Theor Appl Genet (1990) 80:437–448

RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*

J. C. Miller * and S. D. Tanksley

Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA

Received February 13, 1990; Accepted May 3, 1990
Communicated by F. Salamini

Summary. Forty single-copy, nuclear probes of known chromosomal position were used to examine restriction fragment length polymorphism in the tomato genus *Lycopersion*. The probes were from three libraries: one cDNA, and two genomic libraries – one genomic made with EcoRI and the other with PstI. Total DNA from 156 plants representing eight species was cut with five different restriction enzymes and scored in 198 probe-enzyme combinations. Genetic distances between accessions (populations) and species were calculated from the resultant restriction patterns and proportion of shared bands. Accessions belonging to the same species largely clustered together, confirming their current classification. However, one mountain accession, classified as *L. peruvianum* var. *humifusum* (LA2150), was sufficiently distinct from the other accessions of *L. peruvianum* that it may qualify as a separate species *L. esculentum* and *L. pimpinellifolium* were the least clearly differentiated, possibly reflecting introgressive hybridization, known to have been promoted by man in recent history. Dendrograms constructed from cDNA versus genomic clones were nearly identical in their general grouping of species. The dendrograms revealed two major dichotomies in the genus: one corresponding to mating behavior [self-compatible (SC) versus self-incompatible (SI) species] and the other corresponding to fruit color (red versus green-fruited species). The ratio of within- versus between-accession diversity was much lower for SC species, indicating that most of the diversity within these species exists between populations, rather than within populations. Overall, the amount of genetic variation in the SI species far exceeded that found in SC species. This result is exemplified by the fact that more genetic variation could be found *within* a single accession of one of the SI species (e.g., *L. peruvianum*) than *among* all accessions tested of any one of the SC species (e.g., *L. esculentum* or *L. pimpinellifolium*). Results from this study are discussed in relationship to germ plasm collection/utilization and with regard to the use of RFLPs in tomato breeding and genetics.

Key words: *Lycopersicon* – Phylogeny – RFLP – Genetic variation – Breeding system

Introduction

Muller (1940) published one of the first modern taxonomic treatments of the genus *Lycopersicon*, which includes the cultivated tomato and its wild relatives. The many previously described "species" were reduced to six botanical species and six additional subspecies and varieties. Luckwill (1943) had been simultaneously reorganizing the genus and published a similar version 3 year later. Rick (1979) produced a classification based upon crossability relationships. A more detailed overview of the taxonomic history of *Lycopersicon* can be found elsewhere (Warnock 1988).

Two molecular systematic studies have been reported in *Lycopersicon*, both based on organellar DNA. These previous studies are not in perfect agreement with each other or with previous taxonomic data. The study of Palmer and Zamir (1982), based on chloroplast DNA (cpDNA), places *L. chmielewskii* close to *L. peruvianum*, while crossability data (Rick 1979) suggests a more distant relationship between these two. McClean and Hansen's (1986) mitochondrial DNA (mtDN)A phylogeny places *L. chmielewskii* closest to *L. pennellii*, and also places *L. hirsutum* as *L. esculentum's* closest relative

* Current address: Department of Agronomy, University of Wisconsin, Madison, WI 53706, USA

and the two other red-fruited species, *L. pimpinellifolium* and *L. cheesmanii,* closer to the green-fruited species than to the cultivated tomato.

Organellar introgression may be one possible explanation for the disparate results of these studies. Ferris et al. (1983) showed that mtDNA can move from one species into another across a sympatric zone. Organelles also account for only a fraction of a plant's total genetic material. For these reasons and others, we decided to conduct a study of genetic variation in the genus *Lycopersicon* using RFLP probes, and we present the first phylogeny based wholly on nuclear DNA.

Materials and methods

Plant DNA

Total DNA was isolated from leaves of the plants listed in Table 1. The procedure used was the same as that of Bernatzky and Tanksley (1986a), but the CsCl/ethidium bromide centrifugation was omitted. Accessions of *Lycopersicon* spp. were kindly supplied by Dr. C. M. Rick, University of California Davis, and were chosen to encompass the diversity typically found in each species. The genus contains species with three types of mating behavior: self-pollinating, facultative outcrossing, and self-incompatible. Two plants were individually analyzed from each accession of the self-pollinating species *L. esculentum* Mill., *L. pimpinellifolium* (Jusl.) Mill., *L. cheesmanii* Riley, four plants of each accession of *L. chmielewskii* Rick, Kes., Fob. & Holle (facultatively outcrossing) and *L. parviflorum* Rick, Kes., Fob. & Holle (self-pollinating), and five plants of each accesion of the self-incompatible species [*L. pennellii* (Corr.) D'Arcy, *L. hirsutum* Humb. & Bonpl., *L. peruvianum* (L.) Mill]. Two replacement plants, one of *L. hirsutum* accession LA1918, and one of *L. esculentum* cultivar VF36, brought the total number of plants to 156.

Clones

Total DNA from leaves of tomato cultivar VF36 was digested separately with EcoRI and PstI (Bethesda Research Laboratories and New England Biolabs), and the resultant fragments from the two digests were separated on sucrose gradients (Maniatis et al. 1982). The 1.5–3.0 kb fractions were collected, and the fragments were ligated into EcoRi- or PstI-cut pUC8, which was then used to transform competent TB1 cells of *E. coli.* Colonies with inserts were selected by their growth on media containing X-gal and IPTG. Organellar and high-copy clones were screened out via colony hybridization with radiolabelled organellar (chloroplast and mitochondrial) and nuclear DNA. Plasmids from both the EcoRI and PstI libraries were isolated according to Birnboim and Doly (1979). Ten EcoRI and 15 PstI clones were then linkage-mapped against more than 200 previously mapped loci in an F_2 population of *L. esculentum* ("VF36") × *L. pennellii* (LA716) (Bernatzky and Tanksley 1986b and unpublished data). Map distances were calculated from the maximum likelihood equations of Allard (1954). Fifteen previously mapped, single-copy cDNA clones were also used. Locations of the loci corresponding to all 40 clones used in this study are shown in Fig. 1.

Hybridizations

DNAs of the 156 plants were cut with DraI, EcoRI, EcoRV, HindIII, and XbaI according to the supplier's instructions (Bethesda Research Laboratories and New England BioLabs), electrophoresed in 0.9% agarose gels, and Southern-blotted to GeneScreen Plus hybridization membrane (Biotechnology Systems, NEN Research Products) as per Bernatzky and Tanksley (1986c). Bacteriophage lambda and ΦX174, cut with HindIII and HaeIII, respectively, were included as size standards on each gel, as was "VF36" DNA. The Southern blots were probed with random, hexamer-labelled insert (Feinberg and Vogelstein 1983) isolated from each of the 40 clones. Hybridization and wash conditions were identical to those of Bernatzky and Tanksley (1986a), except that both the second and third wash were 1 × SSC (0.15 *M* NaCl, 0.025 *M* citrate), 0.1% SDS. CD72 and CD77 were not stored on XbaI-cut DNA, as the fragments were not well separated. This brought the number of probe-enzyme combinations from 200 to 198.

Calculations

The bands revealed by autoradiography were assigned molecular weights based upon their positions relative to the lambda and ΦX174 molecular weight standards. Genetic distance matrices were then generated for comparisons among individuals (or groups of individuals) based on the proportion of shared restriction fragments, using equations 5.53 through 5.55 from Nei (1987). Genetic distances were expressed as the proportion of nucleotide substitutions (Nei 1987). Analyses were performed on an IBM PC/AT Personal Computer using dBase III software by Ashton-Tate.

Dendrograms, based on genetic distances, were constructed via the unweighted pair-group method with arithmetic averages (Sneath and Sokal 1973). Standard errors for the nodes were determined by deleting data from each clone in turn, and then redrawing the trees each time to locate the nodes.

The number of unique restriction fragments and number of unique restriction patterns were also calculated for each accession, or group of accessions, and used as an estimate of the amount of overall genetic variation (Table 1).

Results and discussion

EcoRI and PstI libraries

After chloroplast and mitochondrial clones were screened out via colony hybridization, the PstI library was found to compris 92% single-cope clones, whereas only 36% of the EcoRI clones were single copy. PstI is sensitive to C-methylation and will not cleave at its recognition sequence of 5′-CTGCAG-3′ if the 5′ cytosine is methylated. In maize, it has been found that cloning with methylation-sensitive enzymes like PstI also yields libraries enriched for single-copy sequences (Burr et al. 1988).

Genetic variation: within- versus between-accession diversity

L. esculentum and other self-compatible species. The amount of genetic variation found within accessions versus between accessions (estimated as the average genetic distance between individuals within accessions versus between accessions) differed greatly among the species (Fig. 2). At the low end were the modern *L. esculentum*

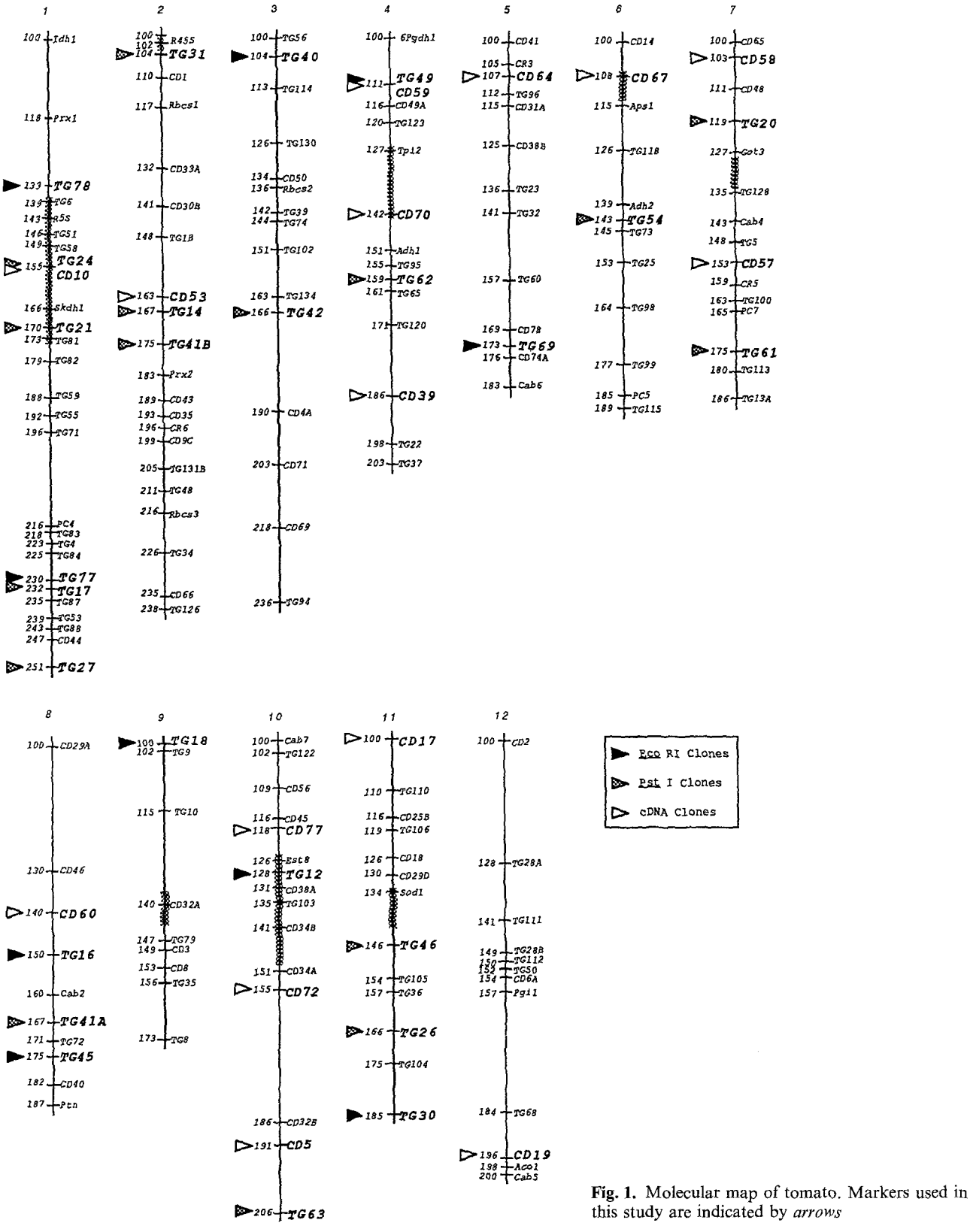

Fig. 1. Molecular map of tomato. Markers used in this study are indicated by *arrows*

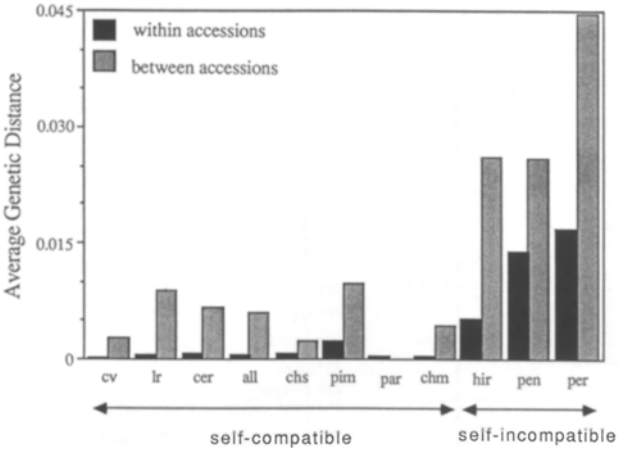

Fig. 2. Average genetic distance between individuals within accessions (*black*) versus between individuals from different accessions (*grey*)

cultivars. Here, less than 10% of the variation could be accounted for by within-accession variation. Most of the genetic variation was found between accessions. In contrast 2–3 times as much within-accession variation was found in *L. esculentum* var. *cerasiforme* and other landraces of *L. esculentum* from Latin America and South America. The lower diversity observed in the modern cultivars may reflect popular breeding methods (often single-seed descent or pedigree selection), which promote genetic uniformity.

As a group, accessions of the botanical variety *L. esculentum* var. *cerasiforme,* as well as other *L. esculentum* landraces, had levels of between-accession variation several-fold higher than detected in the modern cultivars (Fig. 2). The lower overall genetic diversity of modern cultivars may reflect the genetic "bottleneck" to which modern tomato cultivars were subjected during their introduction from Latin America to Europe (and later to the United States) by Spanish explorers in the middle part of this millenium. Presumably, only limited numbers of seeds (and accessions) were carried back and served as the basis of today's modern cultivars (Rick 1976).

Even though only a limited amount of variation was found between modern cultivars, it was still possible to distinguish all cultivars on the basis of one or more unique RFLP. Previous attempts to "fingerprint" tomato cultivars on the basis of protein profiles or isozymes had not been entirely successful due to the lack of sufficient variation. RFLPs may thus provide a method of uniquely identifying tomato cultivars, both for legal and practical purposes (e.g., testing for seed mixtures and establishing the identify of questionable seed material).

Self-incompatible species. The SI species had both within- and between-accession variation far in excess of that found for any of the SC species (Fig. 2). This result agrees with other studies associating increased genetic diversity with obligate outcrossing in tomato (Rick 1979, 1983) and other plant species (Gottlieb 1981). However, until now, standardized comparisons have not been made of genetic diversity both within and between species in the genus *Lycopersicon.* The highest level of genetic variation was found in *L. peruvianum,* which gave an average genetic distance between accessions of 0.045 – nearly tenfold higher than observed in any of the SC species. The other two SI species, *L. hirsutum* and *L. pennelii,* gave values nearly fivefold greater than the SC species.

It is striking to note that, on average, more genetic variation (defined either as the number of unique restriction fragments, Table 1, or as average genetic distance among individuals, (Fig. 2) can be found in a single accession from one of the SI species (e.g., *L. peruvianum*) than in *all* accessions (samples in this experiment) of any one of the SC species. This result emphasizes the importance of SI in maintaining genetic variation, and points out the necessity for using large population sizes, both for propagating and screening SI accessions at genetic repositories. These results reinforce previous studies demonstrating the tremendous genetic potential (for tomato improvement) existing in SI species (Rick 1982).

All of the species in the *Lycopersicon* genus can be crossed with the cultivated tomato and thus serve as potential sources of novel genes. From the perspective of tomato breeding, it of interest to ask where novel mutations (not currently found in tomato cultivars) exist within the genus. To answer its question, it was first necessary to remove, from each taxon data set, the restriction fragments held in common with *L. esculentum* cultivars. This was done, and the results are displayed in Fig. 3. Together, the three SI species accounted for 75% of the novel RFLPs found in this study. *L. peruvianum* was the highest and was estimated to contain 33% of the novel fragments. *L. pimpinellifolium* also contains many novel mutations and had the highest rating (6.9%) of any of the SC taxa. The latter species is also considered to be the species most closely related to the cultivated tomato, and genetic variation in this species is easily accessible to tomato breeders.

Rate of gain in genetic diversity versus number of accessions sampled

Given limited financial resources, collectors and curators of germ plasm banks face a number of difficult questions, including how much of one taxon to collect/maintain relative to another and what the value is of one group of material relative to another. These questions must be addressed in the context of the proposed use for the germ plasm bank, and undoubtedly there are a number of issues to consider in searching for the answers.

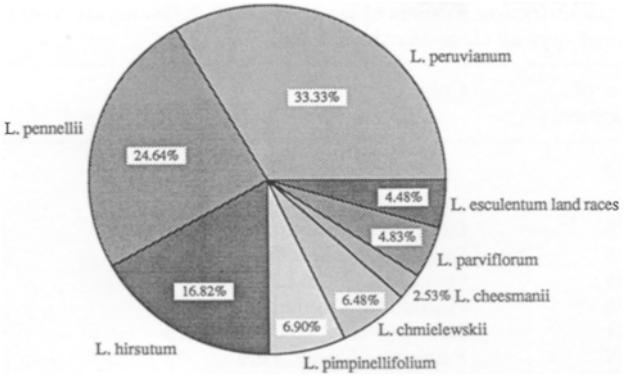

Fig. 3. Percentage unique restriction fragments found in each taxon but not found in modern *esculentum* cultivars. *L. esculentum* landraces include both var. *cerasiforme* and other landraces listed in Table 1

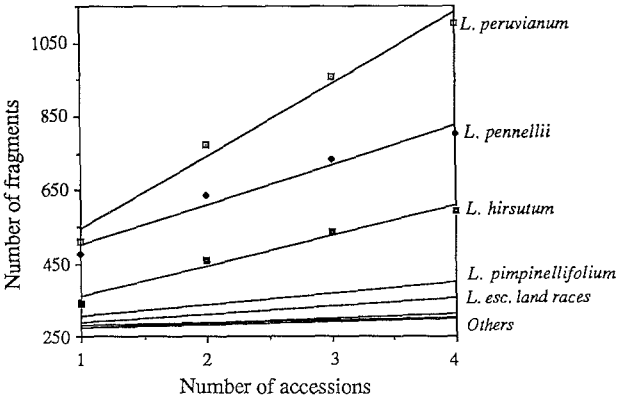

Fig. 4. Plot of average number of novel restriction fragments gained by sampling one, two, three or four accessions of each *Lycopersicon* taxon

One parameter that can be estimated from RFLP data, and that may have a bearing on these questions, is the likelihood that a particular accession will contribute new alleles to a collection. For example, adding another *L. cheesmanii* accession is associated with a certain probability of adding new alleles to a germ plasm bank beyond those that already exist in the collection. Species that are geographically diverse and polytypic, like *L. peruvianum,* are more likely to provide novel alleles to the gene bank. A quantitative estimate of the differences can be obtained by comparing the average genetic distances among accessions within taxa (Fig. 2) or by comparing the slopes of the plots of novel mutations gained as a function of sampling additional accessions of each species (Fig. 4). By either method, we estimate that adding a single *L. peruvianum* accession was 20 times more likely to add new alleles to the collection than adding a single accession of the cultivated tomato. These results can be extrapolated to all of the SI species, which, on average, had values at least twofold greater than any SC species (Fig. 4). Among the red-fruited, SC species (which are most closely related to the cultivated tomato), *L. pimpinellifolium* and *L. esculentum* landraces were the most diverse. Adding a single *L. pimpinellifolium* or *L. esculentum* landrace was 3–4 times more likely to add novel alleles than adding another *L. esculentum* cultivar or accession of *L. cheesmanii.*

Percentage polymorphic loci between L. esculentum cv VF 36 and other taxa

One of the primary motivations for constructing an RFLP map is its potential use in breeding and genetics. In order for the markers to segregate and thus be informative, they must show allelic differences between the parents used in the cross. For this reason, we have calculated the percentage of polymorphic loci between a standard tomato cultivar (VF36) and other accessions of both cultivated and wild tomato species (Table 2). These values give some indication of what proportion of the RFLP loci might be expected to segregate in these crosses. Crosses between modern cultivars show very little polymorphism. For a single probe-enzyme combination, only 6% of the probes were polymorphic on average, using a single restriction enzyme. By using additional restriction enzymes (five in this case), the percentage increases to 20%, still relatively low. These estimates are based on a sample of cultivars that include fresh market, processing, and cherry tomatoes as well as some older, obsolete varieties. If one examines the variation within each of these subtypes, the values are likely to be even lower.

As one begins to make comparisons between landraces or other species, the percentage of polymorphic probes rises dramatically. The value for landraces is double that found for modern cultivars and nearly five times greater in comparison with the closely-related, red-fruited, SC species (*pimpinellifolium* and *cheesmanii*). In crosses with green-fruited species (*parviflorum, chmielewskii, hirsutum, pennellii,* and *peruvianum*), nearly all of the probes reveal polymorphism. Thus, in breeding programs involving introgression from wild relatives, RFLP markers should be highly informative and may significantly reduce the time required to achieve the breeding goals (Tanksley et al. 1989).

The very narrow genetic base of modern tomato cultivars, combined with the SC nature of this species, probably accounts for the sparcity of polymorphism. Because of lack of genetic variation, tomato breeders have frequently relied upon land races and wild species as a source of important genes, such as those for disease and insect resistance (Rick 1982). As a result, while other segments of the genome may have only low levels of polymorphism among modern cultivars, the regions around these introgressed genes are highly polymorphic

Table 1. Plant materials used in this study. No. patt = number of unique restriction patterns observed over all clone-enzyme combinations. No. fragments = number of unique restriction fragments scored over all clone-enzyme combinations

Species/Accession (total)	No. of individuals	No. of patt.	No. of fragments	Collection site
L. esculentum (total)	46	196	428	
var. *cerasiforme*	22	145	401	
LA1455	2	43	278	Gral Teran, Nuevo Leon, Mexico
LA1511	2	40	276	Sete Lagoas, Brazil
LA1385	2	51	305	Quincemil, Cuzco, Peru
LA1334	2	42	276	Pescaderos, Arequipa, Peru
LA1324	2	44	281	Hda. Potrero, Apurimac, Peru
LA1312	2	40	279	Paisanato, Cuzco, Peru
LA1307	2	43	288	San Francisco, Ayacucho, Peru
LA1286	2	42	278	S. Martin de Pangoa, Junin, Peru
LA1226	2	39	277	Sucua, Morina-Santiago, Ecuador
LA1205	2	38	271	Copan, Honduras
LA1025	2	42	278	Wahiawa, Oahu, Hawaii, USA
average =		42	280	
other landraces: (total)	6	80	338	
LA 393	2	43	278	Chiclayo, Lambayeque, Peru
LA1568	2	46	280	Yucatan, Mexico
LA 408	2	41	274	Guayaquil, Ecuador
average =		43	277	
cultivars (total):	18	78	311	
Stone	2	40	272	
Marmande	2	41	274	
San Marzano	2	35	267	
M82	2	39	268	
New Yorker	2	42	270	
T5	2	43	275	
Vendor Tm-2[a]	2	42	267	
VFNT Cherry	2	42	266	
VF36	2	40	271	
average =		40	270	
average (all) =		42	276	
L. chmielewskii	16	92	309	
LA1316	4	49	268	Ocros, Ayachucho, Peru
LA1325	4	47	265	Puente Cunyac, Apurimac, Peru
LA1330	4	48	267	Hde. Francisco, Apurimac, Peru
LA2663	4	47	266	Tijtohaia, Cusco, Peru
average =		48	267	
L. pimpinellifolium	20	228	456	
LA 369	2	49	243	La Cantuta, Lima, Peru
LA 373	2	41	279	Culebras, Ancash, Peru
LA 411	2	56	314	Pichilingue, Los Rios, Ecuador
LA1246	2	46	280	La Toma, Loja, Ecuador
LA1332	2	55	290	Nazca, Ica, Peru
LA1579	2	47	303	Col. Pto. Cuatro No. 1, Lambayeque, Peru
LA1589	2	42	277	Viru-Calunga, La Libertad, Peru
LA1601	2	63	325	La Providencia, Ancash, Peru
LA1689	2	56	316	Piura-Castilla No.1, Piura, Peru
LA2176	2	45	284	Timbaruca, Cajamarca, Peru
average =		50	296	
L. cheesmanii	8	69	298	
var. *minor*	4	51	280	
LA1400	4	40	274	Isabela, Galapagos, Ecuador
LA 317	2	47	277	S. Bartolome, Galapagos, Ecuador
average =		44	276	

Table 1. (continued)

Species/Accession (total)	No. of individuals	No. of patt.	No. of fragments	Collection site
others	4	57	288	
LA1412	2	44	274	San Cristobal, Galapagos, Ecuador
LA 434	2	46	278	Sta. Cruz, Galapagos, Ecuador
average =		45	276	
L. parviflorum	4	47	276	
LA2133	4	47	276	Ona, Azuay, Ecuador
L. pennelli	20	525	802	
LA1272	5	156	513	Pisaquera, Lima, Peru
LA1657	5	167	539	Casma-Yuatan, Ancash, Peru
LA1809	5	136	469	El Horador, Piura, Peru
LA1940	5	118	386	Rio Atico km 26, Arequipa, Peru
average =		144	497	
L. hirsutum	20	296	593	
LA1264	5	96	345	Bucay, Chimborazo, Ecuador
LA1736	5	98	388	Pucutay, Piura, Peru
LA1361	5	81	340	Pariacoto, Ancash, Peru
LA1918	5	57	285	Liauta, Ayacuchio, Peru
average =		83	340	
L. peruvianum	20	358	1104	
var. *humifusum*	5	138	449	
LA2150	5	138	449	Pte. Muyuna, Cajamarca, Per
var. *glandulosum*	5	83	481	
LA1292	5	83	481	San Mateo, Lima, Peru
others	10	277	795	
LA 371	5	151	602	Supe, Lima, Peru
LA 462	5	130	507	Azapa, Tarapaca, Chile
average =		141	555	

Table 2. Percentage of probes detecting a plolymorphism between *L. esculentum* cultivar "VF36" and other accessions. Single enzyme: average value over the five restriction enzymes used in this study; all enzymes: locus considered polymorphic if one or more of the enzymes detected a polymorphism

VF36 compared to accessions of:	Single enzyme	All enzymes
L. esculentum ev	6	20
L. esculentum landraces	13	41
L. cheesmanii	27	67
L. pimpinellifolium	33	78
L. parviflorum	56	95
L. chmielewskii	62	95
L. hirsutum	80	100
L. pennelli	85	99
L. peruvianum	84	100

facilitating identification of diagnostic RFLP markers for these genes (Young et al. 1988; Osborn et al. 1987; Tanksley and Hewitt 1988; Sarfatti et al. 1989 and unpublished data this laboratory).

Because of the narrow genetic base of modern cultivars, it may not be possible to find RFLP markers segregating over all chromosome regions for any given cross between cultivars. In fact, it is likely that large tracts of genomes between cultivars may be common by recent descent. However, the chromosomal regions that do show polymorphism may be the ones that contribute most substantially to the characters differentiating the varieties.

Relationships among accessions

The dendrogram drawn from the genetic distances among accessions is shown in Fig. 5 (the distance matrix itself was not included in the interest of space). All but one of the accessions cluster in groups corresponding to previously established species. Accession LA2150, classified as *L. peruvianum* var. *humifusum*, groups with the SC species rather than with the other collections within *L. peruvianum* or with any of the other SI species. LA2150 was collected in Cajamarca, in mountainous northern Peru. The other three accessions of *L. peruvianum* were collected further south, two in Lima, one in northern Chile. *L. peruvianum* populations growing in the mountains in northern Peru have smaller flowers and leaves, and crossability studies divide *L. peruvianum* into two

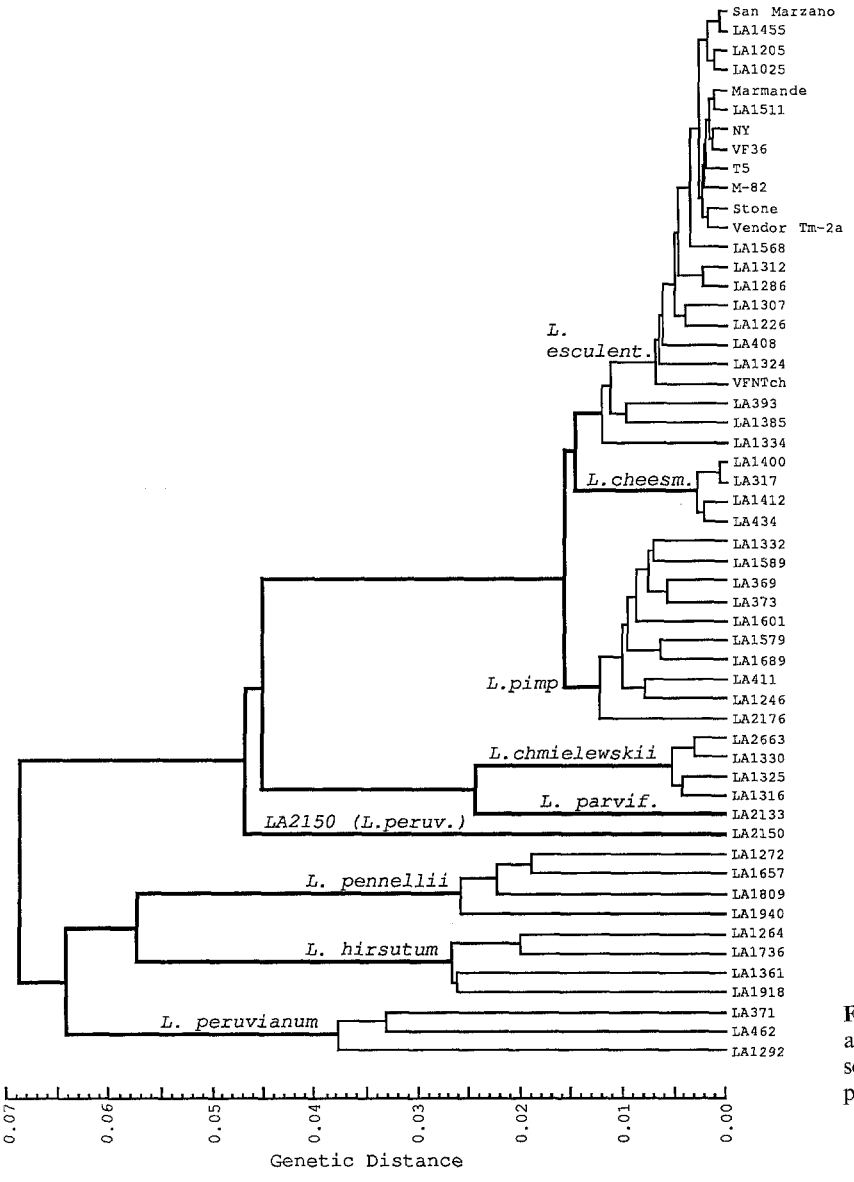

Fig. 5. Dendrogram of 54 *Lycopersicon* accessions based on 40 cloned nuclear DNA sequences. Bold lines indicate branching pattern of the species

groups, a "humifusum" group and a "peruvianum" group (Rick 1963, 1986).

Isozyme data show *L. peruvianum* var. *humifusum* to be more closely related to *L. parviflorum* (a green-fruited SC species) than to other populations of *L. peruvianum* (S. D. Tanksley and C. M. Rick, unpublished data). The findings reported in this study appear to support this notion. Genetic distances (based on RFLP data) between LA2150 and individual accessions of *L. parviflorum, L. chmielewskii,* and *L. peruvianum* are given in Table 3. According to these estimates, LA2150 is more closely related to *L. parviflorum* (distance = 0.0438) than to either *L. chmielewskii* (average distance = 0.0503) or *L. peruvianum* (average distance = 0.0515). Moreover, genetic distances between LA2150 and the other 53 *Lycopersicon* accessions ranged from 0.0435 to 0.0717 (data not shown), suggesting that LA2150 is at least as divergent from other *peruvianum* accessions as it is from other *Lycopersicon* species. More accessions of *L. parviflorum, L. chmielewskii,* and *L. peruvianum* var. *humifusum* should be examined to better determine the exact genetic relationships among these taxa.

Relationships among species

To make a species phylogeny, we pooled the data from the accessions into species. Since data from our studies showed that *L. peruvianum* accession LA2150 is distinct from the other accession of *L. peruvianum,* we decided to treat it as a separate taxon. The distance matrix based upon the data from all 198 probe-enzyme combinations

Table 3. Genetic distance values (× 100) between *L. peruvianum* accession LA2150 and accessions of *L. parviflorum*, *L. chmielewskii*, and other accessions of *L. peruvianum*

	L. parviflorum	*L. chmielewski*				*L. peruvianum*		
	LA2133	LA1316	LA1325	LA1330	LA2663	LA371	LA1292	LA462
LA2150	4.38	5.04	5.09	4.99	5.02	5.07	5.34	5.04
LA2133	–	2.44	2.50	2.42	2.45	6.36	6.07	6.10
LA1316	–	–	0.42	0.58	0.41	6.64	6.54	6.18
LA1325	–	–	–	0.65	0.45	6.76	6.46	6.24
LA1330	–	–	–	–	0.31	6.77	6.52	6.21
LA2663	–	–	–	–	–	6.67	6.52	6.21
LA 371	–	–	–	–	–	–	3.74	3.30
LA1292	–	–	–	–	–	–	–	2.79

Table 4. Genetic distance values used for calculating phylogenetic relationships among species Proportion of fragments shared (upper right) and estimated number of nucleotide substitution per 100 base pairs (lower left). Numbers in parentheses are the number of fragments found within each species

	esc	*pim*	*che*	*par*	*chm*	LA2150	*pen*	*hir*	*per*
L. esculentum	(428)	0.805	0.705	0.474	0.421	0.461	0.358	0.376	0.392
L. pimpinellifolium	1.223	(456)	0.679	0.470	0.424	0.475	0.382	0.378	0.403
L. cheesmanii	1.995	2.217	(298)	0.484	0.445	0.420	0.273	0.314	0.307
L. parviflorum	4.379	4.438	4.252	(276)	0.633	0.475	0.299	0.343	0.322
L. chmielewskii	5.127	5.084	4.778	2.638	(309)	0.430	0.279	0.313	0.317
L. peruvianum LA2150	4.561	4.370	5.131	4.379	4.988	(449)	0.326	0.365	0.375
L. penellii	6.147	5.739	7.895	7.304	7.746	6.738	(802)	0.434	0.433
L. hirsutum	5.830	5.806	6.976	6.417	7.010	6.025	4.925	(593)	0.417
L. peruvianum	5.574	5.401	7.133	6.825	6.919	5.178	4.904	5.178	(956)

is given in Table 4, and Fig. 6 shows the dendrogram drawn from this data.

Classically, fruit color has been the major taxonomic character used to divide the genus into red-fruited (subgenus *Eulycopersicon*) and green-fruited (subgenus *Eriopersicon*) species. The first branching in Fig. 6, however, separates the SI species (*L. pennellii, L. hirsutum,* and *L. peruvianum*) form the SC species (*L. esculentum, L. pimpinellifolium, L. cheesmanii, L. parviflorum,* and *L. chmielewskii*), with the exception of LA2150. Division on the basis of fruit color is seen at the next node separating *L. esculentum, L. pimpinellifolium,* and *L. cheesmanii* (red-fruited) from the rest of the genus (green-fruited).

LA2150 is SI, yet it groups, albeit at the farthest node, with SC species. The branches between LA2150, the *L. parviflorum*/*L. chmielewski* complex, and the red-fruited species are very close. It is unlikely that this accession's similarity to the SC taxa is due to recent introgressive hybridization, since it is currently isolated in the mountains of northern Peru. It may, however, represent the closest living SI relative of SC species. Since the SC species cluster together, the loss of self-incompatibility may be the result of a single, monophyletic event.

The three SI species (*hirsutum, pennellii,* and *peruvianum*) are well-separated from the other taxa. Because of the high level of within-species variation, the branching order of these three species from each other cannot be clearly discerned. It seems unlikely that the addition of more low-copy clones to the data set will resolve this issue. We regard this as an effective "trifurcation." Nei (1987) states that this method of calculating distance should be used only for closely related taxa, specifically where the distance calculated is less than 0.05. It should be noted that the three SI species are at this limit of resolution. Sequencing, or examination of conserved portions of the genome, may be better suited to studying these branch points.

Comparison of accession and species trees

The topologies of the accession and species trees are slightly different. In the accession tree (Fig. 5), *L. esculentum* seems to be more closely related to *L. cheesmanii,* while in the species tree (Fig. 6), *L. pimpinellifolium* is closer to the cultivated tomato. This discrepancy may

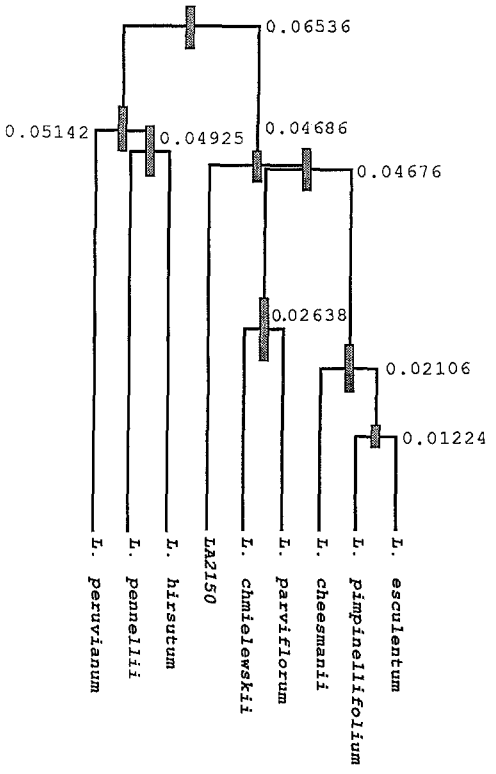

Fig. 6. Dendrogram of *Lycopersicon* species based on all 40 probes. Shaded bars represent one standard error on either side of the node

result from pooling the data into species groups. *L. pimpinellifolium* and landraces of *L. esculentum* are sympatric in Central America, and interspecific hybridization seems likely, especially since they are grown opportunistically for food. A few restriction fragments shared due to recent crossing between *L. esculentum* and *L. pimpinellifolium* would carry as much weight as the same number of ancestral similarities between *L. esculentum* and *L. cheesmanii*. They would represent recent mating, rather than shared ancestry. If these recent similarities are present in only a few accessions, and these accessions are pooled together with many others in which they are absent, then these introgressed fragments would bias the analysis. Wild landraces of both *L. esculentum* and *L. pimpinellifolium* occur throughout Central and South America and would have had opportunity to hybridize, while *L. cheesmanii* is reproductively isolated on the Galapagos Islands. Retriculation may be occurring between *L. esculentum* and *L. pimpinellifolium* due to the lack of reproductive isolation.

Although the accessions of each red-fruited species clustered together, differentiation required use of all the data. The *L. esculentum* and *L. pimpinellifolium* accessions only separated into discrete species clusters with the addition of the last two or three clones. Removal of any one of TG14, TG16, TG17, or CD58 from the data set is enough to cause *L. pimpinellifolium* accession LA2176 to group with the *L. esculentum* accessions, supporting the idea that these species are not well-differentiated and that some natural hybridization between these groups is likely to have occurred.

Possible effects of insertions/deletions on data interpretation

Available data suggests that some and perhaps much of the genetic variation uncovered by RFLP analysis in plants is due to insertions/deletions or other types of DNA rearrangements (Schwarz-Sommer et al. 1985; McCouch et al. 1988; Roth et al. 1989). A point mutation is likely to be detected only by the specific restriction enzyme whose recognition site has been affected. In contrast, deletions/insertions may be detected by all enzymes whose restriction sites bound the insertion/deletion site. This is an important consideration for phylogenetic studies since RFLPs detected by the same probe, but with different restriction enzymes, may not represent independent mutational events. Ideally, for phylogenetic studies, all detected mutations should be independent.

In the study described here, we used 40 different probes, each with five different restriction enzymes. One of the reasons for using more than one enzyme was to test the efficiency of various restriction enzymes in detecting polymorphism (Miller and Tanksley 1990). However, because more than one enzyme was used, the possibility exists that some of the detected mutations are not independent. To hedge against potential biases engendered by this situation, we generated separate genetic distance matrices (and dendrograms) based on the 40 probes for *each* enzyme (thus five dendrograms were produced). Since each dendrogram was based on only a single enzyme, all mutational events should be independent. The dendrograms produced by this exercise (not shown), while based on less total data, gave overall similar groupings as the dendrogram based on the entire data set and support the conclusion drawn from this study.

Effects of clone source

Figure 7a–c shows dendrograms based upon the EcoRI clones, the PstI clones, and the cDNA clones, respectively. Coding sequences are commonly believed to be highly conserved, and therefore a tree based on cDNA clones might exhibit reduced internode distances. However, this was not the case, as the three trees were not appreciably different from one another in overall length or topology. The dendogram based on the EcoRI clones has an overlap in the standard errors of the nodes of the red-fruited taxa, and the cDNA tree has a minor change in topology among the SI species. The errors of the SI taxa overlap under all circumstances, however, so this minor switch is probably not important.

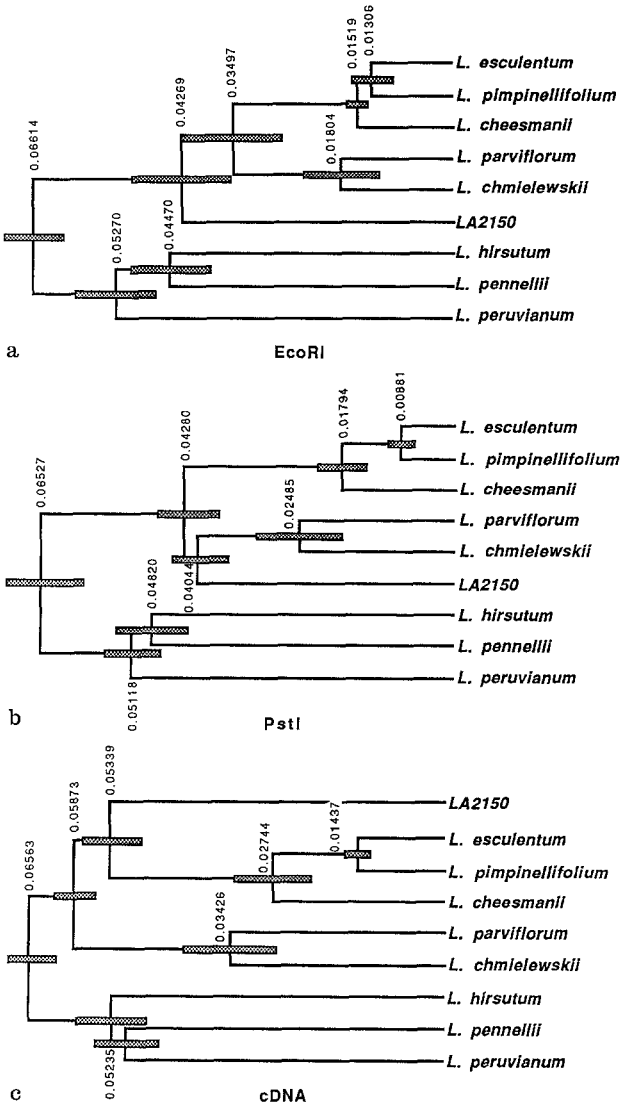

Fig. 7a–c. Dendrograms of *Lycopersicon* species based on: *Eco*RI clones only (**a**), *Pst*I clones only (**b**), cDNA clones only (**c**)

The branch separating the SC groups occurs much earlier in the tree produced by the cDNA clones than in the other two. In the EcoRI and PstI trees, LA2150 and the SC species begin to separate at distances of 0.0427 and 0.0428, respectively (Fig. 7). This divergence in the cDNA library occurs earlier, at 0.0534. In spite of this, the overall lengths of the dendrograms derived from each of the three libraries remain similar (0.0661, 0.0653, and 0.0656).

When the data were subdivided by library, the position of LA2150 did not vary significantly. In the tree generated by the EcoRI clones, it lay just outside of the SC species, while the PstI clones showed it to be more closely related to the *L. parviflorum/L. chmielewskii* group. The cDNA library placed it close to the red-fruited taxa. In all three libraries there was overlap in the standard errors of the nodes separating the red-fruited from green-fruited species and the branch removing LA2150. Here, as in the three SI species, there seems to be an effective trifurcation leading to the red-fruited species, the *L. parviflorum/L. chmielewskii* complex, and LA2150.

Conclusions

Using a set of 40 nuclear probes for RFLP analysis, it has been possible to estimate the extent and distribution of genetic variation within and between taxa in the genus *Lycopersicon*. Results indicate that the level of DNA polymorphism within accessions and species is highly correlated with mating system. The SI species harbor, on average, more than tenfold more variation within accessions compared with SC species. In terms of total variation it is estimated that the three SI, and thus obligately outcrossing species (*hirsutum, peruvianum* and *pennellii*), together contain nearly three times as much genetic variation as the four SC, self-pollinating species (*esculentum, pimpinellifolium, cheesmanii,* and *parviflorum*) combined. Remarkably, a single accession from one of the SI species (e.g., *peruvianum*) is likely to contain more detectable variation at the DNA level than all tested accessions of any one of the SC species.

The RFLP-based dendrogram produced for the *Lycopersicon* genus is largely consistent with much of the previous classification based on morphology and crossability and matches the natural history of these species. No significant differences were seen in the structure of dendrograms independently derived from cDNA or genomic clones (PstI or EcoRI) . From the fact that dendrograms based on independent data look so similar, one can be confident that the deduced phylogenetic relationships accurately reflect evolutionary relationships in this genus.

The species dendrogram suggests that two of the most obvious dimorphisms in the genus, fruit color (green versus red) and compatibility reaction (SI versus SC), are most likely monophyletic in origin. The fact that all SC species (both red- and green-fruited) cluster together forming a dichotomy in the dendrogram supports this hypothesis. Likewise, within the SC group, the red-fruited species (*esculentum, pimpinellifolium,* and *cheesmanii*) form a tight, cohesive cluster.

A possible exception to current classification in *Lycopersicon* was manifested by a single accession of *L. peruvianum* var. *humifusum* (LA2150) from northern Peru. This accession (and other related accessions of this race) are morphologically distinct from and exhibit a high degree of cross-incompatibility with other *L. peruvianum* accessions (Rick 1963, 1986). Data from this study indicate that this accession is no more closely related (at

the DNA level) to other *L. peruvianum* accessions than to other species in the genus. In fact, *L. parviflorum*, a SC, green-fruited species, showed the closest genetic affinity with LA1250. While *L. parviflorum* and LA1250 are morphologically distinct, they both differ from other *peruvianum* accessions in having more diminutive leaf and floral structure (Rick 1976, 1986). It is tempting to speculate that LA1250 and related *L. peruvianum* var. *humifusum* accessions are descendents of the SI genetic pool that gave rise to the SC *L. parviflorum* and its sibling SC species *L. chmielewskii*.

Acknowledgements. This work was supported by grants from the USDA/CRGO, the National Science Foundation, and the Binational Agricultural Research and Development Fund (BARD No. US-1388-87) to SDT. Thanks to C. M. Rick for advice and comments, M. Bonierbale, W. Burnquist, and M. Sorrels for editorial comments, and to Steffie David for typing the manuscript.

References

Allard RW (1954) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278

Bernatzky R, Tanksley SD (1986a) Genetics of actin-related sequences in tomato. Theor Appl Genet 72:314–321

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

Bernatzky R, Tanksley SD (1986c) Methods for detection of single or low copy sequences in tomato on southern blots. Plant Mol Biol Rep 4:37–41

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513

Burr B, Burr RA, Thompson KH, Albertsen MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Ferris SD, Sage RD, Huang C-M, Nielson JT, Ritte U, Wilson AC (1983) Flow of mitochondrial DNA across a species boundary. Proc Natl Acad Sci USA 80:2290–2294

Gottlieb LD (1981) Electrophoretic evidence and plant populations. Phytochemistry 7:1–46

Luckwill LC (1943) The genus *Lycopersicon*, an historical, biological, and taxonomic survey of the wild and cultivated tomatoes. Aberdeen University Studies, Scotland, No. 120

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

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

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Miller JC, Tanksley SD (1990) Effects of restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor Appl Genet 80:437–448

Muller CH (1940) A revision of the genus *Lycopersicon*. United States Department of Agriculture, Miscellaneous Publication No. 382

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Osborn TC, Alexander DC, Fobes JF (1987) Identification of restriction frgament length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73:350–356

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

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

Rick CM (1976) Tomato. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 262–273

Rick CM (1979) Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In: Hawkes JC, Lester RN, Skelding AD (eds) The biology of the Solanaceae. Academic Press, New York, pp 1–27

Rick CM (1982) The potential of exotic germ plasm for tomato improvement. In: Vasil IK, Scowcroft WR, Frey KJ (eds) Plant improvement and somatic cell genetics. Academic Press, New York, pp 478–495

Rick CM (1983) Evolution of mating systems: evidence from allozyme variation. In: Genetics: new frontiers. Proc. XV Inter. Congress Genetics, pp 215–221

Rick CM (1986) Reproductive isolation in the *Lycopersicon peruvianum* complex. In: D'Arcy WG (ed) Solanaceae biology and systematics. Columbia Press, New York, pp 477–495

Roth EJ, Frazier BL, Apuya NR, Lark KG (1989) Genetic variation in an inbred plant: variation in tissue cultures of soybean [*Glycine max*(L) Merrill]. Genetics 121:359–368

Sarfatti M, Katan J, Fluhr R, Zamir D (1989) An RFLP marker in tomato linked to the *Fusarium oxysporum* resistance gene *I2*. Theor Appl Genet 78:755–759

Schwarz-Sommer Z, Gierl A, Cuypers H, Peterson PA, Saedler H (1985) Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J 4:591–597

Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman, San Francisco

Tanksley SD, Hewitt J (1988) Use of molecular markers in breeding for soluble solids content in tomato – a re-examination. Theor Appl Genet 75:811–823

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technol 7:257–264

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

Young ND, Zamir D, Ganal MW, Tanksley SD (1988) Use of isogenic lines and simultaneous probing to identify DNA markers tightly linked to the *TM-2a* gene in tomato. Genetics 120:579–585