Theor Appl Genet (2001) 102:741–750

V. Lefebvre · B. Goffinet · J. C. Chauvet · B. Caromel
P. Signoret · R. Brand · A. Palloix

Evaluation of genetic distances between pepper inbred lines for cultivar protection purposes: comparison of AFLP, RAPD and phenotypic data

Received: 25 March 2000 / Accepted: 11 July 2000

Summary We evaluated concordance of AFLP and RAPD markers for estimating genetic distances of 47 pepper inbred lines belonging to five varietal types. It enabled us to see the efficiency of these markers for identification, estimation of distances between varieties and variety discrimination. Genetic distance and multidimensional scaling results showed a general agreement between AFLP and RAPD markers. Based on pattern scores, dendrograms were produced by the UPGMA method. Phenetic trees based on molecular data were consistent with the classification of variety group. The precision of the estimation of the genetic distance was given. The molecular genetic distances were correlated with distances based on a set of discriminating agronomic traits measured for identification and distinctiveness tests. The relationship between molecular and morphological distances appeared to be triangular. These results and their implications in the cultivar protection purposes of pepper hybrids are discussed.

Keywords *Capsicum annuum* · Agronomic trait · Molecular marker · Genetic distance index

Introduction

Estimation of genetic distances between varieties of a cultivated crop is of interest in cultivar registration and protection. A newly released cultivar must be distinct from all previously released cultivars to be either listed eligible for protection. Cultivar identification and distinctiveness in pepper is based on morphological and disease resistance criteria. New techniques for DNA profiling are also powerful tools for cultivar identification. Their use for distinctiveness purposes has been discussed by the UPOV (Union pour la Protection des Obtentions Végétales) but not recommended for the moment due to the danger in reducing the minimum distances between registered varieties.

In parallel to the distinctiveness test, seed company, associations such as ASSINSEL, and official authorities charged with the examination of varieties have proposed evaluating quantitatively the genetic relationship between cultivars in order to obtain tools for determining the essential derivation rights between closely related cultivars [The 1991 Act of the UPOV Convention, chapter 5, Article 14(5)]. Phenotypic descriptors do not always allow the quantification of the genotypic difference or similarity between cultivars as do genetic distances based on DNA polymorphism. Soller and Beckmann (1983) proposed using molecular markers as an additional tool for varietal description. DNA markers have the advantage of being independent of environmental effects and providing direct information on the genome of each individual.

Most of the present-day cultivars of pepper are F_1 hybrids of *Capsicum annuum* L. species. In Europe, the Mediterranean areas and North America, these are mainly large-fruited types. The genetic diversity among and within *Capsicum* species has been investigated in several studies using morphological, cytogenetical and molecular markers such as isozymes, 1-dimensional and 2-dimensional PAGE seed proteins, restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) (McLeod et al. 1983; Panda et al. 1986; Pickersgill 1988; Loaiza-Figueroa et al., 1989; Conicella et al. 1990; Livneh et al. 1990; Lefebvre et al. 1993; Posch et al. 1994, Prince et al. 1992, 1995; Rodriguez et al. 1999). Isozyme variability has been found to be very low in *C. annuum* species. Studies with RFLP markers detected a lot of polymorphism between

Communicated by H.C. Becker

V. Lefebvre (✉) · J.-C. Chauvet · B. Caromel · A. Palloix
INRA, Génétique et Amélioration des Fruits et Légumes, BP94, 84143 Montfavet cedex, France
e-mail: veronique.lefebvre@avignon.inra.fr

B. Goffinet
INRA, Biométrie et Intelligence Artificielle, BP27, 31326 Castanet Tolosan cedex, France

P. Signoret · R. Brand
GEVES, BP1, 84300 Les Vignères, France

small-fruited and large-fruited cultivars, but these were ineffective in discriminating large-fruited pepper inbred lines; cultivated large-fruited peppers possess little genetic diversity, particularly the hybrid varieties. Paran et al. (1998) and Ballester and de Vicente (1998) evaluated the relative effectiveness of polymerase chain reaction (PCR)-based markers to detect polymorphism among closely related sweet pepper cultivars. These markers need low quantities of DNA, and the protocols are rapidly carried out and easily automated. These authors concluded that despite their frequently dominant inheritance, PCR-based markers could be efficient in the routine assessment of F_1 hybrid seed purity. Nevertheless, there is no information available regarding the suitability of PCR-based markers to analyse genetic distances among closely related large-fruited pepper inbred lines that are agronomically well-characterised and currently being employed in breeding programmes of pepper.

The aim of the study presented here was to evaluate the potential of phenotypic and PCR-based DNA markers in analysing the relationships among closely related pepper inbred lines used as the parents of European and North American commercial F_1 hybrids. We calculated a genetic distance index (GDI) and studied the relationships between GDI estimated with molecular markers and a genetic distance based on morphological traits. The effectiveness of RAPD and AFLP markers for estimating GDI was compared.

Materials and methods

Plant material

Five private seed companies supplied a total of 36 inbred lines (Table 1), the parents of 19 pepper F_1 hybrids. For reasons of confidentiality, inbred lines were coded. The relationships between them were unknown. These inbred lines belong to three main sweet pepper varietal types: the long type (coded E), the half-long type (coded D) and the blocky type. The blocky types were further subdivided into three subgroups according to their geographic origin: the North American (US) (coded A), the Dutch (NL) (coded B), and the Italian (IT) (coded C) blocky type. Six ancestral inbred varieties, frequently used as progenitors of the present varietal types, were also included in this work. We added 5 small-fruited inbred lines frequently used in breeding programmes as sources of disease resistance traits. INRA's inbred lines (coded 37–47) were selfed in insect-proof cages to eliminate outcrossing. Individuals within inbred lines supplied by private seed companies were assumed to be homozygous.

Phenotypic descriptions

The 47 inbred lines were grown in experimental glasshouses of GEVES (in the south-east of France). They were planted in three complete independent randomised blocks. Each block was composed of a single plot of 6 individuals per inbred line. A total of 41 agronomic traits were measured in each plot (Table 2). These traits are the ones usually recommended by UPOV (Union pour la Protection des Obtentions Variétales) for assessing distintiveness, uniformity and stability criteria, and are used by national authorities. Three traits were assessed qualitatively, 22 by means of a discrete scale and 16 on a continuous quantitative scale. Depending on the trait, measurements were made on 15 fruits sampled in each block, or on the 6 plants of each block, or evaluated on the whole block, as indicated in Table 2.

Table 1 Descriptions of pepper inbred lines

Number	Code	Type
1	A1	US blocky
2	A2	US blocky
3	A3	US blocky
4	A4	US blocky
5	A5	US blocky
6	A6	US blocky
7	A7	US blocky
8	A8	US blocky
9	B9	NL blocky
10	B10	NL/IT blocky
11	B11	NL/US blocky
12	B12	NL blocky
13	B13	NL/IT blocky
14	B14	NL blocky
15	C15	IT/US blocky
16	C16	IT blocky / long
17	C17	IT blocky
18	C18	IT blocky
19	C19	IT blocky/half-long
20	C20	IT blocky
21	C21	IT blocky
22	D22	Half-long
23	D23	Half-long
24	D24	Half-long
25	D25	Half-long
26	D26	Half-long
27	D27	Half-long/blocky
28	E28	Long
29	E29	Long
30	E30	Long
31	E31	Long
32	E32	Long/half-long
33	E33	Long/half-long
34	E34	Long/half-long/US blocky
35	E35	Long
36	M51	Half-long / IT blocky
37	**Yolo Wonder**[a]	US blocky
38	**Florida VR2**[a]	US blocky
39	**Novi**[a]	NL blocky
40	**Vania**[a]	Half-long
41	**Lamu**[a]	Long
42	**Valencia Largo de Rheus**[a]	Long
43	**CM334**[a]	Inbred chilli
44	**Perennial**[a]	Inbred chilli
45	**H3**[a]	Inbred chilli
46	**PM687**[a]	Inbred chilli
47	**SC81**[a]	Inbred chilli

[a] Inbred lines supplied by INRA: 37–42, ancestral inbred varieties; 43–47, small-fruited inbred lines

DNA analyses

Total DNA was extracted from young leaves as described by Caranta et al. (1996) for both RAPD and AFLP analyses. RAPD analyses were performed, as described by Lefebvre et al. (1995), by BIOCEM (Angers, France), a private consulting company. A total of 136 10-mer primers were used (Operon Technologies, Alameda, Calif.). Among them, 46 primers are known to generate 75 RAPD markers well-dispersed throughout the pepper map (Lefebvre et al. 1997).

AFLP reactions were performed, as described by Vos et al. (1995), with *Eco*RI+3 and *Mse*I+3 primer combinations (PC). The *Eco*RI primers end-labelled with [^{33}P]-ATP allowed visualisation of the selectively amplified products on autoradiograms. A total of ten PC was used.

Table 2 List of morphological traits

Number designated to trait	Trait	Type of trait[a]	Number of measurements in each block	Mean	Variance	Coefficients of determination[b]		
						R^2_I	R^2_b	R^2_{Ib}
1	Anthocyanin coloration of the hypocotyl	*1*	1	1.98	0.02	100.0*	0.0	–
2	Adult growth habit	*2*	1	3.68	2.10	88.3*	0.7	–
3	Main stem length (from cotyledons to the first branching)	*3*	6	287.12	3604.88	82.9*	0.7*	5.1*
4	Leave number of the main stem	*3*	6	12.50	6.38	70.2*	0.0	5.3*
5	Anthocyanin coloration on nodes and stems	*2*	1	4.59	0.96	81.7*	1.7*	–
6	Stem pilosity	*2*	1	3.34	1.97	89.8*	1.8*	–
7	Leave habit	*2*	1	4.19	1.93	78.4*	1.8*	–
8	Leaf blade length	*3*	6	185.43	1058.73	63.1*	1.0*	12.2*
9	Leaf blade width	*3*	6	96.23	396.44	66.3*	0.1	10.4*
10	Ratio length/width of the leaf blade	*3*	6	1.95	0.05	70.5*	1.2*	4.6*
11	Leaf blade shape	*2*	1	2.42	0.29	75.2*	0.3*	–
12	Leaf blade size	*2*	1	5.01	1.46	84.3*	0.2	–
13	Leaf blade colour	*2*	1	4.88	3.71	92.0*	1.2*	–
14	Curling leaves	*2*	1	4.97	2.19	80.0*	1.8*	–
15	Attitude of peduncles (on young fruits)	*2*	1	1.66	0.43	87.4*	0.7	–
16	Fruit colour before maturity	*2*	1	2.00	3.33e-29	32.9	1.5	–
17	Fruit colour intensity before maturity	*2*	1	6.33	2.73	53.2*	3.3*	–
18	Fruit length	*3*	15	119.02	1791.10	90.4*	0.1*	1.1*
19	Fruit diameter	*3*	15	91.01	662.65	98.7*	0.2*	1.1*
20	Ratio length/diameter of the fruit	*3*	15	1.48	0.62	92.1*	0.0	0.7*
21	Shape of the fruit longitudinal section	*1*	1	1.11	0.10	90.1*	0.3	–
22	Tendency at the trapezoidal shape of the fruit	*2*	1	4.24	7.25	88.9*	0.4	–
23	Shape of the fruit transversal section	*2*	15	2.16	0.24	26.9*	0.1	4.3*
24	Sinuosity of the pericarp	*2*	1	3.27	6.00	92.3*	0.6*	–
25	Surface texture of the fruit	*2*	1	1.13	0.11	74.5*	1.1	–
26	Fruit colour at maturity	*2*	1	2.36	0.88	100.0*	0.0	–
27	Fruit colour intensity at maturity	*2*	1	5.14	1.11	68.3*	1.9	–
28	Fruit brightness	*2*	1	4.95	1.14	71.5*	2.0*	–
29	Depth of the stalk cavity	*2*	15	2.81	1.03	44.1*	0.2*	4.8*
30	Shape of the fruit apex	*2*	1	6.17	1.51	87.8*	0.3	–
31	Depth of the fruit interlocularary grooves	*2*	1	5.39	4.36	82.1*	0.3	–
32	Cavity number per fruit	*3*	15	3.42	0.53	36.5*	0.0	3.1
33	Flesh thickness	*3*	15	6.98	3.94	84.0*	0.2*	1.4*
34	Peduncle length	*3*	15	47.22	175.84	75.3*	0.0	2.5*
35	Peduncle thickness	*3*	15	11.84	9.61	78.5*	0.2*	2.7*
36	Calix aspect	*1*	1	1.06	0.06	100.0*	0.0	–
37	Capsaicin in the placenta	*2*	1	1.14	0.21	97.7*	0.0	–
38	Number of days between sowing and flowering	*3*	6	69.01	27.04	67.1*	0.0	7.1*
39	Fruit maturity in number of days relative to the earlier flowering plant	*3*	6	26.21	144.69	82.1*	0.4*	5.0*
40	Fruit mean weight	*3*	15	255.70	11746.84	86.0*	0.6*	2.1*
41	Plant height	*3*	1	111.52	836.76	92.8*	0.3	–

* Significative at $P<0.05$

[a] 1, Qualitative trait; 2, semi-quantitative trait; 3, quantitative trait

[b] Phenotypic variance (in percentage) among inbred lines accounted for by: R^2_I, the inbred effect; R^2_b, the block effect; R^2_{Ib}, the inbred line×block interaction effect. –, Not calculated

For the RAPD and AFLP analyses, only clear and unambiguous bands were visually scored as either present (1) or absent (0) for all inbred lines, resulting in two binary data matrices. Each band was interpreted as one allele. Bands with the same mobility were assumed to be homologous (Rouppe van der Voort et al. 1997). Each marker was treated as an independent unit character.

Data analyses

Analyses of variance of phenotypic data were performed with inbred line, block and inbred line×block interaction effects. Fractions of the variation due to inbred line, to block and to inbred line×block interaction effects were estimated by coefficients of determination, R^2_I, R^2_b and R^2_{Ib}, (except for traits with only 1 measurement per block), respectively.

A phenotypic Mahalanobis distance matrix (Mahalanobis 1936), based on 36 traits, was computed for each pair of inbred lines. For calculating the Mahalanobis distance, we discarded 5 traits, denoted nos.1, 16, 26, 31, and 36 in Table 2, among the 41 initial ones, because the residual variance-covariance matrix was not definitely-positive when using these traits. The mean effect μ_i of each inbred line i was estimated by $\hat{\mu}_i$ as the additive inbred line effect in the model inbred line+block, and the residual variance-covariance matrix Σ_E was estimated in the model with inbred line×block interaction. The unbiased Mahalanobis distance d^2 (i,i') between inbred lines i and i' was computed as:

$$d^2(i,i') = (\hat{\mu}_i - \hat{\mu}_{i'})' \sum_E^{-1} (\hat{\mu}_i - \hat{\mu}_{i'}) - p \times (n_i + n_{i'}) / (n_i \times n_{i'})$$

where p=36 is the number of traits and n_i (respectively $n_{i'}$) is the number of degrees of freedom in the estimation $\hat{\mu}_i$ of μ_i (respectively $\mu_{i'}$). The reason for the correcting term $p\times(n_i+n_{i'})/(n_i\times n_{i'})$ is that $(n_i\times n_{i'})/(n_i+n_{i'})\times d^2(i,i')$ follows a χ^2 distribution with p degrees of freedom and a non-centrality parameter of $(n_i\times n_{i'})/(n_i+n_{i'})\times D^2(i,i')$. $D^2(i,i')=(\mu_i-\mu_{i'})\Sigma_E^{-1}(\mu_i-\mu_{i'})$ is the distance computed with the true value of the means μ_i and $\mu_{i'}$. These expressions are correct, as the number of repetitions used in the estimation of Σ_E is sufficiently great to consider that it is the true matrix. The variance of $d^2(i,i')$ is then computed as $2\times p\times(n_i+n_{i'})^2/(n_i\times n_{i'})^2$. In this experiment, the value n_i depends on the trait as shown in Table 2. Neglecting the missing values, there are three categories of traits: p_1=19 traits with n_i=n1=3 values (1 value per block×3 blocks), p_2=7 traits with n_i=n2=18 values (6 plants×3 blocks), and p_3=10 traits with n_i=n3=45 values (15 fruits×3 blocks). We considered these three categories of traits as independent from one another. The bias is therefore computed as $2\times p1/n1+2\times p2/n2+2\times p3/n3=13.89$, and the variance as $4\times p1/(n1)^2+4\times p2/(n2)^2+4\times p3/(n3)^2=8.55$.

As each marker band was interpreted as one allele, the Nei's genetic distance $Nei(i,i')$ between the inbred lines i and i' is: $Nei(i,i')=-\log(A/B_{(i)}\times B_{(i')})=-\log(1-p(i,i'))$ with $A=\sum_{k=1}^{n}\sum_{j=1,2} p_i(k,j)\times p_{i'}(k,j)$ and $B_{(i)}=\sum_{k=1}^{n}\sum_{j=1,2} p_{(i)}(k,j)\times p_i(k,j)$ (respectively $B_{(i')}$). $p_i(k,j)$ is the frequency of the allele j at the locus k in the individual i. $p(i,i')$ is the percentage of bands differing between the two inbred lines i and i' (Nei 1973). When $p(i,i')$ is small, $Nei(i,i')$ can be assimilated as $p(i,i')$. Supposing that all the bands are sampled at random over the genome, an approximation of the sampling variance of $Nei(i,i')$ is therefore $Nei(i,i')\times(1-Nei(i,i'))/N$, where N is the number of bands considered in the analysis. The Nei distances were constructed using the GENDIST procedure of the PHYLIP phylogenetic analysis software package (Felsenstein 1993). A dendrogram was constructed using the neighbour-joining method of Saitou and Nei (1987) contained in the NEIGHBOR procedure of PHYLIP. The analysis was also made on 100 bootstrap re-sampling of the data.

Binary molecular data matrices were also employed to perform principal component analyses using the computer software SPLUS (1993).

Marker frequencies were calculated for each molecular marker among the five varietal types. Diagnostic markers were defined as markers present in all inbred lines of one varietal type but absent from all inbred lines of another varietal type.

Results

Agronomic traits

Block and/or inbred line×block interaction effects were detected on 25 of the measured agronomic traits (P<0.05). The inbred line effect was always important, with environmental effects being of a very small magnitude when compared to the inbred line effect (Table 2). Only the trait fruit colour before maturity (no. 16) had no significant inbred line effect. Further analyses were performed based on the residuals of the model with only the block effect, i.e. on the phenotypic effect of the inbred line. We considered the mean values for all 47 inbred lines.

Table 3 indicates the number of variables for which there is a significant difference between two varietal types (P<0.05). As expected, the blocky types (A, B and C) are very close but distinct. The US and the NL blocky types are significatively different for variable no. 38 (number of days between sowing and flowering), the US and the IT blocky types for variable no. 25 (surface texture of the fruit) and the NL and the IT blocky types for variable no. 23 (shape of the fruit-transversal section). They are distant from the half-long (D) and the long (E) types. The half-long (D) types are close to the long (E) types. They differ for the fruit length (no. 18), depth of the stalk cavity (no. 29) and shape of the fruit apex (no. 30).

Table 3 Number of variables (among the 41) for which two varietal types[a] are significatively different (P<0.05)

	B	C	D	E
A	1	1	9	14
B		1	10	16
C			8	15
D				3

[a] A, US blocky types; B, Dutch blocky types; C, Italian blocky types; D, half-long types; E, long types

Randomly amplified polymorphic DNA

Analysis of 136 RAPD primers yielded 1,204 amplification products (an average of 8–9 bands per primer) of which 544 (corresponding to 130 primers) were polymorphic (45.2%) among the 47 inbred lines. An average of 6 patterns was observed with each primer among the 47 inbred lines, with a maximum of 19 different patterns revealed by Operon P11 primer. The 136 primers produced a total of 806 distinct patterns on the 47 inbred lines. 142 patterns were unique, that is to say, they were noticed in only one inbred line among the 36 parental inbred lines of the hybrids. Consequently, 32 of the 36 parental inbred lines had at least 1 specific pattern. The inbred line D22 had 17 specific patterns.

Amplified fragment length polymorphism

Analysis of 10 AFLP PC yielded 863 selectively amplified DNA fragments of which 378 were polymorphic markers (43.8%) among the 47 inbred lines. Between 62 and 122 distinguishable bands were observed after selective amplification with each PC; 24–51 of these AFLP bands were polymorphic among the 47 inbred lines. Neglecting the markers for which 1 of the 47 inbred lines had missing data, the 10 PC generated 327 distinct patterns from the 47 inbred lines (calculated on 767 bands). Between 28 and 45 patterns (45 with the E42-M55 PC) were revealed per PC from the 47 inbred lines, with a mean of 36 patterns per PC. 202, patterns were unique that is to say, then were noticed in only one inbred line among the 36 parental inbred lines of the hybrids. Consequently, 36 parental inbred lines had at least 2 specific patterns. Inbred lines C17, D25 and E33 had 9 specific patterns; inbred line D22 had 8.

Table 4 Range values of genetic distances between pepper inbred lines

	Number of inbred lines	Phenotypic traits	RAPDs	AFLPs
Whole set of pepper inbred lines	47	480.30[a]	0.096	0.089
	(nos.1–47)	3.45[b]	0.006	0.009
		2906.00[c]	0.245	0.227
		8.55[d]	0.96 10^{-4}	1.02 10^{-4}
. Ancestral and modern large-fruited inbred lines	42	202.00	0.075	0.068
	(nos. 1–42)	3.45	0.006	0.009
		688.70	0.123	0.131
		8.55	0.78 10^{-4}	0.82 10^{-4}
.. Modern breeding inbred lines	36	205.10	0.075	0.065
	(nos. 1–36)	12.60	0.006	0.009
		600.70	0.123	0.131
		8.55	0.78 10^{-4}	0.79 10^{-4}
... Blocky type (NL+IT+US) (A+B+C)	21	132.00	0.052	0.044
	(nos. 1–21)	14.20	0.007	0.011
		491.70	0.101	0.087
		8.55	0.56 10^{-4}	0.54 10^{-4}
US blocky type (A)	8	144.00	0.038	0.032
	(nos. 1–8)	23.10	0.007	0.016
		484.90	0.066	0.047
		8.55	0.41 10^{-4}	0.40 10^{-4}
NL blocky type (B)	6	124.10	0.046	0.033
	(nos. 9–14)	35.70	0.020	0.014
		284.80	0.072	0.051
		8.55	0.50 10^{-4}	0.41 10^{-4}
IT blocky type[e] (C)	7	131.90	0.069	0.064
	(nos. 15–21)	52.40	0.040	0.037
		255.30	0.094	0.087
		8.55	0.72 10^{-4}	0.78 10^{-4}
... Half-long type[e] (D)	6	197.90	0.078	0.077
	(nos. 22–27)	35.20	0.023	0.034
		331.60	0.116	0.119
		8.55	0.81 10^{-4}	0.92 10^{-4}
... Long type (E)	8	153.90	0.056	0.061
	(nos. 28–35)	16.60	0.006	0.009
		435.70	0.104	0.093
		8.55	0.60 10^{-4}	0.74 10^{-4}
.. Ancestral inbred varieties	6	196.90	0.080	0.082
	(nos. 37–42)	66.60	0.023	0.022
		302.30	0.115	0.114
		8.55	0.83 10^{-4}	0.97 10^{-4}
. Small-fruited inbred lines	5	1208.90	0.182	0.166
	(nos. 43–47)	477.30	0.153	0.129
		1781.20	0.212	0.203
		8.55	1.69 10^{-4}	1.79 10^{-4}

[a] Mean genetic distance
[b] Minimum (the most related inbred lines)
[c] Maximum (the most distant inbred lines)
[d] Mean sampling variance
[e] Breeders classified the inbred line M51 (no. 36) as a half-long or IT blocky inbred line

Distance matrices, dendrograms and principal component analysis

Genetic distance indices (GDIs) and corresponding sampling variances for each of the 47*46/2=1081 pairs of inbred lines were calculated from binary molecular data matrices and from phenotypic data (Table 4). As expected, whatever the mean used to measure the GDI, the small-fruited pepper inbred lines had a larger mean GDI than the large-fruited ones. Modern large-fruited inbred

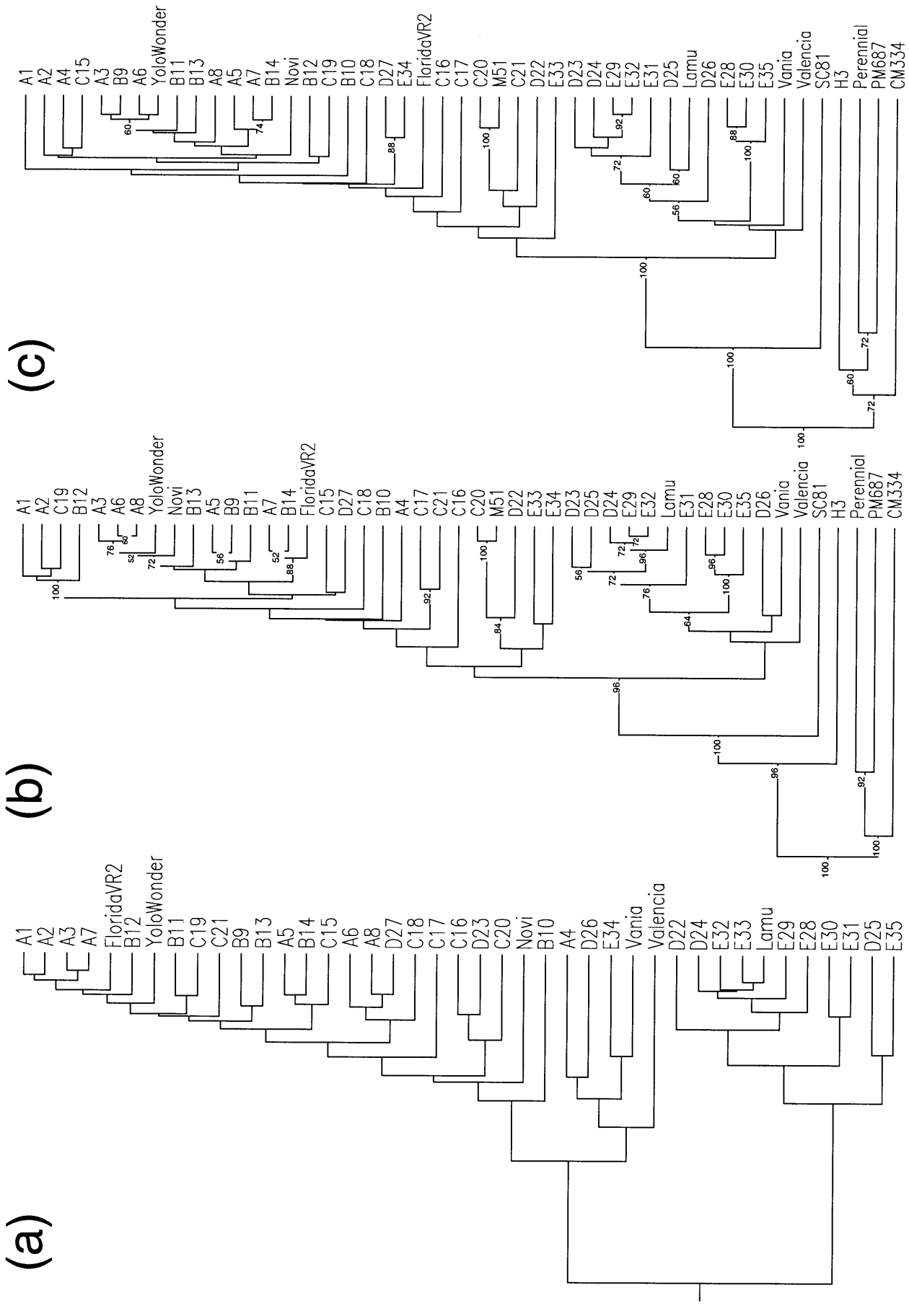
(a)
A1
A2
A3
A7
FloridaVR2
B12
YoloWonder
B11
C19
C21
B9
B13
A5
B14
C15
A6
A8
D27
C18
C17
C16
D23
C20
Novi
B10
A4
D26
E34
Vania
Valencia
D22
D24
E32
E33
Lamu
E29
E28
E30
E31
D25
E35
(b)
A1
A2
C19
B12
A3
A6
A8
YoloWonder
Novi
B13
A5
B9
B11
A7
B14
FloridaVR2
C15
D27
C18
B10
A4
C17
C21
C16
C20
M51
D22
E33
E34
D23
D25
D24
E29
E32
Lamu
E31
E28
E30
E35
D26
Vania
Valencia
SC81
H3
Perennial
PM687
CM334
(c)
A1
A2
A4
C15
A3
B9
A6
YoloWonder
B11
B13
A8
A5
A7
B14
Novi
B12
C19
B10
C18
D27
E34
FloridaVR2
C16
C17
C20
M51
C21
D22
E33
D23
D24
E29
E32
E31
D25
Lamu
D26
E28
E30
E35
Vania
Valencia
SC81
H3
Perennial
PM687
CM334

Table 5 Diagnostic markers and their frequencies among the varietal types

Marker type	Locus	Varietal type			Inside the blocky type		
		Blocky	Half-long	Long	US	NL	IT
RAPD	AF14_3.00	***0.00***	0.40	**1.00**	***0.00***	***0.00***	***0.00***
AFLP	E38-M48_10	0.14	**1.00**	**1.00**	***0.00***	***0.00***	0.43
RAPD	D11_0.95	0.19	0.83	**1.00**	***0.00***	***0.00***	0.57
RAPD	AH09_1.05	0.82	***0.00***	***0.00***	**1.00**	**1.00**	0.57
RAPD	D11_0.85	0.81	0.17	***0.00***	**1.00**	**1.00**	0.43

lines had a larger mean phenotypic distance than the ancestral ones, whereas the mean molecular distances of these two groups were similar. Based on phenotypic, RAPD and AFLP analyses, the half-long types were more divergent than the blocky and long ones. But the classification inside the blocky type differed between the phenotypic and molecular analyses. The US blocky types were the more divergent with respect to phenotypic data, whereas the IT blocky types had the largest GDI with respect to RAPD or AFLP data.

RAPD as well as AFLP data did not reveal any duplicate inbred lines. The pair of inbred lines E29 and E32, the most related inbred lines, differed for 7 RAPD primers and for 6 AFLP primer combinations. The RAPD data appeared more discriminating than the AFLP data, if we considered only the band information instead of the primer or PC information. RAPD distances were generally larger than AFLP distances, probably due to the higher quantity of data obtained with the 136 RAPD primers than with the 10 AFLP PC.

The ratio between the mean sampling standard deviation for the whole set of inbred lines and the mean genetic distance is $\sqrt{8.55}/480.3=0.01$ for the phenotypic traits, $\sqrt{0.96\cdot10^{-4}}/0.096=0.10$ for the RAPD and $\sqrt{1.02\cdot10^{-4}}/0.089=0.11$ for the AFLP. Estimation of the genetic distance is therefore more precise using phenotypes than using molecular markers. Nevertheless, it is not the same genetic distance.

UPGMA clustering analysis was carried out in order to represent graphically the genetic distances among pepper inbred lines (Fig. 1). The resulting dendrograms showed, as expected, that small-fruited pepper inbred lines are very divergent from ancestral and modern large-fruited inbred lines. All the dendrograms split pepper large-fruited inbred lines into two main groups. The first group included all the blocky types (A, B and C) together with Yolo Wonder, Florida VR2 and Novi, the three blocky representatives. The second one comprised long types (E) together with Lamu, the long-type representative. The half-long types (D) split up into the two groups. The 41 phenotypic traits did not enable the three blocky types to be classified, whereas AFLP markers allowed us to distinguish the group of IT blocky types from the NL and US blocky types. NL and US blocky types are difficult to distinguish from each other whatever the molecular or phenotypic analyses used. To display the relationships among the pepper inbred lines in terms of their position relative to coordinate axes, we performed principal component analyses (data not shown). In agreement with the results of the dendrograms, the half-long types were located in an intermediate position between the blocky types and the long types. The NL blocky types were close to the US blocky type cluster. The US and NL blocky types were less dispersed than the IT ones, indicating a larger genetic variability among the IT blocky types than among the US and NL ones.

One AFLP and 4 RAPD markers were identified as diagnostic markers (Table 5). They generally allowed two varietal types to be distinguished, but no diagnostic marker clearly identified only one varietal type; that is to say, the marker would be systematically present in one varietal type but totally absent in all other varietal types. The RAPD marker AF14_3.00 is totally absent in all the blocky types and totally present in all the long-type inbred lines, while the half-long types had an intermediate frequency probably because of their pedigree. This marker may be in linkage disequilibrium with a gene influencing the length of the fruit. Whereas the characterisation of the long types seemed easy with the extreme frequency diagnostic markers, the characterisation of the blocky types was hampered by the fact that they included three types, particularly the IT ones which often exhibited intermediate frequencies. The long types and the NL and US blocky types had opposite patterns for the 5 diagnostic markers. No reliable diagnostic marker was identified between NL blocky types and US ones. Likewise, no reliable marker clearly allowed identification between the half-long types and the IT blocky types.

◀ **Fig. 1** Associations among pepper inbred lines revealed by UPGMA clustering analysis on the basis of phenotypic (**a**), RAPD (**b**) and AFLP (**c**) genetic distance values. Small-fruited pepper inbred lines were discarded from the phenotypic analysis (**a**) to observe more precisely the phenotypic variation between large-fruited pepper inbred lines. *Numbers* at the *forks* indicate the number of times the group consisting of the varieties which are to the *right* of that *fork* occurred among the trees, out of 100 trees. Only numbers above 50 are indicated. Note that C15, initially declared to be an IT type by breeder partner, was phenotypically categorised later as belonging to a US type when breeder partners visited the assay

Correlation between distance matrices

RAPD and AFLP GDIs were highly correlated (r=0.95, P<0.001; Fig. 2a). The phenotypic distance matrix was

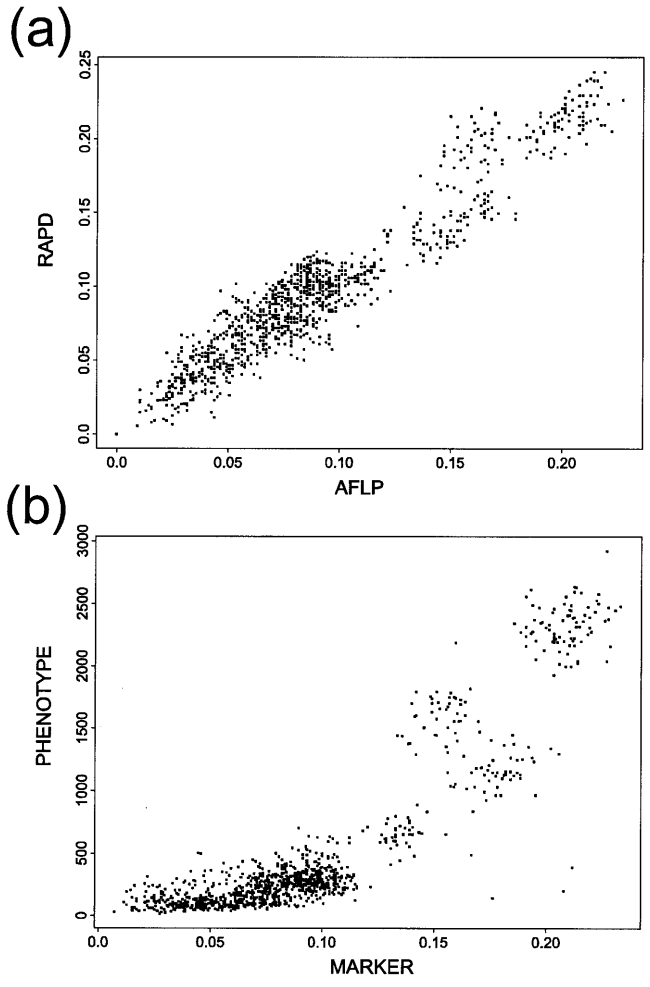

Fig. 2 Regression analysis to compare the genetic pairwise distances obtained from RAPD and AFLP (**a**) and from all the molecular markers and phenotypic traits (**b**)

compared to a molecular genetic distance matrix based on all molecular markers (Fig. 2b). A regression analysis indicated a high correlation between molecular and phenotypic data in their ability to detect genetic relationships between pepper inbred lines (r=0.62, P<0.001). The relationship between molecular and phenotypic distances is clearly not linear but displays a "triangular" shape, as Dillmann et al. (1997) and Burstin and Charcosset (1997) already demonstrated. Low-molecular-marker distances were systematically associated with low phenotypic distances. On the contrary, high-molecular-marker distances were associated either with low or high phenotypic distances. Thus, the molecular marker GDI was not exactly representative of the phenotypic GDI. However, molecular markers did frequently allow not phenotypically distinct cultivars to be distinguished.

Discussion and conclusion

Identification, description

Assessment of genetic variability between cultivars is of interest not only for cultivar protection but also for practical applications such as conservation of genetic resources and broadening the genetic basis of the varieties. For genetic resource conservation, it is useful to know whether 2 individuals phenotypically similar display a similar gene combination. For breeding purposes, genetic distances between 2 inbred lines help to predict their ability to combine (Charcosset et al. 1998).

The analysis presented here assumed that the inbred lines are perfectly homozygous. Replicates in a double-blind protocol would allow this hypothesis to be evaluated, as performed by Dillmann et al. (1997). Discrepancy between replications may come either from experimental errors or from residual heterozygosity of inbred lines. Hence, the residual errors as well as the sampling of the markers throughout the genome and the number of markers were sources of variation for the estimation of $Nei(i,i')$ from the molecular data. One can consider that the standard deviation of experimental errors on the estimated distance between 2 inbred lines is smaller that the smallest molecular distance (0.006 with RAPD markers, 0.009 with AFLP).

Molecular GDIs were computed in this study on the basis of band information instead of locus information. A GDI calculated on band information may introduce a bias since all the bands are not completely independent. On the one hand, molecular GDI may be increased because of a redundancy of band information in the case of allele bands of the same locus (Lefebvre et al. 1993) or genetic linkage between molecular markers. The clustering of RAPD and AFLP markers is well-documented in pepper (Lefebvre et al. 1995; unpublished data of our laboratory, INRA-GAFL). On the other hand, it may be lowered because of different alleles sharing the same band mobility. Our molecular dataset was supposed to reliably represent the pepper genome; markers were numerous and some of them are known to be well-distributed throughout the pepper map. Nevertheless, we are mapping new markers used in this study and we are developing a formula of GDI which takes the position of the markers on the genetic map into consideration.

Variety discrimination

With respect two distinctiveness studies, the objective is to perform routine studies on large datasets, with readily interpretable molecular data. Repeatable standard protocols and undisputed results are required. PCR-based markers, particularly AFLP ones, are well-adapted for this objective. The number of markers will depend on the precision to be reached for the estimation of genetic distance since the precision of the estimate increases with the square root of the number of markers. Above 100

markers, the increase is low (Dillmann et al. 1997). Likewise, as illustrated by the formula, the sampling variance of *Nei*(*i*,*i*′) decreases as the number of markers increases. The use of fingerprinting markers, like AFLP ones, is effective in displaying numerous loci in a short time and, consequently, in decreasing the sampling variance.

Relationship between molecular marker genetic distance and distance based on morphological data

The triangular relationship between molecular and phenotypic distances is linked to the polygenic inheritance of traits used to compute phenotypic distance. Burstin and Charcosset (1997) demonstrated (1) that the magnitude of the correlation coefficient between phenotypic and molecular distances depends on the association between marker loci and quantitative trait loci (QTLs); (2) that the correlation necessarily decreases as the number of loci involved in the variation of quantitative trait increases. Poor associations between marker loci and QTLs lead to a low correlation between distances.

The relationship between molecular distances and phenotypic distances shows that inbred lines with different phenotypes also differ with respect to markers. Thus, the risk of discriminating between a candidate inbred line close to a protected inbred line based on the phenotypic distinction study, but with a very low genome divergence, is low. On the contrary, we cannot only use molecular markers for such studies because a candidate variety, by definition, must differ from protected ones on the basis of phenotypic traits. However, molecular markers provide complementary information in cases of phenotypic similarity or proximity. Indeed, different gene combinations originating from unrelated genetic resources may govern the same phenotype, whereas identical marker genotypes necessarily display the same phenotype. Bar-Hen et al. (1995) proposed first to compare inbred lines at marker loci and declare them distinct if their distance is above a predetermined threshold and second, to compare the selected pair of inbred lines for phenotypic traits. However, there is a very low probability of distinguishing two near-isogenic lines in this way; so a new inbred line with a very advantageous trait may escape.

Grouping of varieties

Despite the low level of molecular polymorphism in *C. annuum* genome (Lefebvre et al. 1993; Prince et al. 1995; Paran et al. 1998; Rodriguez et al. 1999), PCR-based DNA molecular markers have proved to be informative enough to allow for representative varietal types of cultivated pepper to be distinguished. The classification obtained with phenotypic traits and molecular markers is consistent with the commercial classification of varieties. Molecular markers provide additional information.

Genetic distance and essential derivation

Molecular genetic distances are useful for the question of intellectual property protection, in particular for essential derivation rights. Indeed, the molecular marker GDI estimated from experimental data may be compared to a minimum genetic distance above which 2 varieties are declared independent (or not essentially derived). The probability of declaring the molecular GDI above or below the minimum distance threshold depends on the sampling variance of the estimate, on the variance of experimental error and on the repartition of the markers on the genetic map of the species (Lombard et al. 2000). The use of evenly spaced markers with a known position on a genetic map allows their genetic determinism to be considered, which is more pertinent to calculate a molecular marker GDI because it avoids redundancy. Therefore, we advise preferentially using codominant markers selected on the basis of their discriminating power in a reference collection and computing the GDI using locus information instead of band information.

Acknowledgments The RAPD analyses were committed to the care of P. Gadille (BIOCEM). We thank Claire Baril for her helpful comments on the manuscript, and Lynda S. Hagen for editing the grammar of the manuscript. This work was supported by a grant from CTPS (Comité Technique Permanent de la Sélection des Plantes Cultivées) of the French Ministry of Agriculture and conducted in collaboration between INRA, GEVES and five private seed companies (ASL, Clause Semences, Gautier, Tézier, Seminis Vegetable Seeds). These experiments comply with the current laws of France.

References

Ballester J, de Vicente MC (1998). Determination of F_1 hybrid seed purity in pepper using PCR-based markers. Euphytica 103:223–226

Bar-Hen A, Charcosset A, Bourgoin M, Guiard J (1995) Relationship between genetic markers and morphological traits in maize inbred lines collection. Euphytica 84:145–154

Burstin J, Charcosset A (1997) Relationship between phenotypic and marker distances: theoretical and experimental investigations. Heredity 79:477–483

Caranta C, Palloix A, Gebre-Selassie K, Lefebvre V, Moury B, Daubèze AM (1996) A complementation of two genes coming from susceptible *Capsicum annuum* L. lines confers a new and complete resistance to pepper veinal mottle virus. Phytopathology 86:739–743

Charcosset A, Bonnisseau B, Touchebeuf O, Burstin J, Dubreuil P, Barrière Y, Gallais A, Denis JB (1998) Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. Crop Sci 38:38–44

Conicella C, Errico A, Saccardo F (1990) Cytogenetic and isozyme studies of wild and cultivated *Capsicum annuum*. Genome 33:279–282

Dillmann C, Bar-Hen A, Guérin D, Charcosset A, Murigneux A (1997) Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. Theor Appl Genet 95:92–102

Felsenstein J (1993) PHYLIP (phylogeny inference package). Department of Genetics, University of Washington, Seattle, Wash

Lefebvre V, Palloix A, Rives M (1993) Nuclear RFLP between pepper cultivars. Euphytica 71:189–199

Lefebvre V, Palloix A, Caranta C, Pochard E (1995) Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled haploid progenies. Genome 38:112–121

Lefebvre V, Caranta C, Pflieger S, Moury B, Daubèze AM, Blattes A, Ferrière C, Phaly T, Nemouchi G, Ruffinatto A, Palloix A (1997) Updated intraspecific maps of pepper. *Capsicum* Eggplant Newsl 16:35–41

Livneh O, Nagler Y, Tal Y, Harush SB, Gafni Y, Beckmann JS, Sela I (1990) RFLP analysis of a hybrid cultivar of pepper (*Capsicum annuum*) and its use in distinguishing between parental lines and in hybrid identification. Seed Sci Technol 18:209–214

Loaiza-Figueroa F, Ritland K, Laborde Cancino JA, Tanksley SD (1989) Patterns of genetic variation of the genus *Capsicum* (*Solanaceae*) in Mexico. Plant Syst Evol 165:159–188

Lombard V, Baril CP, Zhang D (2000) Improvement of the precision of genetic distance estimates between rapeseed cultivars based on AFLPs by using information from their position on a consensus linkage map. In: Heller SR (ed) Int Conf Status Plant Anim Genome Res, VIII (abstracts). San Diego, Calif., USA p 186

Mahalanobis PC (1936) On the generalized distances in statistics. Proc Natl Inst Sci India 2:49–55

McLeod MJ, Guttman SI, Eshbaugh WH, Rayle RE (1983) An electrophoretic study of evolution in *Capsicum* (*Solanaceae*). Evolution 37:562–574

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Panda RC, Aniel Kumar O, Raja Rao KG (1986) The use of seed protein electrophoresis in the study of phylogenetic relationships in Chili pepper (*Capsicum* L.). Theor Appl Genet 72:665–670

Paran I, Aftergoot E, Shifriss C (1998) Variation in *Capsicum annuum* revealed by RAPD and AFLP markers. Euphytica 99:167–173

Pickersgill B (1988) The genus *Capsicum*: a multidisciplinary approach to the taxonomy of cultivated and wild plants. Biol Zent Bl 107:381–389

Posch A, van den Berg BM, Duranton C, Görg A (1994) Polymorphism of pepper (*Capsicum annuum* L.) seed proteins studied by two-dimensional electrophoresis with immobilized pH gradients: methodical and genetic aspects. Electrophoresis 15:297–304

Prince JP, Loaiza-Figueroa F, Tanksley SD (1992) Restriction fragment length polymorphism and genetic distance among Mexican accessions of *Capsicum*. Genome 35:726–732

Prince JP, Lackney VK, Angeles C, Blauth JR, Kyle MM (1995) A survey of DNA polymorphism within the genus *Capsicum* and the fingerprinting of pepper cultivars. Genome 38:224–231

Rodriguez JM, Berke T, Engle L, Nienhuis J (1999) Variation among and within *Capsicum* species revealed by RAPD markers. Theor Appl Genet 99:147–156

Rouppe van der Voort JNAM, van Zandvoort P, van Eck HJ, Folkertsma RT, Hutten RCB, Draaistra J, Gommers FJ, Jacobsen E, Helder J, Bakker J (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255:438–477

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67:27–33

SPLUS (1993) Statistical Sciences, S-plus programmer's manual, version 3.2. StatSci, a division of Mathsoft, Seattle

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414