

Genetic variability in *Lycopersicon* **species and their genetic relationships**

M. P. Bret6, M. J. Asins, and E. A. Carbonell

Instituto Valenciano de Investigaciones Agrarias, Apartado Oficial, 46113 Moncada, Valencia, Spain

Received August 5, 1992; Accepted September 3, 1992 Communicated by P. M. A. Tigerstedt

Summary. Genetic diversity has to be described and measured in order to establish breeding strategies and manage genetic resources. It is also fundamental to develop a comparative intraspecific study before attempting to discuss and conclude any phylogenetic relationship. The genetic variability of *Lycopersicon* species was studied using starch gel electrophoresis of 11 enzymatic systems in a hierarchical fashion. The species with the greatest genetic variability are *L. chiIense, L. peruvianum* and *L. pennellii,* mainly due to the within-line component. *L. ehmielewskii, L. parviflorum* and *L. pimpinellifolium* show an intermediate total variability and their between-component clearly predominates over the within-component. The least variable species are *L. cheesmanii* and *L. esculentum.* Cluster analysis resulted in three main groups: one formed by the cultigen, *L. pimpinellifolium, L. cheesmanii* and *L. peruvianum;* another by two species with self-incompatibility systems, *L. pennelli* and *L. chilense;* and another by two autogamous species *L. chmielewskii* and *L. parviflorum.* With respect to *L. esculentum* the farthest related species is *Solanum rickii* and the closest, *L. pimpinellifolium.*

Key words: *Lycopersicon -* Isozymes - Genetic diversity - Genetic distance - Salt tolerance

Introduction

Genetic erosion, mainly caused by the action of man, is being stopped by the collecting and preserving of samples of representative genetic variability from those species that are or might be useful to mankind. In this context, and also to obtain important information for the germ plasm bank users, FAO, the Commission of the European Communities (1991) and other national and international organizations demand adequate studies on genetic diversity in many if not all these species. A vast array of genetic variation contained in the cultivated tomato and related taxa has been collected, increased and maintained by the Tomato Genetics Stock Center (TGSC) at the Department of Vegetable Crops, University of California (Davis) (Rick 1987).

The genus *Lycopersicon* contains nine species that are distributed throughout Central and South America. Rick and his associates (reviewed by Doebley 1989) have conducted extensive analysis of isozyme variation among these species, including studies on breeding systems, genetic control, linkage and populational variation. Much of their work focused on three red-fruited species, *L. cheesmanii, L. pimpinellifolium* and *L. eseulentum,* which together form a natural assemblage within the genus. The isozyme analysis of 14 loci by Rick and Fobes (1975) indicated that *L. pimpinellifolium* (or its ancestral form) gave rise to *L. cheesmanii* (wild taxon endemic of the Galápagos Islands). These authors also concluded that *L. esculentum* var 'cerasiforme' (the weedy form of the species) gave rise to the cultigen. A general and more recent view of the *Lyeopersicon* genus is reported by Hanson and McClean (1987), where a tomato mtDNA phenogram is constructed from restriction fragment hybridization data. In this study, *L. hirsutum* was found to be the most closely related species to *L. esculentum,* and *L. pennelli* groups with *L. chmilewskii* and *L. peruvianum.*

Our ultimate goal is to introduce tolerance to high concentrations of salts $(171.1 \text{ mM of NaCl, minimum})$

Correspondence to: M.J. Asins

into the cultivated tomato. After screening several well-known salt-tolerant cultivars for this character, we have focused on their wild relatives. It can be deduced from Harlan (1976) that the closer the wild relative, the better it is for breeding purposes. Therefore, genetic diversity has to be described and measured in order to establish breeding strategies and manage genetic resources (Lefort-Buson etal. 1988; Asins and Carbonell 1989). Because a large number of samples may need to be screened, markers should be inexpensive and easily scored. To date, allozyme electrophoresis has been the genetic technique most widely used to study the genetic structure of populations.

The objective of investigation discussed in the present paper was to obtain a general view of the genetic relationships among most *Lycopersicon* species, based on their pattern of variation as shown by allozyme electrophoresis. In order to do so, we consider it to be fundamental to develop a comparative intraspecific study in *Lycopersicon* before attempting to discuss and conclude any phylogenetic relationship.

Materials and methods

The material studied is shown in Table l. The entries were kindly supplied by C. M. Rick from the 'Tomato Genetics Stock Center', by F. Nuez (Univ. Politécnica de Valencia, Spain), the FAL (FRG) and the N. I. Vavilov Institute.

Several anthers from each of five plants per entry were homogenized in 0.1 M TRIS-HCl, pH 8.5, and 1% glutation. Only one plant from *S. rickii* flowered, therefore this species was not included in the variability study. Crude extracts were electrophoresed in 12% horizontal starch gels following methods described in Wendel and Weeden (1989). The enzymatic systems examined were: triose phosphate isomerase (TPI), glutamate axaloacetate transaminase (GOT), phosphoglucoisomerase (PGI), phosphoglucomutase (PGM), aconitase (ACO), malic enzyme (ME), 6-phosphoglucose dehydrogenase (6PG), glutamic acid dehydrogenase (GDH), superoxide dismutase (SOD), esterase (EST) and malic acid dehydrogenase (MDH). The staining methods are described in Vallejos (1983) for ACO, 6PG,

Table 1. Plant material. Mating system taken from Rick (1987)

Species		of entries	Symbol Number Mating system				
L. esculentum	esc	15	Autogamous				
L. esculentum var cerasiforme	cer		Autogamous				
L. chilense	chi	5	Allogamous $(SI)^a$				
L. cheesmanii	che	5	Autogamous				
L. chmielewskii	chm		Facultative (SC)				
L. pennellii	pen	5	Facultative (SC/SI)				
L. parviflorum	par		Autogamous				
L. pimpinellifolium	pim	5	Facultative (SC)				
L. peruvianum	per	5	Facultative (SC/SI)				
S. rickii	ric		9				

a SC, Self-compatible; SI, self-incompatible

PGI, PGM; Tanksley and Rick (1980) for EST; Shaw and Prasad (1970) for TPI; Goodman et ai. (1982) for GOT, ME, GDH, MDH; and Baum and Scandalios (1979) for SOD.

The tomato cv 'Madrigal' was used as a control in every analysis. Enzymatic bands were numbered with respect to their relative mobility in the gels relative to the anodal front and following the genetic control data by Tanksley and Rick (1980), Tanksley (1983) and Rick (1983). However, since the genetic control of SOD, EST and MDH is not very clear, the phenotypic patterns have been used as variables for these systems in the analysis. In all other cases, variables have been defined in terms of the presence/absence of a determined allele at a locus. In the case of the esterases, two zones of independent variation were observed; therefore the patterns for each zone have been considered separately.

Genetic variability has been studied in a hierarchical fashion as follows: (1) between species and (2) within species. These were further subdivided into: (a) between entries per species, and (b) within entries. The statistical methodology used to measure these variability components has been described and discussed in Asins and Carbonel (1986, 1987, 1989). Variability components were measured as follows:

1) correspondence analysis (Benzecri and Benzecri 1980; Greenacre, 1984) using data obtained by summing up all of the variables in each species. Dendrograms were based on the chi-square distance of Benzecri (1970), and the aggregation method was that of Lance and Williams (1967). The analysis was run twice, with and without *S. rickii.*

2) number of polymorphic loci (P), mean number of alleles per locus (A), coefficient of variation of the mean gene frequency (CV), mean percentage of heterozygotes (H), considering only those loci without null alleles, and mean chi-square distance obtained by averaging all chi-square distance between plants from a given species (d_{ws}^2 , or WS) as measures of total variability within that species; (a) d_{be}^2 (named BE) estimated by the different $d_{ws}^2 - d_{ws}^2$; (b) mean of the chi-square distances between plants of a given entry, for a given species, by its overall mean (d_{max}^2) or WE).

Results

The observed frequencies for each isozyme or enzymatic pattern per *Lycopersicon* species and the allozyme relative mobilities (Rm) can be obtained from the authors. Pictures of all of the enzymatic systems assayed are found in Fig. 1, and a schematic representation of EST, MDH and SOD patterns are shown in Figs. 2, 3 and 4, respectively.

The comparative amount of genetic variability per species is presented in Table 2, and its distribution into within- and between-line components is graphically represented in Fig. 5. The species with the greatest genetic variability are *L. chilense, L. peruvianum* and *L. pennellii,* mainly due to the within-line component. The lowest genetic variability was shown by *L. esculentum* and *L. cheesmanii,* where the contribution of both components was similar. No heterozygote at any of the 13 enzymatic loci was found in *L. parviflorum, L. pimpinellifolium* or *L. esculentum* ssp. *cerasiforme*, while the highest percentages of heterozygotes were found in *L. peruvianum, L. chilense* and *L. pennelli.*

The distribution of percentage of heterozygotes

Fig. 1. Profiles of the enzymatic systems

and coefficients of variation of the allozyme frequencies per locus and species is shown in Table 3. The higher the number of alleles at a locus, the higher the percentage of heterozygotes and coefficient of variation of its mean genic frequency. The ranking of species with

respect to both parameters is the same for the first four species with higher values *(L. chilense, L. peruvianum, L. pennelli* and *L. chmielewskii),* but not for the others. This is also true for the number of polymorphic loci or mean number of alleles per locus. There are

Fig. 2. Combinations of esterase patterns from zones 1 and 2

Table 2. Components of variability per species using different criteria (see Materials and methods for definition of abbreviations)

	che	chi	chm	pen	per	par	pim	cer	esc
WS	1.43	22.18	2.31	12.83	20.14	4.43	1.09	1.14	0.97
WE	0.92	19.81	0	8.84	18.14	1.01	0.07	0.75	0.54
BE	0.51	3.37	2.31	3.99	2	3.42	1.02	0.39	0.43
P	4	13		9	15	4			Q
A	0.27	0.71	0.30	0.44	0.73	0.25	0.24	0.23	0.37
CV	20.15	102.34	41.47	57.29	104.01	40.96	28.11	0	20.56
H	0.81	20.19	4.10	13.85	23.08	0	0	0	1.34

Table 3. Percentage of heterocygotes *(bold upper values)* and coefficient of variation *(lower values)* per species and locus (loci with null alleles are not included)

	$Tpi-1$	$Tpi-2$					Got-1 Got-2 Got-3 Pgi-2 Pgm-2 Aco-1 Aco-2 6Pg-1 6Pg-2 6Pg-3						$Gdh-1$
che ^a						5	5						
						39	148						75
chi	12 40	44 65		131	$\qquad \qquad$	25 77	37 151	44 243	62 216	6 165	6 95	19 117	6 30
chm				7 35	$\overline{}$ 50	—	7 50	\sim 117	7 53	176			33 58
pen		33 58	$\overline{}$ $\overline{}$	7 35	$\overline{}$ —	20 108	$\overline{}$ $\overline{}$	47 180	÷. $\qquad \qquad$	27 122	33 62	13 180	
per	82	-	4 59	26 116	9 28	30 85	43 172	57 224	39 189	30 161	43 69	9 107	9 60
par			$\overline{}$	75			185	169					103
pim								188			39	139	
cer													
esc		3 21				90	4 41	4 48			4 20	-	48
NA^b	3	3	$\boldsymbol{2}$	4	$\overline{2}$	6	8	11	9	$\overline{7}$	$\overline{2}$	6	3

 $-=0$

^a For definition of symbols, see Table 1

^o NA, Number of alleles that were distinguished

Fig. 4. Superoxide dismutase patterns

Fig. 5. Total variability (WS), within-entry (WE) and betweenentry (BE) variability components. The *X axis* represents the ranking of species (see Table 1 for symbols) by the coefficient of variation of the mean gene frequency (CV)

Fig. 6. Graphic representation of species on the plane defined by the first two factors of the correspondence analysis *(S. rickii* included). See Table 1 for definition of symbols

only three species where the locus showing the highest value of percentage of heterozygotes also shows the highest value of the coefficient of variation; these are: *L. peruvianum, L. penellii* and *L. cheesmanii.*

When *S. rickii* is considered, the first three factors generated by the correspondence analysis absorbed 27.25%, 21.45% and 14.65% of the total variance, respectively. The variables contributing most to the first axis (mainly related to *L. esculentum)* is *Got-2.3;* to the second axis (mainly related to *L. parviflorum, L. chmielewskii* and *L. chilense)* are *Me-2.2* and *Me-2.3.* The graphic representation of species on the plane defined by the first two factors is shown in Fig. 6; the dendrogram of the species is shown in Fig. 7. When *S. rickii* is removed, the correspondence analysis results in minor changes, and the clustering remains the same.

From all of the *Lycopersicon* species considered in this study *L. pimpinellifolium* is the most closely related to *L. esculentum,* followed by *L. cheesmanii.*

Discussion

In spite of the great importance of intraspecific genetic variability studies in the interpretation or discussion of phylogenetic relationships (Lewontin 1985; Templeton et al. 1981), very few authors have studied intraspecific variability in crops. In their studies on *Lycopersicon* C.M. Rick and his associates have considered it, but they have not used it to build up the corresponding dendrogram of species. Disregarding intraspecific variability, Hanson and McLean (1987) obtained a phenogram involving most *Lycopersicon* species in which several unexpected similarities were found; this phenogram is in disagreement with previous data reported by Rick (1979) and Palmer and Zamir (1982) and with our results.

Intraspecificic variability

The nine known species of the genus *Lycopersicon* can be classified in four groups according to their mating systems (Rick 1987). As shown in Table 1, the mating system ranges from autogamy to self-incompatibility. Our results on the distribution of genetic variability (Table 2) indicate that the three species with selfincompatibility systems show the highest values of total variability and are also those species in which the within-component is clearly greater than the between-component *(L. chilense, L. peruvianum* and *L. pennelli).* As Levin (1975) concluded, outbreeding annual species are expected to be more heterozygous and polymorphic per population than their inbreeding counterparts.

When the measures of total intraspecific variability we have used are compared (Table 2), all of them except for WS, are somehow affected by the number of lines screened (especially the number of polymorphic

loci, P). If the lines studied per species are considered to be representative of the species, the measure H indicates *L. esculentum* ssp. *cerasiforme, L. parviflorum* and *L. pimpinellifolium* as the species with the lowest index of allogamy because the within-component of variability is due to the presence of several homozygotes in the population and no heterozygote has been detected. Hence, the relatively high level of variability found in *L. parviflorum* and *L. pimpinellifollium,* two inbreeders, can be explained, following Nevo (1978), by their condition of being largely generalists, given their wide geographic distribution (Rick 1983).

In order to compare the measures of variability, CV and H, Table 3 shows the separate contribution of 13 enzymatic loci. For the species with the greatest genetic variability those loci with the greater number of alleles, *Aco-1* and *Aco-2,* also resulted in higher values of those indexes. As Gottlieb (1981) suggested, monomers might show more variability than oligomers. However, *Aco-2* is variable only in *L. chilense, L. chmielewskii* and *L. peruvianum,* while *Gdh-1* (three alleles and a likely oligomer) is variable in most *Lycopersicon* species.

In general, our results concerning the study of genetic variability mostly agree with those reviewed by Rick (1983). For tomato breeding purposes and taking into account the ecogeographic data, our results indicate that *L. parviflorum, L. chmielewskii* and *L. pimpinellifolium* would be very good resources to start with initial screenings, given that the mating system of the three most variable species may make traditional methods of plant breeding difficult.

Genetic relationships among species

The genetic richness of *L. peruvianum* and *L. chilense* makes their characterization a frequency matter because they show almost any allele. Therefore, methods to quantify relatedness as the one used by Hanson and McClean (1987) to construct a tomato phenogram are not adequate because intraspecific variability exists and variable species are not well characterized by their pool of genes but by the distribution of their gene frequencies. A coefficient of similarity will never unveil these kind of differences. We do consider that *S. rickii* is not well represented in our study, but whether it is included or not does not affect the relationships among the other species. Its distant position from all *Lycopersicon* species supports its inclusion in another genus.

Cluster analysis resulted in three main groups: one formed by the cultivated species, *L. pimpinellifolium, L. cheesmanii* and *L. peruvianum;* a second formed by two species with self-incompatibility systems, *L. pennellii* and *L. chilense;* and a farther group formed by two autogamous species, *L. parviflorum* and *L. chmielewskii.* This grouping agrees with extensive data reviewed by Rick (1983), but disagrees with those reported by Hanson and McClean (1987). We think that this disagreement is mainly due to two reasons: (1) species that are very variable are not well represented in the study of Hanson and McClean (1987), and (2) the way interspecific relationships (a coefficient of similarity) are measured in the study of these authors. Several facts regarding these groups deserve further discussion for future research.

Rick (1983) reported *L. chmielewskii* and *L. parviflorum* to be sibling species. These autogamous species of intermediate genetic variability group together in the analysis because they characteristically share several genes. The only other species that has shown most of these common genes is *L. peruvianum.*

L. chilense is highly polymorphic, and some alleles are only present in this species. This clarifies the fact of it being the most variable regarding the WS measure of total variability but the second for other measures such as P or CV. This also makes it very difficult to aggregate *L. chilense* to any other species. Its closeness to *L. pennetli* is mainly explained by the high frequency found for one common allele. The high frequency of two other genes makes these species related to each other and also to *L. peruvianum, L. chmielewskii* and *L. parviflorum.* In agreement to what Palmer and Zamir reported (1982), our results show that *L. pennelli* belongs with no doubt to the genus *Lycopersicon.*

The main cultivar responsible for the variability of *L. esculentum* is the Egyptian cv 'Edkawi'. This cultivar is one of the most salt tolerant within the cultivated species and, noteworthy, it shows some characteristic *L. pennellii* isozymes, such as TPI-2.1, GOT-2.1, PGM-2.2 (also characteristic of *L. cheesmanii),* 6PG-2.1 and the pattern K of MDH. There must be an introgresive origin involved in this polymorphic Egyptian cultivar. Besides this, the cultivated species is practically monomorphic. Its closest related species is *L. pimpinellifolium.*

The wild *Lycopersicon* species closest to *L. cheesmanii* is *L. pimpinellifolium* which is in agreement with the conclusions of Rick (1983), Palmer and Zamir (1982) and Hanson and McClean (1987). However, the rare allele *Pgm-2.2* fixed in the *L. pennellii* entries we have screened is also the one most commonly found in *L. cheesmanii.* This fact and the hydric stress habitats of these two species allows us to either hypothesize a specific role of *L. pennellii* in the origin of *L. cheesmanii* through *L. pimpinellifolium* or conclude that we are dealing with an example of convergent evolution.

The wide distribution of *L. pimpinellifolium* (Rick 1983), its variability and its close situation to the cultigen makes this species a good candidate to be the progenitor of *L. esculentum.* This fact along with its considerable variability makes this species a good choice for germ plasm screening when looking for desirable characteristics to be introduced into the cultivated species.

Surprisingly, *L. peruvianum,* one of the two most variable species included in this study and one with a completely different mating system, was included in the cultivated species group. The explanation for this is that the isozymes shown by *L. esculentum* are, in general, the ones most frequently found in *L. peruvianum.* Previous data involving mtDNA (Palmer and Zamir 1982) consider each entry of this species separately in order to build up the dendrogram instead of characterizing the species as a whole set of entries. We have chosen this second approach because we think it is closer to reality. This could be the main source of disagreements between both papers.

As mentioned above, *L. peruvianum* is also related to species of the other two groups; in fact, correspondence analysis places this species near the middle point of the graphic representation of species and variables (Fig. 6). We think that *L. peruvianum* is a good candidate to be the ancestor, or its nearest relative, of the *Lycopersicon* species. Evolution of the mating system and adaptation to specific habits must have played major roles in the speciation processes within the *Lycopersicon* genus.

Acknowledgements. We wish to thank Drs. C. M. Rick and F. Nuez for providing the plant material, and Ms. A. Sanchis and Mr. F. Bimbo for technical assistance. This work has been partially funded by grants from INIA (MJA) and Conselleria de Cultura, Educaci6 i Ciencia (MPB).

References

Anonymous (1991) European workshop on biotechnology for improvements in the assessment, conservation and utilisation of biological diversity. Commission of the European Community, Brussels

- Asins MJ, Carbonell EA (1986) A comparative study on variability and phylogeny of Triticum species. Theor Appl Genet 72:551-558
- Asins MJ, Carbonell EA (1987) Concepts involved in measuring genetic variability and its importance in conservation of plant genetic resources. Evol Trends Plants 1:51-62
- Asins MJ, Carbonell EA (1989) Distribution of genetic variability in a durum wheat world collection. Theor Appl Genet 77:287-294
- Baum JA, Scandalios JG (1979) Developmental expression and intracellular localization of superoxide dismutases in maize. Differentiation 13:133-140
- Benzecri JP (1970) Distance distributionelle et metrique de chi-deux en analyse factorielle des correspondences. Laboratoire de Statistique Mathematique, Paris
- Benzecri JP, Benzecri F (1980) Practique de l'analyse des données. 1. Analyse des correspondences. Exposé elementaire. Dunod, Paris
- Doebley J (1989) Isozymic evidence and the evolution of crop pants. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Hong Kong, pp 165-191
- Goodman MM, Stuber CW, Newton KJ (1982) Isozyme loci in maize. In: Sheridan WF (ed) Maize for biological research. Plant Mol Biol Assoc, Charlottesville, N.C., pp 53-60
- Gottlieb LD (1981) Electrophoretic evidence and plant populations. Prog Phytochem 7:1-46
- Greenacre \overline{MJ} (1984) Theory and application of correspondence analysis. Academic Press, London
- Hanson MR, McClean PM (1987) Tomato plastid and mitochondrial genomes: change and conservation. In: Nevins DJ, Jones RA (eds) Tomato biotechnology. Alan R. Liss, New York, pp 215-228
- Harlan JR (1976) Genetic resources in wild relatives of crops. Crop Sci 16:329-333
- Lance GN, Williams WT (1967) Mixed-data classificatory programs. 1. Agglomerative system. Aust Comput J 1 : 15-20
- Lefort-Buson M, Hebert Y, Damerval C (1988) Les outils

d'evaluation de la diversite genetique et phenotypique. Agronomie 8 : 173-178

- Levin DA (1975) Genetic correlates of translocation heterozygosity in plants. BioScience 25:724-728
- Lewontin RC (1985) Population genetics. Annu Rev Genet 19:81-102
- Nevo E (1978) Genetic variation in natural populations: patterns and theory. Theor Popul Biol $13:121-177$
- Palmer JD, Zamir D (1982) Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon.* Proc Natl Acad Sci USA 79:5006-6010
- Rick CM (1979) Tomato. In: Hawkes JG, Lester RN, Skelding AD (eds) The biology and taxonomy of Solanaceae. Academic Press, London, pp 667-677
- Rick CM (1983) Tomato *(Lycopersicon).* In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Elsevier Sci Publ, Amsterdam, pp 147-165
- Rick CM (1987) Genetic resources in *Lycopersicon.* In: Nevins DJ, Jones RA (ed) Tomato biotechnology. Alan R. Liss, New York, pp 17-26
- Rick CM, Fobes JF (1975) Allozymes of Galápagos tomatoes: polymorphism, geographic distribution, and affinities. Evolution 29: 443-457
- Shaw CR, Prasad R (1970) Starch gel electrophoresis of enzymes. A copilation of recipies. Biochem Genet 4:297-320
- Tanksley SD (1983) Enzyme-coding genes in tomato *(Lycopersicon esculentum).* Isozyme Bull 18:43-45
- Tanksley SD, Rick CM (1980) Isozymic gene linkage map of the tomato: applications in genetics and breeding. Theor Appl Genet 57:161-170
- Templeton AR, Salle R, Walbot V (1981) Speciation and inferences on rates of molecular evolution from genetic distances. Heredity 47:439-442
- Vallejos CE (1983) Enzyme activity staining. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetic and breeding. Elsevier, Amsterdam, pp 469-516
- Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Hong Kong, pp 5-45