Theor Appl Genet (2003) 106:363–373
DOI 10.1007/s00122-002-1076-0

C. He · V. Poysa · K. Yu

Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among Lycopersicon esculentum cultivars

Received: 19 March 2002 / Accepted: 16 May 2002 / Published online: 4 September 2002

Abstract The simple sequence repeat (SSR) or microsatellite marker is currently the preferred molecular marker due to its highly desirable properties. The aim of this study was to develop and characterize more SSR markers because the number of SSR markers currently available in tomato is very limited. Five hundred DNA sequences of tomato were searched for SSRs and analyzed for the design of PCR primers. Of the 158 pairs of SSR primers screened against a set of 19 diverse tomato cultivars, 129 pairs produced the expected DNA fragments in their PCR products, and 65 of them were polymorphic with the polymorphism information content (PIC) ranging from 0.09 to 0.67. Among the polymorphic loci, 2–6 SSR alleles were detected for each locus with an average of 2.7 alleles per locus; 49.2% of these loci had two alleles and 33.8% had three alleles. The vast majority (93.8%) of the microsatellite loci contained di- or tri-nucleotide repeats and only 6.2% had tetra- and penta-nucleotide repeats. It was also found that TA/AT was the most frequent type of repeat, and the polymorphism information content (PIC) was positively correlated with the number of repeats. The set of 19 tomato cultivars were clustered based on the banding patterns generated by the 65 polymorphic SSR loci. Since the markers developed in this study are primarily from expressed sequences, they can be used not only for molecular mapping, cultivar identification and marker-assisted selection, but for identifying gene-trait relations in tomato.

Keywords Simple sequence repeat · Molecular marker · *Lycopersicon esculentum* · Gene diversity

Communicated by G. Wenzel

C. He (✉) · V. Poysa · K. Yu
Greenhouse and Processing Crops Research Center,
Agriculture and Agri-Food Canada, Harrow, ON N0R 1G0, Canada
e-mail: hec@ba.ars.usda.gov
Tel.: +1-301-504-6649, Fax: +1-301-504-5728

Present address:
C. He, Soybean and Alfalfa Research Laboratory, USDA-ARS,
Building 006, 10300 Baltimore Ave., Beltsville, MD 20705, USA

Introduction

Molecular markers can provide an effective tool for efficient selection of desired agronomic traits since they are based on the plant genotypes and thus are independent of environmental variation. The use of molecular markers can facilitate tomato breeding by means of marker-assisted selection (MAS) to improve agronomically important traits such as yield, fruit quality and disease resistance. In the last decade, molecular markers such as RFLP (Van Ooijen et al. 1994; Sandbrink et al. 1995; Stevens et al. 1995), RAPD (Stevens et al. 1995; Qian et al. 2001), ISSR (Zietkiewicz et al. 1994; Joshi et al. 2000) and AFLP (Vos et al. 1995) have been developed in tomato and other crops. However, the use of RFLP for breeding purposes is limited because it requires the use of radioactivity and is labour intensive; RAPD, ISSR and AFLP markers either identify only dominant alleles or are sensitive to PCR amplification conditions.

Simple sequence repeats (SSRs) or microsatellites are short (mostly 2–4 bp) tandem repeats of DNA sequences. It is hypothesized that the variation or polymorphism of SSRs are a result of polymerase slippage during DNA replication or unequal crossing-over (Levinson and Gutman 1987). SSRs are not only very common but also hypervariable among the types of tandem repetitive DNA in the genomes of eukaryotes (Hamada et al. 1984; Edwards et al. 1991; Vosman and Arens 1997; Rallo et al. 2000; Van der Schoot et al. 2000). SSR markers are becoming the preferred molecular markers in crop breeding because of their properties of genetic co-dominance, high reproducibility and multiallelic variation. They are the most practical markers for genomic mapping, variety identification and marker-assisted selection.

In tomato, some microsatellite markers have been developed (Smulders et al. 1997; Areshchenkova and Ganal 1999), but the number of SSR markers available for molecular breeding is still small and only a limited number of SSR markers have been mapped to the tomato genome (Broun and Tanksley 1996; Areshchenkova and Ganal 1999). It is desirable, therefore, to develop more

SSR markers for genetic mapping and marker-assisted selection, since the SSR markers developed to-date are not evenly distributed and do not cover the entire genome.

The objectives of the present study included: (1) to develop and characterize more SSR markers for *Lycopersicon esculentum*; and (2) to determine the genetic relationships among a set of tomato varieties with different geographical origins using these SSR markers.

Materials and methods

Plant materials and DNA isolation

Seventeen *L. esculentum* varieties representing geographically different tomato germplasm obtained from Agriculture and Agri-Food Canada, Harrow, Ontario, and the two parental lines, DRS-Ben and DRS-Bosch, obtained from De Ruiter Seeds Inc., Holland, were used in this study to detect polymorphisms in simple sequence repeats (Table 1). Genomic DNA of the two parental lines was kindly provided by Rene Hofstede of De Ruiter Seeds Incorporated, and genomic DNA of the 17 lines was isolated from young leaves following the method described by Yu and Pauls (1994) with some modifications. For each sample, four fresh leaf disks, obtained by punching leaves with the cap of a 1.5-ml Eppendorf tube, were put into 400 µl of DNA extraction buffer (200 mM Tris–HCl, pH 7.4, 250 mM of NaCl, 25 mM of EDTA, pH 8.0, 0.5% SDS) and homogenized with a plastic pestle (Mandel Scientific Company Ltd.). Then 400 µl of 24:1 chloroform/isoamyl alcohol was added to the homogenized solution, vortexed and left at room temperature for 30 min. The homogenate was spun in a microcentrifuge at a speed of 10,500 rpm for 2 min and 350 µl of the supernatant were transferred into a new Eppendorf tube. For DNA precipitation, an equal volume (350 µl) of isopropanol was added to the tube that was left at room temperature for 5 min and then spun at 11,000 rpm for 5 min. Then, the DNA pellet was air-dried at room temperature for 30 to 60 min before it was dissolved in 200 µl of water at 4 °C overnight. The supernatant was collected after microcentrifugation at 1,300 rpm for 2 min, yielding about 25 ng/µl of DNA.

Search of DNA sequences and primer design

A list of about 1,000 solanaceae microsatellites (the majority were *L. esculentum*) showing the GenBank database accession numbers with their motifs, and the number of repeats was kindly provided by Andreas Matern, Cornell University, Ithaca, New York. The entire DNA sequence for each accession was searched, retrieved from the GenBank database and verified for the presence of SSRs. If the SSR was not at, or very close to, either the 5′ or 3′ end, the sequence was collected. Prior to primer design, all the saved DNA sequences were analyzed using the program DNASIS (Hitachi America Ltd., San Bruno, Cal.) for homologous sequences. Each sequence was compared with the rest of the DNA sequences. If homologous sequences were found, only one unique sequence was kept for primer design while the rest of the homologous DNA sequences were eliminated because of their redundancy.

PCR primers (forward and reverse) flanking the repeat sequence were designed using the computer program GENE RUNNER (Hastings Software, Inc., N.Y.). The core parameters used in the primer design include the following: (1) the primer length is between 18 bp and 25 bp, (2) the percentage of GC is between 35% and 60%, (3) the *Tm* of the primers is over 40 °C which was calculated using $Tm = 59.9 + 0.41\,(\%G+C) - (675/\text{primer length})$ based on the standard PCR conditions at a salt concentration of 50 mM (Sharrocks 1994), and (4) the predicted PCR products range from 100 to 350 bp in length with a preference of between 100 bp and 250 bp. In addition, the primer internal structures, such as hairpin loops, possible primer dimers, length of single base pair run at the 3′ end and the number of short repeats (such as CT, GA etc.) were also taken into consideration. When two or more SSRs were located in the same DNA sequence but were at different sites, two flanking primers were designed separately for each of the SSRs. All designed oligonucleotides were synthesized commercially by Sigmagenosys, Incorporated.

Table 1 The plant materials and their origins used in the identification of simple sequence repeats (SSRs) and the study of genetic diversity

Number	Name	Origin
1	Borbas	Hungary
2	Bulgaria 436-76	Bulgaria
3	CC218	Canada
4	Cocabul	France
5	Cornell-1010	USA
6	FM 6203	USA
7	Heinz 916010	Canada
8	L2024	South Africa
9	N1190	Canada
10	NC EBR-111	USA
11	Ohio 8245	USA
12	Purdue 812	USA
13	S-11-83-4	China
14	Saljut	Russia
15	Sandpoint	USA
16	Scorpio	Australia
17	White Fruit	?
18	DRS-Ben	Holland
19	DRS-Bosch	Holland

PCR amplification and product electrophoresis

PCR reactions were performed in 96-well plates using either the Perkin Elmer GeneAmp PCR system 9600 (PE Biosystems) or the TECHNE Genius themal cycler (Techne Ltd., U.K.) with the same amplification program. Each 10-µl reaction mixture contained about 25 ng of tomato genomic DNA, 0.3 µM of forward and reverse primers, 300 µM of each dNTP, 1 µl of 10 × PCR buffer containing 100 mM of Tris–HCl, pH 8.3, 500 mM of KCl, and 1 unit of *Taq* DNA polymerase. The PCR amplification conditions were programmed as one cycle of denaturation at 94 °C for 2 min, followed by 35-cycles amplification with a 25 s denaturing at 94 °C, a 25 s annealing at the *Tm* (*Tm* varies for the individual primers) and a 25 s extension at 68 °C.

After PCR amplication, the products were mixed with 3 µl of stop buffer (97% deionized formamide, 0.3% each bromophenol blue and xylene cyanol FF and 10 mM of EDTA, pH 8.0) and then denatured at 94 °C for 5 min in a PCR machine. Four microlitres of each denatured PCR product were used for fragment separation on a DNA sequencing gel (6% polyacrylamide, 8 M urea and 1 × TBE buffer) running at a constant power of 55 W for 2–2.5 h, using an S2 sequencing-gel apparatus (GIBCO BRL). A 1-kb-plus DNA size marker was also loaded along with the samples for each run to estimate the fragment sizes of the separated DNA fragments. After each run, the gel was placed in 10% glacial acetic-acid fixation solution for 20 min with gentle shaking, silver-stained for 30 min and then immediately developed in a 3% sodium carbonate solution according to the DNA silver-staining kit (Promega).

Nomenclature of SSR markers

The nomenclature of the SSR markers was based on the method described by Yu et al. (2000). The SSR name was prefixed with LE or LH, standing for *L. esculentum* or *Lycopersicon hirsutum*, followed by the repeat motif in lowercase and a number starting from 001 for each distinct repeat motif. For example, LEaat001 and LEaat002 represent, respectively, the SSR markers at two different loci with the same repeat motif “aat”. For the imperfect or compound repeats, such as $(AAG)_3T(TGA)_7$, only the motif with the highest repeat number, in this case TGA, is used. When two or more different repeats such as the SSR locus $(CT)_{12}(GATA)_{12}(AT)_2(AC)_{10}$ have the same number of repeats, the repeat motif at the 5′ end is used. Thus, the SSR name for $(CT)_{12}(GATA)_{12}(AT)_2(AC)_{10}$ is designated as LEct rather than LEgata. This SSR nomenclature system can be applied to any newly developed microsatellites and provides a simple way to track SSR loci for use in a breeding program.

Genetic analysis

All 19 genotypes from different geographic origins were used to screen the SSR primers for PCR amplification and product-length polymorphism. For primers that produced the expected fragments after PCR reactions, the number of alleles was recorded and the polymorphism information content (PIC) of an SSR locus was calculated as described by Saal and Wricke (1999):

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

where p_i is the frequency of the ith allele out of the total number of alleles at an SSR locus, and k is the total number of different alleles for that locus.

For phylogenetic analysis, only the data for the polymorphic SSR loci were entered for all DNA samples, and a “1” or “0” was used if an allele was present or absent for a genotype, respectively. The data were analyzed using the computer program TREECON (Van de Peer and De Wachter 1994). The genetic-distance estimation was based on the method described by Nei and Li (1979). All 19 different tomato genotypes were clustered based on the estimated genetic distance, and the phylogenetic tree topology was inferred with the clustering method of the Unweighted Pair Group Method Using Arithmetic Average (UPGMA).

Results

DNA sequence retrieval from database and design of the SSR primers

Each of the accession numbers showing putative microsatellites with a minimum of six repeats for trinucleotide SSRs, or nine repeats for dinucleotide SSRs, totalling a minimum of 18 nucleotides within the microsatellite regions, was entered into the GenBank for DNA sequence retrieval. However, four microsatellites with the accession numbers of L19762, M13938, X13437 and Z15141, which had 4 or 5 repeats or fewer than 18 nucleotides in total within the SSR region, were also searched in the GenBank for primer design, because these four SSRs were reported to be polymorphic among four *Lycopersicon* species (Smulders et al. 1997).

In total, 500 *L. esculentum* DNA sequences, as well as one from *L. hirsutum* and two from *Lycopersicon pimpinellifolium*, were searched and checked for the presence of SSRs. After each of the DNA sequences was checked for the presence of one or more microsatellites, sequence homology or duplication, 127 (25.4%) DNA sequences had the SSR at either the 5′ or 3′ end and 41 (8.2%) DNA sequences were redundant duplicates or homologous to other sequences. One hundred and ninety three (38.6%) were short sequences or contained a high A/T content from which no suitable primers could be designed. After these unsuitable DNA sequences were eliminated, a total of 139 (27.8%) DNA sequences, both genomic sequences and ESTs, were found suitable for designing primers flanking the microsatellites. The names of these microsatellites, their locus names, core motifs, the primer sequences (forward and reverse) with their melting temperatures (*Tm*) and expected sizes of the PCR products are listed in Table 2. In addition to the 139 primer pairs designed, 15 primer pairs published by Areshchenkova and Ganal (1999) and four primer pairs published by Smulders et al. (1997) were also used in this study as these 19 primer pairs generated two or more SSR alleles among different *L. esculentum* cultivars and among four *Lycopersicon* species, repectively. Thus, 158 SSR primer pairs were available for PCR reactions.

Allelic variation and SSR characterization

All of the 158 SSR primers were used to screen a set of 19 diverse tomato cultivars or lines from different countries (Table 1). Of the 158 SSR primer pairs, 129 were able to produce the expected DNA fragments in their PCR products while the other 29 primers failed to amplify the expected PCR fragments. Of the 129 amplified primer pairs, 65 were polymorphic and 64 were monomorphic among the 19 tomato cultivars.

For the polymorphic SSR loci, 2–6 alleles were detected and the expected fragment sizes varied from 100 to 385 bp (Table 2). The variation of PCR fragment sizes among different alleles within the individual SSR locus tested in this set of 19 tomato cultivars ranged between 2 and 74 bp. The polymorphism information content (PIC) ranged from 0.09 for the primers LEaat003 (AW035051) and LEtca001 (AW035615), to 0.67 for the primer LEta019 (X90770). Among the 65 polymorphic SSR loci, 32 (49.2%) of them showed two alleles and 22 (33.8%) had three alleles (Table 3). The average number of alleles per locus was 2.7 for the polymorphic primers. For the 129 SSR loci which produced the expected PCR products, a total of 242 SSR alleles were amplified. Most (93.8%) of the SSR loci for tomato contained di- (55%) and tri-nucleotide (38.8%) repeats and only eight (6.2%) had tetra- and penta-nucleotide repeats (Table 4). Of the 71 (55.0%) SSR loci with dinucleotide repeats, 40 (56.3%) of them were polymorphic with an average PIC of 0.38. For the 50 (38.8%) SSR loci with trinucleotide repeats, 22 (44.0%) of them were polymorphic with an average PIC of 0.34. Among the 129 SSRs characterized, the TA/AT repeat was the most common type (41.1%), followed by the AAT/ATA (10.1%) and GA/CT (8.5%) repeats. The percentages of polymorphic loci for these three repeat types were 52.8%, 46.2% and 72.7%, respectively (Table 5).

SSR markers and cultivar differentiation

Table 6 lists the allelic profiles of the 19 cultivars at five SSR loci with a range of 2 to 4 alleles. The number of

Table 2 The simple sequence repeat markers, their locus names, core motifs and the flanking primer sequences, melting temperatures, allele numbers, expected fragment sizes of the PCR products and their polymorphic information content (PIC)

SSR name[a]	Locus	Core motif[b]	Primer sequence (5′~3′)	Tm (°C)	Allele no.	Expected size (bp)	PIC
LEaac001	cLEC32K6	$(aac)_6(ggc)_2$	f: agg aag agc gtg agt ctg aac	49.2	1	110	
(AW034789)			r: tcc tgc gcc act tta gag	45.2			
LEaat001	cLER5E10	$(aat)_{14}$	f: gat gga cac cct tca att tat ggt	48.9	4	136	0.46
(AI773078)			r: tcc aag tat cag gca cac cag c	51.6			
LEaat002	cLES4O3	$(aat)_{12}$	f: gcg aag aag atg agt cta gag cat ag	52.9	3	106	0.55
(AI778183)			r: ctc tct ccc atg agt tct cct ctt c	54.2			
LEaat003	EST279678	$(tct)_5(aat)_6$ imp	f: ctt gag gtg gaa ata tga aca c	46.0	2	189	0.09
(AW035051)			r: aag cag gtg atg ttg atg ag	44.6			
LEaat004	cLEC36E21	$(aac)_3(aat)_6$	f: cag gat cag aac agc gat g	46.0	1	240	
(AW035780)			r: cca ctg gta tcc atc ttt cac	47.3			
LEaat005	cLEE1C1	$(aat)_6$	f: ggt cat gca ggt tgg att ac	46.7	1	129	
(AW036045)			r: aac ctt cct tcc tat tgg c	43.8			
LEaat006	cLET1M11	$(aat)_{12}$	f: gcc acg tag tca tga tat aca tag	48.9	3	174	0.56
(AW037347)			r: gcc tcg gac aat gaa ttg	42.9			
LEaat007	cLET10O9	$(aat)_{12}$	f: caa cag cat agt gga gga gg	48.7	3	100	0.52
(AW039042)			r: tac att tct ctc tct ccc atg ag	48.4			
LEaat008	THox1	$(aat)_{12}$	f: gag tca aca gca tag tgg agg agg	54.0	3	178	0.58
(U76409)			r: cgt cgc aat tct cag gca tg	48.7			
LEac001	CLES13J1	$(ac)_9$	f: tgc ctt cca tct aac caa tc	44.6	1	219	
(AI899556)			r: ctg tgg caa ata tgt ccc taa g	47.9			
LEac002		$(gt)_9(at)_8(ac)_{13}(ga)_{12}$	f: tgt tgg ttg gag aaa ctc cc	46.7	2	180	0.40
(TMS22)			r: agg cat tta aac caa tag gta gc	46.6			
LEact001	cLEC35I20	$(act)_6$	f: aat cat caa ctt taa act gtg aca c	46.0	1	155	
(AW032325)			r: tgc att gag atg agt cgt tgg	47.3			
LEag001	Cleb3O13	$(ag)_{11}$	f: gca cga gca cat ata gaa gag aat ca	51.3	2	161	0.44
(AI491173)			r: cca ttt cat cat atc tct cag ctt gc	51.3			
LEag002	toxb0002L22r	$(ag)_{11}$	f: aga cgc ttc gac ggg gtt ta	48.7	2	184	0.33
(AQ367719)			r: agg aca ggt gaa tgg gtc aaa ga	50.2			
LEag003	cLEE3E15	$(ag)_{11}$	f: acc cta aaa cta acg aca ttc aac g	49.3	1	167	
(AW036506)			r: ttc gtg gac taa tgt atg aag tgt acc	51.6			
LEaga001	cLET1P9	$(aga)_6$	f: ttc ttc act gtt gac aga gag ag	48.4	1	219	
(AW038161)			r: cat tag ttg aga gtg ata ccg c	47.9			
LEagat001	LEMSP450	$(agat)_{10}$	f: tcc aga tag tca gtc aga cag c	49.7	1	270	
(X91107)			r: tct cta tct tta aga gtg gga gaa c	49.3			
LEat001	cLED9E6	$(at)_{12}$	f: gcg cga gct ctc tct gat ctc t	53.4	1	115	
(AI487132)			r: ttg taa ttg cat cgg cca cg	46.7			
LEat002	Cleb1P20	$(at)_9$	f: act gca ttt cag gta cat act ctc	48.9	2	203	0.50
(AI491065)			r: ata aac tcg tag acc ata ccc tc	48.4			
LEat003	cLED34A4	$(at)_{10}$	f: gag aag ttg gtg cat tca taa c	46.0	1	116	
(AI771611)			r: aaa cag taa acc aaa cac ttg c	44.1			
LEat004	cLER2C24	$(at)_{12}$	f: gcc act tga tca tca tca tga gta ca	51.3	1	228	
(AI772305)			r: aga agc caa tga agt gag tgt tgc	50.6			
LEat005	Cles12B1	$(at)_9$	f: tgc agc ctt tgg gta aac	42.9	2	164	0.20
(AI780685)			r: ata gtt tga aga gag gga gaa ag	46.6			
LEat006	Clec10F17	$(at)_{12}$	f: cat aat cac aag ctt ctt tcg cca	48.9	2	166	0.35
(AI895937)			r: cat atc cgc tcg ttt cgt tat gta at	49.7			
LEat007	Clec14J3	$(at)_9$	f: gcc cta gat ctc aca agc c	48.1	1	175	
(AI896276)			r: cac aaa gct gaa tga tac gaa g	46.0			
LEat008	cLED30J1	$(at)_{12}$	f: aag cgc gag ctc tct ctg atc tc	53.7	1	102	
(AI897766)			r: cca cga tct ccg cca tat gc	50.8			
LEat009	toxb0002K08r	$(at)_{15}$	f: gcc cag gta aaa gca atg ttg c	49.7	1	219	
(AQ367308)			r: agc aaa cct agg gac aga tcc ata	50.6			
LEat010	toxb0002H05r	$(at)_{30}$	f: tgg ctc tgc tca act caa gaa cta c	52.6	1	337	
(AQ367511)			r: cac gtg agg tta gcc agt gga tc	53.7			
LEat011	toxb0002K01f	$(at)_{10}$	f: tgg gct gac ttc gag ttt g	46.0	1	160	
(AQ368334)			r: cga gaa agg gca gag aat g	46.0			
LEat012	cLEC30K22	$(at)_{11}$	f: cgg caa agg gac tcg aat tg	48.7	1	110	
(AW033372)			r: gtg gcg gag tag aaa cct tag ga	51.9			
LEat013	Clec14E19	$(at)_{11}$	f: atc aca agc ttc ttt cgc cac a	47.9	2	163	0.27
(AW034465)			r: acc cat atc cgc tcg ttt cg	48.7			
LEat014	Clec11L13	$(at)_9$	f: tgt gtt gcg tca tta cca cta aac	48.9	2	209	0.10
(AW034592)			r: ccc aac cac caa tac ttt cc	46.7			
LEat015		$(at)_{67}$	f: gga ttg tag agg tgt tgt tgg	47.3	3	385	0.62
(TMS23)			r: ttt gta att gac ttt gtc gat g	42.3			
LEat016	lap17.1a	$(at)_9$	f: ccc aaa tgc tat gca ata cac	45.3	4	184	0.35
(Y08305)			r: agt tca gga ttg gtt taa ggg	45.3			

Table 2 (continued)

SSR name[a]	Locus	Core motif[b]	Primer sequence (5′~3′)	Tm (°C)	Allele no.	Expected size (bp)	PIC
LEat017 (Y09371)	LESATTAGA	$(at)_{12}$	f: tga gaa caa cgt tta gag gag ctg r: cgg gca gaa tct cga act c	50.6 48.1	3	206	0.35
LEat018 (TMS39)		$(at)_{29}$ imp	f: cgg cgt att caa act ctt gg r: gcg gac ctt tgt ttt ggt aa	46.7 44.6	5	120	0.65
LEat019 (A21360)	pTOM5	$(at)_{10}$	f: tgc ctc tct tca aag ata aag c r: cgg aaa gtt ctc tca aag gag	46.0 47.3	1	209	
LEat020 (X60441)	LEGTOM5	$(at)_{10}$	f: act gcc tct ctt caa aga taa agc r: acg gaa agt tct ctc aaa gga gtt g	48.9 50.9	1	212	
LEata001 (AI487481)	cLED11G18	$(ata)_8$	f: tgc aac aac tgg ata ggt cg r: tgt gga tga aac gga tgt tg	46.7 44.6	1	187	
LEata002 (AI489079)	cLED19B18	$(ata)_8$	f: tgc aac aac tgg ata ggt cg r: tga aat cac aac tcg aac atc c	46.7 46.0	1	129	
LEata003 (AI490477)	cLED18K18	$(ata)_8$	f: tct gca aca act gga tag gtc r: gtg gat gaa acg gat gtt g	47.3 43.8	1	188	
LEata004 (AI895825)	cLEC10O17	$(ata)_8$	f: caa ctg gat agg tcg atg g r: gat gtg gat gaa acg gat g	46.0 43.8	1	184	
LEata005 (AW035829)	cLEC36G16	$(ata)_6$	f: atg ctg ttt ggc gtg agg r: cgg cgg caa ctt tag aag	45.2 45.2	1	151	
LEatag001 (X91107)	LEMSP450	$(atag)_8(atgg)_7$ $(atag)_1(atgg)_7$	f: ctt att tag atg gtt tgt gtg aga c r: ggc tgt ctg act gac tat ctg g	47.7 51.6	1	278	
LEatct001 (M21775)	TOM2A11	$(atcg)_2(atct)_3$	f: aaa ctc tga ttg cat cgg aat tac c r: tac aga caa cac tat acg cgc aga g	49.3 52.6	1	168	
LEatg001 (AW036481)	cLEE3M11	$(atg)_7$	f: tcc cat tga aga cca agg r: agg tcc ttc aaa gct ctg c	42.9 46.0	1	243	
LEatt001 (AI898209)	Cled32E16	$(att)_9$	f: cca ttg ttc cat gca gaa g r: cca atg ctg att taa tgc g	43.8 41.6	2	118	0.19
LEcaa001 (AW034970)	cLEC31N20	$(caa)_7$	f: aga agg cgt gag agg caa c r: ctt agc act tga tgt tga ttg g	48.1 46.0	2	105	0.33
LEcac001 (AW033878)	cLEC27O13	$(cac)_6$	f: agc tgt tgc tgc agt tgg r: gaa aca tag agt cca tag gtg c	45.2 47.9	1	159	
LEcag001 (AF000142)	LEAF000142	$(cag)_8$	f: atg gtt ctt cat caa cag cag r: aga agt att gag cca agt cgg	45.3 47.3	2	123	0.19
LEcag002 (AW032661)	cLEC23M7	$(cag)_6$	f: ggg tgt ttc tct tct agt gtt tg r: gct cta tta acc ctt gct gc	48.4 46.7	1	114	
LEcag003 (AW034362)	cLEC33E15	$(cag)_7$	f: ccg cct ctt tca ctt gaa c r: cca gcg ata cga tta gat acc	46.0 47.3	3	133	0.42
LEcca001 (AW033946)	cLEC37C20	$(cca)_7$	f: aac acc cgc tac acc atg r: gca cct agc ttg aga gca tc	45.2 48.7	1	102	
LEcccca001 (L19762)	TOMSSF	$(cccca)_4$	f: cgc tct caa gta ccg taa gat ggc r: tct cca acc tac att gac atg acc a	54.0 50.9	1	221	
LEcgg001 (AW034705)	cLEC12D10	$(cgg)_7$	f: gct taa tcc tcc att cga tc r: atc cat ctg gct tca ccg	44.6 45.2	2	131	0.10
LEct001 (AI780156)	cLES10N9	$(ct)_{12}$	f: tcc aat ttc agt aag gac ccc tc r: ccg aaa acc ttt gct aca gag tag a	50.2 50.9	3	111	0.35
LEct002 (M13938)	TOMWIPIG	$(ct)_4(at)_3$	f: gtg gtg cac tct tac aaa ttc act c r: agg taa att ctt tgt gga agt ccc	50.9 48.9	1	236	
LEct003 (TMS4)		$(ct)_{12}(gata)_{12}atat(ac)_1$	f: cga tta gag aat gtc cca cag r: tta cac ata caa ata tac ata gtc tg	47.3 45.0	3	230	0.59
LEct004 (TMS29)		$(ct)_3c_{14}(ct)_{23}$	f: agc cac cca tca caa aga tt r: gtc gca cta tcg gtc acg ta	44.6 48.7	3	354	0.64
LEctat001 (X90937)	LEGATAREP	$(ctat)_8$	f: tgc cca tga cgt tcc atc r: gac aga cag aga gac aga ctt aga g	45.2 52.6	3	292	0.23
LEctt001 (AI897173)	cLED26N22	$(ctt)_9$	f: cct ctc ttc acc tct tta caa ttt cc r: cac tgg tca tta agt cta cag cc	51.3 50.2	2	101	0.39
LEctt002 (AW032327)	cLEC35G20	$(ctt)_6$	f: aaa caa cac cgc aac tcc r: tca gag aaa tag cga gtc cac	42.9 47.3	2	120	0.34
LEctt003 (AW032557)	cLEC8C22	$(ctt)_7$	f: att ccc aac act tgc cac r: ccc acc act atc caa acc c	42.9 48.1	1	219	
LEctt004 (AW038907)	cLET10M5	$(ctt)_6$	f: ccc atg gct tcg tta tcc r: cgc aag aag atg gaa gga ag	45.2 46.7	1	110	
LEga001 (AI898079)	cLED31L15	$(ga)_{29}$	f: cat cac tgg agt ttc tcc ctc r: cac tct cgc tct ctc tca ctc	49.2 51.2	1	173	
LEga002 (AW037298)	cLET1G9	$(ga)_{26}$	f: cct ggt gac tta tgg ttc tcg r: gac att cat gct act cag ttc ag	49.2 48.4	1	121	
LEga003 (TMS26)		$(ga)_{20}$	f: ttc ggt tta ttc tgc caa cc r: gcc tgt agg att ttc gcc ta	44.6 46.7	4	241	0.58

Table 2 (continued)

SSR name[a]	Locus	Core motif[b]	Primer sequence (5′~3′)	Tm (°C)	Allele no.	Expected size (bp)	PIC
LEga004 (TMS33)		$(ga)_{26}$ imp	f: agc atg gga aga aga cac gt r: ttg agc aaa aca tcg caa tc	46.7 42.6	3	267	0.61
LEga005 (TMS43)		$(ga)_{31}(gata)_7$	f: ttg gcc taa tcc ttt gtc at r: aac aat gtg acg tct tat aag gg	42.6 46.6	2	314	0.21
LEga006 (TMS45)		$(ga)_{17}(gt)_8$	f: ccg tcc aga aga cga tgt aa r: caa agt ctt gcc aac aat cc	46.7 44.6	2	248	0.17
LEga007 (TMS37)		$(ga)_{21}(ta)_{20}$	f: cct tgc agt tga ggt gaa tt r: tca agc acc tac aat caa tca	44.6 43.4	6	193	0.56
LEgaa001 (AW033198)	cLEC30M11	$(gaa)_6$	f: tca tct tca acc tca agg c r: tcg gat tcg gat tct tcg	43.8 42.9	1	131	
LEgaa002 (AW038667)	cLET7I23	$(gaa)_7$	f: agc tgc tct aat gtt gtt tct c r: ttc aaa gct act ctc aac atc c	46.0 46.0	1	207	
LEgata001 (TMS6)		$(gata)_{45}$	f: ctc tct caa tgt ttg tct ttc r: gca agg tag gta gct agg ga	43.4 48.7	3	335	0.42
LEgata002 (TMS9)		$(gata)_{26}$	f: ttg gta att tat gtt cgg ga r: ttg agc caa ttg att aat aag tt	40.5 41.2	3	344	0.62
LEgcc001 (AW034775)	cLEC32C6	$(gcc)_6$	f: gtt cct aat ggc act gct g r: gca gcg ttg taa agt tga gc	46.0 46.7	1	110	
LEgt001 (TMS42)		$(at)_{17}(gt)_{18}$	f: aga att ttt tca tga aat tgt cc r: tat tgc gtt cca ctc cct ct	41.2 46.7	4	274	0.23
LEgtc001 (AW035226)	cLEC35A17	$(gtc)_6$	f: tcg gag gca gat atc agc r: cga cag aac gac tct ctt agg	45.2 49.2	2	115	0.13
LEta001 (AI779459)	cLES8C23	$(ta)_{10}$	f: cgt cga gga aca cag aaa c r: act tag ttc ttc tcc aca gtt gag	46.0 48.9	1	129	
LEta002 (AI780401)	cLEC6J12	$(ta)_{13}$	f: gcc tcc cac aac aat cat cta tac a r: tcc tcc gta ctt tga tca tct tgt t	50.9 49.3	1	190	
LEta003 (AI895126)	cLED34K7	$(ta)_9$	f: gct ctg tcc tta caa atg ata cct cc r: caa tgc tgg gac aga aga ttt aat g	52.9 49.3	4	111	0.43
LEta004 (AI898482)	cLES11L23	$(ta)_{13}$	f: aag aat gga tag tca aca acc c r: ctg tga cgt aat tta tca tat cac	46.0 45.4	2	158	0.40
LEta005 (AQ367416)	toxb0001C23r	$(ta)_9$	f: gca aga tga ttt ggt gag atc r: tgt cag ctt gaa atc tcc atc	45.3 45.3	1	203	
LEta006 (AW035731)	cLEC36O1	$(ta)_{20}$	f: ccc tct tgc cta aac atc c r: tct act cgt tgc gaa ttc ag	46.0 44.6	2	167	0.29
LEta007 (AW031453)	cLEC40H9	$(ta)_{20}$	f: gcc gtt ctt ggt gga tta g r: cct cct ttc gtg tct ttg tc	46.0 46.7	3	291	0.34
LEta008 (AW030390)	cLEC20K18	$(ta)_9$	f: atg caa cct cca aac ata ttc c r: gaa cac aca aga tga agt gaa acg	46.0 48.9	2	168	0.10
LEta009 (AW031868)	cLEC38G20	$(ta)_9$	f: tca tgg ctc tca ctg ctc ttt ag r: atc ttt ctt gga tcg gag ctg	50.2 47.3	2	247	0.10
LEta010 (AW036280)	cLEE1L22	$(tg)_{14}(ta)_{15}$	f: cct cct tga aat atc ggc taa aca r: ggg ttg aaa gaa caa aga gag aga aag	48.9 51.6	1	263	
LEta011 (AW038112)	cLET2J3	$(ta)_{14}$	f: cgg tcc agt aag gtt gat gaa agc r: cca atg ttc att aca aga ctc gac aa	52.3 49.7	1	178	
LEta012 (AQ368062)	toxb0001B06r	$(ta)_{19}$	f: tga tcc taa gct ttt tcc gtg agt r: caa gtt cac ctc att tca ccc ct	48.9 50.2	3	254	0.24
LEta013 (M61915)	TOMILV1B	$t_9(ta)_{10}t_5$	f: aaa gag aag ata aac aga ggg taa g r: caa cct gtc ctt taa tct tta gg	47.7 46.6	2	374	0.22
LEta014 (TMS7)		$(ta)_{31}(gata)_{13}$ imp	f: aca aac tca aga taa gta aga gc r: gtg aat tgt gtt tta aca tgg	44.8 41.4	4	170	0.64
LEta015 (U63117)	tomloxA	$(ta)_{15}$	f: ata tgc atg gac aaa tct tga ggg r: ctc gcg cat caa att aat gta tca g	48.9 49.3	2	107	0.49
LEta016 (U81996)	le16	$(ta)_{14}$	f: agg ttg atg aaa gct aaa tct ggc r: caa cca cca atg ttc att aca aga c	48.9 49.3	3	174	0.43
LEta017 (X13437)	LEE8	$(ta)_5$	f: gag cac cca tta att tcg tta cg r: gtg gcg gat cta gaa att taa act g	48.4 49.3	3	182	0.19
LEta018 (X63093)	LEGAST1	$(ta)_{12}$	f: aaa tca ggt gag ccc aaa tg r: cat aat gtt ggc cct tga aac c	44.6 47.9	2	146	0.10
LEta019 (X90770)	LEMSREPRG	$(ta)_{20}$	f: tgt aga taa ctt cct agc gac aat c r: acg gac gga tgg aca aat g	49.3 46.0	5	243	0.67
LEta020 (Y08306)	LELAP17PR	$(ta)_{11}$	f: aac ggt gga aac tat tga aag g r: cac cac caa acc cat cgt c	46.0 48.1	4	175	0.60
LEta021 (M76552)	LELE25	$(ta)_{11}$	f: ttc ttc cgt atg agt gag t r: ctc tat tac tta tta tta tcg	41.6 37.5	3	225	0.20
LEta022 (M88487)	LEACS4A	$(ta)_7$	f: tac aga ata ggg ttt gcc ata r: gtt tta gtg ggt tgt gtt gaa	43.4 43.4	2	128	0.31

Table 2 (continued)

SSR name[a]	Locus	Core motif[b]	Primer sequence (5′~3′)	Tm (°C)	Allele no.	Expected size (bp)	PIC
LEta023 (TMS48)		$(ga)_{24}(ta)_{31}$ imp	f: att gct cat aca taa ccc cc r: ggg aca aaa tgg taa tcc at	44.6 42.6	3	184	0.61
LEta024 (L35306)	LEMDDN	$(tg)_4(ta)_5$	f: taa ata caa aag cag gag tcg r: gag ttg aca gat cct tca atg	43.4 45.3	4	280	0.51
LEta025 (AA824863)	CT149	$(ta)_9$	f: cct cca tcc ata ctt aat ccc r: ggt gta cta aca att tgg gta gg	47.3 48.4	1	211	
LEta026 (U81378)	LEU81378	$(ta)_{15}$	f: ggt caa gat ttg gag tgt tta g r: aat ttg ccc ttg gtc gtc	46.0 42.9	1	229	
LEta027 (X05985)	LERBCS3B	$(tg)_4(ta)_6$	f: ggt gga aga gtc agt tgc atg r: cgt act tct tca tgt taa ttg gtg g	49.2 49.3	1	147	
LEta028 (X14060)	LENIA	$(ta)_9(tg)_5$ imp	f: cag tac ttt gtt gtc aca agt ctt g r: ctt tag gct tgt aat gga gtg c	49.3 47.9	1	184	
LEta029 (X15499)	LELAT59G	$(ta)_{16}$ imp	f: acc cgg aac tct tcg tca tg r: gat cat ctc ctg gtg caa cc	48.7 48.7	1	197	
LEta030 (X59139)	LEACC2G	$(ta)_5(aaat)_3$ imp	f: att gtt ctc gtc cct tcc cag r: ttc aag cta gaa gct aca cgt gag	49.2 50.6	1	160	
LEta031 (X61287)	LECAB9	$(ta)_6(ca)_3$	f: act gtg gtc ctg aag ctg ttt gg r: ccg aag taa ttc aat gtg ttt ccg	51.9 48.9	1	161	
LEta032 (X90937)	LEGATAREP	$(ta)_{10}$	f: cta cct tcc tac cta cct act tac c r: cag aca aac aga cag aaa gac ag	52.6 48.4	1	296	
LEta033 (Z15141)	LECHI3	$(ta)_4(ga)_4$	f: cca aat act gca gcg gaa ag r: ttc taa atg ggc ata cag aat c	46.7 44.1	1	233	
LEtaa001 (AI489275)	Cled17L17	$(taa)_8$	f: tga gag aga tca acc aac tcc r: act act cct gcc tct cta tat cc	47.3 50.2	2	133	0.47
LEtaa002 (AI771867)	cLED38O3	$(taa)_8$	f: tga gag aga tca acc aac tcc r: act act cct gcc tct cta tat cc	47.3 50.2	1	133	
LEtac001 (AW037257)	cLET1G9	$(tac)_6$	f: ccg gtg aag gtg agt ctg ag r: ttt atg cac cgc gac tcg	50.8 45.2	2	127	0.18
LEtat001 (AI484595)	cLED1E23	$(tat)_9$	f: ctg ttg atg atg aac ttg gtc c r: tgt tag ggc att tga tag aag g	47.9 46.0	1	119	
LEtat002 (AI486387)	CLED8F8	$(tat)_{12}$	f: acg ctt ggc tgc ctc gga r: aac ttt att att gcc acg tag tca tga	49.7 48.6	3	196	0.58
LEtat003 (U21085)	LE21085	$(gt)_2(ta)_3(tat)_6$ imp	f: cat ttt atc att tat ttg tgt ctt g r: aca aaa aaa ggt gac gat aca	42.7 41.4	3	104	0.36
LEtatg001 (AW037767)	cLET3J20	$(tatg)_5$	f: act agt agc agc cag ata aac tg r: cca tat agg tgc aaa tcg atc	48.4 45.3	1	227	
LEtc001 (AI896256)	cLEC14F9	$(tc)_9$	f: cct tcc acc ttc cta tcc c r: aac ctg atg atg atg atg tga g	48.1 46.0	1	106	
LEtca001 (AW035615)	cLEC39L12	$(tca)_7$	f: tgc atg gca aca tta aag tc r: cgt gga tgc aac ttc att g	42.6 43.8	2	176	0.09
LEtcc001 (AW032956)	cLEC17F17	$(tcc)_7$	f: gcc aag ctc gaa cct gta c r: att ggc cat tgt tgc tcg	48.1 42.9	2	110	0.20
LEtct001 (AI483067)	cLEB8E24	$(tct)_8$	f: gca cca ggt ttc gtt gaa g r: cag cag aaa taa cag atc ttg g	46.0 46.0	1	238	
LEtct002 (AI778597)	cLES5F24	$(tct)_8$	f: cta tag ctg aaa ctc aac ctg ag r: cca gca gaa ata aca gat ctt g	48.4 46.0	1	202	
LEtct003 (AI897170)	cLED26N14	$(tct)_8$	f: tcg ttg aag aag atg atg gtc r: gag cca cca aag aat aag aag	45.3 45.3	1	207	
LEtga001 (AW037442)	LELEUZIP	$(aag)_3t(tga)_7$	f: cgt ctg cat caa ttt cct c r: gtg ttc cta cat ttc agc tcc	43.8 47.3	1	164	
LEtga002 (Z12127)	cLET4H22	$(tga)_6$	f: ggt ggt gat aat ttg gga ggt tac r: aat gat tcc cgc cgg taa ag	50.6 46.7	2	150	0.19
LEttc001 (AW032445)	cLEC35N13	$(ttc)_6$	f: tga ttc aag gta caa gta gta gtg c r: gga gga ggg tga ata atc g	49.3 46.0	2	236	0.46
LEttc002 (AW033091)	cLEC23E9	$(ac)_3(ttc)_6$ imp	f: ttc tca cac ctg cac aca cc r: agc ggg atg att aca gaa atg	48.7 45.3	1	113	
LEttc003 (M37151)	TOMSODB	$(ttc)_6$	f: acc aca acc agc act acc aat tc r: tag tga cag cat aaa ggg tca aag	50.2 48.9	1	142	
LHaat001 (AJ002235)	LHJ002235	$(aat)_8$	f: tgt gtg tgt ctg cgt gtg c r: taa gtt tgt acg aag cat cct g	48.1 46.0	1	327	

[a] Names in brackets were the accession numbers from GenBank and the ones with a prefix "TMS" were the SSR markers from Areshchenkova and Ganal (1999)

[b] imp means "imperfect" repeat

Table 3 Allelic variation among SSR loci

Number of alleles	Number of SSR loci	% of loci
1	64	49.6
2	32	24.8
3	22	17.1
4	8	6.2
5	2	1.6
6	1	0.8

Table 4 No. nucleotides per repeat and the number of SSR loci

Repeat	No. of SSR loci		Polymorphic SSR loci	
	Total	%	Total	%
Dinucleotide	71	55.0	40	56.3
Trinucleotide	50	38.8	22	44.0
≥Tetranucleotide	8	6.2	3	37.5

Table 5 The major types of SSRs and the number of polymorphic loci

SSR type	No. of SSR loci		Polymorphic SSR loci	
	Total	%[a]	Total	%[b]
TA/AT	53	41.1	28	52.8
AAT/ATA	13	10.1	6	46.2
GA/CT	11	8.5	8	72.7
CTT	7	5.4	3	42.9

[a] % = total SSR loci for the particular type of SSRs/total number of SSR loci
[b] % = polymorphic SSR loci for the particular type of SSRs/total number of loci for that type

cultivars sharing the same banding pattern varies with different loci; for instance, for locus LEcaa001, 15 cultivars had the B allele, for locus LEaat002 and LEga003, only two cultivars had the same A or B alleles, respectively. Particularly, there is a Scorpio cultivar-specific allele of 247 bp at the locus LEga003. Although the use of an individual SSR locus may not differentiate many tomato cultivars, the combination of any two or three SSR loci could increase the efficiency for cultivar differentiation. The combination of all five SSR loci in Table 6 can differentiate all of the 19 tomato cultivars. The average polymorphism information content (PIC) for these five SSR loci was 0.51, higher than the average PIC (0.37) for all the 65 polymorphic SSR loci. In addition, although most of the cultivars had only a single band for a specific SSR locus, the presence of two alleles at the SSR loci in some of the cultivars could help discriminate among the varieties. For example, at the locus LEat002 (AI491065), the breeding line S-11-83-4 had two alleles and the rest of the genotypes had only one allele. At the locus LEaat002 (AI778183), both S-11-83-4 and White Fruit amplifed two alleles while the rest of the 17 cultivars (lines) only amplified one allele.

Phylogenetic analysis

Although 129 microsatellites were able to generate the expected PCR products, only 65 of them could produce polymorphisms among this set of 19 tomato cultivars. Therefore, only these 65 polymorphic SSR markers were used to analyse and group the 19 tomato cultivars using the TREECON computer program (Van de Peer and De Wachter 1994) (Table 2). Based on the genetic distance

Table 6 Allelic profiles of the 19 tomato varieties at five SSR loci

SSR marker	Allele (bp)	Cultivar[a]																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
LEtat002																				
(AI486387)	A (196[b])	+[c]			+	+		+			+		+	+	+					+
	B (199)		+				+		+			+				+		+	+	
	C (205)			+						+							+			
LEat002																				
(AI491065)	A (201)	+	+			+		+		+	+			+		+		+		+
	B (205)			+	+		+		+			+	+	+	+		+		+	
LEaat002																				
(AI778183)	A (103)								+								+			
	B (106)		+		+	+							+	+				+	+	+
	C (109)	+		+			+	+		+	+	+		+	+	+		+		
LEcaa001																				
(AW034970)	A (102)			+		+					+		+							
	B (105)	+	+		+		+	+	+	+		+		+	+	+	+	+	+	+
LEga003																				
(TMS26)	A (235)	+		+			+			+					+					
	B (239)					+										+				
	C (241)		+		+			+	+		+	+	+	+				+	+	+
	D (247)																+			

[a] 1 = Borbas, 2 = Bulgaria 436-76, 3 = CC218, 4 = Cocabul, 5 = Cornell-1010, 6 = FM 6203, 7 = Heinz 916010, 8 = L2024, 9 = N1190, 10 = NC EBR-111, 11 = Ohio 8245, 12 = Purdue 812, 13 = S-11-83-4, 14 = Saljut, 15 = Sandpoint, 16 = Scorpio, 17 = White Fruit, 18 = DRS-Ben, 19 = DRS-Bosch
[b] Estimated fragment size; [c] + indicating the presence of the allele

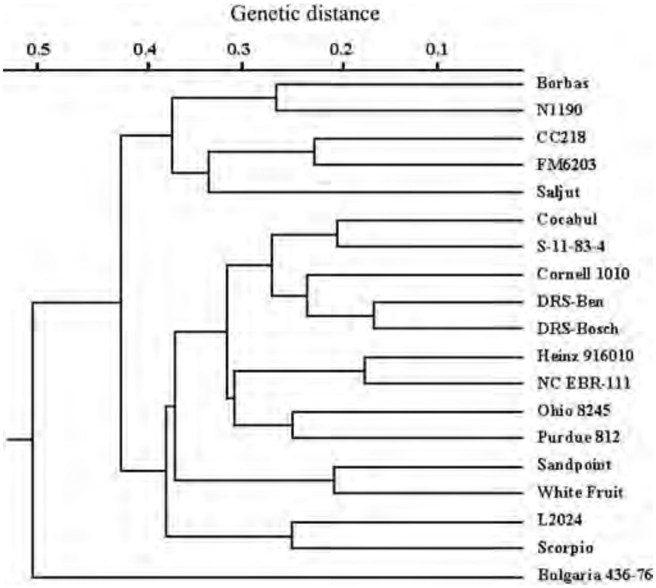

Fig. 1 Dendrogram presenting the association among the 19 tomato genotypes based on the UPGMA cluster analysis of 65 SSR amplification products

of Nei and Li (1979), 19 tomato cultivars were clustered into several different groups while the cultivar Bulgaria 436-76 was in a separate group. As the pedigrees of the majority of the cultivars were unknown, this dendrogram may only partially reflect their genetic relationships or geographic origin (Fig. 1).

Discussion

In this study, 500 tomato DNA sequences have been retrieved from the GenBank, but only 139 (28%) sequences were finally used for designing SSR primers. This was due to the fact that the majority, or 72%, of the DNA sequences were not suitable for primer design since 25% of them had simple sequence repeats at either the 5′ or 3′ end, 8% of them were duplicate or redundant DNA sequences and 39% of them were short sequences or had a high A/T content. This result was consistent with those reported in *Sorghum bicolor* and cassava where 70% and 45%, respectively, of the clones had SSRs too close to the cloning sites, which resulted in the SSRs being located at either the 5′ or 3′ end (Taramino et al. 1997; Mba et al. 2001). Likewise, the redundancy of DNA sequences, consisting of the same SSR locus or showing more than 95% similarity in the flanking sequences, were found to be 20% in cassava (Mba et al. 2001), 16% in perennial ryegrass (*Lolium perenne* L.) (Jones et al. 2001) and 10% in white clover (*Trifolium repens* L.) (Kölliker et al. 2001). Most of these were found to be due to cloning or locus-duplication or allelism, and were from the same SSR enrichment library. In addition, Ashkenazi et al. (2001) also reported that some of the conserved DNA sequences flanking the SSR regions were too short to design an appropriate primer in potato.

There are several advantages that microsatellite markers have over other types of markers such as RFLPs, RAPDs, AFLPs and ISSRs. One of them is the multiallelism of the simple sequence repeats with a range of 1 to 7 alleles for the majority of SSR loci (Loridon et al. 1998; Yu et al. 1999; Li et al. 2000; Ashkenazi et al. 2001; Jones et al. 2001; Kölliker et al. 2001). For instance, Danin-Poleg et al. (2001) found that the average number of alleles detected in melon was 3.5 by SSRs but only two by RFLP. In this study, the average number of SSR alleles/locus was 2.7 for the 65 polymorphic loci and the average PIC was 0.37. Similarly, Smulders et al. (1997) reported, on average, three alleles per locus after testing 30 SSR loci on seven species and cultivars of tomatoes. The present result was higher than the number of alleles detected in cucumber in which an average of 2.4 alleles/locus and a PIC of 0.28 were reported (Danin-Poleg et al. 2001). On the other hand, this average number of alleles was lower than that of potato (4.5) although the average PIC was close to the average heterozygous frequency of potato (0.39) (Ashkenazi et al. 2001). However, SSR loci with greater numbers of alleles might not necessarily have an advantage for determining PIC or differentiating genetic materials (Ashkenazi et al. 2001). In the present study, the majority of the polymorphic SSR loci had two alleles (49%) or three alleles (34%), and they could still be used to effectively differentiate tomato cultivars (Table 6, Fig. 1). In addition, 50% of the amplified SSRs, or 41% of all attempted SSRs, were polymorphic among the 19 tomato cultivars, which was lower than that (88%) found in white clover (*T. repens* L.) (Kölliker et al. 2001). Earlier studies also indicated that tomato cultivars were considered low in DNA polymorphisms based on the studies of SSRs (Broun and Tanksley 1996) and RFLPs (Miller and Tanksley 1990).

Allelic variation may be correlated with the number of repeats within a particular microsatellite locus. In other words, the repeat length may correlate with the polymorphism information content (PIC). A positive correlation ($r = 0.46$, $P<0.001$) was found between the number of repeats and the PIC for this study, which agreed with earlier reports in tomato (Smulder et al. 1997; Areshchenkova and Ganal 1999). Similar results were also found for grapevine (Thomas and Scott 1993), ryegrass (Jones et al. 2001) and white clover (Kölliker et al. 2001), but not in other species such as *Brassica* (Szewc-McFadden et al. 1996), rice (Panaud et al. 1996) and *Cucumis* (Danin-Poleg et al. 2001). No correlation was found in this study between PIC and the number of nucleotides per repeat ($r = -0.06$, $P = 0.61$). The average PIC for the SSR with dinucleotide repeats was 0.38, while the average PIC for the SSR with trinucleotide repeats was 0.34. However, there are reports that the polymorphism level in trinucleotide repeats is lower than that in dinucleotide repeats for rice (Blair et al. 1999) and ryegrass (Jones et al. 2001).

Earlier studies reported that the AT/TA repeat was the most-frequent type of SSR in plants, followed by the CT/GA repeat (Wang et al. 1994; Yu et al. 1999; Danin-

Poleg et al. 2001). In this study, the most-frequent type of microsatellite repeat was the AT/TA repeat (41%), followed by the AAT/ATA repeat (10%), the CT/GA repeat (9%) and the CTT repeat (5%), respectively (Table 5). However, the frequency of a microsatellite repeat may vary with different species. For instance, Ashkenazi et al. (2001) reported that ATT and GT were the most frequent repeats in potato.

To use microsatellite markers for cultivar differentiation, five representative polymorphic SSR loci showing easily scorable alleles along with the allelic profiles of the 19 tomato cultivars or lines were presented (Table 6). A unique banding pattern could be found for all of the 19 tomato cultivars within these five SSR loci, further suggesting that SSR markers are suitable for identifying cultivar-specific markers for tomato which has a low level of DNA polymorphism detected by other types of markers (Miller and Tanksley 1990; Broun and Tanksley 1996; Bredemeijer et al. 1998). In other words, DNA profiles generated by SSR markers can provide a tool for diagnostic fingerprinting of tomato cultivars. Use of these five SSR loci could effectively differentiate all 19 cultivars, which agrees with the previous study by Bredemeijer et al. (1998) where four SSR markers could differentiate 16 tomato cultivars. In potato, Ashkenazi et al. (2001) reported that as few as two markers could characterize 12 cultivars. This is because the average number of alleles per locus for potato is higher than that of tomato (Smulders et al. 1997; Ashkenazi et al. 2001). In addition, Table 6 indicated that the tomato line S-11-83-4 showed two alleles at the LEat002 and LEaat002 loci, while White Fruit had two alleles at the LEaat002 locus. The presence of two alleles in some cultivars for some of the SSR loci suggested that small heterozygous fragments still remain in the genomes of these cultivars during the inbreeding process or that some form of mutation occurred in the SSR regions. Nevertheless, microsatellite markers were demonstrated to be highly polymorphic and efficient for differentiating genetic materials, further suggesting their capacity for practical application in cultivar and seed purity identification and phylogenetic study.

In the phylogenetic analysis, the two tomato cultivars DRS-Ben and DRS-Bosch were clustered together in the dendrogram (Fig. 1). They were both from De Ruiter Seeds Incorporated, in Holland, and thus might have a similar genetic background although DRS-Ben is resistant to powdery mildew while DRS-Bosch is susceptible. The cultivars, CC218 and N1190, were both from Nabisco Ltd, Canada, and FM6203, a cultivar from the former tomato seed company, i.e. Ferry Morse Seed Co., USA, had Nabisco breeding lines in its pedigree (Poysa, personal communication). This could be the reason why CC218 and FM6203 were clustered together and were in the same group as N1190 at the genetic distance of ≤0.38 (Fig. 1). As for other cultivars, since the details on their pedigrees were unknown, the relationships among them could be biased due to the small number of loci being used. Thus, caution should be taken when the relationships among other cultivars are inferred for the choice of genetic materials in tomato breeding.

In conclusion, in this study, we developed and characterized 129 new microsatellite markers for *L. esculentum* in response to the limited number of SSR markers currently available. These SSR markers, combined with other published ones, can provide a supply for use in tomato breeding and research. Because of their advantages, SSR markers are becoming the preferred molecular marker for variety identification, genetic mapping and marker-assisted selection in tomato.

Acknowledgements This project was partially funded by the Ontario Tomato Research Institute and the Matching Investment Initiative (MII) fund of Agriculture and Agri-Food Canada. The authors thank Andreas Matern at Cornell University for providing a list of accessions with SSRs, Rene Hofstede of De Ruiter Seeds Inc. for providing two parental DNAs, Margaret Haffner, Bailing Zhang and Breaden Sweeting at Agriculture and Agri-Food Canada, Harrow, Canada, for their technical assistance.

References

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

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

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

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

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

Danin-Poleg Y, Reis N, Tzuri G, Katzir N (2001) Development and characterization of microsatellite markers in *Cucumis*. Theor Appl Genet 102:61–72

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Hamada H, Petrino MG, Kakunaga T, Seidman M, Stollar BD (1984) Characterization of genomic poly(dT-dG) poly(dC-dA) sequences: structure, organization and conformation. Mol Cell Biol 4:2610–2621

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

Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, Brar DS (2000) Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theor Appl Genet 100:1311–1320

Kölliker R, Jones ES, Drayton MC, Dupal MP, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). Theor Appl Genet 102:416–424

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Li CD, Rossnagel BG, Scoles GJ (2000) The development of oat microsatellite markers and their use in identifying relationships among *Avena* species and oat cultivars. Theor Appl Genet 101:1259–1268

Loridon K, Cournoyer B, Goubely C, Depeiges A, Picard G (1998) Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*. Theor Appl Genet 97:591–604

Mba REC, Stephenson P, Edwards K, Melzer S, Nkumbira J, Gullberg U, Apel K, Gale M, Tohme J, Fregene M (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. Theor Appl Genet 102:21–31

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

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

Panaud O, Chen SR, McCouch R (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.) Mol Gen Genet 252:597–607

Qian W, Ge S, Hong DY (2001) Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. Theor Appl Genet 102:440–449

Rallo P, Dorado G, Martin A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). Theor Appl Genet 101:984–989

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

Sandbrink JM, Van Ooijen JW, Purimahua CC, Vrielink M, Verkerk R, Zabel P, Lindhout P (1995) Localization of genes for bacterial canker resistance in *Lycopersicon peruvianum* using RFLPs. Theor Appl Genet 95:444–450

Sharrocks AD (1994) The design of primers for PCR. In: Griffin HG, Griffin AM (eds) PCR technology: current innovations. CRC Press, Inc, pp 5–11

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 97:264–272

Stevens MR, Lamb EM, Rhoads DD (1995) Mapping the *Sw-5* locus for tomato spotted wilt virus resistance in tomatoes using RAPD and RFLP analyses. Theor Appl Genet 90:451–456

Szewc-McFadden AK, Kresovich SK, Bliek SM, Mitchell SE, McFerson JR (1996) Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. Theor Appl Genet 93:534–538

Taramino G, Tarchini R, Ferrario S, Lee M, Pe ME (1997) Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor Appl Genet 95:66–72

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86:985–990

Van de Peer Y, De Wachter R (1994) TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Applic Biosci 10:569–570

Van der Schoot J, Pospíšková M, Vosman B (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor Appl Genet 101:317–322

Van Ooijen JW, Sandbrink JM, Vrielink M, Verkerk R, Zabel P, Lindhout P (1994) An RFLP linkage map of *Lycopersicon peruvianum*. Theor Appl Genet 89:1007–1013

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP – a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4417

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

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

Yu K, Pauls KP (1994) Optimization of DNA-extraction and PCR procedures for random amplified polymorphic DNA (RAPD) analysis in plants. In: Griffin HG, Griffin AM (eds) PCR technology: current innovations. CRC press, Boca Raton, Florida, pp 193–200

Yu K, Park SJ, Poysa V (1999) Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). Genome 42:27–34

Yu K, Park SJ, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J Hered 91:429–434

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176–183