Available online at www.sciencedirect.com

ScienceDirect

The Crop Journal

Phenotypic and molecular diversity-based prediction of heterosis in pearl millet (*Pennisetum glaucum* L. (R.) Br.)☆

Shashi Kumar Gupta[a,*], *Thirunavukkarasu Nepolean*[a,b], *Chinna Ghouse Shaikh*[a], *Kedarnath Rai*[a], *Charles Thomas Hash*[a,c], *Roma Rani Das*[a], *Abhishek Rathore*[a]

[a]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, India*
[b]*Indian Agricultural Research Institute (IARI), New Delhi 110012, India*
[c]*International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Niamey BP 12404, Niger*

ARTICLE INFO

Article history:
Received 7 July 2017
Received in revised form 24 August 2017
Accepted 25 September 2017
Available online 8 December 2017

Keywords:
Hybrid parents
Agronomic traits
Euclidean distance
Molecular markers
Simple matching distance

ABSTRACT

Genetic distances between hybrid parents based on phenotypic traits and molecular markers were investigated to assess their relationship with heterosis for grain and stover yield and other traits in pearl millet (*Pennisetum glaucum* [L.] R. Br.). Fifty-one hybrids developed using 101 hybrid parents (B and R lines) and showing a wide range of genetic distance between their parents based on eight phenotypic traits and 28–38 SSRs were evaluated in two sets for two seasons. The correlation between Euclidean distance (phenotypic distance, ED) and simple matching distance (molecular distance, SM) for parents of both sets was low but positive and significant ($r = 0.2$, $P < 0.001$).The correlation of ED in parents with better-parent heterosis for grain yield was similar in both sets ($r = 0.38$, $P < 0.05$). SM was not correlated with heterosis for grain yield in either set of hybrids. The results showed that phenotypic distance could be a better predictor of heterosis than molecular distance. The correlation between phenotypic distance and heterosis was not strong enough to permit the use of phenotypic diversity among parents as a major selection criterion for selection of parental lines displaying high levels of heterosis for grain and stover yield in pearl millet.

1. Introduction

Pearl millet (*Pennisetum glaucum* [L.] R. Br.) is a major food and fodder crop for farmers living on marginal agricultural lands in the arid and semi-arid tropics of Africa and Asia (largely India). Its grain serves as staple food and its stover is equally important for livestock in these marginal economies. In India, pearl millet breeding programs have been developing hybrids since the 1960s, and hybrids presently occupy about 5 Mha of the total of >8 Mha under cultivation, especially in higher-yielding environments. Hybrid adoption contributed to a crop productivity increase from 288 kg ha^{-1} during 1951–1955 to 1164 kg ha^{-1} during 2013–2014, registering an improvement of about 300% for pearl millet in India [1]. Although this order of productivity gain is quite impressive for a crop grown under low-input conditions in marginal

☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding author.*
E-mail address: s.gupta@cgiar.org (S.K. Gupta).

https://doi.org/10.1016/j.cj.2017.09.008

environments, greater advances are possible if hybrids are developed based on heterosis prediction using parental information for genetic diversity. The level of genetic diversity between parents has been proposed as a predictor of F_1 hybrid performance and heterosis [2,3]. This predictive method may help to identify more heterotic combinations, thus reducing costs associated with making crosses and field evaluation to select promising hybrids.

Conflicting results with respect to the relationship between genetic distance and heterosis have been reported in various crops. Some earlier work is in agreement with the classical theories of heterosis; Zhang et al. [4] in rice (*Oryza sativa*), Riaz et al. [5] in rapeseed (*Brassica napus*), and Kiula et al. [6] in maize (*Zea mays*) found molecular marker-based genetic diversity to be linked to increased heterosis. In contrast, several other studies reported little or no possibility of predicting heterosis from molecular marker-based genetic distance in other crops [7–9].

In a pearl millet study conducted earlier on a limited number of parental lines with a narrow range of molecular diversity, there was no correlation between molecular marker-based genetic distance and heterosis for grain yield [10]. In our earlier work based on a large number of potential hybrid parents and SSR markers [11–12], we observed a wide spectrum of genetic diversity among the hybrid parents, and markers were well able to group genotypes related by pedigree and traits. Based on these results, hybrids involving parental lines with varying genetic distances were evaluated in this study along with their parental lines for yield and other agronomic traits. The trial data was investigated, with the aim of assessing the relationship between genetic distance based on molecular markers and phenotypic traits including performance *per se*, heterosis for grain and stover yield, and other agronomic traits.

2. Materials and methods

2.1. Experimental material

The basic genetic material for this study comprised two sets of parental lines. The first set (hereafter, referred as Set I) comprised of 213 lines, which involved 98 maintainer parents (designated between 1984 and 2004 at ICRISAT, Patancheru, India) and 115 restorer parents (designated between 1985 and 1995 at ICRISAT, Patancheru, India). The second set (hereafter, referred as Set II) comprised of 166 hybrid parents, which comprised 88 maintainer parents and 78 restorer parents bred at ICRISAT, Patancheru, India since 2004. Genotyping data was generated using 38 SSRs for 213 lines in Set I [11] and its subset of 28 SSRs for 166 lines in Set II [12]. Genotyping data of both the sets were analyzed using Darwin 5.0 [13]. The SSRs were highly polymorphic and 30 of them were distributed over all seven linkage groups in earlier studies [14–17] (Table 1). A dissimilarity matrix was calculated for pairs of maintainer parents (B lines) × restorer parents (R lines) using simple matching [13]. Twenty-two and 29 hybrid combinations were identified for sets I and II, respectively, based on genetic distance between B and R lines. Hybrid combinations were identified, considering that pairs with diverse pedigree parents were selected, and the genetic distances between B and R lines of pairs represented all levels (low, medium and high) of genetic distance. Genetic distance varied from 0.19 to 0.90 between B and R lines of Set I and from 0.17 to 0.93 in Set II lines. Seed of these identified hybrid combinations was produced in summer season of 2008 (for Set I) and 2009 (for Set II). Twenty-two hybrids (20 B × R and 2 R × R) were developed from Set I lines using 20 B lines and 23 R lines. Twenty-nine B × R hybrids were developed from Set II lines using 29 each of maintainer and restorer parents.

2.2. Field trials

Hybrids and their parents were planted in alfisol soils in two seasons (rainy season of 2008 and summer season of 2010 for Set I, and rainy season of 2009 and summer season of 2010 for Set II), in randomized complete block designs with three replications, at ICRISAT, Patancheru, India (17.35° N latitude, 78.27° E longitude). Plots consisted of four rows of 4 m length with inter-row spacing of 60 cm in summer season and 75 cm in rainy season with an interplot spacing of 10–15 cm within rows. The hybrids and parents were planted in separate but adjacent blocks within each replication and randomization was performed separately for crosses and parents. Data were recorded for time to flowering as number of days from sowing to full stigma emergence on the main panicle of 50% plants in a plot. Plant height (cm), number of productive tillers, panicle length (cm), and panicle diameter (mm) were recorded for five competitive plants from the central two rows of a plot. At maturity, panicles were harvested manually, sun-dried for two weeks, and threshed to determine grain weight (g). Remaining plants were cut at ground level and fresh stover weight was recorded (kg). About 1 kg of fresh stover was then chopped and oven-dried to determine plot dry weight (kg). Grain and stover (fresh and dry) yield were converted to kg ha^{-1}. A random sample of 200 kernels for each plot was weighed and multiplied by five to determine 1000-grain weight (g). Data for days to 50% flowering and dry stover yield were available from only one season in Set II of hybrids and parents.

2.3. Data analysis

Euclidean distance (ED) was calculated based on eight phenotypic traits (days to 50% flowering, plant height, productive tillers, panicle length, panicle diameter, 1000-grain weight, grain yield, and dry stover yield) and simple matching distance (SM) was computed using SSR data for both sets of hybrid parents. Population structure analysis was performed with STRUCTURE software version 2.3.4 [18].

The dissimilarity matrices from phenotypic traits and from molecular markers were used to construct dendrograms based on Wars hierarchical agglomerative clustering using R version 3.2.2 [19] and unweighted pair group method with arithmetic mean (UPGMA) using Darwin. Analysis of variance (ANOVA) was performed using SAS 9.4 for Windows [20] to identify significant differences between the F_1s and their parents and among the F_1s. For all the traits, absolute mid-parent heterosis (AMPH), relative mid-parent heterosis

(RMPH), and better-parent heterosis (BPH) were calculated as follows:

$$\text{AMPH} = F_1\text{–MP}; \text{RMPH} = \left(\frac{F_1\text{–MP}}{\text{MP}}\right) \times 100, \text{BPH} = \left(\frac{F_1\text{–BP}}{\text{BP}}\right) \times 100,$$

where, F_1 is trait value for hybrid performance, BP is trait value for better parent, and MP is mid parental trait value.

$$\text{Mid–parent (MP)} = \frac{P_1 + P_2}{2},$$

where, P_1 is trait value for first parent and P_2 is trait value for second parent.

Pearson's correlation coefficients between SM and ED were estimated for both sets of parents separately considering all traits and markers and between SM and ED on one hand and better-parent heterosis, mid-parent heterosis and hybrid performance on the other, for all traits and both sets of hybrids.

3. Results

3.1. SSR polymorphism and parental relatedness

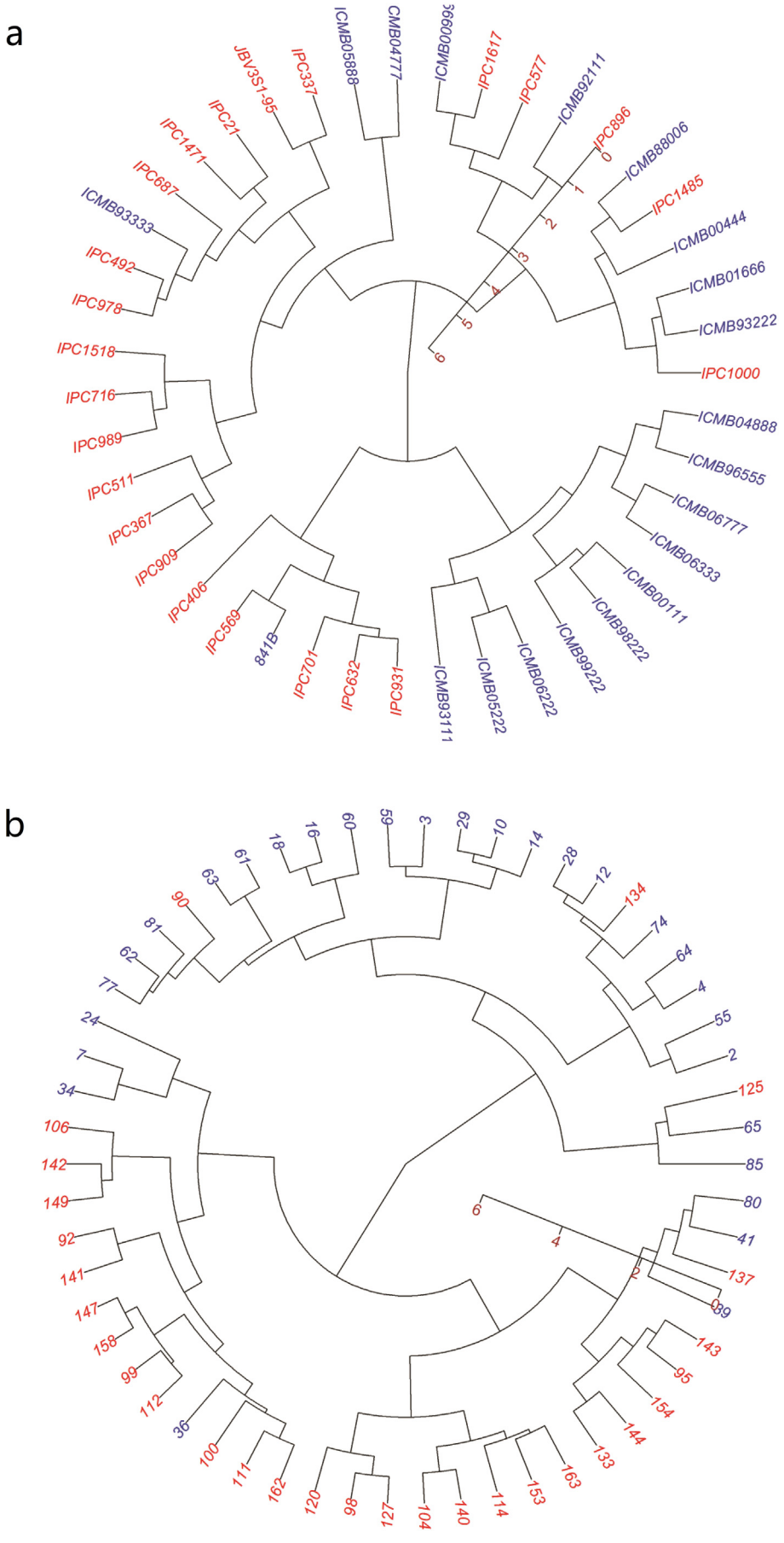
3.1.1. Set I

The 38 SSR loci detected a total of 232 alleles in 43 lines (20 B and 23 R lines), with an average of 6.05 alleles per locus. The number of alleles per locus varied from 2 to 17 (Table 1). Nineteen of the 38 SSRs were highly polymorphic, with PIC values varying from 0.62 to 0.89 and averaging 0.58. Gene diversity varied from 0.09 (*Xicmp3048*) to 0.9 (*Xpsmp2218*) with

Table 1 – Chromosome position, allelic composition, polymorphic information content (PIC), gene diversity, and observed heterozygosity of simple sequence repeat loci based on 101 parents (43 of Set I and 58 of Set II).

Marker	Chromosome number	Allele no.		Gene diversity		Heterozygosity		PIC	
		Set I	Set II	Set I	Set II	Set I	Set II	Set I	Set II
Xctm10	3	10	13	0.85	0.81	0.05	0.13	0.84	0.79
Xctm12	1	7	8	0.82	0.76	0.00	0.02	0.79	0.73
Xpsmp2045	–	3	6	0.54	0.64	0.02	0.03	0.44	0.57
Xpsmp2068	–	10	10	0.86	0.79	0.05	0.02	0.84	0.77
Xpsmp2077	2	5	4	0.51	0.45	0.00	0.18	0.47	0.39
Xpsmp2079	7	12	16	0.84	0.87	0.08	0.02	0.83	0.86
Xpsmp2089	2	11	15	0.84	0.92	0.05	0.11	0.82	0.91
Xpsmp2090	1	5	6	0.77	0.64	0.00	0.00	0.74	0.60
Xpsmp2201	2	5	6	0.42	0.37	0.05	0.00	0.39	0.34
Xpsmp2202	5	2	3	0.39	0.47	0.07	0.02	0.32	0.39
Xpsmp2203	7	9	9	0.78	0.77	0.08	0.03	0.75	0.75
Xpsmp2204	–	7	3	0.78	0.20	0.00	0.02	0.74	0.19
Xpsmp2209	–	5	6	0.62	0.67	0.08	0.04	0.58	0.62
Xpsmp2218	–	16	11	0.90	0.80	0.10	0.02	0.89	0.78
Xpsmp2220	5	10	4	0.85	0.51	0.05	0.02	0.83	0.47
Xpsmp2222	–	4	4	0.32	0.19	0.00	0.06	0.31	0.18
Xpsmp2227	3	3	5	0.37	0.33	0.00	0.00	0.32	0.31
Xpsmp2232	2	7	8	0.74	0.76	0.00	0.00	0.70	0.74
Xpsmp2237	2	5	7	0.67	0.46	0.05	0.00	0.62	0.43
Xpsmp2246	1	4	3	0.57	0.53	0.02	0.02	0.52	0.47
Xpsmp2248	6	5	4	0.61	0.58	0.00	0.03	0.57	0.53
Xpsmp2249	3	3	4	0.33	0.52	0.00	0.03	0.29	0.47
Xpsmp2273	1	9	9	0.83	0.77	0.05	0.02	0.81	0.74
Xicmp3002	6	3	6	0.56	0.63	0.02	0.06	0.47	0.58
Xicmp3032	1	5	5	0.69	0.57	0.07	0.00	0.63	0.53
Xicmp3048	7	2	2	0.09	0.29	0.00	0.02	0.09	0.25
Xicmp3080	1	5	5	0.69	0.57	0.05	0.00	0.64	0.54
Xicmp3088	1	5	7	0.71	0.76	0.00	0.11	0.66	0.73
Xctm8	7	4	–	0.54	–	0.00	–	0.44	–
Xpsmp2070	3	17	–	0.89	–	0.03	–	0.89	–
Xpsmp2086	4	6	–	0.60	–	0.00	–	0.56	–
Xpsmp2207	–	7	–	0.74	–	0.05	–	0.69	–
Xpsmp2211	2	4	–	0.55	–	0.02	–	0.48	–
Xpsmp2212	–	3	–	0.31	–	0.03	–	0.28	–
Xpsmp2214	3	3	–	0.66	–	0.02	–	0.59	–
Xpsmp2251	3	3	–	0.50	–	0.00	–	0.40	–
Xpsmp2267	3	3	–	0.53	–	0.02	–	0.43	–
Xicmp3043	7	3	–	0.46	–	0.00	–	0.42	–
Mean		6	7	0.63	0.59	0.03	0.04	0.58	0.56

–, Not available.


a
ICMB05888
CMB04777
ICMB00090
IPC1617
IPC577
ICMB92111
IPC896
ICMB88006
IPC1485
ICMB00444
ICMB01666
ICMB93222
IPC1000
ICMB04888
ICMB96555
ICMB06777
ICMB06333
ICMB00111
ICMB98222
ICMB99222
ICMB06222
ICMB05222
ICMB93111
IPC931
IPC632
IPC701
841B
IPC569
IPC406
IPC909
IPC367
IPC511
IPC989
IPC716
IPC1518
IPC978
IPC492
ICMB93333
IPC687
IPC1471
IPC21
JBV3S1-95
IPC337
b
59
3
62
10
14
28
12
134
74
64
4
55
2
125
65
85
80
41
137
89
143
95
154
144
133
163
153
114
140
104
127
98
120
162
111
100
36
112
99
158
147
141
92
149
142
106
34
7
24
77
62
81
90
63
61
18
16
60

an average of 0.62. The level of heterozygosity in SSRs across B lines and R lines ranged from 0.02 to 0.09 and averaged 0.029, exceeding 0.05 in six SSRs. Allele sizes for the internal control (Tift23dD_2B_1) were uniform and reproducible for each of the markers, indicating the accuracy of the protocol and reproducibility of allelic data for a given primer across assays in both the sets of parental lines.

All 20 B lines and 23 R lines were diverse in parentage, resulting in a wide range of ED and SM estimates. SM among pairs of B and R lines ranged from 0.04 to 0.95 with a mean of 0.61, and ED varied from 1.22 to 7.44 with a mean of 4.17.

3.1.2. Set II

The 28 SSR loci detected 192 alleles in 58 lines (29 B lines and 29 R lines), with an average of 6.75 alleles per locus. The number of alleles per locus varied from 2 to 16 (Table 1). Twelve of the 28 SSRs were highly polymorphic, with PIC values ranging from 0.60 to 0.91 and averaging 0.56. Gene diversity varied from 0.19 (*Xpsmp2222*) to 0.92 (*Xpsmp2089*). The level of heterozygosity in SSRs across B and R lines ranged from 0.01 to 0.18 and averaged 0.036, exceeding 0.05 in six SSRs.

All 58 lines (29 each of B and R lines) were diverse in parentage; the range of SM was from 0.06 to 0.88 with a mean of 0.58. ED varied from 0.99 to 7.99 with a mean of 3.57.

3.2. Cluster analysis based on phenotypic traits and molecular data

3.2.1. Set I

The dendrograms from cluster analysis based on the ED and SM matrices are presented in Figs. 1-a and 2-a, respectively. The ED-based clustering formed two separate clusters for R lines (with 13 and 5 R lines each) and two separate clusters for B lines (with 10 and two B lines each), and one cluster contained a mixture of six B lines and five R lines (Fig. 1-a). The SM-based dendrogram clearly grouped B and R lines into separate clusters with only one B and R line each found in contrasting clusters (Fig. 2-a). In the structure-based population stratification analysis, B and R lines were clearly separated into two subgroups (Fig. 3).

3.2.2. Set II

The dendrograms from cluster analysis based on ED and SM matrices are presented in Figs. 1-b and 2-b, respectively. The ED-based clustering formed two separate major clusters for R lines (26 R lines) and two clusters for B lines (with 3 and 22 B lines each) (Fig. 1-b). The SM-based dendrogram clearly partitioned B and R lines into separate clusters with two B and three R lines found in alternate clusters (Fig. 2-b). In the structure analysis, the majority of the lines fell into their respective B and R groups, though there was some admixture (Fig. 3).

3.3. Performance *per se* and heterosis

A combined analysis of variance across both the seasons for all phenotypic traits in both the sets of hybrids and parents showed highly significant differences among the parents and the F_1s (results not presented). Parents *vs.* F_1, which tests for heterosis, was also highly significant for all the traits. The means and ranges of heterosis for grain yield and other important traits are presented in Table 2. The extent of heterosis varied considerably for different traits. Grain yield showed the highest RMPH in both sets of hybrids (76.5% in Set I and 86.2% in Set II), followed by plant height, 1000-grain weight, and panicle length in both sets. RMPH for grain yield varied from 37.1% (ICMB 92111 × IPC 1000) to 155.9% (ICMB 04777 × IPC 569) in Set I hybrids, and from 23.1% (B-4 × R-33) to 154.3% (B-12 × R-41) in Set II hybrids. Grain yield showed the highest BPH in both sets of hybrids, with a mean of 56.3% in Set I and a range of 20.7%–122.3% and a mean of 65.3% with range of –19%–118% in Set II.

3.4. Correlation of parental diversity with hybrid performance *per se* and heterosis

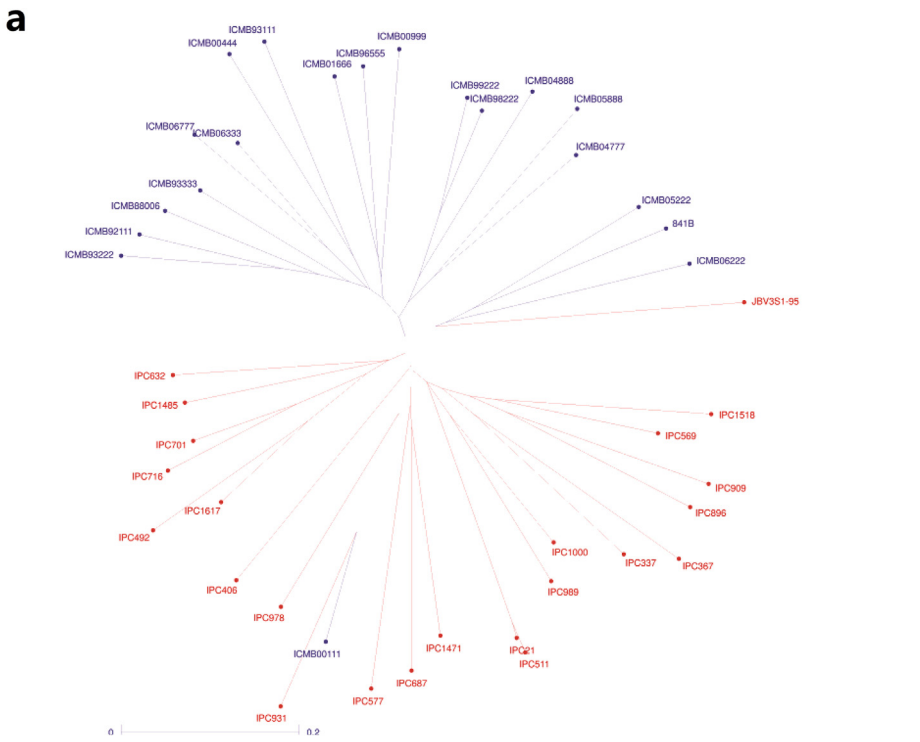
The correlations of ED and SM with hybrid performance, mid-parent heterosis, and better-parent heterosis for different traits in both sets of hybrids are presented in Table 3. ED and SM showed no correlation with hybrid performance for any of the traits in either Set of hybrids, whereas ED showed a significant negative correlation with better-parent heterosis for panicle diameter in Set I and with panicle length in Set II hybrids. ED showed a positive significant correlation ($r = 0.38$; $P < 0.05$) with better-parent heterosis for grain yield in both sets of hybrids. SM showed a significant positive correlation with better-parent heterosis for panicle diameter in Set II hybrids.

ED showed a significant positive correlation with mid-parent heterosis for grain yield ($r = 0.59$ for Set I and $r = 0.50$ for Set II), whereas for plant height and dry stover yield it showed a positive correlation for Set I hybrids only. No significant correlation was found between SM and heterosis for grain yield and dry stover yield, though a positive correlation was found for plant height in Set I hybrids and for productive tillers and panicle diameter in Set II hybrids. The correlation between ED and SM ($r = 0.2$, $P < 0.001$) for parental lines of both the hybrid sets was positive and significant but very low (Fig. 4-a, b).

4. Discussion

Set I, comprising 213 pearl millet hybrid parents (98 B lines and 115 R lines), and Set II. comprising 166 hybrid parents (88 B lines and 78 R lines) were found to be genetically diverse sets of hybrid parents. SSR analysis in each of these two sets showed B and R lines falling in two separate clusters [11,12]. This result indicated that SSRs used in those studies could detect the morphological differences for which B and R lines are bred in ICRISAT's trait-specific breeding program of pearl millet. The structure-based population stratification analysis also explained the grouping pattern between B and R lines in Set I and Set II. Set I was separated into two subgroups, in

Fig. 1 – Clustering pattern of parents based on Euclidean distance based on eight morphological traits using Ward's method. Genotypes shown in red are R lines and in blue are B lines. (a) 43 parents of Set I hybrids, (b) 58 parents of Set II hybrids.


a
ICMB93111
ICMB00444
ICMB00999
ICMB96555
ICMB01666
ICMB99222
ICMB98222
ICMB04888
ICMB05888
ICMB06777
ICMB06333
ICMB04777
ICMB93333
ICMB88006
ICMB92111
ICMB93222
ICMB05222
841B
ICMB06222
JBV3S1-95
IPC632
IPC1485
IPC701
IPC716
IPC1617
IPC492
IPC406
IPC978
ICMB00111
IPC1471
IPC887
IPC577
IPC931
IPC1518
IPC569
IPC909
IPC896
IPC1000
IPC337
IPC367
IPC989
IPC21
IPC511
0
0.2

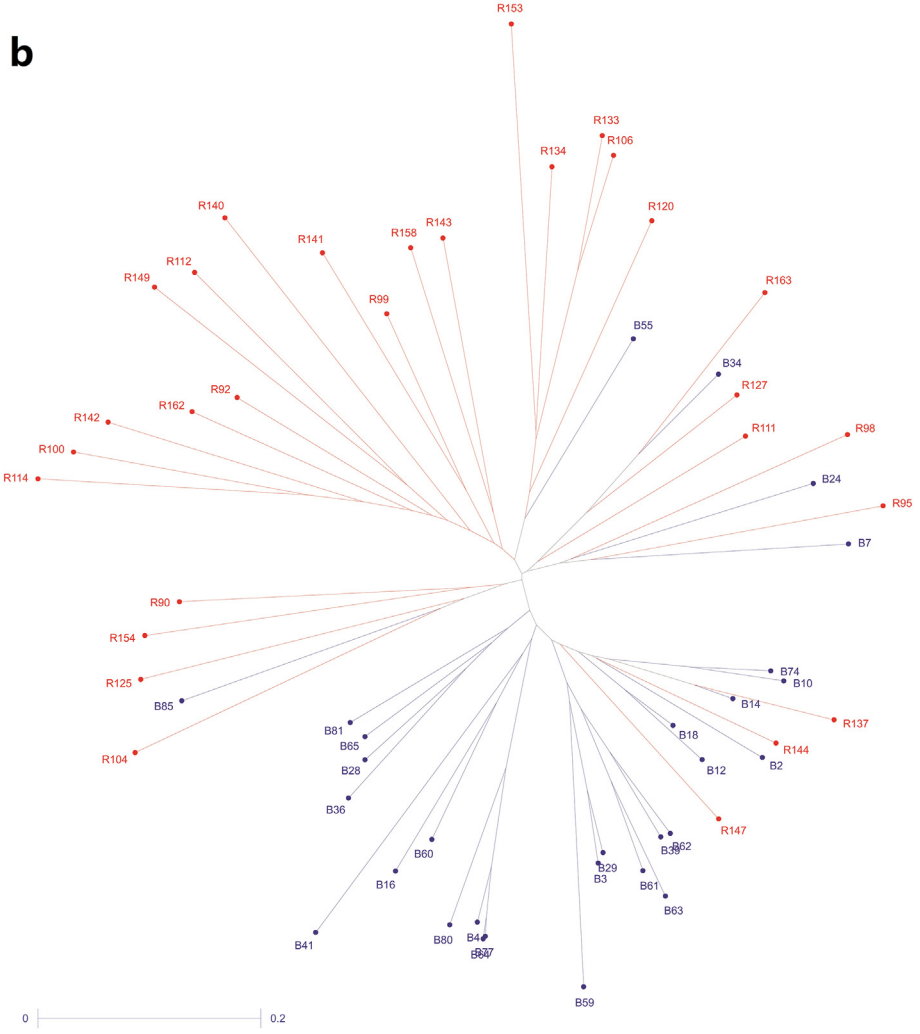

b
R153
R133
R106
R134
R140
R120
R158
R143
R141
R112
R149
R99
R163
B55
B34
R127
R92
R162
R142
R111
R98
R100
R114
B24
R95
B7
R90
R154
R125
B85
R104
B74
B10
B14
R137
B81
B65
B28
B36
B18
R144
B12
B2
R147
B60
B16
B41
B62
B29
B3
B61
B63
B80
B59
0
0.2

Table 2 – Parental and F_1 performance along with mean and range for absolute mid-parent heterosis (AMPH), relative mid-parent heterosis (RMPH), and better-parent heterosis (BPH) for morphological traits in two sets of pearl millet hybrids evaluated in two seasons at ICRISAT, Patancheru, India.

Phenotypic trait	Parental mean	Parental mean range	Hybrid (F_1) performance	Mean AMPH	Mean RMPH (%)	Mean RMPH range (%)	Mean BPH	Mean BPH range
Set I (n[b] = 22)								
Days to 50% flowering	49	43–56	44	−5.11	−10.36	−16.22–6.55	−14.3	−20.61–6.62
Plant height (cm)	120.81	73.73–164.17	157.69	36.14	30.82	8.18–54.49	16.14	−8.08–38.35
Productive tillers (no.)	2.42	1.14–4.44	2.36	−0.05	−0.92	−35.13–38.41	−14.15	−52.50–12.35
Panicle length (cm)	20.23	14.22–39.26	22.63	2.43	12.21	1.77–22.74	1.28	−27.59–17.39
Panicle diameter (mm)	25.3	15.82–33.22	27.33	1.98	7.82	−0.25–14.94	−3.31	−15.75–12.11
1000-grain weight (g)	8.34	5.67–11.86	9.7	1.35	16.51	0.49–47.62	2.7	−16.28–31.06
Grain yield (kg ha^{-1})	1020.2	570.06–1496.33	1765.3	744.59	76.55	37.09–155.94	56.38	20.73–122.30
Fresh stover yield (kg ha^{-1})	4045.43	1686.83–8105.33	5552.32	1443.59	39.88	−0.16–105.07	20.4	−22.04–96.29
Dry stover yield (kg ha^{-1})	1489.92	655.5–2562.17	2015.24	508.12	37.07	5.36–97.27	17.82	−17.13–56.90
Set II (n = 29)								
Days to 50% flowering	50[a]	41–57[a]	46[a]	−3.89[a]	−7.83[a]	−17.20–14.29[a]	−3.49[a]	−6.77–2.02[a]
Plant height (cm)	128.32	79.67–190.83	175.52	47.2	37.41	1.09–66.35	18.75	−15.86–54.71
Productive tillers (no.)	2.27	1.24–3.23	2.14	−0.13	−4.81	−35.28–26.12	−15.76	−42.39–19.60
Panicle length (cm)	21.58	12.87–38.57	25.33	3.75	17.44	−20.06–53.19	4.05	−32.69–44.96
Panicle diameter (mm)	25.52	17.80–36.05	28.28	2.76	10.77	−7.03–41.9	2.45	−18.24–31.59
1000-grain weight (g)	8.96	5.85–12.10	11.17	2.21	25.18	0.40–80.30	16.27	−12.82–69.13
Grain yield (kg ha^{-1})	1072.79	520.58–1549.83	1957.3	884.52	86.2	−13.52–154.25	65.33	−19.07–118.04
Fresh stover yield (kg ha^{-1})	4847.15	2563.67–7049	6652.53	1805.38	38.75	−32.70–78.65	19.88	−33.72–55.60
Dry stover yield (kg ha^{-1})	2102.55[a]	1143.33–3523.33[a]	2364.92[a]	262.37[a]	12.48[a]	−19.53–61.03[a]	−0.29[a]	−6.5–5.66[a]

[a] One season data only.
[b] n represents number of hybrids.

which B and R-lines were clearly separated. In Set II, though there was some admixture, a majority of the lines fell into their respective B and R groups. Cross-breeding between B and R lines may account for the presence of admixture in the lines (12). Thus, 22 hybrid combinations (20 B × R and two R × R) from Set I and 29 (B × R) from Set II, having 0.1 to 0.9 SM between their parents, were evaluated along with the parental lines selected for investigation.

In this study, SM was poorly, though significantly and positively (r = 0.2, P < 0.001), correlated with ED in both sets of hybrids. Earlier studies have also shown both theoretically and experimentally that molecular marker distance does not necessarily correspond to phenotypic trait-based differences [21,22]. According to Burstin and Charcosset [22], polygenic inheritance and linkage disequilibrium could cause such low levels of relationship between the two measures of diversity.

The molecular and phenotypic distance measurements differed in their ability to predict heterosis and F_1 performance. Neither phenotypic (ED) nor molecular genetic distance (SM) showed any correlation with hybrid performance *per se* for grain yield in either set of hybrids. Phenotypic distance was significantly correlated (r = 0.38, P < 0.05) with better-parent heterosis in both sets, and with mid-parent heterosis (r = 0.59, P < 0.01 in Set I and r = 0.50, P < 0.01 in Set II). In contrast, molecular distance was not significantly correlated with either better-parent heterosis, hybrid

Fig. 2 – Clustering pattern of parents based on simple matching distance. Genotypes shown in red are R lines and those in blue are B lines. (a) 43 parents of Set I hybrids using 38 SSRs, (b) 58 parents of Set II hybrids using 28 SSRs.

Table 3 – Correlations of Euclidean distance (ED) and simple matching distance (SM) with hybrid performance, relative mid-parent heterosis, and better-parent heterosis in pearl millet hybrids evaluated in two seasons at ICRISAT, Patancheru, India.

Item	Hybrid set	Grain yield	Dry stover yield	Days to 50% flowering	Plant height	Productive tillers	Panicle length	Panicle diameter	1000-grain weight
ED with hybrid performance	Set I (n = 22)	0.022	−0.072	−0.017	0.070	−0.092	0.389	−0.350	−0.025
	Set II (n = 29)	0.356	0.081[a]	0.011[a]	0.209	0.056	−0.214	0.245	0.124
ED with heterosis over mid-parent	Set I	0.599[**]	0.613[**]	−0.230	0.602[**]	0.030	−0.052	−0.325	0.016
	Set II	0.500[**]	−0.101[a]	−0.099[a]	0.147	0.392[*]	−0.359	0.208	0.010
ED with heterosis over better-parent	Set I	0.380[*]	0.334	−0.296	0.204	−0.031	−0.344	−0.495[*]	−0.219
	Set II	0.380[*]	−0.297[a]	−0.054[a]	−0.103	0.247	−0.399[*]	−0.020	−0.124
SM with hybrid performance	Set I	0.240	−0.028	−0.155	−0.111	−0.062	0.192	−0.202	−0.074
	Set II	0.296	0.212[a]	−0.358[a]	−0.098	0.248	−0.314	0.155	0.063
SM with heterosis over mid-parent	Set I	0.221	0.288	−0.273	0.555[**]	0.059	0.136	−0.357	−0.203
	Set II	0.257	−0.021[a]	−0.358[a]	−0.172	0.378[*]	−0.308	0.436[*]	0.221
SM with heterosis over better-parent	Set I	0.205	0.116	−0.211	0.280	−0.069	0.236	−0.270	−0.157
	Set II	0.344	−0.078[a]	−0.436[a,*]	−0.119	0.328	−0.277	0.432[*]	0.108

[a] Based on one season data only.
[*] Significant at P = 0.05.
[**] Significant at P = 0.01.

performance, or mid-parent heterosis for grain yield in either set of hybrids. Chowdari et al. [10] also found a non-significant correlation between genetic distances based on 20 RAPDs and mid-parent heterosis for grain yield in pearl millet. Similarly, Teklewold and Becker [9] found genetic distance estimation from phenotypic traits to be a better predictor of mid-parent heterosis and F_1 performance than genetic distance estimated from RAPD markers in Ethiopian mustard (*Brassica carinata*). Riday et al. [7] found a significant correlation of heterosis with morphological distance but not with molecular distance based on microsatellite and AFLP markers in two subspecies of *Medicago sativa*.

In contrast to our observation of lack of correlation between molecular marker-based genetic distance and heterosis for grain yield in both sets of hybrids in pearl millet, Knaak and Ecke [23], and Riaz et al. [5] reported the utility of molecular marker-based distance among parental lines in rapeseed to predict heterosis, especially when the parents were genetically related. In our study also, most of the B and R lines fell into clear-cut separate broad-based diverse gene pools. The wide diversity between B and R lines is a consequence of trait-specific breeding, which B and R lines undergo during their development process, and also of the involvement of separate breeding stocks in their parentage, leading to high levels of genetic unrelatedness between B and R lines. This high level of unrelatedness might have resulted in a lack of correlation between genetic distance and heterosis in B × R crosses in both sets. Other likely reasons for low or no correlation between molecular distance and heterosis and/or F_1 performance might be inadequate genome coverage, or due to random dispersion of molecular markers [24]. The presence of multiple alleles [25] and epistasis [2] could also cause the low correlation of SM with heterosis and F_1 performance.

Significant correlation between genetic distance and heterosis was reported in intra-group crosses of inbred lines compared to intergroup crosses in maize [6,26]. Thus, making intra-group crosses in our materials, say B × B or R × R, might reveal a significant linear relationship with heterosis and lead to identification of heterotic crosses. This approach can help hybrid parental line development programs to develop parents (B lines and R lines) with high yield *per se*. Also, there is a need to investigate the relationship between SM and combining ability of parents, an important component of hybrid breeding to enable breeders to predict heterosis based on genetic distances between parents.

5. Conclusions

This study based on phenotypic traits and molecular markers in diverse hybrid parents showed that molecular marker-based distance was not strongly correlated with phenotype-based distance, a conclusion that invites further investigation with a higher number of markers evenly distributed across all linkage groups. Also, it revealed that marker-based distance was not a reliable predictor of heterosis in hybrids produced from crosses between maintainer and restorer parents in pearl millet. This observation might be due to B and R lines behaving as parts of two broad-based diverse and different gene pools, leading to higher levels of genetic diversity where heterosis might not be correlated with diversity. It might also be due to the concentration of the markers used in the study in relatively short segments of chromosomes that lacked linkage with heterosis for grain yield and its component traits. Given that earlier studies have reported higher probabilities of predicting heterosis in intra-group crosses, we suggest that B line × B line and R line × R line intra-group crosses should be investigated in search of a linear relationship between heterosis and genetic distance. This can also help line breeding programs to generate hybrid parents with higher *per se* productivity. However, phenotypic trait-based genetic distance was, to

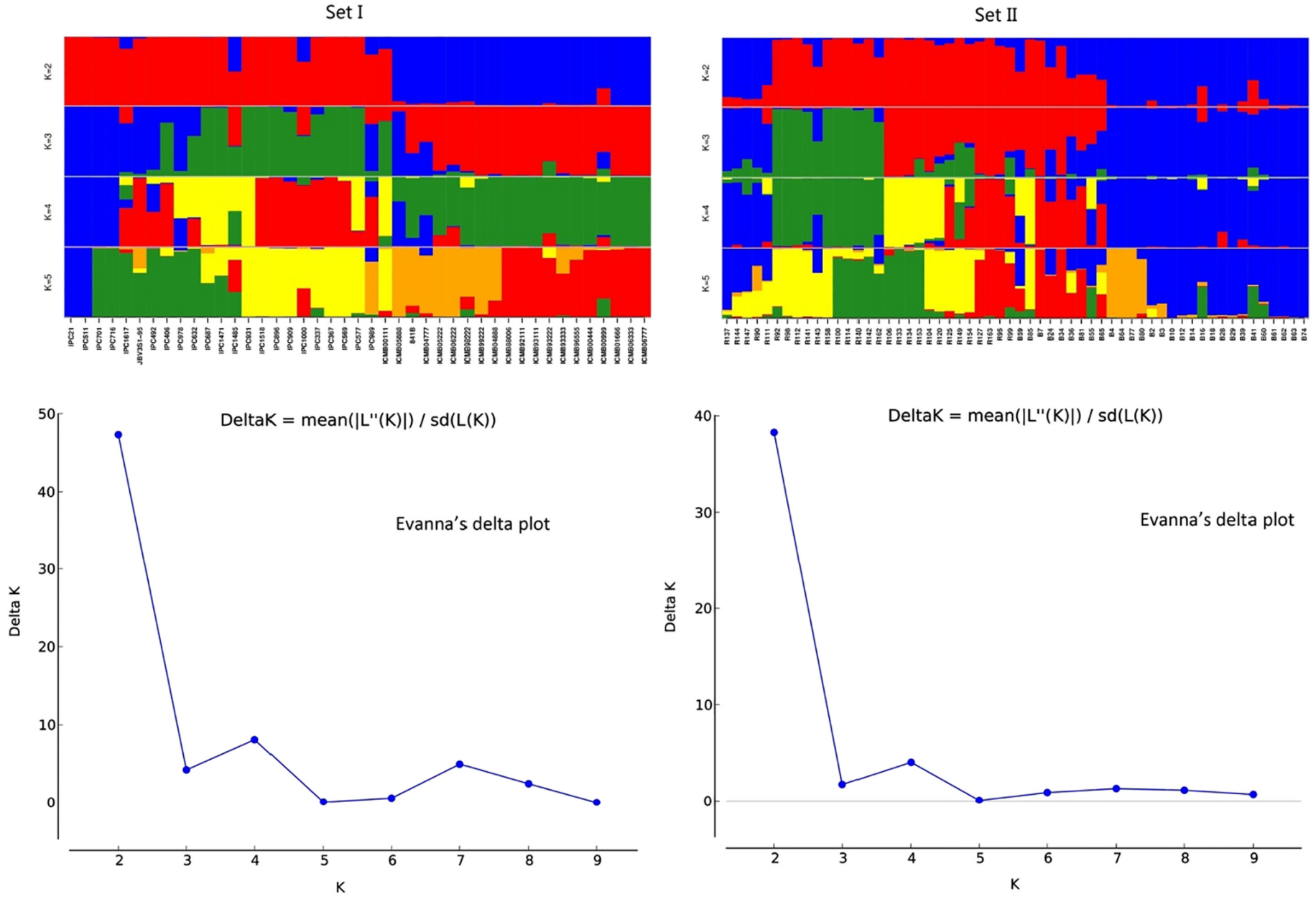


Fig. 3 – Structure-based population stratification analysis of B- and R-lines of Set I and Set II.

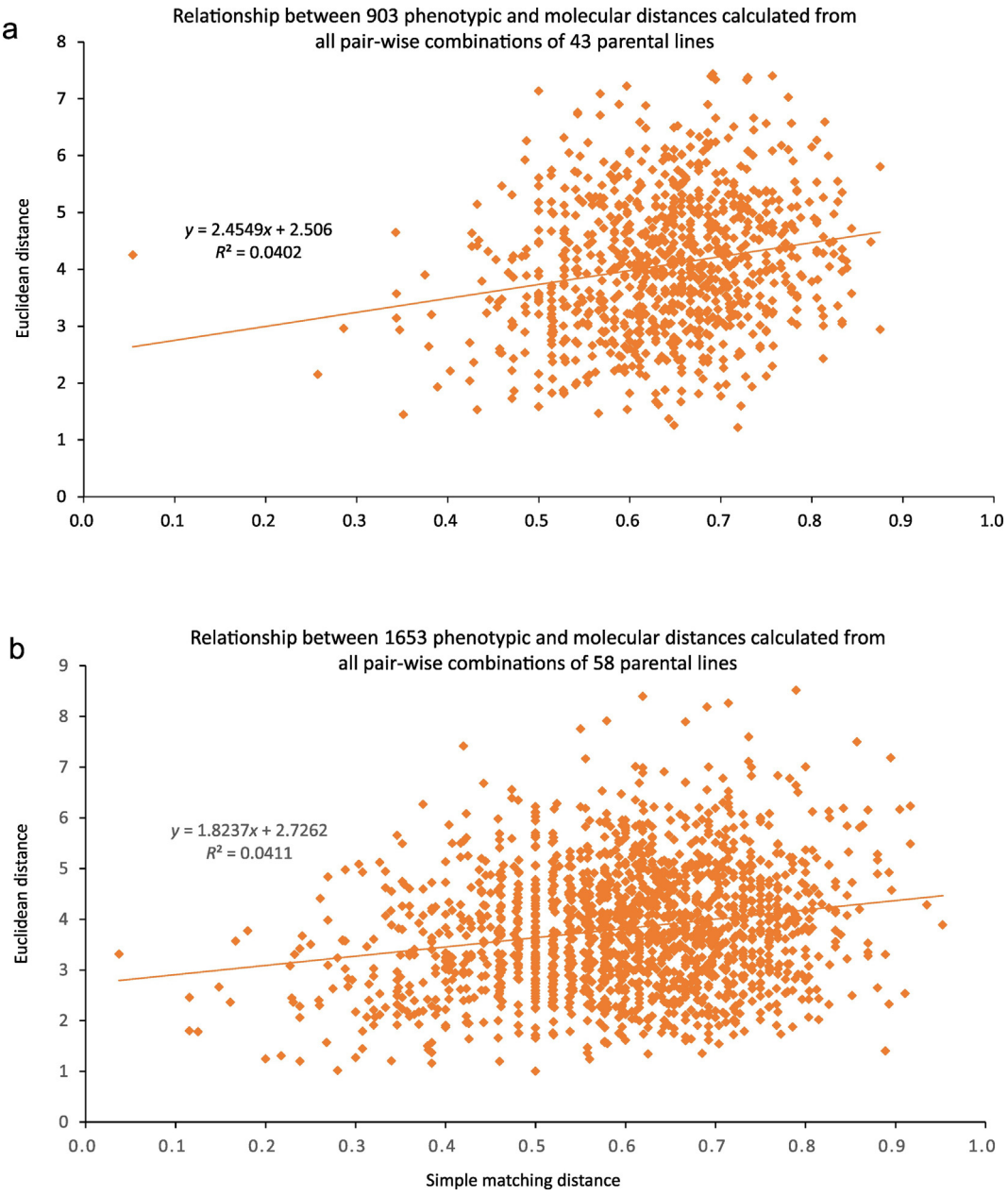


Fig. 4 – Relationship between phenotyping distance (ED) and molecular distance (SM) based on all pairwise combinations of parental lines of (a) Set I, and (b) Set II.

some extent, able to predict mid-parent heterosis and better-parent heterosis for grain yield. Accordingly, it is suggested that the relationship between phenotypic distance and heterosis should be further investigated to determine whether phenotypic distance can be reliably used to select potential parents for heterotic and high-yielding hybrids.

Acknowledgments

This research was supported by the ICRISAT-Sehgal Family Foundation Endowment Fund (YSFF06) and the CGIAR Research Program on Dryland Cereals.

REFERENCES

[1] O.P. Yadav, R.S. Mahala, K.N. Rai, S.K. Gupta, B.S. Rajpurohit, H.P. Yadav, Pearl Millet Seed Production and Processing, All India Coordinated Research project on Pearl millet, Indian Council of Agricultural Research, Mandor, Jodhpur, Rajasthan, India, 2015.

[2] R.H. Moll, J.H. Lonnquist, J.V. Fortuno, E.C. Johnson, The relationship of heterosis and genetic divergence in maize, Genetics 52 (1965) 139–144.

[3] D.S. Falconer, T.F.C. Mackay, Introduction to Quantitative Genetics, 4th edition Longmans Green, Essex, UK, 1996.

[4] Q.F. Zhang, Z.Q. Zhou, G.P. Yang, C.G. Xu, K.D. Liu, Molecular marker heterozygosity and hybrid performance in *indica* and *japonica* rice, Theor. Appl. Genet. 93 (1996) 1218–1224.

[5] A. Riaz, G. Li, Z. Quresh, M.S. Swati, C.F. Quiros, Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance, Plant Breed. 120 (2001) 411–415.
[6] B.A. Kiula, N.G. Lyimo, A.M. Botha, Association between AFLP-based genetic distance and hybrid performance in tropical maize, Plant Breed. 127 (2008) 140–144.
[7] H. Riday, E.C. Brummer, T.A. Campbell, D. Luth, P.M. Cazcarro, Comparisons of genetic and morphological distance with heterosis between *Medicago sativa* subsp. *sativa* and subsp. *falcata*, Euphytica 131 (2003) 37–45.
[8] L.F. Geleta, M.T. Labuschagne, C.D. Viljoen, Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper, Plant Breed. 123 (2004) 467–473.
[9] A. Teklewold, H.C. Becker, Comparison of phenotypic and molecular distances to predict heterosis and F_1 performance in Ethiopian mustard (*Brassica carinata* A. Braun), Theor. Appl. Genet. 112 (2006) 752–759.
[10] K.V. Chowdari, S.R. Venkatachalam, A.P. Davierwala, V.S. Gupta, P.K. Ranjekar, O.P. Govila, Hybrid performance and genetic distance as revealed by the $(GATA)_4$ microsatellite and RAPD markers in pearl millet, Theor. Appl. Genet. 97 (1998) 163–169.
[11] T. Nepolean, S.K. Gupta, S.L. Dwivedi, R. Bhattacharjee, K.N. Rai, C.T. Hash, Genetic diversity in maintainer and restorer lines of pearl millet, Crop Sci. 52 (2012) 2555–2563.
[12] S.K. Gupta, T. Nepolean, S.M. Sankar, A. Rathore, R.R. Das, K.N. Rai, Patterns of molecular diversity in current and previously developed hybrid parents of pearl millet [*Pennisetum glaucum* (L.) R. Br.], Am. J. Plant Sci. 06 (2015) 1697–1712.
[13] X. Perrier, A. Flori, F. Bonnot, Data analysis methods, in: P. Hamon, M. Seguin, X. Perrier, J.C. Glaszmann (Eds.), *Genetic Diversity of Cultivated Tropical Plants*, Science Publishers, Enfield, USA 2003, pp. 43–76.
[14] O.P. Yadav, S.E. Mitchell, A. Zamora, T.M. Fulton, S. Kresovich, Development of new simple sequence repeat markers for pearl millet, SAT eJournal 3 (2007) 34.
[15] S. Senthilvel, B. Jayashree, V. Mahalakshmi, P.S. Kumar, S. Nakka, T. Nepolean, C.T. Hash, Development and mapping of Simple Sequence Repeat markers for pearl millet from data mining of Expressed Sequence Tags, BMC Plant Biol. 8 (2008) 119.
[16] V. Rajaram, T. Nepolean, S. Senthilvel, R.K. Varshney, V. Vadez, R.K. Srivastava, T.M. Shah, A. Supriya, S. Kumar, B.R. Kumari, A. Bhanuprakash, M.L. Narasu, Oscar Riera-Lizarazu, C.T. Hash, Pearl millet [*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs, BMC Genomics 14 (2013) 159.
[17] K.H. Moumouni, B.A. Kountche, M. Jean, C.T. Hash, Y. Vigouroux, B.I.G. Haussmann, F. Belzile, Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach, Mol. Breed. (2015) 35.
[18] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[19] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015.
[20] SAS Institute, Base SAS 9.4 Procedures Guide, SAS Institute, Cary, North Carolina, USA, 2015.
[21] J. Burstin, A. Charcosset, Y. Barriere, Y. Hebert, D. Vienne, C. Damerval, Molecular markers and protein quantities as genetic descriptors in maize. II. Prediction of performance of hybrids for forage traits, Plant Breed. 114 (1995) 427–433.
[22] J. Burstin, A. Charcosset, Relationship between phenotypic and marker distances: theoretical and experimental investigations, J. Hered. 79 (1997) 477–483.
[23] C. Knaak, W. Ecke, Genetic diversity and hybrid performance in European winter oilseed rape (*Brassica napus* L.), *Proceedings of the 9th International Rapeseed Congress, July 4–7, 1995, Cambridge, UK* 1995, pp. 110–112.
[24] R. Bernardo, Relationship between single-cross performance and molecular marker heterozygosity, Theor. Appl. Genet. 83 (1992) 628–634.
[25] C.E. Cress, Heterosis of the hybrid related to gene frequency differences between two populations, Genetics 53 (1966) 86–94.
[26] A. Menkir, A. Melake-Berhan, C. The, I. Ingelbrecht, A. Adepoju, Grouping of tropical mid-altitude maize inbred lines on the basis of yield data and molecular markers, Theor. Appl. Genet. 108 (2004) 1582–1590.