Theor Appl Genet (2005) 111: 95–99
DOI 10.1007/s00122-005-1995-7

ORIGINAL PAPER

H. P. Piepho

Permutation tests for the correlation among genetic distances and measures of heterosis

Received: 4 November 2004 / Accepted: 7 March 2005 / Published online: 7 April 2005

Abstract It is often found that heterosis tends to increase with genetic distance of the parents, though the correlation is not usually very close. It is therefore important to test the null hypothesis that the correlation is zero. The present work shows that standard procedures tend to yield too liberal tests, owing to the lack of independence among genetic distances and among heterosis estimates. A valid alternative is to use a permutation test, which was first suggested by Mantel [(1967) Cancer Res 27: 209–220). This test is well-known among plant breeders and geneticists, who often use it to test the correlation among two distance matrices. Its use is not restricted to the comparison of distance matrices. This is demonstrated in the present work, using two published datasets on marker-based genetic distances of maize inbreds or populations and heterosis of their crosses. It is shown that the test is also applicable in the presence of missing data.

Introduction

In studies on heterosis, it is frequently of interest to assess the relationship between genetic distance among parents and the heterosis observed in their crosses. Most studies have been conducted with the hope of revealing strong relationships that could be of practical interest for hybrid performance prediction. Experimental results have proven highly dependent on the genetic material, and several syntheses (Melchinger 1999) and theoretical studies (Charcosset and Essioux 1994) have clarified the reasons for this variation. Several instances exist, such as hybrids between inbred lines from different heterotic groups, where no relationship is expected.

Most authors use either Pearson's correlation or simple linear regression for assessing the heterosis–distance association (Barbosa et al. 1996, 2003; Betran et al. 2003; Drinic et al. 2002; Gopal and Minocha 1997; Girke et al. 2001; Joyce et al. 1999; Kwon et al. 2002a, 2002b; Lanza et al. 1997; Meng et al. 1996; Parentone et al. 2001; Riaz et al. 2001; Maroof et al. 1997; Reif et al. 2003a, 2003b; Sant et al. 1999; Shieh and Thseng 2002; Xiao et al. 1996). Typically, all pairwise genetic distances and all pairwise measures of heterosis or similar quantities (specific combining ability, etc.) are compiled into symmetric association matrices and related to one another by correlation or regression analysis. Estimates of correlation or the regression coefficient per se are the most crucial pieces of information obtained in such analyses. It is often found that the correlation is not very close, and an important additional question is whether the correlation between genetic distance and heterosis is significantly different from zero.

Genetic distance and heterosis can be regarded as different measures of association among the hybrid parents. Generally, when determining the significance of correlation between two association matrices, it is important to note that elements of either matrix are stochastically dependent (see "Appendix" for two examples). An appropriate procedure in this situation is to use a permutation test (Sokal and Rohlf 1995; Manly 1997), which was first proposed by Mantel (1967) for the correlation among two distance matrices and which is readily available in many packages, e.g., NTSYS-pc (Rohlf 1998). Whereas it is mostly used for comparing two distance matrices, the test is more generally applicable for testing the correlation of any type of association matrix.

The purposes of the present work are (1) to study the effect of ignoring the lack of independence among genetic distances and among heterosis estimates when testing their correlation and (2) to increase the awareness among plant breeders and geneticists that Mantel's

Communicated by H. C. Becker

H. P. Piepho
Bioinformatics Unit, Universität Hohenheim,
Fruwirthstrasse 23, 70599 Stuttgart, Germany
E-mail: piepho@uni-hohenheim.de

permutation test appropriately addresses the lack-of-independence problem. The permutation test is exemplified using two published datasets on maize.

Simulation study

Let d_{ij} and h_{ij} be the associations among parents i and j ($i, j = 1, \ldots, n$) regarding two characteristics. For example, d_{ij} may be a genetic distance among parents i and j, whereas h_{ij} is the mid-parent heterosis in their crosses. Data x_i, y_i, and z_{ij} ($i, j = 1, \ldots, n$) were generated as independent standard normal deviates. A distance was computed as $d_{ij} = |x_i - x_j|$. To study the test of correlation between two distances, a second distance was computed as $h_{ij} = |y_i - y_j|$. For mimicking the correlation among a distance and an estimate of mid-parent heterosis and of better-parent heterosis, $h_{ij} = z_{ij} - (y_i + y_j)/2$ and $h_{ij} = z_{ij} - \max(y_i, y_j)$ were computed, respectively. In either case, the correlation of d_{ij} and h_{ij} was tested using standard procedures (Sokal and Rohlf 1995), assuming independence among the d_{ij} and among the h_{ij}, i.e., $t = \sqrt{n(n-1)/2-2}|r|/.\sqrt{1-r^2}$, where r is the sample correlation among d_{ij} and h_{ij}, was compared to a t-distribution with $n(n-1)/2-2$ degrees of freedom. This standard test assumes normality, which in this simulation will be violated for both d_{ij} and h_{ij}. Thus, for comparison, Spearman's rank correlation, r_S, was computed and tested by the same t-test as Pearson's correlation r. This test is approximate in small samples. For very small n, one should use an exact test in pratice. For brevity, only results for the approximate test are shown, which is valid asymptotically, providing the independence assumption holds. For each setting, 100,000 simulation runs were performed. Results are presented in Table 1.

It is seen that both tests become rather liberal as n increases. This behavior does not even vanish asymptotically, as n increases towards infinity. The main problem here is not lack of normality, otherwise the test based on Spearman's correlation should be valid. The problem rests with the lack of independence of entries in both matrices.

Not too much should be made of the differences among the three types of estimates based on phenotypic data (heterosis and simple difference among parents). Specifically, the results do not lend themselves to identify conditions under which the standard test might still be trusted. The studied settings are just arbitrary examples to demonstrate the general problem over a range of cases.

When entries in one of the two association matrices are stochastically independent, the standard test of the correlation (or regression) coefficient is valid. This can be seen by observing that the t-test of the correlation coefficient is equivalent to that of the regression coefficient. The test of the regression coefficient essentially treats the regressor variable as fixed quantity. Thus, stochastical dependence among observed levels of the regressor variable becomes a non-issue. Taking the association matrix with dependent entries as the regressor variable, we see that the standard test of the regression coefficient, and thus that of the correlation coefficient, must be valid. For example, when simulating $d_{ij} = |x_i - x_j|$and $h_{ij} = z_{ij}$ for $n = 20$, the simulated type I errors at $\alpha = 5\%$ were 0.0509 for Spearman and 0.0517 for Pearson.

It is likely that independence of entries in an association matrix will be rare in practice. An example is F_1 performance in a case where there are no additive gene effects. The presence of additive gene effects will induce dependence among F_1 performances of hybrids involving the same parent. Thus, whereas in some rare cases the standard test may be valid under strong assumptions, it is suggested here that as a rule, Mantel's test (see below) be used instead.

The permutation test

It is convenient to collect d_{ij} and h_{ij} into symmetric association matrixes D and H, respectively. The null hypothesis of independence among d_{ij} and h_{ij} may be tested by a permutation procedure as follows (Mantel 1967; Sokal and Rohlf 1995; Good 2000):

1. Compute correlation r from the upper triangular entries of the observed D and H.
2. Generate a permuted set of labels 1 to n, where n is the dimension of D and H; use this set to permute columns and rows in D; compute r from the upper triangular entries of the permuted D and H (r_p); repeat this M times.
3. Let m_T be the number of times, where $r_p \geq r$. The one-sided permutation P-value is computed as $P = (m_T + 1)/(M + 1)$. A two-sided P-value is computed by the same equation, letting m_T be the number of times, where $|r_p| \geq |r|$.

This procedure may be applied also when some observations in D and H are missing completely at random. Moreover, the test can be used with estimates of the regression coefficient in place of the correlation.

Example

To illustrate the permutation test, data on seven maize populations by Reif et al. (2003a) were used. Heterosis in days to silking and modified Roger's distances based on SSR marker data are shown in Table 2. Pearson's correlation coefficient (r) among genetic distance and heterosis is −0.44. Using the standard two-sided t-test, this has a P-value of 0.0466229, which is just significant at the conventional 5% level. By comparison, the two-sided P-value based on 100,000 permutations is 0.1446286, which is not significant. In this case, using the permutation procedure leads to the opposite conclusion

Table 1 Results of simulation study on standard test of correlation (Pearson and Spearman) among square matrices D and H at a nominal significance level of $\alpha = 5\%$, using 100,000 simulation runs per setting

Sample size	Type I error of standard test ($\alpha = 5\%$) for correlation of distance $d_{ij} = \lvert x_i - x_j \rvert$ with					
	Another distance $h_{ij} = \lvert y_i - y_j \rvert$		Mid-parent heterosis $h_{ij} = z_{ij} - (y_i + y_j)/2$		Better-parent heterosis $h_{ij} = z_{ij} - \max(y_i, y_{jj})$	
n	Pearson	Spearman	Pearson	Spearman	Pearson	Spearman
3	0.0501	0.3339[a]	0.0505	0.3328[a]	0.0490	0.3317[a]
4	0.0726	0.0551	0.0514	0.0590	0.0547	0.0609
5	0.0717	0.0628	0.0530	0.0568	0.0607	0.0626
6	0.0734	0.0645	0.0558	0.0573	0.0667	0.0651
7	0.0764	0.0662	0.0585	0.0571	0.0752	0.0697
8	0.0802	0.0677	0.0613	0.0600	0.0803	0.0734
9	0.0868	0.0717	0.0630	0.0606	0.0893	0.0793
10	0.0957	0.0765	0.0645	0.0604	0.0973	0.0855
11	0.1057	0.0818	0.0675	0.0631	0.1058	0.0929
12	0.1137	0.0881	0.0708	0.0662	0.1106	0.0960
13	0.1247	0.0916	0.0744	0.0687	0.1221	0.1060
14	0.1376	0.0994	0.0770	0.0694	0.1275	0.1083
15	0.1459	0.1047	0.0799	0.0723	0.1376	0.1162
16	0.1583	0.1115	0.0821	0.0730	0.1454	0.1226
17	0.1681	0.1170	0.0826	0.0743	0.1524	0.1274
18	0.1774	0.1226	0.0884	0.0790	0.1618	0.1343
19	0.1881	0.1301	0.0905	0.0803	0.1695	0.1398
20	0.1985	0.1370	0.0955	0.0835	0.1786	0.1474

Entries d_{ij} in matrix D are distances among parents, whereas entries h_{ij} in matrix H are either distances among parents or heterosis estimates of their crosses. Distances and heterosis estimates are computed based on simulated data x_i, y_i, and z_{ij} (for details, see text)

[a] For small n, results are dominated by the discreteness of the null distribution, and an exact test is to be preferred in practice. For $n = 3$, the smallest possible exact P-value is 0.3333, which coincides with the simulated P-value. For $n > 3$, the exact test yields somewhat more conservative simulated P-values (results not shown)

Table 2 Genetic distances for seven maize populations computed from SSR data (*below diagonal* modified Roger's distances) and mid-parent heterosis for days to silking of crosses (*above diagonal*) (reproduced from Reif et al. 2003a)

Population	Pool24	Pop21	Pop22	Pop25	Pop29	Pop32	Pop43
Pool24	–	0.5	−0.4	0.7	−0.3	−0.7	−1.3
Pop21	0.219	–	−0.4	−0.4	−0.4	−0.7	−1.2
Pop22	0.203	0.222	–	−0.6	−1.5	−1.2	−1.8
Pop25	0.224	0.272	0.250	–	−0.9	−0.9	−0.5
Pop29	0.216	0.236	0.233	0.259	–	−0.7	−0.2
Pop32	0.270	0.305	0.284	0.263	0.285	–	−0.9
Pop43	0.248	0.286	0.268	0.278	0.274	0.318	–

compared to the standard test. The observed difference reflects the fact that the standard test is on the liberal side, as was also found in the simulation study.

To exemplify use of the procedure with missing observations, data on 18 maize inbred lines and their hybrids published by Barbosa et al. (2003) were used. Heterosis estimates and genetic distances based on AFLP data and Jaccard's similarity are reproduced in Table 3. There are five missing values for the heterosis estimates. r among genetic distance and heterosis is 0.67. By the standard test, this has a P-value of 8.99×10^{-21}, whereas the two-sided P-value based on 100,000 permutations is 9.99×10^{-6}. Both P-values are highly significant, though there is a notable difference, the standard test being considerably more significant. Again, this confirms that the standard test is on the liberal side. Another reason for the difference is the discreteness of the permutation distribution.

Concluding remarks

The present work has not addressed the computation of power. There does not seem to be a competitive alternative to the permutation approach suggested in this work. Therefore, the use of power calculations would be mainly restricted to the determination of sample size. Such calculations would require an explicit genetic model for heterosis and for the relationship of genetic distance and heterosis. The type of model would be highly dependent on the types of species and of population studied. Also, with heterosis being a very complex phenomenon, this would require a rather complex model to be practically relevant, and it is not easy to decide which of the many possible candidate models is preferable in a given context. Parameterization of all model parameters poses an additional problem. Thus, it does not seem straightforward to devise a simple and practical stragegy for power calculations.

Plant breeders, geneticists, and researchers in many other fields of science routinely use Mantel's permutation test for comparing two distance matrices (Dutilleul et al. 2000), thus appropriately addressing the lack-of-independence problem. The main message of the present

Table 3 Genetic distances for 18 maize inbred lines computed from AFLP data (*below diagonal* Jaccard's similarity coefficient) and mid-parent heterosis of crosses (*above diagonal*) (reproduced from Barbosa et al. 2003)

Lines	105–01	105–02	105–03	105–04	105–05	105–06	105–07	105–08	106–09	106–10	106–11	106–12	106–13	106–14	106–15	106–16	106–17	106–18
105–01	–	4.64	4.25	4.20	5.36	5.27	5.92	3.97	4.82	3.88	4.60	4.33	3.91	4.31	4.99	4.46	4.11	4.00
105–02	0.48	–	4.06	4.35	4.80	4.02	3.55	3.48	4.55	4.14	4.17	4.22	3.30	4.89	5.66	4.14	4.10	4.64
105–03	0.55	0.52	–	3.93	4.00	5.33	4.36	2.50	3.91	3.86	4.37	4.20	3.04	4.51	3.24	2.84	3.85	3.92
105–04	0.52	0.47	0.55	–	4.59	5.12	4.58	3.92	5.03	3.80	4.34	4.07	3.29	4.37	5.87	1.87	4.24	4.18
105–05	0.49	0.46	0.57	0.54	–	5.82	4.92	4.87	3.45	4.66	5.12	5.19	3.86	5.21	6.40	5.38	4.51	4.35
105–06	0.54	0.51	0.57	0.56	0.55	–	2.59	5.69	5.26	5.31	6.04	5.79	4.43	4.54	5.69	5.23	4.81	5.40
105–07	0.51	0.52	0.53	0.58	0.58	0.29	–	4.51	5.23	4.75	5.00	5.47	4.21	4.64	5.32	4.03	4.84	4.60
105–08	0.56	0.56	0.40	0.56	0.62	0.58	0.56	–	4.72	4.46	4.95	4.76	3.89	4.71	4.93	4.66	4.40	4.77
106–09	0.63	0.61	0.67	0.64	0.63	0.67	0.66	0.69	–	3.66	5.08	4.25	3.41	4.87	6.06	–	3.51	5.42
106–10	0.56	0.55	0.63	0.56	0.56	0.62	0.61	0.67	0.55	–	4.69	3.20	4.78	4.10	5.15	–	4.25	4.30
106–11	0.60	0.57	0.66	0.64	0.65	0.63	0.64	0.69	0.66	0.58	–	–	3.38	1.75	4.55	6.03	1.17	0.96
106–12	0.61	0.58	0.67	0.66	0.66	0.65	0.65	0.70	0.67	0.58	0.09	–	3.64	1.32	4.13	5.36	1.63	1.31
106–13	0.62	0.61	0.63	0.60	0.63	0.64	0.64	0.61	0.62	0.60	0.63	0.64	–	4.14	4.11	–	4.01	3.32
106–14	0.60	0.58	0.65	0.64	0.63	0.63	0.63	0.68	0.65	0.58	0.34	0.34	0.63	–	5.30	6.65	1.51	2.05
106–15	0.59	0.61	0.66	0.62	0.63	0.64	0.64	0.66	0.63	0.57	0.53	0.57	0.58	0.56	–	–	4.47	3.74
106–16	0.54	0.52	0.57	0.57	0.55	0.55	0.56	0.62	0.66	0.58	0.67	0.68	0.65	0.67	0.65	–	5.44	4.00
106–17	0.57	0.54	0.64	0.62	0.61	0.62	0.63	0.67	0.65	0.53	0.28	0.31	0.60	0.36	0.56	0.64	–	0.44
106–18	0.58	0.55	0.66	0.62	0.63	0.62	0.63	0.68	0.65	0.54	0.28	0.31	0.61	0.36	0.54	0.65	0.14	–

work is that the same problem occurs with other types of association data, namely with estimates of heterosis (mid-parent or better-patent) or specific combining ability, and that Mantel's test can be used in these cases as well. An SAS/IML macro for the permutation test will be made available on the author's homepage at http://www.uni-hohenheim.de/bioinformatik/.

Appendix

Using results on the distribution of linear functions of random variables, it is straightforward to see that heterosis estimates involving the same parent are stochastically dependent. For example, consider two mid-parant heterosis estimates $h_{ij} = z_{ij} – (y_i + y_j)/2$ (parents i and j) and $h_{\mathrm{ik}} = z_{\mathrm{ik}} – (y_i + y_k)/2$ (parents i and k), where z_{ij}, z_{ik}, y_i, y_j, and y_k are independent normal deviates with constant variance. The correlation of h_{ij} and h_{ik} is 1/6, so the two estimates are stochastically dependent.

Here is a simple example demonstrating that genetic distances involving the same parent are stochastically dependent. Assume that three lines i, j, and k are selected at random from a population of inbred lines, and consider a single locus with alleles a_1 und a_2. Let p be the allele frequency of a_1 and q the allele frequency of a_2 and define a dummy variable, w, indexed by parents. For example, $w_i = 0$ when parent i is homozygous for allele a_1 and $w_i = 1$ when parent i is homozygous for allele a_2. The Euclidean distances of the three parents at the locus are $d_{ij} = |w_i \text{ -} w_j|$, $d_{ik} = |w_i \text{ -} w_k|$, and $d_{jk} = |w_j \text{ -} w_{\mathrm{k}}|$. Table 4 gives the joint distribution of the three pairwise distances. To see that the distances are stochastically dependent, consider the conditional probability of the event $d_{ij} = 0$, given that $d_{ik} = 1$, $P(d_{ij} = 0 | d_{ik} = 1) = (p^2 q + pq^2)/(p^2 q + pq^2 + p^2 q + pq^2) = 0.5$, and the marginal probability of the event $d_{ij} = 0$, $P(d_{ij} = 0) = p^3 + p^2 q + pq^3 + q^3$. Under stochastical independence, both probabilities must be equal. Obviously, $P(d_{ij} = 0 | d_{ik} = 1) \neq P(d_{ij} = 0)$ when $p \neq q$, in which case the distances d_{ij} and d_{ik} are stochastically dependent. For example, when $p = 0.7$, one finds $P(d_{ij} = 0) = 0.58 \neq 0.5 = P(d_{ij} = 0 | d_{\mathrm{ik}} = 1)$.

Table 4 Possible realizations of random draws of z_{i}, z_{j}, and z_{k} with associated genetic distances $d_{ij} = |z_{\mathrm{i}} - z_{\mathrm{j}}|$, $d_{ik} = |z_i - z_k|$, and $d_{\mathrm{jk}} = |z_{\mathrm{j}} - z_k|$ and joint probabilities $P(d_{ij}, d_{ik}, d_{jk})$

Dummy variables			Distances			Joint probability[a]
z_i	z_j	z_k	d_{ij}	d_{ik}	d_{jk}	$P(d_{ij}, d_{ik}, d_{jk})$
0	0	0	0	0	0	p^3
0	0	1	0	1	1	$p^2 q$
0	1	0	1	0	1	$p^2 q$
0	1	1	1	1	0	pq^2
1	0	0	1	1	0	$p^2 q$
1	0	1	1	0	1	pq^2
1	1	0	0	1	1	pq^2
1	1	1	0	0	0	q^3

[a] p Allele frequency of allele a_1, q allele frequency of allele a_2

References

Barbosa NJF, Neto JF, Sorrells ME, Cisar G (1996) Prediction of heterosis in wheat using coefficient of parentage and RFLP-based estimates of genetic relationship. Genome 39:1142–1149

Barbosa AMM, Geraldi IO, Benchimol LL, Garcia AAF, Souza CL, 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

Betran FJ, Ribaut JM, Beck D, de Leon DG (2003) Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. Crop Sci 43:797–806

Charcosset A, Essioux L (1994) The effect of population-structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 89:336–343

Drinic SM, Trifunovic S, Drinic G, Konstantinov K (2002) Genetic divergence and its correlation to heterosis in maize as revealed by SSR-based markers. Maydica 47:1–8

Dutilleul P, Stockwell JD, Frigon D, Legendre P (2000) The Mantel test versus Pearson's correlation analysis: assessment of the difference for biological and environmental studies. J Agric Biol Environ Stat 5:131–150

Girke A, Engqvist G, Becker HC (2001) Predicting heterosis from genetic distances for RFLP markers in resynthesized oilseed rape. In: Gallais A, Dillmann C, Goldringer I (eds) Quantitative genetics and breeding methods: the way ahead. Proceedings of the 11th Meeting of the EUCARPIA section biometrics in plant breeding, Paris, France, 30–31 August–1 September 2000, pp 257–262

Good IJ (2000) Permutation tests: a practical guide to resampling methods for testing hypotheses, 2nd edn. Springer, Berlin Heidelberg New York

Gopal J, Minocha JL (1997) Genetic divergence for cross prediction in potato. Euphytica 97:269–275

Joyce TA, Abberton MT, Michaelson-Yeates TPT, Forster JW (1999) Relationships between genetic distance measured by RAPD-PCR and heterosis in inbred lines of white clover (*Trifolium repens* L.). Euphytica 107:159–165

Kwon SJ, Ahn SN, Jeong EG, Jeon YH, Hwang HG, Choi HC, Moon-HP (2002a) Relationship between genetic divergence and hybrid performance in japonica rice grown in a cold water-irrigated field. Euphytica 128:389–396

Kwon SJ, Ha WG, Hwang HG, Yang SJ, Choi HC, Moon HP, Ahn SN (2002b) Relationship between heterosis and genetic divergence in "Tongil"-type rice. Plant Breed 121:487–492

Lanza LLB, Souza CL, Ottoboni LMM, Vieira MLC, Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94:1023–1030

Manly BFJ (1997) Randomization, bootstrap and Monte Carlo methods in biology, 2nd edn. Chapman & Hall, New York

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Maroof MAS, Yang GP, Zhang QF, Gravois KA (1997) Correlation between molecular marker distance and hybrid performance in U.S. southern long grain rice. Crop Sci 37:145–150

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. Proceedings of an international symposium, CIMMYT, Mexico City, Mexico, 17–22 August 1997. Crop Science Society of America, Madison, Wisconsin, USA, pp 99–118

Meng A, Gong G, Chen D, Zhang H, Qi S, Tang H, Gao Z (1996) DNA fingerprint variability within and among parental lines and its correlation with performance of F1 laying hens. Theor Appl Genet 92:769–776

Parentoni SN, Magalhaes JV, Pacheco CAP, Santos MX, Abadie T, Gama EEG, Guimaraes PEO, Meirelles WF, Lopes MA, Vasconcelos MJV, Paiva E (2001) Heterotic groups based on yield-specific combining ability data and phylogenetic relationship determined by RAPD markers for 28 tropical maize open pollinated varieties. Euphytica 121:197–208

Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Srinivasan G, Bohn M, Frisch M (2003a) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci 43:1275–1282

Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Beck D, Bohn M, Frisch M (2003b) Use of SSRs for establishing heterotic groups in subtropical maize. Theor Appl Genet 197:947–957

Riaz A, Li G, Quresh Z, Swati MS, Quiros CF (2001) Genetic diversity of oilseed Brassica napus inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. Plant Breed 120:411–415

Rohlf FJ (1998) NTSYS-pc: numerical taxonomy and multivariate analysis system. Version 2.0. Exeter Software, Setauket

Sant VJ, Patankar AG, Sarode ND, Mhase LB, Sainani MN, Deshmukh RB, Ranjekar PK, Gupta VS (1999) Potential of DNA markers in detecting divergence and in analysing heterosis in Indian elite chickpea cultivars. Theor Appl Genet 98:1217–1225

Shieh GJ, Thseng FS (2002) Genetic diversity of Tainan-white maize inbred lines and prediction of single cross hybrid performance using RAPD markers. Euphytica 124:307–313

Sokal RR, Rohlf FJ (1995) Biometry, 3rd edn. WH Freeman, New York

Xiao J, Li J, Yuan L, McCouch SR, Tanksley SD (1996) Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. Theor Appl Genet 92:637–643