Theor Appl Genet (1999) 99: 147–156

J. M. Rodriguez · T. Berke · L. Engle · J. Nienhuis

Variation among and within *Capsicum* species revealed by RAPD markers

Received: 1 September 1998 / Accepted: 28 December 1998

Abstract Germplasm characterization is an important link between the conservation and utilization of plant genetic resources. A total of 134 accessions from six *Capsicum* species maintained at the Asian Vegetable Research and Development Center were characterized using 110 randomly amplified polymorphic DNA (RAPD) markers. Ten pairs of potentially duplicated accessions were identified. Multidimensional scaling analysis of the genetic distances among accessions resulted in clustering corresponding to a previous species assignment except for six accessions. Diagnostic RAPDs were identified which discriminate among the *Capsicum* species. The diagnostic markers were employed for improved taxonomic identification of accessions since many morphological traits used in the identification of *Capsicum* are difficult to score. Three *Capsicum* accessions, misclassified based on morphological traits, were reassigned species status based on diagnostic RAPDs. Three accessions, not previously classified, were assigned to a species based on diagnostic RAPDs. Definitive conclusions about the species assignment of three other accessions were not possible. The level of diversity between *Capsicum annuum* accessions from the genebank and the breeding program were compared and no differences were observed either for RAPD variation or diversity. The utilization of genetic resources as a source of variance for useful traits in the breeding program may be the reason for the similarity of these two groups.

Communicated by G. E. Hart

J. M. Rodriguez (✉) · J. Nienhuis
Department of Horticulture, University of Wisconsin,
1575 Linden Drive, Madison, WI 53706, USA

T. Berke · L. Engle
Asian Vegetable Research and Development Center,
Shanhua, Tainan, Taiwan

Key words *Capsicum* · Diagnostic markers · Genetic diversity · Germplasm · RAPD

Introduction

Germplasm characterization is an important link between the conservation and utilization of plant genetic resources. Molecular DNA techniques allow researchers to identify accessions at the taxonomic level, assess the relative diversity within and among species, and locate diverse accessions for breeding purposes. Moreover, the commercial value associated with identifying useful traits places a direct value on genebanks ensuring the long-term preservation of a collection.

The Asian Vegetable Research and Development Center's (AVRDC, Shanhua, Tainan, Taiwan) genebank holds the largest collection of *Capsicum* germplasm in the world (Berke and Engle 1997). As of May 1998, the AVRDC collection contained 6901 accessions from most geographic regions of the world, and is considered representative of much of the variation held *ex situ* for the genus *Capsicum*. The AVRDC also supports a pepper breeding program which develops improved hot and sweet pepper cultivars adapted to tropical and subtropical regions. Genetic resources are utilized as a source of genetic variation for numerous desirable traits needed for breeding new cultivars adapted to tropical environments. The large magnitude of the AVRDC collection maximizes the probability of preserving genetic variability. However, the time and resources required for its characterization, plus the presence of duplicated accessions, simultaneously limit the ability of plant breeders to sample accessions with useful traits. Molecular characterization of the level of genetic diversity between accessions from a breeding program and a genebank may help identify potential new sources of genetic diversity useful in the breeding of improved pepper cultivars.

Flower morphology is among the taxonomic descriptors presently used to characterize *Capsicum* species. Three of the species, *Capsicum annuum*, *Capsicum chinense* and *Capscium frutescens*, form an overlapping complex with a common ancestral gene pool (Pickersgill et al. 1979; Pickersgill 1988). Species derived from this gene pool are based largely on flower color, calyx constriction, and the number of flowers per node. Nevertheless, unambiguous species designation among *C. annuum*, *C. frutescens*, and *C. chinense* using morphological descriptors is difficult since many exceptions to general taxonomic identification exist (Pickersgill et al. 1979; T. Berke, unpublished data). *Capsicum baccatum* is differentiated from the above species by having unique yellow spots on its corolla; *Capsicum pubescens* is differentiated by a purple corolla and black seeds; and *Capsicum chacoense* is differentiated by glabrous plants and long projections on the calyx (IBPGR 1983, key prepared by Paul G. Smith, unpublished). Characterization of accessions within the AVRDC genebank routinely includes these morphological traits in order to provide a species assignment for an accession. However, errors in scoring the small number of traits used for species assignment, or accessions which contain characteristics of two or more species, can confound accurate species assignment (Villand et al. 1996).

Polymorphisms detected with RFLPs, AFLPs, and RAPDs have been reported among and within the domesticated *Capsicum* species (Lefebvre et al. 1995; Paran et al. 1998; Prince et al. 1992, 1993, 1995; Tanksley et al. 1988). RFLPs and RAPDs have supported the morphological classification of *Capsicum* species (Prince et al. 1992, 1995) and can therefore assist in the identification of the correct species assignment of an accession. Isozyme analysis of genetic diversity among Mexican accessions revealed that greater genetic variation is observed among populations, rather than within populations, presumably due to self-pollination and group bottlenecks (Loaiza-Figueroa et al. 1989). Since genetic differentiation was correlated with geographic isolation in Mexico (Loaiza-Figueroa et al. 1989), it may be appropriate to analyze accessions that represent a wide range of geographic origins in order to maximize genetic diversity. Unfortunately, movement of accessions from one environment to another renders the historical derivation of genebank accessions unclear. Therefore, diverse geographic origins among accessions may not be a reliable indicator for the sampling of genetically diverse materials (Skroch et al. 1998); instead, phenotypic or genotypic measures of genetic relationships may be more reliable. In the present study, the diversity of phenotypic traits was used to select accessions which represent the range of genetic diversity at the AVRDC.

The objectives of this study were to: (1) identify duplicated accessions, if any, in the AVRDC genebank; (2) describe the RAPD-based diversity present among and within six *Capsicum* species; (3) identify diagnostic RAPDs which can be used for the taxonomic identification of accessions; and (4) compare the level of genetic diversity between accessions maintained in the AVRDC genebank and accessions used in the AVRDC pepper breeding program.

Materials and methods

Germplasm

A total of 76 accessions from the AVRDC genebank and 58 accessions from the pepper-breeding program were included in this study (Table 1). The selection of accessions from the genebank was based on a previous phenotypic cluster analysis of 1340 *C. annuum* accessions used to help construct the AVRDC core collection (AVRDC 1997). The unweighted pair group method using arithmetic averages (UPGMA) resulted in 14 clusters ranging from 1 to 904 accessions per cluster (AVRDC 1997). For this RAPD analysis, 1–10 accessions were chosen from each of the 14 clusters. In those clusters where ten or fewer accessions were available, all were selected. In those clusters with more than ten accessions, a maximum of ten accessions were selected based on subclusters within the cluster. Within subclusters, accessions were chosen randomly. The selection of accessions from the breeding program were designed to maximize genetic diversity for an array of phenotypic traits including yield, disease resistance, and stress tolerance. The procedures used to select accessions from both the genebank and breeding program were designed to maximize genetic diversity.

Prior to distribution by AVRDC, accessions were grown for seed-increase in cages to eliminate outcrossing, and the seed was bulked. Therefore, individuals within accessions were assumed to be homozygous; however, the variation among individuals within an accession was unknown.

DNA extraction and RAPD reactions

Ten seeds were germinated and immature leaves from six to ten plants were harvested for DNA extraction for each accession. For estimating experimental error, an independent sample of seeds from ten, randomly selected accessions (C00274, C00693, C01464, C01717, C01857, C03854, C03882, PBC0518, PBC0550 and PBC0615) were germinated to harvest leaves and extract DNA. The DNA isolation procedure followed Johns et al. (1997). The RAPD (Welsh and McClelland 1990; Williams et al. 1990) reaction mixtures followed Skroch and Nienhuis (1995 a) while the RAPD cycling conditions followed Johns et al. (1997). The 10-mer primers used in this study were chosen for the ability to generate polymorphic markers well dispersed throughout a *Capsicum* genetic linkage map currently under construction by the authors. The primers employed were AB04, AE19, AG14, AG16, AH17, AI01, AI02, AI11, AK04, AK10, AN19, AO18, AT07, AU03, AU14, AV08, AW07, AW15, AX01, AX17, C1, K16, Q05, R12 and Y02 (Operon Technologies, Inc., Alameda, Calif.). RAPD reactions were separated by electrophoresis in 1.5% agarose gels and stained with ethidium bromide. Results were recorded with Polaroid 667 film while illuminating agarose gels with UV light. A comparison of Polaroid films from this study with a separate study that generated RAPDs on a segregating population from a *C. annuum* intraspecific cross was used to initially select RAPDs to be scored. This allowed the selection of a subset of RAPDs that had been tested for Mendelian segregation and could be considered as monogenic traits. Homology between RAPDs in each study was based on similar molecular-weight estimates. Since this study encompassed six *Capsicum* species, more polymorphic RAPDs were generated in the present study. Therefore, selection of the RAPDs to be scored was also based on

Table 1 Accessions from the Asian Vegetable Research and Development Center (AVRDC) included in the RAPD analysis and their accession name, pedigree name, country of origin, and taxonomic label as listed by the AVRDC. Accessions are sorted first according to their site within the AVRDC, the genebank or the pepper breeding program, and second according to species

Entry	Accession name	Pedigree name	Country of origin	Species
Genebank accessions				
1	C00090	CATIE 9776	Costa Rica	*C. annuum*[a]
2	C00200	Sweet Banana	USA	*C. annuum*
3	C00274	Pepperoncini	–	*C. annuum*
4	C00297	PI 169122	Turkey	*C. annuum*
5	C00330-1	Cili Putih Kelantan	Malaysia	*C. annuum*
6	C00377	No. 79 (F7)	Malaysia	*C. annuum*
7	C00560	Unknown 15	Thailand	*C. annuum*
8	C00692	Astrahanskij 147	USSR	*C. annuum* var. *longum*
9	C00693	Karkovskij	USSR	*C. annuum* var. *longum*
10	C00695	Zolotoj Julilej	USSR	*C. annuum* var. *grossum*
11	C00696	Dar Taskenta	USSR	*C. annuum* var. *grossum*
12	C00697	Ukrainskij gorkij	USSR	*C. annuum* var. *longum*
13	C00707	Pimento	USA	*C. annuum*
14	C00765	Kirsch lormiger	Italy	*C. annuum*
15	C00795	–	Honduras	*C. annuum*
16	C00812-A	Caroussel A	Netherlands	*C. annuum*
17	C00812-B	Caroussel B	Netherlands	*C. annuum*
18	C00845	Chile de Arbol	USA	*C. annuum*
19	C00856	–	El Salvador	*C. annuum*[a]
20	C00895	–	Honduras	*C. annuum*
21	C00963	Quadrato D'asti	Italy	*C. annuum*
22	C01150	Hontaka	Japan	*C. annuum*
23	C01242	–	Germany	*C. annuum*
24	C01283	–	Honduras	*C. annuum*
25	C01294	–	Cuba	*C. annuum*
26	C01295	–	Honduras	*C. annuum*
27	C01311	Fips	Netherlands	*C. annuum*
28	C01381	–	El Salvador	*C. annuum*[a]
29	C01438	–	Mexico	*C. annuum*
30	C01464	–	Switzerland	*C. annuum*
31	C01544	–	Costa Rica	*C. annuum*[a]
32	C01586	–	China	*C. annuum*
33	C01632	Albino	–	*C. annuum*
34	C01717	–	Honduras	*C. annuum*
35	C01803	PI 124540	India	*C. annuum*
36	C01811	PI 127442	Afghanistan	*C. annuum*
37	C01857	PI 148631	Iran	*C. annuum*
38	C01884	PI 159264	USA	*C. annuum*[a]
39	C01888	PI 159271	USA	*C. annuum*
40	C01927	PI 164454	India	*C. annuum*
41	C01953	PI 164773	India	*C. annuum*
42	C01958	PI 164848	India	*C. annuum*
43	C01960-B	PI 164961	Turkey	*C. annuum*
44	C01969	PI 165591	India	*C. annuum*
45	C01999-A	PI 169121	Turkey	*C. annuum*
46	C02004-A	PI 169128	Turkey	*C. annuum*
47	C02008	PI 169132	Turkey	*C. annuum*
48	C02030	PI 172774	Turkey	*C. annuum*
49	C02032	PI 172776	Turkey	*C. annuum*
50	C02088	Paradicsomalaku	Hungary	*C. annuum*
51	C02147	PI 178848	Turkey	*C. annuum*
52	C02153	PI 179197	Turkey	*C. annuum*
53	C02156	PI 179201	Turkey	*C. annuum*
54	C02408	Lami Spiral	Austria	*C. annuum*
55	C02422	Jetta	Austria	*C. annuum*
56	C02639	Num-468	Guatemala	*C. annuum*
57	C02649	Num-923	Guatemala	*C. annuum*
58	C02788	PI 268103	Mexico	*C. annuum*
59	C02856	PI 281389	Mexico	*C. annuum*
60	C03837	Roque Mulato	Mexico	*C. annuum*
61	C03841	V-2 Mulato	Mexico	*C. annuum*
62	C03854	W-C 2419 Jalapeño	Guatemala	*C. annuum*
63	C03862	W-C 2065 Dulce	Mexico	*C. annuum*

Table 1 Continued

Entry	Accession name	Pedigree name	Country of origin	Species
Genebank accessions (continued)				
64	C03869	W-C-2129 XCATIC	Mexio	*C. annuum*
65	C03877	W-C 2154 Carizo	Mexico	*C. annuum*
66	C03882	W-C 2173 Verde	Mexico	*C. annuum*
67	C03931-B	PI 439223	Costa Rica	*C. annuum*
68	C05062	–	Zambia	*C. annuum*
69	C05070	–	Zambia	*C. annuum*
70	C05095	–	Zambia	*C. annuum*
71	C00849-A	Selection A	USA	*C. chacoense*
72	C03894	W-C 2419	Mexico	*C. chinense*
73	C05419	Pili pili	Zaire	*C. frutescens*
74	C00569	Unknown 24	Thailand	*Capsicum* spp.[b]
75	C00579	Pangalengan-7	Indonesia	*Capsicum* spp.[b]
76	C02081	Budai Csipos	Hungary	*Capsicum* spp.[b]
Pepper breeding program				
77	PBC0066	MC4-BW	Malaysia	*C. annuum*
78	PBC0079	8209-1 offspring	USA	*C. annuum*
79	PBC0122	HDA 832	France	*C. annuum*
80	PBC0137	CNPH 703	Brazil	*C. annuum*
81	PBC0142	Pant C-1	India	*C. annuum*
82	PBC0148	Punjab Lal	India	*C. annuum*
83	PBC0178	PI 201234	Mexico	*C. annuum*
84	PBC0186	Cheongryong	Korea	*C. annuum*
85	PBC0275	Early Cal Wonder	USA	*C. annuum*
86	PBC0280	Criollo de Morelos 331	Mexico	*C. annuum*
87	PBC0373	Keriting	Indonesia	*C. annuum*
88	PBC0374	Jatilaba	Indonesia	*C. annuum*
89	PBC0375	Paris Minyak	Indonesia	*C. annuum*
90	PBC0412	Tam Mild Chile-2	USA	*C. annuum*
91	PBC0429	NuMex Sunrise	USA	*C. annuum*
92	PBC0436	Sequeira Mendes	Portugal	*C. annuum*
93	PBC0447	75-3-4-4-1-Bk	USA	*C. annuum*
94	PBC0458	Nacional AG-506	Brazil	*C. annuum*
95	PBC0495	Perennial	India	*C. annuum*
96	PBC0518	PSP-11	India	*C. annuum*
97	PBC0531	ROC 29	Italy	*C. annuum*
98	PBC0534	Tumpang	Indonesia	*C. annuum*
99	PBC0550	Lv 1592	Indonesia	*C. annuum*
100	PBC0613	Prapadaeng	Thailand	*C. annuum*
101	PBC0615	Matikas	Philippines	*C. annuum*
102	PBC0673	Hungarian Hot Wax	Hungary	*C. annuum*
103	PBC0738	Lueng 7	Thailand	*C. annuum*
104	PBC0739	Lueng 8	Thailand	*C. annuum*
105	PBC0781	Feherozon	Hungary	*C. annuum*
106	PBC0787	Agronomico 8	Brazil	*C. annuum*
107	PBC0972	Kulai	Malaysia	*C. annuum*
108	PBC0993	Israel Purple	Israel	*C. annuum*
109	PBC1331	–	Malaysia	*C. annuum* var. *aviculare*
110	PBC1418	Lungo Rosso	Italy	*C. annuum*
111	PBC1525	Hybrid Huarena	Thailand	*C. annuum*
112	PBC0151	IAC Ubatuba Cambuci	Brazil	*C. baccatum*
113	PBC0272	C. baccatum pend. 3-1	France	*C. baccatum*
114	PBC0635	Ambul Miris	Sri Lanka	*C. baccatum*
115	PBC0711	Dedo de moca	Brazil	*C. baccatum*
116	PBC0785	IHR517A	India	*C. baccatum*
117	PBC1351	C. baccatum pend. 3-4	France	*C. baccatum*
118	PBC1355	PI 441561	Brazil	*C. baccatum*
119	PBC0445	–	DDR	*C. chacoense*
120	PBC0627	Chaton 2-2	France	*C. chacoense*
121	PBC0113	Habanero	Mexico	*C. chinense*
122	PBC0206	PI 152225	Peru	*C. chinense*
123	PBC0323-B	PI 159236	Peru	*C. chinense*
124	PBC0562	L-178	USA	*C. chinense*
125	PBC0770	Scotch Bonnet	Jamaica	*C. chinense*

Table 1 Continued

Entry	Accession name	Pedigree name	Country of origin	Species
Pepper breeding program (continued)				
126	PBC0879	Ensalada	Puerto Rico	*C. chinense*
127	PBC0912	–	Tanzania	*C. chinense*[a]
128	PBC0410	UL-3878	Nigeria	*C. frutescens*
129	PBC0459	Malagueta	Brazil	*C. frutescens*
130	PBC0475	Gaberawit	Indonesia	*C. frutescens*
131	PBC0537	Rawit Kutoarjo	Indonesia	*C. frutescens*
132	PBC0913	Bird's Eye	Uganda	*C. frutescens*
133	PBC1426	Rocoto	–	*C. pubescens*
134	PBC1473	Manzano Red Rocoto	Bolivia	*C. pubescens*

[a] Species classification was questioned in this RAPD analysis
[b] Accession was assigned to a species in this RAPD analysis

bold amplification strength and discreet amplification differences among accessions. Homology among RAPDs within this study was based on co-migration of RAPD bands. In cases where co-migration among RAPDs could not be determined, the RAPDs were not scored. Data were scored as the presence (1) or absence (0) of a DNA band for each polymorphic RAPD. The treatment of intensity differences was scored following Skroch and Nienhuis (1995a) where intensity differences were considered polymorphic only in unambiguous cases; in all other cases, missing data were entered.

Identification of duplicate accessions

Differences among duplicated accessions could arise due to experimental error alone and may not reflect true genetic differences. Experimental error was estimated by replicating DNA extractions and performing RAPD reactions on each extract. Experimental error was defined as the failure of an accession to be scored consistently over replications due to seed heterogeneity and random errors in the generation and scoring of data. The number of scored inconsistencies between each replicate pair was divided by the total number of markers scored for that replicate pair. The fraction of scored inconsistencies for potential duplicates was equal to, or less than, the mean experimental error plus one standard deviation. For the ten replicated accessions, the replicate with the least missing data was used in further analyses.

RAPD-based diversity among and within *Capsicum* species

Jaccard's similarity coefficient was computed as the ratio of the number of RAPDs present in both accessions to the number of RAPDs present in at least one accession (Jaccard 1908; Gower 1972). Genetic distances were calculated using the complement to Jaccard's similarity coefficient among all pairwise combinations of the 134 accessions where a genetic distance equalling 0 or 1 indicated maximum similarity or difference between a pair of accessions, respectively. The distance matrix was then converted to two-dimensional coordinates using the multidimensional scaling (MDS) procedure (SAS Institute 1990) in order to visualize the relationships among and within species.

Morphological and RAPD-based taxonomic identification

Routine characterization of accessions at the AVRDC consists of scoring 62 morphological traits, including taxonomic assignment (L.M. Engle, personal communication). The accessions from the genebank used in this study were assigned to a species based on morphological descriptors (L.M. Engle, personal communication); the accessions from the breeding program were assigned to species by T. Berke, AVRDC, based on a morphological identification key prepared by P.G. Smith (unpublished). Accessions were later assigned to species based on clustering results from the MDS analysis. Accessions with concordant AVRDC- and MDS-analysis species assignment were used to identify diagnostic markers for each of the six *Capsicum* species studied: *C. annuum*, *C. frutescens*, *C. chinense*, *C. baccatum*, *C. chacoense*, and *C. pubescens*. Diagnostic marker identification required correct identification of accessions typical for each species. Therefore, discrepancies in species assignment for accessions C00090, C00856, C01381, C01544, C01884, and PBC0912 precluded their use. Marker frequencies were calculated for each RAPD among all accessions verified to belong to each species. A diagnostic marker for each of the six species was defined as a marker with a frequency > 0.50 that was found among accessions of one species, but was absent from all other accessions of all other species.

Comparison of genetic diversity between genebank and breeding program accessions

Neither equal numbers of *Capsicum* species nor equal numbers of accessions representing each *Capsicum* species were sampled from the genebank and the breeding program in this study. Therefore, only *C. annuum* accessions which made up a large portion of accessions sampled from both the genebank and the breeding program were used for comparison. To compare levels of genetic diversity, 65 genebank and 35 breeding program *C. annuum* accessions were employed. Accessions C00090, C00856, C01381, C01544, C01884, and PBC0912 were excluded in this analysis due to discrepancies in species assignment. The 74 RAPDs polymorphic among *C. annuum* were used in the comparison. For each RAPD, marker frequency was calculated as the frequency of the presence of RAPD amplification at a RAPD locus among either genebank or breeding program *C. annuum* accessions. Differences in RAPD frequencies between the genebank and breeding program were tested using the chi-square goodness of fit test (Steel and Torrie 1980). For the null hypothesis, the mean marker frequency became the expected marker frequency. Chi-square values and degrees of freedom were pooled over all polymorphic markers. The Spearman rank correlation between RAPD frequencies for genebank and breeding program *C. annuum* accessions was calculated (SAS Institute 1994). Genetic diversity was estimated as the RAPD diversity for each group using Nei's gene diversity at a locus, $h = (1 - \Sigma x_i^2)n/(n-1) = 2pqn/(n-1)$, where

x_i is the frequency of the ith allele, p is the frequency of the presence and q is the frequency of the absence of RAPD amplification among n accessions in each group (Nei 1987). Comparison of the two groups was done using a t-test for paired observations (Steel and Torrie 1980; Nei 1987).

Results and discussion

The 25 RAPD primers produced 124 scored polymorphic RAPDs. In a separate study including only *C. annuum* genotypes, 48 of these RAPDs were amplified, with Mendelian segregation indicating the repeatability of a subset of RAPDs scored in this study. This study included six *Capsicum* species allowing an additional 76 RAPDs to be scored. Missing data accounted for 1.0% of the data. Between two and ten polymorphic bands per primer were scored with a mean and standard deviation ($\pm$SD) of 4.96 ($\pm$1.90) bands per primer. Between one and six polymorphic bands per primer were scored among only the *C. annuum* accessions in this study with a mean ($\pm$SD) of 3.00 ($\pm$1.26). This number of polymorphic bands per primer is higher than that observed among a population of 34 *C. annuum* accessions where 1–3 polymorphic bands per primer were detected (Paran et al. 1998). However, the germplasm in this study was selected to represent the genetic diversity present among all *C. annuum* accessions in the AVRDC genebank and breeding program. In addition, a rigorous pre-selection of RAPD primers, the amplification conditions, or the larger sample size of RAPDs may explain the observed increased level of polymorphism. Fourteen RAPDs were inconsistently amplified across replications, and these markers were deleted from subsequent data analyses. The scoring errors described by Skroch and Nienhuis (1995 b) were nonrandomly distributed with a large portion being distributed within a relatively small subset of RAPDs. Therefore, the removal of RAPDs with inconsistent amplification across replications was warranted. All analyses were performed on the remaining 110 RAPDs.

Identification of duplicate accessions

Molecular marker fingerprints of duplicated accessions were expected to be 100% similar; however, experimental error including errors from the laboratory procedure, errors in the amplification of DNA, and scoring errors may have contributed to Type-I errors (false differences). In addition, seed heterogeneity could result in the possibility of observing different polymorphic markers among independent samples of individual plants within an accession. The mean experimental error ($\pm$SD) was 1.71 ($\pm$1.58)% (Table 2). This level of experimental error indicates that small differences among accessions may reflect variation among individuals within an accession or false differences among accessions. In the present study, ten pairwise comparisons of accessions had a difference in scored markers $\leq$3.29% (Table 3) and therefore could not be discriminated in this study. Four of these pairs of accessions were suspected as duplicates by the authors at the onset of this study: PBC0738-PBC0739, PBC0272-PBC1351, C01311-PBC0993, and PBC1473-PBC1426. Four other groupings, PBC0447-PBC0275, PBC0148-PBC0495-PBC0142, PBC0972-C00377, and PBC1418-PBC0531 could also be considered duplicates by their common origins. PBC0447 is a BC_7F_5 line derived by backcrossing seven times to PBC0275. PBC0148, PBC0495, and PBC0142 are all from India; in addition, PBC0148 was derived from the cross PBC0495 $\times$ Long Red (Long Red was not included in this study). PBC0972 and C00377 are from Malaysia, and PBC1418 and PBC0531 are from Italy.

The existence of duplicates may be due to multiple donations of the same seed source by independent parties. The identification and elimination of duplications in a collection can save time and money in maintenance due to the reduced number of accessions required. The identification of duplicated accessions reduced the 134 studied accessions to 125 unique accessions. The AVRDC *Capsicum* collection held 6901 accessions as of May 1998. If the number of duplicated accessions occurs at an equal frequency throughout the collection then the AVRDC collection is predicted to contain approximately 450 duplicated accessions. However, the 134 *Capsicum* accessions in this study were not randomly sampled from the whole collection. Instead, they were chosen to represent the diversity present in both the genebank and breeding program, so reducing the probability of sampling duplicates. Therefore, the percent of duplicated accessions in the AVRDC *Capsicum* collection is likely to be higher. Estimates of relationships can be better understood

Table 2 Estimate of experimental error measured as the failure of replicated accessions to be scored consistently

Replicated accession	Number of inconsistencies	Number of scored markers[a]	Percent error
C00274	0	115	0.00
C00693	0	83	0.00
C01464	0	121	0.00
C01717	0	113	0.00
C01857	3	117	2.56
C03854	3	118	2.54
C03882	5	117	4.27
PBC518	3	112	2.68
PBC550	3	100	3.00
PBC615	2	99	2.02
Mean experimental error			1.71

[a] Number of comparisons where no missing data existed for either replicate of the accession. The total possible comparisons involve 124 RAPDs

Table 3 Comparisons of accessions which cannot be distinguished because the fraction of differences in their RAPD scores was less than, or equal to, the mean experimental error plus one standard deviation (0.0329)

Accession comparison	Accession names	GD[a]	Fraction of differences
PBC0738 and PBC0739	Lueng 7 and Lueng 8	0.00	0.000
PBC0272 and PBC1351	*C. baccatum* pend. 3-1 and *C. baccatum* pend. 3-4	0.04	0.008
C01311 and PBC0993	Fips and Israel Purple	0.02	0.008
PBC1473 and PBC1426	Manzano Red Rocoto and Rocoto	0.04	0.016
PBC0447 and PBC0275	75-3-4-4-1-Bk and Early Cal Wonder	0.05	0.016
PBC0148 and PBC0495	Punjab Lal and Perennial	0.02	0.016
PBC0495 and PBC0142	Perennial and Pant C-1	0.02	0.018
PBC0142 and PBC0148	Punjab Lal and Pant C-1	0.05	0.018
PBC0972 and C00377	Kulai and No. 79 (F7)	0.03	0.018
PBC1418 and PBC0531	Lungo Rosso and ROC 29	0.05	0.019

[a] GD, genetic distances were calculated using the complement to Jaccard's similarity coefficient

when analyses from independent data sets and other molecular marker types are compared. For example, when four bell-type cultivars were indistinguishable by 21 RAPD primers, two out of ten AFLP primer pairs were sufficient to distinguish the cultivars from each other (Paran et al. 1998). It would be appropriate to verify the close relationship among these accessions with other techniques before considering them to be duplicated.

Genetic relationships among and within *Capsicum* species

Genetic distances among the 8911 pairwise combinations of accessions resulted in a bimodal distribution ranging from 0.00 to 0.92 with a mean genetic distance equal to 0.47 ($\pm$0.25). The mean genetic distance among the 100 *C. annuum* accessions sampled in this study was 0.35 ($\pm$0.09), which was higher than the mean of 0.07 reported by Paran et al. (1998) using the same genetic distance algorithm. However, the germplasm in this study was selected to represent the genetic diversity present among all *C. annuum* accessions in the AVRDC genebank and breeding program. In addition, a rigorous pre-selection of RAPD primers, amplification conditions, or the larger sample size of RAPDs may explain these increased genetic distance estimates. The relationships among accessions were displayed as a MDS plot of the 134×134 genetic distance matrix (Fig. 1). Six discrete clusters in the MDS plot corresponded to the six *Capsicum* species included in this study (Fig. 1). Clustering of species based on RAPDs in the MDS analysis follows the species assignment of each accession based on morphological data, with the exception of the six accessions C00090, C00856, C01381, C01544, C01884, and PBC0912.

Variety status has been used to distinguish between wild (progenitor or weedy) and domesticated peppers as species pairs (Eshbaugh 1968; Heiser and Pickersgill 1969) and to circumscribe morphological variants based on fruit size (Terpo 1966; Eshbaugh 1980). The

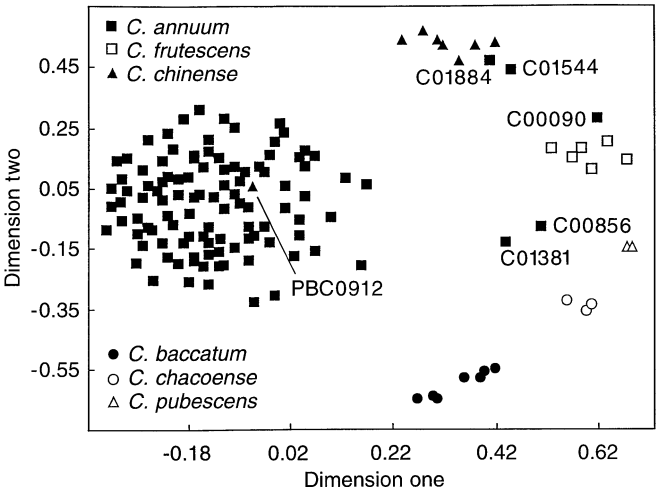

Fig. 1 Multidimensional scaling plot of 134 *Capsicum* accessions based on Jaccard's similarity coefficients from 110 RAPD markers. Accessions were classified according to morphological data

genetic distance matrix was used to identify relationships among accessions assigned to a variety status. No subclusters within *C. annuum* were detected in the MDS analysis. PBC1331, a weedy, small-fruited pepper assigned to var. *aviculare*, was most closely related to PBC0373 (genetic distance = 0.06), a large-fruited cayenne type. C00692, C00693, and C00697, assigned to var. *longum*, did not cluster together in our MDS analysis (mean genetic distance = 0.14), and were more closely related to other accessions such as C01857 and C01311 (mean genetic distance = 0.06), which have relatively small fruits (data not shown). C00695 and C00696, assigned to var. *grossum*, are closely related (genetic distance = 0.04) but also are at a $\leq$0.01 genetic distance unit to other accessions such as C01464 and PBC0787, which were not designated var. *grossum* based on fruit weight (data not shown). The morphological features used to designate varietal status most likely arose from the effects of one to a few discrete loci and it is unknown whether these loci have

been sampled in this analysis. Therefore, the clustering of accessions with different fruit sizes when analyzing markers distributed throughout the genome may be expected. Nevertheless, overall genetic distances derived from multilocus genotypes of these accessions did not support the varietal classifications which have been used in the systematic treatment of domesticated *Capsicum*. In contrast, when *C. annuum* cultivars have been studied with RAPDs by other investigators, sweet, large-fruited peppers were found to be more closely related than small, hot-fruited peppers, with the exception of some paprika and ornamental cultivars (Paran et al. 1998). However, the difference in results may be due to comparing genetic relationships among genotypes to overall fruit phenotype (Paran et al. 1998) rather than solely to one trait, fruit size.

Two accessions, C00856 and C01381, identified according to genebank morphological data as *C. annuum*, appeared as outliers in the MDS plot (Fig. 1). These accessions formed a cluster that is distinct from all other *C. annuum* accessions as well as all other species. Species-specific markers do not reveal a hybrid status in these two accessions (see below). However, the summation of many small differences at many RAPD loci render these two accessions unique from any one species. Since barriers to interspecific hybridization are imperfect in *Capsicum* (Pickersgill 1988), accessions which appear to be the result of hybridization were expected. Nevertheless, we observed only two accessions (C00856 and C01381) which indicated hybridization. We also noted that genetic similarity was greater among accessions within a species than among accessions of different species (Fig. 1). Increasing the amount of seed in cages at the AVRDC may be one mechanism that would minimize the incidence of hybrids in this collection.

Species assignment via diagnostic markers

Only two markers (1.8%) were shared by all six *Capsicum* species. Eight, two, one, two, and one high-frequency marker, (Table 4) served as diagnostic markers for *C. annuum*, *C. chinense*, *C. baccatum*, *C. chacoense*, and *C. pubescens*, respectively (Table 4). A diagnostic marker for each species was defined as a marker with a frequency ≥ 0.50 among accessions of that species, but which was absent from all other accessions from other species. Therefore, the markers are sufficient, but not necessary, for an accession to be typed to a species. The lowest frequency of a diagnostic marker for *C. annuum*, *C. chinense*, *C. baccatum*, *C. chacoense*, and *C. pubescens* was ≥ 0.54, 1.00, 1.00, ≥ 0.67 and 1.00, respectively. A diagnostic marker was not identified for *C. frutescens*. However, marker $AN19_{900}$ was present in all *C. frutescens* and *C. chacoense* accessions examined but not in other species. This marker has a diagnostic value in identifying *C. frutescens* accessions in combination with *C. chacoense*; and these two species can be separated based on other diagnostic markers and morphological data.

Diagnostic markers were used to further characterize six accessions (C00090, C00856, C01381, C01544, C01884 and PBC0912) that did not cluster as expected in the MDS analysis (Fig. 1). This discrepancy in clustering highlighted the potential for either: (1) misclassification of an accession using morphology or (2) inconsistency in species assignment due to accessions having morphological traits of two or more species. Accession PBC0912 has traits of both *C. chinense* and *C. annuum* based on plant, flower, and fruit morphology (T. Berke, unpublished data) and was previously classified as *C. chinense*. However, it strongly clustered with other *C. annuum* accessions in the MDS analysis (Fig. 1), it contained no *C. chinense* diagnostic

Table 4 Diagnostic RAPDs for five *Capsicum* species and their frequencies among *C. annuum*, *C. chinense*, *C. baccatum*, *C. chacoense*, and *C. pubescens* accessions. A diagnostic marker was not identified for *C. frutescens*; however, a marker specific to *C. frutescens* and *C. chacoense* is reported

Marker	*Capsicum*					
	annuum	*frutescens*	*chinense*	*baccatum*	*chacoense*	*pubescens*
$AK10_{1150}$	**1.00**	0.00	0.00	0.00	0.00	0.00
$AX01_{725}$	**1.00**	0.00	0.00	0.00	0.00	0.00
$AW15_{475}$	**0.99**	0.00	0.00	0.00	0.00	0.00
$AV08_{400}$	**0.95**	0.00	0.00	0.00	0.00	0.00
$AU03_{1100}$	**0.94**	0.00	0.00	0.00	0.00	0.00
$AH17_{850}$	**0.85**	0.00	0.00	0.00	0.00	0.00
$AT07_{390}$	**0.68**	0.00	0.00	0.00	0.00	0.00
$AW07_{500}$	**0.54**	0.00	0.00	0.00	0.00	0.00
$AN19_{900}$	0.00	**1.00**	0.00	0.00	**1.00**	0.00
$AE19_{1300}$	0.00	0.00	**1.00**	0.00	0.00	0.00
$AG14_{875}$	0.00	0.00	**1.00**	0.00	0.00	0.00
$AN19_{625}$	0.00	0.00	0.00	**1.00**	0.00	0.00
$Y02_{350}$	0.00	0.00	0.00	0.00	**1.00**	0.00
$K16_{575}$	0.00	0.00	0.00	0.00	**0.67**	0.00
$AX01_{490}$	0.00	0.00	0.00	0.00	0.00	**1.00**

markers, and it contained seven of the eight *C. annuum* diagnostic markers. These results suggest that PBC0912 should be reassigned to *C. annuum*; however, a mixture of two plant types, *C. chinense* and *C. annuum*, have been observed recently by T. Berke when growing this accession in the field; therefore, conclusions about the species identification should not be made until the purity of the seed is verified. Accessions C01884 and C01544 were both classified according to collection data as *C. annuum* accessions. C01884 is an accession from the USDA germplasm collection with the identification number PI 159264. It was also reported as a *C. annuum* accession in the USDA GRIN database (http://www.ars-grin.gov/npgs/). However, C01884 and C01544 clustered with other *C. chinense* accessions in the MDS analysis (Fig. 1), neither had any *C. annuum* diagnostic markers and both possessed the two *C. chinense* diagnostic markers. Recent morphological characterization cannot provide a conclusive species identification for C01884 but C01544 exhibits the campanulate fruit type and annular constriction typical of *C. chinense* (T. Berke and L.M. Engle, unpublished data). These two accessions were apparently misclassified based on morphological descriptors and should be re-assigned to *C. chinense*. Accession C00090 was classified according to collection data as *C. annuum*; however, it clustered with other *C. frutescens* accessions in the MDS analysis (Fig. 1) and possessed no *C. annuum* diagnostic markers. It possessed the marker specific to both *C. frutescens* and *C. chacoense* ($AN19_{900}$) but did not possess the diagnostic markers specific to *C. chacoense* ($Y02_{350}$ and $K16_{575}$). It exhibits the upright fruit type and pale-green flower typical of *C. frutescens* (T. Berke, unpublished data). Apparently, C00090 was misclassified based on morphological descriptors and should be re-assigned to *C. frutescens*.

Diagnostic markers were used to further characterize three accessions (C00569, C00579 and C02081) that were never assigned to a species by the AVRDC (Table 1). All three accessions strongly clustered with other *C. annuum* accessions (Fig. 1). Accessions C00569, C00579, and C02081 contained eight, seven, and seven of the *C. annuum* diagnostic markers, respectively, suggesting that these accessions should be assigned to *C. annuum*. Recent morphological characterization data supports this species identification (L.M. Engle, unpublished data).

Accessions C00856 and C01381 (both collected from El Salvador) appeared to be either of interspecific hybrid origin or outliers from all the other species examined, according to the MDS analysis (Fig. 1). According to collection data, they were classified as *C. annuum*. If a hybrid origin of these two accessions occurred, we expected to observe diagnostic markers from both species distributed among different marker loci. However, C00856 and C01381 possessed one and zero of the eight total *C. annuum* diagnostic markers, respectively, and neither accession possessed any diagnostic markers from any other species. The morphology of these two accessions suggests the occurrence of shared characters between *C. annuum*, *C. chinense*, and *C. chacoense*. C00856 has very small fruits (0.6-cm long $\times$ 0.4-cm wide) which are campanulate in shape. The annular constriction is not clear. Fruits are borne upright and have high capsaicin levels. Flower and plant characteristics resemble *C. annuum* while fruit characteristics resemble *C. chinense* (T. Berke, unpublished data). C01381 also has very small fruits, similar to C00856, but with a clear annular constriction. Flower and plant characteristics, however, resemble *C. chacoense* (T. Berke, unpublished data). These accessions may truly be *C. annuum* accessions which show variability not common in other accessions; alternatively, they may belong to a species not sampled in this study, or may be the result of an interspecific hybrid. If they are interspecific hybrids, the genomic loci revealing evidence for a hybrid origin have not been sampled in this study. Interspecific hybridization was not a common event in this sample of AVRDC germplasm as evidenced by the relative ease in identifying diagnostic markers.

The MDS analysis of the RAPDs detected errors in morphologically based species assignment (i.e., C01884, C01544 and C00090). However, we were able to sufficiently sample the *Capsicum* genome to identify diagnostic markers for each species, except in the case of *C. frutescens*. A weakness of our criteria for defining diagnostic markers to identify species is that it was sampling-dependent and we used few accessions from species outside of *C. annuum*. More non-*annuum* species should be assayed with RAPDs to make our criteria for diagnostic markers more robust. Additional RAPD primers could be screened in order to identify diagnostic markers specific to *C. frutescens*. RAPD diagnostic markers can be cloned and sequenced to develop longer PCR primers (sequence-characterized amplified regions, Paran and Michelmore 1993) with more specificity than RAPDs. Morphologically based taxonomic identification is not always straightforward in the *C. annuum* complex (Pickersgill et al. 1979; T. Berke, unpublished data), and a marker-assisted species assignment can aid curators working in this field. Diagnostic markers could also be used to monitor both natural and artificial hybridization since hybrids in *Capsicum* do not always show intermediate morphology.

Comparison of genetic diversity between genebank and breeding program accessions

C. annuum accessions from the genebank and the breeding program were genetically defined by RAPD frequencies, or analogously by allele frequencies, and any observed differences in these frequencies reflect differences between the groups (Nei 1987). Marker

frequencies were compared using pooled chi-square tests to evaluate the uniqueness of *C. annuum* accessions from the genebank (mean RAPD frequency equal to 0.52) and the breeding program (0.49). The two contrasted groups were not significantly different ($x^2 = 81$, $df = 73$, $P = 0.24$). In addition, the Spearman rank correlation for RAPD frequencies between accessions from the genebank and the breeding program was 0.94 ($P < 0.01$).

Genetic diversity is important to the breeder because the genetic variation present in a breeding program ultimately determines the potential for making gains from selection. Genetic diversity among accessions from the genebank is expected to be greater than genetic diversity among accessions from the breeding program because selection for a small number of economically important traits may have reduced diversity in the AVRDC breeding program. In this study, RAPD-based estimates of genetic diversity were not significantly different ($P = 0.25$) between accessions from the genebank (mean equal to 0.22) and the breeding program (0.20). The lack of significant differences between genebank and breeding program accessions suggests that breeding program has sampled the range of variation available in the genebank *C. annuum* collection. Utilization of genetic resources is an important activity for the breeding program and therefore the use of germplasm representing the range of variation in *C. annuum* is ongoing. Our results verify that the diversity in the genebank was included in the AVRDC's pepper breeding project. Results also indicate that accessions from either collection can be sampled to obtain the range of genetic variation available in the AVRDC *C. annuum* collection as a whole.

References

AVRDC (1997) AVRDC 1996 report. Asian Vegetable Research and Development Center, Shanhua Tainan, Taiwan, pp 49–63

Berke T, Engle L (1997) Current status of major *Capsicum* germplasm collections world-wide. *Capsicum* and Eggplant Newslett 16 : 77–80

Eshbaugh WH (1968) A nomenclatural note on the genus *Capsicum*. Taxon 17 : 51–52

Eshbaugh WH (1980) The taxonomy of the genus *Capsicum* (Solanaceae) – 1980. Phytologia 47 : 153–166

Gower JC (1972) Measures of taxonomic distances and their analysis. In: Weiner JS, Huizinga J (eds) The assessment of population affinities in man. Clarendon Press, Oxford, pp 1–24

Heiser CB, Pickersgill B (1969) Names for the cultivated *Capsicum* species (Solanaceae). Taxon 18 : 277–283

IBPGR (1983) Genetic resources of *Capsicum* – a global plan of action. International Board for Plant Genetic Resources, Rome

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44 : 223–270

Johns MA, Skroch PW, Nienhuis J, Hinrichsen P, Bascur G, Muñoz-Schick C (1997) Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. Crop Sci 37 : 605–613

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

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

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, pp 149–207

Paran I, Michelmore RW (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85 : 985–993

Paran I, Aftergoot E, Shifriss C (1998) Variation in *Capsicum annuum* revealed by RAPD and AFLP markers. Euphytica 99 : 167–174

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

Pickersgill B, Heiser CB, McNeill J (1979) Numerical taxonomic studies on variation and domestication in some species of *Capsicum*. In: Hawkes JG, Lester RN, Skelding AD (eds) The biology and taxonomy of the *Solanaceae*. Academic Press Inc, New York, pp 679–700

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, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. Genome 36 : 404–417

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

SAS Institute Inc (1990) SAS user's guide: statistics ver 6, 4th edition. SAS Inst Inc Cary, North Carolina

SAS Institute Inc (1994) JMP ver 3.1. SAS Inst Inc Cary, North Carolina

Skroch PW, Nienhuis J (1995 a) Qualitative and quantitative characterization of RAPD variation among snap bean (*Phaseolus vulgaris*) genotypes. Theor Appl Genet 91 : 1078–1085

Skroch PW, Nienhuis J (1995 b) Impact of scoring error and reproducibility of RAPD data on RAPD-based estimates of genetic distance. Theor Appl Genet 91 : 1086–1091

Skroch PW, Nienhuis J, Beebe S, Tohme J, Pedraza F (1998) Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. Crop Sci 38 : 488–496

Steel RGD, Torrie JH (1980) Principles and procedures of statistics: a biometrical approach. McGraw-Hill Inc, New York, pp 469–476

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA 85 : 6419–6423

Terpo A (1966) Kritische revision der wildwachsenden arten und der kultivierten sorten der gattung *Capsicum* L. Feddes Rep 72 : 155–191

Villand J, Berke T, Nienhuis J (1996) Characterization of *Capsicum* accessions and cultivars based on RAPD molecular markers and morphological data. Proc Natl Pepper Conf Naples, December 8–11

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18 : 7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18 : 6531–6535