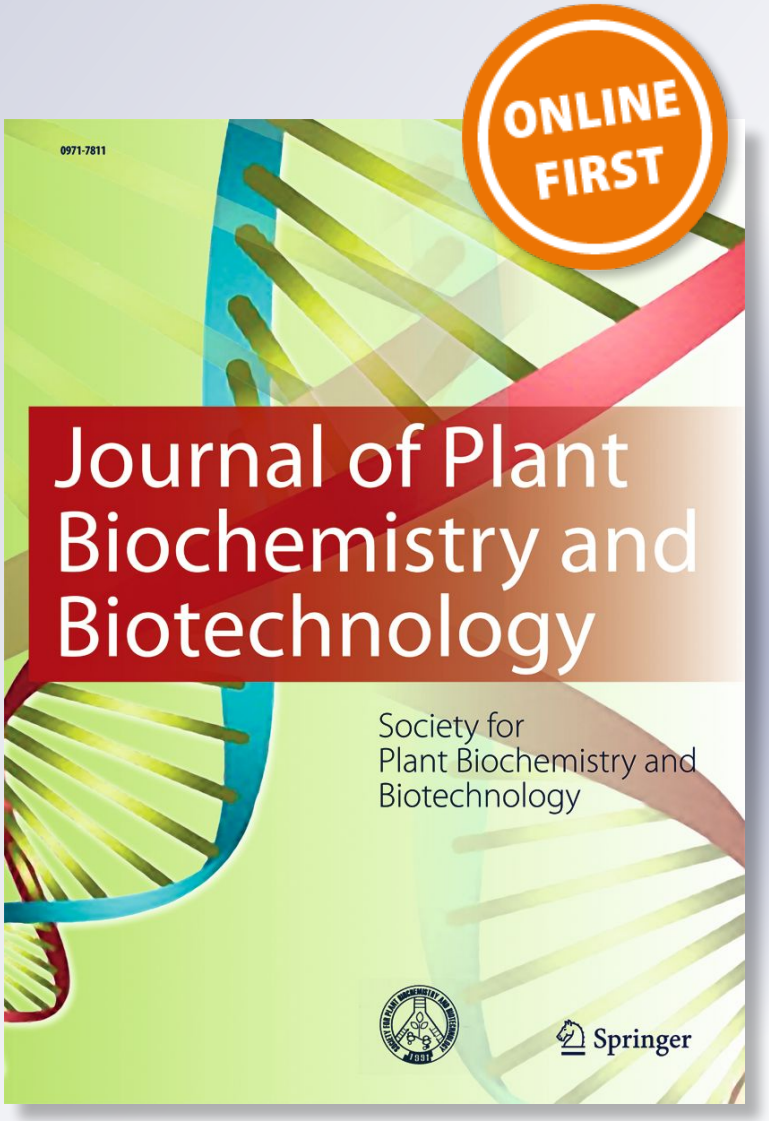
Microsatellite marker-based diversity and population genetic analysis of selected lowland and mid-altitude maize landrace accessions of India

Samanthi K. Wasala & B. M. Prasanna

Journal of Plant Biochemistry and Biotechnology

ISSN 0971-7811

J. Plant Biochem. Biotechnol.
DOI 10.1007/s13562-012-0167-5



Springer

J. Plant Biochem. Biotechnol.
DOI 10.1007/s13562-012-0167-5

ORIGINAL ARTICLE

Microsatellite marker-based diversity and population genetic analysis of selected lowland and mid-altitude maize landrace accessions of India

Samanthi K. Wasala · B. M. Prasanna

Received: 22 February 2012 / Accepted: 10 October 2012

Abstract Maize (*Zea mays* L.) harbours significant genetic diversity not only in its centre of origin (Mexico) but also in several countries worldwide, including India, in the form of landraces. In this study, DNA fingerprinting of 48 landrace accessions from diverse regions of India was undertaken using 42 fluorescent dye-labeled Simple Sequence Repeat (SSR) markers, followed by allele resolution using DNA sequencer and analysis of molecular diversity within and among these landraces. The study revealed a large number of alleles (550), with high mean number of alleles per locus (13.1), and Polymorphism Information Content (PIC) of 0.60, reflecting the level of diversity in the landrace accessions. Besides identification of 174 unique alleles in 44 accessions, six highly frequent SSR alleles were detected at six loci (*phi014*, *phi090*, *phi112*, *umc1367*, *phi062* and *umc1266*) with individual frequencies greater than 0.75, indicating that chromosomal regions harboring these SSR alleles are not selectively neutral. F statistics revealed very high genetic differentiation, population subdivision and varying levels of inbreeding in the landraces. Analysis of Molecular Variance showed that 63 % of the total variation in the accessions could be attributed to within-population diversity, and 37 % represented between population diversity. Cluster analysis of SSR data using Nei's genetic distance and UPGMA revealed considerable genetic diversity in these populations, although no clear separation of accessions was observed based on their geographic origin.

Keywords Genetic diversity · India · Landraces · Population structure · *Zea mays* L

Abbreviations

PIC	Polymorphism Information Content
AMOVA	Analysis of Molecular Variance
UPGMA	Unweighted Pair Group Method using Arithmetic means
H_o	Observed heterozygosity
H_e	Expected heterozygosity
F	Inbreeding coefficient
A	Number of alleles per locus
A_e	Effective number of alleles
GD	Genetic distances
SSR	Simple sequence repeat
SNP	Single nucleotide polymorphism

Introduction

Analysis of the levels and pattern of genetic diversity in crop plants is invaluable as it opens up the possibilities for systematic management, conservation, introgression of desirable genes from diverse genetic sources into the elite material, as well as derivation of novel and useful germplasm. Maize (*Zea mays* ssp. *mays*) was domesticated from its wild species ancestor, teosinte (*Zea mays* ssp. *parviglumis*) about 9,000 years ago. This domestication event took place in the mid-elevations (~1,500 m above sea level, masl)

S. K. Wasala · B. M. Prasanna
Indian Agricultural Research Institute,
New Delhi 110012, India

Present Address:
S. K. Wasala
Plant Genetic Resource Centre,
Gannoruwa, Sri Lanka

Present Address:
B. M. Prasanna (✉)
International Maize and Wheat Improvement Center (CIMMYT),
ICRAF House,
United Nations Avenue, Gigiri,
Nairobi, Kenya
e-mail: b.m.prasanna@cgiar.org

Published online: 06 November 2012

of South Central Mexico, and occurred once starting with the teosinte race Balsas (Matsuoka et al. 2002). Maize then followed a very complicated pattern of introduction out of Mexico into different countries spread across the world, including the North and South America, Europe, Africa and Asia (Rebourg et al. 2003; Dubreuil et al. 2006; Marilyn Warburton, personal communication). Most of such introductions happened several centuries ago, and maize landraces have been selected by the local farmers to adapt these in the new environments, leading to several new derivatives in the process, with significant variation in characteristics like prolificacy, flowering behavior, yield, nutritive value and resistance to biotic and abiotic stresses. However, much of the diversity in maize landraces remains to be efficiently explored and utilized. For better utilization of such gene pools, it is important to generate proper agronomic and genetic knowledge (Nass et al. 1993).

Extensive variability in plant, ear, and tassel characteristics was observed in the North eastern and Northwestern highlands of India (Singh 1977). While considerable attention has been laid on phenotypic and molecular characterization of landraces from the North East Himalayan (NEH) region in India (Singh 1977; Sharma et al. 2010), very little efforts have been made with regard to understanding of diversity of maize landraces from other regions in India (Prasanna and Sharma 2005; Prasanna 2010). Survey of varietal distribution in the major maize growing areas revealed that landraces or local cultivars occupy significant areas in Madhya Pradesh (82 %), Uttar Pradesh (42 %), and Bihar (45 %) during the rainy (*Kharif*) season, while the acreage during winter (*Rabi*) season is comparatively lower (Joshi et al. 2005).

Molecular markers can support detailed, reliable and unambiguous characterization of genetic resources. A vast potential lies in their ability to identify the structure of genetic diversity within and among accessions. The microsatellite or simple sequence repeat (SSR) markers are widely used in maize, as these markers are mapped, PCR-based, genetically codominant, hypervariable, highly polymorphic, robust, reproducible and amenable to automation (Warburton et al. 2002; Dubreuil et al. 2006; Prasanna et al. 2010). The 'population bulk DNA fingerprinting' method, using fluorescent dye-labelled SSR markers and resolution of PCR amplification products through a DNA Sequencer, has been optimized and utilized for studying patterns of genetic diversity in maize (Rebourg et al. 2001; Warburton et al. 2002, 2011; Rebourg et al. 2003; Dubreuil et al. 2006; Sharma et al. 2010), including precise resolution of alleles and estimation of allele frequencies using advanced software. The objective of the present study was to analyze the levels and patterns of molecular diversity in selected landrace accessions from diverse maize-growing regions in India, especially the accessions collected from the non-NEH regions, using microsatellite markers.

Materials and methods

Genetic material

A set of 48 landrace accessions from diverse agro-ecologies in India, excluding the North East Himalayan (NEH) region, was selected as experimental materials for this study. Of these, 29 accessions were obtained from the National Gene Bank, National Bureau of Plant Genetic Resources (NBPGR), New Delhi, India (Table 1; S.No. 1–29), and the rest were collected in November 2005 by the Maize Genetics Unit, IARI, under the Indian Council of Agricultural Research (ICAR) National Fellow Project. Details of these accessions are provided in Table 1. The collection sites ranged from 9° to 34° North longitude and 71° to 87° East latitude, with elevations of the sites of collections varying from 21 masl (IML134 from Gujarat) to 2,621 masl (IML426 from Uttarakhand). Thus, the selected accessions included 28 lowland (<400 masl), 16 mid-altitude (400–1,500 masl) and 4 highland (>1,500 masl) accessions.

DNA analysis

DNA samples from each of the 48 accessions were isolated using the 'population DNA bulk strategy' (Rebourg et al. 2001, 2003). Each bulk was prepared by pooling an equal amount of leaf material from 15 individuals in a population. The bulk strategy allows the estimation of within-population allele frequencies accurately, besides facilitating large-scale analyses of diverse maize populations (Dubreuil et al. 2006). DNA isolation was done after the pooling step, using a modified CTAB procedure according to CIMMYT Applied Biotechnology Center's Manual of Laboratory Procedures (available at http://www.cimmyt.org/ABC/Protocols/manualABC.html).

The 48 accessions were genotyped with a set of 42 fluorescent dye-labeled SSR markers, the PCR conditions of which were optimized under a Generation Challenge Program (GCP) International Collaborative Project on Global Maize Diversity Analysis (GCP15). The SSR loci were mostly chosen with tri- or higher repeat motifs, since such loci were shown to be less sensitive to stuttering phenomena than those with di-repeat motif (Perlin et al. 1995; Daniels et al. 1998). The forward primers were fluorescent dye-labeled (at 5′ end) with either 6-carboxyfluorescein (6-FAM), tetrachloro-6-carboxyfluorescein (TET), or hexachloro-6-carboxyfluorescein (HEX) (Sigma-Aldrich, India). Sequences of the SSR primers used in this study are available in public domain (MaizeGDB; http://www.maizegdb.org). Details of the SSR loci are presented in Table 2.

The PCR amplified products were resolved using a DNA Sequencer (MegaBACE 1000). SSR allele sizes were determined based on migration relative to the internal lane size

Table 1 Details of the selected maize landrace accessions used for the study

S. No.	Gene Bank No.	Code[a]	Village	District	State in India	Elevation (masl)[b]	Kernel type[c]
1	IC 77418	IML118	Jagadhari	Ambala	Haryana	277	Y/W F
2	IC 77433	IML119	Jaisinghpura	Ambala	Haryana	277	Y F
3	IC 77469	IML123	Shamli Mundate	Muzaffarnagar	Uttar Pradesh	239	O F
4	IC 77477	IML124	Bakshi Khola	Almora	Uttarakhand	1468	W/Y F
5	IC 77604	IML133	Akhadra Tonk	Tonk	Rajasthan	303	Y F
6	IC 77611	IML134	Jambuda	Junagarh	Gujarat	21	Y/O F
7	IC 82097	IML143	Chidiyawsa	Banswara	Rajasthan	243	O F
8	IC 97997	IML172	Dharamshala	Hamirpur	Himachal Pradesh	850	O/Y F
9	IC 108151	IML185	Nora	Mandi	Himachal Pradesh	1524	O F
10	IC 199114	IML288	Ramchandrapur	Bhagalpur	Bihar	48	O F
11	IC 199124	IML294	Parham	Munger	Bihar	46	O F
12	IC 199157	IML296	Sonka	Darbhanga	Bihar	55	Y/W
13	IC 251267	IML297	Paheli Shepti	Bankura	West Bengal	91	O F
14	IC 283422	IML321	Bharirwar	Begusarai	Bihar	44	Y F
15	IC 319791	IML324	Jharwar	Mandi	Himachal Pradesh	1524	O F
16	IC 325937	IML325	Dondachia	Nandurbar	Maharashtra	250	Y/O D
17	IC 331544	IML330	Khedarpura	Vaishali	Bihar	54	Y/W F
18	IC 331594	IML331	Sanhoula	Gopalganj	Bihar	71	O/Y F
19	IC 332300	IML332	Gandhinagar	Khargaon	Madhya Pradesh	240	O F
20	IC 395739	IML339	Ramgarh	Panchmahal	Gujarat	128	O D/F
21	–	IML340	Jaunpur	Jaunpur	Uttar Pradesh	81	O F
22	–	IML341	Jhabua	Jhabua	Madhya Pradesh	397	O F
23	–	IML342	Dausa	Dausa	Rajasthan	335	O F
24	–	IML346	Arabhavi	Belgaum	Karnataka	641	Y/W F
25	IC 262988	IML404	Behera	Sonbhadra	Uttar Pradesh	280	Y F
26	IC 263215	IML405	Gazar	Almora	Uttarakhand	1468	O F
27	IC 267464	IML409	Bhusuria	Daltonganj	Jharkhand	426	W/Y F
28	IC 273246	IML413	Sundrel	Khargaon	Madhya Pradesh	240	Y F
29	IC 273281	IML416	Mohan Katia	Dhar	Madhya Pradesh	505	Y/O F
30	IC 274642	IML422	Gada Dongri	Sagar	Madhya Pradesh	515	O F
31	IC 279809	IML426	Naganichhoti	Uttarkashi	Uttarakhand	2621	O/Y F
32	IC 282536	IML431	Karimnagar	Palamau	Jharkhand	426	O F
33	IC 298558	IML434	Sompur	Sabarkantha	Gujarat	38	O F
34	IC 309931	IML436	Lyoagwada	Alwar	Rajasthan	281	O F
35	IC 332276	IML452	Baralyakhurd	Indore	Madhya Pradesh	531	Y/W F
36	IC 333247	IML454	Jamli	Barawani	Madhya Pradesh	415	O F
37	IC 337439	IML459	Ratan pura	Vadodara	Gujarat	43	Y D
38	IC 342653	IML462	Jamukhandi	Kodarma	Jharkhand	325	O F
39	IC 353812	IML467	Orgoundanurpudur	Thiruvannamalai	Tamil Nadu	481	Y/O F
40	IC 370494	IML474	Jhadoli	Sirohi	Rajasthan	471	Y/W F
41	IC 373213	IML475	Jatabal	Nabarangpur	Orissa	922	O/Y F
42	IC 374676	IML476	Gund	Srinagar	Jammu & Kashmir	1837	O D
43	IC 385875	IML480	Sahargaon	Pakaur	Jharkhand	274	O F
44	IC 397864	IML485	Kanfara	Una	Himachal Pradesh	561	O F
45	IC 430635	IML496	Bandarugudem	Khammam	Andhra Pradesh	114	O F
46	IC 436850	IML498	Pittaguda	Adilabad	Andhra Pradesh	283	O F
47	IC565898	IML499	Kadanuru	Bangalore Rural	Karnataka	824	Y Pc
48	–	IML505	Kosipur	Nainital	Uttarakhand	1050	Y F

Longitude and latitude data available from the corresponding author on request

[a] *IML* (Indian Maize Landrace) number given for convenience in presentation

[b] *masl* meters above sea level

[c] *Y* Yellow; *O* Orange; *W* White; *F* Flint; *D* Dent; *Pc* Pop corn

Table 2 Summary of SSR polymorphisms and *F*-statistics across the accessions

SSR locus	Bin	*A*	A_e	Major allele freq.*	PIC	H_o	H_e	*F*	F_{IT}	F_{ST}
umc2047	1.09	9	2.63	0.58	0.59	0.45	0.62	0.27	0.42	0.44
phi308707	1.10	23	6.81	0.23	0.84	0.75	0.85	0.12	0.12	0.16
phi227562	1.11	8	4.02	0.43	0.72	0.47	0.75	0.37	0.43	0.44
phi96100	2.01	23	6.89	0.21	0.84	0.64	0.85	0.25	0.31	0.34
phi083	2.04	26	10.66	0.20	0.90	0.58	0.91	0.36	0.36	0.38
umc2250	2.04	16	2.06	0.69	0.51	0.14	0.51	0.74	0.76	0.79
nc133	2.05	5	2.40	0.57	0.52	0.33	0.58	0.44	0.44	0.47
phi084	2.08	12	2.91	0.55	0.63	0.26	0.66	0.60	0.61	0.58
phi127	2.08	17	2.82	0.58	0.63	0.26	0.64	0.59	0.61	0.63
phi102228	3.04	8	2.53	0.54	0.54	0.34	0.60	0.43	0.46	0.47
phi029	3.04	11	2.39	0.59	0.53	0.38	0.58	0.34	0.37	0.34
phi093	3.05	10	4.77	0.36	0.77	0.34	0.79	0.57	0.68	0.68
umc1266	3.06	8	1.26	0.89	0.20	0.13	0.21	0.39	0.49	0.50
phi046	3.08	9	2.23	0.65	0.53	0.28	0.55	0.49	0.52	0.55
phi079	4.05	13	6.08	0.25	0.82	0.59	0.83	0.30	0.30	0.35
phi090	4.08	11	1.52	0.81	0.33	0.13	0.34	0.61	0.67	0.66
phi076	4.11	7	2.85	0.52	0.60	0.51	0.65	0.21	0.27	0.30
nc130	5.00	8	2.18	0.62	0.48	0.41	0.54	0.25	0.33	0.33
phi109188	5.03	16	2.15	0.67	0.52	0.41	0.53	0.24	0.28	0.30
umc1447	5.03	5	2.39	0.58	0.53	0.42	0.58	0.28	0.34	0.36
phi331888	5.04	13	7.27	0.24	0.85	0.74	0.86	0.15	0.15	0.19
umc1332	5.04	15	3.68	0.40	0.69	0.41	0.73	0.44	0.48	0.48
phi075	6.00	11	2.73	0.50	0.57	0.54	0.63	0.14	0.14	0.20
phi031	6.04	19	7.78	0.28	0.86	0.61	0.87	0.31	0.38	0.38
phi123	6.07	14	3.25	0.46	0.65	0.50	0.69	0.28	0.28	0.31
phi299852	6.07	23	5.57	0.35	0.80	0.50	0.82	0.39	0.44	0.43
umc1545	7.00	16	4.64	0.36	0.76	0.39	0.78	0.50	0.50	0.50
phi112	7.01	18	1.51	0.81	0.33	0.15	0.34	0.56	0.72	0.73
phi034	7.02	8	4.33	0.32	0.74	0.38	0.77	0.51	0.56	0.55
umc1304	8.02	19	3.07	0.47	0.63	0.29	0.67	0.57	0.59	0.60
phi115	8.03	14	2.31	0.54	0.49	0.39	0.57	0.32	0.42	0.41
phi014	8.05	7	1.55	0.79	0.33	0.27	0.35	0.24	0.36	0.38
umc1161	8.06	17	3.44	0.50	0.69	0.41	0.71	0.42	0.42	0.45
phi065	9.03	12	5.00	0.25	0.77	0.36	0.80	0.55	0.56	0.58
umc1917	9.03	10	3.06	0.52	0.64	0.45	0.67	0.33	0.40	0.45
phi108411	9.05	10	2.77	0.54	0.60	0.40	0.64	0.37	0.39	0.41
phi041	10.00	11	2.47	0.61	0.57	0.33	0.59	0.45	0.47	0.52
phi059	10.02	11	2.05	0.67	0.48	0.36	0.51	0.29	0.40	0.42
phi063	10.02	22	2.95	0.56	0.65	0.39	0.66	0.41	0.43	0.41
umc1367	10.03	7	1.34	0.86	0.24	0.20	0.25	0.21	0.21	0.26
phi062	10.04	3	1.30	0.87	0.20	0.16	0.23	0.28	0.38	0.39
umc1196	10.07	25	2.77	0.56	0.61	0.49	0.64	0.23	0.30	0.33
	Mean	13.1	3.85	0.52	0.60	0.39	0.63	0.37	0.42	0.43

A, Number of alleles; *Ae*, Effective number of alleles

*Allele with the highest individual frequency

PIC Polymorphism Information Content; H_o Observed heterozygosity; H_e Expected heterozygosity; *F* Overall Inbreeding coefficient; F_{IT} Inbreeding coefficient at individual loci across total populations; F_{ST} Fixation index at individual loci across sub-populations

standard using FRAGMENT PROFILER Version 2.1 (Amersham Biosciences). Methods used for filtering raw data, allele calling (based on specific thresholds for determining true peaks/alleles), estimation of allele frequencies

using the software 'FreqsR', and conversion of allele frequencies in each population into allele sizes in individuals using the software 'F-to-L' were based on detailed protocols described earlier (Warburton et al. 2002; Dubreuil et al. 2006). 'FreqsR' also removes the stutter bands implementing a simple deconvolution method (LeDuc et al. 1995). Both 'FreqsR' and 'F-to-L' were run on 'R' platform (R Development Core Team 2005) and are available in public domain (http://www.cimmyt.org/english/wpu/biotech/index.cfm).

Statistical analysis

Summary statistics (locus-wise and/or population-wise), including the number of alleles per locus (A), effective number of alleles (A_e), major allele frequency, Polymorphism Information Content (PIC), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F) and number and frequencies of unique ('private') alleles, were estimated using the software PowerMarker version 3.25 (Liu and Muse 2005). F-statistics, including the three fixation indices to evaluate population subdivision: F_{IS} (inter-individual), F_{ST} (sub-populations) and F_{IT} (total population), were calculated using POPGENE version 1.31 (Yeh et al. 1997). The statistical significance of the F statistics was evaluated by permutations of genotypes 1,000 times through bootstrap analysis. Pair-wise F_{ST} comparisons based on genotype differences were used to evaluate genetic diversity among the accessions.

The pair-wise genetic distance among the accessions were estimated using the Nei's (1972) genetic distance, and the resulting genetic distance matrix was subjected to UPGMA (Unweighted Pair Group Method using Arithmetic means) clustering algorithm. To ascertain the statistical strength of genetic relationships identified through this analysis, bootstrapping of the data (500 permutations) was performed. The resulting dendrogram was displayed using the TREEVIEW software (Page 1996). The genetic variability at different hierarchical levels (among and within the accessions) was tested using Analysis of Molecular Variance (AMOVA) implemented in ARLEQUIN Version 3.11 (Excoffier et al. 2005).

Results and discussion

SSR polymorphisms in landrace accessions

The use of fluorescent dye-labelled SSR markers, with allele resolution using DNA sequencer, provided greater precision in allele resolution and sizing, followed by estimation of allele frequencies. All the SSR loci employed in the present study had tri-nucleotide or higher repeat motifs (except *phi075* and *phi112*), which aided in limiting the problem in allele calling due to 'stutter' which is often associated with di-nucleotide SSR loci (Warburton et al. 2002). The details of SSR loci analyzed and the data on polymorphisms, including number of alleles detected per locus (A), effective number of alleles per locus (A_e), major allele frequency, Polymorphism Information Content (PIC), observed heterozygosity (H_o), expected heterozygosity (H_e), overall inbreeding coefficient (F), are presented in Table 2. The total number of SSR alleles identified across the 42 SSR loci was 550. The number of alleles detected at each locus varied considerably, ranging from 3 (*phi062*) to 26 (*phi083*), with a mean of 13.1 alleles per locus. Very high number of alleles per locus (>20) was also observed in *phi063* (22), *phi299852* (23), *phi308707* (23) and *umc1196* (25) (Table 3). This proves a high allelic diversity among the maize landraces in India. Warburton et al. (2002) reported higher levels of allelic diversity in diverse tropical and sub-tropical populations available at CIMMYT, Mexico, as compared to the elite inbred lines. Qi-Lun et al. (2008) detected 6.4 SSR alleles per locus (across 45 loci) in a study of 124 maize landraces (bulked DNA samples)

The mean effective number of alleles (A_e) was estimated to be 3.85 across the loci analyzed, with a range of 1.26 (*umc1266*) to 10.66 (*phi083*). A_e is a function of the proportion of polymorphic loci, the number of alleles per locus, and the evenness of allele frequencies, and is thus a measure of the genetic information in a population or species. A_e allows the comparison of populations where the number and distributions of alleles differ drastically. High values of A_e have been reported in case of highly and outcrossing crops like maize (Berg and Hamrick 1997). The mean A_e of 3.85 across the accessions analyzed in this study was comparable to a value of 4.03, recorded by Qi-Lun et al. (2008) in an analysis of landraces of Southwest China.

Concomitant with the high allele number, the set of SSR loci analysed in this study had high PIC values, ranging from 0.20 (*umc1367*) to 0.90 (*phi083*), with a mean of 0.60, reflecting the presence of high allelic variation in the marker loci and their distribution in the landraces. Several loci showed PIC values higher than 0.80, indicating the informativeness of these SSR markers. The H_o for different SSR loci was lesser than the H_e, and the overall inbreeding coefficient (F) of the marker loci varied from 0.12 (*phi308707*) to 0.74 (*umc2250*) indicating population subdivision and varying levels of inbreeding.

The study identified six highly frequent ('major') SSR alleles in different loci (*phi014*, *phi090*, *phi112*, *umc1367*, *phi062* and *umc1266*) with individual frequencies greater than 0.75. The most frequent SSR allele across the 48 accessions was $umc1266_{134bp}$ with a frequency of 0.89. The frequency of the most common ('major') alleles across the accessions ranged from 0.20 (*phi083*) to 0.89 ($umc1266_{134bp}$). Two out of these SSR loci (*umc1367* and *phi062*) with very high major allele frequencies are located closely on chr.10, while the rest

Table 3 Estimates of population genetic parameters

Accession	Allelic richness	Unique alleles	F_{IS}
IML118	2.62	7	0.01
IML119	2.67	4	0.004
IML123	2.05	2	−0.01
IML124	2.14	4	0.001
IML133	2.76	4	−0.002
IML134	2.76	7	0.02
IML143	2.21	2	−0.01
IML172	2.31	3	−0.04
IML185	2.76	8	0.004
IML288	2.29	4	0.04
IML294	2.67	4	−0.02
IML296	2.19	1	0.03
IML297	2.48	4	−0.03
IML321	2.90	5	0.04
IML324	2.45	3	0.05
IML325	1.81	6	0.01
IML330	2.69	8	0.00
IML331	2.60	3	−0.05
IML332	2.83	2	0.03
IML339	2.64	2	0.03
IML340	2.21	6	0.05
IML341	2.57	8	−0.05
IML342	1.79	2	0.01
IML346	2.67	4	−0.02
IML404	2.57	4	0.04
IML405	2.83	9	−0.01
IML409	2.26	0	−0.04
IML413	2.40	7	−0.03
IML416	2.81	4	−0.02
IML422	2.05	3	−0.02
IML426	2.57	1	0.004
IML431	2.57	0	0.04
IML434	2.38	3	−0.01
IML436	2.10	0	−0.01
IML452	2.74	4	−0.002
IML454	2.52	2	0.01
IML459	2.60	4	0.03
IML462	1.81	0	0.11
IML467	1.90	5	0.06
IML474	2.33	1	0.05
IML475	2.50	7	−0.01
IML476	2.57	3	−0.03
IML480	1.98	2	−0.01
IML485	1.95	3	−0.04
IML496	2.43	1	0.01
IML498	2.24	3	0.01
IML499	2.05	1	0.05
IML505	2.50	4	−0.02
Mean	2.41	3.63	0.01

F_{IS} Inbreeding coefficient across the loci in sub-populations

are on chr.3, chr.4, chr.7 and chr.8. Based on the information available in MaizeGDB (http://www.maizegdb.org), *phi014* tags the gene *rip1* (*ribosomal inactivating protein1*) located on chromosome 8 (bin 8.04), while *phi062* tags the gene *mgs1* (*male gametophyte specific1*) on chromosome 10 (bin 10.04). *phi112* is an SSR marker located within the *opaque2*, a gene the zein biosynthesis in the maize endosperm (Prasanna et al. 2001). A major QTL for anthesis-silking interval was reported in the vicinity of *phi014* and *phi031* (Ribaut et al. 1996). QTL for days to pollen were reported in chromosomal regions near *phi053* and *phi112* (CIMMYT, unpublished data). Thus, three of the six highly frequent SSR alleles identified in this study appear to be associated with some extremely important characters from the farmers' selection point of view, namely pathogen resistance (*rip1*), pollen fertility (*mgs1*), flowering behavior, anthesis-silking interval (which is particularly important for drought stress tolerance) and nutritional quality (*o2*). In a study of diverse maize populations at CIMMYT, Mexico, using the same set of SSR markers, Reif et al. (2004) found that *phi014*, *phi031*, *phi053*, and *phi112* were particularly associated with the structuring of the germplasm. The present study shows the same tendency with respect to *phi014* and *phi112*. Thus, it is clear that the chromosomal regions harboring these SSRs are not selectively neutral. This may be an interesting starting point for further fine-scale and association mapping approaches of the underlying genes.

The presence of 'unique alleles' or 'private alleles', limited to only one of the 48 accessions under study, was investigated (Table 3). The analysis led to identification of a set of 174 such alleles in 39 out of the 42 SSR loci. The maximum number of unique alleles (9) was in IML405, followed by IML185, IML330 and IML341, with 8 alleles each. Only four accessions, namely IML409, IML431, IML436 and IML462, did not reveal any unique allele. In terms of distribution, the maximum number of unique alleles were observed in case of *umc1196* (15), followed by *phi112*, *umc1545* and *umc2250* with 9 unique alleles each. Unique or 'private' alleles are important because they may be diagnostic of particular genotype or for regions of the genome specific to a genotype. The presence of unique alleles may be attributed not only to high rates of mutation in SSR loci (Henderson and Petes 1992), but also possible selections of specific alleles in certain accessions associated with their morphology and area(s) of adaptation. It is also an indicator of potential of these accessions as a reservoir of novel alleles for crop improvement.

Genetic variability within and among populations

Allele richness and *F*-statistics (F_{IT}, F_{IS} and F_{ST}) were estimated to analyze the genetic structure of the accessions and deviation from Hardy-Weinberg Equilibrium (HWE) (Tables 2 and 3). The number of alleles per population varied from 48 (IML422) to 124 (IML339), indicating

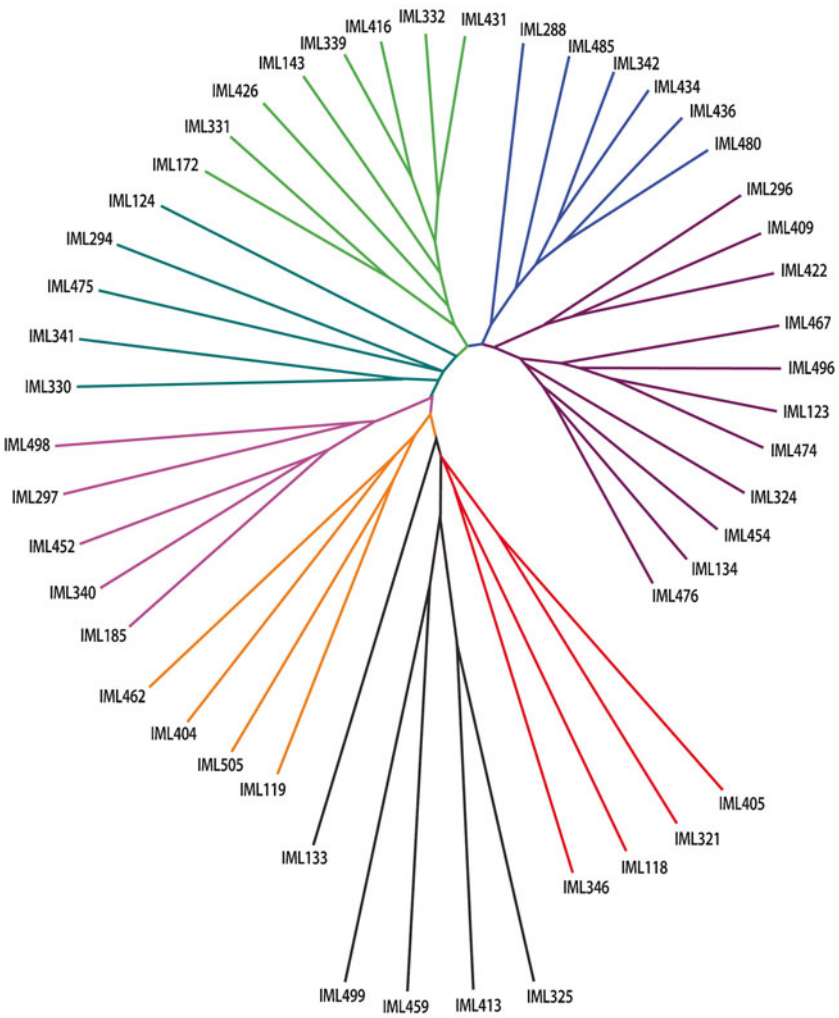


Fig. 1 An unrooted tree showing the genetic relationships among the 48 landrace accessions of maize, based on analysis of SSR dataset using Nei's (1972) genetic distance and UPGMA algorithm

significant genetic diversity. 'Allelic richness', which is the average number of alleles detected across the 42 loci, ranged from 1.79 (IML342) to 2.90 (IML321). The mean inbreeding coefficient (F_{IT}) across the SSR loci was 0.42. The F_{ST} values ranged from 0.16 (*phi308707*) to 0.79 (*umc2250*), with a mean value of 0.43 (Table 2), indicating very high genetic differentiation among the accessions analyzed, with a tendency for fixation of alternative alleles in the sub-populations. Genetic differentiation of the populations was further tested by computing the pair-wise F_{ST} values. Significant variation in the pair-wise F_{ST} values was observed among the accessions. The values ranged from 0.13 (IML340-IML416) to 0.61 (IML467-IML499) with a P value <0.0001.

Deviation from HWE is an indication of the intensity of various forces acting upon the populations, including natural and artificial selections, mutations, genetic drift, gene flow etc. In this study, we observed that the deviation from HWE was primarily due to excess of homozygotes. Although maize landraces are open-pollinated, it was not uncommon to observe excess of homozygosity, as was reported in CIMMYT maize populations established with materials from a single racial complex (Reif et al. 2004), maize landraces from Oaxaca in Mexico (Pressoir and Berthaud 2004), Swiss landraces (Eschholz et al. 2008) and Chinese landraces (Qi-Lun et al. 2008). The predominant type of selection expected on such locals/landraces could be human interference or selection pressures exerted by the farmers in favor of specific characteristics exhibited by these landraces, despite gene flow and genetic drift in the accessions.

Analysis of Molecular Variance (AMOVA) was implemented using ARLEQUIN Version 3.11 to assess the extent of genetic diversity 'within' and 'among' the accessions. Based on bootstrap analysis of the data (with 1,000 permutations), the analysis revealed that genetic variation within the individuals was 63.4 % (P=0.0004) and among the populations was 36.6 % (P=0.0005). Only 0.02 % of the total genetic variation was accounted by variation in individuals within population, which was not statistically significant (P=0.48).

To further analyze the genetic relationships among the 48 accessions, pair-wise genetic distances (GD) were computed using Nei's (Nei 1972) genetic distance. We opted for Nei's (Nei 1972) genetic distance for analysis of pair-wise genetic

distances among the accessions, based on the 'infinite-allele model' (Kimura and Crow 1964). In this model, it was assumed that an ancestral population split into various sub-populations, which diverged due to genetic drift and mutations. Intensive study of microsatellite variability and pattern of mutations in *Zea*, undertaken by Matsuoka et al.(2002), revealed that *Zea* microsatellite alleles does not fit into the 'stepwise mutation model' because much of the observed variation appears to results from InDels in regions flanking the repeat regions. The pair-wise GD values among the various pairs of accessions ranged from 0.20 (IML339-IML416) to 0.98 (IML325-IML346), with an average GD of 0.53, reinforcing the above observations on genetic divergence among the accessions under study. The resulting GD matrix was further subjected to cluster analysis using UPGMA, and the unrooted tree was depicted as Fig. 1.

Cluster analysis of the SSR data showed several groups and sub-groups among the accessions, indicating significant genetic diversity. However, it was not possible to differentiate the accessions on the basis of specific geographical origins. Among the 20 mid-altitude/highland accessions analyzed, seven were in one group, while the rest were dispersed in the dendrogram, indicating considerable genetic drift among these populations. Overall, the study revealed high levels of intra-population as well as inter-population diversity (despite varying levels of inbreeding) in the selected maize landrace accessions from diverse agro-ecological zones in India.

In a study to determine the relationships among common bean landraces in Nicaragua observed that there was no correlation of geographic distribution of those landraces with genetic diversity based on molecular data (Gomez et al. 2004). In the present study too, most of the variation of the landraces at the molecular level was explained by differences within or among landraces but not among agro-ecological zones. The population genetic parameters, particularly F statistics, also clearly revealed high levels of variation not only among the populations but also within individuals of populations.

Thus, the present study wherein 'population DNA bulk strategy' using fluorescent dye-labeled SSR markers was utilized for analysis of intra- and inter-population diversity in selected maize landrace accessions of India, revealed high level of genetic differentiation and presence of several unique alleles in the maize landraces in India. In the context of in situ conservation of landraces, both molecular and phenotypic data are needed for identifying populations for conservation, optimum sites for germplasm collection, and understanding the ongoing changes in the diversity of the landraces in the course of conservation by the farming communities or in the gene banks (Prasanna 2012). Intensive molecular characterization and population genetic analysis of a larger set of maize landraces from diverse agro-ecological regions in India vis-à-vis global collection using high-density SNP platforms, such as genotyping-by-sequencing or GBS (Elshire et al. 2011), shall further aid in understanding the diversity at the haplotype level, and utilizing this information in molecular prebreeding (Prasanna 2012). To achieve this, besides high-density genotyping data, intensive and concerted efforts are also needed for generating multi-location and precision phenotypic data on the selected maize landraces of India for the target traits of interest.

Acknowledgements The study was undertaken as a part of the ICAR National Fellow Project (2005–2010) awarded to BMP. The authors thank the Division of Germplasm Conservation, NBPGR, New Delhi (especially Dr Kalyani Srinivasan) for providing seed material of some landrace accessions used in this study.

References

Berg EE, Hamrick JL (1997) Quantification of genetic diversity at allozyme loci. Can J Forest Res 27:415–424

Daniels J, Holmans P, Williams N, Turic D, McGuffin P, Plomin R, Owen MJ (1998) A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. Am J Hum Genet 62:1189–1197

R Development Core Team (2005) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (http://www.R-project.org)

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Eschholz TW, Peter R, Stamp P, Hund A (2008) Genetic diversity of Swiss maize assessed with individuals and bulks on agarose gels. Genet Resour Crop Evol 55:971–983

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.1: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50

Gomez OJ, Matthew W, Blair P, Bodil E, Frankow-Lindberg C, Urban G (2004) Molecular and phenotypic diversity of common bean landraces from Nicaragua. Crop Sci 44:1412–1418

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol 12:2749–2757

Joshi PK, Singh NP, Singh NN, Gerpacio RV, Pingali PL (2005) Maize in India: production system. Constraints and Research Priorities, CIMMYT, Mexico DF

Kimura M, Crow JR (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

LeDuc C, Miller P, Lichell J, Parry P (1995) Batched analysis of genotypes. PCR Methods Appl 4:331–336

Liu K, Muse SV (2005) PowerMarker: an integrated analysis of environment for genetic marker analysis. Bioinformatics 21:2128–2129

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley JA (2002) Single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Nass LL, Pellicano IT, Valois ACC (1993) Utilization of genetic resources for maize and soybean breeding in Brazil. Brazilian J Genet 16:983–988

Nei M (1972) Genetic distance between populations. Am Natur 106:283–292

Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358

Perlin MW, Lancia G, Ng SK (1995) Towards fully automated genotyping: genotyping microstatellite markers by deconvolution. Am J Hum Genet 57:1199–1210

Prasanna BM (2010) Phenotypic and molecular diversity of maize landraces: characterization and utilization. Indian J Genet 70:315–327

Prasanna BM (2012) Diversity in global maize germplasm and novel initiatives for characterization and utilization. J Biosci 34 (In Press)

Prasanna BM, Sharma L (2005) The landraces of maize (*Zea mays* L.): diversity and utility. Indian J Plant Genet Resour 18:155–168

Prasanna BM, Vasal SK, Kassahun B, Singh NN (2001) Quality protein maize. Curr Sci 81:1308–1319

Prasanna BM, Pixley K, Warburton ML, Xie CX (2010) Molecular marker-assisted breeding options for maize improvement in Asia. Mol Breed 26:339–356

Pressoir G, Berthaud J (2004) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94

Qi-Lun Y, Ping F, Ke-Cheng K, Guang-Tang P (2008) Genetic diversity based on SSR markers in maize (*Zea mays* L.) landraces from Wuling mountain region in China. J Genet 87:287–291

Rebourg C, Gousnard B, Charcosset A (2001) Large scale molecular analysis of traditional European maize populations: relationships with morphological variation. Heredity 86:574–587

Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. Theor Appl Genet 106:895–903

Reif JC, Xia XM, Melchinger AE, Warburton ML, Hoisington D, Beck D, Frisch M (2004) Genetic diversity determined within and among CIMMYT maize populations of tropical sub tropical and temperate germplasm by SSR markers. Crop Sci 44:326–334

Ribaut J-M, Hoisington DA, Deutsch JA, Jiang C, Gonzalez-de-Leon D (1996) Identification of quantitative trait loci under drought conditions in tropical maize: 1. Flowering parameters and the anthesis-silking interval. Theor Appl Genet 92:905–914

Sharma L, Prasanna BM, Ramesh B (2010) Phenotypic and microsatellite-based diversity and population genetic structure of maize landraces in India, especially from the North East Himalayan region. Genetica 138:619–631

Singh B (1977) Races of maize in India. Indian Council of Agricultural Research (ICAR), New Delhi

Warburton ML, Xianchun X, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 42:1832–1840

Warburton ML, Wilkes G, Taba S, Charcosset A, Mir C, Bedoya C, Prasanna BM, Xie CX, Hearne SH, Franco J (2011) Gene flow between different teosinte species and into the domesticated maize gene pool. Genetic Resour Crop Evol 58:1243–1261

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, version 1.2, Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton