HAL
open science

Comparative microsatellite linkage analysis and genetic structure of two populations of F6 lines derived from Lycopersicon pimpinellifolium and L. cheesmanii

Irene Villalta, A. Reina-Sánchez, J. Cuartero, E. Carbonell, M. J Asins

To cite this version:

Irene Villalta, A. Reina-Sánchez, J. Cuartero, E. Carbonell, M. J Asins. Comparative microsatellite linkage analysis and genetic structure of two populations of F6 lines derived from Lycopersicon pimpinellifolium and L. cheesmanii. TAG Theoretical and Applied Genetics, 2005, 110 (5), pp.881-894. 10.1007/s00122-004-1906-3 . hal-02140672

HAL Id: hal-02140672

https://hal.science/hal-02140672v1

Submitted on 8 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Theor Appl Genet (2005) 110: 881–894
DOI 10.1007/s00122-004-1906-3

ORIGINAL PAPER

I. Villalta · A. Reina-Sánchez · J. Cuartero
E. A. Carbonell · M. J. Asins

Comparative microsatellite linkage analysis and genetic structure of two populations of F_6 lines derived from *Lycopersicon pimpinellifolium* and *L. cheesmanii*

Received: 30 September 2004 / Accepted: 11 December 2004 / Published online: 3 February 2005
© Springer-Verlag 2005

Abstract A population of recombinant inbred lines (RILs) has several advantages over its F_2 population counterpart with respect to quantitative trait loci (QTLs) and genomic studies. The objective of the investigation reported here was the comparative characterization by simple sequence repeat (SSR) and sequence characterized amplified region (SCAR) markers of two populations of F_6 lines derived from *Lycopersicon pimpinellifolium* (P population, consisting of 142 lines) and *L. cheesmanii* (C population, consisting of 115 lines) and sharing the female parent, *L. esculentum* var. *cerasiforme*. Almost the same percentage of polymorphic markers was found for each population although a different set of markers was involved. The proportion of SSR primer pairs (93 in total) that resulted in polymorphism for the main band was larger (55–56%) than for SCAR ones (13–16%). The C population showed the largest proportion of markers with zygotic and gametic segregation distortion, which is in agreement with the larger genetic distance reported between *L. esculentum* and *L. cheesmanii* than with the former and *L. pimpinellifolium*. Zygotic distortion corresponded primarily to an excess of heterozygotes in both populations, suggesting that the increment of homozygosity was the main factor limiting viability/self-fertility of the lines. Despite both populations sharing the female parent, *P* alleles were slightly favored in the P population while *E* alleles were the most frequently fixed in the C population. A linkage map for each population was obtained, with the average distances between consecutive markers being 3.8 cM or 3.4 cM depending on the population. Discrepancy between the maps for the location of only four markers on chromosomes 3, 6 and 10 was observed. Two possible causes of this discrepancy were investigated and can not be discarded: (1) the presence of duplicated markers and (2) segregation distortion caused by the selective advantage of gametes carrying one of the two alleles. This marker characterization of both populations will continue and will enable the comparative QTLs and candidate gene analysis of complex traits towards a more efficient utilization of genetic resources and breeding strategies.

Communicated by I. Paran

I. Villalta · E. A. Carbonell · M. J. Asins (✉)
Instituto Valenciano de Investigaciones Agrarias,
Apdo. Oficial, 46113 Moncada-Valencia, Spain
E-mail: mjasins@ivia.es
Tel.: +34-963424067
Fax: 34-96-3424001

A. Reina-Sánchez · J. Cuartero
Consejo Superior de Investigaciones Científicas-Estación
Experimental La Mayora, 29750 Algarrobo-Costa, Spain

Introduction

Tomato (*Lycopersicon esculentum* Mill.) is one of the most important horticultural crops in the world. In terms of human health, the tomato fruit is a major component of daily meals in many countries and constitutes an important source of minerals, vitamins and antioxidant compounds. Carotenoids, in addition to their role in fruit coloring, are an excellent source of vitamin A and antioxidant agents and thus play an important role in preventing cancer and heart diseases (Krinsky 1992). Flavonoids (diphenylpropanes) are antimutagenic and anticarcinogenic (Verma et al. 1988; Francis et al. 1989). In addition to its economical and nutritional importance, the tomato is an ideal research material for physiological, cellular, biochemical and molecular genetic or genomic investigations. It is easy to cultivate, has a short life cycle and is amenable to varied horticultural manipulations, including grafting or cutting, and to genetic transformation. A large number of genes have been described and assigned to specific locations on the 12 chromosomes of its genome, and numerous unigenic mutants are available for a better knowledge of their biochemical and phenotypic effect(s). A vast array of genetic diversity is also available for the cultivated tomato and related taxa from several

germplasm banks. The presence of co-adapted genic complexes in their accessions is a key point that supports the need for germplasm conservation (Brown 1978). Epistatic interactions among loci at two-locus, three-locus and higher order levels have often been shown to produce major effects on adaptability, especially in autogamous species, and have a considerable influence on phenotype (Allard 1988; Pérez de la Vega et al. 1994; Lukens and Doebley 1999). How can we use the continuously growing knowledge on tomato genomics to understand and exploit for plant breeding those co-adapted genic complexes? The possibility to study genotype-by-environment and epistatic interactions through quantitative trait locus (QTL) analysis is envisaged as a valuable tool for this purpose.

The construction of reliable linkage maps based on segregant analysis of co-dominant markers in experimental and breeding populations is a basic requirement for successful comparative studies of the genetic control of traits. Most breeding efforts are centered around locating genes or QTLs conferring resistance or tolerance to biotic and abiotic stress factors. With this aim, breeders have developed many experimental populations derived from interspecific crosses using *L. esculentum* as the receptor parent and various wild relatives as parental donors (Table 1): *L. pimpinellifolium* (Monforte et al. 1996; Tanksley et al. 1996; Chen and Foolad 1999), *L. cheesmanii* (Paran et al. 1995; Monforte et al. 1999), *L. pennellii* (Tanksley et al. 1992; Eshed and Zamir 1995; Haanstra et al.1999), *L. hirsutum* (Monforte and Tanksley 2000), *L. peruvianum* (Fulton et al. 1997) and *L. parviflorum* (Fulton et al. 2000).

Recombinant inbred lines (RILs) have many advantages over other populations that are used for genetic mapping and QTL analysis. Because the genotype is fixed for each line, the whole population can be distributed and replicated for use in experiments in different laboratories and environments. This feature is crucial for quantifying the effect of genotype × environment (G×E) interaction within a QTL analysis of agronomic traits, especially those related to the adaptation or tolerance to abiotic stress factors, such as salinity (Monforte et al. 1997a). A RIL population is more efficient than its F_2 population because fewer individuals are needed to detect linkage of the same magnitude between a marker and a QTL (Simpson 1989), it is genotyped once and as many traits or molecules can be genetically analyzed as needed. In tomato, most experimental populations have been obtained using the advanced backcross design (Table 1), which is not a useful design for detecting epistatic QTLs (Tanksley and Nelson 1996) since every backcross generation greatly reduces the number of genotypic combinations. RILs have been used extensively for genetic mapping in several plants species and in *Arabidopsis* and rice in particular. In tomato, two studies have been reported but just one from an interspecific cross (Paran et al. 1995). Because RILs undergo several rounds of meiosis, the chance that a recombination event will occur between linked loci is greater than in a single-meiosis population of the same size, thereby yielding higher resolution maps. Therefore, RILs enable investigators to obtain more dense maps. For these reasons, the development of this kind of population has been pursued as the first step of the investigation reported here.

Since co-dominant markers are preferred for future QTLs studies and the level of polymorphism for restriction fragment length polymorphism (RFLP) is low, simple sequence repeats (SSRs) or microsatellites are becoming the preferred molecular markers of choice in crop breeding. Due to their properties of co-dominance, high reproducibility and multiallelic variation, they are the most practical markers for genomic mapping and marker-assisted selection. In tomato, many microsatellite markers have been developed (Smulders et al. 1997; Bredemeijer et al. 1998; Areshchenkova and Ganal 1999; He et al. 2003), but only a limited number of SSR markers have been mapped (Areshchenkova and Ganal 2002). We have employed SSR markers, and SCARs (sequence characterized amplified regions) from previously mapped RFLPs to comparatively characterize two populations of F_6 lines derived from two crosses in which *L. pimpinellifolium* and *L. cheesmanii* were the respective male parent and *L. esculentum* var. *cerasiforme* was the common female parent. The choice of parental lines was based on their contrasting salt tolerance and lack of domestication to minimize segregation distortion and fertility problems during the development of RILs.

Materials and methods

Plant material

Two populations of F_6 lines were developed from crosses of *Lycopersicon esculentum* var. *cerasiforme* (line E9) as the female parent with *L. pimpinellifolium* (line L5, P population) and *L. cheesmanii* (line L3, C population), respectively, as male parents. Both populations were developed by single seed descent from 300 (*L. pimpinellifolium*) or 400 F_2 (*L. cheesmanii*) individual plants (Monforte et al. 1997a), with no conscious selection at any generation, under greenhouse or screenhouse conditions. One hundred and forty-two P lines and 115 C lines were obtained at F_6 by the IVIA (Instituto Valenciano de Investigaciones Agrarias).

Marker analysis

DNA was extracted from a bulk of six plants per F_7 line. This bulk reconstituted the genotype of F_6 plants from which seeds were obtained for the subsequent generation. Every F_6 line from population P was genotyped at 140 marker loci (87 co-dominant, 53 dominant) using 15 SCAR and 64 SSR primer pairs. F_6 lines from the C population were genotyped at 117 marker loci (74 co-dominant, 43 dominant) using 14 SCAR and 48 SSR primer pairs.

TG markers are SCARs derived by primer design from the sequences of TG clones (Table 2). These clones were

Table 1 Some of the most important *Lycopersicon* interspecific linkage maps already published

Interspecific cross	Mapping population	n[a]	Markers[b]	Marker type[c]	Esculentum[d]	Wild[e]	H[f]	Percentage segregation distortion[g]	Author
Lycopersicon esculentum cv. NC84173×*L. hirsutum* P1126445	BC_1	145	171	RFLP, SCAR	1*, 3, 4, 5, 6*, 8, 10, 11, 12	–	–	62	Zhang et al. (2002)
L. esculentum cv.E6203× *L. hirsutum* LA 1777	BC_1	149	135	RFLP	1*, 6*, 11, 12	–	1,.2,.3,.10	18 $P<0.01$	Bernacchi and Tanksley (1997)
L. esculentum cv.E6206× *L. hirsutum* LA 1777	BC-recombinant inbred line	111	95	RFLP	1*, 2, 3, 6*, 7, 8, 9, 10, 12	–	2, 3, 4, 7, 8, 9, 10	n.s.	Monforte and Tanksley (2000)
L. esculentum cv. E6203× *L. parviflorum* LA2133	BC_2	170	133	RFLP, SCAR, morphological	4, 3, 6*, 8, 9, 10, 11, 12	–	6, 8, 10	25 $P<0.001$	Fulton et al. (2000)
L. esculentum cv. E6203× *L. pennellii* LA 1657	BC_2	175	110	RFLP	1*,2*, 4, 5, 6*, 7, 11*	–	7, 10, 12	28 $P<0.001$	Frary et al. (2004)
L. esculentum cv. VF36× *L. pennellii* LA716	F_2	67	1030	Isozyme, morphological, RFLP	n.s.	n.s.	n.s.	n.s.	Tanksley et al. (1992)
L. esculentum cv. Vendor× *L. pennellii* LA716	F_2	432	98	RFLP	4, 5, 7	All except for four	1, 2 4, 5, 7, 8, 9, 12	80 $P<0.05$	deVicente and Tanksley (1993)
L. esculentum cv. E6203× *L. peruvianum* LA1708	BC_3	241	174	RFLP, SCAR	1*, 6*, 3, 5, 10	–	5, 12, 9	n.s.	Fulton et al. (1997)
L. esculentum cv. UC82B× *L. chmielewskii* LA1063	BC_1	237	70	RFLP	2, 5, 7	–	2	n.s.	Paterson et al. (1990)
L. esculentum cv. M 821-7× *L. pimpinellifolium* LA 1589	BC_1	257	120	RFLP, RAPD, morphological	5	–	1, 8, 11	8.3 $P<0.05$	Grandillo and Tanksley (1996)
L. esculentum cv. NC84173× *L. pimpinellifolium* LA722	BC_1	119	151	RFLP	6, 7, 11	–	1, 2, 3, 10	9.9 $P<0.05$	Chen and Foolard (1999)
L. esculentum cv. Yellow Stuffer× *L. pimpinellifolium* LA1589	F_2	200	90	RFLP, CAPs	–	7, 9, 10, 11	n.s.	n.s.	Lippman and Tanksley (2001)
L. esculentum cv. UC204B and *L. cheesmanii* LA483	F_2	350	71	RFLP	n.s.	3	2.3	51 $P<0.05$	Paterson et al. (1991)
L. esculentum cv. UC204× *L. cheesmaniii* LA483	F_7	97	132	RFLP	2, 3, 4, 10, 11	3, 6, 7	n.s.	73 $P<0.05$	Paran et al. (1995)
L. esculentum var. *cerasiforme* E9×*L. cheesmanii* L3	F_6	115	114	SCAR, microsatellites	1, 2, 3, 8, 9, 10	4, 6, 12	All	52 $P<0.001$	Present study
L. esculentum var. *cerasiforme* E9×*L. pimpinellifolium* L *5*	F_6	142	132	SCAR, microsatellites	7, 8, 11	2, 4, 10	All	29 $P<0.001$	Present study

[a]Family size
[b]Number of markers used in the map
[c]RFLP, Restriction fragment length polymorphism; SCAR, sequence-characterized amplified region; CAPS, cleaved amplified polymorphic sequence; RAPD, random amplified polymorphic DNA
[d]Chromosomes presenting distortion towards the *E* alleles; n.s., not specified
[e]Chromosomes presenting distortion towards the wild alleles
[f]Chromosomes displaying an excess of heterozygotes
[g]n.s., Not specified
*Segregation distortion caused by artificial selection

Table 2 Primers for markers that are not available at the SGN databases

Marker	Author's code[a]	Forward primer (5′–3′)	Reverse primer (5′–3′)	T_m[b]	Size of the polymorphism (bp)[c]
SSR1	LE20592 (1)	CTGTTTACTTCAAGAAGGCTG	ACTTTAACTTTATTATTGCCACG	45	175 (p,c), 800 (p)
SSR10	LECBPE3 (1)	CCTACAAAAACTGCCTCT	TTATATCAATACAACAACATT	45	130 (p,c)
SSR11	LENIA (1)	TTAAGATTGTATTCATCATGG	CTTTAGGCTTGTAATGGAGTG	45	900 (p)
SSR12	LEGAST1 (1)	ATCTCTATTGTTTTCGACTCG	TCTGTTGTTGCTGCTGCTC	45	140 (p,c), 750 (p,c), 830 (c)
SSR13	LECHSOD (1)	TTATCAATTCATCATTGTGGC	AGGGGTAGTGACAGCATAAAG	45	290 (c), 298 (p), 466 (p), 600 (p)
SSR14	LEGTOM5 (1)	AAAGATAAAGCATGAAATGAA	GGAGTTGAGATAAAGTGAAGA	45	180 (p,c)
SSR16	LELEUZIP (1)	GGTGATAATTTGGGAGGTTAC	CGTAACAGGATGTGCTATAGG	45	396 (p), 600 (p)
SSR17	LEMDDNb (1)	TAAATACAAAAGCAGGAGTCG	GAGTTGACAGATCCTTCAATG	45	320 (p)
SSR19	GA-b (2)	AAGCCTAGACCGTGTGTCATG	TGTAAGTTTCCATCTCCAGCC	45	190 (p), 900 (p)
SSR20	GATA500(3)	ACCCTTGTGTCAGTTCCAGTA	GATCCATGTTTATCCCCTCTA	45	340 (c), 520 (p), 700 (p)
SSR21	GATA450(3)	GACATAGGTATGGGACAAAAC	GTTCGTGCTATTTGATTATTC	45	900 (p,c)
SSR22	GA-c (2)	CGTTCCATACCTTCCAGATAGTC	CATTTCAGAAGTCGGCTGGTCAG	45	200 (p), 325 (p), 344 (p), 517 (c), 530 (c), 950 (p), 1,018 (p,c)
SSR23	LPHFS24 (1)	TTGGATTTACAAGTTCGATGT	GCATTTGACTTGATAGCAGTC	55	396 (p), 600 (p)
SSR24	ATT-a (2)	AGCTGCTTGGTTTGTATTGAC	GTTTCTCTCATTCCACACAGC	45	250 (p,c), 510 (c), 750 (p), 850 (p,c)
SSR26	LEDIH4RE (1)	TTTTGTAATCATCTTGGAAAC	ATTGTGTTATGATGATATTTG	45	344 (p)
SSR27	LESODB (1)	TTATCAATTCATCATTGTGGC	AGTAAGGGGTTTAGGGGTAGT	55	310 (p), 480 (p), 600 (c), 1,015 (c)
SSR29	LELE25 (1)	TTCTTCCGTATGAGTGAGT	CTCTATTACTTATTATTATCG	55	250 (p,c)
SSR3	LECAB9 (1)	TTTATTATCCCAGAAGCCTTC	CCTCACATTTAAACAAATTGC	45	220 (p)
SSR30	GATA332 (3)	CCTACGTACCTACCCATGT	ACATACAAACAGAGAGACAAA	45	300 (p), 375 (c), 450 (p), 520 (p)
SSR31	LE21085 (1)	CATTTTATCATTTATTTGTGTCTTG	ACAAAAAAAGGTGACGATACA	45	130 (p,c)
SSR33	LE2A11 (1)	AATTTTGTAAGGAGAAGACGG	TCATATTCTTCACACCAAAGG	45	396 (c), 900 (c), 1,600 (c)
SSR34	LEHSC80P (1)	CCTGATTAAGAAGCACTCTGA	CACTCATTGGAAACTTCTTTG	45	750 (p)
SSR35	LEMDDNa (1)	ATTCAAGGAACTTTTAGCTCC	TGCATTAAGGTTCATAAATGA	45	298 (p,c), 420 (p)
SSR36	LESSF (1)	TACGCTCTCAAGTACCGTAAG	CCTACATTGACATGACCAAAT	55	320 (p)
SSR37	LPTRYINH (1)	AAGTTTGCTCACATCATTCTG	TAAAAGTTCTTCTCCCTCACC	45	530 (p), 1,200 (p)
SSR38	LEILV1B (1)	GATCGACACATTTGAATTTGT	GGTCACTAATTAATTGATTCC	45	154 (p)
SSR41	LEATPACAb (1)	GTATGTCAAATCTCTCTTGCG	ACTCTCCATCGTCTCTTTCAC	45	200 (p,c), 800 (p)
SSR42	LEE8 (1)	TCTTTAGTAGCTCAGTGGCAG	GGCCAACTAAATCGTTTATTC	55	180 (p,c)
SSR5	LEEF1Aa (1)	AAATAATTAGCTTGCCAATTG	CTGAAAGCAGCAACAGTATTT	50	220 (p)
SSR6	LELAT59G (1)	AAAAGGGGTATGAACATTAGG	GCATCTATCGTCTTGTCACTC	45	180 (p,c)
SSR9	LERBCS3B (1)	AAACCTTGACATTACCTCCAT	AGGAAGGTACGACAGAGTCTC	45	220 (p)
CD40	CD40 (4)	TCTGAAGCCAAATGCAGAC	GCCATTGCCATCAAGATACC	50	530 (c), 700 (p,c)
CT118	CT118 (4)	TTGAATTCCAAGCTAAACCCC	TAACCAAACTCGTGTCCCC	45	330 (c), 470 (c)
CT141	CT141 (4)	GCCTTCCTTTTTCCCCAAC	GCAAGCACAGCATCTCTATC	45	320 (p,c)
CT143	CT143 (4)	AGCTAAGCAATCAGTTACAACC	AGATGTGCAGAGAAGTACAAAG	45	320 (p)
CT149b	CT149 (4)	AAAAATTGCTACAAACCGGC	TTTGTATGAGTTTGACGGAGAG	45	250 (c)
CT156	CT156 (4)	TAAGGCTGCCAGCTCGAATG	GCAGCAATAGATGGCTTGGAG	45	344 (p), 365 (c), 770 (c), 850 (p,c), 1,200 (p,c)
CT167	CT167 (4)	TCCCATTAGCCATAGCAGCC	TCCTCCCACCATTCCAATTTTC	45	340 (p), 530 (p), 650 (c), 700 (p), 800 (p)
CT176	CT176 (4)	AACAAGCAGTGGTTGATGAG	CAACATCTTCCCATACTTCCC	50	280 (p), 296 (p)
CT206	CT206 (4)	TTCAACAGCACCAACAGTC	AAGTCAACCAGCTACTGCC	45	290 (c), 400 (c), 870 (c)1,636 (c)
CT211	CT211 (4)	AATTTTGAGGGGTGGGGTG	GGCTTGTTGAACTTGAAGCTG	45	530 (c), 650 (c)
CT233	CT233 (4)	CTTTCTACCTTCTGAGCCAAC	TCCTCCGATGTAACTATGACG	45	1,700 (p)
CT234	CT234 (4)	CCACAAGAAGACCCCCAAAC	CATCTCCCTACAACACTCTCC	45	320 (p)
CT283	CT283 (4)	GTGAAAGTTCATCAATGGCG	AGTCTAGCAGCTCCTTCAG	45	290 (p), 500 (p), 700 (p), 830 (p), 900 (p)

TG134	TG134 (4)	AATGTTACTGTAGGCGAATG	TCTAGGAGGGTTAAGAGGG	45	118 (p)
TG15	TG15 (4)	GAGATTCATACTCACTGACCAC	GGTCAACTAGTCCAAACCTC	45	480 (c), 900 (c)
TG16	TG16 (4)	TAGATGTGATGGTGGCAG	GATATCCGTGCAAGCAAG	50	380 (p), 440 (c), 507 (c), 530 (p,c), 800 (p,c), 1,000 (p)
TG30	TG30 (4)	ACCGAACTTGAATGAACG	GGATTGACTTGGTATGTGG	45	320 (p,c)
TG35	TG35 (4)	GCACCCAAGAGTGATGTAG	TGACCTTCGTATAATTCCAC	45	506 (c), 517 (c)
TG43	TG43 (4)	ACAGAGTTAGAGAATCACATCC	AAGTAAGCTGTCTGCAACTC	45	600 (c), 750 (p,c), 1,300 (p)
TG48	TG48 (4)	GTGTAATTGGATGAACCAAC	GGCAGCTTGATAGGAATAAG	45	340 (c), 350 (c), 510 (c)
TG69	TG69 (4)	GGATTTGGGGATAGAGGGAC	TCGGACGAGGACAAAGAGAG	55	600 (p,c), 620 (c), 950 (p,c), 1,636 (p)

[a] (1), Smulders et al. (1997); (2), Broun and Tanksley (1996); (3), Vosman and Arens (1997); (4), primers derived from RFLP markers (see Materials and methods

[b] Annealing temperatures (°C) within the amplification profile

[c] (p) and (c) indicate polymorphic markers in the P and/or C family, respectively

kindly provided by Dr. S. Tanksley. CT and CD markers (Table 2) are also SCARs and these were obtained by primer design using cDNA sequences (Ganal et al. 1998) available in the National Center of Biotechnology Information website database (http://www.ncbi.nlm.nih.gov). The PRIME program of the University of Wisconsin Genetics Computer Group (GCG) software package was used for primer design. The PCR products we obtained did not always correspond to the expected length due to the presence of introns in the genomic DNA sequence (Table 2). Additionally, more than one main band was usually obtained due to the amplification of secondary bands. For these reasons their location cannot be assumed to be the same as that reported by Tanksley et al. (1992). We decided not to modify the conditions nor eliminate these additional loci, given that the extra products did not interfere with the main one with respect to genetic interpretation and contributed more polymorphic markers to be included in the genetic maps. SSRW markers correspond to SSR primers available from the International Solanaceae Genomics Network (SGN) website database (http://www.sgn.cornell.edu/) and were chosen based on their map location. Other SSR primers were obtained from different authors and are also summarized in Table 2. Nomenclature used for all marker loci consists of primer name and the size of segregating band in base pairs.

PCR amplification was performed as described by Kijas et al. (1997) with minor modifications, using 300 ng of genomic DNA per 20-µl reaction. Each reaction was overlaid with 25 µl of mineral oil and amplified in a PTC-100 thermal cycler (MJ Research, Waltham, Mass.) under the following conditions; an initial step at 94°C for 3 min; 30 cycles of 1 min at 94°C, 45 s at 45–55°C (depending on the marker, see Table 2) and 1 min and 45 s at 72°C; final steps of 45 s at 55°C and 3 min at 72°C. PCR products were mixed with 5 µl of 5× loading buffer [50% (v/v) glycerol, 1× AE, 10% (v/v) saturated bromophenol blue, 0.2% (w/v) xylene cyanole] and analyzed by electrophoresis in sequencing-type 10% polyacrylamide gels (acrylamide: *N*, *N′*-methylene bisacrylamide, 29:1) under non-denaturing conditions in 1× TBE buffer (90 m *M* Tris borate, 2 m *M*EDTA, pH 8.3). The procedures used for electrophoresis conditions and silver staining are described by Ruíz et al. (2000).

Linkage analysis

Genotype data of both populations were used independently to perform linkage analyses using JOINMAP 3.0 software for Windows (Van Ooijen and Voorrips 2001). A minimum LOD of 3 was set as a threshold to allocate marker loci into linkage groups, and a recombination fraction of 0.5 was used for linkage analysis. The Kosambi function (Kosambi 1944) was used to order markers and to estimate interval distances. Segregation distortion at each marker locus was checked for deviation from the expected F_6 ratio based on a chi-square goodness-of-fit test. Additional chi-square tests were calculated to study gametic

segregation distortion. Linkage between groups with distorted segregation ratios was confirmed using a chi-square (Mather 1957) for the independence of two segregations, conditional on their marginal frequencies.

Cloning and sequencing of problematic markers

In order to investigate any possible duplications of markers presenting high values of segregation distortion within non-distorted chromosomal regions (Frisch et al. 2004), we cloned and sequenced some of these problematic marker bands.

PCR reactions were cloned into the p-GEM- easy vector system (Promega, Madison, Wis.). To evaluate the diversity of inserts from white colonies, we performed PCR reactions using 2 μl of bacterial culture from selected colonies and analyzed these by means of 10% polyacrylamide sequencing type gel electrophoresis. Clones of the expected length were purified using the Concert nucleic acid purification system (Life Technologies, Gaithersburg, Md.). Both strands of the selected clones were sequenced by the IBMCP (Instituto de Biología Molecular y Celular de Plantas, Valencia) Sequencing Service. Sequence analysis and alignment were performed using the SEQUENCHER [Gene Codes Corporation (GCG), Ann Arbor, Mich.] computer program.

Homology searches were done using the online service of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast) and the Solanaceae Genomics Network database (http://www.sgn.cornell.edu/cgi-bin/tools/blast/simple.pl), using the BLASTN search tool (Altschul et al. 1997).

The GCG and SEQUENCHER programs were accessed through the Bioinformatics Service of the University of Valencia.

Results

Marker polymorphism

Preliminary tests were carried out in order to determine the degree of polymorphism between the parents for the two kinds of markers. The percentage of polymorphic markers for SSRs and SCARs was exactly the same for each population, but with different sets of markers. Of 93 SSR primer pairs (79.5%), 74 were polymorphic in the P and C populations. However, when polymorphism was considered only for the main band, the level of polymorphism decreased to 55% and 56% for the P and C populations, respectively. Ultimately, 64 and 50 SSR primer pairs were actually used to genotype the F_6 lines from the P and C populations, respectively. Each of these primer pairs revealed an average of 1.6 polymorphic loci. Of the 42 SCAR primer pairs tested, 23 (54%) were polymorphic in the P and C populations. If polymorphism was considered only for the main amplified band, the level of polymorphism decreased once again to 13% and 16% for the P and C populations, respectively. Of the 23 SCAR polymorphic markers, 15 and 14 were ultimately selected to genotype the P and C populations, respectively. Each of these primer pairs detected an average of 2.3 polymorphic loci.

New alleles (not present in parents) appeared in the P population (Fig. 1) and always involved the same F_6 lines: lines 137 and 189 for markers SSRW63_240 and SSRW344_290 (chromosome 8), SSRW241_200 and SSRW285_290 (chromosome 7) and line 137 for SSRW383_270 (chromosome 9), CD40_700 (chromosome 8) and SSRW450_310 (chromosome 4). These markers were eliminated from these lines during construction of the linkage map. Although apparently the same new allele has been fixed for the common marker loci in both lines, if these marker loci are ignored, the similarity between lines 137 and 189 is 0.329 while the mode of the population corresponds to the 0.4–0.5 class. In fact, the mean fruit weight of these lines was very different, 2.85 g for line 137 and 16.97 g for line 189. The presence of foreign pollen during the development of the lines is a highly unlikely possibility, and the chance that this pollen affected only these loci, in the same lines (they were never contiguous for multiplication), is also very unlikely. These observations negate the hypothesis of a common origin of both lines after the F_2 plants they derive from and support

Fig. 2 Linkage maps obtained for the P (*L. esculentum* × *L. pimpinellifolium*) and C (*L. esculentum* × *L. cheesmanii*) populations. Common markers are connected by *lines*, if the order of marker has changed a *discontinuous line* is drawn. *Framed* linkage groups are those where no common markers with other linkage group has been found. Markers displaying significant ($P < 0.001$) genotypic segregation distortion are shown in *bold*. *Underlined* markers show significant ($P < 0.001$) gametic distortion

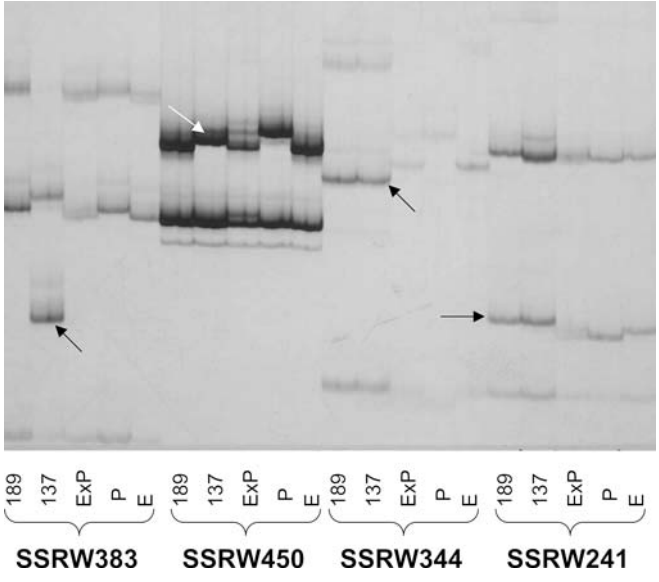
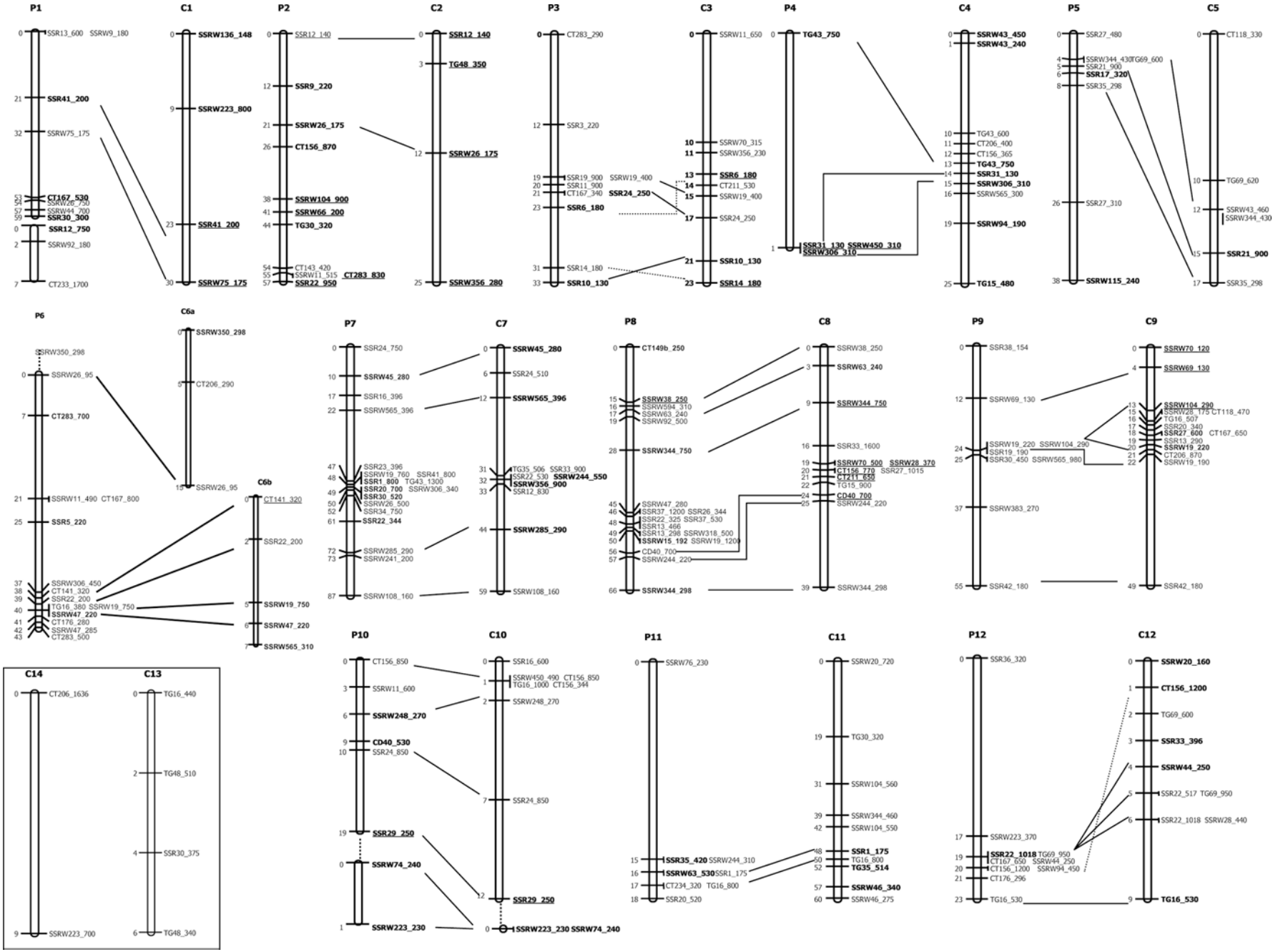


Fig. 1 Examples of markers showing new SSR alleles not present in the parents as revealed by silver-stained polyacrylamide gel electrophoresis in P lines 137 and 189. *E* × P Interspecific hybrid between the E and P parental species, *arrows* new bands


P1
0 SSR13_600 SSRW9_180
21 SSR41_200
32 SSRW75_175
53 CT167_530
54 SSRW26_750
57 SSRW44_700
59 SSR30_300
0 SSR12_750
2 SSRW92_180
7 CT233_1700
C1
0 SSRW136_148
9 SSRW223_800
23 SSR41_200
30 SSRW75_175
P2
0 SSR12_140
12 SSR9_220
21 SSRW26_175
26 CT156_870
38 SSRW104_900
41 SSRW66_200
44 TG30_320
54 CT143_420
55 SSRW11_515 CT283_830
57 SSR22_950
C2
0 SSR12_140
3 TG48_350
12 SSRW26_175
25 SSRW356_280
P3
0 CT283_290
12 SSR3_220
19 SSR19_900 SSRW19_400
20 SSR11_900
21 CT167_340 SSR24_250
23 SSR6_180
31 SSR14_180
33 SSR10_130
C3
0 SSRW11_650
10 SSRW70_315
11 SSRW356_230
13 SSR6_180
14 CT211_530
15 SSRW19_400
17 SSR24_250
21 SSR10_130
23 SSR14_180
P4
0 TG43_750
1 SSR31_130 SSRW450_310
SSRW306_310
C4
0 SSRW43_450
1 SSRW43_240
10 TG43_600
11 CT206_400
12 CT156_365
13 TG43_750
14 SSR31_130
15 SSRW306_310
16 SSRW565_300
19 SSRW94_190
25 TG15_480
P5
0 SSR27_480
4 SSRW344_430 TG69_600
5 SSR21_900
6 SSR17_320
8 SSR35_298
26 SSR27_310
38 SSRW115_240
C5
0 CT118_330
10 TG69_620
12 SSRW43_460
SSRW344_430
15 SSR21_900
17 SSR35_298
P6
SSRW350_298
0 SSRW26_95
7 CT283_700
21 SSRW11_490 CT167_800
25 SSR5_220
37 SSRW306_450
38 CT141_320
39 SSR22_200
40 TG16_380 SSRW19_750
SSRW47_220
41 CT176_280
42 SSRW47_285
43 CT283_500
C6a
0 SSRW350_298
5 CT206_290
15 SSRW26_95
C6b
0 CT141_320
2 SSR22_200
5 SSRW19_750
6 SSRW47_220
7 SSRW565_310
P7
0 SSR24_750
10 SSRW45_280
17 SSR16_396
22 SSRW565_396
47 SSR23_396
48 SSRW19_760 SSR41_800
SSR1_800 TG43_1300
49 SSR20_700 SSRW306_340
SSR30_520
50 SSRW26_500
52 SSR34_750
61 SSR22_344
72 SSRW285_290
73 SSRW241_200
87 SSRW108_160
C7
0 SSRW45_280
6 SSR24_510
12 SSRW565_396
31 TG35_506 SSR33_900
32 SSR22_530 SSRW244_550
SSRW356_900
33 SSR12_830
44 SSRW285_290
59 SSRW108_160
P8
0 CT149b_250
15 SSRW38_250
16 SSRW594_310
17 SSRW63_240
19 SSRW92_500
28 SSRW344_750
45 SSRW47_280
46 SSR37_1200 SSR26_344
48 SSR22_325 SSR37_530
SSR13_466
49 SSR13_298 SSRW318_500
50 SSRW15_192 SSRW19_1200
56 CD40_700
57 SSRW244_220
66 SSRW344_298
C8
0 SSRW38_250
3 SSRW63_240
9 SSRW344_750
16 SSR33_1600
19 SSRW70_500 SSRW28_370
20 CT156_770 SSR27_1015
21 CT211_850
22 TG15_900
24 CD40_700
25 SSRW244_220
39 SSRW344_298
P9
0 SSR38_154
12 SSRW69_130
24 SSRW19_220 SSRW104_290
SSR19_190
25 SSR30_450 SSRW565_980
37 SSRW383_270
55 SSR42_180
C9
0 SSRW70_120
4 SSRW69_130
13 SSRW104_290
15 SSRW28_175 CT118_470
16 TG16_507
17 SSR20_340
18 SSR27_600 CT167_650
19 SSR13_290
20 SSRW19_220
21 CT206_870
22 SSRW19_190
49 SSR42_180
P10
0 CT156_850
3 SSRW11_600
6 SSRW248_270
9 CD40_530
10 SSR24_850
19 SSR29_250
0 SSRW74_240
1 SSRW223_230
C10
0 SSR16_600
1 SSRW450_490 CT156_850
TG16_1000 CT156_344
2 SSRW248_270
7 SSR24_850
12 SSR29_250
0 SSRW223_230 SSRW74_240
P11
0 SSRW76_230
15 SSR35_420 SSRW244_310
16 SSRW63_530 SSR1_175
17 CT234_320 TG16_800
18 SSR20_520
C11
0 SSRW20_720
19 TG30_320
31 SSRW104_560
39 SSRW344_460
42 SSRW104_550
48 SSR1_175
50 TG16_800
52 TG35_514
57 SSRW46_340
60 SSRW46_275
P12
0 SSR36_320
17 SSRW223_370
19 SSR22_1018 TG69_950
CT167_650 SSRW44_250
20 CT156_1200 SSRW94_450
21 CT176_296
23 TG16_530
C12
0 SSRW20_160
1 CT156_1200
2 TG69_600
3 SSR33_396
4 SSRW44_250
5 SSR22_517 TG69_950
6 SSR22_1018 SSRW28_440
9 TG16_530
C14
0 CT206_1636
9 SSRW223_700
C13
0 TG16_440
2 TG48_510
4 SSR30_375
6 TG48_340

the hypothesis of asymmetric recombination at these loci as the source of this new variability.

Linkage maps

The genetic map of the P population (Fig. 2) consists of 132 SSR and SCAR markers distributed over 14 linkage groups (chromosomes 1 and 10 were split in two parts). The average and maximal distances between two adjacent markers were 3.8 cM and 25 cM, respectively. Large gaps (>20 cM) were found on chromosomes 1 (two gaps) and 7 (one gap).

The map of the C population (Fig. 2) contains 114 markers in 16 linkage groups. Chromosomes 6 and 10 were split into two parts, and the chromosomal identity of two linkage groups remains unknown. The average and maximal distances between two individual markers are 3.4 cM and 27 cM, respectively. A large gap (>20 cM) was found only on chromosome 9.

Considerable similarity between the maps from the P and C populations was found. Fourteen homologous linkage groups could be identified based on one to six common markers per linkage group, with a total of 48 common markers. Marker order was the same between maps except in four cases. In two of these, CT156_1200 on chromosome 12 and SSR14_180 on chromosome 3, only 1 cM and 2 cM of localization discrepancy were found, respectively. On chromosome 3, SSR6_180 showed 4 cM of localization discrepancy between maps. Finally, a discrepancy for the ordering of markers within chromosome 10 was solved by ignoring marker SSRW318_298. As we will demonstrate, the sequence analysis of SSRW318_298 revealed more than one sequence for the E and P parents (Table 3).

Genetic structure of populations

E and *P* homozygotes were present in the P population at an average frequency of 0.44 and 0.48, respectively. Of the markers examined, 56.4% showed a higher frequency of *P* homozygotes than *E* homozygotes. Only chromosomes 7, 8 and 11 had large regions where the *E* homozygotes were more abundant than the *P* homozygotes. The average frequency of the heterozygous class was 0.082.

In the C population, average frequencies of *E* and *C* homozygotes at the examined loci were 0.50 and 0.39, respectively. *E* homozygotes were more frequent than the *C* ones at 73.1% of the markers. Only chromosomes 4, 6

Table 3 Variability for the number of AT repeats and nucleotide substitutions among clones isolated from *L. esculentum* (E), *L. pimpinellifolium* (P) and *L. cheesmanii* (c) for markers SSR6_180, SSR29_250, SSRW47_220 and SSRW318_298

Marker	Parent	Size (bp)[a]	Nucleotide substitution	Homologies—e-value
SSR6_180	E	170		LAT59 gene for protein P59 (*L. esculentum*). e-value: 2e−24
	C	162		
		162		
		162		
		162		
SSR29_250	E	222 (1)		TOMLE25 gene ABA regulated, and accumulated in developing seeds and drought-stressed leaves (*L. esculentum*). e-value: 9e−46
		222 (1)		
		220		
		220		
	P	218 (1)		
		218 (1)		
		218 (1)		
		216		
	C	218	A to G at nt 207	
		218	A to G at nt 207	
		218		
		218		
SSRW47_220	E	189 (2)	C to T at nt 1	sgn-u146705 *Lycopersicon* contig homolog to putative pectinesterase (*Oryza sativa*). e-value: 2e−158
		191		
		191		
	P	173		
		173		
	C	203	C to T at nt 48	
		203		
SSRW318_298	E	273		*TP5* gene putative beta-galactosidase (*Nicotiana tabacum*). e-value: 8e−19
		273		
		273		
		275 (1)	G to A at nt 29	
	P	271		
		271	G to A at nt 228	
		271	T to C at nt 83, C to T at nt 220	
		267 (2)		

[a]Size: (1), AT insertion; (2), AT deletion. BLASTN results using the consensus sequence and corresponding e-values

and 12 had large regions at which *C* homozygotes were more abundant than *E* homozygotes. The average frequency of the heterozygous class was 0.11, while the expected frequency of this class in the F_6 plants is 0.03125.

The distribution of the percentage of *E* alleles per marker locus in the P and C populations is presented in Fig. 3a. P population mode was 50% and second major class, 40%. On the other hand, C population major classes were 60% (mode) and 50%.

We observed a clear difference between populations for the distribution of the percentage of *E* alleles per line (Fig. 3b). The population modes are 50% and 60% in the P and C populations, respectively. The same displacement of distributions was observed for the percentage of heterozygotes per marker (Fig. 3c), with the modes being 5.1–7.5% and 10.1–12.5% in the P and C populations, respectively.

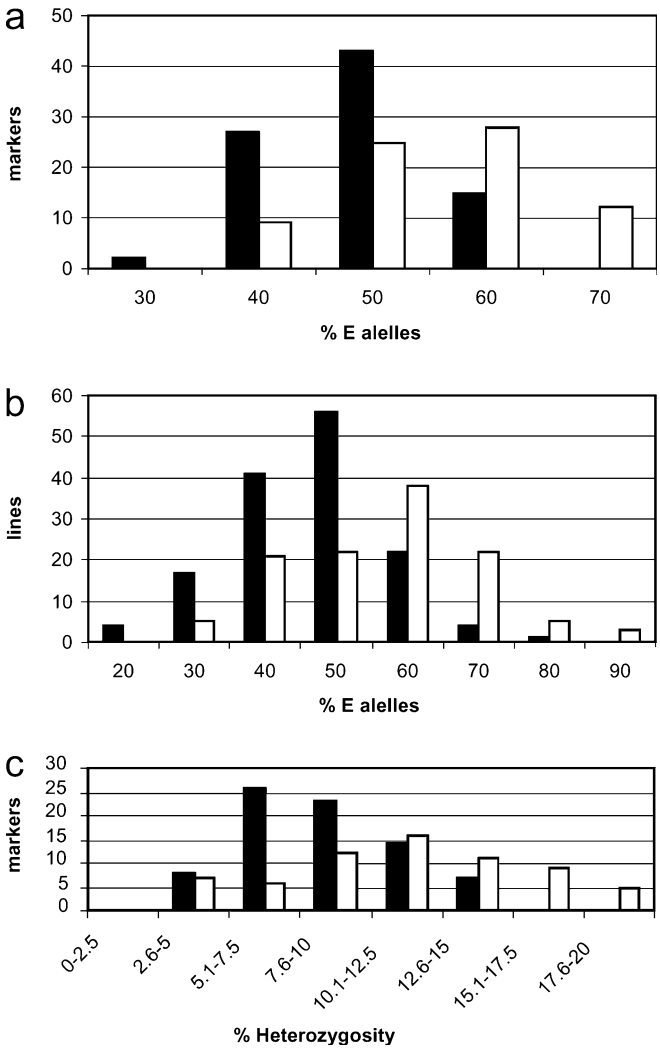


Fig. 3 Distribution of the percentage of *L. esculentum* (*E*) alleles and the percentage of heterozygosity in P (*black bars*) and C (*white bars*) populations. **a** Distribution of the percentage of *E* alleles relative to the number of molecular markers, **b** distribution of the percentage of *E* alleles relative to the number of RILs, **c** distribution of the percentage of heterozygosity of the populations relative the number of molecular markers

Marker segregation distortion and allele composition of populations

The distribution of the allelic ratios observed in both populations is presented in Fig. 4. Of the marker loci examined, 84% showed the expected 1:1 ratio and the *E*:*P* proportion mode was 0.9 in P population. The P population showed a moderate selection favoring *P* alleles at most of the marker loci. Of the markers, 16% showed allelic distortion, while 42 markers (30%) deviated significantly ($P<0.001$) from the F_6 expected genotypic ratio (this percentage was 13.4% in the F_2 generation; data from Monforte et al. 1997a).

In the C population only 70% of markers showed the expected 1:1 allelic ratio, and the *E*:*C* ratio was 1.1. Therefore, the C population presented a strong segregation distortion against *C* alleles. A total of 30% of the markers showed gametic distortion. Sixty markers (51%) deviated significantly ($P<0.001$) from the F_6 expected genotypic ratio that clearly favored *E* alleles. Although this advantage of *E* alleles was also observed in the F_2 generation, only 20% of markers presented genotypic distortion in the latter.

In general, loci with a skewed allelic ratio were scattered throughout the genome, although markers presenting maximum χ^2 values with respect to segregation distortion in both populations were found on chromosome 2: SSRW26_175 and TG30_320 in the P population and SSR12_140 and TG48_350 in the C population. The distortion in these four markers is mainly due to an excess of heterozygotes. In most genomic regions the degree of deviation ($P<0.001$) was similar for all linked markers. However, in a few regions one marker deviated strongly while its nearest neighbor did not. Examples of such markers are: in the C population, SSR6_180 on chromosome 3 ($E=0.64$, $C=0.36$), CT141_320 ($E=0.41$, $C=0.59$) on chromosome 6 and SSR29_250 ($E=0.64$, $C=0.36$) on chromosome 10; in the P population, SSR29_250 on chromosome 10 ($E=0.36$, $P=0.64$).

Linked markers showing segregation distortion ($P<0.001$) were frequently found, especially in the C population. Chromosomes 2, 4 and 10 showed regions with distorted segregation towards the *P* allele in the P population, while in the C population chromosomes 1, 2, 3, 8, 9 and 10 had large regions with segregation distortion towards the *E* allele.

Sequence analysis of problematic markers

The sequences of a few specific markers were analyzed: SSR6_180 and SSR29_250 on chromosomes 3 and 10 because of their high allelic segregation distortion and location discrepancy between maps and SSRW47_220 and SSRW318_298 on chromosomes 6 and 10 due to

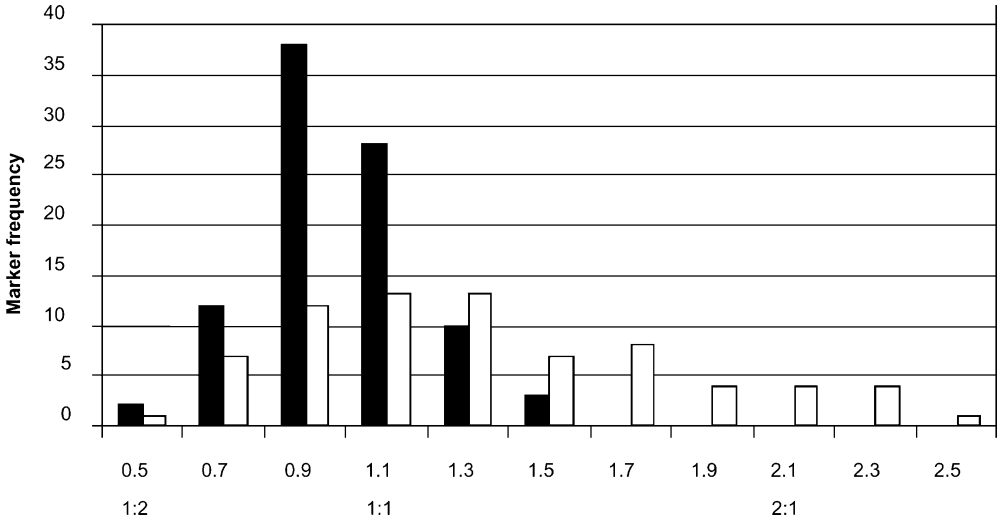


Fig. 4 Distributions of the *E*:*P* and *E*:*C* allele ratios within the P (*black bars*) and C (*white bars*) populations

their high genetic segregation distortion within chromosomal regions that did not show distortion for surrounded markers. The sequence analysis (Table 3) indicated that some of these markers are presumably present in the genome in more than one copy.

Markers SSR29_250, SSRW47_220 and SSRW318_298 showed variability among clones isolated from the same parental line for the number of AT repeats. Four SSR29_250 clones were isolated for the E, P and C parental lines. Two clones from the E parent and three clones from the P parent showed a dinucleotide (AT) insertion at the microsatellite sequence. For SSRW47_220, a dinucleotide (AT) deletion in the microsatellite sequence was found in one of the three clones isolated from the E parental line. For SSRW318_298, one of the four clones from the E parent showed a dinucleotide (AT) insertion. For the P parent, we found that one of the four clones had a dinucleotide (AT) deletion in the microsatellite sequence.

No difference among the four clones isolated from the C parent was found for the SSR6_180 marker.

Indels were not the only variation found among sequences cloned from the same parental line; nucleotide substitutions were also found for SSR29_250 (C), SSR47_220 (E and C) and SSRW318_298 (E and P).

All sequences per marker were aligned, and the consensus sequence was used for BLASTN analysis in the NCBI database and the SGN database. These microsatellites are placed in intragenic regions, and homologies for the flanking regions of the Poly(AT) microsatellite core are summarized in Table 3.

Discussion

Polymorphism of microsatellites and genetic diversity

The co-dominant marker type most commonly used in the construction of earlier linkage maps is the RFLP marker. In comparison to SSR marker technology, RFLP genotyping protocol is very laborious and has a low level of polymorphism. For mapping purposes, a higher level of polymorphism is needed when the initial cross involves genetically closely related species, as in the present study. The level of polymorphism we observed between the species examined was much higher in the present study than in the one reported by Alvarez et al. (2001): the latter found that only 29% of the microsatellites were polymorphic between *L. esculentum* and *L. cheesmanii*, while we found 56% were polymorphic taking into account just the main band. This difference polymorphism level can be explained within the framework of the set of SSR markers used: when we estimated the percentage of polymorphic SSR that just included the 14 primer pairs in common with both the present study and that of Alvarez et al. (2001), the amount of polymorphism changed from 56% to 28%. This indicates that any estimation of genetic differences between species using SSR markers might be highly dependent on the set of SSR markers used. If we take our percentage of polymorphic SSRs as a measure of genetic distance, *L. pimpinellifolium* and *L. cheesmanii* would be equally similar to *L. esculentum*, a result that coincides with the study on genetic relationships reported by Alvarez et al. (2001). Nevertheless, differences between the sizes of SSR alleles are usually much larger between *L. cheesmanni* and *L. esculentum* than between *L. pimpinellifolium* and *L. esculentum* and, moreover, the much higher gametic and genotypic segregation distortion found in the C population (31% and 52%, respectively) relative to that found in the P population (18% and 29%, respectively) suggests that *L. pimpinellifolium* is closer to *L. esculentum* than *L. cheesmanii* is, as has been reported in previous studies using isozymatic markers (Rick and Fobes 1975; Bretó et al. 1993). Therefore, the results of any study of genetic diversity and genetic relationships among species using SSR markers should be considered with caution because diversity does not seem to be exclusively a matter of number of alleles, differences among alleles per SSR marker should be also considered, i.e. not all of the alleles found for a set of species

should be considered to be equally different from one another.

New SSR alleles (not present in parents) arose in the P population (Fig. 1) and always in the same F_6 lines: 137 and 189. After we had discarded errors in the multiplication of these lines, we hypothesized that the genotype of these lines, which were fixed during the selfing generations, increased the occurrence of genetic recombination. Conversely, the fact that new alleles arose in both lines at the same four loci (SSRW63_240 and SSRW344_290 (chromosome 8), SSRW241_200 and SSRW285_290 (chromosome 7) suggests that these loci might be hotspots of recombination. Meiotic recombination is not distributed uniformly throughout eukaryotic genomes, and variation in recombination between different chromosomal regions can be of several orders of magnitude in some species. This variation has given rise to the concept of recombination hotspots. Although genes comprise a small fraction of the genome, they behave in general as recombination hotspots in the sense that intragenic recombination frequencies have been found to be several times greater than recombination between genes (Dooner and Martínez-Férez 1997). In *Lycopersicon*, a recombination hotspot, within an apoplastic invertase, facilitated the precise mapping of the *Brix9-2-5* QTL derived from *L. pennellii* (Fridman et al. 2000). Our finding agrees with these results because SSRW markers were derived from expressed sequences. Investigations currently in progress are focused on testing this hypothesis.

Zygotic and gametic segregation distortions confirm the genetic structure of the populations

Despite having a common female parent, both F_6 populations were found to be quite different with respect to genetic constitution. These populations share only a high frequency of markers that show a genotypic segregation distortion towards an excess of heterozygotes along all chromosomes, with chromosome 2 showing the highest percentage of genotypic and allelic segregation distortion—although in different directions (toward *P* in the P population and toward *E* in the C population). In general, *P* alleles are favored, if present, in the P population, while six chromosomes showed markers with gametic segregation distortion towards the *E* allele in the C population. This observation is in agreement with data on marker segregation distortion at the corresponding F_2 progenies (Monforte et al. 1997a).

There are two main differences between the F_2 and F_6 populations: (1) the large increment in genotypic segregation distortion from generation 2 to generation 6; (2) the origin of this distortion is an excess of homozygotes (for the *P* or *E* alleles) in the F_2 population, while in the F_6 population, it is mostly due to an excess of heterozygotes. Therefore, as the number of generations of self-pollination increases, the viability and/or fertility of homozygotes (at any genomic location) seem to decrease, thereby making allele fixation difficult. Why does it happen in populations derived from autogamous, self-compatible species? Our results suggest that the genes controlling the reproductive system, which are fixed in each of these three species, are not the same, especially between *L. esculentum* and *L. cheesmanii*. A decrease in the frequency of the wild allele and/or the maintenance of a high level of heterozygosity as found in the C population have also been observed in the development of other advanced populations of tomato interspecific crosses (Table 1) involving *L. cheesmanii* (Paran et al. 1995), *L. peruvianum* (Fulton et al. 1997) and *L. hirsutum* (Monforte and Tanksley 2000).

Minor changes in the percentage of markers with gametic segregation distortion or the origin of the favored allele were observed during advancement from the F_2 to the F_6 generation. *E* alleles are favored in the C population, especially on chromosome 2, while gametes carrying the *P* allele in this chromosome are more viable and/or fertile in the P population. This different direction in gametic selection might be explained by differences in the fertility of the pollen on pistils with esculentum cytoplasm that both populations share. This preference in the transmission of the *E* allele occurs in five other chromosomes in population C. Given that all markers that showed gametic distortion also showed genotypic distortion, but not the reverse, it seems clear that it is not only heterozygosity that is related with a higher viability/fertility of lines but also the fixation of *E* alleles in the genomic regions of six chromosomes in population C. In the P population, heterozygosity is the main factor ensuring the viability/fertility of lines, and the presence of the fixation of *P* allele seems to be important only on chromosome 2, i.e. fewer loci controlling self-fertility would be segregating in this population. In fact, although there were 100 fewer F_2 plants in the P population than in C population, more F_6 self-fertile lines were obtained in the former. A locus affecting self-compatibility on chromosome 2 has been reported in other studies (Paterson et al. 1990; Bernacchi and Tanksley 1997; Monforte and Tanksley 2000). The presence of a high degree of heterozygosity in this region of chromosome 2 resulted in significantly enhanced self-fertility in a cross between *L. esculentum* × *L. hirsutum* (Bernacchi and Tanksley 1997) and in a BC_1 derived from *L. esculentum* × *L. chmielewskii* (Paterson et al. 1990). Zamir and Tadmor (1986) also reported an enrichment of homozygotes for *L. pennellii* alleles in a F_2 population derived from a cross between *L. esculentum* × *L. pennellii* for the same region of chromosome 2.

Comparison of linkage maps

Genetic linkage maps are an essential tool for practical applications such as marker-assisted selection and the map-based cloning of target genes in which a correct linear order of loci within linkage groups is essential. Intrachromosomal rearrangements in the genus *Lycopersicon* have been reported for chromosome 9 in a

progeny derived from *L. esculentum* and *L. peruvianum* (Fulton et al. 1997). Considerable similarity between the maps from the P and C populations was observed in the present study (Fig. 2): marker order was the same between maps except in four cases. In two of these cases, CT156_1200 on chromosome 12 and SSR14_180 on chromosome 3, just 1 cM and 2 cM, respectively, of localization discrepancy were found. Loci presenting slight differences in localization discrepancy between maps can be explained by random errors, especially when markers with large segregation distortion are involved. For the other two cases, the location discrepancy was larger and the possibility of duplicate marker loci was investigated.

More than one-third of a typical eukaryotic genome consists of duplicated genes and gene families. The complete genome sequence of *Arabidopsis* has revealed that an estimated 17% of the 25,000 genes are arranged in tandemly repeated segments (The Arabidopsis Genome Initiative 2000), with 60% of the genome contained within large duplicated segments (Blanc et al. 2000; Goff et al. 2002) and almost one-half of the *Arabidopsis* genes within the duplicated segments being conserved. If a duplicated chromosome region contains a DNA sequence that can be used as a molecular marker, the marker alleles at the two duplicated marker loci cannot be distinguished. The equal fragment length results in an identical banding pattern and, consequently, the alleles of duplicated markers are scored in a mapping population as the alleles of one single marker. Frisch et al. (2004) reasoned that segregation distortion caused by gametic/genotypic selection can be distinguished from that caused by a possible duplication event by the fact that segregation distortion for duplicated markers only occurs at the ghost locus (where the four alleles at the two duplicated markers are scored as the alleles of only one marker). In contrast, for zygotic/allelic selection, segregation distortion occurs not only at the locus that is affected by selection but also at closely linked loci. Consequently, we carried out sequencing analysis of markers showing isolated segregation distortion on chromosomes 3 (SSR6_180) and 10 (SSR29_250, SSRW318_298) and 6 (SSRW47_220) in order to test for the presence of duplicated SSR loci as the origin of their distorted segregation and mapping location discrepancy.

Microsatellite sequences of these four markers revealed some differences with respect to the number of dinucleotide (AT) repetitions (Table 3) in accordance with the stepwise mutation model (SMM) proposed by Otha and Kimura (1973) for microsatellite allele diversification. The resolution properties of acrylamide gel electrophoresis enables a 1-bp discrimination between amplification products, but none of the amplified fragments used for genotyping the population showed appreciable differences in fragment length. Nevertheless, differences in the length of inserts from clones of the same genotype could be visualized using acrylamide gel electrophoresis, which suggests only two possible explanations: errors in *taq* polymerase synthesis or preferential amplification of one of the two duplicate SSR loci differing in the number of dinucleotide repeats. In support of the hypothesis of duplicated SSR loci, differences involving simple nucleotide transitions were also found among clones of the same marker loci. However, sequence differences were not as large as those reported by Frisch et al. (2004) for 10 of the 13 amplified fragment length polymorphic (AFLP) bands showing location discrepancy between two maize maps.

Sequence homologies found in the database suggest that the four microsatellites are placed in coding genes. The genes containing the microsatellites pectin esterase (SSR47_220) and β-galactosidase (SSRW318_298) belong to gene families, indicating that duplication events did indeed occur in the past. In the case of SSRW318_298, its elimination from map construction in P population resulted in a common ordering of markers. In addition, pectin esterase, pectate lyase and β-galactosidase are members of a set of genes whose mRNA accumulates late in pollen development, and they are presumed to be stored in readiness for pollen germination. The putative β-galactosidase *TP5* gene, highly homologous to the gene including marker SSRW318_298, seems to play an important role in pollen-tube wall turnover and pollen fertility (Rogers et al. 2001). Therefore, segregation distortion at these loci might be also explained by their effect on pollen fertility during the development of the inbred lines. In the case of marker SSR6_180, for which no difference among *L. cheesmanii* clones was found, the presence of a selective advantage of *E* homozygotes at this locus that includes a pectate lyase gene might explain its large zygotic segregant distortion in linkage group 3C.

Despite the low level of polymorphism usually found when constructing linkage maps using phylogenetically closely related species, the use of microsatellite markers has allowed us to construct medium-density linkage maps for two populations of F_6 lines derived from *L. esculentum* × *L. pimpinellifolium* and *L. esculentum* × *L. cheesmanii*. Efforts are still needed to increase the number of common markers and candidate genes to allow a more complete comparison. The present marker characterization of both populations will continue and will enable comparative QTL and candidate gene analysis of complex traits towards a more efficient utilization of genetic resources and breeding strategies.

Acknowledgements This work was supported in part by grants from IVIA (IV), CSIC (ARS) and INIA (RTA01-113-C2). The authors thank Drs. G.P. Bernet and J. Puchades for technical assistance.

References

Allard RW (1988) Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. J Hered 79:225–238

Altschul SF, Madden TL, Schaffer, AA, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Alvarez AE, Van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theor Appl Genet 103:1283–1292

Areshchenkova T, Ganal MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42:536–544

Areshchenkova T, Ganal M (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229–235

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum*×*L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147:861–877

Blanc G, Barakat A, Guyot R, Cook R, Delseny M (2000) Extensive duplication and reshuffling in the *Arabidopsis* genome. Plant Cell 12:1093–1101

Bredemeijer GMM, Arens P, Wouters D, Visse D, Vosman B (1998) The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor Appl Genet 97:584–590

Bretó MP, Asins MJ, Carbonell EA (1993) Genetic variability in *Lycopersicon* species and their genetic relationships. Theor Appl Genet 86:113–120

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Theor Appl Genet 250:39–49

Brown AHD (1978) Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52:145–157

Chen FQ, Foolad MR (1999) A molecular linkage map of tomato based on a cross between *Lycopersicon esculentum* and *L. pimpinellifolium* and its comparison with other molecular maps of tomato. Genome 42:94–103

deVicente MC, Tanksley SD (1993) QTL analysis of transgressive segregation in an inter specific tomato cross. Genetics 134:585–596

Dooner HK, Martinez-Ferez IM (1997) Germinal excisions of the maize transposon activator do not stimulate meiotic recombination or homology-dependent repair at the bz locus. Genetics 147:1923–1932

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTL. Genetics 141:1147–1162

Francis Ar, Shetty TK, Bhattacharya RK (1989) Modifying role of dietary factors on the mutagenicity of aflatoxin B1: in vitro effect of plant flavonoids. Mutant Res 222:393–401

Frary A, Fulton TM, Zamir D, Tanksley SD (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum*×*L. pennellii* cross and identification of possible orthologs in the Solanaceae. Theor Appl Genet 108:485–496

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA 97:4718–4723

Frisch M, Quint M, Melchinger AE (2004) Duplicate marker loci can result in incorrect locus orders on linkage maps. Theor Appl Genet 108:485–496

Fulton TM, Nelson JC, Tanksley SD (1997) Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into *Lycopersicon esculentum*, followed through three successive backcross generations. Theor Appl Genet 95:895–902

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum*×*L. parviflorum* cross. Theor Appl Genet 100:1025–1042

Ganal MW, Czihal R, Hannappel U, KloosD-U, Polley A, Ling H-Q (1998) Sequencing of cDNA clones from the genetic map of tomato (*Lycopersicum esculentum*). Genome Res 8:842–847

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa L.* ssp. *japonica*). Science 296:92–100

Grandillo S, Tanksley SD (1996) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium*. Theor Appl Genet 92:957–965

Haanstra JPW, Wye C, Verbakel H, Meijer-Dekens F, van den Berg P, Odinot P, van Heusden AW, Tanksley SD, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum*×*L. pennellii* F_2 populations. Theor Appl Genet 99:254–271

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor Appl Genet 106:363–373

Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. Theor Appl Genet 94:701–706

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Krinsky NI (1992) Anticarcinogenic activities of caretenoids in animals and cellular systems. In: Emers I, Chance B (eds) Free radicals and aging. Birkhäuser, Basel, pp 227–234

Lippman Z, Tanksley SD (2001) Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small fruited wild species *Lycopersicon pimpinellifolium* and *L.esculentum* var. *Giant Heirloom*. Genetics 158:413–422

Lukens LN, Doebley J (1999) Epistatic and environmental interactions for QTL involved in maize evolution. Genet Res 74:291–302

Mather K (1957) The measurement of linkage in heredity. Methuen, London

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome 43:803–813

Monforte AJ, Asins MJ, Carbonell EA (1996) Salt tolerance in *Lycopersicon* species. IV. Efficiency of marker-assisted selection for salt tolerance improvement. Theor Appl Genet 93:765–772

Monforte AJ, Asins MJ, Carbonell EA (1997a) Salt tolerance in *Lycopersicon* species V. Does genetic variability at quantitative trait loci affect their analysis?. Theor Appl Genet 95:284–293

Monforte AJ, Asins MJ, Carbonell EA (1997b) Salt tolerance in *Lycopersicon* species. VI. Genotype by salinity interaction in quantitative trait loci detection. Constitutive and response QTLs. Theor Appl Genet 95:706–713

Monforte AJ, Asins MJ, Carbonell EA (1999) Salt tolerance in *Lycopersicon* species. VII. Pleiotropic action of genes controlling earlyness. Theor Appl Genet 98:593–601

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electroforetically detectable alleles in a finite population. Genet Res 22:201–204

Paran I, Goldman IL, Tanksley SD, Zamir D (1995) Recombinant inbred lines for genetic mapping in tomato. Theor Appl Genet 90:542–548

Paterson AH, Deverna JW, Lanini B (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecific cross of tomato. Genetics 124:735–742

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. Genetics 127:181–197

Pérez de la Vega M, Sáenz-de-Miera LE, Allard RW (1994) Ecogeographical distribution and differential adaptedness of multilocus allelic associations in Spanish *Avena sativa* L. Theor Appl Genet 88:56–64

Rick CM, Fobes JF (1975) Allozymes of Galápagos tomatoes: polymorphism, geographic distribution, and affinities. Evolution 29:443–457

Rogers HJ, Maund SL, Johnson LH (2001) A beta-galactosidase like gene is expressed during tobacco pollen development. J Exp Bot 52:67–75

Ruíz C, Bretó MP, Asins MJ (2000) A quick methodology to identify sexual seedlings in citrus breeding programs using SSR markers. Euphytica 112:89–94

Simpson SP (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor Appl Genet 77:815–819

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 94:264–272

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Tanksley SD, Ganal MW, Prince JP, deVicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage map of the tomato and potato genomes. Genetics 132:1141–1160

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J, Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Van Ooijen JW, Voorrips RE (2001) JOINMAP version 3.0, Software for the calculation of genetic linkage maps. Release 3.0. Plant Research International, Wageningen

Verma AK, Johnson JA, Gould MN, Tanner MA (1988) Inhibition of 7,12-dimethylbenz(a)anthracene and *N*-nitrosomethylurea induced rat mammary cancer by dietary flavonol quercitin. Cancer Res 48:5754–5788

Vosman B, Arens P (1997) Molecular characterization of GATA/GACA microsatellite repeats in tomato. Genome 40:25–33

Zamir D, Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Bot Gaz 147:355–358

Zhang LP, Khan A, Nino-Liu D, Foolad MR (2002) A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum*×*Lycopersicon hirsutum* cross. Genome 45:133–146