Theor Appl Genet (2009) 119:973–989
DOI 10.1007/s00122-009-1102-6

ORIGINAL PAPER

LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers

Sheh May Tam · Véronique Lefebvre · Alain Palloix · Anne-Marie Sage-Palloix · Corinne Mhiri · Marie-Angèle Grandbastien

Received: 23 January 2009 / Accepted: 25 June 2009 / Published online: 19 July 2009

Abstract Plant genetic resources often constitute the foundation of successful breeding programs. Pepper (*Capsicum annuum* L.) is one of the most economically important and diversely utilized Solanaceous crop species worldwide, but less studied compared to tomato and potato. We developed and used molecular markers based on two *copia*-type retrotransposons, Tnt1 and T135, in a set of *Capsicum* species and wild relatives from diverse geographical origins. Results showed that Tnt1 and T135 insertion polymorphisms are very useful for studying genetic diversity and relationships within and among pepper species. Clusters of accessions correspond to cultivar types based on fruit shape, pungency, geographic origin and pedigree. Genetic diversity values, normally reflective of past transposition activity and population dynamics, showed positive correlation with the average number of insertions per accession. Similar evolutionary relationships are observed to that inferred by previous karyosystematics studies. These observations support the possibility that retrotransposons have contributed to genome inflation during *Capsicum* evolution.

Communicated by G. Bryan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1102-6) contains supplementary material, which is available to authorized users.

S. M. Tam · C. Mhiri · M.-A. Grandbastien (✉)
Laboratoire de Biologie Cellulaire, UR501,
Institut Jean-Pierre Bourgin, INRA,
78026 Versailles Cedex, France
e-mail: gbastien@versailles.inra.fr

V. Lefebvre · A. Palloix · A.-M. Sage-Palloix
Unité de Génétique et Amélioration des Fruits et Légumes,
UR 1052, INRA, Domaine Saint-Maurice BP 94,
84143 Montfavet Cedex, France

Present Address:
S. M. Tam
School of Science, Monash University Malaysia,
Jalan Lagoon Selatan, 46150 Bandar Sunway,
Selangor Darul Ehsan, Malaysia

Introduction

Pepper (*Capsicum annuum* L.) is one of the most economically important and diversely utilized vegetable crop species worldwide. *Capsicum* is a genus of about 25 species in the Solanaceae family and consists of annual and perennial herbs or shrubs native from South and Central America, including the Galapagos Islands (Walsh and Hoot 2001). *C. annuum* is grown worldwide and four additional species are commonly cultivated in south-central America, the Caribbean and Africa, namely *C. baccatum, C. pubescens, C. frutescens* and *C. chinense* (Pickersgill 1997). Though dominating as a hot spice, pepper has a variety of usages, either consumed fresh or is cooked/preserved and dried, used as food dyes, bred as ornamental plants and also provide the important ingredient (capsaisin) for the drug/chemical industry (Djian-Caporalino et al. 2006). To date, different kinds of molecular markers such as isozymes, RFLP, RAPD, AFLP and SSRs, have all been successfully used to complement traditional morphological studies of *Capsicum* (Kwon et al. 2005; Lefebvre 2005).

Agriculture productivity and sustainability are highly dependent on secured access to genetic resources, as they often constitute the foundation of successful breeding programs. Breeding strategies make use of plant accessions of crops, wild relatives (landraces) or mutants maintained at genetic or germplasm centers with aims to increase crop production and quality, ascertain sources for transfer of

beneficial traits such as disease resistance, abiotic stress tolerance and other useful biological traits. Molecular marker techniques now routinely provide highly reliable information in practical and efficient manners for use in the analysis of genetic diversity; generating invaluable data for plant breeding, crop improvement and germplasm management and conservation (Hammer et al. 2003) such as in cultivar protection, establishment of core collections, choice of parental lines for crossings, marker-assisted selection, variety identification (protection) and hybrid purity in the seed industry. Therefore, developing novel and improving current technical approaches applicable in the characterization and management of biodiversity are always required and ongoing.

Transposable elements are known to play important roles in genome evolution—affecting changes in physical genome size, generating chromosomal rearrangements and natural genetic diversity, altering the expression and function of genes and even serving molecular functions such as telomeres, neo-centromeres and promoters (Bennetzen 2002; Biemont and Vieira 2006; Kashkush et al. 2003; Vershinin et al. 2003). In higher plants, LTR (long terminal repeat)-retrotransposons are abundant, amplifying to such high levels as to compose over 60% and over 80% of the maize and wheat genomes, respectively (Bennetzen 2002, 2005). The predominance of LTR retrotransposons is because they do not excise as part of transposition, and thus generate stable polymorphic insertions that may be consequently exploited as molecular markers (Leigh et al. 2003). SSAP (sequence-specific amplification polymorphism) (Waugh et al. 1997) is the first described and perhaps the most utilized among several recently developed retrotransposon-based molecular marker techniques, which also include copia-SSR, IRAP, REMAP and RBIP (Kalendar and Schulmann 2006). SSAP is a special marker system in its dual-functionality—useful as a tool to assess genetic diversity and, rather uniquely, SSAP data also permit comparative studies on the distribution of retrotransposon insertions, and it is this ability to track genomic changes caused by transposition activity that offers remarkable insights into actual evolutionary processes that generate genomic diversity. In addition, SSAP has been reported to be very useful in studies of genetic diversity, genetic relationships and linkage mapping for certain crop species (Ellis et al. 1998; Gribbon et al. 1999; Leigh et al. 2003; Pearce et al. 2000; Porceddu et al. 2002; Vershinin et al. 2003; Yu and Wise 2000; for review, see Tam et al. 2007a).

Solanaceae is a family with many crop species and is relatively well studied for retrotransposons, from the detection of the first active LTR retrotransposon Tnt1A of tobacco (Grandbastien et al. 1989) to subsequent studies in tomato (Costa et al. 1999; Rogers and Pauls 2000), potato (Manetti et al. 2007) and aubergine (C. Mhiri, personal communication). We had previously found SSAP to be highly useful for studying genomic changes in allotetraploid tobacco (Melayah et al. 2004; Petit et al. 2007), genetic diversity within small collections of tomato and pepper from a private company (Tam et al. 2005) and genetic relationships and linkage mapping in *Solanum* subspecies *Lycopersicon,* revealing interesting insights on the distribution and interaction of these ubiquitous elements with their hosts (Tam et al. 2007b). In this study, we are interested in adapting the utility of the SSAP technique to study a representative set of *Capsicum* accessions maintained at the INRA-Montfavet (France) germplasm center. This particular germplasm collection includes mainly traditional open-pollinated cultivars from all over the world, belonging to the five cultivated species of *Capsicum*, and their wild relatives (Sage-Palloix et al. 2007). Results of this study demonstrate that SSAP is a highly polymorphic, useful multi-locus molecular marker tool for the genus *Capsicum.* Analysis of the distributions of retrotransposon-based markers also contributes to the better understanding of evolutionary relationships and genome evolution of *Capsicum* spp.

Materials and methods

Plant material

A set of 86 accessions (Table 1) from the germplasm collection at INRA Montfavet (France) was used in this study. It comprised of seven species of *Capsicum* with a majority of the accessions belonging to the cultivated species *C. annuum* and wild relatives (*C. annuum* var *glabriusculum*). Cultivated inbred lines of *C. frutescens, C. chinense, C. pubescens, C. baccatum* and two wild species represented by one accession each, respectively (*C. chacoense* and *C. eximium*) were also included in the study. The *C. annuum* accessions were chosen to represent the widest diversity in geographic origins, and fruit shapes were coded according to IPGRI, AVRDC and CATIE (1995). Plants were all grown under similar conditions in the greenhouse and five individuals per accession were collected for DNA extraction. Plant DNA was extracted from dried leaf material following the protocol of the Qiagen DNeasy® 96 Plant Kit (Qiagen SA, Courtaboeuf, France). Isolated total DNAs from the five individuals were pooled per accession for further treatment.

Sequence-specific amplified polymorphism (SSAP)

Three retrotransposon primer sets were generated from two LTR-retrotransposon families, namely Tnt1 and T135,

Table 1 List of *Capsicum* accessions

Code	PM code	Name	Geographic origin	Capsaicin content	Fruit code[a]	L[b]	W[b]	$L \times W$[b]	Fruit size[c]	Status[d]
Capsicum annuum (64 accessions)										
C01	31	Yolo Wonder	USA	Sweet	Square (1)	8	8.5	68	Large	*c*
C02	32	Antibois	France	Sweet	Spherical (6)	4	6	24	Medium	*c*
C03	38	Cerise	France	Pungent	Spherical (6)	1.6	2.2	3.52	Small	*c*
C04	57	Doux d'Alger	Algeria	Sweet	Triangular (3)	18	4	72	Large	*c*
C05	69	Doux d'Espagne	Spain	Sweet	Triangular (3)	14	6	84	Large	*c*
C06	76	Doux des Landes	France	Sweet	Elongated (4)	11	2	22	Medium	*c*
C07	148	Niora	Spain	Sweet	Spherical (6)	3	3.5	10.5	Small	*c*
C08	162	Yolo Y = YRP10	USA	Sweet	Square (1)	9	7	63	Large	*c*
C09	194	Lamu	France	Sweet	Rectangular (2)	13.5	6.5	87.8	Large	*c*
C10	217	Piment 493-1 PI 201234	USA–Mexico	Pungent	Triangular (3)	7	2	14	Medium	*c*
C11	223	Pimiento Morron	Spain	Sweet	Spherical (6)	6.5	6.5	42.3	Medium	*c*
C12	278	Largo de Reus	Spain	Sweet	Rectangular (2)	14.5	6.8	98.6	Large	*c*
C13	343	Collectivist 1962	Russia–Romania	Sweet	Square (1)	8	4.5	36	Medium	*c*
C14	355	Nigrum	Serbia	Pungent	Elongated short (5)	4	2	8	Small	*c*
C15	589	Moura	Brazil	Sweet	Rectangular (2)	8.5	5	42.5	Medium	*c*
C16	597	Podarok Moldavi	Russia	Sweet	Heart-shaped (7)	8.8	6.8	59.8	Large	*c*
C17	600	LP1	USA–Mexico	Pungent	Elongated short (5)	3	1.2	3.6	Small	*c*
C18	602	Avelar PI342940	Brazil	Sweet	Triangular (3)	9.5	5.5	52.3	Large	*c*
C19	609	Ancho Esmeralda	Mexico	Pungent	Triangular (3)	9.1	4	36.4	Medium	*c*
C20	610	Jalapeno Rayado	Mexico	Pungent	Elongated short (5)	5.5	3	16.5	Medium	*c*
C21	611	Mulato Roque	Mexico	Pungent	Triangular (3)	9.2	4.6	42.3	Medium	*c*
C22	612	Passilla Apaseo	Mexico	Pungent	Elongated (4)	13.2	1.7	22.4	Medium	*c*
C23	614	Serrano Vera Cruz S.69	Mexico	Pungent	Elongated short (5)	4.5	1.5	6.75	Small	*c*
C24	621	Agronomico 10-G	Brazil	Sweet	Rectangular (2)	10.8	6.2	67	Large	*c*
C25	641	Turrialba (C81) (var *glabriusculum*)	Costa Rica	Pungent	Elongated short (5)	1.2	0.8	0.96	Small	*Wild*
C26	659	Perennial	India	Pungent	Elongated short (5)	4	0.7	2.8	Small	*c*
C27	662	S20-1 Singh4	India	Pungent	Elongated short (5)	5	1.3	6.5	Small	*c*
C28	687	PI 322719	India	Pungent	Rectangular (2)	6.5	3.5	22.8	Medium	*c*
C29	691	Florida VR2	USA	Sweet	Square (1)	8	7	56	Large	*c*
C30	702	Serrano Criollo de Morelos 334	Mexico	Pungent	Elongated short (5)	6	2	12	Medium	*c*
C31	703	Flambeau SM477P	Tunisia	Pungent	Elongated short (5)	6	2.3	13.8	Medium	*c*
C32	721	Morales PRV121 "Mora"	Mexico	Pungent	Elongated short (5)	6	1	6	Small	*c*
C33	747	Incesu 18	Turkey	Pungent	Elongated (4)	20.5	1.8	36.9	Medium	*c*
C34	750	Carliston 52	Turkey	Sweet	Triangular (3)	11.9	3.2	38.1	Medium	*c*
C35	775	Benxi	China	Sweet	Square (1)	5.5	7.5	41.3	Medium	*c*
C36	799	Novi 3	Netherlands	Sweet	Square (1)	9.5	8	76	Large	*c*
C37	807	H3	Ethiopia	Pungent	Elongated (4)	10	2.4	24	Medium	*c*
C38	960	F1 Zhongjiao No. 4	China	Sweet	Square (1)	7.5	8	60	Large	*c*
C39	973	Carre d'Asti Jaune	Italia	Sweet	Square (1)	13	9	117	Large	*c*
C40	1005	Noras	Spain	Sweet	Spherical (6)	4	5.4	21.6	Medium	*c*
C42	1012	SC81	Cuba	Sweet	Triangular (3)	8.5	3	25.5	Medium	*c*
C45	1041	PLOO71-Cuneo	Italia	Sweet	Square (1)	7.5	9	67.5	Large	*c*

Table 1 continued

Code	PM code	Name	Geographic origin	Capsaicin content	Fruit code[a]	L[b]	W[b]	$L \times W$[b]	Fruit size[c]	Status[d]
C46	1053	Doux Long Lombardie	Italia	Sweet	Elongated (4)	10	2	20	Medium	*c*
C47	1066	Chile de Arbol	Mexico	Pungent	Elongated (4)	9.7	1	9.7	Small	*c*
C48	1072	Lamu VR2	France	Sweet	Rectangular (2)	17.5	8	140	Large	*c*
C50	1086	No. 1106 = Serang	Indonesia	Pungent	Elongated (4)	13	2.3	29.9	Medium	*c*
C51	1107	Kahramanmaras	Turkey	Pungent	Triangular (3)	8	2.5	20	Medium	*c*
C52	1109	Piquin (var *glabriusculum*)	Mexico	Pungent	Elongated short (5)	1.5	0.6	0.9	Small	*Wild*
C53	1125	2112B58	Korea	Pungent	Elongated (4)	7.8	1.8	14	Medium	*c*
C54	1166	Early California Wonder (ECW)	USA	Sweet	Square (1)	5.5	7	38.5	Medium	*c*
C55	1172	12HD	Descent C01 × C09	Sweet	Triangular (3)	16	3.2	51.2	Large	*c*
C56	1194	Saint Remy 1	France	Sweet	Rectangular (2)	8	4	32	Medium	*c*
C57	1202	Penis Shape	–	Pungent	Elongated (4)	7	2.5	17.5	Medium	*c*
C58	1204	Numex Twilight	USA	Pungent	Elongated short (5)	2.5	1.2	3	Small	*c*
C59	1210	Poivron d'Ampuis	France	Sweet	Square (1)	4.5	4	18	Medium	*c*
C60	1215	Beldi	Tunisia	Pungent	Elongated (4)	10.5	3.5	36.8	Medium	*c*
C61	1361	PYC22	India	Pungent	Elongated (4)	8	3.6	28.8	Medium	*c*
C62	1406	HDA 801	Descent C29 × C26	Pungent	Rectangular (2)	7.5	3	22.5	Medium	*c*
C63	1407	Phyo 636	Descent C01 × C10	Sweet	Rectangular (2)	13.2	4.8	63.4	Large	*c*
C64	1408	Milord	France	Sweet	Rectangular (2)	10.3	6.5	67	Large	*c*
C65	1409	Vania	France	Sweet	Rectangular (2)	9.5	6.5	61.8	Large	*c*
C66	1410	HDA 103	Descent C01 × C28	Pungent	Rectangular (2)	8.5	6.5	55.3	Large	*c*
C68	1412	Bastidon	France	Sweet	Rectangular (2)	15	7.5	113	Large	*c*
C69	1486	Brûlant d'Espagne	Spain	Pungent	Spherical (6)	3.6	3.2	11.5	Medium	*c*
Capsicum baccatum (8 accessions)										
C70	319	PEN2 (var *pendulum*)	Peru–Bolivia	Pungent	Elongated (4)	10	2	20	Medium	*c*
C71	325	PEN3-4 (var *pendulum*)	Peru–Bolivia	Pungent	Elongated (4)	5.5	1	5.5	Small	*c*
C72	1022	Cristal Blanco (var *pendulum*)	Peru	Pungent	Elongated (4)	9.5	2.1	20	Medium	*c*
C73	1026	Muaraparte Rojo (var *pendulum*)	Peru	Pungent	Elongated (4)	16	1.8	28.8	Medium	*c*
C74	1203	Christmas Bell (var *pendulum*)	Peru–Bolivia	Pungent	Bell shaped (8)	3	6	18	Medium	*c*
C75	1321	PI 439399 (var *pendulum*)	Peru	Pungent	Elongated (4)	12	2.8	33.6	Medium	*c*
C76	1324	PI 439402 (var *pendulum*)	Peru	Pungent	Elongated (4)	18.5	2	37	Medium	*c*
C91	441	SA 326 (var *baccatum*)	Peru–Bolivia	Pungent	Elongated short (5)	1	0.5	0.5	Small	*Wild*
Capsicum chacoense (1 accession)										
C78	1265	PI 260429	Bolivia–Paraguay	Pungent	Elongated short (5)	0.9	0.7	0.63	Small	*Wild*
Capsicum chinense (5 accessions)										
C79	577	PI 159236	USA	Pungent	Elongated (4)	11	2.5	27.5	Medium	*c*
C80	967	Diego	Madagascar	Pungent	Bell shaped (8)	3	3	9	Small	*c*

Table 1 continued

Code	PM code	Name	Geographic origin	Capsaicin content	Fruit code[a]	L[b]	W[b]	$L \times W$[b]	Fruit size[c]	Status[d]
C81	987	Baili	Tchad	Pungent	Bell shaped (8)	4.5	4	18	Medium	*c*
C82	1093	Chile Habanero (Merida)	Mexico	Pungent	Bell shaped (8)	5.1	4	20.4	Medium	*c*
C83	1363	Notto	Senegal	Pungent	Bell shaped (8)	5	5.5	27.5	Medium	*c*
Capsicum eximium (1 accession)										
C85	1372	EXI2	Bolivia	Pungent	Elongated short (5)	1	0.5	0.5	Small	*Wild*
Capsicum frutescens (5 accessions)										
C86	153	Tabasco	USA	Pungent	Elongated short (5)	4	1.5	6	Small	*c*
C87	262	Filo	Burkina Fasso	Pungent	Elongated short (5)	2	0.7	1.4	Small	*c*
C88	330	Grain de Café	Central America	Pungent	Elongated short (5)	2.5	1.2	3	Small	*c*
C89	357	FRU7 (= CHI7)[e]	Caribbeans	Pungent	Elongated short (5)	5	2.5	12.5	Medium	*c*
C90	1108	Charlette	Central America	Pungent	Elongated short (5)	3.3	0.9	2.97	Small	*c*
Capsicum pubescens (2 accessions)										
C92	374	Rocoto REG 502	Peru	Pungent	Spherical (6)	3.5	3.8	13.3	Medium	*c*
C93	1157	54P	Peru	Pungent	Spherical (6)	4.8	4.5	21.6	Medium	*c*

[a] Fruit types codified according to IPGRI, AVRDC and CATIE (1995)

[b] L fruit length, W fruit width, $L \times W$ = fruit length × fruit width

[c] Fruit sizes classified according to $L \times W$: small ≤11, medium 11–50, large ≥50

[d] *c* cultivated

[e] Reclassified as *C. chinense* by the present study

using the primer combinations Tnt1-C00 (Csp6), Tnt1-E00 and T135-E00 (EcoRI) described in Tam et al. (2005). The SSAP assays were carried out following the protocol previously described (Tam et al. 2005).

Data analysis

All SSAP bands were visually scored as present (1) or absent (0) across the seven species of *Capsicum* (86 accessions) and among accessions within each species. Individual binary matrices were built for each primer combination. For SSAP analyses, detected bands of distinct sizes are treated as independent insertions and scored as a character-type.

Genetic relationships between accessions within and among *Capsicum* species were inferred using genetic distance. Mantel's test (Liedloff 1999) was used to assess the strength of correlation between two observed distance matrices, and the statistic tested for significance against 1,000 random permutations. A combined binary dataset from the merger of the three individual primer datasets was analyzed using PAUP* 4.10 (Swofford 2002). Pairwise distances were computed using the mean character difference option which only takes into consideration mismatches and the Neighbor-Joining (Saitou and Nei 1987) algorithm clustered pairwise distances using the minimum evolution option. Clade support was estimated by 1,000 Neighbor-Joining bootstrap replications. The SSAP–NJ tree was first rooted at midpoint of the longest path connecting two taxa in the network ("midpoint rooting"; Kitching et al. 1998) and second, rooted with *C. eximium* based on results of a previous phylogenetic study using *atpB-rbcL* and *waxy* genes (Walsh and Hoot 2001). We also applied the Neighbor Net network on the distance data using the program SplitsTree 4.0 (Huson and Bryant 2006). This distance-based method displays character conflict (data incompatibilities) that may be used to explore complex within (and inter-) species evolutionary processes that will not be well represented by a strictly dichotomous phylogenetic tree (Morrison 2005).

Bayesian clustering analyses were also performed using the program STRUCTURE version 2.2, which permitted the use of dominant markers (Falush et al. 2007; Pritchard et al. 2000). The program uses multilocus genotypes to infer population structure, identify distinct genetic populations, assign individuals to populations and identify admixed individuals assuming K number of populations (Falush et al. 2007). K was allowed to range from one to ten for the interspecific dataset (all accessions), and 1–22 for the dataset comprising of *C. annuum* accessions only, based on the maximum number of geographic origin of the accessions. Two or more replicate runs were made for each K value using the correlated allele frequency model (prior mean = 0.01; prior SD = 0.05 and Lambda = 1.0), with a

burn-in period of 10,000 iterations followed by a run length of 100,000 iterations.

Retrotransposons insertion frequencies for individual and combined datasets for each species were estimated as percentage polymorphism and gene diversity (H_e), where H_e is the total genetic diversity computed over all loci (Nei 1973) using POPGENE vs 1.31 (Yeh and Boyle 1997). For dominant markers, $H_e = 1 - \sum_i \left(p_{n0}^2 + p_{n1}^2\right)$, with $p_{n0} + p_{n1} = 1$ and where p_{n0} and p_{n1} represent, respectively, the frequencies of the absence and presence of alleles of the marker n, with n being comprised in the set of the i markers (Porceddu et al. 2002). In addition, the number of species-specific insertions (present in accessions of a single species regardless of frequency) was calculated from individual retrotransposon datasets. Phylogenetic independent contrasts (PIC), which provide statistical independence when comparing across related taxa were carried out using CAIC (Purvis and Rambaut 1995) with the option of equal branch lengths to estimate relationships between percentage polymorphism and gene diversity with average number of insertions per accession for individual retrotransposon datasets. Relative numbers of retrotransposon insertions were estimated by directly counting the number of bands on the SSAP gels. The phylogeny of *Capsicum* was coded in CAIC based on *atpB-rbcL* and *waxy* DNA sequence evidence (Walsh and Hoot 2001).

Results

Distribution of Tnt1 and T135 retrotransposon markers in seven *Capsicum* spp.

We performed SSAP with the three primer sets Tnt1-C00, Tnt1-E00 and T135-E00, on a set of 86 cultivated and wild accessions comprised of seven species of *Capsicum* (Table 1). Analysis of the three SSAP primer sets scored across the seven species yielded a total of 195 bands, of which 192 (98.46%) are polymorphic among the 86 accessions (Table 2). Species-specific insertions recorded from each retrotransposon dataset for the seven species studied (Table 2) showed Tnt1-C00 having 11 species-specific insertions from a total of 95 scored bands (11.58%), Tnt1-E00 showed 7 out of 50 (14%), while T135-E00 had 6 of 50 species-specific insertions (12%). Mantel's test of pairwise distances between the accessions revealed significantly high correlations ($P < 0.01$) between the datasets (Tnt1-C00 and T135-E00, $r = 0.9209$; Tnt1-C00 and Tnt1-E00, $r = 0.9074$; T135-E00 and Tnt1-E00, $r = 0.9044$). Therefore, the three datasets were combined and used to infer interspecific relationships of *Capsicum*.

Table 2 SSAP data observed from across the 86 *Capsicum* accessions

	Tnt1-C00	Tnt1-E00	T135-E00	Combined
Nb bands scored	95	50	50	195
Nb polymorphic bands	95	50	47	192
% polymorphic bands	100	100	94	98.46
Nb species-specific bands	11	7	6	24
% species-specific bands	11.58	14	12	12.3

The combined SSAP–NJ tree is well resolved and shows well-supported groups corresponding to the different species. The SSAP–NJ tree rooted using the midpoint rooting function (Fig. 1a) shows two main groups, one consisting of *C. chacoense* sister to *C. frutescens, C. chinense* and *C. annuum*, and a second group consisting of *C. baccatum* as sister taxa to *C. eximium* and together, being more closely related to *C. pubescens*. Species groups on the tree are well defined except for the displacement of two accessions of *C. frutescens* (C88, C89), which clustered with *C. chinense*. Branch lengths between accessions of *C. annuum* are relatively shorter compared to the rest of the species. All *Capsicum* accessions used in this study are cultivated populations or cultivars and have undergone breeding selection, except for five wild accessions (see Table 1). Interestingly, the two wild *C. annuum* var *glabriusculum* accessions (C25, Turrialba and C52, Piquin) appear in basal positions to the cultivated *C. annuum*, and the wild *C. baccatum* var *baccatum* accession (C91) also occupies a basal position compared to cultivated *C. baccatum* var *pendulum* accessions. The SSAP–NJ tree rooted with *C. eximium* (Additional file 1) shows *C. baccatum* as sister to the remainder species studied. *Capsicum pubescens* is sister to *C. chacoense,* which in turn is sister to a distinct group consisting of *C.annuum* together with the sister taxa of *C. chinense* and *C. frutescens*.

Neighbor Net analysis of the 86 accessions (goodness of fit = 94.72) produced a network that is remarkably similar to the NJ tree whereby all accessions generally resolved according to their species (Fig. 2). However, several character conflicts are detected, notably among *C. pubescens, C. baccatum* and within *C. frutescens* where one accession (C89) clearly resolved with *C. chinense*.

When the 86 accessions of *Capsicum* were analyzed for genetic structure, the highest likelihood value (Ln $P(D) = -5{,}204.05$, averaged over two runs) was obtained when five genetically distinct groups were identified ($K = 5$) (Table 3). This reflects the high differentiation that exists between the species and corresponded well to the species delimitation of *C. pubescens, C. baccatum, C. annuum, C. frutescens* and *C. chinense* (Fig. 1b). A total

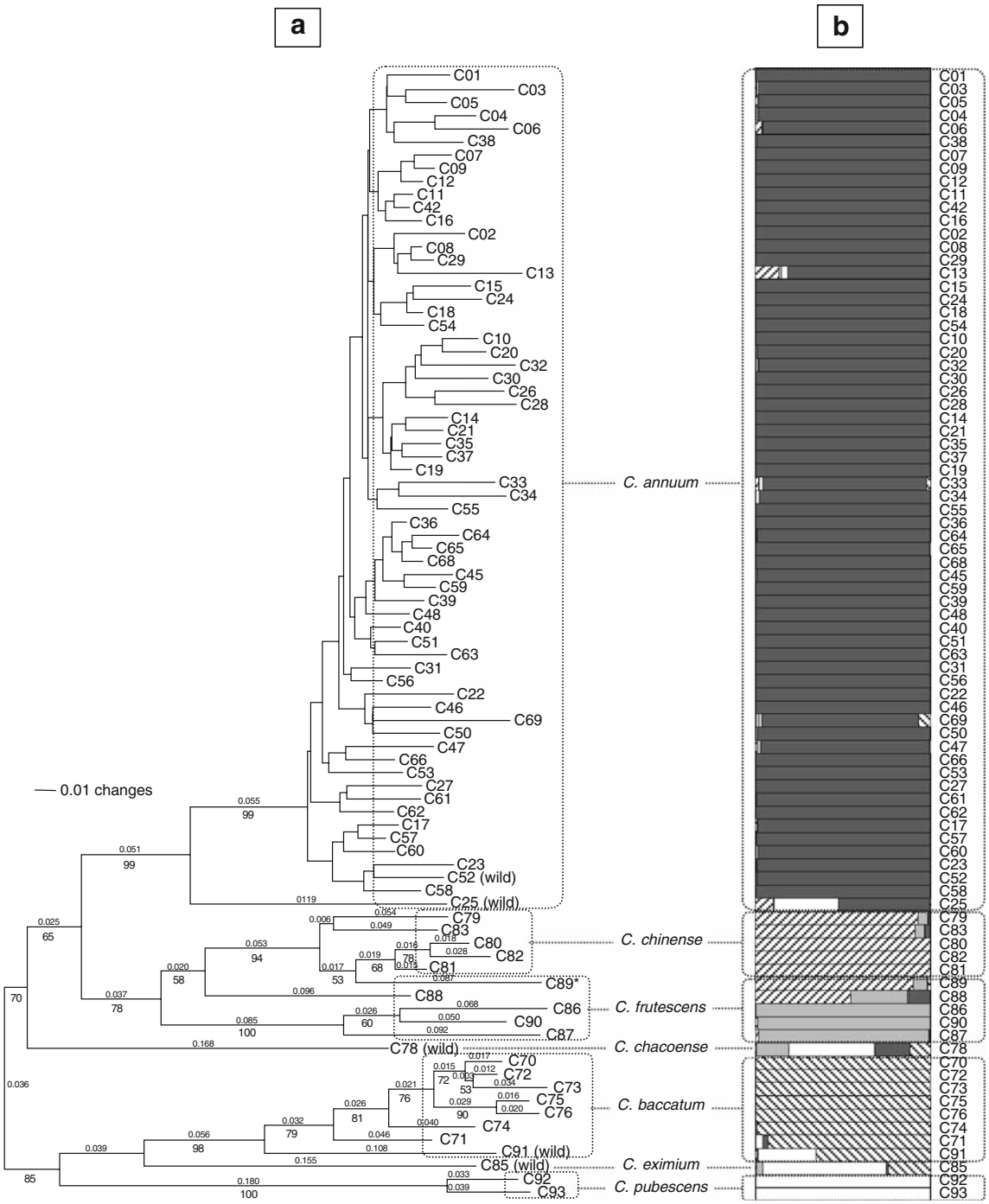

Fig. 1 SSAP–NJ tree and genetic structure of the *Capsicum* collection. **a** Neighbor Joining SSAP tree (SSAP–NJ) depicting genetic relationships among 86 accessions of seven *Capsicum* species, based on the combined SSAP dataset obtained from the three primer combinations Tnt1-E00, T135-E00 and Tnt1-C00. The SSAP–NJ tree was rooted using the midpoint rooting option. *Numbers* above branches indicate genetic change values (for the *C. annuum* accessions see Fig. 3a) and *numbers* below the branches indicate bootstrap support. **b** Genetic structure of the 86 accessions inferred from Bayesian analysis under the admixture model $K = 5$. The five genetic groups are indicated by *white*, *light gray*, *dark gray* and two *diagonally hatched bar* contents

of 15 (17.44%) of the 86 accessions in the sample were categorized as having mixed genome. Of particular interest are *C. baccatum* var *baccatum* (C91), *C. eximium* (C85), *C. chacoense* (C78) and *C. annuum* var *glabriusculum* C25 (Turrialba), which are all wild accessions (Fig. 1b). Interestingly, the *C. frutescens* accession (C89) that resolved in *C. chinense* on NJ and NN trees showed a genetic structure quite similar to *C. chinense* accessions.

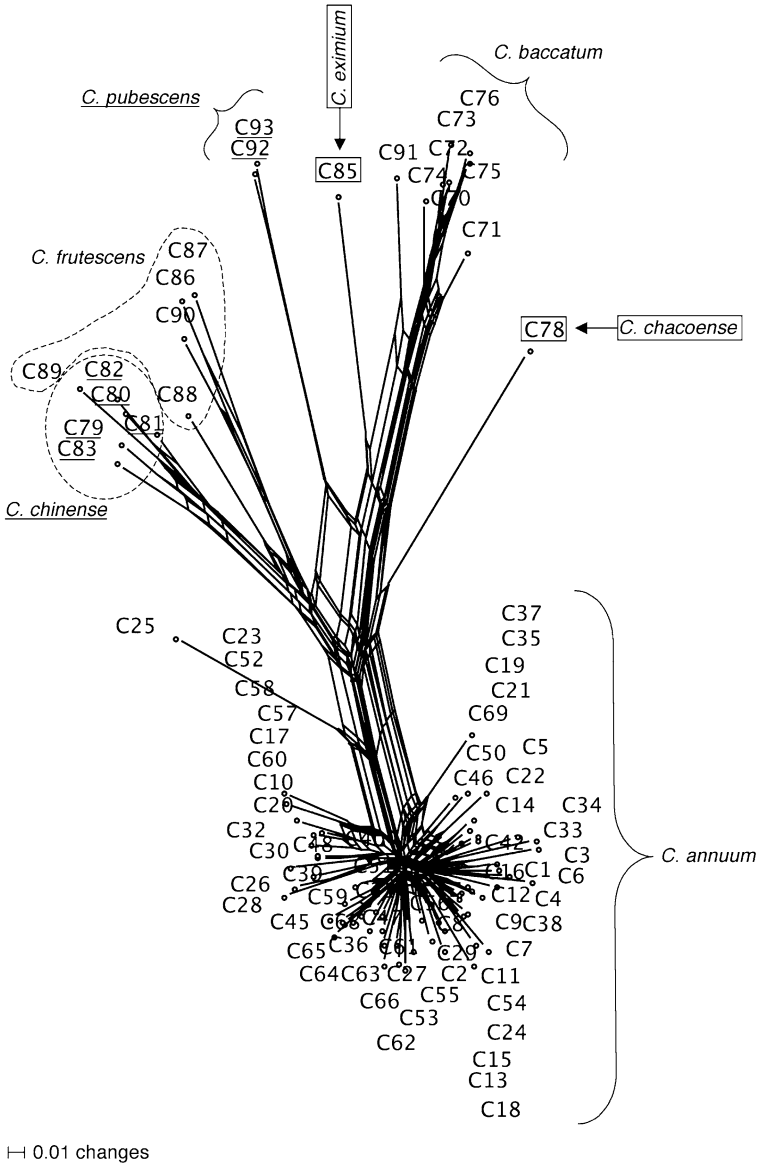

Fig. 2 NeighborNet (NN) SSAP tree of the *Capsicum* collection. The Neighbor-Net (NN) SSAP tree depicts genetic relationships among 86 accessions of seven *Capsicum* species, based on the combined SSAP dataset obtained from the three Tnt1-E00, T135-E00 and Tnt1-C00 primer combinations

Genetic diversity within *Capsicum* species

Genetic diversity values are further compared among four species, namely *C. annuum* (64 accessions), *C. baccatum* (eight accessions), *C.chinense* and *C. frutescens* (five accessions respectively) (Table 4). The number of bands scored from the combined SSAP datasets ranged from 84 for *C. chinense* to 130 for *C.annuum*. The highest percentage polymorphism is calculated from *C. annuum* (82.31%) followed by *C. frutescens, C. baccatum* and lastly *C. chinense* (46.43%). The group displaying the highest gene diversity (H_e) value is *C. frutescens* (0. 276), followed by *C. baccatum, C. annuum* and finally, *C. chinense* (0.172).

Mean genetic distance recorded for *C. annuum* is lowest (0.146), followed by *C. baccatum, C. chinense* and highest for *C. frutescens* (0.390). The maximum genetic distance is estimated from *C. frutescens* (0.557), followed by *C. annuum, C. baccatum* and finally, *C. chinense* (0.321). *Capsicum annuum* has the minimum estimated genetic distance (0.0236), followed by *C. baccatum, C. chinense* and lastly *C. frutescens* (0.189). Higher polymorphism rate (but lower H_e value) in *C. annuum* is probably due to the greater number of accessions sampled from this species. The *C. annuum* collection was divided into two groups: pungent chilli peppers and sweet peppers (Table 4), the former having mean genetic distance estimated as 0.170 and sweet pepper lines, 0.141.

Table 3 Genetic structure analyses based on Bayesian partitioning of ancestry

K[a]	Likelihood value Ln $P(D)$	Variance of Ln $P(D)$
Seven *Capsicum* species (86 accessions)		
1	−8,731.95	546
2	−6,298.55	1,080.4
3	−5,468.5	1,217.45
4	−5,263.65	1,436.6
5	−5,204.05	1,708.55
6	−5,407.15	2,136.15
7	−5,558.3	2,679.25
8	−5,841	2,943.95
9	−6,324.2	4,425.7
10	−6,529.9	4,256.9
Capsicum annuum (64 accessions)		
1	−3,246.6	655.6
2	−3,218.5	946.7
3	−2,970.7	970.03
4	−2,886.67	1,076.47
5	−2,928.4	1,267.1
6	−2,851.75	1,161.85
7	−2,939.15	1,103.95

[a] K = theoretical number of populations; for seven *Capsicum* species, K = 1–10 were tested; for *C. annuum*, K = 1–22 were tested, however, $K > 7$ data are not shown, as no significant improvement was detected

Phylogenetic independent contrast (PIC) test values revealed significant positive correlations between genetic diversity values and average number of insertions per accession for two of the retrotransposon datasets (Table 5). A weak, significant positive correlation is detected between percentage polymorphism and average number of insertions per accession for T135-E00 ($P = 0.084$) while Tnt1-C00 showed significant positive correlation between gene diversity (H_e) and average number of insertions per accession ($P = 0.007$).

Genetic relationships within the *C. annuum* accessions

The SSAP–NJ tree shows accessions of the *C. annuum* resolving as a basal group followed by three distinct clusters (Clus1, which is sister to two more closely related clusters, Clus2 and Clus3; Fig. 3a). Genetic relationships for *C. annuum* inferred from the SSAP–NJ tree show rather good coherence with pepper fruit size and with pepper fruit code (see Table 1). The six other *Capsicum* species studied here are all pungent, small/medium fruited accessions (Table 1), reflecting the lack of selection for sweet large fruits in these species. For *C. annuum*, large-fruited varieties mainly group in Clusters 1 and 3, and small/medium-fruited varieties mainly group in Cluster 2, as well as in the basal group. It is observed that the large sweet peppers in Cluster 1 are very closely related, while the small/medium varieties in Cluster 2 appear more diverse with relatively higher branch lengths (genetic distances). Square (1) and rectangular (2) peppers group in Cluster 1 of large sweet peppers. They are also found in Cluster 3, together with the spherical peppers used mainly for canned food (6 and 7). Elongated peppers (4 and 5), on the other hand, are found in groups of small/medium pungent peppers (basal group and Cluster 2). Finally, triangular peppers (3) are found dispersed among Cluster 2 and Cluster 3, as they are thought to derive either from reduction in size of quadrangular peppers or from proximal enlargement of elongated peppers.

The SSAP–NJ tree also reflects well information of known pepper pedigrees, both in terms of certain geographical origin and breeding programs. For instance, C64 (Milord) and C65 (Vania) belong to the same population derived from parental accessions C68 (Bastidon) and C39 (Carré d'Asti jaune), and all four accessions are closely related within Cluster 1. Cluster 1 also groups sweet peppers lines derived from traditional local populations (C45, C39, C48 and C59) or contemporary selection (C36, C64, C65, C68 and C63) from Western Europe. Cluster 2 groups pungent pepper types from ancient local populations of diverse geographical origins. Traditional small/medium-fruited Mexican cultivars (C19, C20, C21, C30 and C32) group within the same sub-cluster, separated from traditional Turkish cultivars (C33 and C34). Interestingly, however, some traditional Indian (C26 and C28), as well as Ethiopian (C37) and Chinese (C35) varieties resolved close to the Mexican varieties, suggesting that the later four varieties might have origins from Mexican populations now dispersed into diverse regions of the world. Cluster 3 clearly include traditional sweet fruited cultivars from Mediterranean Europe (C05, C04, C06, C07, C09, C12, C11 and C02), and tightly groups three varieties (C15, C24 and C18) issued from a single Brazilian genetic program selecting for viral resistance together with an US accession C54. In addition, Cluster 3 also contains C01 (Yolo Wonder), C08 (Yolo Y), C29 (Florida VR2) and C54 (ECW), all issued from the same genetic pool (American "blocky" types). More disparate is the presence in Cluster 3 of accessions from different geographical origins such as C38 (China), C42 (Cuba), C16 and C13 (Russia), suggesting relationships based on common exchanges.

For the *C. annuum* accessions, the clustering approach implemented by the STRUCTURE program inferred four source populations, with the probability function peaking at $K = 4$ (Ln $P(D) = -2{,}886.7$ averaged over three runs), and values of Ln $P(D)$ did not improve significantly for $K > 4$ (Table 3). The distribution of the inferred four subpopulations appears to correspond to a certain degree

Table 4 Genetic diversity values of *Capsicum* species

Capsicum species	*C. annuum*	*C. baccatum*	*C. chacoense*	*C. chinense*	*C. eximium*	*C. frutescens*	*C. pubescens*
Nb accessions	64	8	1	5	1	5	2
Individual datasets							
Tnt1-C00							
Nb bands scored	63	48	39	38	36	65	30
Nb polymorphic bands	55	26	–	14	–	52	5
% polymorphism	87.3	54.17	–	36.84	–	80	16.67
H_e	0.1878	0.2003	–	0.1404	–	0.2921	0.069
Average nb of bands/accession	33.12	36.5	39	31.2	36	37.8	27.5
Standard deviation	2.585	1.512	–	1.095	–	4.147	0.707
T135-E00							
Nb bands scored	35	33	18	26	17	32	21
Nb polymorphic bands	26	18	–	12	–	21	5
% polymorphism	74.29	54.55	–	46.15	–	65.62	23.81
H_e	0.2064	0.1724	–	0.1716	–	0.2463	0.0986
Average nb of bands/accession	22.39	23.75	18	19.8	17	21.4	18.5
Standard deviation	2.090	1.389	–	1.304	–	3.05	0.707
Tnt1-E00							
Nb bands scored	32	24	16	20	17	25	19
Nb polymorphic bands	26	19	–	13	–	20	4
% polymorphism	81.25	79.17	–	65	–	80	21.05
H_e	0.1378	0.3054	–	0.2341	–	0.2710	0.0872
Average nb of bands/accession	13.53	16.5	16	13.2	17	13.4	17
Standard deviation	1.736	2.39	–	0.837	–	2.191	0
Combined dataset							
Nb bands scored	130	105	73	84	70	122	70
Nb polymorphic bands	107	63	–	39	–	93	14
% polymorphism	82.31	60	–	46.43	–	76.23	20
H_e	0.1805	0.2156	–	0.1724	–	0.2758	0.0828
Genetic distances (gd): all accessions							
Mean gd	0.146	0.218	–	0.233	–	0.390	–
Max gd	0.409	0.390	–	0.321	–	0.557	–
Min gd	0.0236	0.0667	–	0.107	–	0.189	–
Mean sampling variance	0.00094	0.00154	–	0.00205	–	0.00183	–
Genetic distances (gd): pungent chillis							
Mean gd	0.170	–	–	–	–	–	–
Mean sampling variance	0.00111	–	–	–	–	–	–
Genetic distances (gd): sweet peppers							
Mean gd	0.141	–	–	–	–	–	–
Mean sampling variance	0.00109	–	–	–	–	–	–

Table 5 Estimate of relationships between parameters obtained from phylogenetic independent contrasts

Relationships between	Tnt1-C00		T135-E00		Tnt1-E00	
	P	*m*	*P*	*m*	*P*	*m*
% polymorphism and average nb of bands	0.1398	4.11	0.0837	3.48	0.3478	3.22
H_e and average nb bands	0.0071	0.02	0.1871	0.02	0.9173	0.0042

P probability; *m* slope of regression

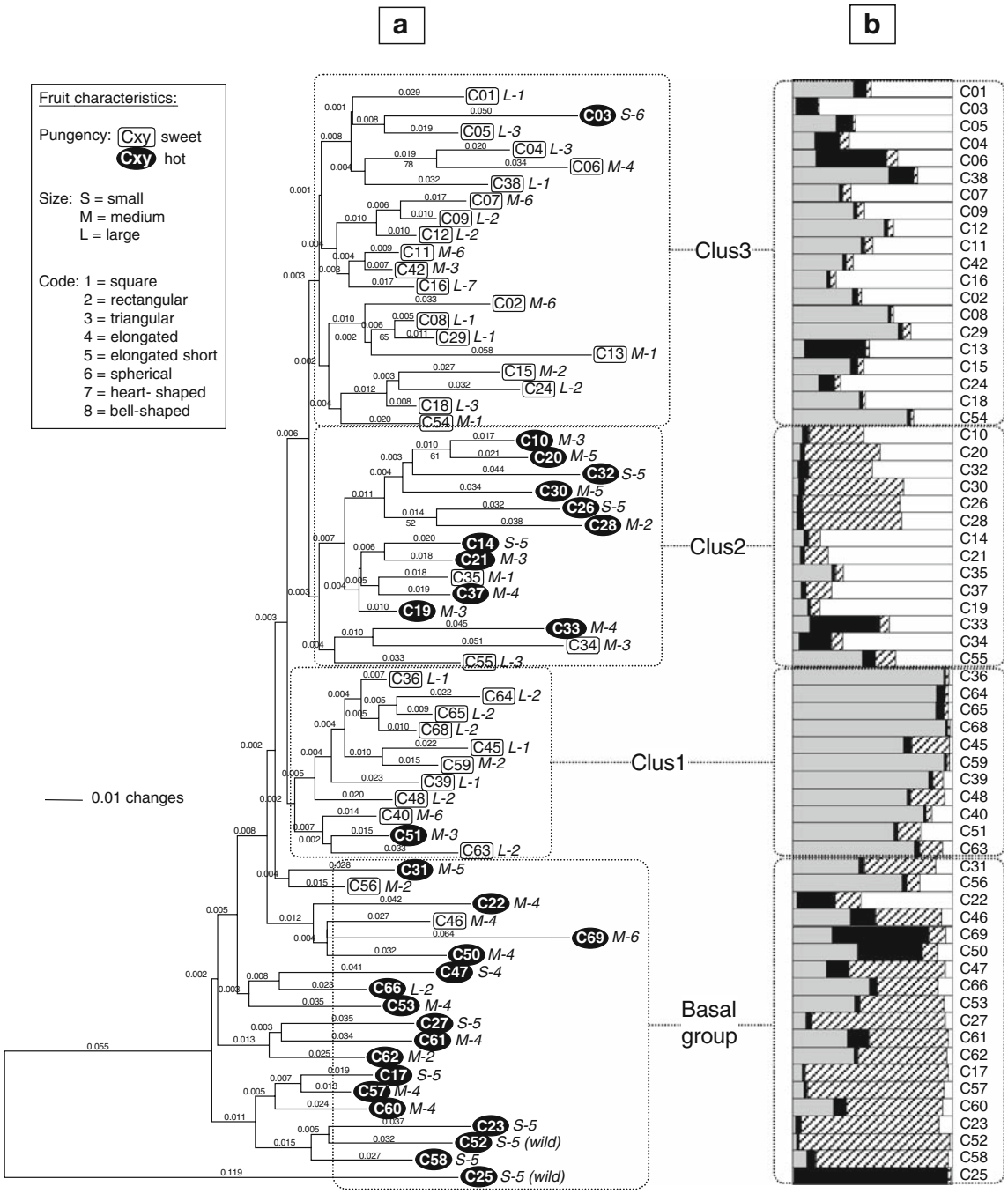

Fig. 3 SSAP–NJ and genetic structure of the *Capsicumannuum* collection. **a** Details of the Neighbor Joining SSAP tree (SSAP–NJ) shown in Fig. 1a, highlighting genetic relationships among 64 accessions of *Capsicum annuum*. Fruits characteristics are indicated on the *left*: fruit shapes were coded according to IPGRI, AVRDC and CATIE, and fruit sizes are based on length × width, small <11, medium 11–50 and large >50. *Numbers* above branches indicate genetic change values. **b** Genetic structure of the 64 *C. annuum* accessions inferred from Bayesian analysis under the admixture model $K = 4$. The four genetic groups are indicated by *black*, *white*, *light gray* and *diagonally hatched bar* contents

with the clusters inferred by the Neighbor Joining method (Fig. 3b). No non-admixed accession was identified; only nearly non-admixed ancestry was inferred for many individuals corresponding to the four source populations. Certain accessions represent the best of the non-admixed source population; they are genetically distant from each other according to the SSAP NJ tree. C25 (Turrialba) is distant from other *C. annuum* accessions, and is considered as the wild ancestor of *C. annuum* species. C52 (Piquin) is the wild accession that could well represent the basal

group. C68 (Bastidon) could represent the best among the large sweet cultivars of Cluster 1. C03 (Cerise) could represent both the branches at the base of Cluster 2 and Cluster 3.

Discussion

Interspecific relationships of domesticated peppers inferred from insertions of Tnt1 and T135 retrotransposons

Three combined SSAP datasets generated using the Tnt1 and T135 families of LTR-retrotransposons successfully inferred interspecific relationships among seven species of *Capsicum*. Molecular phylogeny of *Capsicum* has not been as extensively studied compared to other Solanaceous genera such as *Solanum* subspecies *Lycopersicon* or *Nicotiana.* The SSAP–NJ tree is well resolved and shows several well-supported clusters useful for species delimitation. In general, the tree topology shows good agreement with botanical classifications issued from geographical, ethnobotanical and cytogenetic data (Djian-Caporalino et al. 2006). Results from SSAP analysis is also compatible with previous studies, such as the molecular phylogenetic study of 11 species of *Capsicum* based on combined sequences of *atpB-rbcL* and *waxy* (Walsh and Hoot 2001). The SSAP–NJ tree clearly distinguishes the *C. annuum–C. chinense–C. frutescens* complex from other *Capsicum* species as illustrated by Eshbaugh et al. (1983). *Capsicum chinense* and *C. frutescens* resolved as sister taxa to the highly supported group of *C. annuum*. All accessions clustered according to species except for two accessions of *C. frutescens*, which resolved basal to *C. chinense,* inadvertently supporting past suggestions to combine these two species because of morphological similarities and possible inter crossing between these two species (Djian-Caporalino et al. 2006). It was also proposed that *C. frutescens* could be the putative ancestor of *C. chinense* (Eshbaugh et al. 1983). Based on the position of the *C. frutescens* C89 accession on NJ and NN trees, genetic structure similarity with *C. chinense* accessions and recent morphological re-analysis, we are able to reclassify C89 as *C. chinense* (CHI7) (Table 1). *Capsicum chacoense* resolved on the SSAP–NJ tree as the sister species to the *C. annuum–C. frutescens–C. chinense* complex, and this is in agreement with the morphological similarity of these taxa and evidence based on enzymatic studies and *atpB-rbcL* and *waxy* sequences (McLeod et al. 1982; Onus and Pickersgill 2004; Walsh and Hoot 2001).

The grouping of *C. baccatum, C. eximium* and *C. pubescens* on the SSAP–NJ tree is coherent with previously reported species relationships (Nagy et al. 2007; Park et al. 2000), with *C. baccatum* appearing as a group on its own while *C. eximium* and *C. pubescens* (together with *C. cardenasii*) forming a basal complex group with purple flowers (McLeod et al. 1982; Nagy et al. 2007; Onus and Pickersgill 2004). However, the *atpB-rbcL—waxy* and microsatellite markers (SSR)-based trees positioned *C. baccatum* as sister to *C. chacoense* and together; they formed the sister group to the *C. annuum–C. frutescens–C. chinense* complex which is concordant with crossing compatibility between these species. The SSAP–NJ tree rooted with *C. eximium* showed that *C. pubescens* is sister to *C. chacoense* and to the *C. frutescens–C. chinense–C. annuum* group, and this genetic relationships is different from what was reported by most previous studies, except for an earlier study of six *Capsicum* species using RAPD markers (Rodriguez et al. 1999) that showed agreement with SSAP in species relationships.

To further explore genetic relationships, the Neighbor-Net network was used to depict character-states pattern conflict in the SSAP data. Reticulations in a character display network can be interpreted to represent data uncertainty/error (mistaken homology, model heterogeneity or insufficient data), actual homoplasy in the phylogeny or evolutionary events that involved gene exchange between organisms such as recombination and hybridisation (Huson and Bryant 2006; Morrison 2005). Neighbor-Net indicated prominent reticulate relationships between accessions of *C. pubescens* (C92, C93), *C. baccatum* var. *baccatum* (C91) and *C. frutescens* (C88), in addition to many splits (uncertainty) seen among the *C. annuum* lines. Model-based structure analysis revealed the presence of five populations for the *Capsicum* accessions studied, of which individuals mostly corresponded to the species identified from botanical traits. Interestingly, the different position of *C. pubescens* inferred on the SSAP–NJ tree can be further understood by observations that Neighbor-Net displayed reticulation for C92 and C93, but structure analysis showed both to possess non-admixed genotypes. The genotype of *C. pubescens*, however, was also detected in both *C. eximium* and *C. chacoense*, and this could explain its position on the SSAP–NJ tree which grouped accessions based strictly on genetic similarity in a dichotomous fashion.

Phylogenetic (Neighbor Joining and Neighbor-Net) and model-based structure analyses of the sampled *Capsicum* accessions also showed that the most admixed genotypes correspond to the wild accessions (C78, C85, C91 and C25), which include alleles from two to four distinct species. Another case is the *C. frutescens* accession C88 which displays primitive (wild) traits such as very small and dehiscent fruits; it also showed a admixed genotype. This suggests that wild accessions share diverse alleles, which were subsequently fixed through speciation. SSAP data also show that the germplasm collection cannot be wholly

characterized based on phenotypic (morphological) characters alone. For example, individuals of two populations corresponded well to their assigned species (*C. baccatum* and *C. annuum*), though both showed small groups with admixed genotypes (25 and 9.4%, respectively). The remainder three populations, however, included individuals that were assigned to separate species (especially between *C. chinense* and *C. frutescens*) and accessions of *C. pubescens* were clustered with *C. eximium* and *C. chacoense*.

A possible explanation for the SSAP-inferred genetic relationship is that the evolutionary history of the *Capsicum* species is influenced by the activity of the LTR-retrotransposons, since *Capsicum* is known to have relatively large genomes (3,753–4,763 Mbp) (Bennett and Leitch 2005). Moscone et al. (2003) performed a karyosystematic study on 11 *Capsicum* species and reported that nuclear DNA content measurements in the genus demonstrated significant differences in genome size between taxa. The species studied were delimited as significant groups with characteristic genome sizes (1C value/genome size in Mbp) and karyological data. Four species were included in the same Scheffe group with similar, comparatively low DNA amount and little heterochromatin displayed, namely *C. chacoense* (3.35 pg/3,283 Mbp), *C. frutescens* (3.4 pg/3,332 Mbp), *C. chinense* (3.42 pg/3,352 Mbp) and *C. annuum* var. *annuum* (3.38 pg/3,312 Mbp). The next group (3.71 pg/3,636 Mbp) consisting of *C. baccatum* var. *baccatum* (3.71 pg/3,636 Mbp), *C. baccatum* var. *pendulum* (3.68 pg/3,606 Mbp) and *C. baccatum* var. *umbilicatum* (3.76 pg/3,685 Mbp) showed increased DNA content and more complex heterochromatin banding pattern, corroborating that *C. baccatum* should have a separate placement from the species mentioned. *Capsicum eximium* (4.06 pg/3,979 Mbp) was shown to be intermediate in DNA content between *C. baccatum* and the comparatively larger genome and larger amount of complex patterned heterochromatin of *C. pubescens* (4.47 pg/4,381 Mbp) and thus *C. eximium*, *C. baccatum* and *C. pubescens* were placed into their own respective Scheffe groups. Therefore, it is of interest that the SSAP–NJ tree displayed similar genetic relationships among the *Capsicum* species as suggested by Moscone et al. (2003), reflecting the close relationship of *C. chacoense* to the *C. frutescens*–*C. chinense*–*C. annuum* complex and the position of *C. eximium* in between *C. baccatum* and *C. pubescens*. However, we note that no correlation is apparent between DNA content and copy numbers of the two LTR-retrotransposons studied here.

Genetic diversity and relationships of *Capsicum annuum*

The SSAP technique could be successfully applied to characterize the genetic diversity within *Capsicum* accessions. Highly polymorphic SSAP markers generated DNA profiles that showed no duplicate accession within the collection studied. Identification and elimination of duplicated accessions in a collection is extremely useful in terms of time and cost saved in germplasm maintenance and breeding programs (Rodriguez et al. 1999). Important parameters normally considered in studies relating to crop genetic diversity include the level of genetic variation within populations/groups and the extent of genetic divergence among populations/groups. Molecular markers that detect high levels of polymorphism between cultivars help improve the efficiency and accuracy of genetic similarity estimates (Ribeiro-Carvalho et al. 2004). The SSAP markers show that the pepper collection possesses a high degree of polymorphism (82.31%) and a moderate value of gene diversity ($H_e = 0.1805$).

SSAP shows that the *C. annuum* samples display 1.8-fold higher polymorphism levels than the *C. chinense* samples, while the *C. frutescens* and *C. baccatum* samples are intermediate (1.1-fold and 1.4-fold, respectively), albeit sample sizes being not balanced between species. However, in terms of gene diversity (H_e), *C. frutescens* possess the highest variation, and are 1.3-, 1.5- and 1.6-folds higher than *C. baccatum, C. annuum* and *C. chinense*, respectively. Other studies had also reported that *C. annuum* possess intermediate to moderately high levels of genetic variation for a self-pollinating crop species, using RFLP, RAPD, AFLP and morphological traits (Lefebvre et al. 1993, 2001); AFLP and RAPD (Lanteri et al. 2003); morphology and AFLP (Geleta et al. 2005). High genetic variation could be due to higher levels of outcrossing than normally expected (Tanksley 1984). In a study of an Italian landrace of pepper using RAPD and AFLP, it was observed that most variation was partitioned within, rather than between populations, and this was not consistent with the selfing breeding system of *C. annuum*, but rather indicated a high level of outcrossing in the pepper landrace (Lanteri et al. 2003).

The SSAP–NJ tree shows a basal group closely related to three defined clusters—one cluster of pungent and two clusters of sweet (non-pungent) fruited peppers, respectively. Pungent accessions are mostly small-medium and long fruited, which positioned basal to the sweet large peppers. Sweet large pepper types are separated in two clusters, suggesting origins from two distinct genetic pools, as confirmed by two markedly different types of genetic structure. Altogether, our data are informative and correlate with pepper fruit shape and/or with pedigrees and geographic origins, permitting potential determination of genetic pools. Both distance- and model-based analyses show that many accessions of *C. annuum* carry genetic evidence of some level of admixture within each of the four subpopulations inferred. This could be accounted for by the process of artificial selection and breeding carried out during the

development of these lines. Introgression among different source subpopulations would create intermediate types that will not be easily distinguished at the phenotypic level, and is also difficult to visualize using a Neighbor Joining tree that is forced to represent the data into a simplified hierarchical structure. In particular, we were unable to relate the splitting of sweet-fruited cultivars in two clusters to geographic or pedigree information. Conversely, the subgroups within these clusters were more clearly related with geographic origin and pedigree information. For this pepper study, the results of the splits network and genetic structure analyses definitely present a more realistic and useful representation of the reticulate genetic relationships by allowing the visualization of shared (and conflicting) character states with individuals from other sources/groups.

Additional estimates of mean genetic distance of the pre-determined two groups in the pepper collection is also consistent with those reported by previous studies (Lefebvre et al. 1993, 2001; Paran et al. 1998) where small-fruited cultivars were found to harbor higher degrees of genetic variation compared to the large fruited sweet cultivars, suggesting that small-fruited cultivars were selected from a larger genetic pool and suffered lower genetic bottleneck effects. Similarly, Lanteri et al. (2003) found little genetic variation in a landrace of elongated bell pepper type from North West Italy using RAPD and AFLP. In this study, the pungent chilli group, which contains small and long fruited accessions, has a higher mean genetic distance than sweet pepper lines.

Evolutionary dynamics of Tnt1 and T135 retrotransposons in the genome of *Capsicum*

At present, much is still unknown regarding the distribution of similar types of transposable elements among closely related species (Hawkins et al. 2006) even though transposable elements are ubiquitously present in the genomes of eukaryotic organisms. Our series of ongoing studies have successfully detected and compared insertions of Tnt1 and T135 (originally isolated from tobacco and tomato, respectively) between different species in several genera of the Solanaceae (Manetti et al. 2007; Tam et al. 2005), indicating that these elements have persisted through the duration of speciation for long periods of time. Therefore, we can successfully develop insertion-based molecular markers that are transferable to other Solanaceous species.

One important advantage of the SSAP technique is that genetic diversity values obtained by SSAP indirectly permit detection of activity and maintenance of different retrotransposons in a particular set of hosts where higher diversity values are indicative of more recent transposition activities (Ellis et al. 1998; Gribbon et al. 1999; Kalendar et al. 1999; Vershinin et al. 2003). Elements can transpose at high frequency, with a rate ranging from 10^{-3} to 10^{-5} per element per generation (depending on element) compared with a nucleotide-base substitution rate of 10^{-8}–10^{-9} per nucleotide per generation. Therefore, transposable elements are recognized as powerful sources of genetic diversity (Biemont and Vieira 2006). High rates of SSAP polymorphisms coupled with the identification of species-specific bands (unique bands, regardless of frequency) observed in this study both indicate that retrotransposition events have occurred during the speciation of *Capsicum.* Both Tnt1 and T135 have similar proportion of species-specific bands (12.41 and 12%, respectively), contributing to a notable proportion of natural diversity among genomes of the different *Capsicum* species. This finding supports the important role of retrotransposon activity in generating natural genetic diversity in the various lineages of *Capsicum*.

The evolutionary dynamics of transposable elements essentially reflect interactions between element proliferation and loss as a result of one or more opposing deterministic forces that includes selection acting through the host, genetic drift and self regulatory behavior (Morgan 2001; Sniegowski and Charlesworth 1994; Wright and Schoen 1999). In this study, no apparent difference in lineage-specific amplification of retrotransposon is observed between Tnt1 and T135. No differences were detected in the average number of insertions per accession between the two wild (*C. eximium, C. chacoense*) and five cultivated *Capsicum* species (difference of only 1.03-fold), or between wild *C. chacoense* and the domesticated *C. annuum–C. chinense–C. frutescens* complex (1.06-fold) and between the one self-incompatible (*C. eximium*) and six inbreeding species (1.05-fold). Significant differences in copy number and insertion frequencies between species reflect a strong negative selection against retrotransposon insertions (or even slightly deleterious mutations) as a consequence of higher genetic load associated with inbreeding depression or domestication (Morgan 2001; Wright and Schoen 1999) or, inversely, higher accumulation of elements and at higher frequencies in selfers due to the influences of reduced recombination and effective population size (Wright et al. 2003). Also, retrotransposon insertions would not necessarily always prove useful for inferring host phylogeny. For instance, the Tnt1-ol13 retrotransposon population showed an extremely high number of insertion polymorphisms in *Nicotiana*, but no coherent interspecific phylogeny could be inferred (Melayah et al. 2004). Another example include the analysis of nine selected genotypes of sweet potato by SSAP, which revealed that the majority of the Ty1-copia transposon insertions were unique (33–64%) with only few common bands detected (Berenyi et al. 2002). IRAP markers reportedly showed high levels of genetic diversity, but did not discriminate individuals of *Aegilops tauschii* based on

morphology, possibly due to different activity of retroelements (Saeidi et al. 2008). Hence, our results in *Capsicum* may be taken to indicate the selective neutrality of most of the insertions revealed by SSAP, suggesting that these insertions are well tolerated by the *Capsicum* host genomes.

Interestingly, our study revealed weak but significant positive correlations between genetic diversity values and the average number of insertions per accession for both Tnt1 and T135 in *Capsicum*. In contrast, in an earlier study, no significant correlation was detected between genetic diversity values and retrotransposon abundance in *Solanum* subsection *Lycopersicon* (Tam et al. 2007b). Comparative genomic studies have found significant correlation between retrotransposon abundance and genome size. In *Gossypium,* the major fraction of genome size variation observed within the genus was largely attributed to recent, lineage-specific amplification of one particular group of gypsy-like retrotransposon sequences, Gorge3 within the larger genome *Gossypium* species (Hawkins et al. 2006). For the legume *Vicia pannonica,* it was reported that the significant expansion of its genome size was mediated by the amplification of Ogre, a giant 25 kb Ty3/gypsy element which alone made up 38% of the genome of this species (Neumann et al. 2006). The 2C DNA content of pepper is 2–4-fold greater than that of tomato, but both species have similar chromosome number and gene content (Arumuganathan and Earle 1991). In their study of genome mapping in *Capsicum*, Livingstone et al. (1999) suggested that an increase in the size of pepper genome without apparent increase in gene content involved the expansion of heterochromatin and the increase in the amount of repeated DNA such as retrotransposons. The observed (weak) positive relationship between higher genetic diversity values (reflective of element activity) and increase in the number of insertions (copy number) together with the recorded high percentage of unique (species-specific) insertions (average 12.5%) provide evidence for the active contribution of retrotransposon activities in generating a significant proportion of natural genetic diversity in the genome of *Capsicum.* It further strongly suggests the possibility that retrotransposons might have contributed to genome evolution (expansion) of *Capsicum.*

Acknowledgments We are extremely grateful to Christian Parisod, Miki Okada and Kietsuda Leungwilai for their helpful comments and suggestions. This study was funded by the FP6 European Commission EC-QLRT-1999-31502.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:211–215

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 95:45–90

Bennetzen JL (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica 115:29–36

Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. Curr Opin Genet Dev 15:621–627

Berenyi MS, Gichuki T, Schmidt J, Burg K (2002) Ty1-copia retrotransposon-based S-SAP (sequence-specific amplified polymorphism) for genetic analysis of sweet potato. Theor Appl Genet 105:862–869

Biemont C, Vieira C (2006) Junk DNA as an evolutionary force. Nature 443:521–524

Costa AP, Scortecci KC, Hashimoto RY, Araujo PG, Grandbastien M-A, Van Sluys M-A (1999) Retrolyc1–1, a member of the Tntl retrotransposon super-family in the *Lycopersicon peruvianum* genome. Genetica 107:65–72

Djian-Caporalino C, Lefebvre V, Sage-Daubèze AM, Palloix A (2006) *Capsicum*. In: Singh RJ (ed) Genetic resources, chromosome engineering, and crop improvement series. CRC Press, Florida, pp 185–243

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Eshbaugh WH, Guttman SL, McLeod M (1983) The origin and evolution of domesticated *Capsicum* species. Ethnobiology 3:49–54

Falush D, Stephens D, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578

Geleta LF, Labuschagne MT, Viljoen CD (2005) Genetic variability in pepper (*Capsicum annuum* L.) estimated by morphological data and amplified fragment length polymorphism markers. Biodivers Conserv 14:2361–2375

Grandbastien M-A, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 26:376–380

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of Ty1-copia group retrotransposons in cereal genomes. Mol Gen Genet 261:883–891

Hammer K, Arrowsmith N, Gladis T (2003) Agrobiodiversity with emphasis on plant genetic resources. Naturwissenschaften 90:241–250

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res 16:1252–1261

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

IPGRI, AVRDC, CATIE (1995) Descriptors for *Capsicum* (*Capsicum* spp.). International Plant Genetic Resources Institute, Rome, Italy; the Asian Vegetable Research and Development Center, Taipei, Taiwan; and the Centro Agronomico Tropical de Investigacion y Ensenanza, Turrialba, Costa Rica

Kalendar R, Schulmann AH (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nat Protoc 1:2478–2484

Kalendar R, Grob T, Regina M, Suomeni A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98:704–711

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 33:102–106

Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) Cladistics: the theory and practice of parsimony analysis, 2nd edn. Oxford University Press, Oxford

Kwon YS, Lee JM, Yi GB, Yi SI, Kim KM, Soh EH, Bae KM, Park EK, Song IH, Kim BD (2005) Use of SSR markers to complement tests of distinctiveness, uniformity, and stability (DUS) of pepper (*Capsicum annuum* L.) varieties. Mol Cell 19:428–435

Lanteri S, Acquadro A, Quagliotti L, Portis E (2003) RAPD and AFLP assessment of genetic variation in a landrace of pepper (*Capsicum annuum* L.) grown in North-West Italy. Genet Resour Crop Evol 50:723–735

Lefebvre V (2005) Molecular markers for genetics and breeding: development and use in pepper (*Capsicum* spp.). In: Lörz H, Wenzel G (eds) Molecular marker systems in plant breeding and crop improvement biotechnology in agriculture and forestry. Springer, Berlin, pp 189–214

Lefebvre V, Palloix A, Rives M (1993) Nuclear RFLP between pepper cultivars (*Capsicum annuum* L.). Euphytica 71(18):9–199

Lefebvre V, Goffinet B, Chauvet JC, Caromel B, Signoret P, Brand R, Palloix A (2001) Evaluation of genetic distances between pepper inbred lines for cultivar protection purposes: comparison of AFLP, RAPD and phenotypic data. Theor Appl Genet 102:741–750

Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH (2003) Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques. Mol Genet Genomics 269:464–474

Liedloff A (1999) Mantel nonparametric test calculator, version 2.0. Queensland University of Technology, Brisbane

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. Genetics 152:1183–1202

Manetti M, Rossi M, Costa A, Clausen A, Van Sluys M-A (2007) Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera. BMC Evol Biol 7

McLeod MJ, Guttman SI, Eshbaugh WH (1982) Early evolution of chili peppers (*Capsicum*). Econ Bot 36:361–368

Melayah D, Lim KY, Bonnivard E, Chalhoub B, Dorlhac de Borne F, Mhiri C, Leitch AR, Grandbastien M-A (2004) Distribution of the Tnt1 retrotransposon family in the amphidiploid tobacco (*Nicotiana tabacum*) and its wild *Nicotiana* relatives. Biol J Linn Soc 82:639–649

Morgan MT (2001) Transposable element number in mixed mating populations. Genet Res 77:261–275

Morrison DA (2005) Networks in phylogenetic analysis: new tools for population biology. Int J Parasitol 35:567–582

Moscone EA, Baranyi M, Ebert I, Greilhuber J, Ehrendorfer F, Hunziker AT (2003) Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry and Feulgen densitometry. Ann Bot 92:21–29

Nagy I, Stagel A, Sasvari Z, Roder M, Ganal M (2007) Development, characterization, and transferability to other Solanaceae of microsatellite markers in pepper (*Capsicum annuum* L.). Genome 50:668–688

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

Neumann P, Koblızkova A, Navratilova A, Macas J (2006) Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173:1047–1056

Onus AN, Pickersgill B (2004) Unilateral incompatibility in *Capsicum* (Solanaceae): occurrence and taxonomic distribution. Ann Bot 94:289–295

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

Park Y-K, Park K-C, Park C-H, Kim N-S (2000) Chromosomal localization and sequence variation of 5S rDNA gene in five *Capsicum* species. Mol Cell 10:18–24

Pearce SR, Knox M, Ellis THN, Flavell AJ, Kumar A (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet 263:898–907

Petit M, Lim KY, Julio E, Poncet C, Dorlhac de Borne F, Kovarik A, Leitch AR, Grandbastien M-A, Mhiri C (2007) Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*). Mol Genet Genomics 278:1–15

Pickersgill B (1997) Genetic resources and breeding of *Capsicum* spp. Euphytica 96:129–133

Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F (2002) Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics 267:107–114

Pritchard JK, Stephens M, Donnelly PJ (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Purvis A, Rambaut A (1995) Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. Comput Appl Biosci 11:247–251

Ribeiro-Carvalho C, Guedes-Pinto H, Igrejas G, Stephenson P, Schwarzacher T, Heslop-Harrison JS (2004) High levels of genetic diversity throughout the range of the Portuguese wheat landrace 'Barbela'. Ann Bot 94:699–705

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

Rogers SA, Pauls KP (2000) Ty1-copia-like retrotransposons of tomato (*Lycopersicon esculentum* Mill.). Genome 43:887–894

Saeidi H, Rahiminejad MR, Heslop-Harrison JS (2008) Retroelement insertional polymorphisms, diversity and phylogeography within diploid, D-genome *Aegilops tauschii* (Triticeae, Poaceae) subtaxa in Iran. Ann Bot 101:855–861

Sage-Palloix AM, Jourdan F, Phaly T, Nemouchi G, Lefebvre V, Palloix A (2007) Analysis of diversity in pepper genetic resources: distribution of horticultural and resistance traits in the INRA pepper germplasm. In: Niemirowicz-Szczytt K (ed) Progress in research on capsicum & eggplant. Warsaw University of Life Sciences Press, Warsaw

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

Sniegowski PD, Charlesworth B (1994) Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. Genetics 137:815–827

Swofford DL (2002) PAUP* 4.0 beta 10. Phylogenetic Analysis using Parsimony (and other methods). Sinauer Associates, Sunderland

Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce S, Grandbastien M-A (2005) Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. Theor Appl Genet 110:819–831

Tam SM, Mhiri C, Grandbastien M-A (2007a) Transposable elements and the analysis of plant biodiversity. In: Morot-Gaudry JF, Lea P, Briat J-F (eds) Functional plant genomics. Science Publishers, Enfield, pp 529–558

Tam SM, Causse M, Garchery C, Burck H, Mhiri C, Grandbastien M-A (2007b) The distribution of copia-type retrotransposons and the evolutionary history of tomato and related wild species. J Evol Biol 20:1056–1072

Tanksley SD (1984) High rates of cross-pollination in Chile pepper. HortScience 19:580–582

Vershinin AV, Allnut TR, Knox MR, Ambrose MJ, Ellis THN (2003) Transposable elements reveal the impact of introgression, rather

than transposition in *Pisum* diversity, evolution and domestication. Mol Biol Evol 20:2067–2075

Walsh BM, Hoot S (2001) Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast atpB-rbcL spacer region and nuclear waxy introns. Int J Plant Sci 162:1409–1418

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence specific amplification polymorphisms (SSAP). Mol Gen Genet 253:687–694

Wright SI, Schoen DJ (1999) Transposon dynamics and the breeding system. Genetica 107:139–148

Wright SI, Agrawal N, Bureau TE (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res 13:1897–1903

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157

Yu GX, Wise RP (2000) An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome 43:736–749