, for the remaining loci, overall high  $F_{st}$  values were positively correlated with high  $F_{is}$  values. SSR markers *bnlg2336*, *phi034*, *umc1705*, *phi072*, *umc1153*, *bnlg1520*, *phi061*, *phi127*, *umc1335* and *phi233376* of bins 10.04, 7.02, 5.03, 4.01, 5.09, 2.09, 9.03, 2.08, 1.06 and 8.03, respectively (Fig. 1) recorded higher  $F_{st}$ ,  $F_{is}$  and  $F_{it}$  values than the mean. Collectively, these SSR markers recorded an average fixation index of 0.80. The overall  $G_{st}$  mean of 0.122 when corrected for bias in small populations using Hedrick's standardized  $G_{st}$  ( $G''_{st}$ ) stood at 0.263. As expected, the gene flow ( $N_m$ ) was low for loci with high  $F_{st}$  values.

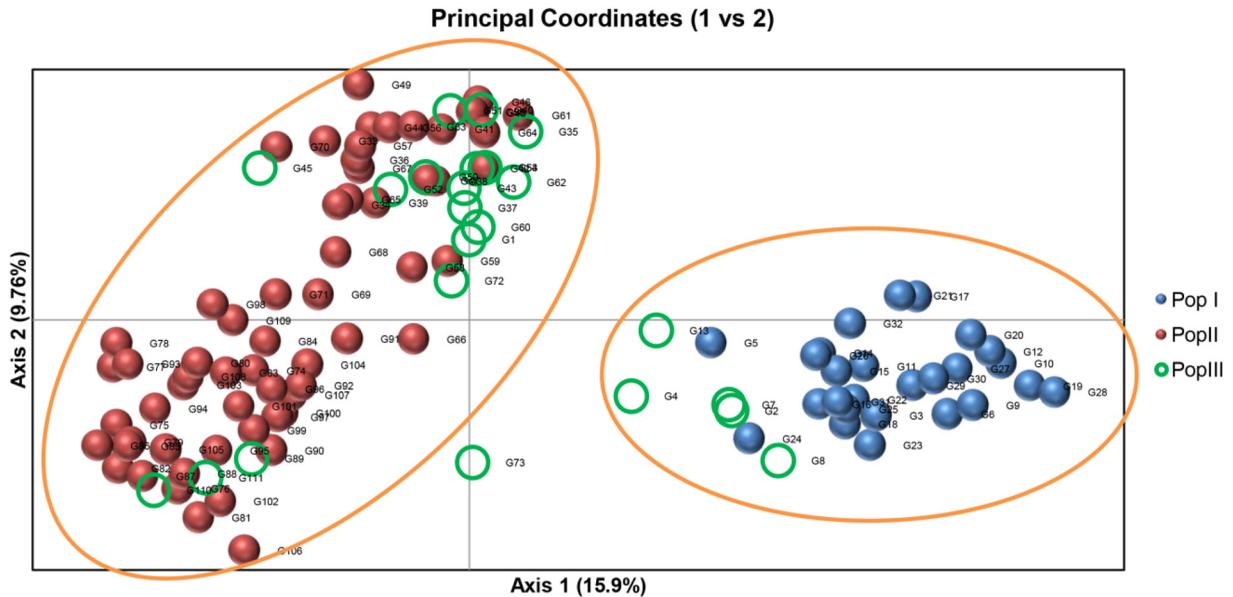


**Fig. 2** (a) Mean likelihood  $l(K)$  and variance per  $K$  value over 20 runs as per STRUCTURE in the subset of 111 individual lines studied for the 38 polymorphic SSR markers. (b) Population subgrouping with membership probabilities  $Q \geq 0.98$  achieved by STRUCTURE in the subset of 111 individual lines studied. Population II further subdivided into Pop-M9 and Pop-T9.

### Population structuring based on SSR genotyping

Subgroup delineation using STRUCTURE based on the optimal value of  $K = 2$  for the 111 lines with a cut-off  $Q \geq 0.98$  differentiated 88 of the 111 lines into two distinct groups. A rigid cutoff value of  $Q \geq 0.98$  was maintained to ensure strict delineation of the lines into different subgroups since no prior information on the heterotic pattern of these lines is known. A visual representation is depicted in Fig. 2(b) with 26 inbreds clustered in Population I (red) and 62 in Population II (green). The remaining 23 lines of mixed ancestry with  $Q < 0.98$  were assigned to a third group, Population III. Members of Population II could be similarly further delineated in two subpopulations Pop-M9 and Pop-T9.

PCoA revealed that the first three axes could explain a total of 31.26% variation, individually accounting for 15.99, 9.75 and 5.52% of the total variation, respectively. The members of Populations I and II were found to occupy distinct quadrants with no overlaps. A majority of the admixtures overlapped members of Population II with relatively fewer individuals overlapping members of Population I (Fig. 3). Comparison of the principal coordinates (PCos) 1 versus 2 revealed that maximum variation reflected in PC1 with an eigen value of 42.96 was accounted by members of Population I. Individuals of Population II comprising lines developed from landraces belonging to M9 (Meghalaya) and T9 (Tripura) were spread across two quadrants along PC2 as per the subgroups obtained using  $Q$  probabilities.



**Fig. 3.** PCoA for plot of axis-1 and axis-2 where Pop I along axis 1 represents Population I, Pop II along axis 2 represents Population II clearly spread across two distinct quadrants and Pop III represents the admixture subgroup.

**Table 1.** AMOVA and  $F$  coefficient values

Source	Estimated variance	% Variance	$F$ -statistics	$P$ -value
Among pops (AP)	1.38	17	$F_{st}$ 0.171	<0.001
Among individuals (AI)	4.43	55	$F_{is}$ 0.661	<0.001
Within individuals (WI)	2.27	28	$F_{it}$ 0.719	<0.001
Total (TOT)	8.10	100	$F_{st\ max}$ 0.644	

$F_{st}$  = inbreeding coefficient within subpopulations relative to the total (AP/TOT).

$F_{is}$  = inbreeding coefficient within individuals relative to the subpopulation (AI/(WI + AI)).

$F_{it}$  = inbreeding coefficient within subpopulations relative to the total ((AI + AP)/TOT).

$F_{st\ max}$  = maximum  $F_{st}$  achievable.

Hierarchical partitioning of total variance using AMOVA indicated a departure from panmixis where 17% variation was the result of population grouping with majority of the variation (55%) accounted for by individuals within populations. Altogether 28% variation was accounted for by differences arising within the lines relative to the entire population taken as a whole. A  $P$  value <0.001 indicated that the variation observed for  $F_{st}$ ,  $F_{is}$  and  $F_{it}$  (Table 1) at 0.171, 0.661 and 0.719 respectively was highly significant implying that six generations of inbreeding has led to the meaningful sub-structuring of the lines into distinct homogeneous units. The highest value of  $F_{st}$  that could be achieved under situations of maximum possible among population diversity ( $F_{st\ max}$ ) was 0.644.

Pairwise calculations of Nei's genetic distance which is based on the assumption of biological changes viz. genetic drift and mutation resulting in differences among populations revealed greater differences between individuals of

Populations I and II compared to Populations II and III or I and III. Similarly, a population  $F_{st}$  value of 0.163 between Populations I and II indicated high genetic differentiation compared to Populations I and III with an  $F_{st}$  value of 0.090 which was indicative of moderate genetic differentiation. In case of Populations II and III, an  $F_{st}$  value of 0.03 indicated very small levels of genetic differentiation. A highly significant deviation from panmixis in all of the three populations for expected ( $H_e$ ) and observed heterozygosity ( $H_o$ ) was also observed following six generations of inbreeding. Allelic richness as indicated by the number of effective alleles ( $N_e$ ) was highest in Population I. A total of 8, 21 and 10 private alleles (PA) were identified in Populations I, II and III respectively. For the 39 PA identified, the frequency of occurrence ranged from 0.008 to 0.2 with the highest number of unique alleles observed for *umc1149* and *phi029* in Populations II and III respectively. Estimates of outcrossing ( $t$ ) calculated from fixation index

**Table 2.** Pairwise genetic distance of  $F_{st}$  values (top diagonal), Nei's unbiased genetic distance (bottom diagonal) and mean values of genetic parameters for the three populations as revealed by the 38 polymorphic SSR markers under study

	Genetic distance			Mean and SE over loci for each pop						
	Population I	Population II	Population III	$N_e$	PA	$H_o$	$H_e$	$F$	$t$	%P
Population I		0.163	0.090	1.67 ± 0.06	8.00	0.18 ± 0.03	0.36 ± 0.02	0.47 ± 0.07	0.361	94.74
Population II	0.226		0.033	1.49 ± 0.06	21.00	0.09 ± 0.02	0.28 ± 0.03	0.66 ± 0.05	0.205	89.47
Population III	0.120	0.020		1.58 ± 0.04	10.00	0.13 ± 0.02	0.34 ± 0.02	0.59 ± 0.05	0.258	94.74

$N_e$ , number of lines;  $N_e$ , number of effective alleles ± standard error; PA, private alleles;  $H_o$ , observed heterozygosity ± standard error;  $H_e$ , expected heterozygosity ± standard error;  $F$ , fixation index;  $t$ , estimated out-crossing; %P, percentage of polymorphism.

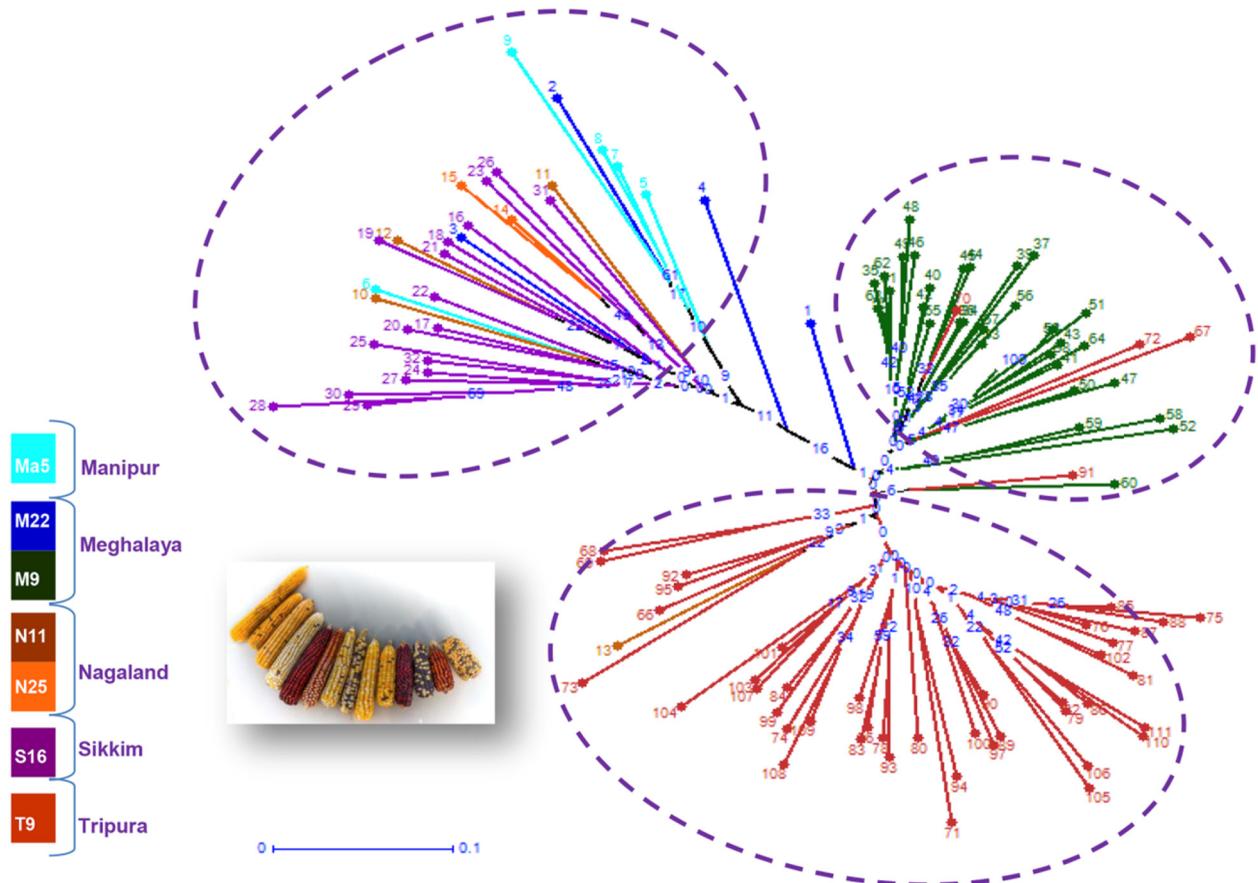
were also lowest in Population II (Table 2). Population I with the highest percentage of polymorphic loci (%P) recorded the lowest fixation index ( $F$ ) which ranged from 0.661 to 0.457 (online Supplementary Fig. S1). Based on SSR genotyping, the overall percent heterozygosity for the individual lines studied ranged from 0 to 30.23% with Population II recording the lowest percent heterozygosity (0 to 19.5%). The observed heterozygosity for the individual lines (online Supplementary Table S4) varied between 7.14 and 25.52% in Population I and between 2.44 and 30.23% was highest for the admixture group constituting Population III.

### Phylogenetic and phenotypic divergence

Of the 26 individuals grouped by STRUCTURE in Population I, all lines originating from landraces S16 and N25 were represented in this subgroup but none of T9 or M9 origin which were mostly clustered in Population II. Population III comprised of a mix of lines derived from M9, M22, Ma5, N11 and T9 but none from S16 and N25. Individual lines in Population I recorded the lowest ear weight, grain yield and also greatly reduced PH and EH when compared to mean values of Populations II and III. For the same traits, the admixture group recorded high values while Population II was the most variable with higher dispersion of individual trait values on either side of the mean (online Supplementary Fig. S2).

The dendrogram generated based on un-weighted neighbour joining also separated the lines studied into three distinct clusters (Fig. 4). Individuals other than M9 and T9 delineated as per STRUCTURE to Populations I and III grouped in a single cluster, while members of M9 and T9 concurrent to the subgrouping achieved as per STRUCTURE were found to occupy distinct clusters with a few exceptions. A minimum genetic distance of 0.049 was recorded for lines numbered 53 and 66 clustering within the M9 group and the maximum genetic distance of 0.587 was observed between lines numbered 28 and 105 originating from landraces S16 and T9, respectively. Maximum intra-cluster divergence (0.541) was observed in the cluster comprising members of Populations I and III. Despite originating from common progenitors, members of the cluster comprising M9 individuals recorded a maximum intra-cluster genetic distance of 0.332 while in case of T9, the maximum intra-cluster genetic distance recorded was 0.439.

With respect to the eight phenotypic traits studied, a highly significant and positive correlation at  $\alpha = 0.01$  was observed for all the traits, with higher correlations detected between EWH, EWWH and GYP when compared to NK and EL. Of the 80.87% total variation explained by the first three principal component (PC) analysis axes, ear traits EWH, EWWH and GYP contributed to 71.26% of the total



**Fig. 4.** Un-weighted neighbour joining tree based on dissimilarity matrix calculated from 38 SSR markers where each tip represents an individual line. The subgroups are indicated in different colours. Majority of the lines originating from M9 and T9 grouped into two distinct clusters with lines originating from landraces M22, Ma5, N11, N25 and S16 grouping into a separate cluster. Inset: A representative photo of maize landraces collected from various parts of NEHR.

variation in PC1. EWH ranged from a minimum of 25 g to a maximum of 195 g with a mean of 100 g while GYP was found to range from 3.1 to 104.3 g with a mean of 42.4 g in the 111 lines studied. The contributions of PH and EH which were highly significantly correlated with each other but not with ear/kernel traits or ASI, were highest in PC2 at 83.42%. ASI was the most variable trait under study ranging from 1 to 10 d and correlated negatively with other variables in PC3 contributing to 96.41% of the total variation accounted for in PC3. Squared cosine values indicated that EWH, EWWH and GYP were the most significant contributors to variation in PC1. Similarly, PH and EH were significant contributors to variation in PC2 and ASI in PC3 (online Supplementary Table S5).

## Discussion

Landraces which are the collective outcome of both natural and human selection constitute a valuable genetic resource for plant breeding (Reif *et al.*, 2005; Villa *et al.*, 2005; Mercer

and Perales, 2010; Casanas *et al.*, 2017). A designated centre of diversity, the NEHR is home to a vast collection of diverse maize landraces that have been maintained under an informal seed system by the tribal farmers of the region over time (Dhawan, 1964; Singh, 1977; Kumar and Sachan, 1994; Prasanna, 2012; Wasala and Prasanna, 2012). The allelic richness of these landraces are yet to be explored to their full potential (Singode and Prasanna, 2010) and the current study to characterize inbreds developed from the indigenous landraces of NEHR was one such attempt.

Genetic divergence studies help to define breeding strategies geared to address challenges of yield and an understanding of the same is a prerequisite for exploitation of heterosis (Yao *et al.*, 2007; Semagn *et al.*, 2012; Aci *et al.*, 2018). For the current subset, a significant increase in percent homozygosity ranging between 70 and 100% as against the expected theoretical average of approximately 88% (Hallauer *et al.*, 2010) on account of inbreeding leading to a deviation from Hardy–Weinberg expectations were observed. For the same SSR loci, moderate to high  $F_{st}$  values were recorded in 22 of the 38 markers implying

that these loci were favoured for genetic differentiation. Meirmans (2006) had reported that the presence of excess average heterozygotes tends to obscure genetic differentiation even in the presence of population structuring a view concurred by Aci *et al.* (2013) from their studies on genetic diversity of Algerian maize accessions. They had observed that loci with high/moderate  $F_{st}$  values contributed to increased genetic differentiation in a manner similar to what was obtained in the current study.

When extrapolated at the population level for defining natural subgroups based on maximum Hardy–Weinberg equilibrium and complete linkage equilibrium within populations (Pritchard *et al.*, 2000; Liu *et al.*, 2003), the 111 near homozygous inbred lines could be distinguished into two distinct and one admixture subgroup. As revealed by AMOVA and  $F_{st}$  coefficients significantly greater than zero at  $P < 0.001$ , the subgrouping was a result of 17% structuring among populations. Since AMOVA and classical  $F$  coefficient interpretations are analogous, both of which are based on the null hypothesis that the population under study is one large random mating entity (Mengoni and Bazzicalupo, 2002; Holsinger and Weir, 2009), an  $F_{st}$  value of 0.17 observed at the population level not only indicated a significant deviation from panmixis but also implied the presence of a high variation in allele frequency (Holsinger and Weir, 2009) between the differentiated subgroups. In fact, hierarchical splitting of the total molecular variance indicated that the subgroups were significantly differentiated at all three levels of stratification with the highest variation accounted for by the Among Individual (AI) within population component. High AI variance among individuals within a subgroup is generally the case when lines under study are derived from a broad genetic base (Reif *et al.*, 2003) and would be advantageous for developing novel allelic combinations in  $F_1$  hybrids for greater expression of heterosis (Springer and Stupar, 2007). The presence of population structuring in the current subset were further validated by a relatively low within individual (WI) component implying a low association between alleles within an individual relative to the entire population taken as a whole.

The distance-based neighbour joining cluster analysis method in a manner almost similar to grouping achieved by STRUCTURE also delineated majority of the individuals of Pop-M9 and Pop-T9 into two distinct clusters, while the remaining lines and the admixtures grouped into a third cluster which also recorded highest intra-cluster divergence. Where Nei's geographic distance and  $F_{st}$  values at the population level indicated that Populations I and II were more divergent, neighbour joining method clearly identified lines originating from landraces belonging to Population I (S16) and Population II (T9) as the ones with the maximum genetic distance between them. These results obtained using both model-based and distance-

based classification approaches revealed an almost similar pattern of population structuring clearly delineating the lines originating from certain landraces to be distinct from one another at the molecular level. Additionally, although over represented, intra-cluster divergence was also observed for individuals belonging to subpopulations Pop-M9 and Pop-T9 which had been revealed by PCoA to occupy distinct quadrants. The members of these subpopulations were also highly variable for the eight yield contributing traits studied. The divergence detected within and between the defined clusters can be attributed to have arisen from accumulated recombinations in the different lines as a result of four rounds of full sib mating prior to selfing. The process of full sib mating is advantageous over selfing in generating variability as it creates better opportunities for selection from crossing over in regions of the genome still heterozygous and give rise to new combinations (Rodrigues *et al.*, 2001; Lee and Kannenberg, 2004).

However, with most of the variation seen in indigenous maize of NEHR arising either from alterations due to human intervention of the four originally grouped races (Singh, 1977) or as a result of hybridization between them over time, moderate levels of genetic differentiation is expected and can be seen to range between 0.049 and 0.587 for the subset under study despite originating from varying altitudes. While the expression of heterosis is correlated with the genetic distance between the parental lines (Hochholdinger and Baldauf, 2018), moderately differentiated lines tend to result in higher heterosis for yield as opposed to extremely divergent parents (Moll *et al.*, 1965; Springer and Stupar, 2007). A partial diallel experiment (unpublished data) involving 10 randomly drawn individuals from the three populations registered better parent heterosis for all of the highly variable ear-related traits in the 41 viable single cross hybrids (H1 to H41) which were evaluated. Encouraging results against the commercial checks for the inter-population crosses were also observed (online Supplementary Table S6) particularly in hybrids H3 and H19 which recorded highly significant superior grain yield/plant over all the three commercial checks in the second sowing window. The parental lines of both hybrids belonged to the moderately differentiated Populations I and III as delineated by STRUCTURE and had recorded genetic distances of 0.41 and 0.47 respectively when using the distance-based clustering approach. Also, while lines in Population I had recorded the lowest values for the highly variable ear-related traits, Population III comprised of individuals with highest values for the same. These results allow us to believe that the defined subgroups which appear to be divergent both at molecular and morphological levels can provide a suitable germplasm base for heterotic grouping.

The high variation observed for ear-related traits in the current study is already known to have a genetic basis for

maize landraces of North East India (Sanjenbam *et al.*, 2018) as a result of seed selection habits of farmers which focuses mainly on ear characteristics as part of their socio-cultural requirements (Prasanna, 2012). In such informal systems of selection, seed management at the individual level also plays a key role in generating variability influenced by (a) sample size, which has a bearing on genetic drift and (b) selection decisions, which are almost always tilted towards the most vigorous plants/cobs and promote heterogeneity (Pressoir and Berthaud, 2004; Bellon and Van Etten, 2013). Over time, a significant amount of individuality within these landraces are generated which continue to be maintained due to space isolation (Berthaud and Gepts, 2004). However, although such a seed selection process maintained under a system of open pollination helps in evolution of the conserved traditional landraces, since it is informal in nature, morphological traits other than seed characteristics may or may not improve in the desired direction of selection. Therefore, interventions by plant breeders involving scientific selection methods become imperative for generating improved varieties (Louette and Smale, 2000).

When such interventions involve utilization of genetically variable landraces as a potential source of inbred development, a successful single cross hybrid development would depend on the identification and utilization of heterotic groups and patterns (Melani and Carena, 2005) where heterotic groups are a collection of related inbreds and presence of genetic variation is fundamental to such grouping (Melchinger and Gumber, 1998; Reif *et al.*, 2003; Semagn *et al.*, 2012). With molecular markers efficiently predicting genetic divergence, there is a general consensus today that such markers can effectively assign individuals to respective heterotic groups (Melchinger *et al.*, 1991; Barbosa *et al.*, 2003) at lower costs (Fernandez *et al.*, 2015; Punya *et al.*, 2018). Heterotic grouping using molecular markers also allows greater number of lines to be evaluated as opposed to traditional methods and are vital in situations where heterotic patterns for developing single cross hybrids are not well established (Gichuru *et al.*, 2016), such as in the current study.

While we seek to take advantage of the locally available diversity for developing a robust hybrid breeding programme, the fear of genetic erosion with the introduction of hybrids also needs to be addressed. Studies have shown that despite introduction of hybrids, cultivation of traditional landraces continues to thrive in areas of crop diversity since it is community based, depends on an informal seed exchange system and is strongly dictated by cultural preferences of the indigenous people involved (Bellon and Hellin, 2011; Fenzi *et al.*, 2015). Also, at the individual level, depending on the farmers' perception there is always a tradeoff between utilization of genetic resources (local/improved varieties) available at his/her disposal.

While hybrids with higher responsiveness to fertilizers are generally grown under favourable ecological conditions, the hardy landraces known to give higher returns under non-optimal situations are preferred in marginal environments (Ficiciyan *et al.*, 2018). Additionally, once heterotic groups are defined, the knowledge of the same can also be utilized for developing cost effective synthetics/composites with improved agronomic traits. Such populations when developed under conditions of open pollination will not only stem genetic drift but also generate new recombinations. Under these circumstances, the genetic diversity may even increase should the improved germplasm be more heterogeneous than the traditional landraces (van Heerwaarden *et al.*, 2009).

## Conclusion

For our current study, genetic analysis of this subset of inbred maize lines originating from different altitudes of NEHR of India using both model and distance-based clustering approaches showed an almost concurrent pattern of population structuring. Each defined population was also associated with a definite pattern of morphological traits leading us to conclude that these populations are distinct. Taking advantage of the promising results of the partial diallel analysis for heterosis, defining heterotic groups of the indigenous NEHR maize lines would be the next step forward. Once achieved, such a grouping would help initiate a productive breeding programme geared to address challenges of yield while utilizing the rich genetic resources local to this centre of Asiatic maize diversity.

## Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262120000246>.

## Acknowledgements

The authors are extremely grateful to the farmers of NEHR of India from whom the landraces for the current study were procured for their generosity. The authors are also grateful to the Central Agricultural University (Imphal), for providing the necessary infrastructure for completing all field and laboratory work and to the Jawaharlal Nehru Memorial Fund (JNMF), New Delhi, for providing financial assistance during the course of the research work. Thanks are also due to Dr Wricha Tyagi, Professor, School of Crop Improvement, for her generous help with the molecular work done. The help rendered by ICAR-NBPGR Regional Station Umroi, India for advancing the generations is also gratefully acknowledged.

## Conflict of interest

None.

## References

- Aci MM, Revilla P, Morsli A, Djemel A, Belalia N, Kadri Y, Khelifi-Saloui M, Ordas B and Khelifi L (2013) Genetic diversity in Algerian maize (*Zea mays* L.) landraces using SSR markers. *Maydica* 58: 304–310.
- Aci MM, Lupini A, Mauceri A, Morsli A, Khelifi L and Sunseri F (2018) Genetic variation and structure of maize populations from Saoura and Gourara Oasis in Algerian Sahara. *BMC Genetics* 19: 1–10.
- Anonymous (2007) Guidelines for the conduct of tests for distinctness, uniformity and stability on maize (*Zea mays* L.). The Protection of Plant Varieties and Farmers Rights Authority, New Delhi. *Plant Variety Journal India* 1: 1–13.
- Barbosa AM, Geraldi IO, Benchimol LL, Garcia AA, Souza CL and Souza AP (2003) Relationship of intra- and interpopulation tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. *Euphytica* 130: 87–99.
- Bedoya CA, Dreisigacker S, Hearne S, Franco J, Mir C, Prasanna BM, Taba S, Charcosset A and Warburton ML (2017) Genetic diversity and population structure of native maize populations in Latin America and the Caribbean. *PLoS ONE* 12: e0173488.
- Bellon MR and Hellin J (2011) Planting hybrids, keeping landraces: agricultural modernization and tradition among small-scale maize farmers in Chiapas, Mexico. *World Development* 39: 1434–1443.
- Bellon MR and van Etten J (2013) Climate change and on-farm conservation of crop landraces in centres of diversity. In: Jackson M, Ford-Lloyd B and Parry ML (eds) *Plant Genetic Resources and Climate Change*. CABI Publishing, Wallingford, UK, pp. 137–150.
- Berthaud J and Gepts P (2004) Assessment of effects on genetic diversity. In: CEC Secretariat of the Governments of Canada, Mexico and of the United States (eds.) *Maize and Biodiversity: The Effects of Transgenic Maize in Mexico*. Montreal: North American Commission on Environmental Cooperation, pp. 1–33.
- Bhagat D, Mishra BK and Priyamvada S (2015) Studies on changing livestock products scenario in Meghalaya. *Journal of Hill Agriculture* 6: 123–127.
- Casanas F, Simo J, Casals J and Prohens J (2017) Toward an evolved concept of landrace. *Frontiers in Plant Science* 8: 145.
- Charlesworth D and Willis JH (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* 10: 783–796.
- Choukan, R, Hossainzadeh, A, Ghannadha, MR, Warburton, ML, Talei, AR and Mohammadi SA (2006) Use of SSR data to determine relationships and potential heterotic groupings within medium to late maturing Iranian maize inbred lines. *Field Crops Research* 95: 212–222.
- Dass S, Kaul J, Manivannan A and Chikkappa GK (2009) Single cross hybrid maize – a viable solution in the changing climate scenario. *Indian Journal of Genetics and Plant Breeding* 69: 331–334.
- Dass S, Kumar A, Jat SL, Parihar CM, Singh AK, Chikkappa GK and Jat ML (2012) Maize holds potential for diversification and livelihood security. *Indian Journal of Agronomy* 57: 32–37.
- Dhawan NL (1964) Primitive maize in Sikkim. *Maize Genetics Co-op Newsletter* 38: 69–70.
- Doyle JF and Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 39–40.
- Duvick DN (2001) Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics* 2: 69–74.
- Duvick DN (2005) The contribution of breeding to yield advances in maize (*Zea mays* L.). *Advances in Agronomy*, 86: 83–145.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611–2620.
- FAO (2018) World Food and Agriculture – Statistical Pocketbook 2018. Rome, pp. 28, Licence: CC BY-NC-SA 3.0 IGO.
- FAO (2019) Voluntary Guidelines for the Conservation and Sustainable Use of Farmers' Varieties/Landraces. Rome.
- Fenzi M, Jarvis DI, Reyes LMA, Moreno LL and Tuxill J (2015) Longitudinal analysis of maize diversity in Yucatan, Mexico: influence of agro-ecological factors on landraces conservation and modern variety introduction. *Plant Genetic Resources* 15: 51–63.
- Fernandez EH, Schuster I, Scapim CA, Vieira ES and Coan MM (2015) Genetic diversity in elite inbred lines of maize and its association with heterosis. *Genetics and Molecular Research* 14: 6509–6517.
- Feroze SM, Raju VT, Singh R and Tripathi AK (2010) Status of livestock sector: a micro study of North Eastern India. *Indian Journal of Hill Farming* 23: 43–51.
- Ficiciyan A, Loos J, Sievers-Glotzbach S and Tschamtker T (2018) More than yield: ecosystem services of traditional versus modern crop varieties revisited. *Sustainability* 10: 2834.
- Gichuru L, Derera J, Tongoona P and Murenga M (2016) Combining ability and molecular genetic distances in tropical maize lines for grain yield and agronomic traits. *South African Journal of Plant and Soil* 34: 175–183.
- Grogan P, Lalnunmawia F and Tripathi SK (2012) Shifting cultivation in steeply sloped regions: a review of management options and research priorities for Mizoram state, Northeast India. *Agroforestry Systems* 84: 163–177.
- Hallauer AR, Carena MJ and Miranda Filho JD (2010) *Quantitative Genetics in Maize Breeding*, vol. 6. New York: Springer Science & Business Media, p. 432.
- Hochholdinger F and Baldauf JA (2018) Heterosis in plants. *Current Biology* 28: R1089–R1092.
- Holsinger KE and Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting  $F_{ST}$ . *Nature Reviews Genetics* 10: 639–650.
- Hossain F, Muthusamy V, Bhat JS, Jha SK, Zunjare R, Das A, Sarika K and Kumar R (2016) Maize. In: Singh M and Kumar S (eds) *Broadening the Genetic Base of Grain Cereals*. New Delhi: Springer, pp. 67–88.
- Kumar M and Sachan JK (1994) Cytological basis of racial diversity in Indian maize: an overview. *Indian Journal of Genetics and Plant Breeding* 54: 409–417.
- Kumar R, Srinivas K and Sivaramane N (2013) Assessment of the maize situation, outlook and investment opportunities in India. Country Report – Regional Assessment Asia (MAIZE-CRP), National Academy of Agricultural Research Management, Hyderabad, India.
- Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M and Jiao Y (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature Genetics* 42: 1027–1030.
- Lee EA and Kannenberg LW (2004) Effect of inbreeding method and selection criteria on inbred and hybrid performance. *Maydica* 49: 191–197.
- Lia VV, Poggio L and Confalonieri VA (2009) Microsatellite variation in maize landraces from North Western Argentina:

- genetic diversity, population structure and racial affiliations. *Theoretical and Applied Genetics* 119: 1053–1067.
- Liu K, Goodman M, Muse S, Smith JS, Buckler E and Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117–2128.
- Louette D and Smale M (2000) Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. *Euphytica* 113: 25–24.
- Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60 : 2399–2402.
- Melani MD and Carena MJ (2005) Alternative maize heterotic patterns for the northern corn belt. *Crop Science* 45: 2186–2194.
- Melchinger AE and Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops. In: Lamkey KR and Staub JE (eds) *Concepts and Breeding of Heterosis in Crop Plants*. Madison, WI: CSSA Special Publication, pp. 29–44.
- Melchinger AE, Messmer MM, Lee M, Woodman WL and Lamkey KR (1991) Diversity and relationships among US maize inbreds revealed by restriction fragment length polymorphisms. *Crop Science* 31: 669–678.
- Mengoni A and Bazzicalupo M (2002) The statistical treatment of data and the Analysis of Molecular Variance (AMOVA) in molecular microbial ecology. *Annals of Microbiology* 52: 95–101.
- Mercer KL and Perales HR (2010) Evolutionary response of landraces to climate change in centers of crop diversity. *Evolutionary Applications* 3: 480–493.
- Moll RH, Lonquist JH, Fortuno JV and Johnson EC (1965) The relationship of heterosis and genetic divergence in maize. *Genetics* 52: 139.
- Morrell PL, Buckler ES and Ross-Ibarra, J (2012) Crop genomics: advances and applications. *Nature Reviews Genetics* 13: 85–96.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70: 3321–3323.
- Nyaligwa L, Hussein S, Amelwork B and Ghebrehiwot H (2015) Genetic diversity analysis of elite maize inbred lines of diverse sources using SSR markers. *Maydica* 60 : 1–8.
- Pavithra S, Boeber C, Shah SA, Subash SP, Birthal PS and Mittal S (2018) Adoption of modern maize varieties in India: insights based on expert elicitation methodology. *Agricultural Research* 7: 391–401.
- Peakall RO and Smouse PE (2012) Genalex 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. *Bioinformatics (Oxford, England)* 28: 2537–2539.
- Portwood JL, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, Braun BL, Dietze M, Dunfee B, Elisk CG, Manchanda N, Coe E, Sachs M, Stinard P, Tolbert J, Zimmerman S and Andorf CM (2018) Maize GDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Research* 47: 1146–1154.
- Prasanna BM (2010) Phenotypic and molecular diversity of maize landraces: characterization and utilization. *Indian Journal of Genetics and Plant Breeding* 70: 315–327.
- Prasanna BM (2012) Diversity in global maize germplasm: characterization and utilization. *Journal of Bioscience* 37: 843–855.
- Pressoir G and Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92: 95–101.
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Punya, Sharma VK and Kumar P (2018) Microsatellite marker dependent genetic divergence assessment within and among heterotic groups of tropical maize inbred lines. *Maydica* 63: 13.
- Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P and Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. *Theoretical and Applied Genetics* 106: 895–903.
- Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Srinivasan G, Bohn M and Frisch M (2003) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Science* 43: 1275–1282.
- Reif JC, Hamrit S, Heckenberger M, Schipprack W, Maurer HP, Bohn M and Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. *Theoretical and Applied Genetics* 111: 906–913.
- Rodrigues MC, Valva FD, Brasil EM and Chaves LJ (2001) Comparison among inbreeding systems in maize. *Crop Breeding and Applied Biotechnology* 1: 105–114.
- Romay MC, Butron A, Ordas A, Revilla P and Ordas B (2012) Effect of recurrent selection on the genetic structure of two broad-based Spanish maize populations. *Crop Science* 52: 1493–1502.
- Sanjenbam M, Sen D, Tyagi W and Chand R (2018) Phenotypic diversity analysis and screening for northern corn leaf blight resistance in maize (*Zea mays* L.) landraces grown in North Eastern hill region of India. *Indian Journal of Plant Genetic Resources* 31: 286–294.
- Semagn K, Magorokosho C, Vivek BS, Makumbi D, Beyene Y, Mugo S, Prasanna BM and Warburton ML (2012) Molecular characterization of diverse CIMMYT maize inbred lines from Eastern and Southern Africa using single nucleotide polymorphic markers. *BMC Genomics* 13: 113.
- Sharma SK and Brahma P (2011) Gene bank curators: towards implementation of the international treaty on plant genetic resources for food and agriculture by the Indian National Gene Bank. In: Frison C, López F and Esquinas-Alcázar JT (eds) *Plant Genetic Resources and Food Security: Stakeholder Perspectives on the International Treaty on Plant Genetic Resources for Food and Agriculture*. Oxon: FAO and Bioversity International, pp. 183–196.
- Sharma L, Prasanna BM and Ramesh B (2010) Analysis of phenotypic and microsatellite-based diversity of maize landraces in India, especially from the north east Himalayan region. *Genetica* 138: 619–631.
- Singh B (1977) *Races of Maize in India*. New Delhi: Indian Council of Agricultural Research.
- Singode A and Prasanna BM (2010) Analysis of genetic diversity in the North Eastern Himalayan maize landraces using microsatellite markers. *Journal of Plant Biochemistry and Biotechnology* 19: 33–41.
- Springer NM and Stupar RM (2007) Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Research* 17: 264–275.
- Tripathi AK, Singh SB and Datta KK (2003) Prospects, problems and scope of maize and pulses in north eastern hill region of India. *Indian Journal of Hill Farming* 16: 48–52.

- van Heerwaarden J, Hellin J, Visser RF and van Eeuw FA (2009) Estimating maize genetic erosion in modernized smallholder agriculture. *Theoretical and Applied Genetics* 119: 875–888.
- Villa TCC, Maxted N, Scholten M and Ford-Lloyd B (2005) Defining and identifying crop landraces. *Plant Genetic Resources* 3: 373–384.
- Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J and Hufford MB (2017) The interplay of demography and selection during maize domestication and expansion. *Genome Biology* 18: 215.
- Wasala SK and Prasanna BM (2012) Microsatellite marker-based diversity and population genetic analysis of selected lowland and mid-altitude maize landrace accessions of India. *Journal of Plant Biochemistry Biotechnology* 22: 392–400.
- Yang Q, Balint-Kurti P and Xu M (2017) Quantitative disease resistance: dissection and adoption in maize. *Molecular Plant* 10: 402–413.
- Yao Q, Yang K, Pan G and Rong T (2007) Genetic diversity of maize (*Zea mays* L.) landraces from Southwest China based on SSR data. *Journal of Genetic and Genomics* 34: 851–860.