Theor Appl Genet (2005) 111: 291–312
DOI 10.1007/s00122-005-2023-7

ORIGINAL PAPER

Anne Frary · Yimin Xu · Jiping Liu · Sharon Mitchell
Eloisa Tedeschi · Steven Tanksley

Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments

Received: 2 February 2005 / Accepted: 23 March 2005 / Published online: 31 May 2005
© Springer-Verlag 2005

Abstract Tomato and potato expressed sequence tag (EST) sequences contained in the solanaceae genomics network (SGN) database were screened for simple sequence repeat (SSR) motifs. A total of 609 SSRs were identified and assayed on *Solanum lycopersicum* LA925 (formerly *Lycopersicon esculentum*) and *S. pennellii* LA716 (formerly *L. pennellii*). The SSRs that did not amplify, gave multiple band products, or did not exhibit a polymorphism that could be readily detected on standard agarose gels in either of these species were eliminated. A set of 76 SSRs meeting these criteria was then placed on the *S. lycopersicum* (LA925) × *S. pennellii* (LA716) high-density map. A set of 76 selected cleaved amplified polymorphism (CAP) markers was also developed and mapped onto the same population. These 152 PCR-based anchor markers are uniformly distributed and encompass 95% of the genome with an average spacing of 10.0 cM. These PCR-based markers were further used to characterize *S. pennellii* introgression lines (Eshed and Zamir, Genetics 141:1147–1162, 1995) and should prove helpful in utilizing these stocks for high-resolution mapping experiments. The majority of these anchor markers also exhibit polymorphism between *S. lycopersicum* and two wild species commonly used as parents for mapping experiments, *S. pimpinellifolium* (formerly *L. pimpinellifolium*) and *S. habrochaites* (formerly *L. hirsutum*), indicating that they will be useful for mapping in other interspecific populations. Sixty of the mapped SSRs plus another 49 microsatellites were tested for polymorphism in seven tomato cultivars, four *S. lycopersicum* var. *cerasiforme* accessions and eight accessions of five different wild tomato species. Polymorphism information content values were highest among the wild accessions, with as many as 13 alleles detected per locus over all accessions. Most of the SSRs (90%) had accession-specific alleles, with the most unique alleles and heterozygotes usually found in accessions of self-incompatible species. The markers should be a useful resource for qualitative and quantitative trait mapping, marker-assisted selection, germplasm identification, and genetic diversity studies in tomato. The genetic map and marker information can be found on SGN (http://www.sgn.cornell.edu).

Communicated by R. Hagemann

A. Frary · Y. Xu · J. Liu · E. Tedeschi
Department of Plant Breeding, Cornell University, Ithaca, NY 14853, USA

Present address: A. Frary
Department of Biology, Izmir Institute of Technology, Urla, Izmir, 35430, Turkey

S. Mitchell
Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA

S. Tanksley (✉)
Departments of Plant Breeding and Plant Biology, Cornell University, Ithaca, NY 14853, USA
E-mail: sdt4@cornell.edu
Tel.: +1-607-2551673
Fax: +1-607-2556683

Introduction

In addition to its worldwide agricultural and economic importance as a crop, tomato (*Solanum lycopersicum*, formerly *Lycopersicon esculentum*) is a preeminent model system for genetic studies in plants. Genetic mapping of morphological traits began during the first decades of the last century (Jones 1917) and, by the 1990s, tomato was one of the first plants for which a high-density DNA-based molecular map was available (Bernatzky and Tanksley 1986; Tanksley et al. 1992). This map, consisting primarily of restriction fragment length polymorphisms (RFLPs), facilitated many other "firsts" in the species. Extensive mapping of qualitative traits such as various disease resistances (e.g., Young et al. 1988; Martin et al. 1991) allowed tomato breeders

to be among the first to use powerful marker-assisted selection (MAS) techniques to develop improved plant cultivars. In addition, tomato was the first species in which quantitative trait loci (QTLs) mapping for a complete genome was conducted in a single segregating population (Paterson et al. 1988). Qualitative and quantitative trait mapping was further simplified by the development of a set of genetically fixed, molecularly characterized, overlapping introgression lines (ILs; Eshed and Zamir 1995) that are used as a permanent resource to identify and fine-map genes more rapidly and at less expense than transient populations. Moreover, the first plant disease resistance gene to be cloned by a map-based approach, *Pto* (Martin et al. 1993), as well as the first plant QTL, *fw2.2* (Frary et al. 2000), were both isolated in tomato.

Since publication of the original high-density RFLP map (Tanksley et al. 1992), additional types of molecular markers, including microsatellites (Broun and Tanksley 1996; Grandillo and Tanksley 1996a; Areshchenkova and Ganal 1999, 2002; Suliman-Pollatschek et al. 2002), random amplified polymorphic DNA (RAPDs; Grandillo and Tanksley 1996a), amplified fragment length polymorphisms (AFLPs; Haanstra et al. 1999; Suliman-Pollatschek et al. 2002) and single nucleotide polymorphisms (SNPs; Suliman-Pollatschek et al. 2002) have been utilized in tomato. All of these alternative markers are based on the PCR; as a result, many are considered to be cheaper, faster, less labor intensive, and more widely applicable than RFLPs. However, with the exception of a combined RFLP-AFLP map (Haanstra et al. 1999), there has been no PCR-based map that provides complete coverage of the tomato genome. The availability of a map containing highly reproducible, locus-specific, codominant PCR-based markers would facilitate many types of studies in tomato (including gene mapping, QTL analysis, and MAS) particularly in research and breeding programs with limited resources.

The goals of the present study were to develop a set of PCR-based anchor markers for research in tomato genetics and breeding. These markers consist of microsatellite and cleaved amplified polymorphism sequence (CAPs) markers, primarily based on single-copy/coding regions. These markers encompass the entire tomato genome, are placed at regular intervals, and are anchored in the tomato high-density molecular linkage map that currently comprises 1,579 markers (http://www.sgn.cornell.edu). Moreover, a priority was given to markers for which polymorphism could be readily detected on standard agarose gels to facilitate their use in laboratories throughout the world. Microsatellites or simple sequence repeats (SSRs) are short (usually two to four nucleotide), tandemly repeated DNA sequences. Because of their repetitive nature, the length (i.e., the number of repeat units) of SSRs are vulnerable to mutation caused by replication slippage and unequal cross-over during meiosis (Valdes et al. 1993). As a result, SSRs are often highly polymorphic. However, because they represent simple repeated sequences, SSRs are likely to be located in regions of repetitive DNA, such as centromeres, so that random selection of SSRs from genomic DNA can result in a distorted map in which such markers are clustered around the pericentric heterochromatin (Areshchenkova and Ganal 1999). By screening for SSRs in and around coding regions, it is more likely that a set of markers anchored in the coding (especially euchromatic) portions of the chromosome will be identified and fuller genome coverage will be attained. Full genome coverage of SSRs can be complemented with CAPs assays to allow uniform coverage of the entire coding portion of the genome.

In the SSR technique, primers that are complementary to the conserved regions that flank a microsatellite are used to amplify the repeat. Length polymorphisms are then detected by gel electrophoresis using either standard agarose or sequencing gels (Akkaya et al. 1992). In the CAPs technique, sequence-specific primers are used to amplify a given region of the genome. Amplification products are then cleaved with an appropriate restriction enzyme and separated by electrophoresis to reveal polymorphism among individuals (Konieczny and Ausubel 1993).

The primary disadvantage of both SSRs and CAPs is their dependence on a priori DNA sequence information, however, a considerable amount of sequence information currently exists for tomato and the closely allied species potato. The recent development of an expressed sequence tags (EST) database containing nearly 300,000 tomato and potato sequences (Van der Hoeven et al. 2002; http://www.sgn.cornell.edu) has made it easier to identify and exploit SSRs in the species. In addition, the sequences of most of the tomato RFLP markers, the majority of which correspond to coding regions, are publicly available (http://www.sgn.cornell.edu) and provide an excellent resource for CAPs development. The current high-density molecular map of tomato contains more than 1,500 markers, 576 of which correspond to conserved ortholog set (COS) markers derived from ESTs that are conserved and single or low copy in both the tomato and Arabidopsis genomes as well as a subset of RFLP and other markers from the first high-density map (Tanksley et al. 1992; Fulton et al. 2002; http://www.sgn.cornell.edu). The sequences of these COS markers provide another source for CAPs marker development.

In order to be considered for use on the map, the anchor SSR and CAPs markers had to meet certain criteria: detection of polymorphism between the parents of the high-density mapping and IL populations, *S. lycopersicum* and *S. pennellii*; visualization of the polymorphism on standard agarose gels to allow use by as many laboratories as possible; and placement of the markers on the map at intervals of less than 20 cM. Additional SSR markers were genotyped in the mapping population with a sequencer and are also included on the map. Moreover, 109 SSR markers were surveyed for polymorphism on a range of tomato germplasm including seven cultivated lines, four *S. lycopersicum*

var. *cerasiforme* lines, and eight wild accessions. This information should be of interest for those wishing to use these SSRs for mapping with other parental lines and for genetic diversity studies.

Materials and methods

Plant material

An F_2 population of 83 individuals from the cross *Solanum lycopersicum* (LA925) × *S. pennellii* (LA716) was used for map construction. This population was previously used to construct a map containing more than 1,500 markers derived from genes that are conserved and single or low copy in both tomato and Arabidopsis (Fulton et al. 2002). A subset of markers was also mapped on the *S. pennellii* ILs (Eshed and Zamir 1995). These lines contain defined introgressions of *S. pennellii* (LA716) in a *S. lycopersicum* cv. M82 background. Nineteen tomato lines were used for the polymorphism study, including seven cultivated lines (*S. lycopersicum* var. *esculentum*), four accessions of the putative progenitor of cultivated tomato (*S. lycopersicum* var. *cerasiforme*), and eight accessions of five other wild species. The tomato lines used in the study are described in Table 1. The SSRs were also tested on potato (*Solanum tuberosum* B11B), eggplant (*S. melongena* MM738) and pepper (*Capiscum annuum* R Naky). Wild tomato accessions were provided by the C.M. Rick Tomato Genetics Resources Center, University of California, Davis, Calif., USA. Other lines were obtained from D. Zamir, Hebrew University, Jerusalem, Israel (TA210); P& P Seed Co., Collins, N.Y., USA (TA1486); Clause Seed Co., San Juan Bautista, Calif., USA (TA1487); M. Jahn, Cornell University, N.Y., USA (R Naky); M.-C. Daunay, INRA, Montfavet, France (MM738). The DNA from single plants of each cultivated or wild accession was extracted as described by Fulton et al. (1995).

Identification and amplification of microsatellites

In May 2001, tomato and potato EST sequences contained in the Solanaceae Genomics Network (SGN; http://www.sgn.cornell.edu) were screened for SSR motifs of two to six nucleotides with a total repeat length of at least 18 nucleotides using analytical tools developed by Genomics Edge Technologies (St. Louis, Mo.). In addition to perfect SSRs, the program also identified imperfect and compound repeats. Thus, 609 candidate SSRs were selected for further analysis. Primer pairs for each SSR were designed using the PRIMER3 program (Rozen and Skaletsky 1997). Primer sequences and repeat motifs for SSR markers mapped on agarose gels are listed in Table 2 and are available from SGN (http://www.sgn.cornell.edu). One previously published tomato microsatellite, SSR638, was also used in the analysis (Areshchenkova and Ganal 1999).

The DNA from TA209 and LA716 was tested for amplification from each primer pair using the MJ Research Research Peltier Thermal Cycler (PTC-225; Waltham, Mass.). Each 25-µl reaction mixture contained 30–50 ng of template DNA, 1 pmol of each forward and reverse primer, 1× PCR buffer (50 m *M* KCl,

Table 1 Description of plant material used and number of accession-specific alleles identified for each line

Species[a]	Accession	Name	Type/origin	Mating system[b]	Number of Unique alleles[c]
S. lycopersicum var. *esculentum*	TA209	E6203	Processing inbred	Autogamous SC	2
	TA210		Freshmarket inbred		3
	TA491	Sunpear	Freshmarket variety		2
	TA503	Yellow Pear	Freshmarket variety		0
	TA1143	M82	Processing variety		1
	TA1486	Monster (Giant Red) Tomato	Freshmarket variety		1
	TA1487		Freshmarket inbred		3
S. lycopersicum var. *cerasiforme*	LA1226		Morona-Santiago, Chile	Autogamous SC	5
	LA1388		Junin, Peru		8
	LA1455		Nuevo Leon, Mexico		7
	LA1574		Lima, Peru		6
S. pimpinellifolium	LA373		Ancash, Peru	Autogamous SC	25
	LA411		Los Rios, Ecuador		28
	LA1246		Loja, Ecuador		22
	LA1589		La Libertad, Peru		19
S. neorickii	LA2133		Azuay, Ecuador	Autogamous SC	34
S. habrochaites	LA1777		Ancash, Peru	Allogamous SI	54
S. pennellii	LA716		Arequipa, Peru	Facultative SC	61
S. peruvianum	LA1708		Cajamarca, Peru	Allogamous SI	30

[a] Formerly *L. esculentum* var. *esculentum*, *L. esculentum* var. *cerasiforme*, *L. pimpinellifolium*, *L. parviflorum*, *L. hirsutum*, *L. pennellii*, and *L. peruvianum*, respectively

[b] SC, Self-compatible; SI, self-incompatible

[c] Number of unique (accession-specific) alleles identified in each line

Table 2 Description of the agarose gel-based SSR and CAPs markers on the map (*NA* not applicable)

Marker[a]	Chromosome	Repeat type and length	SGN ID no.[b]	Forward primer (5′-3′)	Reverse primer (5′-3′)	Product size[c] (bp)	Polymorphic band sizes[c] (in bp)		Amplification
							LA716/E6203 (enzyme)	LA1589/ E6203 (enzyme)	St/Sm/ Ca[d]
SSR478	1	$(GGTG)_2$ $(CT)_{16}$	SGN-E226034	GCAGCATAT ATCACCTTGGCT	CGTGCTCTC CAATAGTTCACC	450	450/450 + 400	NA	ND
TG58	1	NA	NA	TGTGATACGGA ACTTTGAACCTCC	CCTGGATTTG GTCAGCTCCTTAG	900	900/600 (*Dra*I)	NA	NA
TG184*	1	NA	NA	TTGTTGGGAT AACGAGTCCA	ATTTGTCA TGCCCGAGCTTA	1100	500/600 (*Alu*I)	300/330 (*Ssp*I)	NA
SSR192	1	$(ATC)_6$	SGN-E288219	ACAACATGG GAAGCACTTGA	ATTAAATT GGGCCATGGTGA	169	169/165	NA	y/n/n
SSR51	1	$(ACAA)_6$	SGN-E360433	CTACCCTGGTC TTGGTGGAA	AAAGGATGCTC TAGCTTCTCCA	148	148/128	NA	y/y/y
SSR270	1	$(GAA)_5$ $(GGAGAA)_7$	None	AGCTCAAGGC TTCTGTTGGA	AACCACCTCA GGCACTTCAT	244	255/244	NA	y/y/n
SSR316	1	$(AG)_6$ $(TTGCAG)_2$	None	CCACCGCAACA AACCTTATT	GGGTGGTGA GAAGGATCTGA	217	234/217	NA	y/n/n
SSR135	1	$(ATT)_6$	SGN-E210325	TGATCGCTTG TGTCCACCTA	AAAGGAAGTGA TGGAAAGCG	133	128/133	NA	y/y/y
T1409	1	NA	NA	ATGTGATACC TCCGCCGTTA	TTTCTTCCATTTC CCAACCA	500	250/500 (*Sac*I)	NA	NA
SSR222	1	$(TCT)_7$	None	TCTCATCTG GTGCTGCTGTT	TTCTTGGAG GACCCAGAAAC	163	175/163	160/163	y/n/n
TG83	1	NA	NA	TTATGGCACTC AGGATGGTG	AGGCATGAAA ACCACAAAGG	1500	800 + 700/900 + 600 (*Hpa*II)	NA	NA
CT267	1	NA	NA	TGGGTATCCT AGAAGGCCAGT	ATGGTAGCCC TTGGTTGGAT	900/800	900/800	NA	NA
SSR150	1	$(CTT)_7$	SGN-E338056	ATGCCTCGCT ACCTCCTCTT	AATCGTTCG TTCACAAACCC	217	211/217	NA	y/y/y
SSR308	1	$(TA)_{12}$	None	TTTCCCTGTTT CAGCCTTTG	GGCACGAGAAT TTAGCCACT	293	293/300	NA	y/n/n
SSR117	1	$(TC)_{11}$	None	AATTCACCTT TCTTCCGTCG	GCCCTCGAA TCTGGTAGCTT	241	249/243	NA	y/y/y
SSR156	1	$(TCT)_7$	None	CACGCCTATG CACCTTTCTT	CTTCAAGGCT AAACCTCCGA	167	150/167	NA	y/y/n
TG27	1	NA	NA	GCCGAGTTCTC CATCTTCTCT	GCGGTTTTCA ACAACCAAAA	2500/2100	2500/2100	1300 + 1100/ 2200 (*Acc*I)	NA
cLEC7P21	2	NA	NA	TGAACAGAAA GCACGAGTGG	GACAGTTCTTC GAAGCGTTTG	350	300 + 50/350 (*Hae*III)	NA	NA
T1616	2	NA	NA	TTGGAAGAAGA GGAGGACGA	CCCAGAAAGATC CAAGTCCA	1500	900 + 600/1500 (*Hae*III)	NA	NA
SSR586	2	$(AAC)_6$	SGN-E362741	TCCATCTAAG GTCTTTGCCG	ACAAAGGAA GTGGGAGAGCA	120	124/120	NA	y/y/n
SSR40	2	$(AC)_7(GC)_7$	SGN-E282208	TGCAGGTATG TCTCACACCA	TTGCAAGAAC ACCTCCCTTT	146	180/146	NA	y/y/y
SSR66	2	$(ATA)_8$	SGN-E235275	TGCAACAACTG GATAGGTCG	TGGATGAAACG GATGTTGAA	185	174/185	183/186	y/y/y

SSR96	2	$(AT)_{12}$	SGN-E273474	GGGTTATCAATGAT GCAATGG	CCTTTATGT CAGCCGGTGTT	222	214/224	NA	y/y/y
SSR356	2	$(AT)_{20}$	None	ACCATCGAGG CTGCATAAAG	AACCATCCA CTGCCTCAATC	259	200/259	NA	y/y/n
SSR349A	2	$(ATAAAA)_2(TA)_{11}$	None	GAGTGATCATCC ATCCTCTCA	GGAAGAGACTT TGGACTAAGGGA	242	226/242	234/242	y/y/n
SSR605	2	$(CAA)_6$	SGN-E308255	TGGCCGGCTTC TAGAAATAA	TGAAATCAC CCGTGACCTTT	196	176/196	NA	y/n/n
T562	2	NA	NA	ATCCTCTGCC CTTTCTTTCC	CAATTGCAGG GGATATGCTT	1099	1099/980 (*Nsi*I)	NA	NA
SSR26	2	$(CGG)_6$	SGN-E300265	CGCCTATCGATAC CACCACT	ATTGATCCGTT TGGTTCTGC	179	173/179	NA	y/y/y
SSR32	2	$(TTC)_7$	SGN-E367187	TGGAAAGAAGC AGTAGCATTG	CAACGAACA TCCTCCGTTCT	187	187/175	NA	y/y/n
T1494	2	NA	NA	GACTGAAGTTT GTGGCGAAGG	TGCTGAATTGGC TGCTGCTA	1000	450 + 600/ 1000 (*Cfo*I)	NA	NA
T1480	2	NA	NA	ACCACCTTGG ATGAATACCG	TGCAACAGCTT TTCCCTCTC	1300	550/450 (*Alu*I)	550/450 (*Alu*I)	NA
Fw2.2	2	NA	NA	GCGAGAGCGA GTTGAGTGTATATC	CAGAAGAGAG AAGCTGC AAAGCAG	1300	1300/810 + 490 (*Taq*I)	NA	NA
cLER17N11	2	NA	NA	AACTTGTCTGG ACCCATTGC	TAAATGACCGC CTTTCTTGC	828	550/750 (*Rsa*I)	NA	NA
T634	2	NA	NA	TTCGCAGTTTTG GGAGTACC	GCGCAACAGTT CTTCTTTCC	1300	425 + 900/400 + 850 (*Rsa*I)	NA	NA
T1201	2	NA	NA	CATGCACTTA GGCAAAAGCA	TCCCAAATGT CCACCAATCT	1100	500 + 600/1100 (*Msp*I)	NA	NA
TG324	3	NA	NA	CACTTGGTTGAT GGATAGTG	CTTCTAGTA GTCCAACAG CAACTG	1300	1300/1200 (*Eco*RV)	NA	NA
cLPT5E7	3	NA	NA	ACAGCCAA GTCGAAGTCCTC	CCCAAGGGGG TTTAAGAGAA	1800	1400/800 + 600 (*Taq*I)	NA	NA
TG130	3	NA	NA	CGGAACCCACT TTGTTTTTC	ATCAACCCTG CAAGCTCAAC	1500	400/500 (*Alu*I)	NA	NA
TG585	3	NA	NA	TGGAAAGCCAG ACACACAGA	CAGGGGTATCA GTAGGCAGTG	491	491/300 + 191 (*Hin*fI)	506/350 + 191 (*Dra*I)	NA
T1388	3	NA	NA	GCGATTTGG CTATCTGGGTA	AACCGAAAG GCTTTTCCAAG	1000	480 + 270/350 + 200 (*Hin*fI)	700 + 300/700 + 200 (*Hin*cII)	NA
cLPT2E21	3	NA	NA	CGAAGATGTT GCTTGATTGC	AAGCAGGAGC TGGACACAAT	1250	1250/1050 + 200 (*Eco*RI)	NA	NA
TG74	3	NA	NA	CATGCTTGAA AAGCAGTGGA	GATTATACGA GGCCTCAAGGA	2300	2300/1900 (*Hpa*II)	NA	NA
SSR231	3	$(TA)_{10}$	SGN-E300800	TGCCAATCCACT CAGACAAA	TGGATTCACCA AGGCTTCTT	169	181/169	NA	n/n/n
SSR111	3	$(TC)_6(TCTG)_6$	None	TTCTTCCCTTC CATCAGTTCT	TTTGCTGCT ATACTGCTGACA	188	188/200	NA	y/y/y
SSR22	3	$(AT)_{11}$	SGN-E274235	GATCGGCAGT AGGTGCTCTC	CAAGAAACAC CCATATCCGC	218	206/218	NA	y/n/y
T794	3	NA	NA	CGCAAAATGGAC CTAGAAGC	GAGAAGC GGCAGAAATCTTG	1600	750/700 (*Hin*fI)	NA	NA

Table 2 (Contd.)

Marker[a]	Chromosome	Repeat type and length	SGN ID no.[b]	Forward primer (5′-3′)	Reverse primer (5′-3′)	Product size[c] (bp)	Polymorphic band sizes[c] (in bp)		Amplification
							LA716/E6203 (enzyme)	LA1589/ E6203 (enzyme)	St/Sm/ Ca[d]
T1621	3	NA	NA	GACTGGTGAGG ACGATGATG	CGGCAATC TCTTCGTCAAA	1300	600/550 (*Hin*fI)	600/550 (*Hin*fI)	NA
SSR320	3	$(AT)_{12}$	None	ATGAGGCAATC TTCACCTGG	TTCAGCTGAT AGTTCCTGCG	172	179/172	165/173	y/n/n
SSR14	3	$(ATA)_9$	SGN-E307729	TCTGCATCTGG TGAAGCAAG	CTGGATTGCC TGGTTGATTT	166	158/166	160/166	y/y/y
SSR11	3	$(CAG)_6$	SGN-E305953	CCTTCAATTGA CCTCCCTCA	GCATCTGGAAAT TAGAGGCG	235	228/235	NA	y/y/y
SSR300	3	$(TTC)_{10}$	None	AATGGCAGCT ATGATGAGCC	ACCCGACTTC ATTTCACCTG	650	590/650	NA	y/y/n
SSR27	3	$(TC)_6(ATGT)_2$	None	CCCAAATCA AGGTTTGTGGT	TCAGATGCCAC CACTCTCAG	149	161/149	NA	y/y/y
TG244	3	NA	NA	GCGATGACT AAGGCCATTCT	ATAAAGCAT GCCCACCAAAG	450	250 + 200/325 + 125 (*Eco*RV)	NA	NA
T707	4	NA	NA	TCGTGGATTA TGGGCTTCTT	GGTAAGGCTG CAACACATCA	458	400/460 (*Apo*I)	NA	NA
SSR43	4	$(TAC)_7$	SGN-E284323	CTCCAAATTGG GCAATAACA	TTAGGAAGTT GCATTAGGCCA	237	228/237	NA	y/y/y
SSR593	4	$(TAC)_7$	SGN-E284323	TGGCATGAACA ACAACCAAT	AGGAAGTTG CATTAGGCCAT	295	285/295	296 + 160/294 + 163	y/y/n
SSR603	4	$(GAA)_8$	SGN-E307794	GAAGGGACAATT CACAGAGTTTG	CCTTCAACTT CACCACCACC	251	Null/251	173/251 + 179	ND
SSR450	4	$(AAT)_7$	SGN-E226350	AATGAAGAACCA TTCCGCAC	ACATGAGCCCA ATGAACCTC	265	255/265	NA	y/n/n
SSR306	4	$(ATT)_7$	None	ACATGAGCCCA ATGAACCTC	AACCATTCCG CACGTACATA	258	248/258	NA	y/n/n
SSR310	4	$(TGA)_9$	None	GCGATGAGG ATGACATTGAG	TTTACAGGC TGTCGCTTCCT	149	139/149	NA	n/n/n
SSR638	4	$(GT)_9(AT)_8(AC)_{13}(GA)_{12}$	None	TGTTGGTTGGAG AAACTCCC	AGGCATTTAAA CCAATAGGTAGC	180	180/240	NA	y/n/n
T1405	4	NA	NA	CACCAACAAC TAGCCCTTGA	AAGCAATTC CTCCAGCTTCA	535	535/375 + 160 (*Apo*I)	NA	NA
cLEC7B23	4	NA	NA	GGAGAACACG GCTACCTCAG	AGCTGGAAA TGAGGTTTTGC	600	600/520 (*Cfo*I)	NA	NA
SSR146	4	$(AT)_7(CAT)_5$	SGN-E252704	TATGGCCATG GCTGAACC	CGAACGCCA CCACTATACCT	243	239/243	NA	y/n/n
SSR188	4	$(AT)_{11}$	None	TGCAGTGAGTC TCGATTTGC	GGTCTCATTGC AGATAGGGC	140	148/140	138/140	y/n/n
cLEX13I3*	5	NA	NA	TGAAGAAGCTA GTCATGGCAAC	TCATCCACGATC TCGATTACC	800	450 + 350/600 (*Taq*I)	400/430 (*Rsa*I)	NA
SSR13	5	$(AAG)_6$	SGN-E307674	GGGTCACATACA CTCATACTAAGGA	CAAATCGCGAC ATGTGTAAGA	102	97/102	NA	y/y/y
SSR115	5	$(AT)_{16}$	SGN-E331883	CACCCTTTATT CAGATTCCTCT	ATTGAGGGTA TGCAACAGCC	211	211/223	197/213	y/n/n

T1640	5	NA	NA	TGGACAGACCCT GATTCCAT	TTCCGACTGGGA AAAGAAGA	2000	2000/1300 + 700 (*Kpn*I)	NA	NA
T1601	5	NA	NA	TGCAACGACGT ACAATGAGG	GCATATTTACG AGCCGCATC	800	800/700 (*Hae*III)	NA	NA
CLEX13G5	5	NA	NA	GCAGCATTTAG GCTCAGAGG	TTCTCAGTGG ATCGTGATGG	420	420/300 + 120 (*Eco*RI)	NA	NA
TG23	5	NA	NA	GGTGGTTCAAA TCCTTATTGG	AAAGTGTTGGGG TAACGCAC	1700	1600/1200 (*Hae*III)	1,700/900 + 800 (*Bgl*II)	NA
SSR590	5	$(TC)_6(AC)_4$	SGN-E362409	TCTCAAAGTCG TTCCTTCTTGA	GGAAGAGAAA CGCGGACATA	161	161/158	NA	y/n/n
T633	5	NA	NA	GATGGGCTATG CTTGCTGTT	ACATCCCCAATG TTGTTGTG	600 + 375	375/350 (*Mse*I)	NA	NA
SSR47	6	$(AT)_{14}$	SGN-E277474	TCCTCAAGAAATG AAGCTCTGA	CCTTGGAGATAA CAACCACAA	191	180/191	NA	y/y/y
T892*	6	NA	NA	TGGCTCTTCGGA CTTTAGTGA	AGCACCTTCT GCGTTCATCT	1200	700/1200 (*Taq*I)	NA	NA
T507	6	NA	NA	CCTTTTATCT CCTCCGGTGT	TCTGTCCACT CACATGGATCA	800	800/700 (*Dra*I)	NA	NA
SSR128	6	$(CAG)_6(CAA)_3(CAG)_7$	SGN-E312046	GGTCCAGTTCA ATCAACCGA	TGAAGTCGTCT CATGGTTCG	123	123/106	NA	y/y/y
SSR578	6	$(AAC)_6(ATC)_5$	SGN-E344355	ATTCCCAGCA CAACCAGACT	GTTGGTGGATG AAATTTGTG	294	289/294	292/295	y/y/n
TG275	6	NA	NA	CCTGCAGCTTC TGTTTTGTG	ATCAGCTCAA TGCTGCAAGT	480	250/320 (*Rsa*I)	NA	NA
TG279*	6	NA	NA	GTAGAATCCG CTGTCGCTTC	TATGTCCACGAA GTCGGTGA	750	750/450 + 300 (*Dra*I)	350/320 (*Taq*I)	NA
T1053	6	NA	NA	AACGGTGATT CTCGCTGAGT	GCAACCACATC TTCCTCCTC	1800	950/1100 (*Hin*fI)	NA	NA
SSR350	6	$(AT)_{13}$	None	GGAATAACCTC TAACTGCGGG	CGATGCCTTCA TTTGGACTT	267	256/267	268 + 146/270 + 150	y/y/n
SSR241	7	$(AAT)_{13}$	None	TCAACAGCATAGT GGAGGAGG	TCCTCGGTAATTG ATCCACC	200	185/200	NA	y/n/n
SSR285	7	$(TTAT)_2(AT)_6$	None	AGTGGCTCTCA CCTACTGCG	CAATTCTCAG GCATGAAACG	276	265/276	273 + 159/277 + 159	y/n/n
SSR52	7	$(AAC)_9$	SGN-E360594	TGATGGCAGCA TCGTAGAAG	GGTGCGAAGG GATTTACAGA	202	202/190	NA	y/n/n
FW7	7	NA	NA	ACAGCCAGACCC TTCTCATACT	GGATCCTAAA AGAATGTGCAGT	750	750/700 (*Msp*I)	NA	NA
TG252*	7	NA	NA	GCCAGTGTCAC TTTTATTTTGC	GCTCCGTCTGA CCTATCTCG	1500	600 + 400/600 + 500 (*Alu*I)	400/370 (*Rsa*I)	NA
TG174*	7	NA	NA	TTCCAAGATCTT TTAGCGTCTC	CTGTTGCGGA TGTGATCATT	1500	900 + 600/1500 (*Msp*I)	750/1500 (*Rsa*I)	NA
TG217	7	NA	NA	CGTTGCTTCC TGATCCTACC	AGCTAGTGATG ATCCTGGCG	800	750/600 (*Taq*I)	NA	NA
SSR557	7	$(ATCT)_7$	None	GCCACAAGAAAC ATTGCTGA	TACGCGCACG TGCATAAATA	239	Null/239	NA	ND
SSR45	7	$(AAT)_{14}$	SGN-E279510	TGTATCCTGGTG GACCAATG	TCCAAGTATC AGGCACACCA	246	275/248	NA	y/n/n
T1738	7	NA	NA	GGTTGGTATGG AAGGTCTGC	CGGCTTCCACCA GTGATATT	800	800/600 + 200 (*Dpn*II)	NA	NA

Table 2 (Contd.)

Marker[a]	Chromosome	Repeat type and length	SGN ID no.[b]	Forward primer (5′-3′)	Reverse primer (5′-3′)	Product size[c] (bp)	Polymorphic band sizes[c] (in bp)		Amplification
							LA716/E6203 (enzyme)	LA1589/ E6203 (enzyme)	St/Sm/ Ca[d]
T1255	7	NA	NA	TTTGCTTTGCT TCTCCTTCA	ATTCAACTCGA GCAACGTCA	400	350 + 220/320 + 200 (*Alu*I)	NA	NA
CT114	7	NA	NA	ATTGAAGAAT GGCGGTGAAG	ATGCCAACTTCT TGGCAAAC	1125	800 + 325/800 + 300 (*Dra*I)	NA	NA
SSR344	8	$(AT)_{12}$	None	TGTTGCTCGAA CTCTCCAAA	CATAGGAGAG GTAACCCGCA	275	245/275	NA	n/n/n
SSR244	8	$(TA)_{14}$	SGN-E85210	GCGATGGTCTGA GACACTGA	CAGCTGGTGA TCCTCCTCTT	193	Null/193	NA	ND
SSR327	8	$(AAT)_7$	None	TCAGGATCAGG AGCAGGAGT	TGGACTTGTTC CATGAACCC	195	180/195	NA	y/n/n
TG302	8	NA	NA	CTCTCCGGGTG GCTATTACA	TCTTGGGACT CCTCCTTTTCT	750	600 + 150/350 + 250 + 150 (*Alu*I)	600 + 150/400 + 150 (*Alu*I)	NA
SSR38	8	$(TCT)_8$	SGN-E285849	GTTTCTATAGCTG AAACTCAACCTG	GGGTTCATCAAA TCTACCATCA	237	215/237	NA	y/y/n
SSR594	8	$(TCT)_8$	SGN-E285849	TTCGTTGAAGAAG ATGATGGTC	CAAAGAGAAC AAGCATCCAAGA	293	273/293	288/294	y/y/n
TG510	8	NA	NA	ATGGTGATGC TCGCTAATCC	CTCTTGCGTGA TTGCAGGTA	1800	900 + 400/1100 + 380 (*Dpn*II)	1,500 + 1,400 + 1,100 + 420 + 200/1500 + 1,100 + 420 + 100 (*Rsa*I)	NA
T1359	8	NA	NA	TTTGAGAG GCATGA TGGTCA	TCCCACCGGT TAAACTCATC	850	515 + 335/400 + 335 (*Aci*I)	NA	NA
CT68	8	NA	NA	CAGACGATTG TCCGAGATCA	TGTTGAGCGCA TTTGAAGAG	550	375 + 425/550 (*Apo*I)	NA	NA
TG294	8	NA	NA	ATTGGCTGC AATGATGGATT	CTAAGCAGG ACGGCCATCTA	900/800	900/800 (*Alu*I)	NA	NA
TG254*	9	NA	NA	GACTTCGGGGC AATTATCTG	AAACGAGCACT GCATTCATG	1700	1700/1200 (*Dra*I)	400/430 (*Eco*RI)	NA
TG18*	9	NA	NA	AAGGGTTGTT GATTCCGTCA	GCACCAGGTT TTCCATCTGT	468	200/468 (*Rsa*I)	270/300 (*Rsa*I)	NA
SSR73	9	$(AG)_2(AGA)_7$ $(TAGTGA)_2$	SGN-E344864	TGGGAAGATCC TGATGATGG	TTCCCTTTC CTCTGGACTCA	700	1300/700	NA	y/n/n
TG223	9	NA	NA	CAAGAAAATAT TGTGTAGTGT TCTCCA	TCCCCCTCT TCATCAAATTC	700/800	700/800	NA	NA
SSR69	9	$(TAT)_7$	SGN-E226973	TTGGCTGGA TTATTCCTGTTG	GCATTTGA TAGAAGG CCAGC	127	Null/127	NA	ND
SSR70	9	$(AT)_{20}$	SGN-E346864	TTTAGGGTGT CTGTGGGTCC	GGAGTGCGCAG AGGATAGAG	120	93/120	NA	y/y/y
SSR237	9	$(AT)_{11}$	None	GTGGTAACGG CAAAGGGACT	CTTATGGCCTT AGCAGCCAG	191	160/191	NA	y/n/n
SSR19	9	$(AT)_{16}$	SGN-E304225	CCGTTACCTTGG TCCATCAC	GGGAGATGCC ACATCACATA	186	174/186	178/186	y/n/n

SSR155	9	$(TAT)_9$	None	TACATGGTGC TCAAACTCGC	TGGTTGAGAA GGACAGGTGA	200	Null/200	NA	ND
SSR383	9	$(AT)_{11}$	SGN-E223034	ATTGTACAAAG ACCCGTGGC	GTTGCACACTG GATCAATGC	188	188/230	NA	y/n/n
T1190	9	NA	NA	GCGTTCTCGT TACTGGTGCT	GTTGCATGGT TGACATCAGG	1800	1800/1100 + 700 (*Dra*I)	900 + 700 + 500/1300 + 700 + 600 + 400 (*Rsa*I)	NA
T1519	9	NA	NA	TGCACAGACAC AAACTGCAA	CACCCTGGTAA TGCCAAACT	466	466/400 (*Hin*cII)	NA	NA
T156	9	NA	NA	GCGGTTGATTC ACATCGTAA	CCTGTAGCACC CAAAGGATG	1000	500/750 (*Msp*I)	450/750 (*Dpn*II)	NA
T1065	9	NA	NA	GACGGTGAAGG GTACCAAGA	CAGGAGT GCATGGGTA GGTT	407	350/200 (*Tsp*509I)	NA	NA
TG230	10	NA	NA	TTGCAGAAGC AACCCTTGAC	TACTTCTCC CCATTCCATGC	1200	470/900 (*Dpn*II)	NA	NA
SSR4	10	$(CGG)_7$	SGN-E311098	TTCTTCGGAG ACGAAGGGTA	CCTTCAATC CTCCAGATCCA	168	181/168	NA	y/y/y
TG303*	10	NA	NA	CGTAAAGGGTT GTTCTTGTGC	TGTTTTCG AGTGGGGTTCAT	400	300/280 (*Alu*I)	450/500 (*Hae*III)	NA
SSR34	10	$(GA)_7(TCGA)_2$	None	TTCGGATAAAG CAATCCACC	TCGATTGTG TACCAACGTCC	184	186/184	192/184	y/y/y
SSR218	10	$(TCA)_7$	SGN-E265605	GTGGTTATCC CAAGACCCAA	CGCCAGTC TTCCTCTGACTT	127	143/127	NA	y/y/y
SSR301	10	$(TTGGT)_2(TA)_{15}$	None	TTTCCACCT CAAACCACTCC	CCCTTTGAC CTGTGCCA	181	171/181	NA	y/y/y
SSR248	10	$(TA)_{21}$	SGN-E330194	GCATTCGCTG TAGCTCGTTT	GGGAGCTTCA TCATAGTAACG	251	221/251	NA	y/y/y
SSR85	10	$(TAA)_7$	SGN-E248079	ATCCGTTAGC TATTGTGCCG	TTGCCATGC ACTTATCTTCG	179	170/179	NA	y/y/n
T1682	10	NA	NA	CCTCCCTCACC ATCCAATTA	CTGCTTTA ACCACCGGATTC	804	804/490 + 314 (*Eco*RV)	NA	NA
SSR223	10	$(TCT)_7$	None	TGGCTGCCTC TTCTCTGTTT	TTTCTTGAA GGGTCTTTCCC	191	190/193	198/192	y/y/n
TG233	10	NA	NA	CATGCCTTTT TCTTGGGATG	TGGAACCCCTT TAACTGTGC	500	430/500 (*Alu*I)	500/450 (*Hin*cII)	NA
TG497	11	NA	NA	CGGAGAGTGA AGATGCATTG	AAGTTCCAGA GGGAGCACAA	800	400 + 300/500 + 300 (*Taq*I)	500 + 350/500 + 290 (*Taq*I)	NA
SSR136	11	$(CAG)_7$	SGN-E206926	GAAACCGCCT CTTTCACTTG	CAGCAATGAT TCCAGCGATA	149	137/149	NA	y/y/y
SSR80	11	$(TTTCAA)_2$ $(GTACAA)_2$ $(CAA)_7$	SGN-E340444	GGCAAATG TCAA AGGATTGG	AGGGTCATGT TCTTGATTGTCA	186	175/186	NA	y/y/y
SSR67	11	$(AGA)_2(AAG)_7$	SGN-E232901	GCACGAGAC CAA GCAGATTA	GGGCCTTTCC TCCAGTAGAC	900	1000/900	NA	n/n/n
SSR76	11	$(CGG)_7$	SGN-E345310	ACGGGTCGTC TTTGAAACAA	CCACCGGATT CTTCTTCGTA	199	220/199	NA	ND
SSR46	11	$(AT)_{14}$	SGN-E277505	CCGAGGCG AATCTTGAATAC	GCACCATCT CTTGTGCCTCT	375	400/375	NA	y/y/y

Table 2 (Contd.)

Marker[a]	Chromosome	Repeat type and length	SGN ID no.[b]	Forward primer (5′-3′)	Reverse primer (5′-3′)	Product size[c] (bp)	Polymorphic band sizes[c] (in bp)		Amplification
							LA716/E6203 (enzyme)	LA1589/ E6203 (enzyme)	St/Sm/ Ca[d]
TG400*	11	NA	NA	TCCAAATCCAC CACCTATCC	AGCATTGCT CCCTGCTAAAG	404	404/300 +104 (*Hin*fI)	220/250 (*Bgl*II)	NA
CLEC24C3	11	NA	NA	AGATCGGCAAA TGATCCAAG	ACTTGTGGC GAAAAATGAGG	1250	1250/950 (*Eco*RI)	NA	NA
T302	11	NA	NA	TGGCTCATC CTGA AGCTGATA GCGC	AGTGTACATCC TTGCCATTGACT	950/850	950/850	600+300/600 +200 (*Msp*I)	NA
CLET24J2	11	NA	NA	CAACCATCCT AGC AATGAAATCT	GAGGCATTCA CTCTCTTCGATAC	400	400/220 +180 (*Kpn*I)	NA	NA
TG393	11	NA	NA	TGGATTTGAT TAGCCGAAGG	CCAAGAA TCCCAGAAGGAGA	750	325/400 (*Dpn*II)	500+350/500 +400 (*Dpn*II)	NA
TG180	12	NA	NA	TCTCAGTGGA CTAAGGGGTCA	TCACAGCAG ACATGTCGGAC	1000	1000/900 (*Dra*I)	1000/900 (*Dra*I)	NA
T659	12	NA	NA	AATGGTCATG GAATGGGAAA	TGCCAAGGATA AAGGAGCAG	1100	900/1100 (*Acc*I)	NA	NA
TG68	12	NA	NA	TCCACCTAGG ATGAGTTTGGA	CATGTCAAGG GGATTGAACA	440	200/440 (*Dpn*II)	200/440 (*Dpn*II)	NA
cLET8K4	12	NA	NA	CACTTTGTG GCAATCGACAT	TGCCTTATGC CAAACAGAAA	1100	900/1000 (*Dpn*II)	800+500/500 +350 (*Taq*I)	NA
T1736	12	NA	NA	ATTCTCGATCA ACGGACCAC	ACACTGAGC AATGCGAATCA	1050/1200	1050/1200	NA	NA
SSR20	12	$(GAA)_8$	SGN-E303929	GAGGACGACAA CAACAACGA	GACATGCCACT TAGATCCACAA	157	148/157	NA	y/y/y
SSR124	12	$(CACC)_2(GA)_7$	SGN-E323915	TCAATCCATCA CACCTTGGA	GAGGAAGAA GACCACGCAAA	1300	1270/1300	NA	y/n/n
SSR44	12	$(GA)_{54}$	None	TCATCTGCAAT TCATGGCTC	AGGTCAAGGA TGTGCTTCCC	238	238/180	NA	ND
T801	12	NA	NA	GAGCCGGAA GGTATGATTGA	GACCTTGT GGTAGGGCATGT	589	600/400 (*Acc*I)	NA	NA
T1305*	12	NA	NA	CCTTGTACCC CCAAATCTGA	CCTGAATTTT GGCTCCAGAC	1250	1250/1050+200 (*Eco*RI)	230/200 (*Hae*III)	NA
T800	12	NA	NA	GCCTCGAAAAT TGCAGAAGA	GCGTGTCTTGC AATCACATT	600	420+180/600 (*Eco*RV)	NA	NA

[a]Underlined markers are also polymorphic for *Solanum lycopersicum* (E6203) and *S. pimpinellifolium* (LA1589). Markers followed by an asterisk require different primers to detect the LA1589/E6203 polymorphism. These primer sequences can be found on SGN

[b] SGN ID no., Sequence identity number from the Solanaceae Genomics Network

[c]The PCR product sizes for the listed primers, polymorphic band sizes and restriction enzymes for each comparison (*S. pennellii* (LA716)/E6203 and LA1589/E6203) are also included

[d]Amplification results in potato (St), eggplant (Sm), and pepper (Ca). NA, Not applicable; ND, not done; y, amplification; n, no amplification

10 m *M* Tris-HCl, pH 8.3, 1.5 m *M* $MgCl_2$), 0.2 m *M* dNTPs, and 0.5 U *Taq* polymerase. Amplification consisted of an initial denaturation for 5 min at 94°C, followed by 35 cycles of amplification with denaturation at 94°C for 30 s, annealing at 50°C or 55°C for 45 s, and extension at 72°C for 45 s, with a final cycle of 72°C for 5 min. Following amplification, products were separated on 2–3% agarose gels in 1× TAE buffer for 3–5 h at 100 V and 4°C.

Mapping of microsatellites

The 89 SSRs showing single-band polymorphism between TA209 and LA716 on regular agarose gels were then mapped on the 83 individuals of the F_2 mapping population. The same population was also used to map an additional 51 SSRs that could only be discriminated on the sequencer. The methods used for amplification, separation, and data analysis of these SSRs are described below. Sequence, polymorphism and other information for these markers can be found on the SGN website (http://www.sgn.cornell.edu). Linkage analysis was performed using MAPMAKER ver. 2.0 software (Lander et al. 1987). The GROUP and ORDER commands were used to define linkage groups and the linear order of markers within groups. The RIPPLE command was used to confirm marker orders at LOD≥3.0. Genetic distances were calculated using the Kosambi mapping function (Kosambi 1944).

Development and mapping of CAPs markers

Newly developed CAPs and CAPs for a known QTL and its homolog (*fw2.2*, Nesbitt and Tanksley 2002; *fw7*, unpublished) were used to fill the gaps between SSRs. Thus, previously mapped tomato RFLP and EST markers (Fulton et al. 2002) were selected at intervals of 20 cM or less between SSRs to be converted into CAPs assays. Primers for the markers were designed using PRIMER3 (Rozen and Skaletsky 1997) and synthesized by Integrated DNA Technologies (Coralville, Iowa). The PCR of each marker was performed on both parents (TA209 and LA716) in 100-µl reactions containing 100 ng DNA, 10 pmol of each forward and reverse primer, 1× PCR buffer (50 m *M* KCl, 10 m *M* Tris-HCl, pH 8.3, 1.5 m *M* $MgCl_2$), 0.2 m *M* dNTPs, and 1 U *Taq* polymerase. Thermocycling was as described for the microsatellites. The PCR products of the parent lines were digested with different restriction enzymes including *Alu*I, *Eco*RI, *Eco*RV, *Dra*I, *Hae*III, *Rsa*I, *Hin*fI, *Msp*I, *Kpn*I, *Taq*I, *Dpn*II, and *Tsp*509I and electrophoresed through 2% agarose to identify polymorphisms. The CAPs markers were then assayed on the mapping population as described for the SSRs. Sequences of the primers used for the CAPs assays are listed in Table 2 and are available on the SGN website (http://www.sgn.cornell.edu).

Survey of SSR polymorphism in tomato and wild relatives

The SSRs that amplified single bands for both TA209 and LA716, including all of the loci mapped on agarose gels, were surveyed for polymorphism on the 19 cultivated and wild tomato accessions as described by Matsuoka et al. (2002). Forward primers were labeled with HEX, FAM, or TET fluorescent tags (ABI Biosystems, Foster City, Calif.). After amplification, 0.5-µl samples of PCR product were mixed with 0.1 µl GeneScan 500XL size standard (PE Biosystems, Foster City, Calif.) and 1.0 µl loading buffer, heated at 92°C for 5 min, then placed on ice. The denatured samples (0.6 µl) were loaded on 5% denaturing (6 *M* urea) 36-cm LongRanger (FMC BioProducts, Rockland, Me.) gels in 1× TBE buffer and electrophoresed for 3 h at 3,000 V on an automated PE Biosystems, model 377 DNA Sequencer. Fragment sizes were determined based on migration relative to the size standards using GENESCAN ver. 2.1 software. Data were scored and compiled using GENOTYPER ver. 3.0 (PE Biosystems). For each SSR marker, fragments were classified into bins representing different alleles using the method of Matsuoka et al. (2002). In general, fragments differing by 1 bp or more in length were considered to be distinct alleles. Heterozygous individuals were characterized by the presence of two distinct amplification products. The polymorphism information content (PIC) value for each marker was determined for the cultivars (*S. lycopersicum* var. *esculentum*), for the *S. lycopersicum* var. *cerasiforme* accessions, for the wild species, and for all 19 accessions combined. These values were calculated as previously described (Saal and Wricke 1999):

$$\mathrm{PIC} = 1 - \sum_{i=1}^{k} p_i^2$$

where p_I is the frequency of the *I*th allele, and *k* is the total number of different alleles for the locus. Statistical analyses were performed with STATVIEW software for MacIntosh (SAS Institute, Raleigh, N.C.).

Results

Genetic map

A total of 152 PCR-based markers were mapped on 83 F_2 individuals from the cross *S. lycopersicum* LA925 × *S. pennelli* LA716 using standard agarose gels. An additional 51 SSRs were also mapped in the population using the sequencer. These markers are shown in Fig. 1, which also includes a framework of markers from the high-density map developed from

the same population (http://www.sgn.cornell.edu). Included among the agarose gel-based PCR markers were 76 SSRs and 76 CAPs. Sixty-nine of these SSRs were identified from tomato ESTs and seven from potato. Forty-two of the CAPs were based on EST markers, 32 were derived from RFLP markers (29 derived from genomic DNA and three from cDNAs), and two were for a known QTL and its homolog: *fw2.2* on chromosome 2 and *fw7* on chromosome 7. Most of the mapped SSRs were trinucleotide repeats (47%), with fewer dinucleotide (26%) and compound (24%) repeats and only two tetranucleotide repeats. All but two of the dinucleotide repeats (90%) had the AT motif, whereas both ATT (33%) and CTT (30%) motifs comprised the majority of the trinucleotide repeats. Different repeat motifs appeared to be randomly distributed across chromosomes with no obvious clustering of particular motifs.

The resulting map covered 1,397 cM, which corresponds to 95% of the total distance encompassed by the high-density map (Fulton et al. 2002). When compared with this previous map, marker coverage on each chromosome ranged from 86% for chromosome 7 to 100% for chromosomes 1, 4, and 10. Each chromosome contained from 9 (chromosomes 5 and 6) to 18 (chromosome 3) agarose gel-based PCR markers, with an average genetic distance of 10.0 cM between these markers. Overall, 84% (128) of the markers were 20 cM or less apart, and 70% (106) were 15 cM or less apart. The largest gap between markers was the 33.5 cM between SSR146 and SSR188 on chromosome 4. A total of 122 of the agarose gel-based PCR markers were also mapped on the *S. pennelli* ILs (Eshed and Zamir 1995; Fig. 1).

With the exception of 13 SSRs that gave unclear results on the polymorphism survey, all of the mapped markers were also checked for polymorphism between *S. lycopersicum* E6203 and the closely related species *S. pimpinellifolium* LA1589. *S. lycopersicum* × *S. pimpinellifolium* LA1589 is a common cross used for mapping of both major genes and QTLs (Grandillo and Tanksley 1996b; Tanksley et al. 1996; Bernacchi et al. 1998b; Ku et al. 1999; van der Knapp and Tanksley 2001; Doganlar et al. 2002; van der Knapp et al. 2002). In all, 46 (73%) of the SSRs and 28 (37%) of the CAPs showed polymorphism for this species combination (includes the underlined markers in Fig. 1). Of the 46 polymorphic SSR markers, ten had null genotypes in LA1589 and one had a null phenotype in E6203. The SSR markers were also examined for polymorphism between E6203 and *S. habrochaites* LA1777, another wild species that has been the subject of many mapping and QTL studies (Bernacchi et al. 1998a, b; Monforte and Tanksley 2000a, b). Fifty-nine (94%) of the markers were polymorphic, including nine markers with null genotypes in LA1777 and one with a null genotype in E6203. In addition, 57 of the CAPs markers were surveyed on *S. lycopersicum* and *S. habrochaites* LA1721, and 48 (84%) were polymorphic.

Fig. 1 Molecular linkage map of the tomato genome showing the positions of the PCR-based markers in *bold* within the framework of the high-density map (http://www.sgn.cornell.edu). The PCR-based markers to the *right* of each chromosome were distinguished on agarose gels, those to the *left* were distinguished on a sequencer. *Underlined* markers are polymorphic between *Solanum lycopersicum* and *S. pimpinellifolium*. *Underlined* markers that are not in *bold* are PCR markers developed especially for mapping in *S. lycopersicum* × *S. pimpinellifolium* (LA1589) populations and were not mapped in the *S. pennelli* F_2 population. (Primer sequences and other information for these markers can be found on the SGN website.) *Vertical lines* drawn to the right of each chromosome show the positions of introgressions contained in the *S. pennelli* introgression lines (IL). *Dashed horizontal lines* indicate which markers were mapped in the ILs. Microsatellite markers are prefixed *SSR*. The CAPs markers are named according to the RFLP marker or gene from which they were derived. Map scale is in centiMorgans (cM), and approximate positions of the centromeres are indicated by *gray bars*

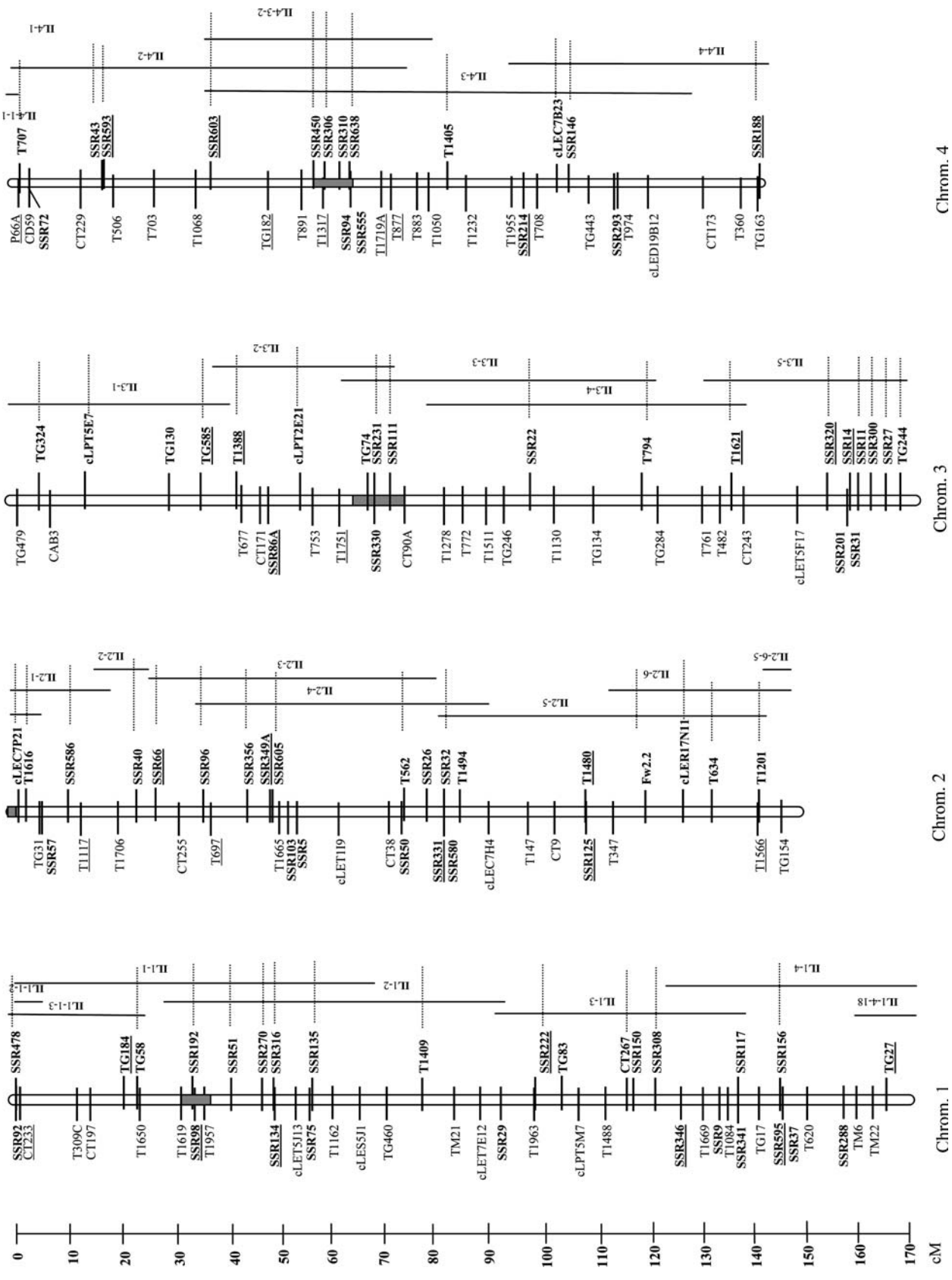
Characterization of SSRs

A total of 122 microsatellites were surveyed for polymorphism on seven tomato cultivars and 12 accessions of six different wild species using the automated sequencer (Table 1). This set of SSRs included the 76 markers mapped on agarose gels, 22 of the markers mapped using the sequencer, and an additional 24 unmapped markers. As indicated above, 13 of the mapped SSRs gave unclear results in the polymorphism study and were eliminated from the analysis. Of the 109 microsatellites that were analyzed, most (46%) were trinucleotide repeats, with nearly equal numbers of dinucleotide repeats (27%) and repeats containing more than one motif (compound repeats, 25%). Less than 3% of the SSRs were tetranucleotide repeats. The majority (83%) of dinucleotide microsatellites consisted of the TA/AT motif, with the remainder as TC/CT repeats. Similarly, among the loci with trinucleotide repeats, AAT/ATA and CTT/TCT were the most common motifs (24 and 26%, respectively), followed by TTG/TGT repeats (16%). The remaining trinucleotide motifs each comprised less than 8% of the total.

The total length of the SSR repeats ranged from 18 to 84 nucleotides, with an average length of 27 nucleotides. Most of the longest repeats were compound or AT repeats. For example, the longest locus used in the polymorphism study, SSR638, had the motif $(GT)_9(AT)_8(AC)_{13}(GA)_{12}$ (Areshchenkova and Ganal 1999). The longest dinucleotide repeat, $(AT)_{39}$, was 78 nucleotides long (SSR63), while the longest trinucleotide repeat, $(AAT)_{14}$, was 42 nucleotides long (SSR45).

SSR polymorphism and allelic variation among tomato accessions

Mean PIC values and number of alleles were calculated for the cultivars, *cerasiforme* lines, wild species


Chrom. 1
Chrom. 2
Chrom. 3
Chrom. 4
cM
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
SSR478
TG184
TG58
SSR192
SSR51
SSR270
SSR316
SSR135
T1409
SSR222
TG83
CT267
SSR150
SSR308
SSR117
SSR156
TG27
SSR92
CT233
T309C
CT197
T1650
T1619
SSR98
T1957
SSR134
cLET5J13
SSR75
T1162
cLES5J1
TG460
TM21
cLET7E12
SSR29
T1963
cLPT5M7
T1488
SSR346
T1669
SSR9
T1084
SSR341
TG17
SSR595
SSR37
T620
SSR288
TM6
TM22
cLEC7P21
T1616
SSR586
SSR40
SSR66
SSR96
SSR356
SSR349A
SSR605
T562
SSR26
SSR32
T1494
T1480
Fw2.2
cLER17N11
T634
T1201
TG31
SSR57
T1117
T1706
CT255
T697
T1665
SSR103
SSR5
cLET119
CT38
SSR50
SSR331
SSR580
cLEC7H4
T147
CT9
SSR125
T347
T1566
TG154
TG324
cLPT5E7
TG130
TG585
T1388
cLPT2E21
TG74
SSR231
SSR111
SSR22
T794
T1621
SSR320
SSR14
SSR11
SSR300
SSR27
TG244
TG479
CAB3
T677
CT171
SSR86A
T753
T1751
SSR330
CT90A
T1278
T772
T1511
TG246
T1130
TG134
TG284
T761
T482
CT243
cLET5F17
SSR201
SSR31
T707
SSR43
SSR593
SSR603
SSR450
SSR306
SSR310
SSR638
T1405
cLEC7B23
SSR146
SSR188
P66A
CD59
SSR72
CT229
T506
T703
T1068
TG182
T891
T1317
SSR94
SSR555
T1719A
T877
T883
T1050
T1232
T1955
SSR214
T708
TG443
SSR293
T974
cLED19B12
CT173
T360
TG163

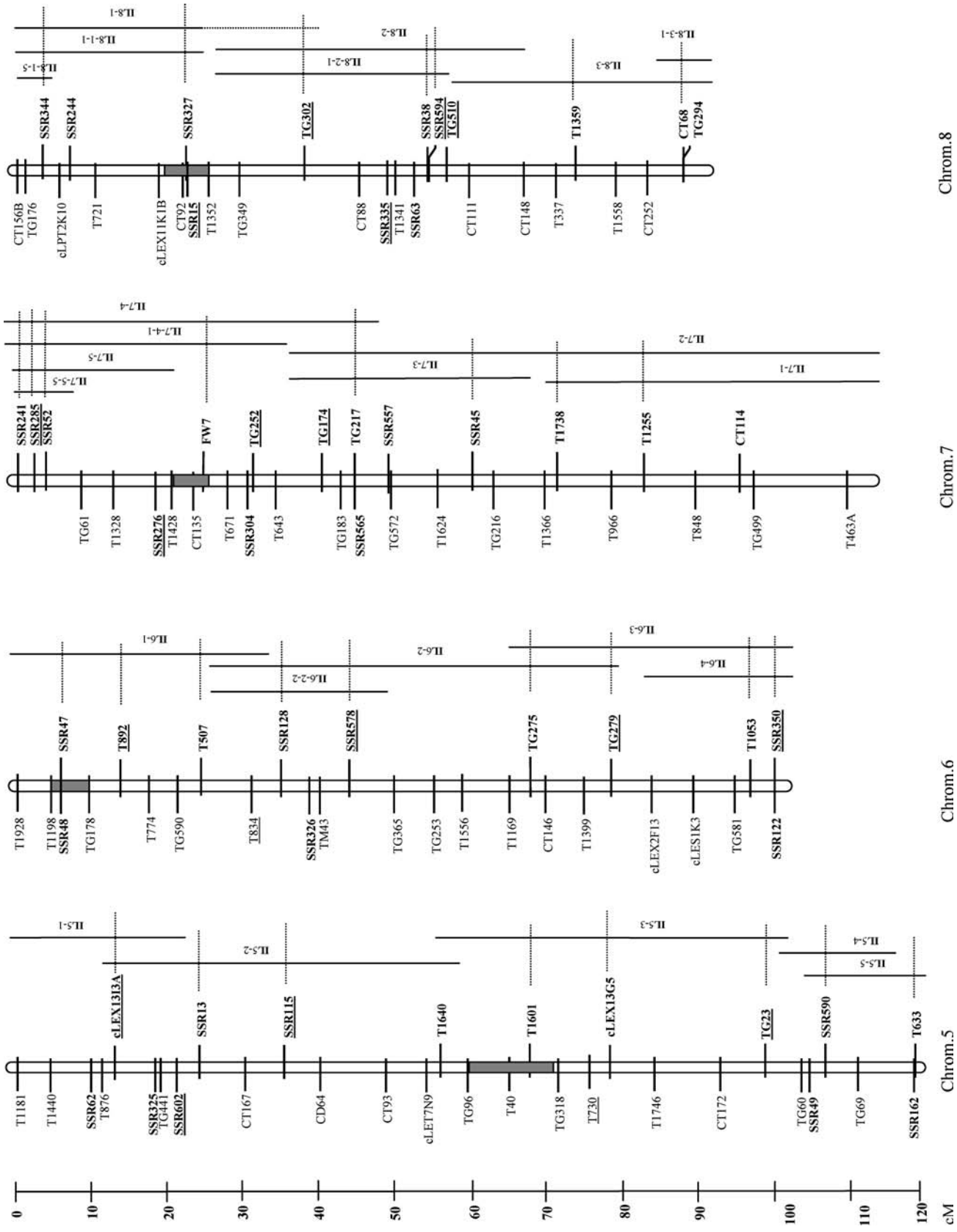

Chrom.8
Chrom.7
Chrom.6
Chrom.5
cM
0
10
20
30
40
50
60
70
80
90
100
110
120

Fig. 1 (Contd.)

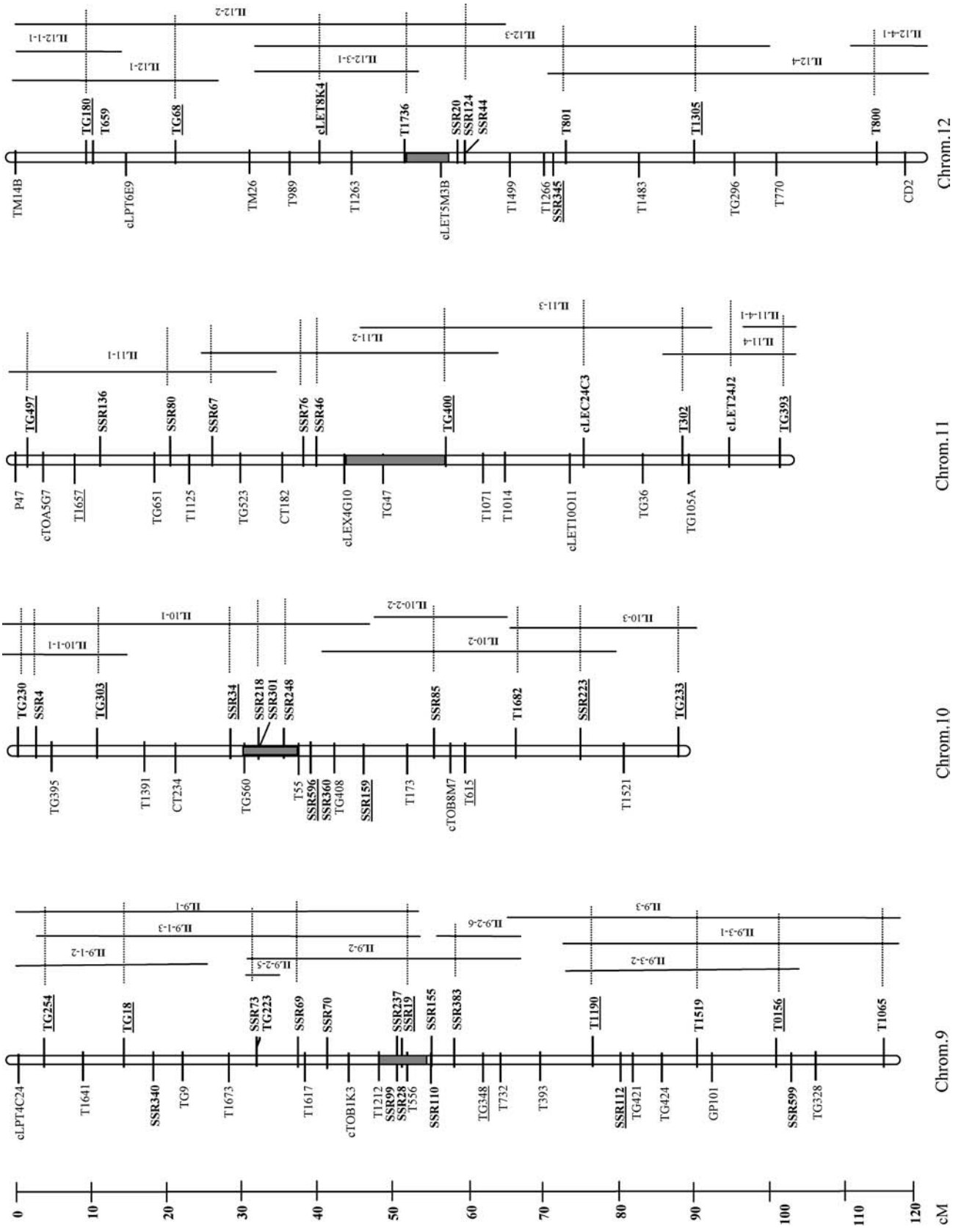

Chrom.12
TG180
T659
TG68
cLET8K4
T1736
SSR20
SSR124
SSR44
T801
T1305
T800
TM14B
cLPT16E9
TM26
T989
T1263
cLET5M3B
T1499
T1266
SSR345
T1483
TG296
T770
CD2
Chrom.11
TG497
SSR136
SSR80
SSR67
SSR76
SSR46
TG400
cLEC24C3
T302
cLET24J2
TG393
P47
cTOA5G7
T1657
TG651
T1125
TG523
CT182
cLEX4G10
TG47
T1071
T1014
cLET10O11
TG36
TG105A
Chrom.10
TG230
SSR4
TG303
SSR34
SSR218
SSR301
SSR248
SSR85
T1682
SSR223
TG233
TG395
T1391
CT234
TG560
T55
SSR596
SSR360
TG408
SSR159
T173
cTOB8M7
T615
T1521
Chrom.9
TG254
TG18
SSR73
TG223
SSR69
SSR70
SSR237
SSR19
SSR155
SSR383
T1190
T1519
T0156
T1065
cLPT4C24
T1641
SSR340
TG9
T1673
T1617
cTOB1K3
T1212
SSR99
SSR28
T556
SSR110
TG348
T732
T393
SSR112
TG421
TG424
GP101
SSR599
TG328
0
10
20
30
40
50
60
70
80
90
100
110
120
cM

Fig. 1 (Contd.)

accessions, and for all 19 lines combined (Tables 1, 3). The PIC values were lowest when comparisons were limited to *S. lycopersicum* var. *esculentum* or var. *cerasiforme* lines (0.17 and 0.25, respectively) and highest for the wild species (0.64). In fact, only 44% (61) of the SSRs were polymorphic among cultivars, whereas 99% of the loci were polymorphic among the wild species. Because many of the markers used for the polymorphism study were chosen for mapping, most (92%) exhibited polymorphism between *S. pennellii* and two standard *S. lycopersicum* var. *esculentum* lines, TA209 and TA1143. In addition, 88% of the SSRs exhibited polymorphism between the *S. lycopersicum* lines and *S. pimpinellifolium* LA1589 and between *S. lycopersicum* and *S. habrochaites* LA1777.

Overall, the 109 SSRs identified 172 different alleles in the cultivated species, thus giving an average of 1.6 alleles per locus with most SSRs having only one allele per locus (the mode; Table 3). Similarly, 186 alleles were identified in the *cerasiforme* accessions, with an average of 1.7 alleles per locus (mode = 1). In contrast, 501 alleles were detected in the wild accessions, which is an average of 4.6 alleles per locus (mode = 4); and 602 alleles were identified in all accessions combined, with an average of 5.5 alleles per locus (mode = 6). Within the cultivars, between one and four alleles were detected per locus, while as many as 13 alleles were detected for the comparisons across all species (Table 3). Forty-three (39%) of the loci had null alleles (i.e., no amplification in at least one accession), with a combined total of 92 null alleles in the 19 lines, representing 15% of the total number of alleles.

Most of the SSR loci (90%) had at least one accession-specific allele for the tested lines, with a total of 310 unique alleles identified for the 109 loci. This value corresponds to 16% of the total number of alleles sampled in the study and 51% of the different alleles identified in the 19 lines. In general, the wild species' accessions had the greatest number of unique alleles. Only 2% of the alleles detected in the cultivars (12 of 602 alleles in total) were unique, while 45% of those detected in the wild species (273 of 602) were unique. A portion of this difference may be attributed to the fact that fewer cultivars than wild accessions were sampled in the study. The *S. pennellii* accession had the greatest number of unique alleles—61 (Table 1). Overall, 59% of the alleles identified in this species were accession-specific (Fig. 2a) with *S. pennellii* accounting for 20% of the unique alleles identified in this study (Fig. 2b). *S. habrochaites* LA1777 had 54 accession-specific alleles, corresponding to 47% of the alleles identified in this species and 17% of the total number of unique alleles identified in all 19 accessions. *S. neorickii* and *S. peruvianum* had similar numbers of accession-specific alleles, with each accounting for approximately 23% of the alleles identified in each species and approximately 9% of the total number of unique alleles. Cultivars such as Yellow Pear, M82, and Monster (Giant Red) Tomato had very few unique alleles. Only 2% of the alleles identified in the cultivars were unique, with the cultivars contributing only 1% of the unique alleles identified in the study.

Table 3 Mean PIC values and numbers of alleles per locus (± standard error) for each of the comparisons

Sample	Mean PIC[a] (range)	Mean no. alleles[a] (range; mode)
Cultivars	0.17 ± 0.02 a (0–0.72)	1.6 ± 0.1 a (1–4; 1)
cerasiforme	0.25 ± 0.03 b (0–0.75)	1.7 ± 0.1 a (1–4; 1)
Wild accessions	0.64 ± 0.02 c (0–0.91)	4.6 ± 0.2 b (1–10; 4)
All lines	0.54 ± 0.02 d (0.05–0.90)	5.5 ± 0.2 c (2–13; 6)

[a]Values that are followed by a different letter are significantly different at $P \leq 0.05$ as determined by Fisher's least significant difference procedure

Most individuals were homozygous for SSR alleles, however, there were some exceptions. For example, the *S. habrochaites* accession was heterozygous for 22 SSRs (20%), and *S. peruvianum* LA1708 was heterozygous for 17 loci (16%) (data not shown). In all, 41 (38%) of the SSRs displayed heterozygous genotypes in at least one accession.

Predictors of SSR polymorphism

Based on ANOVA analysis, it was found that the number of nucleotides in the SSR repeat motif had a significant effect on PIC value (Table 4). Thus, loci with dinucleotide repeats had significantly higher PIC values than the trinucleotide SSRs for all four comparisons. However, the decrease in PIC value between the trinucleotide and tetranucleotide repeats (e.g., from 0.50 to 0.30 from the comparison across all lines) was not significant, perhaps because only three tetranucleotide SSRs were assayed. There was a significant negative correlation ($P < 0.0001$, $r = -0.42$ to -0.55) between the number of nucleotides per repeat unit and PIC value. The number of repeats and total length of the SSRs also had significant effects on PIC values. A greater number of repeat units was positively correlated with higher PIC values ($P < 0.0001$, $r = 0.42$–0.52). Similarly, longer SSRs tended to have higher PIC values for all three comparisons ($P = 0.001$, $r = 0.30$–0.38). No significant differences among PIC values were identified when SSRs were categorized based on motif (AT- or CT-based) or between PIC values for simple versus compound repeats (data not shown).

In addition to PIC value, the number of nucleotides in the repeat motif also had a significant effect on the number of alleles for a given locus. For all comparisons, dinucleotide SSRs had significantly more alleles than trinucleotide and tetranucleotide repeats (Table 5). The number of repeats and total length of the SSRs also had significant effects on the number of alleles identified.

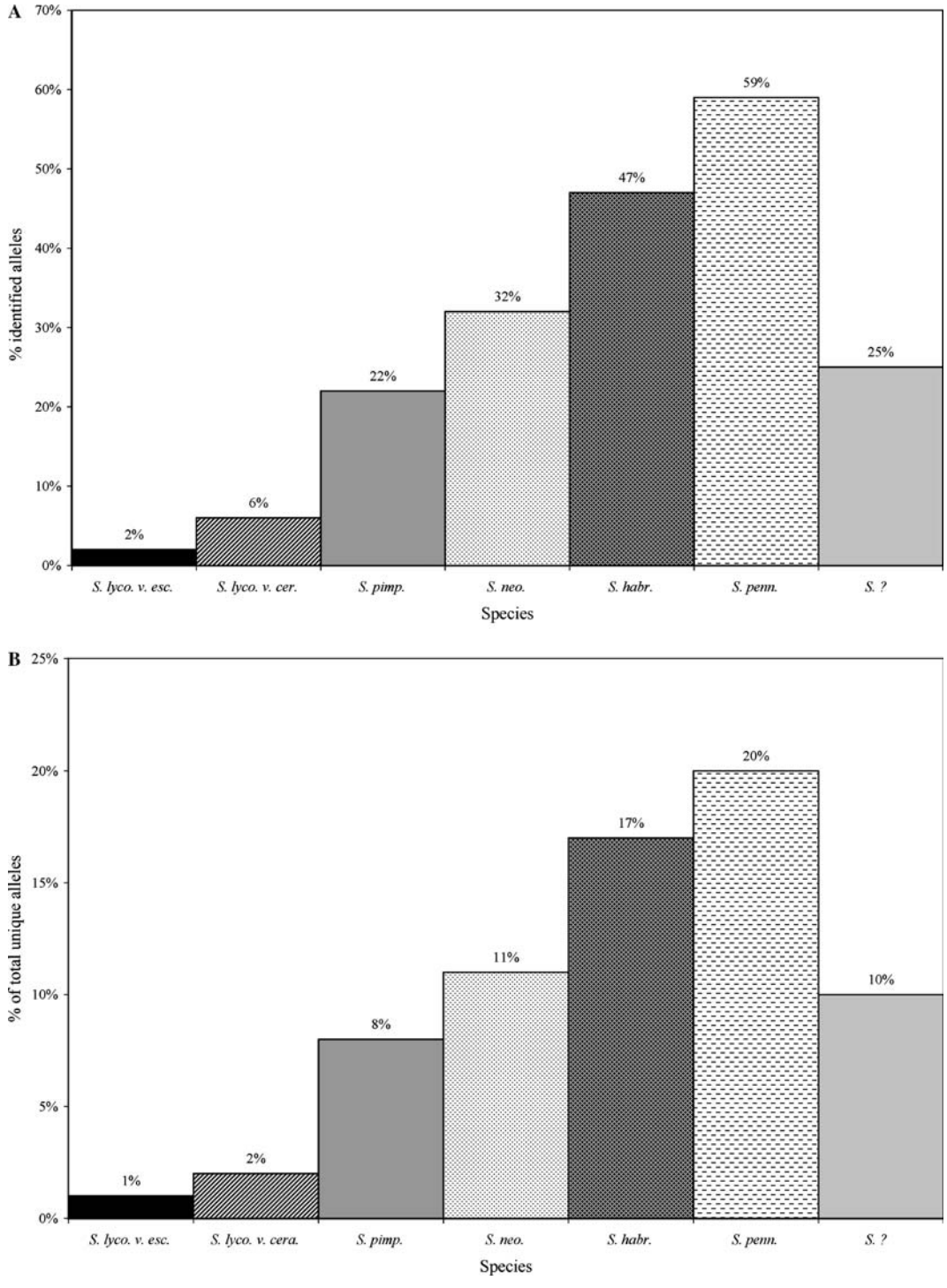


Fig. 2 Allelic variation of the 109 surveyed SSRs. **a** Percentage of the alleles identified for each species that were unique (accession-specific). Values were averaged across accessions for species that had multiple accessions sampled. **b** Contribution of each species to the total number of unique (accession-specific) alleles detected in this study. Values were averaged for species that had multiple accessions sampled

Thus, there were strong positive correlations between the number of alleles and the number of repeat units per locus ($P < 0.0001$, $r = 0.39$–0.54) and total length of the SSR ($P \leq 0.001$, $r = 0.31$–0.35). No significant differences among numbers of alleles were identified when SSRs were classed by motif or for simple versus compound repeats (data not shown).

Regression analysis indicated a highly significant, positive ($P < 0.0001$, $r = 0.44$) correlation between the PIC values for the cultivated and *cerasiforme* lines. However, there was no significant correlation between PIC values for the cultivated and wild accessions. The mapped markers were also classified based on genomic

Table 4 Relationship between number of nucleotides per repeat motif and PIC values

Repeat	PIC values[a]			
	Cultivars	*cerasiforme*	Wild accessions	All accessions
Dinucleotide	0.30 ± 0.04 a	0.38 ± 0.05 a	0.74 ± 0.03 a	0.68 ± 0.03 a
Trinucleotide	0.12 ± 0.03 b	0.19 ± 0.03 b	0.59 ± 0.03 b	0.50 ± 0.03 b
Tetranucleotide	0 ± 0 b	0 ± 0 b	0.53 ± 0.16 a,b	0.30 ± 0.10 b

[a]Within comparisons, PIC values (± standard error) that are followed by a different letter are significantly different at $P \leq 0.05$ as determined by Fisher's least significant difference procedure

Table 5 Relationship between number of nucleotides per repeat motif and number of alleles

Repeat	Number of alleles[a]			
	Cultivars	*cerasiforme*	Wild accessions	All accessions
Dinucleotide	2.0 ± 0.1 a	2.1 ± 0.2 a	5.7 ± 0.3 a	7.4 ± 0.5 a
Trinucleotide	1.4 ± 0.1 b	1.5 ± 0.1 b	4.2 ± 0.2 b	4.8 ± 0.3 b
Tetranucleotide	1.0 ± 0.0 b	1.0 ± 0.0 b	3.3 ± 0.7 b	3.3 ± 0.7 b

[a]Within comparisons, PIC values (± standard error) that are followed by a different letter are significantly different at $P \leq 0.05$ as determined by Fisher's least significant difference procedure

location as proximal to the centromere (within 10 cM of each side of the centromere) or telomere (within 10 cM of the chromosome ends) and interstitial. Together, the centromeric and telomeric locations were likely to contain heterochromatic DNA, while the interstitial regions were likely to contain euchromatic DNA. Based on ANOVA, mean PIC values for these different classes of markers were only significantly different for the comparison among cultivars. This comparison gave means of 0.10, 0.21, and 0.30 for interstitial, centromeric, and telomeric markers, respectively, with a significant difference ($P = 0.01$) between the PIC values for the interstitial and telomeric markers only.

Amplification of microsatellites in other solanaceous species

A set of 121 SSRs was tested for amplification in the related species potato, eggplant, and pepper. Most (92%) of the markers amplified a product in potato, while 58% and 34% amplified products in eggplant and pepper, respectively. Included among these markers were 68 of the SSRs that were mapped on agarose gels. The results of these amplifications are provided in Table 2. Of this subset of markers, 38% amplified products in all three species, while 60% amplified products in at least two species. The success of amplification appeared to be related to the genetic/phylogenetic distance between tomato and each of the other species. Thus, SSRs that did not amplify a product in potato, the closest relative of tomato, never amplified a product in either eggplant or pepper. Similarly, only one marker that amplified a product in pepper gave unsuccessful amplification in eggplant.

Discussion

PCR-based anchor map of tomato

The SSR and CAPs markers were used to develop a PCR-based anchor map of tomato using the same *S. lycopersicum* (LA925) × *S. pennellii* (LA716) F_2 population that was used to create the most recent high-density molecular map of tomato (http://www.sgn.cornell.edu). As compared to the high-density map, the PCR-anchor map, containing 152 markers, provides coverage of 95% of the genome, with a nearly uniform spacing of one marker every 10.0 cM and a maximum interval distance of 33.5 cM. These PCR-anchor markers are codominant, locus-specific, and highly reproducible. Moreover, they provide a faster, easier, and cheaper alternative to other types of markers because they are based on PCR, require little template DNA, and polymorphisms can be distinguished on ordinary agarose gels. Thus, they are economical and practical for use in a wide variety of research laboratories. Of these PCR-anchor markers, 122 were also used to characterize the *S. pennellii* IL lines (Eshed and Zamir 1995) that are now widely used in genetic mapping in tomato (e.g., Fridman et al. 2000, 2002; Ronen et al. 2000; Sela et al. 2001; Tadmor et al. 2002). Placement of the PCR-anchor markers on these IL lines should further facilitate their use in tomato genetics research. For example, the PCR-anchor markers can be used to verify the introgression contained in each of these stocks and as

flanking markers to select recombinants in fine-mapping studies.

In addition to their usefulness for mapping in populations derived from crosses between cultivated tomato and *S. pennelli*, many of the markers are also polymorphic for other wild species. For example, 44 of the agarose gel-based markers (16 SSRs and 28 CAPs) and 16 sequencer-distinguished SSRs have also been mapped in a *S. lycopersicum* E6203 × *S. pimpinellifoliu m* LA1589 population (underlined markers in bold in Fig. 1). An additional 15 CAPs markers have also been especially developed for this population (underlined markers in Fig. 1; primer sequence available on the SGN website). Although some of the SSR polymorphisms are due to null alleles and are therefore dominant, most (75%) are codominant and can be easily applied to fixed populations such as the *S. pimpinellifolium* inbred backcross lines developed by Doganlar et al. (2002). Similarly, 89% of the tested loci were polymorphic between *S. lycopersicum* and *S. habrochaites* and can be used for mapping in the *S. habrochaites* near-isogenic and inbred backcross lines described by Monforte and Tanksley (2000a). The null genotypes (presence/absence polymorphisms) were probably due to primer mismatch, especially in the wild species. It is possible that alternative primers could be designed to amplify both alleles of these markers, thereby rendering them codominant. A set of the SSR markers was also tested in the related solanaceous crop species, potato, eggplant and pepper. As expected, the proportion of markers that successfully amplified products in these species decreased with increasing evolutionary distance. Thus, potato gave the highest percentage of amplification, while pepper gave the lowest. These preliminary results suggest that a considerable portion of the SSR markers developed for tomato will be useful in related species.

Previous mapping of SSRs in tomato suggested that microsatellites, especially those with longer repeat and GATA motifs, tended to cluster at the centromeres (Broun and Tanksley 1996; Grandillo and Tanksley 1996a, b; Areshchenkova and Ganal 1999, 2002; Suliman-Pollatschek et al. 2002). Such clustering was not apparent in the present study. Different repeat motifs and lengths appeared to be randomly distributed along the chromosomes. This was not surprising because the SSRs used in this work were identified from ESTs and, therefore, many are associated with expressed genes, which are primarily found in euchromatin. In contrast, most of the previous mapping of SSRs in the tomato genome used markers that were identified from genomic DNA. Interestingly, however, it was found that, among the cultivars, markers that mapped to telomere proximal regions had significantly higher PIC values than those that mapped in interstitial regions ($P = 0.01$). This difference was not apparent in the wider cross-species comparisons and may reflect the greater homogenity of euchromatin in cultivated material relative to wild species.

Given their ease of use and adaptability to other tomato populations and solanaceous species, the PCR-anchor markers are ideal for various types of genetic studies, including qualitative and quantitative trait mapping. In addition, they are a valuable resource for breeding programs using marker-assisted selection. Because the map is set within a framework of COS and RFLP markers from the high-density map, chromosomal locations can be extrapolated between the two maps, thereby allowing the ready identification of additional markers for finer mapping of any gene of interest.

Survey of SSRs for polymorphism across tomato varieties and species

Of the 109 SSRs that were selected for the polymorphism study, TA/AT (22%) was the most frequent type of repeat, followed by CTT/TCT (12%) and AAT/ATA (11%). Several other studies have also reported the high incidence of AT-based repeats in tomato (Smulders et al. 1997; Areshchenkova and Ganal 2002; Suliman-Pollatschek et al. 2002; He et al. 2003), potato (Ashkenazi et al. 2001), and other plants (Morgante and Olivieri 1993; Wang et al. 1994). In addition, the present study identified a considerable proportion (25%) of compound or mixed repeats. Even higher percentages of compound SSRs were isolated when markers were preferentially selected to contain more than 30 repeat units (Areshchenkova and Ganal 1999).

Similar to previous studies, a high percentage (99%) of the SSRs was found to be polymorphic across all seven species, while only 44% were polymorphic within cultivated tomato (Suliman-Pollatschek et al. 2002; He et al. 2003). This level of genetic variation among tomato cultivars is quite high compared with other types of markers like RFLPs and RAPDs. For example, Miller and Tanksley (1990) found that only 20% of the RFLP markers were polymorphic among nine tomato cultivars. In addition, most of the SSR loci reported herein (88–92%) were polymorphic between *S. lycopersicum* and three wild species that have been widely exploited for mapping and QTL studies: *S. pennellii* (LA716), *S. pimpinellifolium* (LA1589, LA1246, LA411, and LA373) and *S. habrochaites* (LA1777). Such high levels of polymorphism have not been previously reported in tomato; however, the current data are somewhat biased because the initial selection criterion for 76 of the loci (i.e., the mapped SSRs) was polymorphism between *S. lycopersicum* and *S. pennellii*. Despite this caveat, the results suggest that many of the SSRs that have not already been mapped on the F_2 population may be useful PCR-based markers for the *S. pennellii* and other wild species-derived populations.

The PIC value of a marker is a measure of its informativeness for genetic studies. In this study, PIC values

for each marker were calculated among cultivars, among *cerasiforme* accessions, among wild species, and for all 19 accessions. In general, PIC values for the cultivars were quite low, with an average PIC of only 0.17. However, if only the 48 SSRs that were polymorphic among cultivars were considered, the average PIC value rose to 0.39. This value is comparable to that obtained by He et al. (2003), who obtained a PIC of 0.37 for 65 polymorphic loci tested on 19 tomato cultivars. Similarly, although the cultivars were chosen to represent diverse processing and freshmarket germplasm, the number of alleles detected among cultivars was low (1.6). This value rose to 2.3 alleles per locus when non-polymorphic markers were excluded. This result is quite similar to previously reported values of 2.5 and 2.7 alleles per locus (Areshchenkova and Ganal 2002; He et al. 2003). The PIC values for the comparisons across *cerasiforme* lines, wild species, and all accessions were significantly higher as were the numbers of alleles per locus. On average, 5.5 alleles per SSR were detected for the seven different species examined. In similar studies, three (Smulders et al. 1997; Suliman-Pollatschek et al. 2002) and 8.5 (Alvarez et al. 2001) alleles per locus were identified in four and ten different tomato species, respectively. Thus, the present results are within the range expected for such a comparison. Despite this general concurrence, our value is probably an underestimate for outcrossing species such as *S. habrochaites* and *S. peruvianum*, because only one individual for each accession was sampled. Surveying of multiple individuals in these species would undoubtedly reveal additional alleles.

The number of nucleotides in the repeat motif (dinucleotide and trinucleotide) was significantly correlated with both the PIC value and the number of alleles per locus such that SSRs consisting of dinucleotide repeats had significantly higher PIC values and numbers of alleles. Although such correlations have been reported in rice (Blair et al. 1999) and ryegrass (Jones et al. 2001), previous studies in tomato did not detect such a relationship (Smulders et al. 1997; Suliman-Pollatschek et al. 2002; He et al. 2003). Total number of repeats and total length of the SSRs were significantly and positively correlated with both PIC value and number of alleles per locus. This association between longer microsatellites and greater polymorphism has also been noted by several other researchers (Smulders et al. 1997; Areshchenkova and Ganal 1999; He et al. 2003). Because of their higher polymorphism, longer SSRs may be preferable for mapping studies. Although Areshchenkova and Ganal (1999) found that long (>30 repeat units) dinucleotide SSRs were clustered at the centromere, no relationship between microsatellite length and map location was detected in the present study.

A high proportion of the SSRs had accession-specific alleles, suggesting that these loci may be useful for germplasm identification. As expected, the species-mating system was a fairly good predictor of allelic diversity. Thus, allogamous, self-incompatible (*S. habrochaites* and *S. peruvianum*) and facultative self-compatible (*S. pennellii*) species had the highest numbers of unique alleles. Similarly, the two self-incompatible species were found to have the most heterozygous loci. With the notable exception of *S. neorickii*, which ranked third in total number of unique alleles, autogamous self-compatible species had fewer unique alleles and were rarely heterozygous. This relationship between mating system and genetic diversity of microsatellites was also reported by Alvarez et al. (2001).

Conclusions

We have developed a set of PCR-anchor markers, based on SSRs and CAPs, that covers the entire tomato genome at regular intervals and can be readily analyzed on standard agarose gels. This resource should be valuable for those wishing to map genes quickly, easily, and cheaply. Because the map is anchored in a high-density map containing more than 1,500 markers, identification of markers for fine mapping is also facilitated. In addition, we have surveyed 109 EST-derived SSRs for polymorphism in cultivated tomato and other species in the genus *Solanum*. The results indicate that many of the loci will be useful for mapping in wild species-derived populations and suggest that the markers may also be appropriate for germplasm fingerprinting/identification, studies of species relationships, taxonomy, and breeding. The data generated by this work, including an interactive map and primer sequences, are available on the SGN website (http://www.sgn.cornell.edu).

Acknowledgements We thank Dr. Amy Frary for valuable comments on the manuscript and Dane Rusçuklu and Evrim Balci for help with data analysis. This project was supported by grants from the National Science Foundation (DBI-0116076), the U.S. Department of Agriculture Plant Genome Program (9701552), and the Binational Agricultural Research and Development Fund (IS-3337-02).

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Alvarez AE, van de Wiel CC, Smulders MJ, 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 MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229–235

Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001) Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome 44:50–62

Bernacchi D, Beck-Bunn T, Eshed Y, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (1998a) Advanced backcross quantitative trait locus analysis of tomato. I. Identification of QTLs for traits of agronomic importance from *L. hirsutum*. Theor Appl Genet 97:381–397

Bernacchi D, Beck-Bunn T, Emmatty D, Eshed Y, Inai S, Lopez J, Petiard V, Sayama H, Uhlig J, Zamir D, Tanksley SD (1998b) Advanced backcross quantitative trait locus analysis of tomato. II Evaluation of near isogenic lines carrying single-donor introgressions for desirable wild QTL alleles derived from *L. hirsutum* and *L. pimpinellifolium*. Theor Appl Genet 97:1191–1196

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random complementary DNA sequences. Genetics 112:887–898

Blair MW, Panaud O, McCouch SR (1999) Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). Theor Appl Genet 98:780–792

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

Doganlar S, Frary A, Ku HK, Tanksley SD (2002) Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45:1189–1202

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 QTLs. Genetics 141:1147–1162

Frary A, Nesbitt TC, Frary A, Grandillo S, Van der Knaap E, Cong B, Lui J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

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

Fridman E, Liu YS, Carmel-Goren L, Gur A, Shoresh M, Pleban T, Eshed Y, Zamir D (2002) Two tightly linked QTLs modify tomato sugar content via different physiological pathways. Mol Genet Genomics 266:821–826

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

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

Grandillo S, Tanksley SD (1996b) Quantitative trait locus analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theor Appl Genet 92:935–951

Haanstra JP, Wye C, Verbakel H, Meijer-Dekens F, van den Berg P, Odinot P, van Heusden AW, Tanksley S, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* × *L. pennelli* 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

Jones DF (1917) Linkage in *Lycopersicon*. Am Nat 51:608–621

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

van der Knapp E, Tanksley SD (2001) Identification and characterization of a novel locus controlling early fruit development in tomato. Theor Appl Genet 103:353–358

van der Knapp E, Lippman ZB, Tanksley SD (2002) Extremely elongated tomato fruit controlled by four quantitative trait loci with epistatic interactions. Theor Appl Genet 104:241–247

Konieczny A, Ausubel FM (1993) A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

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

Ku HM, Doganlar S, Chen KY, Tanksley SD (1999) The genetic basis of pear-shaped tomato fruit. Theor Appl Genet 99:844–850

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Martin GB, Williams JG, Tanksley SD (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc Natl Acad Sci USA 88:2336–2340

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262:1432–1436

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea*—variability, patterns of mutations and use for evolutionary studies. Theor Appl Genet 104:436–450

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

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

Monforte AJ, Tanksley SD (2000b) Fine mapping of a quantitative trait locus from *L. hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Nesbitt TC, Tanksley SD (2002) Comparative sequencing in the genus Lycopersicon: implications for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162:365–379

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors using a complete linkage map of restriction fragment polymorphism. Nature 335:721–726

Ronen G, Carmel-Goren L, Zamir D, Hirschberg J (2000) An alternative pathway to β-carotene formation in plant chromoplasts discovered by map-based cloning of Beta (*B*) and old-gold (*og*) color mutations in tomato. Proc Natl Acad Sci USA 97:11102–11107

Rozen S, Skaletsky HJ (1997) PRIMER3. http://www-genome.wi.mit.edu/genome_software/other/primer3.html

Saal B, Wricke G (1999) Development of simple sequence repeats in rye (*Secale cereale* L.). Genome 42:964–972

Sela M, Budai-Hadrian O, Qilin P, Carmel-Goren L, Vunsch R, Zamir D, Fluhr R (2001) Genome-wide analysis for Fusarium resistance in tomato reveals multiple complex loci. Mol Genet Genomics 265:1104–1111

Smulders MJ, 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 97:264–272

Suliman-Pollatschek S, Kashkush K, Shats H, Hillel J, Lavi U (2002) Generation and mapping of AFLP, SSRs and SNPs in *Lycopersicon esculentum*. Cell Mol Biol Lett 7:583–597

Tadmor Y, Fridman E, Gur A, Larkov O, Ravid U, Zamir D, Lewinsohn E (2002) Identification of *malodorous*, a wild species allele affecting tomato aroma that was selected against during domestication. J Agric Food Chem 50:2005–2009

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

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

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Van der Hoeven R, Ronning C, Giovannoni J, Martin G, Tanksley S (2002) Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large expressed sequence tag collection and selective genomic sequencing. Plant Cell 14:1441–1456

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

Young ND, Zamir D, Ganal MW, Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *Tm-2a* gene in tomato. Genetics 120:579–586