Theor Appl Genet (2001) 103:1283–1292

A.E. Alvarez · C.C.M. van de Wiel · M.J.M. Smulders
B. Vosman

Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*

Received: 19 February 2001 / Accepted: 26 March 2001

Abstract In order to determine how informative a set of microsatellites from tomato is across the genus *Lycopersicon*, 17 microsatellite loci, derived from regions in and around genes, were tested on 31 accessions comprising the nine species of the genus. The microsatellite polymorphisms were used to estimate the distribution of diversity throughout the genus and to evaluate the efficacy of microsatellites for establishing species relationships in comparison with existing phylogeny reconstructions. Gene diversity and genetic distances were calculated. A high level of polymorphism was found, as well as a large number of alleles unique for species. The level of polymorphism detected with the microsatellite loci within and among species was highly correlated with the respective mating systems, cross-pollinating species having a significantly higher gene diversity compared to self-pollinating species. In general, microsatellite-based trees were consistent with a published RFLP-based dendrogram as well as with a published classification based on morphology and the mating system. A tree constructed with low-polymorphic loci (gene diversity <0.245) was shown to represent a more-reliable topology than a tree constructed with more-highly polymorphic loci.

Keywords SSR · Tomato · Microsatellites · Genetic diversity · Species relationships

Introduction

Cultivated tomato (*Lycopersicon esculentum)* is a species where genetic markers like isozymes (Breto et al. 1993; Foolad et al. 1993) and RFLPs (Miller and Tanksley 1990, Foolad et al. 1993; Williams and St. Clair 1993) yielded limited information. This lack of variability is a consequence of its self-pollinating nature in combination with the narrow genetic base of the modern cultivars. Van der Beek et al. (1992) using a combination of 195 probes and six restriction enzymes for RFLP analysis found that polymorphisms among tomato cultivars were scarce (e.g. with only three polymorphisms between two cultivars that did not contain introgressed areas from wild species). In contrast, Vosman et al. (1992) reported high levels of polymorphism in *L. esculentum* cultivars using oligonucleotide fingerprinting with a $(GATA)_4$ hybridization probe. In line with this, Kaemmer et al. (1995) found that three out of seven oligonucleotide motifs, $(GATA)_4$, $(CCTA)_4$ and $(GGAT)_4$, appear suitable for the identification and differentiation of cultivars and breeding lines that are otherwise difficult to distinguish. Moreover, Rus-Kortekaas et al. (1994) showed a GACA-containing probe to be more effective in cultivar identification than the four most-optimal RAPD primers selected. However, contrary to the RAPDs, the GACA probe failed to establish species relationships due to the lack of bands in common between the species samples.

The genus *Lycopersicon* is presently recognized as having nine species, eight wild ones and one containing both wild and domesticated material, but there is no general agreement on the phylogeny of the genus (Warnock 1988). To our knowledge, from the 1980s up till now, there have been three different approaches to reconstruct phylogenetic relationships within the genus with the aid of molecular tools, two based on organellar DNA and one based on nuclear DNA (Palmer and Zamir 1982; MacClean and Hanson 1986; Miller and Tanksley 1990). These studies are not in overall agreement with each other nor with the taxonomic data based on morphological traits or based on mating systems as reviewed by Rick (1979).

The informativeness of sequence-tagged microsatellite sites (STMSs), also called simple sequence repeats (SSRs), as a genetic marker has already been shown with great success in several plant species. Germplasm assessment for the amount of genetic diversity and cultivar identification have been the main applications in

Communicated by H.F. Linskens

A.E. Alvarez
Universidad Nacional de Salta, Facultad de Ciencias Naturales, Buenos Aires 177, 4400-Salta, Argentina

C.C.M. van de Wiel (✉)
Plant Research International B.V., P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: c.c.m.vandewiel@plant.wag-ur.nl
Tel.: +31-317-476981, Fax: +31-317-418094

crop plant species (for a review see Powell et al. 1996). In case they are transferable over a set of related species, microsatellites might also be useful for establishing species relationships where other commonly used characteristics, like direct sequencing of selected genomic regions, may fail to generate sufficient phylogenetically informative sites. However, the usually high variability of microsatellites might lead to inconsistencies due to the high chance of independently arising, equally sized alleles (homoplasies). At present, a good set of microsatellite loci in the genus *Lycopersicon* is available. Smulders et al. (1997), by screening EMBL and Genebank databases, identified 36 primer pairs that yielded well-scorable fragments, or groups of fragments, in *L. esculentum* cultivars and accessions of *Lycopersicon* species. These microsatellites were present in, or close to, coding regions and generally consisted of short (less than 12) repeat units. Such STMS markers may generate polymorphisms useful for the analysis of genetic diversity and species relationships within this genus.

The objectives of the present study, therefore, are to determine how informative a set of microsatellites from tomato is across the genus *Lycopersicon*, to establish the applicability of each microsatellite locus, depending upon its degree of polymorphism detected within and among accessions, to estimate the distribution of genetic diversity across the genus and to evaluate the feasibility of using microsatellites for studying species relationships within the genus.

Material and methods

Plant material

Two cultivars of *L. esculentum* and 29 accessions of the genus *Lycopersicon* were obtained from the tomato collection of the Centre for Genetic Resources (CGN, part of Plant Research International), The Netherlands (see Table 1).

DNA extraction

Total genomic DNA was extracted from 2-week-old seedlings. Each seedling was treated individually. The extraction was performed according to Fulton et al. (1995). The DNA concentration per sample was measured in a DNA-fluorometer model TKO 100, Hoefer Scientific instruments, San Francisco.

Microsatellite loci

Seventeen microsatellite loci were selected (see Table 2) out of 44 available (see Smulders et al. 1997). The selection was based on the scale proposed by Smulders et al. (1997), the loci chosen having quality 1 (strong amplification with only two bands, shadow and stutter bands weak, allele size easy to determine on silver-stained PAGE) and 2 (as 1 but shadow and stutter bands relatively strong, but determination of the allele size still possible). Both loci polymorphic among *L. esculentum* cultivars and loci not polymorphic among cultivars were included.

PCR amplifications

PCR amplifications were carried out in a 15-µl reaction volume: 6 µl of DNA template (about 1ng/µl), 1.5 µl of PCR GIBCO buffer 10×[200 mM Tris-HCl (pH 8.4), 500 mM KCl, 1.5 mM $MgCl_2$, 0.05% (v/v) polyoxyethylene ether (W-1)], 2.7 µl of water, 1.5 µl of a deoxyribonucleotides mixture (1 mM) (Life Technologies), 1.5 µl of forward and reverse primer (20 ng/µl) (Isogen, Maarssen, the Netherlands) and 0.3 µl of Taq DNA polymerase (1U/µl) (Life Technologies). Amplifications were performed in microtitre plates using a Hybaid Omni Gene thermal cycler. Basically, the amplification conditions were: 1 cycle of 94°C for 3 min, 25 to 30 cycles of 50 to 55°C for 45 s, 72°C for 1 min 45 s, and 94°C for 45 s (the annealing temperature and the number of cycles employed varied depending on the primer pair used). Last elongation step: one cycle of 50 or 55°C for 45 s and 72°C for 3 min (Arens et al. 1995).

Detection of microsatellite polymorphisms

The samples were prepared for PAGE electrophoresis by adding an equal volume of urea loading-buffer (8 M urea, 10 mM of NaOH and 0.05% bromophenol blue) to the reaction mixtures. Samples were denatured at 80°C for 5 min, followed by quenching on ice. Samples were electrophoresed on 6% denaturing polyacrylamide sequencing gels (gels were prepared in Tris-borate buffer, pH 8, 8 M urea,) using a sequencing-gel electrophoresis apparatus (Bio Rad Sequigen). Each gel was run during 2–3 h at a 110-W constant current. The DNA bands were visualized by silver staining according to the Silver sequence DNA sequencing system (Promega). A permanent record of the gels was made using EDF films (Kodak). The sizes of the PCR products were estimated by comparison to an accompanying sequence reaction ladder using plasmid pGEM-3Zf(+) and the 24-mer pUC forward primer. The allele sizes were recorded in base pairs. For this, the most intense, upper band was used.

Data analysis

Diversity measure

Gene diversity was calculated per locus, accession and species. $D=1-(1/m)\sum_l\sum_u p^2_{lu}$, where p_{lu} is the frequency of the *u*th allele at the *l*th locus and m is the number of loci (Weir 1996).

Genetic distance measures

Three different methods for measuring genetic distances were calculated. The proportion of shared alleles distance (Dps) (Bowcock et al. 1994), measures the similarity between the multiple locus genotype of two individuals.

Nei's standard genetic distance (Dns) (Nei 1972), is based on genetic identity (I) which is the ratio of the proportions of alleles that are alike between and within populations.

The Delta mu-squared distance (Ddm) for microsatellites, proposed by Goldstein et al. (1995), is based on the variance in the repeat number (difference between allele size). The genetic diversity index, genetic distances, and bootstrap re-sampling of 100 replicates were calculated using the computer programme MICROSAT version 1.5 (Goldstein et al. 1995).

Dendrograms

PHYLIP (Phylogeny Inference Package) Version 3.57c by Felsenstein (1995) was used to construct trees from the genetic distances Dps, Dns and Ddm by the Neighbor-Joining (NJ) method (unrooted trees) (Saitou and Nei 1987) using the outgroup option. To test the robustness of tree topology, the trees were compiled by CONSENSE (part of the PHYLIP package). Linear regressions were calculated using SYSTAT.

Table 1 Plant material used

Species	Accession No. (CGN)	Address and number	Collection site	No. of individuals	Total No. individuals	Mating behavior[a]
Lycopersicon esculentum Mill. (cultivars)	Money maker			5		
	Calypso			5	15	SC/FO
Lycopersicon esculentum var. *cerasiforme* (Dunal) Alef.	922421[b]	INOCRO		5		
L*ycopersicon cheesmanii* Riley	923679	LA 1409	Isabela: Punta Albermarle, Galapagos Islands, Ecuador	5		
	927134	LA 1448	Santa Cruz: Puerta Ayora, Pelican Bay, Galapagos Islands, Ecuador	5	15	SC/A
	15916	LA 1402	Fernandina: W of Punta Espinoza, Galapagos Islands, Ecuador	5		
Lycopersicon pimpinellifolium (Juslen.) Mill.	923479	LA 414 (UNCDUC)	Daular, Guayas, Ecuador	5		
	927148[b]	SOFGPB		6	16	SC/FO
	18401	LA 419 CAAES	Calderon, Manabi, Ecuador	5		
Lycopersicon parviflorum C.M. Rick et al.	923492	LA 735	Huanuco-Cerro de Pasco, Huanuco, Peru	9	9	SC /A
Lycopersicon chmielewskii C.M. Rick et al.	15816	SOFGPB	Ecuador	10	10	SC/FO
Lycopersicon hirsutum Humb. & Bonpl.	923544[b]	PI 127826 (BARCPI)		5		
	923673[b]	PI 308182 (BARCPI)		5	15	SC/FO/SI
	15370[b]	RUITER		5		
Lycopersicon pennellii (Correll) D'Arcy	927171[b]	ARSNPI		5	15	SI
Lycopersicon pennellii var. *puberulum* (Correll) D'Arcy	923549	LA 750	Ica-Nazca, Ica, Peru	5		
	15819	LA 1926	Agua Perdida (Rio Ingenio), Ica, Peru	5		
Lycopersicon peruvianum var. *humifusum* C.H. Müller	923453	LA 2334	San Juan, Cajamarca, Peru	5		
Lycopersicon peruvianum (L.) Mill. (accessions called *L.per*/Caj in this study)	15392	LA 2172	Cuyca, Cajamarca, Peru	5		
	15798	LA 1708	Chamaya, Cajamarca, Peru	5	20	SI
	15801	LA 2157	Tunel Chotano, Cajamarca, Peru	5		
Lycopersicon peruvianum (L.) Mill.	17046	LA 372	Culebras, Ancash, Peru	5		
	17047	LA 462	Azapa, Tarapaca, Chile	5		
	17048	LA 1333	Loma Camana, Arequipa, Peru	5		
	17049	LA 1373	Asia, Lima, Peru	5	35	SI
	17050	LA 1274	Pacaibamba, Lima, Peru	5		
	17051	LA 1945	Caraveli, Arequipa, Peru	5		
	17052	LA 1955	Matarani, Arequipa, Peru	5		
Lycopersicon chilense Dunal	15530[b]	MPITU		5		
	15531[b]	LHTUIN		5	15	SI
	15532[b]	LHTUIN		5		
Total number of plants					165	

[a] SI, self incompatible. FO, facultative outcrossing. SC, self compatible. A, autogamous
[b] Collection site unknown

Results

Microsatellite amplification across species

A representative set of microsatellite markers that gave strong amplification products of quality 1 and 2 in *L. esculentum* was selected from the list of Smulders et al. (1997).

Not all primer sets gave good amplification in all accessions analysed. In some species, blanks were found indicating the presence of null-alleles, possibly through mutations in the primer binding sites. The number of these null-alleles was dependent on the accession as well as on the microsatellite used. Most were found in accessions of *Lycopersicon parviflorum , Lycopersicon chmielewskii, Lycopersicon hirsutum, Lycopersicon pennellii* and *Lycopersicon peruvianum* from Cajamarca. The loci LESSRPSPGA and LEATPACAb were most-prone to giving null-alleles whereas loci like LECHSOD, LEACC2G, LECHI3 and LEEF1Ab amplified in almost every plant under the conditions employed (data not shown).

Microsatellite polymorphism in the genus *Lycopersicon*

The number of alleles varied widely among the 17 loci analysed, ranging from 1 to 17 alleles (see Table 2). Polymorphism was detected in 16 out of the 17 loci analysed. A total of 144 different alleles were found among the 31 accessions of *Lycopersicon.* There was no correlation between the degree of polymorphism (counted as the absolute number of different alleles per microsatellite) and the total length of the microsatellite repeat (data not shown).

Five out of 16 polymorphic microsatellites showed allele sizes in accordance with the Stepwise Mutation Model (SMM) (Ohta and Kimura 1973). This model implies that an allele mutates only by losing or gaining a single tandem repeat. An allele in state *i* -an allele with *i* repeats, is assumed to mutate to an allele either in state *i*+1 or *i*–1 with an equal probability. LECHSOD, LECHI3, LEDIH4RE, LELEUZIP and LECAB9 have allele sizes that differ by one or more repeat units only. Locus LEACC2G has most of the alleles with the expected sizes plus one extra non-fitting allele of 147 bp. Five microsatellites appeared to have two allele series (LEGTOM5, LPHFS24, LEATPACAb, LEGAST1 and LEWIPIG), e.g. the two LEWIPIG series are: 249-251-253-255-257-259-261-265-269 and 250-252-254-256-258-260-264-266. The remaining microsatellites had many alleles of unexpected sizes occurring across the genus. Thus, 31% of the microsatellites analysed appear to be concordant with the single-step mutation model and another 31% possibly follow this model with minor modification, although the presence of these series by itself does not prove that they arose by mutation with one repeat unit at a time.

Taking into account that five out of nine species of *Lycopersicon* are selfing species, the variation in these species can be expected to be represented by the presence of different homozygotes rather than heterozygotes. Therefore, the gene diversity index is a more-sensitive measure of variability than the heterozygosity index, which measures the number of heterozygotes observed (Weir 1996). The microsatellites analysed showed different levels of polymorphism, measured as gene diversity (D) (Table 2). The degree of gene diversity per locus showed a linear correlation with the number of alleles (r^2=0.53, F=15.83, α=0.001).

Distribution of genetic diversity across the genus

Among the 16 polymorphic microsatellites, 14 had at least one allele unique to a species (see Table 2). In some cases, all the individuals within a species had the same unique allele, whereas in other cases an allele was specific to a particular accession within the species. Sixty six alleles (46%) were found to be specific to a certain species, 35 of which (24%) were specific to a certain accession within the species. All the species, except *L. esculentum,* had at least one unique allele. *L. pennellii, L. peruvianum* and *Lycopersicon chilense* were the species with the highest proportion of unique alleles (Table 3).

The average gene diversity of species within the *Lycopersicon* genus, calculated with microsatellite data (see Table 3), is in agreement with the description of their respective mating systems (Rick 1979). The self-compatible species (*L. esculentum, Lycopersicon cheesmanii, L. parviflorum* and *L. chmielewskii)* had in general a lower gene diversity than the outcrossing species, reflecting their autogamous mode of reproduction. The accessions of *Lycopersicon pimpinellifolium,* treated separately, had diversity ”0”, but when the accessions are pooled together the diversity went up to 0.197 for the species (see Table 3). Thus, for this species the diversity was found exclusively between accessions. The average diversity of *L. pimpinellifolium* was the highest of any of the self-compatible species. This is important for breeding, because *L. pimpinellifolium* is closely related to the cultivated tomato; therefore, the genetic variation found in this species is easily accessible for *L. esculentum.* This result is in agreement with previous RFLP studies (Miller and Tanksley 1990). The low values of *L. pennellii* and *L. hirsutum* compared to *L. chilense* and *L. peruvianum* can be due to the fact that some biotypes of these species are self-compatible. The outcrossing, self-incompatible species, *L. chilense* and *L. peruvianum,* had the highest gene-diversity value. These species contribute a substantial amount of genetic variation to the genus. The *L. peruvianum* accessions from Cajamarca have less genetic diversity than the rest of *L. peruvianum,* which may be due to the presence of some self-compatible accessions within this group. Rick (1982) also reported a northern race of *L. peruvianum* which was self-compatible and appears to be naturally inbreeding.

Table 2 Microsatellite core sequences, allele size, number of alleles per locus, gene diversity (D), number of specific alleles per locus, specific alleles present only in one accession within a species and specific alleles present in more than one accession within the species

Locus	Core sequence[b]	Allele size	No. alleles	D	No. specific alleles	Accession-specific alleles[e]	Species-specific alleles[f]
LEHSC80P	$(GAA/G)_5(GAA)_{5\text{-}1}$	190	1	–	–	–	–
LERNALX	$(ATT)_{6\text{-}1}$	128-129	2	0.0442	0	–	–
LECHSOD[a]	$(CTT)_6$	198-201-207	3	0.1917	1	207/*L. per* 923453	–
LEACC2G	$(AAAT)_3$	147[c]-149-153-157	4	0.1200	1	157/*L. per* 17051	–
LECHI3[a]	$(TA)_{6\text{-}1}(GA)_4$	124-126-128-130-132	5	0.1435	1	128/*L. per* 17050	–
LEDIH4RE[a]	$(AAT)_5(AAG)_2$	89-92-95-98-104	5	0.1192	3	98/*L. chi* 15532	89/*L. per*-104/*L. chi*
LELEUZIP[a]	$(AAG)_{6\text{-}1}TT(GAT)_7$	98-101-104-107-110	5	0.3098			
LEEF1Ab	$(TTA)_{4\text{-}1}$ A(TTA)A$(TTA)_4$	241-242-243-244-247-252	6	0.1230	3	252/*L. per* 17046-241/*L. per*17049	244/*L. per*
LECAB9[a]	$(TA)_6(CA)_3$	118-120-122-124-126-128-130-134	8	0.2590	1	134/*L. chi* 15532	–
LE2A11	$(ATCT)_{5\text{-}1}$	143-157-159-165-170-171-172-176	8	0.1520	6	165/*L. per*17050-170/*L. hir* 923673-171/*L. per*17047	143/*L. per*-172/ *L. hir*-176/*L. pen*
LE21085	$(TA)_2(TAT)_{9\text{-}1}$	90-93-96-99-102-105-108-111-118-119-127-132	12	0.2759	6	102/Lchi15530-119/*L. pim*927148-127/*L. per*17046-132/*L. chi*15532	111/*L. che* -118/*L. par*
LEGTOM5[d]	$(TA)_{10}$	162-174-*175*-176-*177*-178-180-*181*-182-*183*-184-186-188	13	0.4986	6	174/*L. hir*923673-175/ *L. pen* 923549	162,183/*L. chm*-181/ *L. par*-184*L. pen*
LPHFS24[d]	$(TA)_6$	149-*150*-151-*152*-153-*154*-157-*158*-159-*160*-161-*162*-170	13	0.2251	6	153/*L. per*17052-154/ *L. pen*927171-161/*L. chi*15532-162/ *L. per*15392-170/*L. per*923453	149/*L. per*
LEATPACAb[d]	$(GA)_7$	169-175-179-183-*184*-185-*186*-187-*188*-189-193-205-213	13	0.3047	5	169/*L. chi*15530-175/ *L. per*17051–193/*L. per*17050-205/ *L. chi*15531	213/*L. per*.Caj
LEGAST1[d]	$(TA)_{12}(TG)_{8\text{-}2}$	122-*127*-128-*129*-130-*131*-132-*133*-134-*135*-*137*-*139*-*145*-*151*	14	0.3852	7	134/*L. per*17047-151/*L. pim*18401	122,127,132/*L. per*-129/ *L. hir*-139/*L. che*
LESSRPSPGa	$(TATT)_5$	205-210-214-215-216-217-219-221-222-234-242-243-244-245-249	15	0.2560	8	205/*L. chi*15530-210/ *L. per*17050-214/*L. per*17046	215,249/*L. per*-222/ *L. per*.Caj-242/ *L. hir*-244/*L. chi*
LEWIPIG[d]	$(CT)_4(AT)_4$	*249*-250-*251*-252-*253*-254-*255*-256-*257*-258-*259*-260-*261*-264-*265*-266-*269*	17	0.3431	12	250/*L. pim*927148-251/ *L. chi*15530-255/ *L. per*17051-260,266/*L. pen*15819-	252,253,264,265,269/ *L. pen* -254/ *L. chi*-258/*L. per*

[a] Loci with allele sizes in accordance with the Stepwise Mutation Model (SMM)
[b] Core sequence from *Lycopersicon* EMBL database sequences (Smulders et al. 1997)
[c] Allele size deviating in a series otherwise in accordance to the SMM
[d] Loci with two series following the SSM, second series indicated by italics
[e] Allele size/accession for which it is unique
[f] Allele size / species for which it is unique

Table 3 Diversity (D) per accession and per species. Number of specific alleles that are present only in one accession within a species (A), number of specific alleles that are present in more than one accession within a species (B) and the proportional number of specific alleles relative to the number of individuals analyzed per species (C)

Species	Accession No. (CGN)	D per accession	Average D per species	No. specific alleles		Proportion of specific alleles per species (C)
				A	B	
Lycopersicon esculentum (cultivars)	Money maker	0.0000		0		
	Calypso	0.0313	0.0343	0	0	0
Lycopersicon esculentum var. *cerasiforme*	922421	0.0263		0		
Lycopersicon cheesmanii	923679	0.0200		0		
	927134	0.1456	0.1122	0	2	0.13
	15916	0.0000		0		
Lycopersicon pimpinellifolium	923479	0.0000		0		
	927148	0.0000	0.1968	2	0	0.18
	18401	0.0000		1		
Lycopersicon parviflorum	923492	0.0665	0.0665	*	2	0.22
Lycopersicon chmielewskii var. *minutum*	15816	0.0571	0.0571	*	2	0.20
Lycopersicon hirsutum	923544	0.0000		0		
	923673	0.0625	0.1697	2	3	0.29
	15370	0.0113		0		
Lycopersicon pennellii	927171	0.0700		1		
Lycopersicon pennellii var. *puberulum*	923549	0.1440	0.2401	1	7	0.65
	15819	0.1334		2		
Lycopersicon peruvianum var. *humifusum*	923453	0.1680		2		
Lycopersicon peruvianum	15392	0.1303	0.3631	1	2	0.25
(called *L. per*/Caj in this study)	15798	0.1129		0		
	15801	0.0583		0		
Lycopersicon peruvianum	17046	0.3525		3		
	17047	0.3630		2		
	17048	0.2156		0		
	17049	0.3277	0.5685	1	10	0.69
	17050	0.3728		4		
	17051	0.2888		3		
	17052	0.2171		1		
Lycopersicon chilense	15530	0.4367		4		
	15531	0.3475	0.5172	1	3	0.80
	15532	0.4252		4		

Relationships among accessions

Dendrograms based on three different methods of measuring genetic distance, will be confronted here with previous biosystematic studies of the genus *Lycopersicon* and with previous dendrograms reconstructed from molecular-genetic analyses. The dendrograms drawn from the genetic distances among accessions, using the proportion of shared alleles (Fig. 1a) and the Nei standard genetic index (data not shown), have similar topologies. Most of the accessions clustered in groups corresponding to previously established species, although often with low bootstrap values (Fig. 1a). The highest bootstrap values were shown by the group containing all the accessions that belong to the closely related and self-compatible species, *L. esculentum, L. cheesmanii* and *L. pimpinellifolium* (supported by 92% bootstrap re-sampling), and the group containing all accessions from *L. hirsutum* (supported by 86% bootstrap re-sampling). The accessions of *L. esculentum, L. cheesmanii* and *L. pimpinellifolium* were not well-differentiated into species, which may be explained by introgression between these closely related species (Rick 1979). The *L. peruvianum* var. *humifusum* accession (CGN 923453) and three accessions of *L. peruvianum* (CGN 15392, 15798 and 15801), which all came from the Cajamarca region in northern Peru, clustered closer to the *L. esculentum, L. cheesmanii* and *L. pimpinellifolium* group and to *L. parviflorum* and *L. chmielewskii* than to the rest of *L. peruvianum*, which come from southern Peru and northern Chile. Miller and Tanksley (1990) found the same relationship with an accession belonging to *L. peruvianum* var. *humifusum* using RFLP data. Rick (1963) found a complete barrier between var. *humifusum* and most of the other *L. peruvianum* in hybridization experiments.

A dendrogram reconstructed for accessions with a distance measure specifically developed for microsatellite data, the Delta mu-squared distance (Ddm) (Fig. 1b), shows differences with the former tree and failed in grouping accessions from some of the species, e.g. the three self-

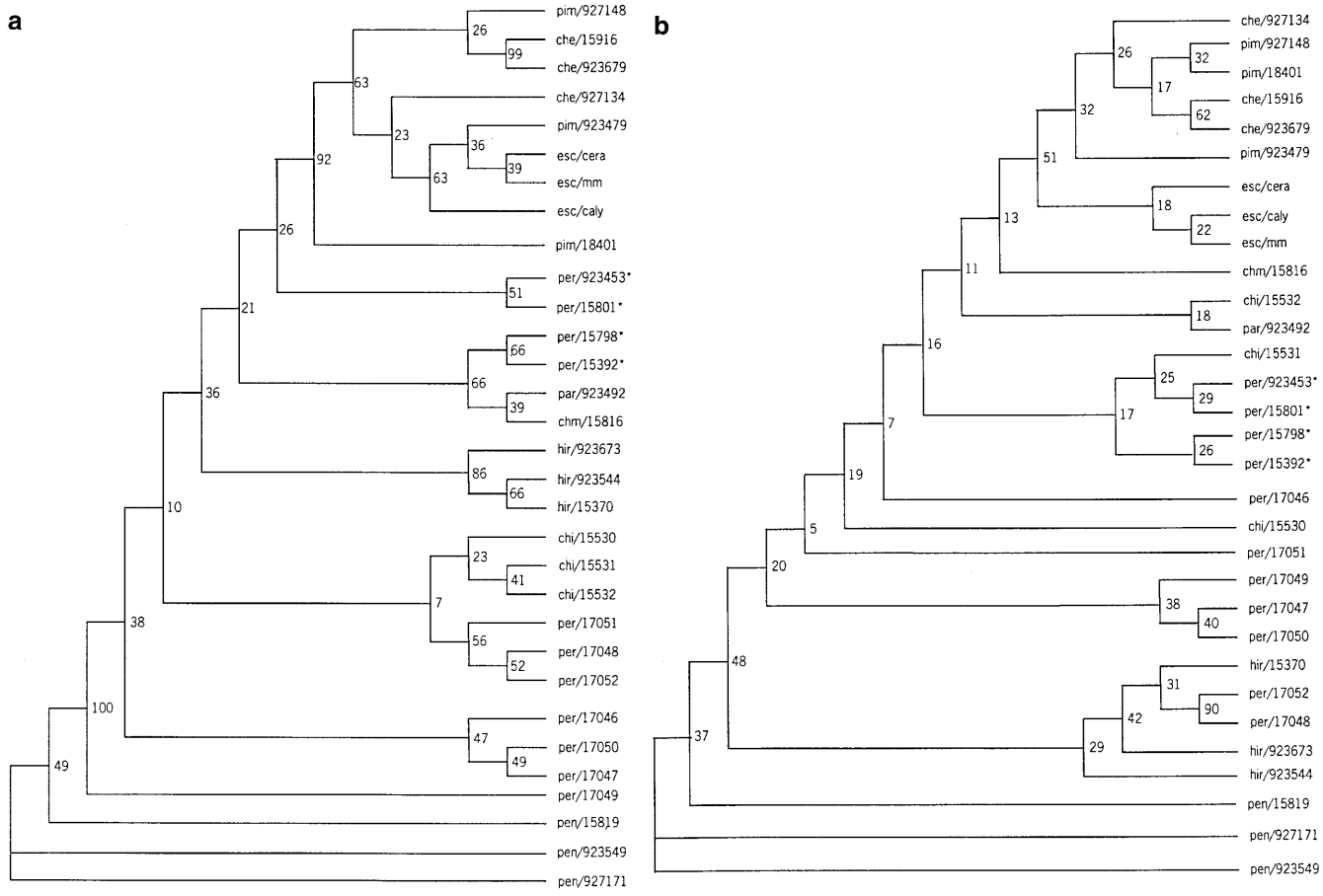

Fig. 1a, b Neighbor-Joining (*NJ*) dendrograms of *Lycopersicon* accessions based on 16 microsatellites. *L. pennellii* accession number 927171 was used as an outgroup option. **a** Using a Proportion of shared-alleles (Dps) matrix; **b** Using a Delta mu-squared distance (Ddm) matrix. The *numbers* at the nodes are bootstrap values for 100 bootstrap re-samplings. **L. peruvianum* accessions from Cajamarca, Peru

incompatible *L. chilense* accessions were scattered across separate clusters. One of these accessions, *L. chilense* CGN15532, was grouped even far more closely with the self-compatible species *L. parviflorum* than the latter's self-compatible sibling species, *L. chmielewskii* (cf. Rick 1979), is (Fig. 1b). Also the bootstrap re-sampling values were generally smaller than in the Dps tree, indicating a less-reliable tree topology (compare Fig. 1a with b).

Relationships among species

Species dendrograms were made by pooling the data from the accessions into species. Based on the accession trees in Figs. 1a and b and evidence from other sources (Miller and Tanksley 1990; Rick 1963, as discussed in the previous paragraph), the four *L. peruvianum* accessions from Cajamarca were treated as a separate taxon named *L. peruvianum*/Caj.The bootstrap percentage values for the species tree were generally higher than when treating the accessions individually.

Dps and Dns (the latter not shown) trees showed a major division of the genus into two clusters (Fig. 2a). One major cluster grouped the self-compatible species, *L. pimpinellifolium, L. cheesmanii, L. esculentum, L. parviflorum* and *L. chmielewskii* (49% of bootstrap re-sampling). Within this cluster of selfing species, the red-fruited species, *L. pimpinellifolium, L. cheesmanii* and *L. esculentum*, were subsequently separated from the green-fruited *L. parviflorum* and *L. chmielewskii.* The combination of the red-fruited species was even supported by 100% of bootstraps. The other major cluster consisted of the self-incompatible species *L. peruvianum, L. chilense* and *L. peruvianum*/Caj. Within this group, *L. peruvianum* was more closely related to *L. chilense* (the two grouped with 99% of bootstrap re-sampling) than to *L. peruvianum*/Caj. By and large, the tree topology is also congruent with the main division in crossability groups, which shows a major split between the self-incompatible species, on the one hand, and the self-compatible species together with *L. hirsutum* and *L. pennellii*, on the other (Rick 1979). In comparison, the Ddm tree (Fig. 2b) clustered the red-fruited self-compatible species *L. esculentum, L. pimpinellifolium* and *L. cheesmanii* too, but with a bootstrap support of only 77%. The closely related sibling species *L. chmielewskii* and *L. parviflorum* show a less closer relationship than in the Dps tree. As in the accession trees (Fig. 1a and b), the Ddm species tree showed lower bootstrap support than the Dps species tree (compare Fig. 2a and b).

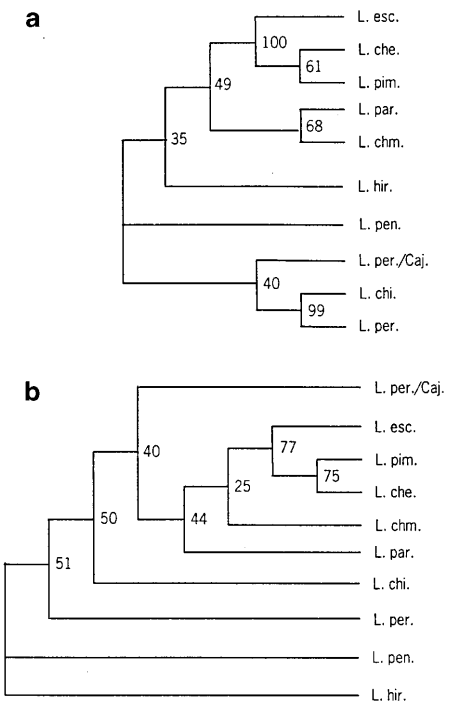

Fig. 2a, b NJ dendrograms of *Lycopersicon* species based on 16 microsatellites. *L. pennellii* was used as an outgroup option; **a** Using a Proportion of shared-alleles (Dps) matrix, **b** Using a Delta mu-squared distance (Ddm) matrix. The *numbers* at the nodes are bootstrap values for 100 bootstrap re-samplings. *L. per*/caj: *L. peruvianum* accessions from Cajamarca, Peru

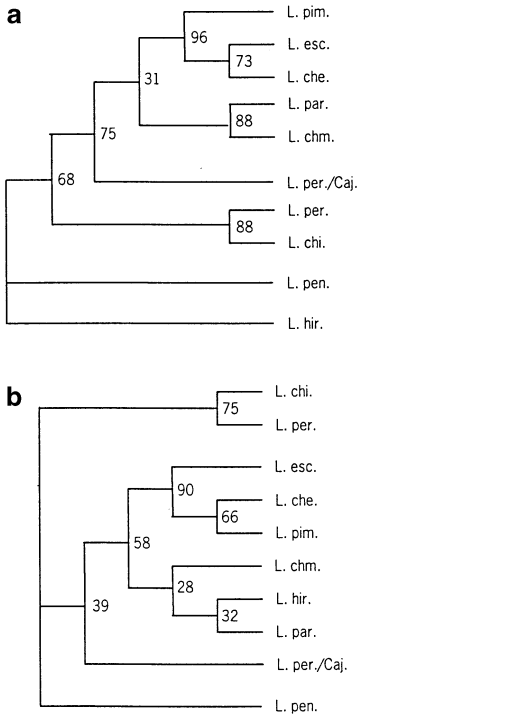

Fig. 3a, b NJ dendrograms of *Lycopersicon* species based on microsatellites differing in polymorphism rate. *L. pennellii* was used as an outgroup option; **a** Tree made with low-polymorphic (D<0.245) microsatellites (*lp.ms*), **b** Tree made with highly polymorphic (D>0.245) microsatellites (*hp.ms*). The *numbers* at the nodes are bootstrap values for 100 bootstrap re-samplings. *L.per/*Caj*: L. peruvianum* accessions from Cajamarca, Peru

Comparison between trees based on high-polymorphic microsatellites and those based on low-polymorphic microsatellites

In an attempt to verify the efficiency of microsatellites in studying species relationships according to their differences in polymorphism rate all over the genus, the microsatellites were classified into two groups based on the gene diversity index (D) (see Table 2). A Dps matrix was constructed using data from microsatellites with a high degree of polymorphism (hp.stms, microsatellites with D>0.245), and one using data from microsatellites with a lower degree of polymorphism (lp.stms, microsatelles with D<0.245). The lp.stms tree (Fig. 3a) showed one main group consisting of a cluster containing *L. cheesmanii*, *L. esculentum* and *L. pimpinellifolium* (96% bootstrap support) together with a cluster containing *L. chmielewskii* and *L. parviflorum* (88% bootstrap support); *L. peruvianum* from Cajamarca, Peru, was added to this main group at 75% bootstrap value. Another cluster relates *L. chilense* with *L. peruvianum* (supported by 88% bootstrap re-sampling).

In general, the tree based on lp.stms showed higher values for bootstrap re-sampling (with just one value below 50) than the hp.stms tree. Moreover, the tree based on hp.stms clustered *L. parviflorum* most closely to L. *hirsutum* (see Fig. 3b), whereas *L. parviflorum* was reported to be most closely related to *L. chmielewskii* on the grounds of mating system and morphological characters (Rick 1979). The close relationship of *L. chmielewskii* and *L. parviflorum* was also reflected in the results based on the RFLPs of Miller and Tanksley (1990). In the lp.stms tree, on the other hand, *L. hirsutum* had a more-separate position closer to *L. pennellii*, which is also more in accord with the RFLP tree from Miller and Tanksley (1990). Taken together, these observations suggest that a tree based on low-polymorphic microsatellites shows a more-reliable topology than a tree based on high-polymorphic microsatellites.

Discussion

Microsatellites were used to evaluate genetic diversity and species relationships within the genus *Lycopersicon*. In the following, the feasibility of using microsatellites for these two purposes is discussed under separate headings.

Genetic diversity

The 17 microsatellite primer sets analysed gave good amplification across the nine species of *Lycopersicon*

with 16 microsatellites being polymorphic, leading to an average of eight alleles per microsatellite. Null-alleles appeared in some accessions, that is the microsatellite did not amplify from all the plants in a particular species, but nevertheless information could be extracted from the greater part of the microsatellites from all species. The locus LEATPACAb, which was the most prone to give null-alleles, still amplified in 26 out of 31 accessions. This is relatively favourable, since the transferability of microsatellites across species is often quite low; for instance, in the reasonably intercrossable species group of *Lactuca sativa*/*Lactuca serriola*, *Lactuca saligna* and *Lactuca virosa* only half of the microsatellite PCR primer sets developed for *L. sativa* amplified a product in the latter two species (Van de Wiel et al. 1999). The amount of variation found in *Lycopersicon* (counted as the number of polymorphic microsatellites between species) ranged from 29% of microsatellites being polymorphic between *L. esculentum* and *L. cheesmanii* to 94% of microsatellites being polymorphic between *L. esculentum* and *L. peruvianum*. Variation within species ranged from 1 microsatellite out of the 16 showing polymorphisms between two *L. esculentum* cultivars to 15 out of 16 microsatellites showing polymorphisms between *L. peruvianum* accessions. Thus, this set of microsatellites appears to be suitable for use as genetic markers across the genus *Lycopersicon* for several purposes. For instance, the large amount of variation exemplified by the common occurrence of alleles unique to accessions (24% of the alleles) and to species (46% of the alleles) makes the microsatellites a versatile tool for germplasm management, with some microsatellites being useful for the identification of accessions and others for the identification of (sub)species. The usefulness for studying species relationships is discussed in the next section.

The average gene diversity per species of the *Lycopersicon* genus calculated with microsatellite data matches the description of the mating systems by Rick (1979). As expected, most of the variation was found in the outcrossing species, *L. peruvianum* being the species with the highest genetic variation, followed by *L. chilense* and *L. pennellii*. *L. pimpinellifolium* showed the highest genetic variation among the selfing species. These results are in line with previous RFLP-based results published by Miller and Tanksley (1990).

Species relationships

In the present study, the proportion of shared alleles distance (Dps) and the Nei's standard distances (Dns) (based on the variance in allele frequencies) calculated from 16 microsatellites produced similar dendrograms, which matched previous RFLP-based dendrograms published by Miller and Tanksley (1990) and the biosystematics of the genus *Lycopersicon* based on crossability, mating system and morphological characters published by Rick (1979). As with these authors, the results deviated from organellar DNA-based trees (Palmer and Zamir 1982; MacClean and Hanson 1986). Organellar introgression is one of the possible explanations for this (Miller and Tanksley 1990).

The bootstrap support of Dps trees was generally higher than the trees made with a distance measure that was specifically designed for microsatellites, the delta mu-squared distance (Ddm). The Ddm is based on a step-wise mutation model (SMM) for microsatellite variation and thus produces an ordering of alleles according to size between two populations; i.e. the distance is based on the variance in repeat number (difference between allele sizes). In contrast, Dps and Dns distances are based on the variance in allele frequencies and assume an infinite-allele mutation model (IAM) in which a mutation can involve any number of tandem repeats and always results in an allele not encountered before in the population. In the case studied here, where 69% of the loci analyzed did not strictly follow the predictions from the SMM, i.e. the allele size variation was not in accord with variation being limited to the basic repeat number, the Ddm can be expected to perform poorly. By using only the loci behaving according to the SSM, a better result might be obtained, but in this case the number of microsatellites left (5 out of 16) would be too small to make reliable tree reconstructions. According to Goldstein et al. (1995), distances like Ddm will do progressively worse as the mutation model becomes more like the infinite alleles' model. For a mixed model the proportion of single-step mutations is reduced (and the proportion of arbitrary sizes goes up) so the performance of this distance will decline. Takezaki and Nei (1996), using computer simulation, investigated the efficiencies of different distance measures in obtaining the correct tree topology using microsatellite data. These authors demonstrated that the Ddm distance is generally less useful than traditional distance measures for phylogenetic inference even for microsatellites, for which this distance measure was specifically designed.

Since this study shows sizeable variation in the variability of individual microsatellites, it can be hypothesized that not all of them might be equally valuable for reconstructing species relationships. Low-polymorphic microsatellites will have a lower mutation rate and be more stable over longer periods of time, i.e. show less homoplasies and therefore may contain more phylogenetic information at the species level. Indeed in this study, a tree based on relatively low polymorphic microsatellites (lp.stms) has been shown to have a generally higher bootstrap support than one based on relatively highly polymorphic microsatellites (hp.stms). The lp.stms tree also showed a topology more in line with trees based on previous biosystematic studies (Rick 1979) and RFLP data (Miller and Tanksley 1990). Taken together, this implies that microsatellites can be used for establishing relationships between related species, but that taking into account their level of variability may improve their effectiveness.

Acknowledgements The authors thank Dr. M. Koornneef for support and critically reading the manuscript, Ir. I. Boukema for providing seeds from the CGN collection, especially those not yet taken up in the collection (all those having six-digit numbers), and additional information about the accessions, and Dr. G. Bredemeijer for advice on the microsatellite primer sets.

References

Arens P, Bredemeijer G, Smulders MJM, Vosman B (1995) Identification of tomato cultivars using microsatellites. Acta Hort 412:49-57

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution human evolutionary trees with polymorphic microsatellites. Nature 368:455-457

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

Felsenstein J (1995) PHYLIP, phylogeny inference package (version 3.5c). Department of Genetics, University of Washington, Seattle

Foolad MR, Jones RA, Rodriguez RL (1993) RAPD markers for constructing intraspecific tomato genetic maps. Plant Cell Rep 12:293-297

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

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471

Kaemmer D, Weising K, Beyermann B, Börner T, Epplen JT, Kahl G (1995) Oligonucleotide fingerprinting of tomato DNA. Plant Breed 114:12-17

McClean PE, Hanson MR (1986) Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics 112:649-667

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 (1972) Genetic distance between populations. Am Nat 106:283-291

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

Palmer JD, Zamir D (1982) Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proc Natl Acad Sci USA 79:5006-5010

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222

Rick CM (1963) Barriers to interbreeding in *Lycopersicon peruvianum*. Evolution 17:216-232

Rick CM (1979) Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In: Hawkes JG, Nester RN, Skelding AD (eds) The biology and taxonomy of the solanaceae. Linnean Society Symposium Series 7. Academic Press, London, pp 667-678

Rick CM (1982) A new self-compatible wild population of *L.peruvianum*. TGC Report 32:43-44

Rus-Kortekaas W, Smulders MJM, Arens P, Vosman B (1994) Direct comparison of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. Genome 37:375-381

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

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

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144: 389-399

Van de Wiel C, Arens P, Vosman B (1999) Microsatellite retrieval in lettuce (*Lactuca sativa* L.). Genome 42:139-149

Van der Beek H, Verkerk R, Zabel P, Lindhout P (1992) Mapping strategy for resistance genes in tomato based on RFLPs between cultivars: CF9 (resistance to *Cladosporium fulvum*) on chromosome 1. Theor Appl Genet 84:106-112

Vosman B, Arens P, Rus-Kortekaas W, Smulders MJM (1992) Identification of highly polymorphic DNA regions in tomato. Theor Appl Genet 85:239-244

Warnock SJ (1988) A review of taxonomy and phylogeny of the genus *Lycopersicon*. HortScience 23:669-673

Weir BS (1996) Genetic data analysis. II. Methods for discrete population genetic data. 2nd edn. Sinauer Associates, Sunderland

Williams CE, St. Clair DA (1993) Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Genome 36:619-630