Institut für Nutzpflanzenwissenschaften und Ressourcenschutz
Professur für Speziellen Pflanzenbau und Pflanzenzüchtung
PD Dr. K. Pillen

Comparative genetic linkage map for *Solanum ochranthum* and *S. juglandifolium* and genetic diversity and population structure in *S. lycopersicoides* and *S. sitiens*

Inaugural-Dissertation
zur
Erlangung des Grades
Doktor der Agrarwissenschaften
(Dr. agr.)

der
Hohen Landwirtschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt am 14.02.2008

von
Dipl.-Ing. agr. Elena Albrecht
aus Detmold

(D98)

Erster Berichterstatter: **PD Dr. Klaus Pillen**

Zweiter Berichterstatter: **Prof. Dr. Georg Noga**

Tag der mündlichen Prüfung: **16.05.2008**

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Erscheinungsjahr: **2008**

Abstrakt

Die vier Nachtschattengewächse *Solanum ochranthum*, *S. juglandifolium*, *S. lycopersicoides* und *S. sitiens* bilden die untersten Ränge im Stammbaum der Tomatengewächse und gelten als Verbindungsglied zwischen Kartoffel (*S. tuberosum*) und Tomate (*S. lycopersicum*). Alle vier besitzen morphologische Charakteristika, die typisch für Kartoffel sind, und sind von der Tomate durch Kreuzungsbarrieren getrennt. Während *S. ochranthum* und *S. juglandifolium* in den tropisch feuchten Regionen Kolumbiens, Ecuadors und Perus vorkommen, sind *S. lycopersicoides* and *S. sitiens* ausschließlich in trockenen, eng begrenzten Gebieten im Süden Perus und Norden Chiles zu finden.

In den Genomen von *S. lycopersicoides* und *S. sitiens* fehlt eine der parazentrischen Inversionen, die Tomate von der Kartoffel differenzieren. Um die Genomstruktur in *S. ochranthum* und *S. juglandifolium* zu untersuchen, wurde an 66 F_2 Pflanzen einer interspezifischen Kreuzung eine Gen-Kartierungsanalyse durchgeführt. Insgesamt wurden 132 molekulare Marker (96 CAPS, 19 RFLPs und 17 Mikrosatelliten) eingesetzt. Abweichungen von dem erwarteten Aufspaltungsmuster wurden bei einem Drittel der molekularen Marker beobachtet. Dreizehn Loci, die die Aufspaltung möglicherweise beeinflussen, wurden auf neun der zwölf Chromosomen identifiziert. Die Länge der Genkarte umfasste 790 cM, was einer 42 %-igen Reduktion der Rekombinationsereignisse im Vergleich zur Referenz-Genkarte für Tomate entsprach. Wie erwartet war der Grad der Kollinearität mit Tomate hoch. Die Analyse deutete auf eine reziproke Translokation zwischen den Armen zweier Chromosomen (8 und 12) in einer der beiden Eltern-Spezies hin.

Das Ausmaß der genetischen Diversität sowie deren strukturelle Verteilung wurden in 14 *S. lycopersicoides* und sieben *S. sitiens* Populationen anhand von insgesamt elf Isoenzymen und 15 Mikrosatelliten bestimmt. Die genetische Diversität war geringer auf dem Protein-Level verglichen mit dem DNA-Level. Auf Basis der Mikrosatelliten erwiesen sich *S. lycopersicoides* Populationen diverser als die der Schwester-Spezies *S. sitiens*, die Isoenzyme-Analyse jedoch zeigte das Gegenteil. Insgesamt war der Grad der Übereinstimmung zwischen den beiden Markersystemen in *S. lycopersicoides* recht gering. Die Analyse zeigte einen hohen Grad an Fragmentierung in *S. sitiens* Populationen sowie Anzeichen, dass sich in ihnen in jüngerer Vergangenheit genetische Flaschenhals Ereignisse zugetragen hatten. Strukturen der Isolierung durch Distanz waren in beiden Spezies deutlich, und eine Reihe von Allelen sowie Diversitäts-Parameter zeigten Korrelationen mit geographischen Eigenschaften (sog. Clines), vor allem mit dem Breitengrad. In den Stammbaumanalysen beider Spezies wurden jeweils drei primäre Populations-Gruppen deutlich; eine nördliche, eine zentrale und eine südliche.

Abstract

The four nightshades *Solanum ochranthum*, *S. juglandifolium*, *S. lycopersicoides* and *S. sitiens* compose the basal ranks of the tomato clade (*S.* sect. *lycopersicum* and *S.* sect. *juglandifolium*), representing a link between cultivated potato (*S. tuberosum*) and cultivated tomato (*S. lycopersicum*). All four exhibit potato-like morphological features and are isolated by strong reproductive barriers from tomato. *S. ochranthum* and *S. juglandifolium* occupy wet, tropical regions in Colombia, Ecuador and Peru while *S. lycopersicoides* and *S. sitiens* are narrowly endemic in the arid south of Peru and Northern Chile.

The *S. lycopersicoides* and *S. sitiens* genome lacks one of the major whole-arm paracentric inversions that differentiate tomato from potato. To investigate the genomic structure in *S. ochranthum* and *S. juglandifolium*, a genetic linkage map was constructed from 66 F_2 plants of an interspecific mapping population using 96 CAPS, 19 RFLPs and 17 microsatellites. Segregation distortion affected one third of the markers, and 13 putative segregation distorter loci were identified on nine out of twelve chromosomes. Total map length spanned 790 cM, representing 42 % length reduction relative to the tomato reference map. As expected, the degree of collinearity with the tomato genome was high. Evidence was found for a reciprocal whole-arm translocation among the parental species involving chromosome 8 and 12.

Levels of genetic diversity and genetic structure were investigated for the two narrowly endemic tomatoes *S. lycopersicoides* and *S. sitiens*. Fourteen and seven populations, respectively, were analyzed with a total of 11 allozyme and 15 microsatellite markers. Less variability was detected at the protein compared to the DNA level. *S. lycopersicoides* appeared more diverse than *S. sitiens* from the microsatellite analysis, whereas the opposite picture was presented by the allozyme analysis. Congruence between the two marker systems was low in the former species. Populations of *S. sitiens* were characterized by severe population fragmentation and exhibited signs of recent bottleneck events. A pattern of isolation by distance was evident in both species, and several alleles and diversity estimates exhibited geographic clines, primarily across the latitudinal range. Phylogenetic analyses revealed three major population clusters for each species; a northern, a central and a southern.

TABLE OF CONTENTS

1. Introduction

The focus of the present dissertation is a group of four tomato-like nightshades *Solanum ochranthum* (Dunal), *S. juglandifolium* (Dunal), *S. lycopersicoides* (Dunal) and *S. sitiens* (IM Johnston).

1.1 Tomato and the *Solanaceae*: facts & numbers

The *Solanaceae* or nightshade family is highly diverse and one of the largest and economically most important angiosperm families (Knapp et al. 2004, http://www.nhm.ac.uk/solanaceaesource). The estimated 2300 species of 96 genera (D'Arcy 1991) are distributed worldwide except Antarctica. The center of diversity lies in central and South America. Secondary centers of diversity are found in Australia and Africa (http://www.nhm.ac.uk/solanaceaesource). The *Solanaceae* family ranks 3rd in economic importance among cultivated plants (http://www.sgn.cornell.edu). Plants are utilized in a broad spectrum of ways, mainly as food (potato, tomato, eggplant, pepper), ornamentals (petunia, floripondio, velvet tongue, butterfly flower) or as sources of drugs (deadly nightshade, jimson weed, tobacco, henbane). Some (potato, tomato, tobacco, petunia) serve as model systems for scientific research (Knapp et al. 2004; Labate et al. 2007; http://www.nhm.ac.uk/solanaceaesource). With ca. 1,100 to 1,500 species, the genus *Solanum* is the largest genus within the *Solanaceae* family (D'Arcy 1991; Knapp et al. 2004) and one of the largest among flowering plants (Knapp et al. 2004; http://www.nhm.ac.uk/solanaceaesource). One of its subgenera is *Potatoe* which contains the crop species potato (*S. tuberosum* L.), tomato (*S. lycopersicum* L. = *Lycopersicon esculentum* Mill.) and pepino (*S. muricatum*; D'Arcy 1972; Spooner et. al 1993; Peralta and Spooner 2005). *Solanum* species display an astonishing diversity in terms of morphology, life forms and environmental adaptations. Habitats range from sea level to over 4,500 m, and from arid deserts to wet tropical forests. The majority of species occurs in the Andean region of western South America, secondary centers of diversity include western North America, Mesoamerica, eastern Brazil, the West Indies, Africa including Madagascar and Australia (Child 1990; http://www.nhm.ac.uk/solanaceaesource).

1.2 Tomato utilization

The name 'Tomato' derived from 'tomatl', a word in the Aztec language which was used to refer to several different plants with fleshy fruit (Gould 1983). Miller (1754) established the botanical name '*Lycopersicum esculentum*', a compound term of the Greek word 'lykos' (= wolf) and the Latin words 'perscium' (= peach) and 'esculentum' (= edible), hence literally meaning

'edible wolf-peach'. The name has its roots in a German legend, according to which the fruits were used by witches to attract werewolves (Hammerschmidt and Franklin 2005). After the first introduction to Europe tomatoes were believed to be poisonous, and Miller intended to diffuse that notion by stressing the edibility in the name (Miller 1754). Among the solanaceous vegetable crops tomato is the second most important. It ranks third in world production after the staple potato and the cucurbits, amounting to over 120 Mt in 2005 (FAO 2005). Main producers are China, the United States, Turkey, India, Egypt and Italy. The annual per capita consumption of tomatoes increased over the past decades, and lies at an estimated 12.1 kg/cap/year (FAO 2002). During the period from the early 1990s to the early 2000s the consumption of fresh market tomatoes in the US increased 15 % to almost 8 kg per person while that of processed tomatoes declined 9 % to 31 kg (fresh weight; USDA-ERS 2006). Tomato is a beneficial dietary source of vitamin A and C, and other antioxidants. Its glycoalcaloid tomatin lowers the contents of LDL cholesterol in the blood (reviewed in Labate et al. 2007).

Tomato has been used as a model system to study a vast array of research topics, including but not limited to the evolution of plant breeding systems (Rick 1979), the genetic architecture of plant morphological aspects such as fruit and seed shape (Doganlar et al. 2000; Frary et al. 2000; Van der Knaap and Tanksley 2003), the genetic characterization of fruit ripening (Vrebalov et al. 2002), abiotic stress (Maskin et al. 2004) and pathogen responses (Xiao et al. 2001). Tomato possesses a series of features that make it particularly attractive as a model species: it is a diploid with a small- to medium-sized genome (950 Mbp; Arumuganathan and Earle 1991), has a comparatively short life cycle, and genetic transformants can be obtained with relative ease. Genetically better characterized than most other crop species, many genetic tools are available, including germplams collections, various genetic stocks, a dense linkage map and genomic and cDNA libraries. Finally, genetic information can often be cross-utilized in related solanaceous species. (http://www.sgn.cornell.edu; http://tgrc.ucdavis.edu).

Currently ten countries collaborate under the umbrella of the International Tomato Sequencing Project in an effort to sequence and annotate the tomato genome (initially only the euchromatin fraction) and to create a bioinformatics platform for interconnected *Solanaceae* species (Mueller et al. 2005; http://www.sgn.cornell.edu).

1.3 Taxonomy

Early taxonomic treatments were solely based on morphology while today phylogenies based on molecular data are consulted to deduce relationships among taxa.

No.	Name[1,2]	Section[3]	Subsection[3]	*Lycopersicon* equivalent	Breeding system[4,5,6,7]	Fruit color[7]	Crossability with cultivated tomato[8,9,10]	Mapping analyses
1	*Solanum juglandifolium* Dunal	*Juglandifolium*	-	*Lycopersicon juglandifolium* (Dunal) Shaw	allogamous	green	?	NA
2	*Solanum ochranthum* Dunal	*Juglandifolium*	-	*Lycopersicon ochranthum* (Dunal) Shaw	allogamous	green	?	NA
3	*Solanum sitiens* Johnston	*Lycopersicon*	*Lycopersicoides*	*Lycopersicon sitiens* (Johnst.) Shaw	allogamous	green	UI, EC	11
4	*Solanum lycopersicoides* Dunal	*Lycopersicon*	*Lycopersicoides*	*Lycopersicon lycopersicoides* (Dunal in DC) Child ex Shaw	allogamous	green-black	UI, EC	11,12,13
5	*Solanum pennellii* Correll	*Lycopersicon*	*Lycopersicon*	*Lycopersicon pennellii* (Correll) D'Arcy	allogamous/ fac. autog.	green	UI	14,15,16,17,18,19,20,21
6	*Solanum habrochaites* Knapp & Spooner	*Lycopersicon*	*Lycopersicon*	*Lycopersicon hirsutum* Dunal	allogamous/ fac. autog.	green	UI	22,23
7	*Solanum chilense* (Dunal) Reiche	*Lycopersicon*	*Lycopersicon*	*Lycopersicon chilense* Dunal	allogamous	green	UI, EC	NA
8	*Solanum peruvianum* L.	*Lycopersicon*	*Lycopersicon*	*Lycopersicon peruvianum* (L.) Miller	allogamous/ fac. autog.	green	UI, EC	NA
9	*Solanum huaylasense* Peralta	*Lycopersicon*	*Lycopersicon*	Part of *Lycopersicon peruvianum* (L.) Miller	allogamous	green	UI, EC	NA
10	*Solanum corneliomuelleri* Macbr.	*Lycopersicon*	*Lycopersicon*	*Lycopersicon peruvianum* f. *glandulosum* Mull.	allogamous	green	UI, EC	NA
11	*Solanum arcanum* Peralta	*Lycopersicon*	*Lycopersicon*	Part of *Lycopersicon peruvianum* (L.) Miller (incl. var.*humifusum* and Marathon races)	allogamous	green	UI, EC	24,25
12	*Solanum neorickii* (Rick, Kesicki, Fobes & Holle) Spooner, Anderson & Jansen	*Lycopersicon*	*Lycopersicon*	*Lycopersicon parviflorum* Rick, Kesicki, Fobes & Holle	autogamous	green	UI	26
13	*Solanum chmielewskii* (Rick, Kesicki, Fobes & Holle) Spooner, Anderson & Jansen	*Lycopersicon*	*Lycopersicon*	*Lycopersicon chmeilewskii* Rick, Kesicki, Fobes & Holle	fac. allogamous	green	UI	27
14	*Solanum cheesmaniae* (Riley) Fosberg	*Lycopersicon*	*Lycopersicon*	*Lycopersicon cheesmaniae* L. Riley	autogamous	yellow	BC	28,29
15	*Solanum galapagense* Darwin & Peralta	*Lycopersicon*	*Lycopersicon*	*Lycopersicon cheesmaniae* f. *minor* L. Riley	autogamous	orange	BC	NA
16	*Solanum pimpinellifolium* L.	*Lycopersicon*	*Lycopersicon*	*Lycopersicon pimpinellifolium* (L.) Miller	fac. allogamous	red	BC	30,31,32
17	*Solanum lycopersicum* L.	*Lycopersicon*	*Lycopersicon*	*Lycopersicon esculentum* Miller	autogamous	red	NA	33, the ref. above (exc. 11)

Table 1. Tomato species.

UI = unilateral incompatibility (cross succeeds only when cultivated tomato is used as female parent); EC = embryo culture required to obtained interspecific hybrid (and often first backcross); BC = bilateral compatibility (cross succeeds in either direction); ? = crossability unknown.

References: 1: Peralta et al. 2004; 2: Peralta et al. 2005; 3: Spooner et al. 2005; 4: Rick 1963; 5: Rick 1979; 6: Rick 1986c; 7: TGRC; 8: Rick 1979; 9: Pertuzé et al. 2003; 10: Canady et al. 2005; 11: Pertuzé et al. 2002; 12: Chetelat and Meglic 2000; 13: Chetelat et al. 2000; 14: Tanksley et al. 1982; 15: Bernatzky and Tanksley 1986; 16: Tanksley et al. 1992; 17: Broun and Tanksley 1996; 18: Pillen et al. 1996; 19: Haanstra et al. 1999; 20: Areshchenkowa and Ganal 1999; 21: Areshchenkowa and Ganal 2002; 22: Bernacchi and Tanksley 1997; 23: Zhang et al. 2002; 24: Van Ooijen et al. 1994; 25: Fulton et al. 1997; 26: Fulton et al. 2000; 27: Paterson et al. 1988; 28: Paterson et al. 1991; 29: Paran et al. 1995; 30: Grandillo and Tanksley 1996b; 31: Tanksley et al. 1996; 32: Doganlar et al. 2002c; 33: Saliba-Colombani et al. 2000.

Taxonomic classification of tomato and its wild relatives (including the four tomato-like nightshade species) has been the subject of frequent controversy from the time they were first described by Linnaeus in 1753. At the center of the dispute stood the question whether Lycopersicon represented an independent genus separate from genus *Solanum*. This was not an isolated case in the *Solanaceae* family: taxanomic circumscriptions have been undergoing constant change. For example, in 1979 estimates arrived at 83 genera and 2671 species (D'Arcy 1979) for the family and a decade later those numbers had changed to 96 genera and 2300 species (D'Arcy 1991). Linnaeus treated tomatoes within the genus *Solanum* due to their five-parted flowers. Only a year later Philip Miller (1754) removed the group from genus *Solanum* to form the independent genus *Lycopersicon* on the basis of unique features (i.e. the fused anthers and sterile anther tips). The dispute over the generic circumscription continued over the following one and a half centuries. Those in favor of an independent genus *Lycopersicon* were motivated by mostly practical, traditional but also biological arguments. Some referred to the validity of the first official recognition of *Lycopersicon esculentum* as taxon name at the Fourteenth International Botanical Congress in Berlin 1879 (Terrell 1983; D'Arcy 1991). Lester (1991) e.g. pointed out the practical value of the name but admitted that the classification was "artificial" from a phylogenetic viewpoint. Similarly, Symon (1981) advocated a "utilitarian component in nomenclature", arguing that based on the importance of the crop the group should be classified in a distinct genus. Rick (1979) emphasized biological aspects, i.e. the reproductive isolation of *Lycopersicon* from *Solanum*.

The opponents on the other hand argued that, for phylogenetic consistency, if *Lycopersicon* was granted genus-status, genus *Solanum* would either have to be accepted as paraphyletic or split into several monophyletic genera (reviewed in Spooner et al. 1993). Likewise, if taxonomic rankings

were established on crossing relationships, additional subdivisions within *Solanum* would need to ensue (Spooner et al. 1993).

Both *S. juglandifolium* and *S. ochranthum* were first described by Dunal (1816) and later by Bitter (1912). *S. lycopersicoides* was named by Dunal in 1852 in reference to its resemblance to tomato. Johnston was the first to describe *S. sitiens* in 1929. In 1961 Correll, believing that Rick had been the first to discover the species, classified it as *S. rickii*. The nomenclatural ambiguities were clarified by Marticorena and Quezada in 1977 (although it took several years before knowledge about their publication had spread through the scientific community), and the earlier name is now recognized as having precedence. Series *Juglandifolia* was described within genus *Solanum* sect. *Tuberarium* (Dun.) Bitt., subsect. *Hyperbasarthrum* Bitt. by Rydberg in 1924.

As the genus *Lycopersicon* was defined primarily by the presence of yellow flowers and sterile anther tips, Correll (1958) classified *S. pennellii*, which lacks the sterile appendages, within the series *Juglandifolia* (Correll 1958). However, pedicel articulation in *S. pennellii* is close to the base, hence more tomato-like, whereas pedicels of the four tomato-like nightshades are articulated below the calyx (Correll 1958). This circumstance motivated Correll to remove *S. pennellii* from the group only a few years later and treat it as the sole representative of the newly created section *Neolycopersicon* in genus *Lycopersicon*, i.e. as *L. pennellii* Correll (Correll 1962). The species' closer association with *Lycopersicon* rather than with series *Juglandifolia* was also supported by crossability tests and other evidence (Rick 1979). Correll (1962) first described all four, *S. juglandifolium*, *S. ochranthum*, *S. lycopersicoides* and *S. rickii*, together in series *Juglandifolia* Rydb. and listed as common features of this series: woody, non-tuber bearing plants with yellow corollas and pedicels articulated well above the base (Correll 1958, 1962). He also remarked the morphological hiatus among the two groups which were grouped together in one series and disagreed with the treatment under subsect. *Hyperbasarthrum* because of the morphological resemblance of this group to *Lycopersicon*. As an alternative he proposed the inclusion of the series as a distinct subsection in sect. *Tuberarium*. D'Arcy (1972) shared Correll's opinion and moved series *Juglandifolia* to *Solanum* sect. *Petota* Dumort subsect. *Potatoe* G Don. Child (1990) finally split the two groups and removed *S. lycopersicoides* and *S. sitiens* from series *Juglandifolia* (Rydb.) Hawkes to form the new subsection *Lycopersicoides* Child within section *Lycopersicum* (Mill.) Wettst. He treated *S. juglandifolium* and *S. ochranthum* under section *Juglandifolium* (Rydb.) Child (Child 1990) in *Solanum* subgenus *Potatoe* (*Tuberarium*; Dun.) Bitter.

More recent molecular phylogenies (Spooner et al. 1992, 2005) show the tomatoes deeply nested within genus *Solanum,* which led to the relegation of *Lycopersicon* to a sectional status and sparked a comprehensive revision and reclassification of the entire group. Supported by molecular data and

morphological characterizations the tomato clade is now treated within genus *Solanum* subgenus *Potatoe*. Thirteen species are segregated into two sections (*Lycopersicon* and *Juglandifolium*) and one subsection (*Lycopersicoides*; Spooner et al. 2005; Peralta et al. 2005). Former genus *Lycopersicon* is equivalent to section *Lycopersicon* subsection *Lycopersicon* with the exception that two taxa (*L. peruvianum* and *L. cheesmaniae*) were subdivided into four and two species, respectively (Peralta et al. 2005; Darwin et al. 2003).

1.4 Phylogenetic relationships

A large number of studies investigated phylogenetic relationships within *Solanum* subgenus *Potatoe*. The most comprehensive analyses were based on morphological aspects (Dunal 1852; Bitter 1912; Correll 1958, 1962; D'Arcy 1972, 1979; Child 1990; Hawkes 1990; Lester 1991) but amino acid sequences data (Boulter et al. 1979), serological studies (Lester 1991), occurrence of steroidal alkaloids (Tétényi 1987), crossing relationships (Rick 1979), somatic hybrid formation (Wann and Johnson 1963), distribution of gametophytic self-incompatibility (Whalen and Anderson 1981) comparative mapping analyses (Tanksley et al. 1992; Perez et al. 1999; Pertuzé et al. 2002; Doganlar et al. 2002a; Gebhardt et al. 2003), nucleic acid sequence data (Peralta and Spooner 2001) and nucleic acid restriction sites (Spooner et al. 1993) were also used to establish phylogenetic relationships.

The four tomato-like nightshades have long been recognized as linking the potatoes and the tomatoes (Correll 1962; Rick 1988; Child 1988). Analyses of antisera revealed that *Solanum* sect. *Juglandifolium* was closely associated with sect. *Petota* and sect. *Etuberosum* and that all three were only slightly distinct from sect. *Lycopersicon* (Boulter et al. 1979). Later, as molecular markers became available, phylogenetic evidence in support of this notion was obtained from different sources including analysis of cpDNA restriction sites (Spooner et al. 1993), GBSSI sequence data (Peralta and Spooner 2001), ITS sequence data (Marshall et al. 2001) and AFLP cladistic analysis (Spooner et al. 2005).

Of greater uncertainty are the relationships among the four species. Few phylogenetic studies contain sufficient species sampling to address this question. Most results (GBSSI sequence, morphology and AFPLs) support sect. *Juglandifolium* and subsect. *Lycopersicoides* as separate monophyletic groups (Peralta and Spooner 2001; Spooner et al. 2005); only an analysis based on cpDNA sequence data identified *S. sitiens* as sister to *S. ochranthum* and both as sister to *S. lycopersicoides* (Spooner et al. 1993). Whether sect. *Juglandifolium* or subsect. *Lycopersicoides* is the closest outgroup to tomato has been subject of much speculation. Prior to molecular evidence (i.e. based on morphology and crossing relationships) *S. lycopersicoides* and *S. sitiens* were widely

assumed to be the closest relatives to sect. *Lycopersicum*, separated by only a short divergence time (Rick 1979). To date only two informative molecular analyses are available that elucidate hierarchical structuring among the groups, one is based on GBSSI sequence data (Peralta and Spooner 2001) and the other on AFLPs (Spooner et. al. 2005). Both support *S. ochranthum* and *S. juglandifolium* as closest relatives to sect. *Lycopersicon*, subsect. *Lycopersicoides* (*S. sitiens* and *S. lycopersicoides*) basal to the two groups and sect. *Petota* (the potatoes) as sister to the entire group.

1.5 Genome characterization

The base chromosome number of x = 12 defines a monophyletic group within the *Solanaceae* that comprises the *Anthocercidae* tribe, genus *Nicotiana* and subfamily *Solanoideae*, which includes genus *Solanum* (Olmstead et al. 1999). The great majority of members in genus *Solanum*, subgenus *Potatoe* are diploid, only about one third of sect. *Petota* is polyploid (3x - 6x; Hawkes 1990; Child 1990). Tomatoes are diploid with the exception of some spontaneous tetraploids (Rick 1976). The genomes of the species within the tomato clade (sect. *Lycopersicon* subsect. *Lycopersicon*) are homologous and highly collinear (Rick 1979; Tanksley et al. 1992; Paran et al. 1995; Grandillo and Tanksley 1996; Bernacchi and Tanksley 1997; Fulton et al. 1997), differentiated only by small rearrangements (van der Knaap et al. 2004).

1.6 Genetic diversity

Each domestication event constitutes a bottleneck (i.e. a sudden reduction in population size; Tanksley and McCouch 1997; Lowe et al. 2004), therefore it is no surprise to find a dearth of genetic variation among many crop species, for example soybean, wheat, rice, cotton and tomato (Rick and Fobes 1975). Without genetic diversity, breeding efforts remain ineffective (Rick 1988; Zamir 2001) and susceptibility towards disease and insect epidemics is enhanced (Tanksley and McCouch 1997). As a consequence many plant breeders turned to crop related wild species for germplasm enhancement (Zamir 2001). A large number of solanaceaeous wild species have contributed to the improvement of their cultivated relatives in breeding programs (e.g. Rick and Chetelat 1995; http://www.nhm.ac.uk/solanaceaesource). The major limitations associated with the use of wild species are 1) linkage drag (i.e. the introduction of undesirable characteristics along with the target trait), 2) complex inheritance of many traits (QTLs), determined by multiple loci and epistatic effects, and 3) crossing barriers (Bonierbale et al. 1988; Zamir 2001).

A prominent domestication trait in tomato is the inserted stigma (i.e. recessed within the anther cone), which secures selfing and therefore fertilization in the absence of pollinating insects, e.g. under unfavorable environmental conditions. Stigmas are well exposed in the wild ancestors, but

style length was shortened progressively during the evolution of the crop. Thus, Latin American cultivars exhibit more exserted stigmas than European cultivars, which, in turn, have stigmas that are more exserted than those of modern Californian cultivars (Rick 1976).

The wild cherry tomato, *S. lycopersicum* var. *cerasiforme*, is thought to be the direct wild ancestor to tomato (Rick and Holle 1990). On the basis of very limited archeological and ethnobotanical evidence it is believed that the domestication occurred in Mesoamerica (Jenkins 1948). Genetic diversity is diminished among populations of *S. lycopersicum* var. *cerasiforme* in that region, presumably as a result of founder events that accompanied the migration from the center of origin, the north central Andes (Rick 1976; Rick and Holle 1990). Two major bottleneck events followed in the course of the cultigen's history: the first during the introduction to Europe, which started in the 16th century (Rick 1976; Labate et al. 2007), as evidenced by the early herbalist Matthiolus (1544), and the second as the vegetable was reintroduced to Amercia, this time North America, from Europe in the late 18th century (Rick 1986b). This, in combination with its inbreeding nature (cultivars were maintained as pure lines) led to little if any increase in average yields of varieties bred in the early 20th century (Rick 1976, 1988). The utilization of wild tomato relatives as a new source of variation, starting around 1940, marked a turning point in this trend (Rick 1988). Fusarium wilt resistance from *S. pimpinellifolium* was the first economic trait of exotic origin to be introgressed into an elite variety. It also represented the first high-level resistance in tomato known at that time (Bohn and Tucker 1940; Rick 1988). Since then many valuable sources for disease and insect resistances, fruit quality traits and abiotic stress tolerances have been found among wild species and bred into tomato cultivars (Rick 1988). The result was a four to five-fold yield increase over the subsequent decades. By 1995 resistances to at least 42 major diseases had been discovered among related wild species and half of those had been bred into horticultural tomatoes (Rick and Chetelat 1995). The use of exotic resources did not remain restricted to qualitative traits: With the development of advanced backcross QTL (AB-QTL) analysis (Tanksley and Nelson 1996) a new technique became available to a) efficiently screen exotic germplasm for desirable quantitative traits and b) to introgress the newly discovered alleles directly into elite varieties. In *S. habrochaites*, for example, valuable QTLs have been identified for yield, fruit flavor and chilling tolerance (Monforte and Tanksley 2000; Fulton et al. 2002b; Goodstal et al. 2005), in *S. pimpinellifolium* and *S. peruvianum* for fruit size and weight (Tanksley et al. 1996; Fulton et al. 1997), and in *S. pennellii* for yield and antioxidant content (Rousseaux et al. 2003; Frary et al. 2004).

The tomato clade is comprised of 13 species which are divided among two sections and two subsections within the genus *Solanum* (Spooner et al. 2005; Peralta et al. 2005; Table 1). According to the concept of Harlan and de Wet (1971) the crop's primary genepool (i.e. within which gene

transfer is easily achieved) comprises the nine species *S. lycopersicum*, *S. pimpinellifolium*, *S. cheesmanii*, *S. galapagense* (formerly *Lycopersicon cheesmanii* f. *minor*), *S. chmielewskii*, *S. neorickii* (formerly *L. parviflorum*), *S. habrochaites* (formerly *L. hirsutum*), *S. pennellii* and *S. arcanum* (formerly *L. peruviuanum* var. *humifusum*). The secondary genepool (i.e. within which gene transfer is more difficult) contains the four species *S. peruvianum*, *S. corneliomuelleri* (formerly *L. peruvianum* f. *glandulosum*), *S. huaylasense* (formerly part of *L. peruvianum*) and *S. chilense*. The two genepools correspond to the '*esculentum* complex' and the '*peruvianum* complex', respectively, of *Solanum* sect. *Lycopersicon* subsect. *Lycopersicon* (Rick 1976). The tertiary genepool (i.e. within which gene transfer is accomplished only with extreme difficulty) includes the two species within *Solanum* subsect. *Lycopersicoides* (i.e. *S. lycopersicoides* and *S. sitiens*; Rick 1976, 1988).

Wild tomato species are adapted to a wide range of habitats and, as a consequence, possess substantial allelic variability for agronomically important traits (Correll 1962; Rick 1976; Rick 1979). The most polymorphic among them is *S. peruvianum* with diversity estimates that surpass those of outcrossers such as maize and loblolly pine (Rick 1979; Städler et al. 2005). A single *S. peruvianum* accession typically harbors more genetic diversity than those of the self-compatible tomato species combined (Miller and Tanksley 1990). Thus, it is believed that several tomato species originated from a common ancestor to *S. peruvianum* only fairly recently via a combination of lineage sorting and postspeciation events (Baudry et al. 2001). Within each wild species, maximum levels of diversity are typically found at the center of the distribution, and decline towards the edges of the species range. Diversity measures are positively correlated with the degree of outcrossing, as expected (Rick et al. 1977, 1979).

Over the past century seed banks have been established for many crops in order to preserve genetic diversity and provide accessibility of genetic material, including wild relatives, for agricultural breeding programs (Tanksley and McCouch 1997). More than 75,000 tomato accessions are maintained in over 120 countries (reviewed by Robertson and Labate 2007). One of the most diverse germplasm collections for tomato is held at the CM Rick Tomato Genetics Resource Center (TGRC) at the University of California at Davis. About one third of the over 3,600 tomato accessions are collections of wild species, the rest are monogenic mutants or other types of genetic stocks, such as cultivars (modern, vintage and Latin American), prebred lines (introgression lines, backcross recombinant inbreds, alien substitution lines, monosomic alien addition lines), stress tolerant stocks, cytogenetic stocks (translocations, trisomics, autotetraploids), cytoplasmic variants, genetic markers, etc. The TGRC supplies more than 5,000 seed samples per year to researchers in over 34 countries, representing a utilization rate of nearly 150 %. TGRC stocks are heavily used for

research or breeding of disease resistances and fruit color (carotenoids, flavonoids) and, to a lesser extent for genetic, physiological, biosystematic or diversity studies (http://tgrc.ucdavis.edu; Chetelat 2006). Other larger tomato genebanks are held at the Plant Genetic Resource Unit (PGRU) of the United States Department of Agriculture, Agricultural Research Service (USDA-ARS) in Geneva, NY and the Asian Vegetable Research and Development Center (AVRDC) in Taiwan.

1.7 Habitat preferences

Wild tomatoes are originally native to western South America, from Ecuador to northern Chile, including the Galápagos Islands. Habitats range from coastal regions to over 3,600 m above sea level in the Andes (reviewed in Peralta and Spooner 2007). The wild cherry tomato, *S. lycopersicum* var. *cerasiforme* – presumed ancestor of cultivated *S. lycopersicum* – has spread into Mesoamerica, and is now widely distributed in tropical and subtropical regions globally (Rick and Holle 1990). Cherry tomatoes occur as weeds or volunteers in many regions, often as a consequence of garden escape (Rick 1976).

Within the vast genus *Solanum*, the most tomato-like taxa outside sect. *Lycopersicon* subsect. *Lycopersicon* are the four species *S. ochranthum*, *S. juglandifolium*, *S. lycopersicoides*, and *S. sitiens*. They form two groups of sibling species, *S. ochranthum* and *S. juglandifolium* in sect. *Juglandifolium* and *S. lycopersicoides* and *S. sitiens* in sect. *Lycopersicon* subsect. *Lycopersicoides*. The two groups occupy distinct regions and habitat types: *S. ochranthum* and *S. juglandifolium* occur exclusively at mid-elevations in rainforest regions from Colombia to southern Peru; *S. lycopersicoides* and *S. sitiens* are restricted to mid- to high elevations, extremely arid regions of southern Peru and northern Chile. *S. ochranthum* and *S. juglandifolium* are distributed over a narrow, elongated range in the tropical, northwestern part of the Andes (Rick 1988). The two species are largely sympatric, frequently found within only a few kilometers from each other in Ecuador and Colombia. Collections of both were made at, e.g. the Ecuadorian/Colombian border in Maldonado, Carchi, Ecuador. Levels of diversity have never been assessed for these two species. The highest number of population occurrences are found in Ecuador (Smith and Peralta 2002), with an additional center of diversity for *S. ochranthum* located in the region around Cusco, Peru. The distribution of *S. juglandifolium* is continuous, it occurs on both sides of the Andean cordillera in Ecuador and on all three major mountain ranges in Colombia: the coastal Cordillera Occidental, the Cordillera Central and the Cordillera Oriental. The distribution range of *S. ochranthum* also follows the Andean cordillera but covers a greater latitudinal range reaching from Colombia far south down to Peru's Cusco region. Populations are fragmented and further scattered out than those of *S. juglandifolium*, with large gaps in central Peru and central Ecuador and two clusters at the

Ecuadorian/Colombian and Ecuadorian/Peruvian frontiers. There are no obvious environmental factors that explain the distribution patterns, which may therefore also reflect differences in collection efforts, accessibility by roads, herbivory pressures or other factors (Smith and Peralta 2002; http://tgrc.ucdavis.edu). Habitat preferences of *S. ochranthum* and *S. juglandifolium* appear to be similar: both occur in regions with frequent and high rainfall, in swampy, poorly drained areas, and at riparian sites (Rick 1988). Plants may be found among a diverse array of vegetation types, ranging from thickets of primary cloud forests to the clearings at road or forest margins, where they often climb trees (Correll 1962). The only difference between the two may be the higher temperature requirements of *S. juglandifolium* (Rick 1988), which grows at slightly lower elevations than *S. ochranthum* (1200 - 3048 m versus 1400 - 3658 m, respectively)[1] as well as at a more equatorial latitudes (~ 7.5˚ N – ~ 4.5˚ S versus ~ 6˚ N – ~ 14˚ S; Correll 1962; Smith and Peralta 2002). Their habitat ranges overlap with those of *S. habrochaites* and *S. neorickii* in southern Ecuador and northern Peru (http://tgrc.ucdavis.edu).

S. lycopersicoides and *S. sitiens* on the other hand are allopatric, separated by a distance of ~ 300 km, and confined to relatively small areas in southern Peru and northern Chile (Rick 1988; Smith and Peralta 2002). *S. lycopersicoides* prefers high elevation, more mesic sites, and *S. sitiens* lower elevations and extremely dry sites (Correll 1958; Rick 1988; Smith and Peralta 2002). *S. lycopersicoides* occurs in a narrow region along the western part of the main Andean cordillera in southern Peru (province of Tacna) and northern Chile (Region I, Tarapacá), which is located between ~ 17˚ and 19˚ S latitude (i.e. spanning a north/south line of ca. 250 km). Plants typically grow on south-facing slopes, along drainages, among boulders and in quebradas (Correll 1962). *S. lycopersicoides* is distributed at higher elevations than any other tomato species (up to 3,800 m)[1], where it is regularly exposed to frosts (Rick 1988). Two tomato species, *S. chilense* and *S. peruvianum* are sympatric with *S. lycopersicoides* at a few mid- to high elevation locations in Tarapacá, Chile (http://tgrc.ucdavis.edu).

Unfortunately populations of both species have been experiencing a rapid decline over the past decades and have recently been classified as ‘endangered and rare’ (*S. lycopersicoides*) and ‘vulnerable and rare’ (*S. sitiens*) by the Chilean government. Grazing animals like llamas, alpacas and goats are posing the greatest threat to the existence of *S. lycopericoides* while *S. sitiens* is primarily suffering from hostile climate conditions and habitat loss through mining activities (Chetelat and Pertuze 2006). *S. sitiens* grows under the most arid conditions reported for any tomato species (Rick 1988). It is native to the northern areas of the Atacama desert (~ 22˚ - 24˚ S) in Chile’s Region II (Antofogasta), one of the driest places on earth with annual rainfalls below 10 mm (Johnston 1929; Caviedes 1973; Alpers and Brimhall 1988; Smith and Peralta 2002). Its

distribution is restricted to a small area spanning about 230 km from north to south and 2500 – 4000[1] m in altitude, mostly on the slopes of the Cordillera de Domeyko, a minor cordillera between the coast ranges and the main Andean crest (Rick 1988; Smith and Peralta 2002). Scarcely vegetated quebradas, rocky ravines and dry mountains are among the preferred habitats of this species. A local name is 'uva minera' (= miner's grape), which alludes to the grape-like shape and color of its fruits, and tendency to be found growing near mine sites (Rick 1988). There are few perennial plants that can survive the hyperarid habitat of *S. sitiens*. Associated flora consists primarily of herbaceous xerophytes, i.e. *Calandrinia crassifolia*, *Adesmia atacamensis* and *Nolana* ssp. (Rick 1988; http://tgrc.ucdavis.edu). The only other tomato species in this region is *S. chilense*, which grows along the coast and in the Andes of Chile and Peru, but is never found alongside *S. sitiens*, presumably because that climate it is too dry.

1.8 Morphology

All members of the tomato clade (*Solanum* sect. *Lycopersicon*) are biennial or perennial herbs with pinnately segmented leaves, sympodia of two or three leaves, cymose inflorescences, yellow corollas, yellow anthers that are held together by interlocking marginal hairs to form a tube around the pistil, and berry-like fruits (Müller 1940; Rick 1979; Smith and Peralta 2002).

The four tomato-like nightshades share a number of morphological traits that place them in an intermediate position between tomato and potato. On the one hand they possess tomato-like characteristics that set them apart from most other *Solanum* species, such as the yellow corolla (flowers in subsect. *Petota* are mainly white, purple or blue), and they lack certain tomato traits, such as sterile anther appendages and longitudinal pollen dehiscence (Instead, pollen typically dehisce through terminal pores, spreading longitudinally along anthers.). Their anthers are free (i.e. unattached), floral bracts are missing, pedicels are articulated below the calyx, and finally strong reproductive barriers separate them from the tomato group (Correll 1958, 1962; Rick 1988; Child 1990; Stommel 2001; Smith and Peralta 2002). All four species are diploid (2n = 24), woody perennial shrubs or vines (Correll 1962; Rick 1988). Characteristic are highly compound inflorescences and the ubiquitous presence of glandular-pubescent hairs. Foliage and flowers of each species display distinctive fragrances (Correll 1962). The foliage odor of *S. sitiens* and *S. lycopersicoides* are easily distinguishable, and flowers are generally scented (nectar or honey-like fragrance), unlike the tomatoes which lack scent.

[1] Distributions were inferred from passport data of herbarium specimens in the ecogeographic survey by Smith and Peralta (2002) and the TGRC database (http://tgrc.ucdavis.edu).

Apart from the few shared features the two groups have little in common. *S. ochranthum* and *S. juglandifolium* resemble one another so closely that they may be confounded outside their natural environment (Appendix 1A, 1B). Both are woody, shrubby climbing vines that draw the observer's attention for their showy, bright yellow flowers. In fact, their utilization as ornamentals has been proposed. Further characteristics are walnut-like leaves (hence the name '*Juglandifolium*') with entire, elliptic to lanceolate leaflets (Correll 1962; Rick 1988).

S. ochranthum forms robust branches that reach 8 - 10 cm in diameter and up to 30 m in length. Leaves are odd-pinnately segmented and up to 32 cm long. In contrast to *S. juglandifolium* leaves bear pseudostipules. The leaflets, of which there are typically 9 - 13, are mostly sessile (or near sessile) and more closely spaced than in the sister species. Leaflets are smooth and velvet-like on the surface. The *ochranthum* inflorescence can measure 30 cm across and is composed of large flowers (up to 3.5 cm in diameter) on long peduncles. The corollas are more rotate than those of *S. juglandifolium*. The only plant parts that are diminished in size with respect to their *S. juglandifolium* counterpart are the pedicels, which rarely exceed the length of 1 cm. Fruits require an extraordinarily long ripening period (at least 8 - 9 months when grown at UC Davis). Fruit size (4 - 6 cm in diameter) far exceeds that of any other wild tomato. Pericarps are woody and thick. The interior is composed of 3 - 4 locules, in contrast to the wild tomato species, which bear almost exclusively bilocular fruit. Fruits emit an apple-like fragrance when ripe (Child 1990). *S. ochranthum* seeds are large and winged (Correll 1962; Rick 1979; Rick 1988; Child 1990).

S. juglandifolium appears like a miniature version of *S. ochranthum*. Vines grow up to 5 m long. Leaves are scabrous and densely pilose on the reverse side. The corolla is stellate and up to 3.7 cm in diameter. Pedicels are up to 3 cm long. In Colombia common names are 'riñon' or 'riñonada' (Correll 1962, Rick 1988).

S. sitiens and *S. lycopersicoides* are also morphologically similar to one another, although to a lesser extent than *S. ochranthum* and *S. juglandifolium*. The plants are erect with highly pinnatifid leaves, stellate corollas of ca 2 cm in diameter, and white to cream-colored anthers.

S. lycopersicoides is a woody bush that grows up to 2.5 m in height (Appendix 1C). Leaves are up to 13 cm long and herbaceous-like thin. The main difference to its sister species are the fruits, which are smaller (6 mm) and turn black from anthocyanin accumulation towards maturity. A thick pericarp allows the interior to retain its moisture for a long period of time after ripening (Correll 1958, 1962; Rick 1988; Peralta and Spooner 2005). Correll (1962) noted the high level of morphological variability among populations.

S. sitiens is a herbaceous plant with a woody stem (Appendix 1D). It is smaller than *S. lycopersicoides* with a maximum height of ~ 50 cm. Leaves are fleshy-coriaceous, about 4 cm in

length and composed of small, almost needle-like leaflets. Fruits are patchy-green, globose, ~ 1 cm in diameter and insulated only by a thin pericarp, which desiccates on the vine, eventually attaining a papery eggshell texture (Correll 1962; Rick 1988; Peralta and Spooner 2005).

1.9 Useful traits

Considering their unique ecological adaptations, these four tomato-like nightshades are expected to harbor traits not found elsewhere among tomato relatives (Rick 1988). Furthermore, novel traits, often observed in crosses of diverse genetic backgrounds, are likely to arise from hybridizations with the cultigen (Rick 1967, 1982a, 1988, 1995).

Of the four species, only *S. lycopersicoides* has been subjected to anything near a thorough evaluation for beneficial traits. As a result, a number of fungal and viral resistances, insect and abiotic stress tolerances were identified. Known disease resistances include resistances to tomato mosaic virus and cucumber mosaic virus (Phills et al. 1977a), early blight (Bamberg et al. 1994), Fusarium wilt (*Fusarium oxysporum* f. sp. *lycopersici*; Phills et al. 1977b), Phytophtora root rot (*Phytophtora parasitica*) and grey mold (*Botrytis cinerea*; Gradziel and Robinson 1989; Chetelat et al. 1997; Guimarães et al. 2004; Zhao et al. 2005). In addition, foliage and fruit were found to be rarely attacked by parasitic Lepidoptera or leaf miners (*Liriomyza trifolii*; Rick 1988; Chetelat et al. 1997). A repellent or antibiotic effect may be conferred by the high levels of glycoalkaloids in the leaves (Oleszek et al. 1986). Tolerance of chilling temperatures has been confirmed in several reports (Robinson and Kowaleski 1974; Wolf et al. 1986; Zhao et al. 2005).

Little is known about potentially valuable traits in *S. sitiens* except that it can withstand extreme drought and to a degree also low temperatures, as has been attested by field and greenhouse observations (Rick 1986b; Rick and Chetelat 1995). Likewise, the mechanisms that convey cold or drought adaptations in either species are unknown. Both have the ability to send up knew shoots from the root system, which may allow them to quickly recover after a period of stress (e.g. drought, freezing or grazing). At the driest sites *S. sitiens* plants tend to be stunted and have darker, narrower leaves that are folded along the veins, presumably to reduce transpiration (Chetelat, pers. comm.).

S. ochranthum and *S. juglandifolium* are expected to display traits associated with wet conditions, such as tolerance of waterlogged soils and resistances to moisture-related fungal or bacterial diseases of both the root system and upper plant parts (Rick 1988, 1990). In addition they have been reported to be remarkably free of virus symptoms (Rick 1988). In *S. ochranthum* resistance to late blight (*Phytophtora infestans*; Kobayashi et al. 1994) and root knot nematodes (*Meloidogyne* spp.; Bamberg et al. 1994) has been observed. The type-B glandular trichomes in this species, which

cover the entire plant surface, have been shown to confer resistance to a number of small, "soft-bodied" insects in other *Solanum* species (Tingey et al. 1981), including aphids (*Myzus persicae*), larvae of the Colorado potato beetle (*Leptinotarsa decemlineata*), leaf miner (*Liriomyza trifolii*; Moretti 1990), potato leafhopper (*Empoaseae fabae*) and flea beetle (*Epitrix* ssp.; Bamberg et al. 1994). However, a general resistance against aphids was not observed (Moretti 1990). The type-B stalks manifest a physical barrier in the defense (Neal et al. 1989) while, besides its function as a deterrent (Lapointes and Tingey 1984), a viscous, adhesive exudate hampers feeding and mobility (Tingey and Gibson 1978).

1.10 Crossing relationships

The tomatoes display a broad range of mating systems, including nearly complete autogamy (*S. lycopersicum, S. cheesmanii, S. neorickii*), facultative outcrossing (*S. chmielewskii, S. pimpinellifolium*) and obligate outcrossing (*S. chilense, S. habrochaites, S. peruvianum, S. huaylasense, S. corneliomulleri, S. arcanum, S. pennellii*; Rick 1963, 1979, 1986c; Table 1). Self-compatible forms have been found among all the outcrossing types, often, but not always among marginal populations (Rick 1979; Rick et al.1979; Rick and Tanksley 1981; Rick 1986c). Self-incompatibility is based on the gametophytic system, and is controlled by a single, multiallelic *S* gene, located on chromosome 1 of tomato (Tanksley and Loaiza-Figueroa 1985). This SI-system is widespread in subgenus *Potatoe*; other *Solanum* subgenera are predominantly self-compatible (Whalen and Anderson 1981; Child 1990). To date there are no reports of self-compatible forms among populations of any of the four tomato-like nightshades (Peralta and Spooner 2001).

Rick (1979) defined the relationships among wild tomatoes on the basis of their ability to hybridize. Within the *Lycopersicon* clade, species segregate into either the '*esculentum* complex' (*S. lycopersicum, S. pimpinellifolium, S. cheesmanii, S. chmielewskii, S. neorickii, S. habrochaites, S. pennellii*) or the '*peruvianum* complex' (*S. peruvianum, S. corneliomuelleri, S. huaylasense, S. arcanum* – considered by Rick to represent one species - and *S. chilense*). Species within each complex are experimentally intercrossable with varying ease but crosses between both complexes require use of embryo culture (Rick 1979). It is important to note that natural hybridization of any of these species is extremely rare, due to pre- and postzygotic reproductive barriers, including problems in the F_1 or later generations. Species barriers within sect. *Lycopersicon* are maintained via reproductive and geographic isolation (Rick 1979).

Among the four tomato-like nightshades crosses only succeed between *S. lycopersicoides* and *S. sitiens* (Rick 1979) and between *S. ochranthum* and *S. juglandifolium,* the latter combination only with the aid of embryo culture (observed in the present study). Though overlapping somewhat in

their natural range, *S. ochranthum* and *S. juglandifolium* do not hybridize spontaneously (Rick 1988; Smith and Peralta 2002). There are no apparent phenological differences, both flower periodically throughout the year. Little is known about their pollinators, but their similar flower structures do not suggest the presence of specialized pollen vectors. A separation on the basis of diurnal specificities is also unlikely given that both species are bee- and therefore daytime-pollinated (Correll 1962; Smith and Peralta 2002). Post-fertilization incompatibilities act either during the prezygotic phase in pollen-stigma or pollen-pistil interactions or the postzygotic phase, from the time of fertilization onwards (Grant 1975). Fruits develop upon artificial interspecific pollination, but seeds are inviable (Rick 1979) and do not normally survive without embryo culture. The evidence therefore suggests that these species are isolated by postzygotic barriers, primarily.

Crosses between *S. sitiens* and *S. lycopersicoides* on the other hand are readily obtained in both directions, and the interspecific hybrid is relatively fertile (Rick 1979; DeVerna et al. 1990). According to observations from plant collectors (Chetelat, pers. comm.), the flowering times of the two are similar, therefore spontaneous hybridizations would be possible in areas of sympatry (Smith and Peralta 2002). The two are thought to have diverged only recently and geographic isolation appears critical for the maintenance of the species barrier (Pertuzé et al. 2002; Smith and Peralta 2002). Smith and Peralta (2002) speculated that different adaptations, i.e. towards extremely cold or extremely dry conditions, may have contributed to their present geographic separation.

Of the four tomato-like nightshades only *S. lycopersicoides* is unilaterally compatible with sect. *Lycopersicon* (Rick 1979). *S. sitiens* and *S. ochranthum* are incompatible with sect. *Lycopersicon* (Rick 1979). *S. juglandifolium* has not been thoroughly tested, other than in crosses with *S. chilense*, which failed in both directions (Graham 2005), and until this point there has been no reason to expect it to be more compatible with the other tomatoes.

Among the four tomato-like nightshades *S. lycopersicoides* has always been the most hopeful candidate for introgressions to tomato, as it is the only species that can be crossed directly to the cultigen, using embryo rescue technique (occasional seeds are formed without embryo culture). Rick obtained a viable F_1 hybrid but backcrosses to tomato failed due to male sterility and stylar incompatibility to tomato pollen, thereby preventing any form of germplasm transfer (Rick 1951). Further progress was not reported until some 35 years later, when a set of alien addition lines was generated from synthetic sesquidiploids using *S. pennellii* as a bridge to overcome incompatibility (Rick et al. 1986; DeVerna et al. 1987). Two additional, independent approaches succeeded to produce backcross (BC_1) material without polyploidization. Gradziel and Robinson (1989) circumvented stylar incompatibility by application of a bud pollination technique, and Chetelat et al. (1989) employed *S. pennellii*–derived bridging lines. However, each of these attempts had its

shortcomings and did not produce the desired introgressions of *S. lycopersicoides* in the *S. lycopersicum* background. Somatic hybrids (*S. esculentum* + *S. lycopersicoides* and (*S. lycopersium* x *S. pennellii*) + *S. lycopersicoides*) did not overcome fertility problems either (Handley et al. 1986; Guri et al. 1991; Hossain et al. 1994). Years later the fortunate discovery of a *S. lycopersicum* × *S. lycopersicoides* hybrid with an unusually high male fertility, which could be used as staminate parent in backcrosses – thereby avoiding stylar incompatibility problems – provided a breakthrough for genetic transfer (Chetelat et al. 1997). A *S. lycopersicoides* introgression library with ~ 96 % genome representation has since been generated (Canady et al. 2005).

S. sitiens can be crossed indirectly to tomato using the sesquidiploid *S. lycopersicum* × *S. lycopersicoides* (LLS) hybrid as a bridging line (DeVerna et al. 1990). Introgression lines representing a portion of the *S. sitiens* genome in the background of cultivated tomato are currently being developed at the TGRC (Pertuzé et al. 2003).

Unfortunately, *S. ochranthum* is reproductively isolated from all *Solanum* relatives except sister species *S. juglandifolium*, and therefore strategies similar to those implemented for *S. lycopersicoides* and *S. sitiens* cannot be used to breach the reproductive barrier to tomato (Rick 1979; Stommel et al. 2001).

Tomato (*S. lycopersicum*) and potato (*S. tuberosum*) have never been hybridized by conventional sexual means (protoplasts fusions are viable – and sterile, as mentioned below). Various crossing attempts have been made between the tomato group and other sections of the genus *Solanum* or different *Solanaceae* genera, but without success (Wann and Johnson 1963; Omidiji 1979; Rick 1979). Wann and Johnson (1963) stated that among the various *Solanum* species, only *S. ochranthum* and *S. etuberosum* showed signs of cross-compatibility with tomato. Ovule development was triggered in *S. ochranthum* upon pollination with *S. peruvianum* and *S. chilense* and in tomato after pollination with *S. etuberosum*, but in both cases no fertilization took place. In an attempt to transfer agronomic traits among sexually incompatible species, many plant breeders resorted to somatic hybridization techniques. Successes have been reported, yet fewer with increasing genetic distance between the participants. Somatic hybrids have been generated between tomato and members of other sections within subgenus *Potatoe*, albeit with varying degrees of success. However, unlike the progress reported in conventional introgression, no backcross progeny was attained from somatic hybrids between *S. lycopersicum* + *S. sitiens* (O'Connell and Hanson 1986). Somatic hybrids between sect. *Lycopersicon* and sect. *Etuberosum* were successfully produced by Gavrilenko et al. (1992) but, as with *S. sitiens*, backcross generations could not be recovered. Melchers (1978) was the first to generate viable somatic hybrid plants between tomato and potato. The sterility of first generation fusion hybrids was overcome sixteen years later by

Jacobsen et al. (1994) who used hexaploid (tomato + potato) somatic hybrids in backcrosses with tetraploid potato. Somatic hybrids between *S. lycopersicum* + *S. ochranthum* were obtained by Kabayashi et al. (1996). Unfortunately, in spite of all efforts reciprocal backcrosses of both the allotetra- and allohexaploid fusion products failed, suggesting that the potential for gene transfer through this route is limited (Stommel et al. 2001). Somatic hybridization attempts were also conducted with a *S. lycopersium* × *S. pennellii* hybrid (Olmstead and Palmer 1997) and the more distantly related eggplant (*S. melongena*; Guri 1991) which is reproductively isolated from tomato (Miwa et al. 1958); these manipulations failed as hybrid growth was arrested after formation of leaf primordia. Even broader fusions, like those of *Petunia hybrida* and *S. peruvianum*, have been attempted but without any success (Tabaeizadeh et al. 1985).

1.11 Genetic mapping in tomato

Early linkage maps, so called "classical maps" were derived primarily from morphological mutants, as well as allozymes and disease resistance genes. The first linkage analyses in tomato date back to the beginning of the 20th century (Hedrick and Booth 1907). MacArthur constructed the first linkage map in the 1920's (MacArthur 1926, 1934). Later cytogenetic analyses were introduced as a tool to characterize genomes (Afify 1933; Rick and Butler 1956; Khush and Rick 1968). The discovery, in the 1960's, of protein-based isozyme markers, represented a major advance for linkage analyses, because these are – in contrast to most morphological markers - codominant, show few epistatic interactions, and are essentially neutral to selection (Chetelat and Li 2006). Beginning in the 1970's Tanksley and Rick generated a tomato isozymic genetic linkage map based on existing classical maps, with the difference that most marker variation came from interspecific crosses, as little variation was present among strains of cultivated tomato (Tanksley and Rick 1980; Tanksley 1985). The advent of DNA-based, molecular markers (RFLPs) in the 1980ies revolutionized the field of genetic mapping, marking the starting point of a proliferation that has been ongoing until this date (Botstein et al. 1980).

Genetic linkage mapping in tomato has been conducted using RFLPs (Bernatzky and Tanksley 1986), AFLPs (Haanstra et al. 1999), RAPDs (Saliba-Colombani et al. 2000), SSRs (Areshchenkowa and Ganal 2002), SCARs (Fulton et al. 1997) and CAPS (Yang et al. 2004).

Genetic linkage maps exist for most of the tomato species and – with the inclusion of the present work – the two pairs of tomato-related nightshades in subsect. *Lycopersicoides* and sect. *Juglandifolium*: the primary reference maps, all available at http://www.sgn.cornell.edu, are based on F_2 *S. lycopersicum* × *S. pennellii* (Tanksley et al. 1992), BC_1 *S. lycopersicum* × *S. habrochaites* (Bernacchi and Tanksley 1997) and BC-recombinant inbred *S. lycopersicum* × *S. pimpinellifolium*

(Grandillo and Tanksley 1996a; Tanksley et al. 1996; Doganlar et al. 2002c). The Tomato-EXPEN 2000 map (from F_2 *S. lycopersicum* LA925 × *S. pennellii* LA716; Fulton et al. 2002a) was used in the present study as a reference; it is comprised of over 2222 markers (RFLPs, CAPS, SSRs and SNPs) with an average spacing of ~ 0.6 cM.

Additional tomato genome maps include cytogenetic maps based on fluorescence in-situ hybridization (FISH; Zhong et al. 1996; Wang et al. 2006) and a T-DNA insert map (Gidoni et al. 2003). Physical maps of the tomato genome were generated from BAC libraries (Budiman et al. 2000; http://www.sgn.cornell.edu).

1.12 History of comparative genetic mapping

Comparative mapping was introduced as a new tool in the 1980s, when DNA based markers (RFLPs) started to be cross-mapped, thereby enabling comparisons of linkage relationships across related taxa. The first comparative linkage maps were generated for *Solanaceae* species (Bonierbale et al. 1988; Tanksley et al. 1988) and grasses (Hulbert et al. 1990). Later, RFLPs were complemented or replaced by PCR-based molecular markers, such as RAPDs, AFLPs, SSRs, SCARs and SNPs. Of particular interest for comparative studies are markers pertaining to the conserved ortholog set (COS). These are single- or low-copy sequences preserved above the family level (Fulton et al. 2002a; Wu et al. 2006). Currently more than half of the ~ 1000 COS markers (conserved between tomato and *Arabidopsis thaliana*) and several hundred of the ~ 2900 identified COSII (conserved within the Asterid I clade of flowering plants, e.g. the *Solanaceae* and the *Rubiaceae*, e.g. coffee) markers have been mapped. More recently, DNA sequence information and genomics approaches are implemented for comparative genome analyses (reviewed in Labate et al. 2007).

A common finding of comparative mapping analyses was that gene order had been remarkably preserved over millennia of evolution among related species (reviewed in Devos and Gale 2000). With the intent to capitalize on collinearity and extrapolate information from well-characterized species (e.g. model organisms) to poorly-characterized species, comparative linkage maps were generated for most agronomically important plant families, including the *Poaceae* (reviewed in Devos and Gale 2000), *Brassicaceae* (Lagercrantz 1998), *Solanaceae* (Tanksley et al. 1992), *Fabaceae* (Boutin et al. 1995), *Compositae* (Burke et al. 2004), *Pinaceae* (Krutovsky et al. 2004) and *Rosaceae* (Dirlewanger et al. 2004). Comparative mapping efforts did not remain restricted to families and were even carried out among species as distantly related as monocots and dicots (Paterson 2000).

1.13 Comparative mapping in tomato and the *Solanaceae*

The first steps toward comparative mapping in tomato were undertaken with isozyme-based data within sect. *Lycopersicon* (Tanksley and Rick 1980) and among tomato and pepper (*Capsicum annuum*; Tanksley 1984). Over the past two decades comparative analyses with DNA-based markers were conducted between tomato and potato (Bonierbale et al. 1988; Gebhardt et al. 1991; Tanksley et al. 1992), tomato and eggplant (Doganlar et al. 2002a), tomato and pepper (Tanksley et al. 1988; Prince et al. 1993; Livingstone 1999), as well as the related potato and *S. etuberosum* (Perez et al. 1999). All studies revealed a high level of homosequentiality among solanaceous species, in particular the presence of conserved marker order organized in reshuffled blocks. For over a decade it was believed that the genomes of tomato and potato (*S. tuberosum*) were separated by five major paracentric inversions involving 5S, 9S, 10L, 11S and 12S. However, an additional whole-arm paracentric inversion on 6S was revealed just recently by Bai et al. (2007). Twenty-three paracentric inversions and five translocations differentiate the genomes of tomato and eggplant (*S. melongena*), while the pepper (*Capsicum annuum*) genome is more extensively rearranged. Direct comparisons of the solanaceous genomes allowed the reconstruction of the divergence of the *Solanaceae* lineages from a hypothetical common ancestor (Livingstone et a. 1999; Doganlar et al. 2002a). Comparative analyses identified conserved segments even beyond the family level; among tomato and *Arabidopsis* (Ku et al. 2000), potato and *Arabidopsis* (Gebhardt et al. 2003) and coffee and tomato (Lin et al. 2005).

Gene order within the tomatoes is highly preserved, as revealed by comparisons of genetic maps based on interspecific crosses between tomato and *S. chmieleswskii* (Paterson et al. 1990), *S. pennellii* (Tanksley et al. 1992), *S. peruvianum* (van Oojien et al. 1994; Fulton et al. 1997), *S. galapagense* (Paran et al. 1995), *S. pimpinellifolium* (Grandillo and Tanksley 1996b), *S. habrochaites* (Bernacchi and Tanksley 1997) and *S. neorickii* (formerly *L. parviflorum*; Fulton et al. 2000). An exception represented the finding of an inversion on 7S between the genomes of tomato and *S. pennellii* (Van der Knaap et al. 2004). The genome of subsect. *Lycopersicoides* (*S. lycopersicoides* and *S. sitiens*) lack the inversion of 10L that separates tomato from potato, making it a cytotaxonomic marker (Pertuzé et al. 2002).

1.14 Applications for comparative information

Comparative linkage information finds applications in plant breeding as well as developmental and evolutionary genetic research (Gale and Devos 1989). One of the great advantages of comparative mapping is that it allows a thorough comparison of genomes among species that are not cross-compatible. Furthermore, comparisons with model organisms provide

access to genomes that are essentially intractable, e.g. large genomes such as wheat that do not lend themselves well to chromosome walking or genome sequencing (Gale and Devos 1998; Devos and Gale 2000). Finally, comparative mapping provides comprehensive insights into genome evolution, genome organization and phylogenetic relationships among taxa (Paterson et al. 2000), as well as to gene function (Labate et al. 2007). Comparative mapping enables the localization of orthologous genes in related taxa. Its predictive power decreases with increasing phylogenetic distance, it is most valuable among species within one family (Paterson et al. 2000). Thus, the extrapolation of functional information from taxa as phylogenetically distant as *Arabidopsis* and potato will most likely be highly error-prone (Gebhardt et al. 2003). Today there are numerous reports where the knowledge about linkage or sequence conservation across species led to the discovery and isolation of favorable monogenic traits and even QTLs in related species. A comparative map-based cloning approach in rice led to the isolation of the *Ph1* gene that controls chromosome pairing in wheat (Foote et al. 1997). Conserved QTLs for fruit size were detected among tomato and pepper (Zygier et al. 2005). In potato the *R3a* late blight resistance gene was isolated based on genomic information about the *I2* complex in tomato. Both loci are members of an ancient locus involved in the defense against oomycetes and fungal pathogens (Huang et al. 2005). A list of physiological/morphological QTLs and disease resistance genes that were the subject of comparative mapping analyses in the *Solanaceae* can be found in Labate et al. (2007).

Stretches of collinearity are even conserved among the two major subclasses of flowering plants, monocots and dicots, which diverged from a common ancestor between 130 and 240 myr (Wolfe et al. 1989; Paterson et al. 1996; Chaw et al. 2004). Transferring linkage information between the two clades seems feasible at the gene level, but not necessarily at the genome level, indicating that collinearity may be restricted only to certain regions (Gale and Devos 1998). Another prominent example is the isolation of the ‘green revolution genes’ where homologues of the *Arabidopsis thaliana* dwarfing gene *GAI* (gibberellic acid insensitive gene) were detected in wheat and maize using rice ESTs (Peng et al. 1999). However, chances to find economically relevant traits for grasses in *Arabidopsis* are deemed extremely small (Devos and Gale 2000).

1.15 Karyotype evolution

Plant genomes are forged by counteracting forces: those that preserve integrity and those that introduce variation (Paterson et al. 2000). Among the most interesting findings that emerged from comparative mapping analyses was the high degree of consensus in terms of mechanisms of genome evolution and sequence conservation even across remotely related species. Common mechanisms by which synteny is eroded include segmental duplications, deletions, chromosomal

rearrangements, mutations and/or genome expansion brought on by transposable element activity (for review see Paterson et al. 2000). A common finding is that genes are typically organized as collinear linkage blocks or gene islands, which are separated by stretches of repeats. The segments generally vary in size among related taxa, but do so in proportion to their genome sizes (i.e. segments in wheat are larger than those in rice; Barakat 1997; Lagercrantz 1998; Paterson et al. 2000; Doganlar et al. 2002a). Collinear blocks are even conserved among species that differ greatly in genome content (Livingstone et al. 1999), but may assume different locations in the genomes among taxa – analogous to a brick-like type of rebuilding - while their internal gene orders remains preserved (Lagercrantz 1998; Paterson et al. 2000; Doganlar et al. 2002a). Both the degree of collinearity and the size of the interspersed stretches most likely vary with the region in the genome (Devos and Gale 1998). Gene order may be conserved in syntenic blocks as a result of selective constraints to ensure the joint expression of adjacent loci that control parts of the same network (Gebhardt et al. 2003). Preserved gene functions are highly common. Houskeeping genes are particularly well conserved, even across kingdoms (Gebhardt et al. 2003), while disease resistance genes are fast evolving (Pan et al. 2000). Studies in the grasses and the *Solanaceae* showed orthologous relationships of QTLs associated with domestication traits (seed dispersal and shattering in the former, fruit weight, shape and color in the latter) among related taxa (Devos and Gale 2000; Doganlar et al. 2002b).

The rates of structural changes vary greatly among taxa, and are strongly influenced by biological characteristics, life history traits and ecogeographic factors. The amount of rearrangements is independent of phylogenetic distance, and closely related lineages can assume very distinct evolution rates (Devos and Gale 2000). The highest rates ever observed in a eukayotic genome occurred in *Drosophila* with more than 100 paracentric inversion between *D. melanogaster* and *D. replete* (Ranz et al. 2001). In plants 0.2 - 2.5 rearrangements happen on average per million years. The highest rates have been reported for genus *Helianthus* (sunflower) with 5.5 - 7.3 rearrangements per million years (Burke et a. 2004). Chromosomal restructuring is more common in annual herbs and decreases in related perennials and woody plants (Grant 1981). However, evolutionary estimates are based on years, not generations and may therefore be underestimated in organisms that reproduce less than once per year such as bi- and perennial plants (Lagercrantz 1998; Burke et al. 2004). Polyploid genomes are generally more dynamic and exhibit faster rates of repatterning (Gale and Devos 1998; Lagercrantz 1998). Theory suggests that polyploid genomes may be more tolerant of rearrangements, because genomic redundancy protects against deleterious effects (Rieseberg et al. 2001). In addition, aberrant meiotic pairing between homoeologous chromosomes promotes the occurrence of rearrangements in polyploids (Lagercrantz 1998).

However, mapping analyses in recent polyploids are generally more complex due to a high number of paralogs (Gale and Devos 1998). Rearrangements are often only deleterious in a heterozygous state (Levin 2002) and therefore may have a greater chance of survival in self-compatible species (Lande 1979; Lagercrantz 1998). Finally, demographic conditions are likely to have an impact on the establishment of structural changes which are more frequent in marginal (Levin 2002) or small populations, presumably because newly arisen rearrangements encounter better prospects to become fixed via genetic drift (Lande 1978).

The mode of genome evolution that predominates in a lineage is often a characteristic feature for specific plant families (Paterson et al. 2000). This is not unexpected as evolutionary mechanisms are a function of life history traits and genome specificities (Lagercrantz 1998). Speciation in the *Brassicaceae* family involved several auto- and allopolyploidization events, as classically illustrated in U's triangle, that shows how members of the genus *Brassica* arose from natural hybridization events (U 1935). Polyploidization also played a significant role in the evolution of the grasses. Hexaploid wheat (*Triticum aestivum*) is comprised of 3 subgenomes (McFadden and Sears 1946) that underwent ancient regional gene duplications in an ancestral genome (Hart 1983), and extensive autopolyploidization characterizes the genus *Saccharum*, where haploid chromosome counts may reach 70 and vary substantially even within single species (Ming et al. 1998). Sunflower (*Helianthus*) and species in the *Solanaceae* family are primarily differentiated by chromosomal rearrangements (Tanksley et al. 1992; Livingstone 1999). Finally, transposable element activity has had a strong impact on the formation of the pepper genome (Livingstone et al. 1999) and the maize genome (*Zea maize*, family *Poaceae*) where mobile element duplications account for over 50 % of the DNA content (SanMiguel et al. 1996).

1.16 Future trends of comparative genome analyses

While comparative mapping using conventional markers is likely to retain significance as powerful tool for coarser, genome-wide characterizations, especially of less well-studied genomes, the trend is toward comparative genomics, i.e. to focus in on particular DNA sequences in order to infer patterns of evolution and dissect biological networks (Ku et al. 2000; Fei et al. 2004). A recent impressive example was given by the exhaustive black cottonwood (*Populus trichocarpa*)/*Arabidopsis thaliana* comparison that revealed – among an array of other intriguing findings – an ancient "eurosid" duplication event that predated the divergences of the two lineages (Tuskan et al. 2006). The field of comparative analyses will fast be taken to the next level as new technologies are on the horizon that will facilitate high-throughput sequencing, gene expression profiling and genome annotation (Caicedo and Purugganan 2005).

1.17 Genetic diversity and genetic structure analyses in natural populations

Knowledge about levels and partitioning of genetic diversity provides insights into a species' ecological status and evolutionary history, and is essential for the development of conservation strategies (Ellstrand and Elam 1993). Miller and Tanksley (1990) suggested the evaluation of germplasm accessions according to 'the likelihood that they contribute new alleles to a collection'. In order to obtain a frame scale they proposed a comparison of the average genetic distances among accessions within a given species.

The amount of genetic variability comprised within species was a matter of much speculation prior to the introduction of protein markers in the 1960ies. The first diversity assays published in 1966 (Johnson et al. 1966, Harris 1966), represented the starting point of the production of vast amounts of genotypic surveys. Later, in the 1980ies, as DNA-based markers became available, RFLP fingerprints from hypervariable minisatellites (VNTRs; variable number of tandem repeats) were discovered as a tool to characterize natural populations (Jeffreys et al. 1985; Hill 1987). The invention of the polymerase chain reaction (PCR) in the late 1980ies (Mullis et al. 1986) opened an entire new field of marker methodologies and applications for genetic diversity characterizations. Among those were the microsatellite technique (STR; Tautz 1989), single-strand conformational polymorphisms (SSCP; Orita et al. 1989) and random amplified polymorphic DNA (RAPD) in the early 1990ies (Williams et al. 1990), followed by the invention of the amplified fragment length polymorphism (AFLP) technology in the mid-1990ies (Vos et al. 1995). Over the past few years due to technological improvements sequencing analyses have become more affordable and quickly advanced to become the golden standard for diversity structure characterizations in species. A significant bonus unique to the analysis of DNA sequences is the ability to directly infer phylogenies (due to the fact the data can be ordered), thereby enabling phylogeographic assays. However, a drawback associated with sequencing experiments is that the intense investigation needs to remain limited to just a few loci. Therefore, to obtain good genome coverage, combinations with PCR-based marker systems often constitute the approach of choice.

1.18 Diversity analysis in tomato

In wild tomato species genetic diversity has often been characterized using allozymes (Rick and Fobes 1975; Rick et al. 1976, 1977, 1979; Rick and Tanksley 1981; Rick and Holle 1990; Breto et al. 1993), but also RFLPs (Miller and Tanksley 1990), RAPDs (Egashira et al. 2000; Ercolano et al. 2005), microsatellites (Alvarez et al. 2001) and DNA sequences (Caicedo and Schaal 2004; Roselius et al. 2005). A consensus finding of these various genetic diversity analyses was that *S.*

peruvianum represented the most diverse species within the tomato clade, followed by *S. chilense*, and that the self-compatible species are more genetically depleted.

1.19 Nature of genetic diversity

Levels and distribution of genetic diversity are determined by extrinsic factors such as ecological processes, demographic events (e.g. population subdivision), the presence of soil seed banks (i.e. a repository of viable seed in the soil), natural selection and genetic drift, or intrinsic factors like the mating system, population turnover rates, local chromosomal recombination rates, mutation rates and gene flow (for reviews see Ellstrand and Elam 1993; Avise 2004). The importance of genetic drift increases with decreasing effective population sizes (N_e) relative to that of mutation (Avise 2004). There is typically a positive correlation between the amount of genetic diversity and the level of outcrossing in plants (Gottlieb 1981), which has also been observed in tomato many times and with a variety of approaches (e.g. Rick 1979; Miller and Tanksley 1990; Stephan and Langley 1998; Alvarez 2001). A positive correlation has been reported for nucleotide diversity and local crossing-over rates in various plant and animal species (Begun and Aquadro 1992), including tomato (Stephan and Langley 1998). Recombination counteracts the homogenizing effects of selective sweeps ('hitchhiking effects') or background selection (Stephan and Langley 1998; Roselius et al. 2005). A major challenge in the interpretation of molecular diversity data lies in the distinction of historical demographic events from recent occurrences of mutations, drift or migration. Only DNA sequence data, genealogically ordered and analyzed in a 'phylogeographic' fashion – i.e. against the backdrop of geographic information – (phylogeographic analyses; Avise 2000) allow to discern between the two.

1.20 Allozymes

Allozymes are still used for genetic diversity studies but have often been criticized due to shortcomings in comparison with other methodologies. Among the most common critique points are the following: the number of available loci is small compared to that of other marker systems, and it is unlikely that many new loci will be discovered in the future (Gao et al. 2002). To date a total of ca. 100 allozyme systems exist, but typically only 10 - 30 are available for a given species (molecular marker review in Avise 2004). The lack of allelic variability often further limits their usefulness in genetic diversity analyses (Gao et al. 2002). Allozyme estimates yield lower levels of diversity than DNA-based markers due to smaller effective mutation rates. Only the fraction of mutations that is manifested at the protein level (i.e. non-synonymous or replacement substitutions), and that results in proteins with altered electrophoretic properties will create new alleles (Avise

2004). The allozyme mutation rate was estimated 10^{-6} - 10^{-7} per locus and generation (Kahler et. al 1984). Furthermore, as many of the enzyme systems pertain to specific biological areas such as the glycolytic pathway or the citric acid cycle, doubts exist whether they truly reflect genome-wide diversity (Avise 2004). In addition, proteins are targets of natural selection, and non-neutrality of allozyme markers has been observed on many occasions (e.g. Karl and Avise 1992; Dhuyvetter et a. 2004). To date allozyme analyses have largely been supplanted by more modern methodologies that offer a greater return of information for a smaller amount of labor and cost investment plus greater detection sensitivities, less reporting bias, broader genome representation (both spatially and qualitatively) and higher prospects for selective neutrality. However, allozymes are still useful for comparisons with older literature, for example in tomato.

1.21 Microsatellites

Microsatellites or 'simple sequence repeats' (SSRs) are short stretches of DNA composed of up to 60 repeat units of one to six basepair motifs (Goldstein and Pollock 1997). Mutation rates are several orders of magnitude higher than those of non-repetitive eukaryotic sequences, an estimated 10^{-3} to 10^{-5} vs. 10^{-9} per locus and generation, respectively (Weber and Wong 1993; Schug et al. 1997; Ellegren 2000; Vigouroux et al. 2002). The genetic instability has been attributed to their propensity for replication slippage (Levinson and Gutman 1987) and unequal crossing over events (Jakupciak and Wells 2000).

Microsatellite motifs occur in the genomes of many eukaryotes. Mutation rates and patterns are substantially heterogeneous with respect to species type, sex and age (reviewed in Ellegren 2000; Azaiez et al. 2006), locus position and flanking regions (di Rienzo et al. 1994), repeat motif class (di-, tri- or tetranucleotide) and base composition (Broun and Tanksley 1996), microsatellite type (perfect, compound or interrupted), allele length (Ellegren 2000; Balloux and Lugon-Moulin 2002) and insert orientation (Azaiez et al. 2006). In plant genomes the by far most predominant motif is the dinucleotide AT (Morgante and Olivieri 1993) followed by mono-, tri- and tetranucleotides (Wang et al. 1994). Microsatellite repeats were shown to be three times more abundant in dicotyledons compared to monodicotyledons (Wang et al. 1994). Clustering of mono-, di-, tri- and tetranucleotides in centromeric regions was observed in several plant species including Arabidopsis (Brandes et al. 1997), sugar beet (Schmitdt and Heslop-Harrison 1996) and tomato, where they appear associated with the high concentration of repetitive sequences (Areshchenkova and Ganal 2002). However, this distribution pattern may not be universal. In many grasses, for example, both microsatellites and other repetitive sequences are more randomly distributed throughout the genome (Areshchenkova and Ganal 2002). Base composition and motif length influence SSR distribution

and variability. Units composed of the bases C and G have been shown to be more mutable than those composed of the bases A or T in yeast (Harfe and Jinks-Robertson 2000). In tomato 'AT' motifs tend to be more evenly distributed throughout the genome (Areshchenkova and Ganal 2002). Other types of dinucleotides, especially long dinucleotide repeat motifs and tetranucleotides are clustered around centromeres (Arens et al. 1995; Broun and Tanksley 1996; Areshchenkova and Ganal 1999, 2002), whereas trinucleotides are the prevailing microsatellite class in tomato coding regions (Smulders et al. 1997). SSRs are rare in coding regions (Wang et al. 1994) with the exception of tri- and hexanucleotides (Toth et al. 2000), most likely because these repeat types do not result in frameshift mutations, and are thus less detrimental (Malpertuy et al. 2003). SSR mutations occur asymmetrically; alleles tend to mutate into larger alleles (Amos et al. 1996; Jarne and Lagoda 1996). However, a general finding in many organisms is that microsatellite length does not increase indefinitely, i.e. that there is a "length ceiling" for microsatellite alleles (Garza et al. 1995). The underlying causes remain uncertain. Long microsatellite repeats may be subject to negative selection (Li et al. 2002). Alternatively, the size reduction of longer microsatellites may simply be a function of stochasticity; i.e. the larger the allele, the higher the probability that a mutation will reduce its size (Vigouroux et al. 2002). Size limitations may also be caused by rare large deletions (Amos et al. 1996) or interruptions that inhibit replication slippage (Smulders et al. 1997), thereby 'stabilizing' the microsatellite (Vosman and Arens 1996).Variability has been found to decrease with repeat unit size, i.e. dinucleotides show higher mutation rates than tri- or tetranucleotides (Schug et al. 1998; Vigouroux 2002). This may be a consequence of the size constraint (i.e. the size ceiling), allowing dinucleotide SSRs to accumulate higher numbers of repeats than those composed of larger units (Schug et a. 1998). However, the opposite, i.e. greater variability associated with larger repeat unit sizes, has also been observed (Weber and Wong 1993). Variability has often been shown to increase with the number of repeat motifs (Vigouroux et al. 2002; Frary et al. 2005; Azaiez et al. 2006), albeit not in a linear fashion (Goldstein and Pollock 1997). In tomato, both a positive correlation between number of repeat motif and variability (Smulders et al. 1997; Areshchenkova and Ganal 1999; Frary et al. 2005) as well as no correlation (Broun and Tanksley 1996) have been observed.

The mutation modus has been a long-standing issue of debate that continues to this date (for review see Balloux and Lugon-Moulin 2002). Two contrasting mutation models dominate microsatellite interpretation: the infinite alleles model (IAM; Kimura and Crow 1964) and the stepwise mutation model (SMM; Otha and Kimura 1973). According to the IAM each mutation results in a novel allele, i.e. the occurrence of homoplasy is excluded. In the SSM on the other hand, a mutation will cause either the deletion or duplication of a repeat motif and thereby generate a ladder-like increase

or decrease in allele size. Alleles that are similar in size are assumed to be more closely related than those that are more different in size. Thus, the SMM incorporates additional information. Modified or extended versions exist for both models: the *K*-allele model (KAM) was derived from the IAM to account for homoplasy (Kimura 1968), and the two-phase model (TPM), a modification of the SMM, allows a certain proportion of mutations to involve more that one repeat unit and to create larger sized 'steps' (Di Rienzo et al. 1994). The analytical output may vary considerably depending on the mutation model that is implemented (Lugon-Moulin et al. 1999), therefore caution is warranted regarding the interpretation.

None of the existing models achieves to fully analyze microsatellites, as these do not follow a simple mutation pattern (Ellegren 2000). The true mutation modus may best be described by a mixture of these models. Due to the extreme complexity of the mechanisms it is questionable whether there will ever be a model that captures the nature of microsatellites appropriately (Anderson et al. 2000; Balloux and Lugon-Moulin 2002). Most mutations appear to follow the stepwise modus, but that proportion may vary substantially (Awadalla and Ritland 1997; Anderson et al. 2000; Vigouroux et al. 2002). Additions or deletions of single repeat motifs are more frequent than those of several motifs (Weber and Wong 1993; di Rienzo et al.1994). In animal species an estimated 4 to 74 % are multistep changes (reviewed in Ellegren 2000). Non-stepwise SSR allelic variability is generated by mutations in the flanking regions of the repeat segment (Buteler et al. 1999; Matsuoka et al. 2002).

SSR markers have been developed for e.g. wheat (Röder et al. 1998), barley (Liu et al. 1996), maize (Taramino and Tingey 1996), rice (McCouch et al. 1997), mouse (Dietrich et al. 1996), cattle (Ihara et al. 2004) and humans (Dib et al. 1996). In tomato 188 mapped SSRs are available at the SGN website (http://www.sgn.cornell.edu), most of which include primer information. The advantages of microsatellite makers are their codominant nature, their universal abundance (Morgante and Olivieri 1993), good reproducibility, locus-specificity (Frary et al. 2005), a high sensitivity that allows distinctions even between closely related individuals, a high proportion of single-fragment amplification – which is especially attractive for species with genetic redundancies (Bindler et al. 2007) –, predominant selective neutrality (reviewed in Li et al. 2002) and cost-effectiveness (Loridon et al. 2005). Although microsatellite markers are generally regarded as selectively neutral they may assume regulatory or protein-coding functions and have been proposed as intrinsic source of genetic variation, providing the ability for evolutionary adaptations (Kashi et al. 1997; Li et al 2002; Nevo et al. 2005). Problems associated with microsatellite applications are those generally encountered with PCR-based techniques. Null alleles may result as a failure of amplification, and homoplasy as a consequence of recurrent mutations. The hypervariability of microsatellites makes

the occurrence of size homoplasy especially likely, however, this is not expected to become a concern for the interpretation of population genetic studies that focus on shorter evolutionary distances such as at the below-species level (Jarne and Lagoda 1996) unless the mutation rates are exceptionally high, population sizes are large and size constraints on alleles are strong (Estoup et al. 2002). Another drawback of SSR markers is that primer development is laborious, requiring genetic library generation, screening and clone sequencing (Areshchenkova and Ganal 2002). Recently, however, with the accumulating amount of sequence data, SSRs may also be detected through in silico mining of publicly available databases (Feingold et al. 2005). As a second advantage, the screening of EST sequences for SSRs avoids an overrepresentation of centromeric and telomeric regions (Frary et al. 2005). A further disadvantage of SSRs is the typically low transferability across species (Van de Wiel et al. 1999) which can be mitigated by developing primers from conserved sequences (Frary et al. 2005).

1.22 Objectives

Two research projects focusing on four species belonging to section *Juglandifolium* and subsection *Lycopersicoides* of the genus *Solanum* are reported in the present thesis; the construction of a comparative genetic linkage map for *S. ochranthum* and *S. juglandifolium* and assessment of genetic diversity and genetic structure in wild populations of *S. lycopersicoides* and *S. sitiens*.
A comparative genetic mapping analysis was conducted to compare the genome of *S. ochranthum* and *S. juglandifolium* vis-à-vis to that of tomato and related solanaceous species, with the ultimate aim to

1) estimate their potential as a source for tomato germplasm enhancement, and to
2) provide insights into genome evolution in the genus *Solanum*.

The principle aim of estimating levels of genetic diversity and genetic structure in wild populations of *S. lycopersicoides* and *S. sitiens* was to acquire information on the amount of genetic diversity and its geographic structuring both on the protein and the DNA level in order to

3) provide guidelines for conservation strategies, and to
4) elucidate the species' demographic histories.

2. Materials and methods

2.1 Comparative mapping analysis

A comparative mapping analysis based on an F_2 mapping population derived from an interspecific cross was conducted to characterize the genomes of *S. ochranthum* and *S. juglandifolium* relative to that of tomato. The following chapter describes the generation of the mapping population, fertility measurements, DNA isolation, marker analysis and map construction.

2.1.1 Plant material

An F_2 mapping population was generated from an interspecific cross of *S. ochranthum* LA3650, collected by Charles Rick and Miguel Holle at Choquemaray, Apurimac, Peru and *S. juglandifolium* LA2788, collected by Charles Rick and Miguel Holle at Quebrada La Buena, Antioquia, Colombia. Seeds were provided by the C. M. Rick Tomato Genetics Resource Center (TGRC), Department of Plant Sciences, University of California, at Davis (USA). Due to self-incompatibility of both parental species and their interspecific hybrid, two F_1 plants were crossed to obtain 66 pseudo-F_2 progeny. All plants were grown at the greenhouse facility at UC Davis at approximatetly 21˚ C day temperature, 16˚ C night temperature and 65 % relative humidity. Daylength periods corresponded to the natural day lengths at Davis. Light intensity was artificially elevated during winter months by overhead light sources.

2.1.2 Embryo culture

All F_1 and F_2 plants were obtained via embryo culture. Fruits were harvested ~ 45 days post-pollination, surface-sterilized for 10 min in 70 % (v/v) ethanol and ~ 1.25 % (w/v) sodium hypochlorite and rinsed prior to the extraction of the ovules. Embryos were cultured first on the HLH medium of Neal and Topoleski (1983), then transferred after 10 - 14 days to Gamborg's B-5 basal media with minimal organics (Sigma, St. Louis, MO), prepared according to Sacks et al. (1997). After 3 - 7 weeks of development *in vitro* plantlets were transferred to soil, acclimatized and moved to the greenhouse.

2.1.3 Pollen stainability

Anthers of five flowers per F_1 plant (01L5288-1 and 01L5311-1) were squashed in acetocarmine (1 % w/v in 50 % v/v glacial acetic acid) on three different days. The number of grains that were presumed to be viable (i.e. grains that were intact, full-sized and well-colored) out of a total of 100 grains was counted under the microscope.

2.1.4 Chromosome pairing

Chromosome pairing was examined during meiosis in the F_1 hybrid 01L5288-1 using the acetocarmine sqash method (Khush and Rick 1963). Developing floral buds were soaked in fixative (3:1 95 % EtOH : glacial acetic acid with $FeCl_3$) for 24 hours, washed 3x with 70 % EtOH and stored at 4° C. The number of uni-, bi- and multivalents was recorded in eight individual cells using a Zeiss Axioskop compound microscope.

2.1.5 DNA isolation

Young leaves were harvested from mature plants of both parental accessions (for a better representation of the parental genotypes samples from five individuals per species were bulked), the two F_1 hybrids and 66 F_2 plants. DNA was extracted using to the polytron method (Chetelat et al. 1995; Chetelat and Meglic 2000).

2.1.6 Marker analysis

Single copy markers, previously mapped in interspecific crosses with cultivated tomato (*S. lycopersicum*) and the wild species *S. pennellii* (Tanksley et al. 1992; Fulton et al. 2000; Frary et al. 2005) with LOD scores ≥ 2 were selected for whole genome coverage at an average distance of 10 cM (Appendix 2). Tomato-EXPEN 2000 and in some cases Tomato-EXPEN 1992 were used as reference maps. Both maps are based on F_2 *S. lycopersicum* cv. VF36 × *S. pennellii* LA716 and are available at the SOL Genomics Network database (http://www.sgn.cornell.edu).

Marker types included RFLPs, COS markers, COSII markers and microsatellites. RFLPs are based on tomato genomic clones ('TG'; Zamir and Tanksley 1988; Miller and Tanksley 1990) and tomato leaf epidermal cDNA clones ('CT'; Yu et al. 1991). COS (Conserved Ortholog Set) markers represent single or low copy genes conserved among tomato and Arabidopsis (Fulton et al. 2002a). COSII markers are based on genes that have a single homologous match in Arabidopsis (Wu et al. 2006). Simple sequence repeats (SSRs) are derived from tomato ESTs (Frary et al. 2005).

CAPS (Conserved amplified polymorphic sequence). Primers were designed from sequence information available at the SOL Genomics Network database (http://www.sgn.cornell.edu) using the primer design program Primer3 (Rozen and Skaletsky 2000; http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi), and assembled by Sigma Genosys and Operon Technologies. Primers were developed from RFLP and unigene sequence information. COS markers were developed from SGN unigene build sequences that were assembled from ESTs of tomato and a number of closely related species (i.e. species within *Solanum* sect. *Lycopersicum*). Primer pairs were selected to

amplify ~ 1000 bp (+/- 700 bp). Preference was given to regions supported by a high number of member sequences. In cases where unigene builds were relatively short (≤ 600 bases) or composed of only few EST member sequences, alternative COS sites were chosen in the genomic vicinity.

The majority of the COSII loci were amplified using the *Universal Primers for Asterid Species* (UPA), available at http://www.sgn.cornell.edu (Wu et al. 2006). A few primers were designed from sequence information according to the procedure described above. PCR amplifications were conducted in a total volume of 20 µl in a thermal cycler (GeneAmp; Applied Biosystems). Each cycling reaction contained reaction buffer (1x final concentration; Applied Biosystems), $MgCl_2$ (1.5 mmol/l; Applied Biosystems) dNTPs (200 µmol/l; Applied Biosystems), bovine serum albumine (400 µg/µl; New England Biolabs), *Taq* polymerase 0.05 U/µl Applied Biosystems, 0.3 mmol/l primer (forward + reverse) and template DNA (100 ng). Amplification consisted of an initial denaturation for 5 min at 94˚ C, followed by 40 cycles of amplification with denaturation at 94˚ C for 30 s, annealing at 55 - 62˚ C for 1 min and extension at 72˚ C for 1 min, and a final extension at 72˚ C for 10 min. The optimal annealing temperature was determined for each primer pair on parental genomic DNA in a gradient PCR (Techne). Primer performance was evaluated by electrophoresis on 1.8 - 2.0 % agarose (Amresco) followed by ethidium bromide staining. The locus identity was evaluated by comparing PCR product sizes obtained from genomic DNA of *Lycopersicum esculentum* cv. VF36 to those predicted by the *Primer3* program.

Primer pairs that resulted in single-band amplification were digested with a set of eight frequently cutting restriction enzymes (*Bso*BI, *Hinf*I, *Msp*I, *Ban*I, *Hae*III, *Sty*I, *Hae*II and *Hha*I; New England Biolabs, Promega). Digests were carried out according to the manufacturer. Digested DNA fragments were separated on 1.8 - 2 % agarose gels. If none of the restriction enzymes yielded polymorphic banding patterns among the parental genotypes a second set of digests was performed using eight additional restriction enzymes (*Ava*II, *Dra*I, *Nci*I, *Rsa*I, *Alu*I, *Dpn*I and *Dde*I). Successful primer pair/restriction enzyme combinations were confirmed on both F_1 hybrids and 4 - 7 F_2 plants to rule out, e.g. residual polymorphism, before they were used in the screen of the entire F_2 population.

SSRs (Simple Sequence Repeats). Primer sequence information was obtained from the SOL Genomics Network database. Primer assemblage and reaction mix preparation were as described above. Thermocycling reaction conditions were based on those employed by Frary et al. (2005): after an initial denaturation for 5 min at 94˚ C, 40 cycles of amplification consisting of 30 s denaturation at 94˚ C for, 45 s annealing at 55˚ C and for 45 s extension at 72˚ C followed by a final extension of 72˚ C for 10 min. Depending on the size difference between parental fragments,

samples were resolved on either 1.8 - 2 % agarose and stained with ethidium bromide or on 5.2 % polyacrylamide using a LiCor (4200) sequencing apparatus to detect fluorescently labeled fragments. The tailed primer method was used to incorporate fluorescence dye (IR-700 or IR-800) into PCR amplicons via a labeled M13 primer (TTTCCCAGTCACGACGTT; MWG-Biotech) that was added at 0.05 μg/μl to the PCR reaction.

RFLPs (Restriction fragment length polymorphism). Probe identity had been confirmed previously by comparison of insert sizes and restriction fragment banding patterns with those from the SolGenes database (http://ukcrop.net/perl/ace/search/SolGenes; Pertuzé et al. 2002). Total genomic DNA of both parents, the interspecific hybrids and 66 F_2 individuals were digested with a panel of six restriction enzymes (*Eco*RI, *Eco*RV, *Hinf*I, *Xba*I, DraI, *Hae*III; New England Biolabs; Promega). Digestions were performed according to the manufacturer's instructions. Samples were separated electrophoretically on 0.8 % agarose and subsequently blotted onto nylon membranes (Hybond-N+; Amersham). Probes were labeled with [^{32}P]-dCTP and [^{32}P]-dATP using the random hexamer primer method (Feinberg and Vogelstein 1983) and washed three times to a final stringency of 0.5x SSC before exposure to X-ray film (Kodak BioMax MS; Fuji Super RX) at -80° C. Surveys were conducted to identify probe/restriction enzyme combinations that produced polymorphisms between parental genotypes.

2.1.7 Map construction

The χ^2 goodness-of-fit statistic was employed to test for deviations from expected Mendelian segregation ratios (1:2:1) at $P < 0.05$. Linkage analysis and map construction were conducted with MapMaker version 2.0 for Macintosh (Lander et al. 1987). Linkage groups were assigned with LOD $\geq$ 4 and a recombination fraction $\leq$ 0.3. The stringency was raised to LOD = 6 in order to resolve the spurious association of two chromosomes (see results). Kosambi's mapping function was used to convert recombination fractions into map units [cM] (Kosambi 1944):

$$cM = -\ln[(1 + 2y)/(1 - 2y)]$$

(y = recombination fraction)

2.2 Genetic diversity analysis

Wild populations of *S. lycopersicoides* and *S. sitiens* were analyzed for their content and structure of genetic diversity. Experimental procedures, from generation of plant material to marker and data analysis are described below.

2.2.1 Plant material

Fourteen *S. lycopersicoides* and seven *S. sitiens* accessions[2] were selected according to a) geographic distribution (with the aim to sample the entire distribution range), b) population and sample size (prioritizing larger populations and samples), and c) availability of passport data (selecting accessions with more detailed collection information.; Table 2, Figure 1). One *S. chilense* accession (LA2773), sympatric with *S. lycopersicoides*, was included as a reference. Samples of *S. lycopersicum* (cv. VF36), *S. pennellii* (LA0716) and F_1 *S. lycopersicum* × *S. lycopersicoides* (LA3857) served as gel scoring standards. Seeds and collection data were obtained from the C. M. Rick Tomato Genetics Resource Center (TGRC). Each population was represented by 11 - 29 plants, composed of five half-sib progeny from five different mother plants, depending on the availability of collection information and seed material. Seeds were sown from 'original' seed (i.e. seed collected from plants growing *in situ*). After soaking in 2.5 % sodium hypochlorite for 40 min, seeds were rinsed under running water for 15 min and incubated on moist absorbent paper inside transparent plastic boxes in a germination chamber (25˚ C, 12 h photoperiod). Those that failed to germinate within 10 days were nicked (i.e. the seed coat was opened with a scalpel in proximity of the radicle) to facilitate radicle emergence. At the cotelydon stage seedlings were transplanted into flats filled with 'desert soil mix'[3] in order to accommodate the requirements of these species – especially *S. sitiens* – for dry conditions. Small plantlets were transplanted into five gallon pots of the same soil medium type. *S. sitiens* accessions were grafted onto F_1 *S. lycopersicum* × *S. pennellii* rootstock to obtain satisfactory plant growth. Plants were grown at the UC Davis greenhouse facilities at the conditions as described above. Young leaf tissue was harvested for DNA extractions, young shoot material for allozyme analyses. Not all plants were genotyped with both sets of markers; 85 % and 56 % of plants of *S. lycopersicoides* and *S. sitiens*, respectively, were

[2] Accessions LA4110 and LA4111 were located at less than 3 km from each other in the wild and therefore considered as a single population (they are maintained by the TGRC as one population). They are referred to as 'LA4111' in the present text.

[3] The 'desert soil mix' is composed of 10 l peat moss, 10 l coarse sand and 10 l redwood bark to which 0.215 kg dolomite and 0.160 kg fertilizer (14-14-14) are added subsequent to steam sterilization.

Population number	Accession number	Population name	g.l.	Location	Province	Country	Latitude	Longitude	Altitude (m)	Population size	Collection year
	S. lycopersicoides										
1.	LA4018	Aricota #1	1	Lago Aricota	Tacna	Perú	-17.333	-70.250	2888	many	1988
2.	LA2387	Aricota #2	1	Lago Aricota (Tarata)	Tacna	Perú	-17.355	-70.313	2852	NA	1981
3.	LA1964	Chupapalca	2	Chupapalca	Tacna	Perú	-17.761	-69.912	3459	50	1979
4.	LA1966	Palca	2	Palca	Tacna	Perú	-17.767	-69.950	3134	60	1979
5.	LA2781	Putre #1	3	Desvío a Putre	Tarapacá	Chile	-18.199	-69.540	3736	large	1986
6.	LA2777	Putre #2	3	Putre	Tarapacá	Chile	-18.203	-69.564	3462	large	1986
7.	LA2776	Perquejeque	3	Catarata Perquejeque	Tarapacá	Chile	-18.210	-69.596	3115	immense	1986
8.	LA2772	Zapahuira	3	Zapahuira	Tarapacá	Chile	-18.271	-69.580	3416	many	1986
9.	LA4320	Lluta	3	Alto Río Lluta	Tarapacá	Chile	-18.318	-69.805	1509	small	2005
10.	LA4130	Pachica	4	Pachica (Camarones)	Tarapacá	Chile	-18.908	-69.604	2672	> 20	2001
11.	LA4131	Esquina	4	Esquina (Camarones)	Tarapacá	Chile	-18.926	-69.551	2341	> 68	2001
12.	LA4126	Nama	5	Camiña – Nama	Tarapacá	Chile	-19.287	-69.396	3156	> 50	2001
13.	LA4123	Camiña	5	Camiña	Tarapacá	Chile	-19.306	-69.421	2599	> 50	2001
14.	LA2730	Moquella	5	Moquella	Tarapacá	Chile	-19.404	-69.600	1719	NA	1985
	S. sitiens										
15.	LA4116	Paqui	1	Quebrada de Paqui	Antofagasta	Chile	-22.159	-68.782	2935	> 100	2001
16.	LA4114	Carbonatera	1	Pampa Carbonatera	Antofagasta	Chile	-22.191	-68.757	2736	> 35	2001
17.	LA4113	Cere	1	Estación Cere	Antofagasta	Chile	-22.235	-68.762	2652	> 20	2001
18.	LA4112	Limón Verde	2	Aguada Limón Verde	Antofagasta	Chile	-22.617	-68.948	2780	> 47	2001
19.	LA4331	Quimal	3	Cerro Quimal	Antofagasta	Chile	-22.969	-68.821	3074	15-20	2005
20.	LA4110/4111	San Juan	4	Mina San Juan	Antofagasta	Chile	-23.098	-69.033	2718	> 60	2001
21.	LA4105	Escondida	5	Mina La Escondida	Antofagasta	Chile	-24.211	-69.241	2618	> 20	2001

Table 2. *S. lycopersicoides* and *S. sitiens* accessions used in the diversity analysis.

Passport data were obtained from http://tgrc.ucdavis.edu. More detailed information is available on the website. 'Population size' is the observed number of individuals in the wild. 'g.l.' indicates the primary geographic locations as referred to in the text: in *S. lycopersicoides* 1) Lago Aricota, 2) Palca, 3) Putre, 4) Camarones and 5) Camiña; in *S. sitiens* 1) north of Chuquicamata, 2) Aguada Limón Verde, 3) Cerro Quimal, 4) Mina San Juan and 5) Mina La Escondida.

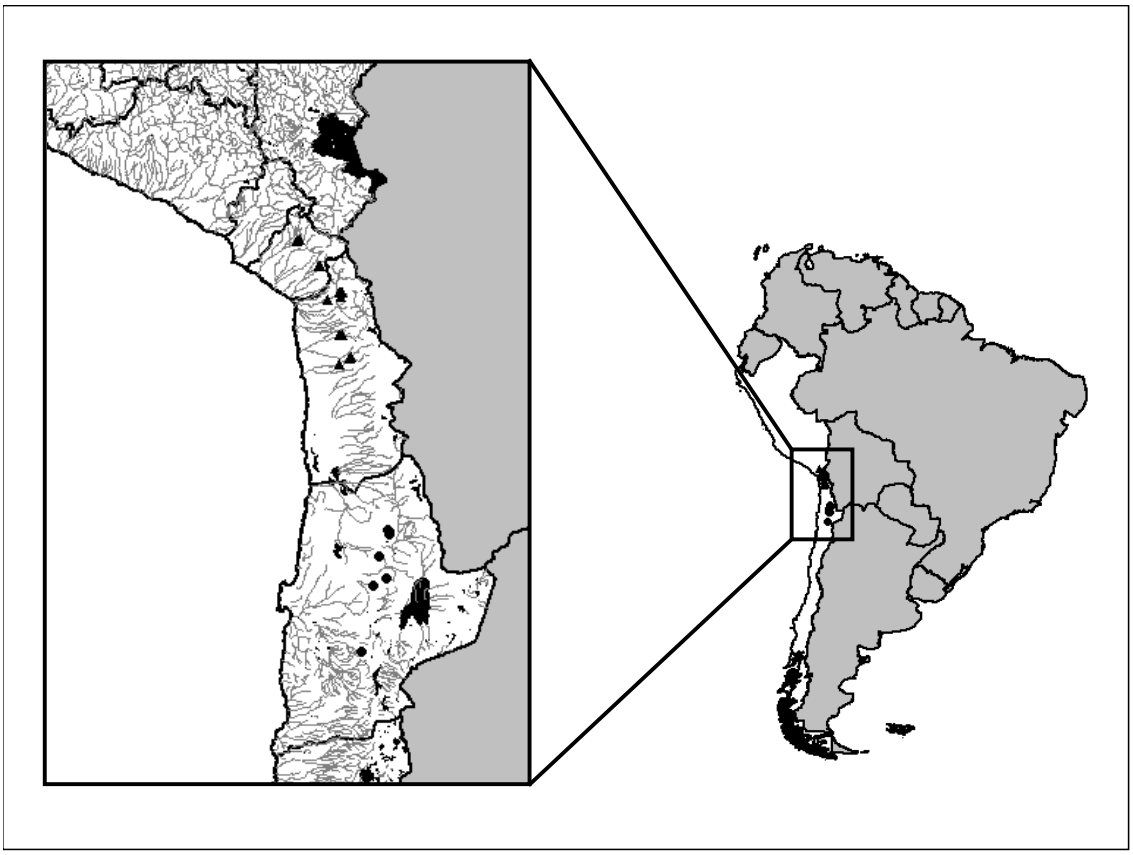


Figure 1. Geographic distribution of *S. lycopersicoides* and *S. sitiens* in southern Peru and northern Chile.

S. lycopersicoides populations are indicated by black triangles, *S. sitiens* populations by black circles.

genotyped for both marker types. Disregarding the three additional populations in the SSR dataset, portions of identical plants amounted to 94 % and 65 %, respectively. To prevent discarding of valuable data, calculations were based on the complete as well as the reduced, 'common' set of individuals for direct comparisons and combined data analyses.

2.2.2 DNA isolation

DNA was extracted according to the 'microprep protocol' (Fulton et al. 1995). DNA of *S. chilense* LA2773 was kindly provided by Elaine B. Graham.

2.2.3 Marker analysis

2.2.3.1 Allozymes

Plant shoot sample preparation, horizontal slab gel electrophoresis and staining procedures were carried out as described in Chetelat et al. (1997). Eight out of ten tested enzyme systems were chosen yielding 14 polymorphic out of a total of 16 loci (Table 3). These are *Aco-1* (aconitase), *Aco-2*, *6-Pgdh-1* (6-phosphogluconate dehydrogenase), *6-Pgdh-2*, *6-Pgdh-3*, *Idh-1* (isocitrate dehydrogenase), *Adh-1* (alcohol dehydrogenase), *Adh-2*, *Fdh-1* (formate dehydrogenase), *Got-2* (glutamate oxaloacetate transaminase), *Got-3*, *Pgi-1* (phosphoglucose isomerase), *Pgm-1* (phosphoglutomutase) and *Pgm-2*. Enzymes were resolved on 12 % potato starch gels (StarchArt) using pH 7.0 (*Aco-1*, *Aco-2*, *6-Pgdh-1*, *6-Pgdh-2*, *6-Pgdh-3* and *Idh-1*) or pH 7.8 (*Adh-1*, *Adh-2*, *Fdh-1*, *Got-2*, *Got-3*, *Pgi-1*, *Pgm-1* and *Pgm-2*) buffer systems. Alleles (unique banding patterns) were recorded as differences in migration (in mm) relative to the tomato (*S. lycopersicum*) reference on the same gel. Control samples *S. pennellii* (LA716) and F_1 *S. lycopersicum* × *S. lycopersicoides* (LA3857) were employed to facilitate gel interpretation.

2.2.3.2 Microsatellites

Marker information including primer sequences of tomato EST-derived SSRs were obtained from the SOL Genomics Network database (http://www.sgn.cornell.edu; Table 3). Selection criteria were a) even genomic distribution and the avoidance of an overrepresentation of centromeric or telomeric regions, b) amplification quality, c) polymorphism degree and d) scorability (i.e. sufficient spatial separation of alleles on the polyacrylamide gel).

Thermocycling reactions including fluorescent labeling were performed as described above in the section on comparative mapping. All fragments were separated on 5.2 % polyacrylamide and visualized with a LiCor (4200) sequencing apparatus. *S. lycopersicum* 'VF36' was used as the control. SSR identity was confirmed by comparison of the observed and predicted (from SGN) fragment sizes for the control genotype. A total of 15 out of 37 tested markers were used for the population genotyping. Alleles were recorded according to their fragment sizes (in base pairs). Putative rare alleles were validated by repeating the SSR genotyping of selected individuals.

2.2.4 Data analysis

Prior to the computational analyses data were tested for agreement with Hardy-Weinberg expectations in order to avoid the inclusion of loci that were not exhibiting neutral genetic variation due to factors such as selection or scoring errors, and would violate the assumptions of most test statistics.

	Marker name	Chr	Position (S/L or cM)	LOD score	*S. lycopersicoides* *N*	*S. sitiens* *N*
Allozymes						
1.	*Aco-1*	12	L	NA	-	150
2.	*Aco-2*	7	L	NA	296	150
3.	*6-Pgdh-2*	12	NA	NA	208	149
4.	*6-Pgdh-3*	5	NA	NA	272	146
5.	*Idh-1*	1	S	NA	279	146
6.	*Adh-2*	6	L	NA	285	124
7.	*Fdh-1*	2	NA	NA	297	-
8.	*Got-2*	7	L	NA	298	142
9.	*Pgi-1*	12	NA	NA	298	148
10.	*Pgm-1*	12	S	NA	-	147
11.	*Pgm-2*	4	L	NA	298	150
	Mean				281.2	145.2
	Total				298	150
Microsatellites						
1.	SSR125	2	106.6	I	315	154
2.	SSR15	8	22.7	I	306	152
3.	SSR320	3	158.0	I	317	153
4.	SSR325	5	18.5	I	314	154
5.	SSR341	1	137.5	I	313	154
6.	SSR345	12	72.5	I	311	149
7.	SSR43	4	14.0	F	316	155
8.	SSR50	2	70.5	I	316	155
9.	SSR578	6	44.0	CF	317	-
10.	SSR599	9	103.0	F	315	155
11.	SSR74	10	74.0	I	315	155
12.	SSR76	11	38.0	I	316	155
13.	SSR80	11	20.0	I	316	155
14.	SSR85	10	55.0	I	316	153
15.	SSR98	1	31.8	CF	317	155
	Mean				314.7	153.7
	Total				317	155

Table 3. List of allozyme and microsatellite markers.

Chromosomal positions and LOD scores are from the tomato (*S. lycopersicum*) isozyme map (Tanksley 1985) and tomato-EXPEN 2000 (available at http://www.sgn.cornell.edu). S/L = short/long chromosome arm; *N* = number of individuals analyzed. Shaded areas indicate 'common' markers (i.e. markers used in the analysis of *S. lycopersicoides*, *S. sitiens* and the control species *S. chilense*).

A Marcov-chain algorithm (Guo and Thompson 1992) with 1000 dememorizations, 100 batches and 1000 iterations per batch was implemented to test the probability of deviations from Hardy-Weinberg equilibrium using the 'exact Hardy-Weinberg test' (H0 = random union of gametes; Haldane 1954; Weir 1990; Guo and Thompson 1992) and the single and multiple-samples version of the score test ('U test'; H1 = heterozygote deficiency or H1 = heterozygote excess) was employed (Raymond and Rousset 1995). The latter test statistic is the more powerful and was given priority in the analyses. Both tests were performed in GENEPOP on the web (http://genepop.curtin.edu.au; Raymond and Rousset 1995). The Bonferroni correction for multiple testings (Shaffer 1995) was applied in the statistical analyses, if appropriate.

The total number of alleles (*k*), the number of private alleles (i.e. alleles that were unique to one accession), allele frequencies, percent polymorphic sites (*P*), allelic richness per locus and sample (R_S) and over all samples (R_T), Nei's gene diversity (H_E), Wright's inbreeding coefficient per population (F_{IS}) and the population pairwise F_{ST} were computed in FSTAT ver. 2.9.3 (Goudet 2001). For each locus and population the number of alleles, the number of private alleles and the proportion of polymorphic loci (*P*; i.e. all sites with allele frequencies below 99 %) were recorded from the FSTAT ver. 2.9.3 (Goudet 2001) output. Allelic richness was calculated per locus and sample (R_S) and over all samples (R_T). The measure was independent of sample size; standardizations were realized via an adaptation of the rarefaction method (Hulbert 1971; El Mousadik and Petit 1996). All estimates were based on the smallest number of individuals typed at a given locus within a sample. The inbreeding coefficient per population (F_{IS}) was estimated according to Nei (1987; i.e. not requiring the different weighting to adjust for different population sizes). The inbreeding coefficient per locus and the population pairwise F_{ST} were calculated as Weir's & Cockerham's (1984) unbiased estimators Smallf (*f*) and theta (θ), respectively which account for small and unequal population sizes.

Significant deviations of the inbreeding coefficient per population (F_{IS}) from the null hypothesis (i.e. panmixia) were tested with randomizations at the 5 % level (i.e. randomizing alleles among individuals within samples and multi-locus genotypes between two samples, respectively). Significance of the population pairwise F_{ST} was evaluated with the log-likelihood statistic G (Goudet et al. 1996).

Locus-specific inbreeding coefficients (F_{IS} , F_{ST} and F_{IT}), Nei's gene diversity (Nei 1987) as total (H_T) and mean expected heterozygosity at each locus (H_E), as well as the mean observed heterozygosity (H_O) per locus and per population were estimated using the software program ARLEQUIN ver. 3.11; Excoffier et al. 2005). Significance of the inbreeding coefficients (F_{IS} and F_{IT}) and the fixation index (F_{ST}) per locus was tested via non-parametric permutations. The

hypothesis of isolation by distance (IBD) and other matrix-based correlations were tested in the Mantel test option in ARLEQUIN using 10000 permutations (Mantel 1967; Smouse et al. 1986). For IBD tests transformations of Weir's & Cockerham's (1984) unbiased estimator θ via the formula $F_{ST}/(1- F_{ST})$ and of geographic distances into ln-distances were obtained in GENEPOP, following the method described by Rousset (1997). A single negative pairwise F_{ST} value in the *S. lycopersicoides* allozyme dataset was set to 0.0001 in order to perform the transformation.

A conventional *t*-test and the nonparametric Mann-Whitney test were implemented to test for significant differences between allozyme and microsatellite diversity estimates. In a reduced, common dataset (i.e. composed of only those individuals that were tested with both marker systems) pairs of global population and locus estimates were compared using either population or locus means as replicates. The correlation of the two datasets was investigated via a Mantel test on the pairwise F_{ST} (θ) matrices. Spearman's rank test was used to evaluate correlations of various diversity estimates and allele frequencies with geographic components and population size estimates. The statistical tests were conducted in STATISTICA ver. 6.0.

The amount of gene flow between populations (*Nm*) and the average frequency of private alleles were estimated in GENEPOP. Gene flow was calculated according to Barton and Slatkin's parameter (Barton and Slatkin 1986) which is based on the distribution of rare alleles assuming the island model of migration. Wright's traditional equation $Nm \equiv (1 - F_{ST})/4\ F_{ST}$ (Wright 1951) was also included as a comparison. Populations were investigated for signatures of recent bottlenecks in BOTTLENECK ver. 1.2.02 (Piry et al. 1999) with 10000 iterative runs assuming the infinite alleles model (IAM) via two detection methods, in the absence of historical population information or reference population data. One is based on the principle that the number of alleles (in particular rare alleles) declines faster than the gene diversity after a bottleneck event, resulting in an excess of heterozygosity relative to that expected at mutation-drift equilibrium. Significant heterozygosity excess was tested using the Wilcoxon sign-rank test because of its robustness with few (less than 20) polymorphic loci (Cornuet and Luikart 1996; Luikart et al. 1997a; Piry et al. 1999). In addition, the graphical method described by Luikart et al. (1998) was implemented, designed to detect bottlenecks by visualizing mode-shift distortions (i.e. an over-proportional frequency reduction of rare alleles). This test, however, cannot be evaluated for its significance and requires sample sizes larger than those of the present study (more than 30 individuals) in order to return confident results. It was therefore viewed as inferior to the Wilcoxon sign-rank test and only included herein to provide additional information from a different perspective.

In order to investigate the partitioning of genetic variability among populations global and locus-by-locus analyses of molecular variance (AMOVA) were conducted in ARLEQUIN ver. 3.01 (Excoffier et al. 1992, 2005; Weir and Cockerham 1984).

Significance tests were based on non-parametric permutations (16000) on the covariance components associated with the structural levels (within and among individuals, within and among populations, and within and among groups of populations, if these were specified).

Cluster analyses were performed using the programs SEQBOOT, GENDIST, NEIGHBOR and CONSENSE from the PHYLIP software package version 3.6 (Felsenstein 2005). Allele frequency tables were generated in CONVERT ver. 1.31 (Glaubitz 2004). Three genetic distances were computed: Reynolds' distance (Reynolds et al. 1983), Cavalli-Sforza's chord distance (Cavalli-Sforza and Edwards 1967) and Nei's genetic distance (Nei 1972). The first two assume that genetic differences are caused by drift alone while the latter attributes a role to the effects of both mutation and drift in generating genetic differences. Tree reconstruction was achieved by implementation of the UPGMA (Sokal and Sneath 1963) and the neighbor-joining (Saitou and Nei 1987) method. Bootstrapping was carried out over 10000 replicates. Strong branch support was indicated by bootstrap values above 70 %. A consensus tree was built according to the extended majority rule and rooted with the respective sister species as outgroup. Phylogenetic trees were printed in TREEVIEW ver. 1.6.6 (Page 1996). Principle component analyses (PCA) were conducted on Weir and Cockerham's pairwise θ using the statistics software program STATISTICA ver. 6.0. Individuals were assigned to populations via a model-based, clustering algorithm provided by STRUCTURE ver. 2.2 (Pritchard et al. 2000; Falush et al. 2003, 2007). A burnin and run length of 10000 steps each was used. The admixture model, which assumes that individuals share portions of their genome with individuals of other populations (i.e. exhibit mixed ancestry) and the correlation model, which assumes that allele frequencies among individuals are correlated because they have derived via drift from a common ancestor, were chosen to optimally analyze the underlying data. No prior population information was added to the computations. The true number of populations was estimated from the posterior probability of the log-likelihood distribution and its second order rate of change as described by Pritchard et al. (2000) and Evanno et al. (2005), respectively. Geographic maps were generated in DIVA-GIS ver. 5.4 (Hijmans et al. 2001).

3. Results

3.1 Comparative mapping analysis

A comparative genetic linkage map based on an interspecific F_2 was constructed and analyzed in comparison to other solanaceous genomes. First, the significant aspects regarding the mapping population (e.g. generation and fertility) are highlighted. Next, results from the genotyping procedure, such as marker numbers, amplification and polymorphism rates as well as segregation distortion are reported. The comparative genetic linkage map of the *S. ochranthum* and *S. juglandifolium* genome vis-à-vis to that of tomato (*S. lycopersicum*) is presented, differences in map length are investigated and the degree of collinearity is evaluated.

3.1.1 Generation of the mapping population

Cross-incompatibility between the parental species *S. ochranthum* and *S. juglandifolium* was overcome by embryo rescue. The two F_1 plants closely resembled each other and were intermediate between the two parents. A 'pseudo-F_2' population comprising 66 plants was generated by intercrossing of two F_1 plants in order to circumvent self-incompatibility reactions. The genotype of F_1 and F_2 plants was confirmed by segregation analysis using molecular markers. Further evidence of hybridity and normal segregation was that F_2 plants varied for parental morphological characteristics, such as leaf size and dimensions (length × width), surface texture (smooth vs. rugose), degree of pubescence and number of pseudostipules (data not shown).

3.1.2 Pollen fertility and chromosome pairing

Pollen fertility averaged 38 % in the two F_1 hybrids (39 % in 01L5288-1 and 37 % in 01L5311-1). The majority of the chromosomes in the meiotic pollen cell in the F_1 hybrid 01L5288-1 paired as bivalents (5 - 8 bivalents/cell) and several as univalents (3 - 6 univalents/cell). Six out of eight cells contained one trivalent. Only one cell contained two, and another no trivalents but instead a multivalent formation involving seven chromosomes.

3.1.3 Marker analysis

A genetic linkage map was constructed with 132 markers (Appendix 2); the map comprised 96 (73 %) CAPS, 19 (14 %) RFLPs and 17 (13 %) microsatellites. The majority (51 %) were COSII markers, followed by 24 % TG probes, 13 % SSRs, 11 % COS markers and 2 % CT probes. Average marker spacing was 6 cM between markers, ranging from 0 to 31.7 cM. Of the 269 markers that were evaluated 51 % did not give satisfactory results. Marker failure was due to either failure of PCR amplification (CAPS and microsatellites), lack of polymorphisms (all marker types)

or complications during the F_2 screen, e.g. amplification problems or allelic polyploidy (all marker types). Among the different marker types used as CAPS COSII markers yielded the highest number of informative loci (62 %), followed by TG probes (35 %) and COSI loci (19 %).

3.1.4 PCR amplification

Of all PCR-based markers 76 % (204 out of a total of 269) yielded single amplicons. PCR amplification was most successful for COSII markers (94 %; 101 of the 108 tested primer pairs generated single bands) and similar for 'TG' or 'CT' probe derived markers (50 %; 31 out of 62) and COSI markers (42 %; 45 out of 107). Among SSR markers 77 % (27 out of 35) yielded single fragments. Amplicon sizes were either equal to or deviated only slightly from those predicted for tomato.

3.1.5 Polymorphism rate

RFLPs. Sixty out of 289 (21 %) tested probe/restriction enzyme combinations were polymorphic. Polymorphism rates ranged from 0 % to 80 % for individual probes. Among restriction enzymes *Eco*RV and *Hae*III yielded the highest (both 35 %) and *Alu*I the lowest (3 %) polymorphism rate.

CAPS. A total of 3.5 % of the markers that yielded single fragments were polymorphic with respect to their amplicon sizes, all of those were COSII markers. A total of 2052 restriction digestion reactions were carried out. Overall polymorphism rate was 12 % with the highest (17 %) among COSI markers followed by 'TG' and COSII sequences which showed similar polymorphism rates (14 % and 13 %, respectively). Polymorphism rates were not corrected for fragment length, therefore those of 'TG' probes, which were generally shorter (data not shown), were likely underestimated.

SSRs. Polymorphism yield among SSRs that amplified as single fragments was 63 % (17 out of 27).

3.1.6 Segregation distortion

The F_1 and F_2 state of the plants was confirmed by genotypic analysis; this is 100 % heterozygosity in the F_1 and presence of all 3 genotypes (J/J, J/O and O/O) in each F_2 plant. Overall, genome-wide segregation matched the expected 1:2:1 ratio. Goodness-of-fit statistic detected significant segregation distortion at a confidence level of $P < 0.05$ at one third (32 %) of all loci, a 6-fold increase of the number of loci expected to be affected by chance alone (Figure 2, Appendix 3). Proximal and distal chromosome positions were affected equally. A total of 12 putative

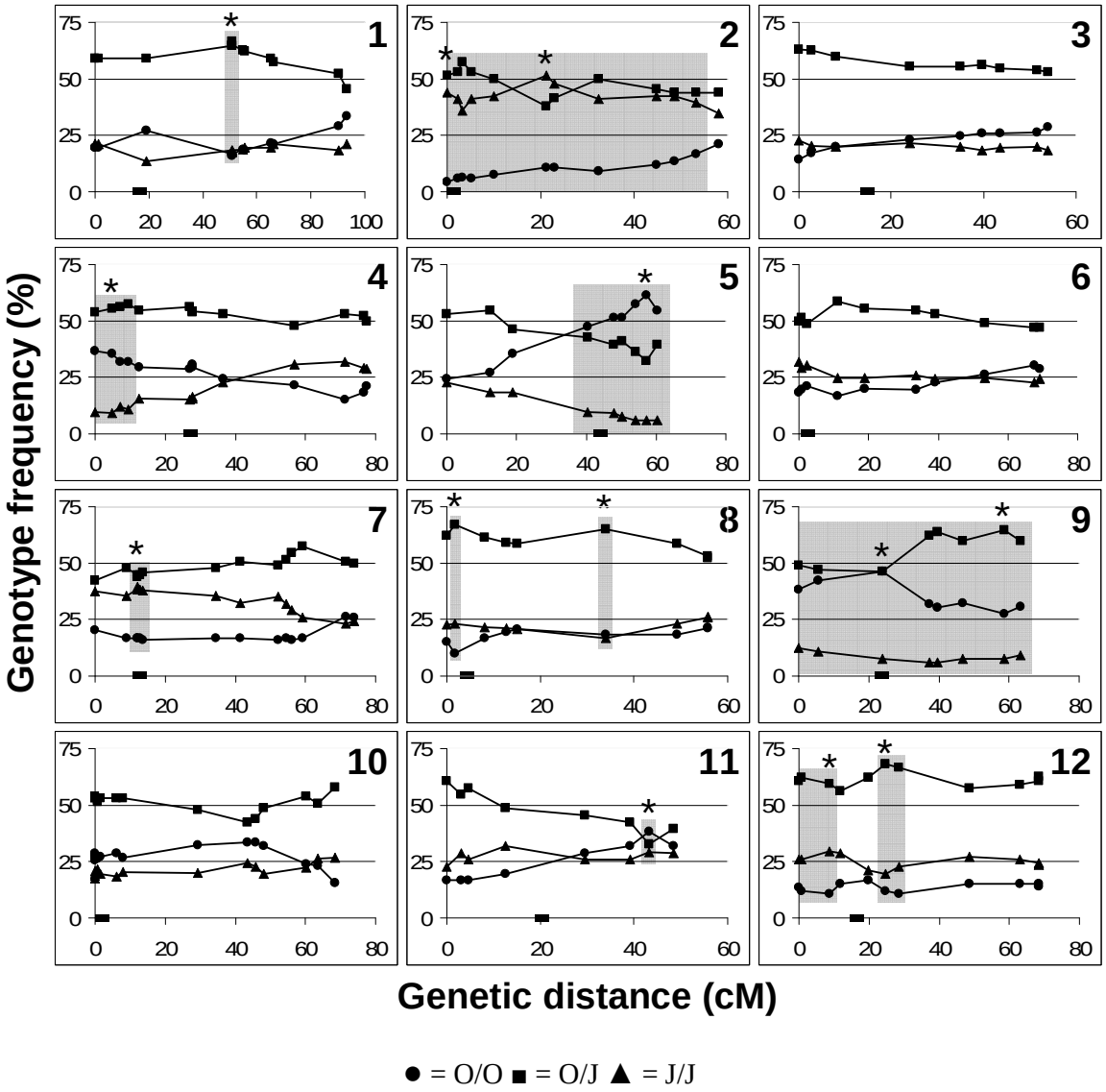


Figure 2. Genotype frequencies across the twelve chromosomes of F_2 *S. ochranthum* × *S. juglandifolium.*

Shaded areas indicate markers showing a significant segregation distortion ($P < 0.05$). Markers associated with the strongest deviation in a region were marked as segregation distorter loci (*). The expected frequencies were 50 % for the heterozygotes and 25 % for the homozygotes, indicated by horizontal lines. The X-axis shows the genetic distance (cM) from the distal short arm to the distal long arm end of each chromosome. Black boxes show putative centromere positions.

segregation distorter loci were observed on nine of the 12 chromosomes. On two chromosomes (chr. 2 and 9) the deviation extended nearly across the entire length of the chromosomes. Segregation distortion was most severe ($P < 0.0001$) on three chromosomes: at two loci on chr. 2, six loci on chr. 5 and one locus on chr. 9. One putative segregation distorter locus was located at the

upper end of chr. 2 at a position predicted for the centromere (TG608; *sd2.1*), and a second towards the middle of the chromosome (C2_At4g20410 and C2_At4g30930; *sd2.2*) where the *S. jugl.* homozygote reached > 50 % of the genotypic distribution at the expense of *S. ochr.* homozygotes. All loci on this chromosome were distorted in the same fashion. The bias decreased towards the end of the long arm and was no longer significant at the most distal locus (C2_At4g37280).

The reverse situation was found on chr. 5. An excess of *S. ochr.* homozygotes was accompanied by a deficiency of *S. jugl.* homozygotes and a less pronounced decline of heterozygotes ($P < 0.0001$). The distortion peaked around T1584 (*sd5.1*). Prominent was the partition of markers into those with highly significant segregation distortion and non-significant segregation distortion on this chromosome. The sharp drop occurred between markers C2_At4g24830 and SSR115 and may therefore coincide with the centromeric position, which was predicted to lie in proximity of C2_At4g24830, albeit on its 'south' side.

An over-representation of *S. ochr.* homozygotes indicated the presence of another segregation distorter locus of high significance ($P < 0.0001$), colocalizing with the centromere on chr. 9 (TG291; *sd9.1*). While *S. jugl.* homozygotes were suppressed throughout the entire range of chr. 9, ratios for the heterozygous state were as expected on 9S but exceeded normal values on 9L. An increase of heterozygotes accompanied by a decline of *S. ochr.* homozygotes was observed across the entire length of the long arm of chr. 9 and pointed to a second putative segregation distorter locus around C2_At3g24050 (*sd9.2*).

Further putative segregation distorter loci with weaker effects ($P < 0.05 \geq 0.001$) were observed at the following locations: On chr. 1 around C2_At2g45620 the heterozygous state was favored over both homozygous genotypes (*sd1.1*). The two linkage groups that represented chr. 1 (linkage groups are depicted in Figure 3) showed similar segregation patterns with the exception of a (non-significant) drop of *S. jugl.* homozygotes and an increase of *S. ochr.* homozygotes at the end of the shorter linkage group in close proximity to the predicted centromere. The distal end of the short arm of chr. 4 showed an excess of *S. ochr.* homozygotes associated with a reduction of *S. jugl.* homozygotes (SSR43; *sd4.1*). *S. jugl.* homozygotes outnumbered *S. ochr.* homozygotes at the centromeric region on chr. 7 (C2_At4g33250; *sd7.1*). The number of heterozygous genotypes was also slightly lower than expected, but the reduction of *S. ochr.* homozygotes was more pronounced. Two putative distorter loci were found on chr. 8: the first was located at the distal end of the short arm (C2_At5g46630; *sd8.1*) and the second in the middle of the long arm (TG510; *sd8.2*). At both loci the heterozygous state was favored, accompanied by a reduction of *S. ochr.* homozygotes at the first and a reduction of both homozygotes at the second locus. The same pattern was evident as a trend at adjacent loci. On chr. 11 a significant reduction of heterozygotes was observed only at

T0142 while the remainder of loci segregated in compliance with the expected Mendelian ratios. This, combined with the fact that a relatively high number of data were missing at this locus (16.7 %), suggested that the deviation could be an artifact and not caused by the presence of a distorter locus. Finally, two additional putative segregation distorter loci were located on chr. 12. The first was centered on the short arm around C2_At5g19690 (*sd12.1*) and the second on the long arm around TG394 (*sd12.2*). The first locus showed an increase of heterozygotes and, to a lesser degree, *S. jugl.* homozygotes at the expense of *S. ochr.* homozygotes. At the second site only the heterozygote state was favored and the number of *S. ochr.* homozygotes was reduced. The same patterns were apparent as trends at adjacent loci.

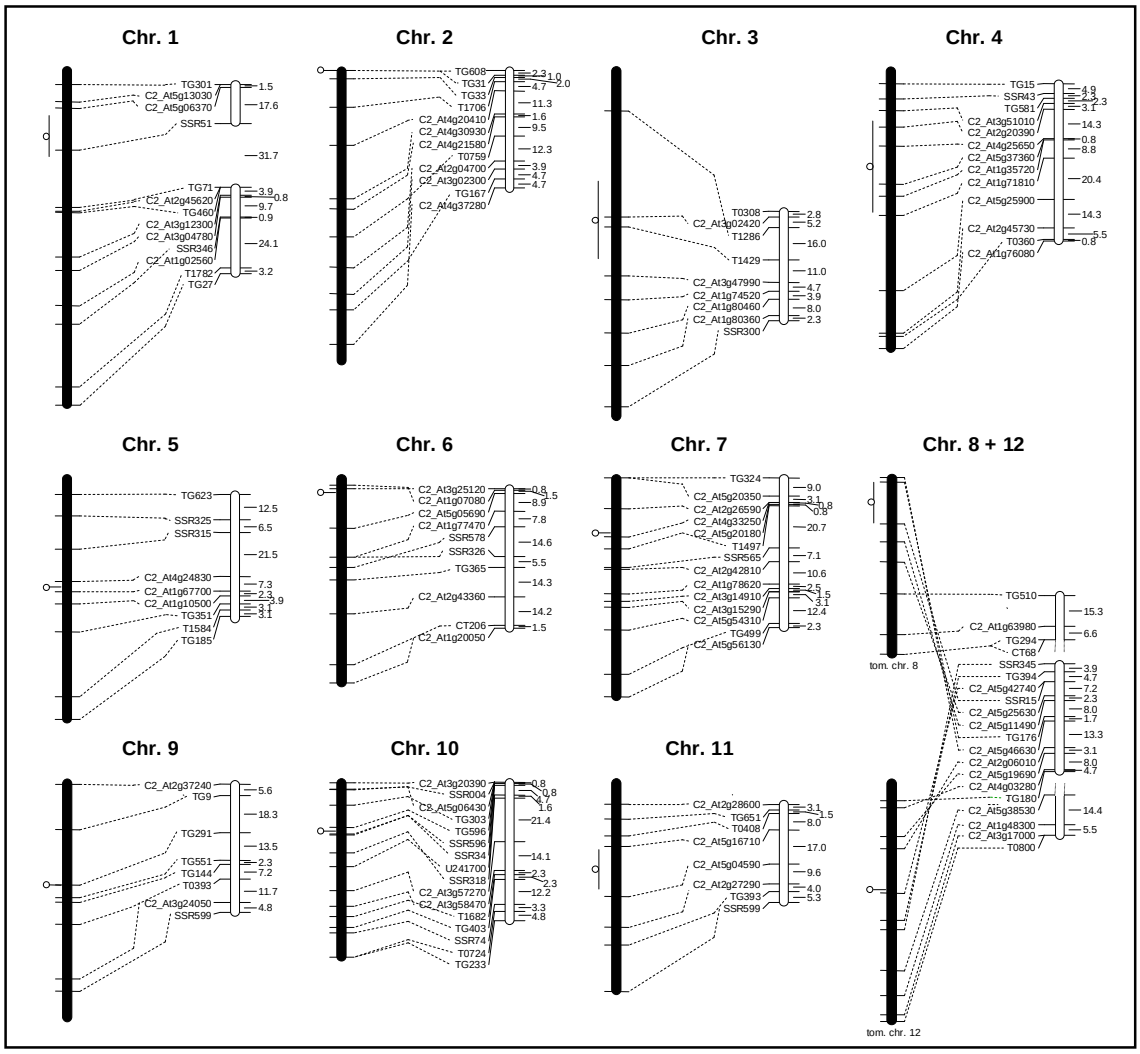
Chr. 8 and 12 displayed similar patterns of segregation distortion. Along the entire ranges of both chromosomes the number of heterozygotes was above, and the number of the *S. ochr.* homozygotes below the expected values. Segregation patterns in the translocation version (see next paragraph) of these chromosomes therefore corresponded to those of the non-translocated counterparts.

3.1.7 Comparative genetic inkage map

3.1.7.1 Linkage groups

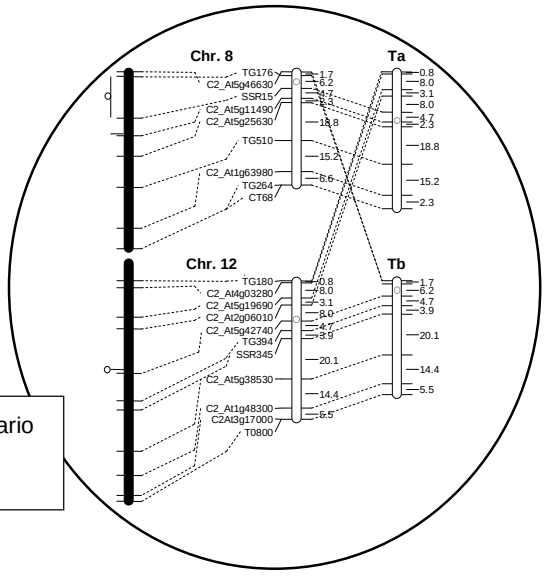
The 132 markers on the *S. ochranthum* × *S. juglandifolium* linkage map were distributed over 12 linkage groups (LOD ≥ 4, recombination fraction (RF) ≤ 0.3), which corresponded to the 12 chromosomes reported for tomato with the following exceptions (Figure 3): Tomato chr. 1 corresponded to two linkage groups on the *S. ochranthum* × *S. juglandifolium* linkage map, which were joined by a LOD of 1.98 and a RF of 0.28 and did not show linkage to any other linkage group. Tomato chr. 5 and 9 were connected to a single large linkage group within which both corresponding tomato chromosomes formed clearly delimited subgroups (i.e. loci order of the two subgroups was unaltered), suggesting a spurious association. The subgroups were split manually by raising the stringency to LOD = 6. Tomato chr. 8 and 12 also emerged as a single linkage group which was almost twice as long as the average linkage group on the *S. ochr.* × *S. jugl.* linkage map, thus appearing as a spurious fusion product. However, marker order was ambiguous along the merged linkage groups and higher stringencies did not result in a division into two balanced chromosomes but revealed the presence of three linkage subgroups (with total lengths of 22 cM, 53 cM and 20 cM), each with clearly determined marker order.

The subgroup (C2_At5g38530-T0800) that corresponded to the distal portion of tomato chromosome 12L disassociated at LOD = 4.9, the one that mapped to tomato chr. 8L (TG510 -



S. ochranthum x *S. juglandifolium*

Tomato-EXPEN 2000
S. lycopersicum x *S. pennellii*



Putative translocation scenario for *S. ochranthum* and *S. juglandifolium*

Figure 3. Comparative genetic linkage map of F_2 *S. ochranthum* × *S. juglandifolium*.

Positions of corresponding markers are shown on the reference map of tomato (tomato-EXPEN 2000) from (http://www.sgn.cornell.edu). Common markers are joined by dotted lines. Circles indicate putative centromere positions (Pillen et al. 1996). Map distances are in Kosambi centi Morgans (cM). The fused linkage group of tomato chromosome 8 + 12 is shown as its three linkage subgroups. Ta and Tb indicate the chromosome pair resulting from a putative reciprocal translocation in either *S. ochranthum* or *S. juglandifolium*.

CT68) disintegrated at LOD = 5.4, leaving a "core" group of 12 tightly linked markers that maintained its integrity up to LOD = 13. Within this subgroup linkage was strongest between markers SSR15 and C2_At42740 (LOD = 27, RF = 0). A reciprocal translocation in one of the parents of a mapping population is known to create pseudolinkage between markers in proximity to the interchange breakpoints (Burnham 1991). On the tomato map SSR15 and C2_At42740 are located near the centromere, a region with high propensity for chromosome breakage in many species (Tanksley et al. 1992; Moore et al. 1997). Assuming this to be the interchange point, linkage groups were rearranged manually resulting in two putative, balanced chromosome pairs: a) tomato chr. 8 and 12 and b) the translocation pair Ta and Tb. All four were similar in length to the rest of the linkage groups. Chromosome pairs with the translocated and the tomato-like configurations produced near equal amounts of overall mapping units (124.0 cM and 124.2 cM, respectively). Within each group computational analysis identified unambiguous marker positions that were also supported by LOD tables generated in MAPMAKER.

3.1.7.2 Map length

When the length of the fused linkage group (8 + 12) was considered as sum of its three subgroups (i.e. 22 + 53 + 20 cM) the total map distance covered by the *S. ochr.* × *S. jugl.* linkage map spanned 761 cM, which represented a 44 % reduction relative to the distance covered by the same markers on the tomato reference map (1363 cM). Under consideration of the putative translocation and the tomato-like configuration scenarios (i.e. after manual splitting of the large linkage group) the total map length amounted to 790.4 cM and 790.2 cM, respectively, which constituted a reduction of 42 % compared to the tomato reference map tomato-EXPEN 2000.

Linkage groups were heterogeneous with respect to map expansions and reductions, but all 12 displayed a net shrinkage compared to the tomato counterparts. Distal, proximal and intermediate regions were affected similarly. Clustering of markers around (putative) centromeres was observed on J2, J4, J6, J7 and J10. Individual linkage groups varied between 63 % (J3) and 21 % (J10) size reduction relative to their tomato counterparts. No linkage group showed a net map expansion

across its entire range, but small regions of map expansions were observed on all but one linkage group (J2). Average marker spacing was 6 cM.

The strongest reduction in mapping distance was observed for linkage group 3 (63 %), mostly due to the nearly complete omission of the short arm (98 % reduction) as a result of inverted positions of two markers, relative to the tomato EXPEN 2000 map, accompanied by a 10-fold size reduction: C2_At3g02420 and T1286 are separated by over 50 cM on the reference map but by only 5 cM on F_2 *S. ochr.* × *S. jugl.*. Severe over-all length reductions were also observed on J2 (57 %), J11 (48 %), J5 (46 %) and J1 (41 %). In four cases recombination suppression was concentrated on one of the two chromosomal arms: J3S (98 %; as mentioned above), J10S (96 %), J1S (87 %) and J5L (76 %).

3.1.7.3 Collinerarity with the tomato genome

Overall marker order was highly conserved among F_2 *S. ochr.* × *S. jugl.* and tomato-EXPEN 2000. Six of the markers used in this study (TG71, TG608, TG342, TG291, TG144 and TG596) were not mapped on tomato-EXPEN 2000 but on tomato-EXPEN 1992. Values extrapolated from adjacent markers with known locations on both tomato-EXPEN 1992 and tomato-EXPEN 2000 were used as a proxy for their reference positions. No duplicated loci were detected as only markers with single banding patterns (RFLPs) or single amplicons (CAPS, SSRs) were included in the study.

Inverted marker positions were observed among seven marker pairs on 7 out of 12 linkage groups, these were: SSR346 and C2_At1g02560 on J1L, TG31 and TG33 in the distal region of J2S, C2_At3g02420 and T1286 on J3S, C2_At3g25120 and C2_At1g07080 on J6S, SSR578 and SSR326 in the centromeric region of J6, C2_At4g33250 and C2_At5g20180 in the proximal region of J7L, TG176 and C2_At5g46630 at the distal end of J8S, and finally SSR596 and SSR034, and U241700 and SSR318 on J10L. In order to infer the correct marker order at the 5 % confidence level markers needed to be spaced at > 3 cM, a criteria that was met in four of the cases (J3S, in the centromeric region of J6, on J8S and on J10L).

Two markers mapped to different chromosomes: TG581, located on tomato 6L was placed on 4S in *S. ochr.* × *S. jugl.* . T0308, on 10L in tomato, mapped to the distal end of 3S on *S. ochr.* × *S. jugl.*, demarcating the most distal position of the chromosome arm that appeared inverted and severely truncated as a result of the above mentioned inversion. The presence of a whole-arm paracentric inversion on J3S may have represented an artifact as tomato 3S was represented by only two markers in this study. Efforts to increase that number were unsuccessful; no markers with polymorphisms among the parental species were found for that area. A similar situation was found

on J6S: the only two markers (C2_At3g25120 and C2_At1g07080) on this chromosome arm revealed inverted positions, calling for a whole-arm paracentric inversion scenario. However, the two loci are located at very close positions on both maps and therefore the inversion could well be an artifact.

3.2 Genetic diversity analysis

To evaluate levels and partitioning of genetic diversity across the distributional range 14 *S. lycopersicoides* and seven *S. sitiens* wild populations (Table 2), represented by a total of 333 and 195 plants, respectively, were analyzed using allozymes and microsatellites (Table 3). Results are presented in the following order: First, the conformity of loci and population estimates to Hardy-Weinberg expectations is reported. Second, levels of genetic diversity (allele numbers and various descriptive diversity parameters) and recent population bottlenecks are highlighted, including statistical correlations among diversity estimates (statistical differences among allozyme and microsatellite data and overall correlations among parameters). Third, population genetic structure as revealed by partitioning of genetic diversity via AMOVA analysis is described. Spatial organization of diversity estimates, such as patterns of isolation by distance and gene flow estimates are illustrated. Forth, population genetic clustering identified via three routes: phylogenetic tree construction, principle component analysis and a model-based structure analysis is presented, and subsequently compared to geographic patterns. Finally, for each population cluster levels and partitioning of genetic diversity are reported.

3.2.1 Hardy-Weinberg disequilibria

3.2.1.1 Allozymes

Thirteen out of a total of 16 loci in eight allozyme systems were polymorphic in *S. lycopersicoides*. Four of these were excluded from further analyses: *6-Pgdh-1* did not fulfill the criteria of > 1 mm band separation, and *Aco-1*, *Got-3* and *Pgm-1* showed significant ($P < 0.042$) heterozygote deficiencies at > 50 % of the polymorphic locus/population combinations according to both test statistics, the ‘exact Hardy-Weinberg test’ and the ‘U test’. This left a total of nine loci that were informative and met the above criteria (Table 3).

Thirteen out of 16 loci were polymorphic in *S. sitiens*. As in *S. lycopersicoides* not all alleles at *6-Pgdh-1* could be resolved at > 1 mm distance, and the locus was removed from the analyses. *Adh-1* and *Got-3* showed significant ($P < 0.008$) deviations from Hardy-Weinberg expectations according

to both test statistics in form of heterozygote deficiencies at ≥ 50 % of the polymorphic loci, leaving a total of 10 informative loci for the subsequent analysis.

After removal of loci that showed scoring uncertainties or deviated from Hardy-Weinberg disequilibria none of the *S. lycopersicoides* and only two of the *S. sitiens* populations tested significantly positive for departures from expected Hardy-Weinberg ratios. Cere (LA4113) was determined as deviant by both tests, exhibiting a significant excess of heterozygotes ($P < 0.006$). Paqui (LA4116) was deviating according to the 'exact Hardy-Weinberg test' only and showed both heterozygote excess and deficiency at the underlying loci ($P < 0.005$).

3.2.1.2 Microsatellites

As allelic sizes within loci did not vary in a strict stepwise fashion, the infinite alleles model (IAM; Kimura and Crow 1964) rather than the stepwise mutation model (SSM; Otha and Kimura 1973) was assumed to more accurately represent the microsatellite mutation mode.

At eight of the 37 SSR loci tested no PCR fragment was obtained, another five loci were monomorphic and nine exhibited ambiguous banding patterns in either of the two species (in most cases loci showed similar behavior in both species), leaving a total of 15 informative markers for the *S. lycopersicoides* and 14 for the *S. sitiens* analyses (Table 3).

In *S. lycopersicoides* none of the fifteen loci showed more than a few significant ($P < 0.036$) deviations in individual populations. The 'exact Hardy-Weinberg test' detected two populations (Pachica - LA4130 and Lluta - LA4320) that deviated significantly ($P < 0.003$) from the expected equilibrium, mostly due to heterozygote deficiencies, the multi-sample version of the 'U-test' one population (LA1966; $P < 0.003$), showing heterozygote deficiencies.

In *S. sitiens* significant ($P < 0.007$) departures from Hardy-Weinberg expectations were restricted to a small portion of locus/population combinations, therefore none of the markers was removed from the dataset. In one population (Escondida - LA4105) a significant ($P < 0.004$) heterozygote deficiency was evident according to both tests statistics.

3.2.2 Genetic diversity

3.2.2.1 Number of alleles, number of private alleles and percentage of shared alleles

S. lycopersicoides. Overall genetic variability was lower at allozyme than at microsatellite loci. A total of 25 alleles were found at nine allozyme loci among 298 individuals in 12 populations of *S. lycopersicoides* (Table 4). The average population sample size was 24.8. The highest number of alleles (5) was counted at locus *Pgm-2* with an average of 2.8 per population (Appendix 4A). Among populations the highest number of alleles at all loci (16) was observed in Camiña (LA4123),

	Population name	Accession number	N_{wild}	N		k		$k_{private}$		P		R_S		H_E		H_O		F_{IS}		θ	
				A	M	A	M	A	M	A	M	A	M	A	M	A	M	A	M	A	M
1.	Aricota #1	LA4018	50	25	25	13	44	0	1	44.4	86.7	1.32	2.76	0.089	0.344	0.076	0.360	0.146	-0.045	0.193	0.342
2.	Aricota #2	LA2387	NA	20	20	14	39	0	0	55.6	80.0	1.46	2.58	0.137	0.407	0.142	0.447	-0.037	-0.102	0.101	0.290
3.	Chupapalca	LA1964	50	19	19	14	42	0	0	44.4	93.3	1.51	2.79	0.136	0.417	0.137	0.379	-0.008	0.094	0.102	0.270
4.	Palca	LA1966	60	25	25	15	54	0	2	44.4	86.7	1.59	3.45	0.171	0.462	0.194	0.394	-0.136	0.149	0.116	0.270
5.	Putre #1	LA2781	75	25	24	12	46	0	1	33.3	86.7	1.31	2.98	0.110	0.351	0.084	0.385	0.242	-0.023	0.254	0.294
6.	Putre #2	LA2777	75	25	25	13	45	0	1	33.3	86.7	1.43	2.87	0.148	0.370	0.131	0.396	0.120	0.074	0.111	0.248
7.	Perquejeque	LA2776	100	28	23	14	49	1	1	44.4	93.3	1.45	3.21	0.131	0.461	0.121	0.506	0.083	-0.100	0.088	0.199
8.	Zapahuira	LA2772	50	29	25	14	54	0	3	44.4	100	1.37	3.49	0.128	0.476	0.126	0.472	0.014	0.009	0.191	0.240
9.	Lluta	LA4320	20	-	22	-	53	-	1	-	100	-	3.47	-	0.484	-	0.530	-	-0.098	-	0.225
10.	Pachica	LA4130	20	-	11	-	39	-	3	-	73.3	-	NA	-	0.371	-	0.389	-	-0.052	-	0.268
11.	Esquina	LA4131	68	26	24	12	43	0	0	22.2	86.7	1.26	2.78	0.063	0.380	0.050	0.385	0.212	-0.018	0.345	0.284
12.	Nama	LA4126	50	25	25	14	43	1	0	33.3	100.	1.41	2.77	0.080	0.428	0.067	0.453	0.177	-0.059	0.185	0.253
13.	Camiña	LA4123	50	24	24	16	43	1	0	55.6	93.3	1.63	2.81	0.154	0.465	0.146	0.447	0.050	0.039	0.112	0.222
14.	Moquella	LA2730	NA	27	25	15	50	1	0	55.6	100	1.31	3.15	0.052	0.486	0.055	0.523	-0.058	-0.078	0.229	0.252
		Mean	55.7	24.8	22.6	13.8	46.0	0.33	0.93	42.6	90.5	1.42	3.01	0.117	0.422	0.111	0.433	0.067	-0.015	0.169	0.261
		SD		(2.9)	(3.9)	(1.2)	(5.2)			(10.4)	(8.2)	(0.12)	(0.31)	(0.038)	(0.051)	(0.044)	(0.057)	(0.117)	0.079	(0.079)	(0.036)
		Total		298	317	25	90	4	13												

Table 4. Genetic diversity in populations of *S. lycopersicoides*.

N_{wild} = estimated population size in the wild; N = number of individuals; k = number of alleles per population; P = percentage of polymorphic sites; R_S = allelic richness per population; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient per population; θ = average pairwise genetic distance; A = allozyme analysis; M = microsatellite analysis. Standard deviations (SD) are reported in brackets. Shaded areas indicate significant values ($P \leq 0.05$ adjusted level).

and the average was 13.8 total alleles per population. Four private alleles (i.e. those that were not shared by other populations of the same species) were detected in *S. lycopersicoides*: in Perquejeque (LA2776; at *Pgi-1*), in Nama (LA4126; at *Fdh-1*), in Camiña (LA4123; at *Idh-1*), and in Moquella (LA2730; at *Pgm-2*). The average frequency of private alleles per population was 0.33.
Ninety SSR alleles were detected at 15 loci in 317 individuals of 14 *S. lycopersicoides* populations, ranging from 39 in Aricota #2 (LA2387) to 54 in Palca (LA1966) and Zapahuira (LA2772), with an average of 46 alleles per population. Individual loci had maxima of two (SSR578) to twelve (SSR15, SSR50, SSR341) alleles in *S. lycopersicoides* (mean = 6.0). A total of 13 private alleles were observed among populations of *S. lycopersicoides*: in Aricota #1 (LA4018; at SSR50), Palca (LA1966; at SSR50, SSR98), Putre #1 (LA2781; at SSR341), Putre #2 (LA2777; at SSR50), Perquejeque (LA2776; at SSR74), Zapahuira (LA2772; at SSR15, SSR341, SSR345), Lluta (LA4320; at SSR85) and Pachica (LA4130; three private alleles at SSR341). The average private allele frequency per population was 0.93.

S. sitiens. A total of 31 alleles were detected at 10 allozyme loci among 150 individuals in six *S. sitiens* populations (Table 5). The loci with the highest number of alleles (4) were *6-Pgdh*, *Adh-2* and *Pgm-1*. Loci showed on average 3.1 alleles among all populations (Appendix 4B). The population that contained the largest number of alleles (25) was Paqui (LA4116), and the population average was 21.2 alleles. Four private alleles were found among *S. sitiens* populations: two in Paqui (LA4116; at *Idh-1* and *Pgm-2*), and two in Escondida (LA4105; at *6-Pgdh-2* and *Idh-1*). The average private allele frequency per population was 0.67.
Sixty alleles were identified at 14 SSR loci among 155 individuals in seven *S. sitiens* populations, ranging from 31 in Carbonatera (LA4114) to 45 in Limón Verde (LA4112), with a mean of 36.3 alleles per population. Individual loci had two (SSR320, SSR599) to eight (SSR15) alleles in *S. sitiens* (mean = 4.3). Eight private SSR alleles were detected among *S. sitiens* populations: in Paqui (LA4116; at SSR43, SSR98), Limón Verde (LA4112; at SSR15, SSR345), San Juan (LA4111; at SSR85), and Escondida (LA4105; at SSR15, SSR50, at SSR80), leading to an average private allele frequency of 1.14 per population.

Comparison among species. In order to be able to compare the three species (including the control species *S. chilense*) directly, the population average for allele numbers and estimates of the main descriptive parameters were recalculated from a common set of markers, containing eight allozyme (*Aco-2*, *6-Pgdh-2*, *6-Pgdh-3*, *Idh-1*, *Adh-2*, *Got-2*, *Pgi-1* and *Pgm-2*) and 14 SSR loci

	Population name	Accession number	N_{wild}	N		k		$k_{private}$		P		R_S		H_E		H_O		F_{IS}		θ	
				A	M	A	M	A	M	A	M	A	M	A	M	A	M	A	M	A	M
15.	Paqui	LA4116	100	25	19	25	36	2	2	100	85.7	2.40	2.55	0.390	0.377	0.383	0.376	0.019	0.003	0.229	0.181
16.	Carbonatera	LA4114	35	27	22	21	31	0	0	60.0	78.6	2.07	2.16	0.299	0.289	0.272	0.289	0.089	-0.002	0.229	0.207
17.	Cere	LA4113	20	28	21	24	32	0	0	70.0	78.6	2.25	2.25	0.309	0.308	0.272	0.313	0.121	-0.017	0.247	0.186
18.	Limón Verde	LA4112	47	24	23	23	45	0	2	70.0	100	2.25	3.07	0.338	0.452	0.307	0.452	0.099	0.000	0.225	0.158
19.	Quimal	LA4331	17	-	25	-	37	-	0	-	85.7	-	2.54	-	0.376	-	0.379	-	-0.008	-	0.190
20.	San Juan	LA4111	70	22	20	17	36	0	1	60.0	85.7	1.70	2.55	0.228	0.345	0.204	0.358	0.108	-0.041	0.275	0.198
21.	Escondida	LA4105	20	24	25	17	37	2	3	60.0	71.4	1.64	2.61	0.163	0.371	0.145	0.324	0.113	0.128	0.476	0.304
		Mean	44.1	25	22.1	21.2	36.3	0.67	1.14	70.0	83.7	2.05	2.53	0.288	0.360	0.264	0.356	0.092	0.009	0.280	0.203
		SD		(2.2)	(2.3)	(3.5)	(4.5)			(15.5)	(9.0)	(0.32)	(0.29)	(0.061)	(0.053)	(0.082)	(0.054)	(0.037)	(0.055)	(0.098)	(0.047)
		Total		150	155	31	60	4	8												

Table 5. Genetic diversity in populations of *S. sitiens*.

N_{wild} = estimated population size in the wild; N = number of individuals; k = number of alleles per population; P = percentage of polymorphic sites; R_S = allelic richness per population; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient per population; θ = average pairwise genetic distance; A = allozyme analysis; M = microsatellite analysis. Standard deviations (SD) are reported in brackets. Shaded areas indicate significant values ($P \leq 0.05$ adjusted level).

(under exclusion of SSR578; Figure 4, Table 3). The results from the 'common' dataset did not diverge much from those based on the 'unique' dataset. The total number of alleles at allozyme loci was similar in both species (22 in *S. lycopersicoides* and 24 in *S. sitiens*), but due to the smaller number of populations in *S. sitiens* the population mean was higher in *S. sitiens* (15.7 vs. 11.8 in *S. lycopersicoides*). The locus average in *S. lycopersicoides* (1.5) was lower than that observed in *S. sitiens* (2.0). Forty percent of all alleles, 48 % of those found in *S. lycopersicoides* and 36 % of those in *S. sitiens* were shared by the two species. In addition, a putative null-allele at *Adh-2* was detected in both species, but because the identity could not be confirmed it was scored as missing data. In contrast to the allozyme analyses, fewer alleles were found among populations of *S. sitiens* than *S. lycopersicoides* using SSRs. At the 14 common loci, *S. lycopersicoides* showed almost 50 % more alleles (88 vs. 60) which translated into a 25 % increase per population (44.3 vs. 36.3 alleles). As a comparison, 61 alleles were found among the 23 individuals of the single *S. chilense* control population. Of the 98 alleles scored at the 14 common loci, 38 were unique to *S. lycopersicoides*, 10 to *S. sitiens*. The two species shared 58.0 % and 83.3 % of their alleles, respectively. *S. lycopersicoides* shared 36.4 % of the alleles with *S. chilense*, *S. sitiens* 45.0 %, or, reversely described, of the alleles found in *S. chilense*, 52.5 % were shared by *S. lycopersicoides*, 44.3 % by *S. sitiens*.

3.2.2.2 Descriptive diversity parameters in *S. lycopersicoides*

Proportion of polymorphic sites. The proportion of polymorphic sites (*P*) was 42.6 % in *S. lycopersicoides* (Table 4). The highest rates were observed at *Fdh-1* and *Aco-2* (100 %), the lowest at *Pgi-1* (8.3 %; Appendix 4A). Among populations Aricota #2 (LA2387), Camiña (LA4123) and Moquella (LA2730) showed the highest rates (55.6 %), Esquina (LA4131) the lowest (22.2 %). The mean *P* at 15 SSR loci in *S. lycopersicoides* was 90.5 %, ranging from 64.3 % at SSR125 to 100 % at SSR15, SSR325, SSR345, SSR50, SSR599, SSR74, SSR80 and SSR85. Among populations percentage rates varied from 73.3 % in Pachica (LA4130) to 100 % in Zapahuira (LA2772), Lluta (LA4320), Nama (LA4126) and Moquella (LA2730).

Allelic richness. Calculations of allozyme allelic richness independent of population size were based on the minimum sample size of eight diploid individuals in *S. lycopersicoides*. Allelic richness over all samples (R_T) ranged from 1.08 (*6-Pgdh-3*, *Pgi-1*) to 2.89 (*Aco-2*) with a mean of 1.54. Average R_S over populations reached from 1.26 (Esquina - LA4131) to 1.63 (Camiña - LA4123) with a mean of 1.42. At SSR loci calculations of allelic richness were based on a minimum sample size of 17 diploid individuals. Population Pachica (LA4130) was excluded from the dataset in order to avoid a downward bias caused by the exceptionally low sample size (N = 11

vs. mean N = 23.5 in the remainder of populations). Allelic richness over all populations (R_T) ranged from 2.00 (SSR578) to 7.88 (SSR15) with a mean of 4.43 per locus. Average R_S over loci spanned from 2.58 (Aricota #2 - LA2387) to 3.49 (Zapahuira - LA2772) with an average of 3.01.

Nei's gene diversity. Nei's total gene diversity (H_T) averaged 0.139 at allozymes, the mean over loci and populations, i.e. the expected heterozygosity (H_E) 0.117 (84 % of H_T). The least diverse loci were *Pgi-1* and *6-Pgdh-3* (H_E = 0.009), the most diverse locus *Aco-2* (H_E = 0.422). Among populations the genetic diversity estimates were smallest in Moquella (LA2730; H_E = 0.052) and largest in Palca (LA1966; H_E = 0.171). Overall observed heterozygosity across loci (H_O = 0.111) ranged from 0.009 at *Pgi-1* to 0.411 at *Aco-2* and among populations from 0.050 in Esquina (LA4131) to 0.194 in Palca (LA1966). At SSRs the total average gene diversity (H_T) was 0.569, the average gene diversity over loci and populations (H_E) 0.422 (74 % of H_T). H_E varied substantially among loci (from H_E = 0.140 at SSR578 to H_E = 0.612 at SSR15 and SSR80). The least diverse populations were Aricota #1 (LA4018; H_E = 0.344) and Putre #1 (LA2781; H_E = 0.351), the most diverse Moquella (LA2730; H_E = 0.486). Overall observed heterozygosity across loci (H_O) was 0.433. Estimates at individual loci ranged from 0.145 (SSR578) to 0.615 (SSR80) and among populations from 0.360 in Aricota #1 (LA4018) to 0.530 in Lluta (LA4320).

Inbreeding coefficient per population. In *S. lycopersicoides* the average F_{IS} per population at allozyme loci ranged from -0.136 in Palca (LA1966) to 0.242 in Putre #1 (LA2781) with an average of 0.067. None of the values were significantly different from zero ($P < 0.006$). F_{IS} estimates from microsatellite data were lower than those obtained from allozymes. The average value of the inbreeding coefficients across populations was close to zero (F_{IS} = -0.015). The population with the highest level of inbreeding was Palca (LA1966; F_{IS} = 0.149), the one with the lowest Aricota #2 (LA2387; F_{IS} = -0.102). The value of the former was significantly different from zero ($P < 0.0033$).

3.2.2.3 Descriptive diversity parameters in *S. sitiens*

Proportion of polymorphic sites. Seventy percent of all allozyme sites were polymorphic in *S. sitiens* (Table 5). The rates were highest (100 %) at the loci *6-Pgdh-2*, *6-Pgdh-3*, *Adh-2*, *Pgm-1* and in the population Paqui (LA4116; 100 %), lowest at the locus *Pgm-2* (16.7 %) and in the populations Carbonatera (LA4114; all 60 %), San Juan (LA4111) and Escondida (LA4105). Eighty-four percent of all sites were polymorphic at 14 SSR loci in *S. sitiens*, spanning from 28.6 % at SSR85 to 100 % at SSR125, SSR320, SSR325, SSR341, SSR345, SSR76 and SSR98. Population values ranged from 71.4 % in Escondida (LA4105) to 100 % in Limón Verde (LA4112; Appendix 4B).

Allelic richness. R_T over allozyme loci ranged from 1.49 (*Idh-1*) to 3.75 (*Adh-2*) with a mean of 2.64. R_S over populations spanned from 1.64 (Escondida - LA4105) to 2.40 (Paqui - LA4116) with a mean of 2.05. Values were based on a sample size of 14 diploid individuals. Allelic richness at individual SSR loci over all populations (R_T) ranged from 1.81 (SSR85) to 6.50 (SSR15) with a mean of 3.41. Average allelic richness over loci (R_S) reached from 2.16 in Carbonatera (LA4114) to 3.07 in Limón Verde (LA4112) with a mean of 2.53 (base number = 17 individuals).

Nei's gene diversity. The average total gene diversity (H_T) was 0.393 and the average gene diversity over loci and populations (H_E) was 0.288 (73 % of H_T) at allozyme loci. The smallest gene diversity was estimated for *Idh-1* (H_E = 0.038), the greatest for *Pgm-1* (H_E = 0.488). Among populations the smallest average gene diversity was observed in Escondida (LA4105; H_E = 0.163), the largest in Paqui (LA4116; H_E = 0.390). The average overall H_O (0.264) was smaller than H_E in *S. sitiens*. Observed heterozygosity was lowest at *Idh-1* (H_O = 0.028) and highest at *Pgm-1* (H_O = 0.441), ranging among populations from 0.145 in Escondida (LA4105) to 0.383 in Carbonatera (LA4114). The average H_T at SSRs was 0.448 and the mean H_E 0.360 (80 % of H_T). Again, variations were high among loci. The lowest level of gene diversity was observed at SSR85 (H_E = 0.058), the highest at SSR76 (H_E = 0.643). Among populations Carbonatera (LA4114) was the least diverse (H_E = 0.289), Limón Verde (LA4112) the most diverse (H_E = 0.452). H_O was similar to H_E (H_O = 0.356), among loci lowest at SSR85 (H_O = 0.048) and highest at SSR76 (H_O = 0.692) and among populations lowest in Carbonatera (LA4114; H_O = 0.289) and highest in Limón Verde (LA4112; H_O = 0.452).

Inbreeding coefficient per population. The population average F_{IS} at allozymes in *S. sitiens* was estimated 0.092, ranging from 0.019 in Paqui (LA4116) to 0.121 in Cere (LA4113). None of the results per population were significant ($P < 0.005$). The average degree of inbreeding among populations was F_{IS} = 0.009 at SSR loci. All populations showed F_{IS} values close to zero except Escondida (LA4105) where F_{IS} = 0.128. The values were not significant ($P < 0.0036$).

3.2.2.4 Descriptive diversity parameters: comparison among species

Proportion of polymorphic sites. Considering only the eight common allozyme loci, the average *P* (over all locus/population combinations) in *S. lycopersicoides* (34.4 %) was about half of that in *S. sitiens* (66.7 %; Figure 4). Comparisons at the 14 common SSR loci revealed that the over-all proportion of polymorphic sites was only slightly reduced in *S. sitiens* relative to *S. lyopersicoides* (*P* = 83.7 % vs. *P* = 92.4 %, respectively). The value in the latter was as high as in the single *S. chilense* population (*P* = 92.9 %).

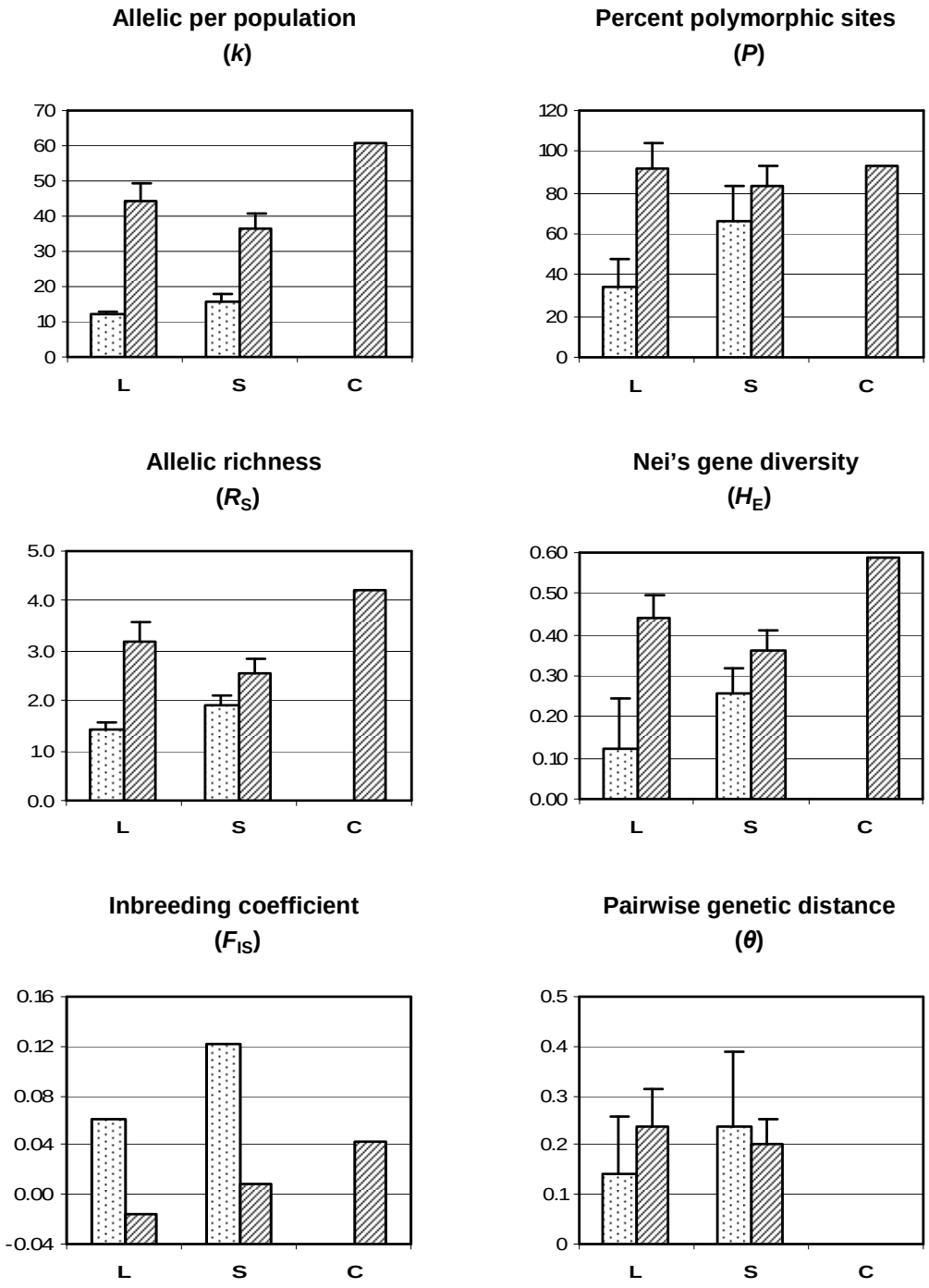


Figure 4. Comparison of population genetic diversity among *S. lycopersicoides*, *S. sitiens* and *S. chilense*.

L = *S. lycopersicoides*; S = *S. sitiens*; C = *S. chilense* (C). Values are based on a common set of markers (see Table 3) in twelve (allozyme) or 13 (SSR) *S. lycopersicoides* populations, 6 (allozyme) or 7 (SSR) *S. sitiens* populations and one *S. chilense* population (23 individuals). Dotted areas represent allozyme data, striped areas microsatellite data. k = number of alleles per population; P = percentage of polymorphic sites; R_S = allelic richness per population; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient per population; θ = average pairwise genetic distance.

Allelic richness. Based on a minimum sample size of 14 diploid individuals the average R_S was higher at the eight common allozyme loci in *S. sitiens* (1.90) than in *S. lycopersicoides* (1.43). Compared on the basis of a common set of loci with a base number of 17 individuals (again excluding *S. lycopersicoides* population Aricota #1 - LA4018) the average allelic richness per population (R_S) was highest in the single *S. chilense* population (R_S = 4.20); almost twice as high as in *S. sitiens* (R_S = 2.53) and intermediate to *S. lycopersicoides* (R_S = 3.21).

Nei's gene diversity. H_T was almost 3.5 times, H_E twice as high in *S.sitens* compared to *S. lycopersicoides* at the subset of eight common allozyme loci (H_T = 0.333 and H_E = 0.256 vs. H_T = 0.097 and H_E = 0.122, respectively). At the 14 shared SSR loci H_T and H_E in *S. sitiens* (H_T = 0.448 and H_E = 0.360) reached only 79 % and 82 % of that in *S. lycopersicoides* (H_T = 0.577 and H_E = 0.442). Expected heterozygosity (H_E) was highest, however, in the single *S. chilense* population (H_E = 0.585).

Inbreeding coefficient per population. The average inbreeding coefficient F_{IS} per population estimated at each of the eight common loci was twice as high in populations of *S. sitiens* (0.121) compared to *S. lycopersicoides* (0.062) at the allozyme loci. The average inbreeding coefficient per population (F_{IS}) measured at the 14 common SSR loci was close to zero and non-significant in both species; -0.016 in *S. lycopersicoides* and 0.009 in *S. sitiens*. The single *S. chilense* population showed an F_{IS} value of 0.042.

3.2.2.5 Signatures of recent population bottlenecks

Results derived from allozyme and microsatellite data were generally not in concordance. Gross incongruencies were also observed between the two statistical detection methods. Only one *S. lycopersicoides* (Aricota #2 - LA2387) and one *S. sitiens* population (Limón Verde - LA4112) were detected as bottlenecked by both approaches, however, the former with microsatellite, the latter with the allozyme data only.

At allozyme loci most of the recent bottleneck signatures, three *S. lycopersicoides* populations (Palca - LA1966, Putre #1 - LA2781, Putre #2 - LA2777) and three *S. sitiens* populations (Paqui - LA4116, Limón Verde - LA4112 and San Juan - LA4111), were in form of mode-shift distortions. Heterozygosity excess was only significant ($P < 0.005$) in Limón Verde (LA4112).

At microsatellites mode-shift distortions were detected in the *S. lycopersicoides* populations Aricota #2 (LA2387) and Chupapalca (LA1964), but only the latter was accompanied by a significant heterozygosity excess ($P < 0.0033$).

Of the *S. sitiens* populations only Escondida (LA4105) exhibited a significant heterozygosity excess ($P < 0.0036$). Cere (LA4113) showed a mode-shift distortion.

3.2.2.6 Correlations between the estimated census population size and genetic diversity estimates

Correlations between the estimated census size in the wild and genetic diversity estimates (k, P, R_S, H_E, F_{IS} per population and average pairwise θ; Table 4, 5) were determined (Appendix 5). Gross estimates of population sizes in the natural habitat were obtained from observations made at the time of collection. Imprecise descriptions such as 'large', 'small', etc. were converted into numerical values (75, 20 individuals, respectively). Approximations for *S. sitiens* are probably more accurate; plants are easier to spot in the terrain, and numerical population size estimates were available for all accessions. Thus, the average census population in the wild was estimated to comprise 56 individuals in *S. lycopersicoides*, ranging from ca. 20 (e.g. Pachica - LA4130) to ca. 100 (Perquejeque - LA2776). *S. sitiens* populations were smaller, containing 44 individuals on average, from ca. 17 (Quimal - LA4331) to ca. 100 (Paqui - LA4116). The estimated census population size in the wild showed significant correlations only with P at allozyme loci and θ at SSR loci in *S. lycopersicoides*, and with both estimators at both locus types in *S. sitiens*.

3.2.2.7 Statistical differences between allozyme and microsatellite data

In order to investigate the concordance between the two marker systems within-species comparisons were based on the common set of individuals (i.e. those that were genotyped with both marker systems, allozymes and SSRs. Estimates were highly similar to those based on the complete set of individuals. Of the parameters that were tested for significant differences between the two marker systems in each population (k, P, R_S, H_E and F_{IS} per population), and at each locus (k, P, R_T, H_E, F_{IS}, F_{ST} and F_{IT} per locus) all but the F_{IS} and the F_{IT} per locus and the F_{IS} per population revealed significant differences ($P < 0.05$) in *S. lycopersicoides* with both test statistics, the conventional t-test and the nonparametric Mann-Whitney test (Appendix 6, 7).

In contrast, in *S. sitiens* none of the loci means and only the average number of alleles (k), the average allelic richness per population (R_S) and the inbreeding coefficient per population (F_{IS}) showed significant differences according to both test statistics.

Mantel tests on pairwise F_{ST} (θ) values indicated significant ($P < 0.01$) correlations between the two marker systems in both species, but the correlation was substantially weaker ($r = 0.363$) in *S. lycopersicoides* than in *S. sitiens* ($r = 0.916$).

3.2.3 Population genetic structure

3.2.3.1 Analysis of molecular variance

The partitioning of genetic variation within each species was determined via a hierarchical analysis of molecular variance (AMOVA). Most genetic variation was found within and less among populations at both marker types (Table 6). Global levels of inbreeding were presented as weighted average over loci: The inbreeding coefficients describe the deviation from expected heterozygosity due to assortative mating within populations (F_{IS}) or within individuals relative to the total population (F_{IT}), or within populations relative to the entire population as a result of selection or drift (F_{ST}; Lowe et al. 2004).

S. lycopersicoides. Allozyme analysis revealed genetic variation was partitioned to 84.7 % within populations and only to 15.3 % among populations of *S. lycopersicoides*. The proportion of genetic variation among individuals within populations (F_{IS}) equaled zero (i.e. was slightly negative which represented an artifact due to heterogeneity in population sizes). Global inbreeding (F_{IT}) measured 0.123, fixation among populations (F_{ST}) 0.153 and inbreeding within populations (F_{IS}) -0.036. The F_{ST} and F_{IT} values were highly significant ($P < 0.0005$ and $P = 0$, respectively). In agreement with allozyme data microsatellite analysis revealed a far greater amount of genetic diversity residing within populations compared to among populations (73.2 % vs. 26.9 %). Nearly

Source of variation	df	SSD	VC	% variation	Inbreeding coefficients		P-value
Allozymes							
Among populations	11	46.2	0.077	15.3	F_{ST}	0.153	0.000
Among individuals within populations	286	116.5	-0.015	-3.1	F_{IS}	-0.036	0.817
Within populations	298	130.5	0.438	87.7	F_{IT}	0.123	0.000
Total	595	293.2	0.499				
Microsatellites							
Among populations	13	713.9	1.148	26.9	F_{ST}	0.269	0.000
Among individuals within populations	303	926.7	-0.068	-1.6	F_{IS}	-0.022	0.920
Within populations	317	1012.5	3.194	74.7	F_{IT}	0.253	0.000
Total	633	2653.2	4.274				

Table 6. Analysis of molecular variance (AMOVA) and global levels of inbreeding in *S. lycopersicoides*.

df = degrees of freedom; SSD = sum of square deviations; VC = variance components; F_{ST} = fixation index; F_{IS} = the inbreeding coefficient; F_{IT} = the overall inbreeding coefficient. Values pepresent the weighted average over loci.

zero variation was partitioned among individuals within populations (-1.6 %; the negative value was an artifact as a consequence of heterogeneous sample sizes). Results were highly significant (P = 0).

Locus-by-locus AMOVA in S. lycopersicoides. Weighted average results over loci revealed F_{IS} values (estimated as Weir & Cockerham's '*smallf*') ranging from -0.148 (*Idh-1*) to 0.289 (*Pgm-2*), F_{ST} (*theta*) values from 0.008 (*6-Pgdh-3*) to 0.222 (*Fdh-1*) and F_{IT} (*Capf*) values from -0.011 (*Idh-1*) to 0.324 (*Pgm-2*; mean F_{IT} = 0.139; Appendix 8A). None of the values were significantly different from zero ($P < 0.004$ per locus, $P < 0.006$ per population). *Fdh-1* was the locus at which differentiation among populations was highest (22.2 %), *Pgm-2* and *6-Pgdh-2* showed extraordinarily high levels of differentiation among individuals within populations (27.5 % and 20.8 %, respectively) and at four loci (*6-Pgdh-3, Idh-1, Got-2* and *Pgi-1*) the entire genetic variation was partitioned among individuals of the whole species. More than half of the values were highly significant ($P < 0.004$).

At SSR loci F_{IS} values ranged from -0.177 (SSR98) to 0.054 (SSR85) and were not significant ($P < 0.004$). F_{ST} values were highly significant (P = 0) at all loci, indicating differentiation among populations. Values ranged from 0.069 at SSR98 to 0.692 at SSR578 with an average of 0.269. F_{IT} values were significant (P < 0.004) at all loci except one (SSR98) and spanned from -0.096 at SSR98 to 0.703 at SSR578.

S. sitiens. In *S. sitiens* a higher percentage of variation was found among population (30.3 %) and accordingly less within populations (69.7 %) at allozyme loci (Table 7). Fixation among individuals within populations was very low (F_{IS} = 0.049). The level of fixation among populations (F_{ST}) was similar to the overall amount of inbreeding (F_{IT} = 0.337). Both global F_{ST} and F_{IT} were highly significant (P = 0). A far greater amount of genetic diversity within populations compared to among populations was also revealed by the microsatellite analysis (78.5 % vs. 21.5 %), and essentially zero (0.7 %) among individuals within populations. However, unlike in *S. lycopersicoides*, among-population differentiation was less pronounced than at allozyme loci. Results were highly significant (P = 0).

Locus-by-locus AMOVA in S. sitiens. At individual loci weighted average results indicated greater variation among inbreeding levels in *S. sitiens* compared to *S. lycopersicoides* (Appendix 8B). F_{IS} values spanned from -0.455 (*Pgm-2*) to 0.346 (*Adh-2*). The former value was significantly different from zero (P < 0.005). F_{ST} values ranged from 0.063 (*Idh-1*) to 0.585 (*Got-2*) and F_{IT} values from -0.062 (*Pgi-1*) to 0.585 (*Got-2*). Among loci *Got-2* showed the greatest amount of differentiation among populations (59.0 %), *Adh-2* among individuals within populations (29.5 %) and both *Pgm-2* the highest among all individuals (100 %). Most of the F_{ST} and F_{IT} values were

Source of variation	df	SSD	VC	% variation	Inbreeding coefficients		P-value
Allozymes							
Among populations	5	152.8	0.584	30.2	F_{ST}	0.302	0.000
Among individuals within populations	144	203.3	0.066	3.4	F_{IS}	0.049	0.063
Within populations	150	192.0	1.280	66.3	F_{IT}	0.337	0.000
Total	299	548.1	1.930				
Microsatellites							
Among populations	6	196.4	0.683	21.5	F_{ST}	0.215	0.000
Among individuals within populations	148	373.3	0.021	0.7	F_{IS}	0.008	0.358
Within populations	155	384.5	2.481	77.9	F_{IT}	0.221	0.000
Total	309	954.2	3.185				

Table 7. Analysis of molecular variance (AMOVA) and global levels of inbreeding in *S. sitiens*.

df = degrees of freedom; SSD = sum of square deviations; VC = variance components; F_{ST} = fixation index; F_{IS} = the inbreeding coefficient; F_{IT} = the overall inbreeding coefficient. Values pepresent the weighted average over loci.

highly significant (P = 0).

At SSR loci results were very similar to those of *S. lycopersicoides*. F_{IS} values ranged from -0.173 (SSR98) to 0.418 (SSR80). Apart from SSR80 significant inbreeding was detected at two additional loci across the combined populations: SSR50 (F_{IS} = 0.159) and SSR15 (F_{IS} = 0.152). F_{ST} values were highly significant at all loci except for SSR98, reaching from 0.019 (SSR98) to 0.367 (SSR599). F_{IT} values varied from -0.151 (SSR98) to 0.612 (SSR80) and were significant (P < 0.007) at 8 out of 15 loci.

3.2.3.2 Pairwise genetic distance and Mantel tests

Correlations between genetic distance and both geographic and elevational distance were evaluated with Mantel tests. Matrices with pairwise geographic distances are shown in Appendix 9, matrices with pairwise genetic distances in Appendix 10A and 10B for *S. lycopersicoides* and *S. sitiens*, respectively.

S. lycopersicoides. Among *S. lycopersicoides* populations the geographic distance was smallest between Putre #2 (LA2777) and Perquejeque (LA2776; 2.5 km) and largest between Aricota #2 (LA2387) and Moquella (LA4018; 240.2 km) with a mean of 99.1 km between any two populations. Elevational distances varied substantially among populations, ranging from just 3 m

between Chupapalca (LA1964) and Putre #2 (LA2777), and to 2227 m between Putre #1 (LA2781) and Lluta (LA4320) and with a mean of 743.7 m.

Pairwise genetic differences were estimated as Weir's & Cockerham's θ. With allozyme data the greatest genetic distance (θ = 0.532) was observed between populations Putre #1 (LA2781) and Esquina (LA4131), the smallest genetic distance (θ = -0.001) between the population pair Putre #1 (LA2781) and Zapahuira (LA2772). The species mean was 0.169. Most values were significant at the adjusted 5 % nominal level. The average pairwise θ was highest for the population Esquina (LA4131; θ = 0.345) and lowest for population Perquejeque (LA2776; θ = 0.088).

The correlation between allozyme genetic diversity and geographic distance was low (r = 0.223) but significant (P < 0.05). The correlation between genetic distance and elevational distance was stronger (r = 0.381, P < 0.05).

At SSRs θ measures ranged from 0.060 for the pair Nama (LA4126)/Camiña (LA4123) to 0.464 for the pair Aricota #1 (LA4018)/Esquina (LA4131) with a mean of 0.261. The northernmost population Aricota #1 (LA4018) displayed the highest average θ (0.342), which was almost twice as high as the lowest average value (0.199) in Perquejeque (LA2776) at the center of the distribution range, indicating isolation by distance. All estimates were highly significant (P = 0).

In contrast to the allozyme analyses the Mantel test detected isolation by distance (IBD). A highly significant positive correlation (r = 0.680, P = 0) was observed between genetic variation and geographic distance. Only a very weak and non-significant negative correlation was observed between genetic and elevational distance (r = -0.033).

S. sitiens. Geographic distances among *S. sitiens* populations were similar to those of *S. lycopersicoides*, ranging from a mere 4.3 km between Paqui (LA4116) and Carbonatera (LA4114) to 232.9 km between Paqui and Escondida (LA4105; mean 95.2 km). Overall elevational distances among *S. sitiens* populations were less than a quarter of those in *S. lycopersicodes*, ranging from 18 m between Carbonatera (LA4114)/San Juan (LA4111) to 456 m between Quimal (LA4331)/Escondida (LA4105). The mean distance was 162.9 m.

At allozymes pairwise θ values ranged from 0.070 for the pair Paqui (LA4116)/Carbonatera (LA4114) to 0.535 between Carbonatera (LA4114) and Escondida (LA4105; mean θ = 0.280). All pairwise θ values were highly significant (P = 0). The average pairwise θ was highest for the southernmost population Escondida (LA4105; θ = 0.476) and lowest for Limón Verde (LA4112; θ = 0.225) at the center of the distribution, as expected under isolation by distance.

The correlation between allozyme genetic distance and geographic distance was much higher (r = 0.785, P < 0.001) than in *S. lycopersicoides*. The correlation with the elevational distance was close

to zero (r = 0.006) and non-significant, not surprising given the relatively limited elevational range of this species.

At SSR loci θ-values varied among population pairs from 0.048 between Carbonatera (LA4114) and Paqui (LA4116) to 0.370 between Carbonatera (LA4114) and Escondida (LA4105) with a mean of 0.203. All values were significantly different from zero (P = 0). As with isozyme data the average F_{ST} was substantially higher in the southernmost population Escondida (LA4105; θ = 0.304), ca. 70 % increased compared to the next lower average (θ = 0.207) in Carbonatera (LA4114). The lowest average θ showed Limón Verde (LA4112; θ = 0.158) at the center of the distribution. As in *S. lycopersicoides* findings were in line with the presence of IBD. The Mantel test revealed a high correlation between genetic and geographic distance (r = 0.847, P < 0.001), similar to the allozyme estimate. The correlation between genetic and elevational distance was not significant.

3.2.3.3 Gene flow

The average amount of gene flow among populations was derived by means of the private allele method (Barton and Slatkin 1986) and from Wright's formula (Wright 1951):

$$Nm \equiv (1 - F_{ST})/4\, F_{ST}$$

The private allele method estimated a very high migration rate at allozymes; 2.54 migrants on average per population after correction for population size. As a comparison, Wright's formula estimated fewer migrants (Nm = 1.23). At SSRs the average number of migrants was estimated 0.72 after correction for population size and – very similar – 0.71 according to Wright's formula.

In contrast to *S. lycopersicoides* the estimated gene flow was fairly low at allozyme loci, only 0.36 migrants per population. Wright's formula estimated almost twice as many migrants (Nm = 0.64) from the same dataset. A slightly higher migration rate (Nm = 0.64) was estimated at SSRs. Wright's formula returned a higher value (Nm = 0.98).

3.2.3.4 Spatial trends of genetic diversity: correlations between geographic locations and genetic diversity parameters

The genetic diversity parameters (k, P, R_S, H_E, F_{IS} per population and average pairwise θ) were tested for correlations with geographic parameters (latitude, longitude, elevation and isolation measured as average distance to other populations (Appendix 5).

Latitude. In *S. lycopersicoides* P and θ were negatively correlated with latitude (the former with SSR data only), i.e. both parameters increased towards the south. The amount of significant

correlations was similar on the allozyme and SSR levels, whereas in *S. sitiens*, again, allozymes showed stronger signals of geographic clines.

In *S. sitiens* most genetic diversity parameters (k, P, R_S and θ) were correlated with latitude in the allozyme dataset, but only θ in the SSR dataset. Except for θ correlations were positive, reflecting a decrease of genetic diversity and an increase of genetic distance towards the south.

Longitude. P and θ were correlated with longitude in *S. lycopersicoides*, and with the exception of P in the allozyme dataset in a positive mode. In *S. sitiens* both P and θ (allozyme dataset) or just θ (SSR dataset) showed a correlation with longitude, of which the one with θ was negative.

Elevation. In *S. lycopersicoides* P increased with altitude in the SSR dataset, but P and θ decreased with altitude in the allozyme dataset. In *S. sitiens* P was positively and θ negatively correlated with altitude in both datasets.

Isolation. In *S. lycopersicoides* the average geographic distance was positively correlated with P (allozyme data) and θ (SSR data), negatively correlated with the number of alleles (SSR data). In *S. sitiens* the estimates of number of alleles (allozyme data only) and P (both datasets) declined with increasing average geographic distance, i.e. towards the margins of the distribution while θ increased with average geographic distance (both datasets).

3.2.3.5 Spatial trends of genetic diversity: Statistical correlations between geographic parameters and allele frequencies

Most of the correlations between allele frequencies and geographic parameters were evident with latitude, followed by longitude and to a lesser extend elevation (Appendix 11A, 11B). The highest number of significant correlations was observed with the *S. lycopersicoides* SSRs dataset where one third (32 %) of the alleles showed a trend with latitude, one quarter (26 %) with longitude and just 3 % with elevation. At allozyme loci 8 % of allele frequencies were correlated with latitude or longitude and 4 % with elevation.

In comparison, only 13 % and 12 % of the *S. sitiens* allele frequencies at SSR loci were correlated with latitude and longitude, respectively. No correlations were observed with elevation. A greater number of correlations was observed at allozyme loci: 23 % showed trends with latitude, only 10 % with longitude but as many as 20 % with elevation.

3.2.4 Population genetic clusters

Population clusters were estimated via three different routes: a) from a phylogeny based on genetic distance estimates b) by PCA based on Weir's and Cockerham's θ and c) via a model-based

(Bayesian) clustering algorithm provided by the software program STRUCTURE. AMOVA and genetic diversity analyses were subsequently carried out for the major groupings.

3.2.4.1 Phylogenetic trees

Phylogenetic trees were constructed using the software package PHYLIP. Three genetic distances (Reynolds' distance, Cavalli-Sforza's chord distance and Nei's genetic distance) in combination with two tree construction methods (UPGMA and neighbor-joining) were compared for each of the four datasets (i.e. a total of six dendrograms per dataset) in order to identify the most informative approach. Evaluations were based on bootstrap values, the degree of congruence among tree topologies as well as the ability to detect geographic groupings.

Reynolds' distance and Cavalli-Sforza's chord distance were expected to be best suited for allozyme data because they attribute all genetic differences to genetic drift and ignore mutations, while Nei's genetic distance had been anticipated as the distance of choice for the SSR data because it factors in both mutation and drift. However, Cavalli-Sforza's chord distance generated the most likely result only with the *S. lycopersicoides* SSR data. Topologies derived from Reynold's distance were largely similar to those obtained from Nei's genetic distance, yet generally slightly inferior, and therefore Nei's genetic distance were chosen for tree construction from allozyme data and the *S. sitiens* SSR dataset. Topologies generated by the UPGMA method were supported by higher bootstrap values than those derived by the neighbor-joining approach. Also, the hierarchies and clusters produced by the latter were often ambiguous and sometimes contradictory to geographic arrangements.

S. lycopersicoides. In *S. lycopersicoides* bootstrap support was very low in the allozyme dendrogram, and the topology hardly reflected the geographic distribution (Figure 5). This outcome was not unexpected due to the weak IBD pattern. In contrast, the phylogenetic tree derived from SSR data was highly supported and showed population groupings in concordance with geographic patterns (even reflecting drainages; Figure 6). Topologies from the two other genetic distance estimators were largely identical, with the exception that Nei's genetic distance identified Perquejeque (LA2776) and Putre #2 (LA2777) as most closely related and Lluta (LA4320) as basal to these, but less well supported (data not shown). The divide into 'Peruvian' (cluster A) and

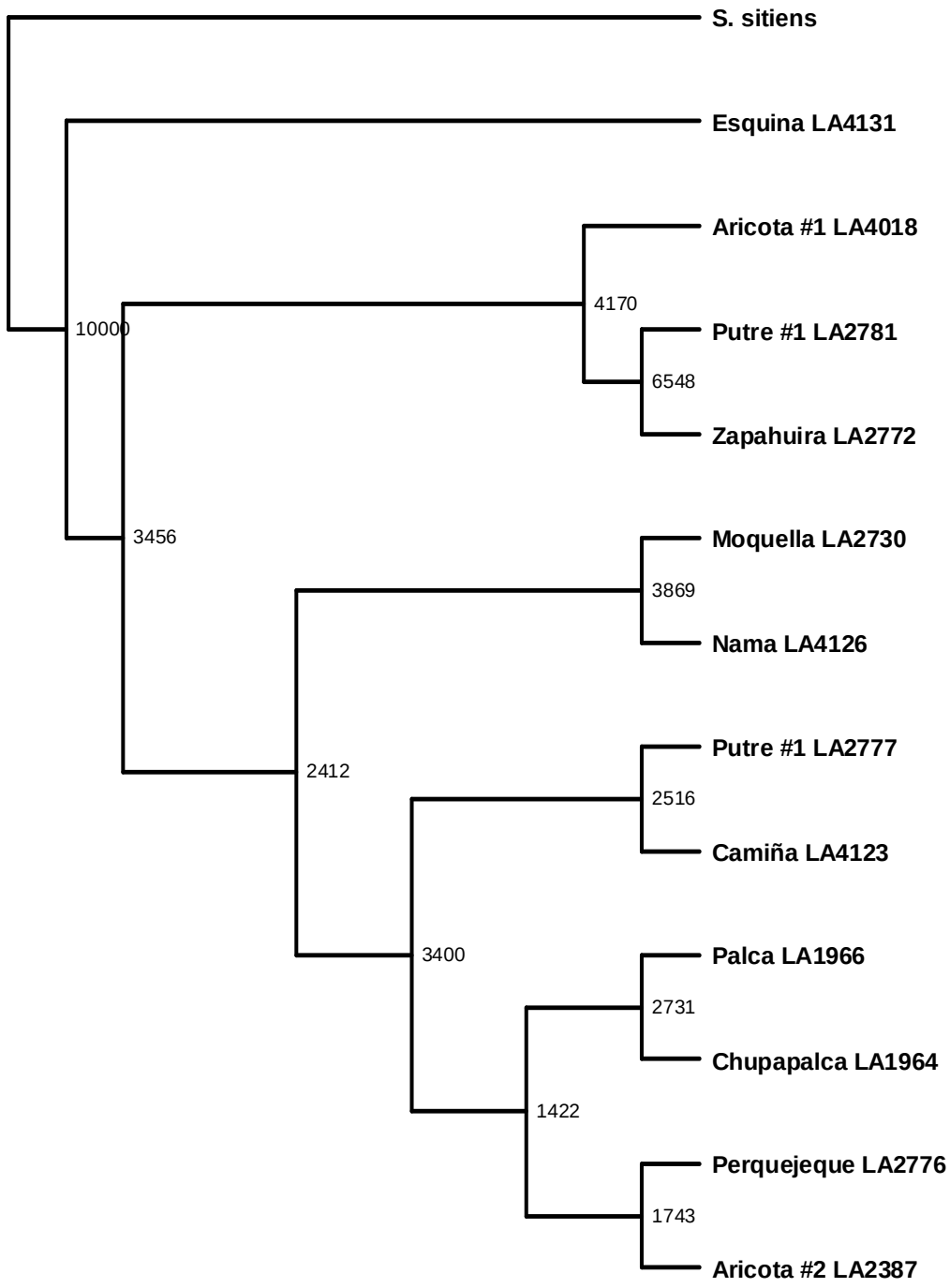


Figure 5. Allozyme phylogeny of *S. lycopersicoides* based on Nei's genetic distance.

Node numbers indicate the number of times a node occurred among 10,000 replications.

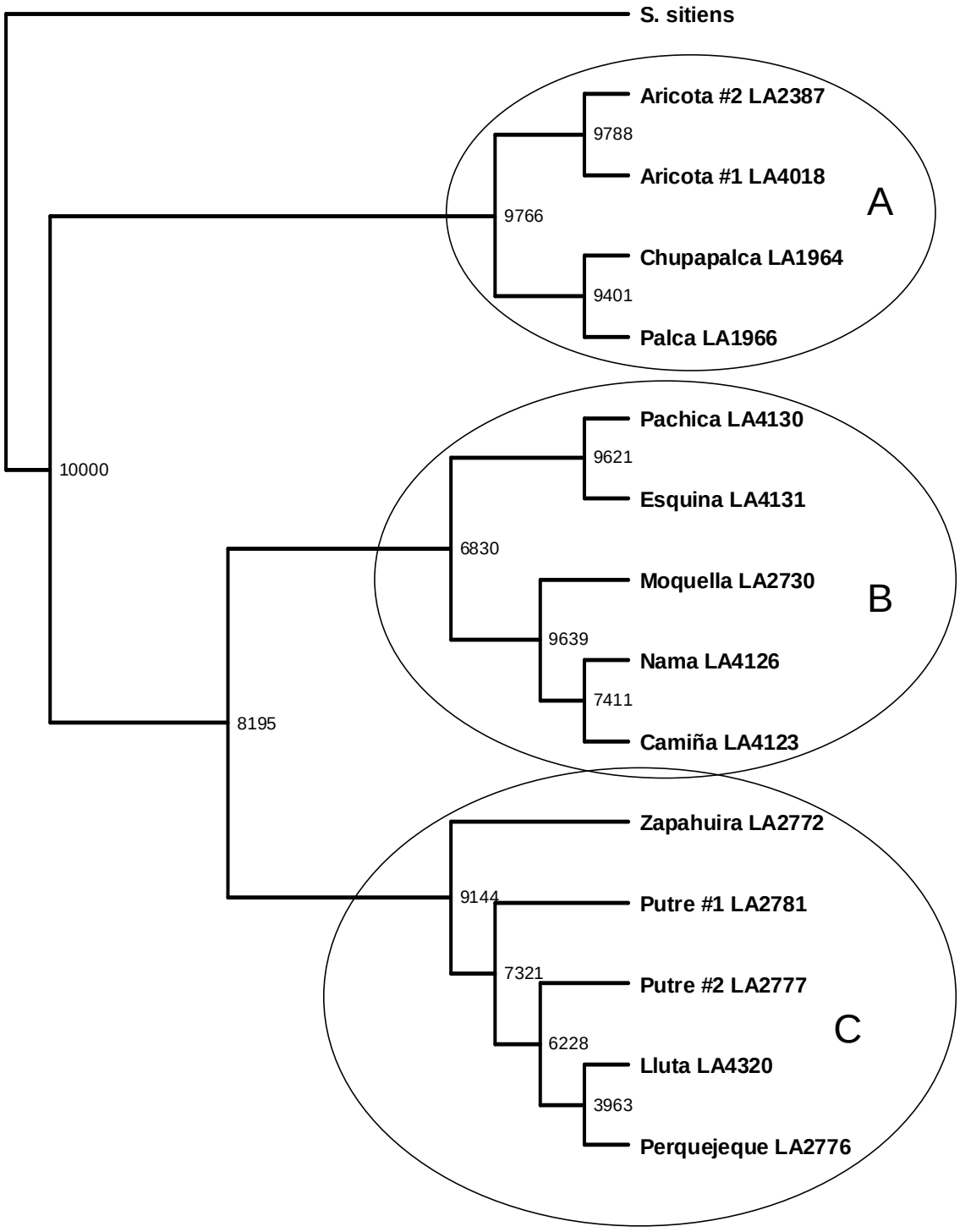


Figure 6. Microsatellite phylogeny of *S. lycopersicoides* based on Cavalli-Sforza's chord distance.

Node numbers indicate the number of times a node occurred among 10,000 replications.

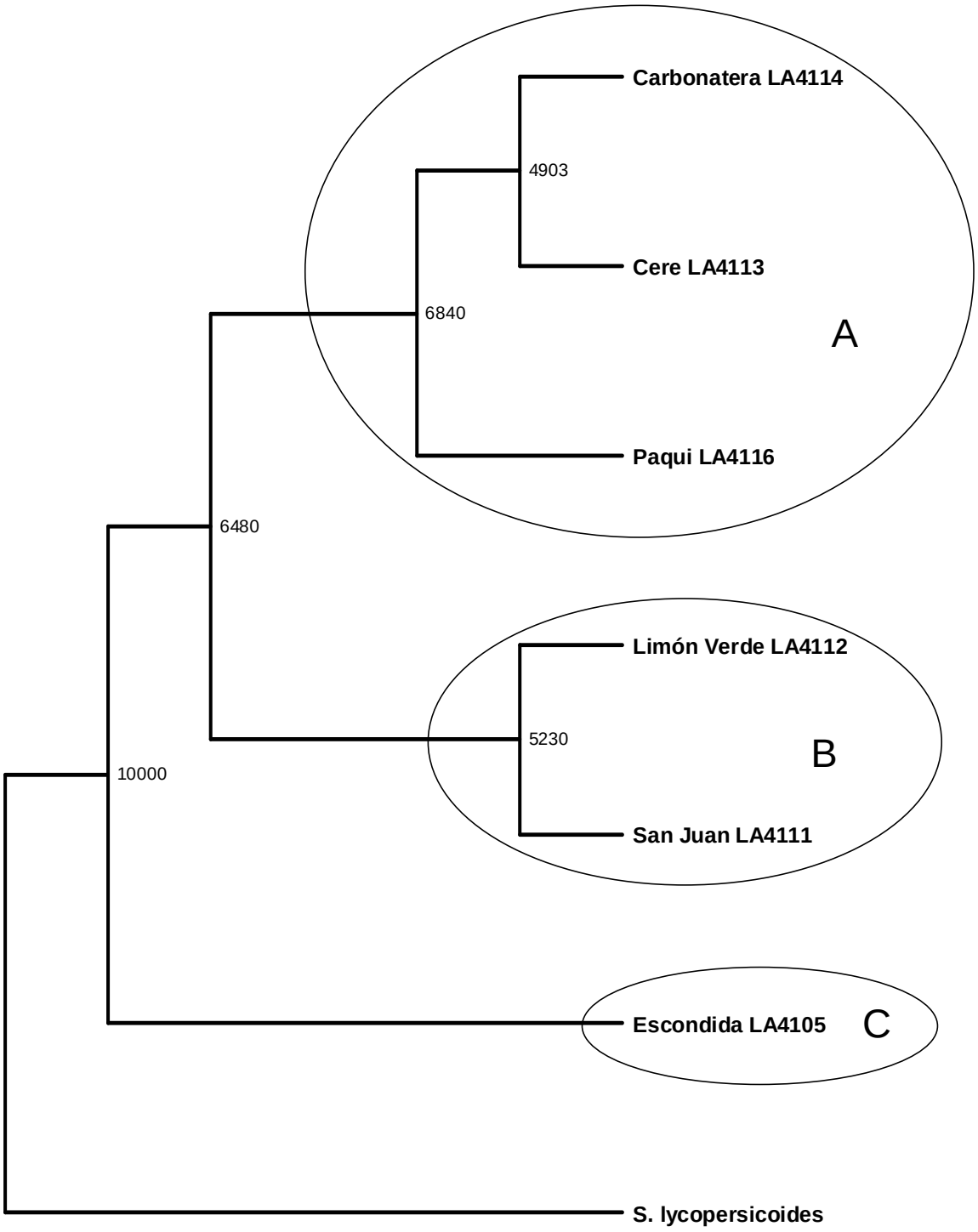


Figure 7. Allozyme phylogeny of *S. sitiens* based on Nei's genetic distance.

Node numbers indicate the number of times a node occurred among 10,000 replications.

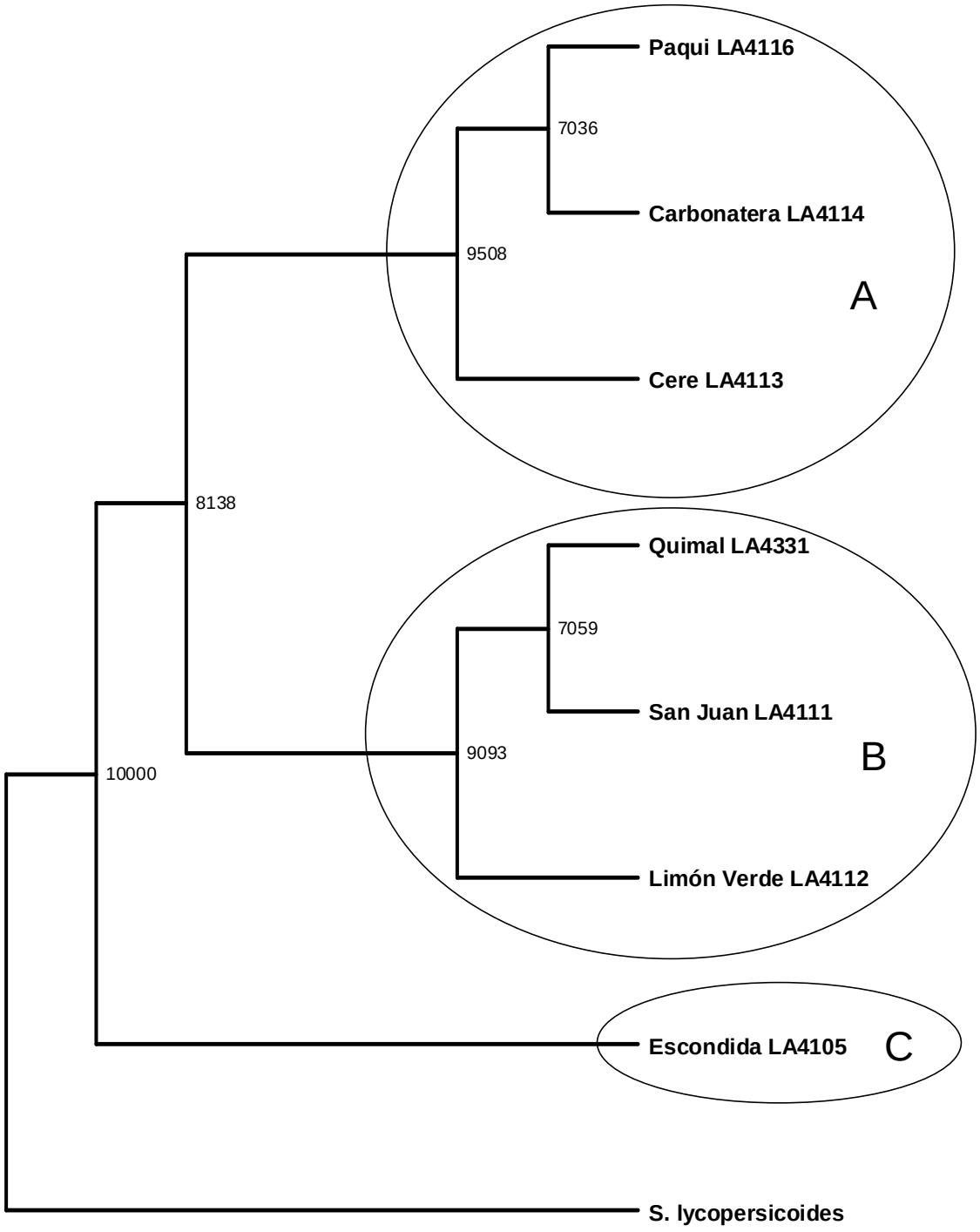


Figure 8. Microsatellite phylogeny of *S. sitiens* based on Nei's genetic distance.

Node numbers indicate the number of times a node occurred among 10,000 replications.

'Chilean' (clusters B and C) populations was evident in the hierarchy, the latter was further partitioned into a 'central' (cluster C) and 'southern' group (cluster B).

S. sitiens. Allozyme and SSR dendrograms were in good agreement in *S. sitiens* (Figure 7, 8). Both identified clusters that corresponded to the geographic provenances of the populations. The only differences compared to the depicted topology were that, disregarding Escondida (LA4105; branch C), San Juan (LA4111) emerged as basal to the remainder of populations in the tree based on Cavalli-Sforza's chord distance, and that Reynold's distance identified Cere (LA4113) instead of Paqui (LA4116) as basal among the three populations of the 'northern cluster' (cluster A). The same configurations were evident in the hierarchy derived from SSR data.

Topologies generated from the three genetic distance estimators were identical for the *S. sitiens* SSR data, but the best support was given by Nei's genetic distance (Figure 8). Escondida (LA4105) clearly emerged as an 'outlying' population (branch C), in agreement with its geographic isolation from the other accessions, and its relatively distinct morphology (data not shown). The remainder of populations clustered into a 'northern' group (composed of three populations in close geographic proximity; cluster A) and a 'central' group (composed of three populations that were more scattered out; cluster B).

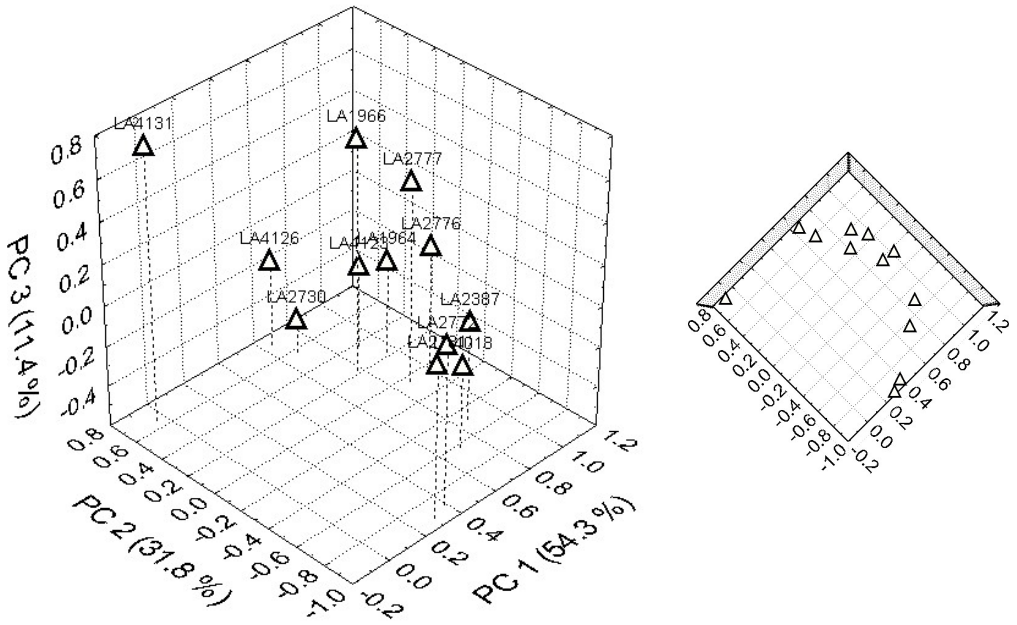
3.2.4.2 Principal component analysis

Topologies identified in the cluster analysis could be confirmed in a PCA performed on Weir and Cockerham's pairwise θ (Figure 9, 10). Again, allozyme and SSR results did not correspond well in the case of *S. lycopersicoides*. Also noteworthy was the continuous population structure at the first two dimensions at allozyme loci in contrast to the clear identification of three clusters at the SSR level.

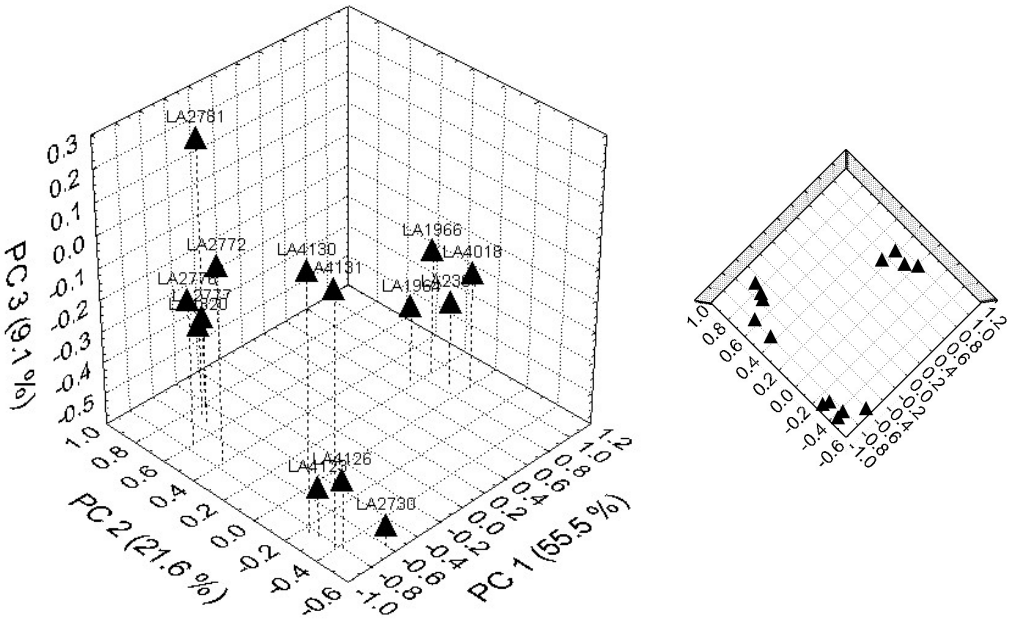
3.2.4.3 STRUCTURE analysis

On the basis of allele frequencies genotypes were assigned to populations in STRUCTURE analyses, assuming that a) populations were admixed and b) allele frequencies were correlated as a consequence of shared ancestry and/or migration.

S. lycopersicoides. Allozyme data failed to reveal any structural pattern in the STRUCTURE analysis (Figure 11), in agreement with the weak signal of IBD. As number of 'true' populations K = 2, 3 or 12 were assumed. Subsequent computations were therefore carried out on SSR data only



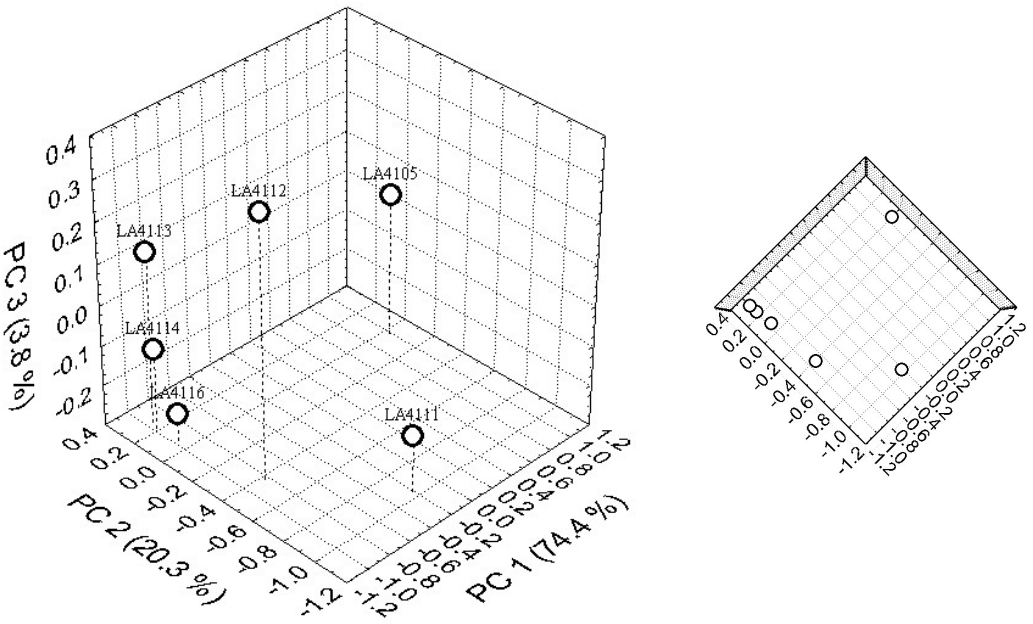
Quality of representation = 97.5 %, Multiple R(z/xy) = 0.510, p = 0.259



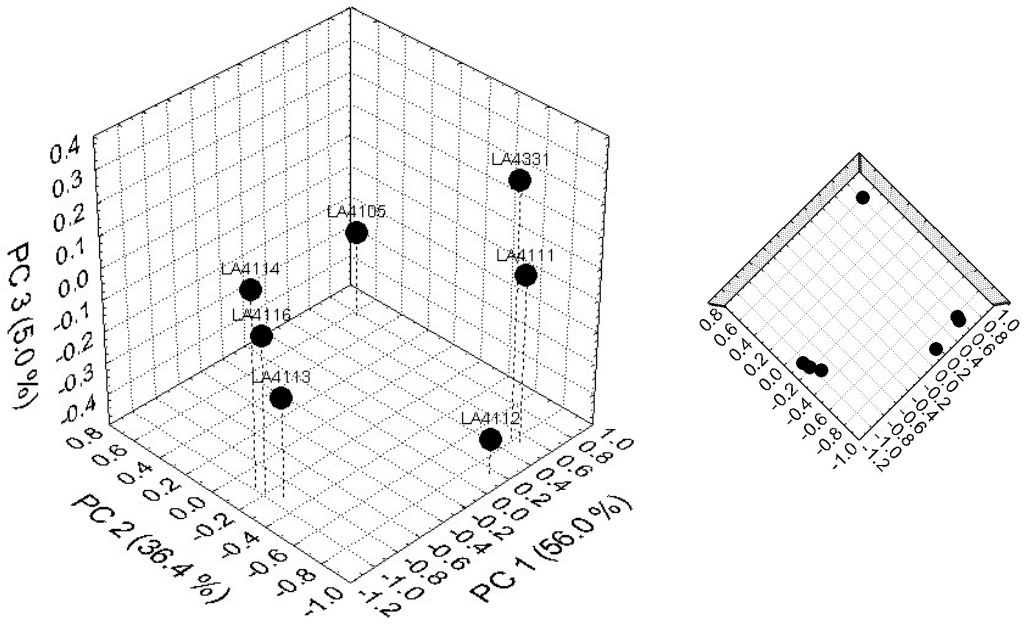
Quality of representation = 86.2 %, Multiple R(z/xy) = 0.364, p = 0.459

Figure 9. PCA in *S. lycopersicoides*.

Principle component analysis of 317 and 298 *S. lycopersicoides* plants, from 12 and 14 wild populations, using 9 allozyme (empty triangles) and 15 SSR loci (black triangles), respectively.



Quality of representation = 98.5 %, Multiple R(z/xy) = 0.214, p = 0.999



Quality of representation = 97.4 %, Multiple R(z/xy) = 0.182, p = 0.935

Figure 10. PCA in *S. sitiens*.

Principle component analysis of 150 and 155 *S.sitiens* plants, from 6 and 7 wild populations, using 10 allozyme (empty circles) and 14 SSR loci (black circles), respectively. Graphs are depicted in three dimensions (left) and viewed from above (right).

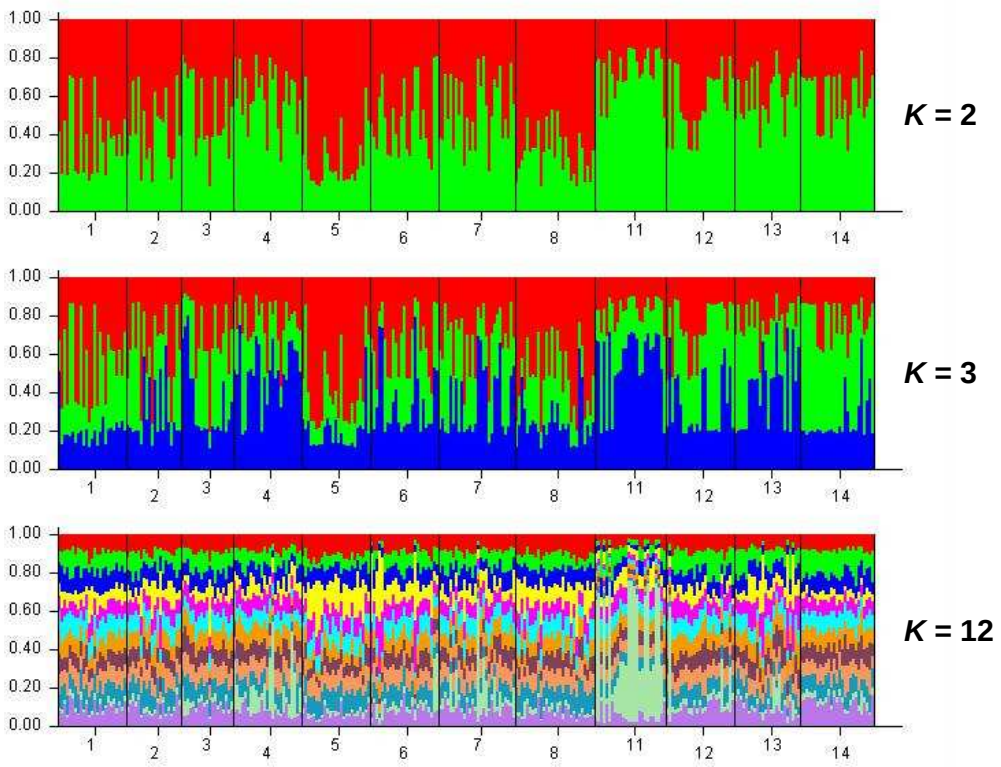


Figure 11. STRUCTURE analysis in *S. lycopersicoides* based on allozyme data.

Membership structure of 317 *S. lycopersicoides* plants based on allozyme allele frequencies, assuming two (K = 2), three (K = 3) and twelve groups (K = 12). Horizontal numbers correspond to the population numbers in Table 2, membership coefficients (Q) are depicted vertically for each individual.

(Figure 12). Assignment patterns from SSR data were generated by three runs each from K = 1 to K = 15 and analyzed in order to determine clustal relationships. Populations and geographic groupings were largely reflected in the result images with the following exceptions and specificities: Principal, unambiguous assignments were made at K = 2 and K = 3, resulting in a northern 'Peruvian' (Lago Aricota and Palca - LA4018, LA2387, LA1964, LA1966)/central-southern 'Chilean' (remainder of locations) subdivision, or a northern (Lago Aricota and Palca)/central ('Putre' – LA2781, LA2777, LA2776, LA2772, LA4320)/southern (Camarones and Camiña – LA4130, LA4131, LA4126, LA4123, LA2730) subdivision, respectively. The northern group (Lago Aricota and Palca) retained its integrity up to K = 12. The northernmost accessions from Lago Aricota (LA4018 and LA2387) consistently formed one group up to K = 15, demonstrating distinctness from the other populations. The first single population that emerged was Zapahuira (LA2772) at K = 4, which showed associations to Putre #1 (LA2781) and Pachica (LA4130)/Esquina (LA4131), both at Camarones. The second single population that emerged was Putre #1 (LA2781) at K = 5. Of the remainder of

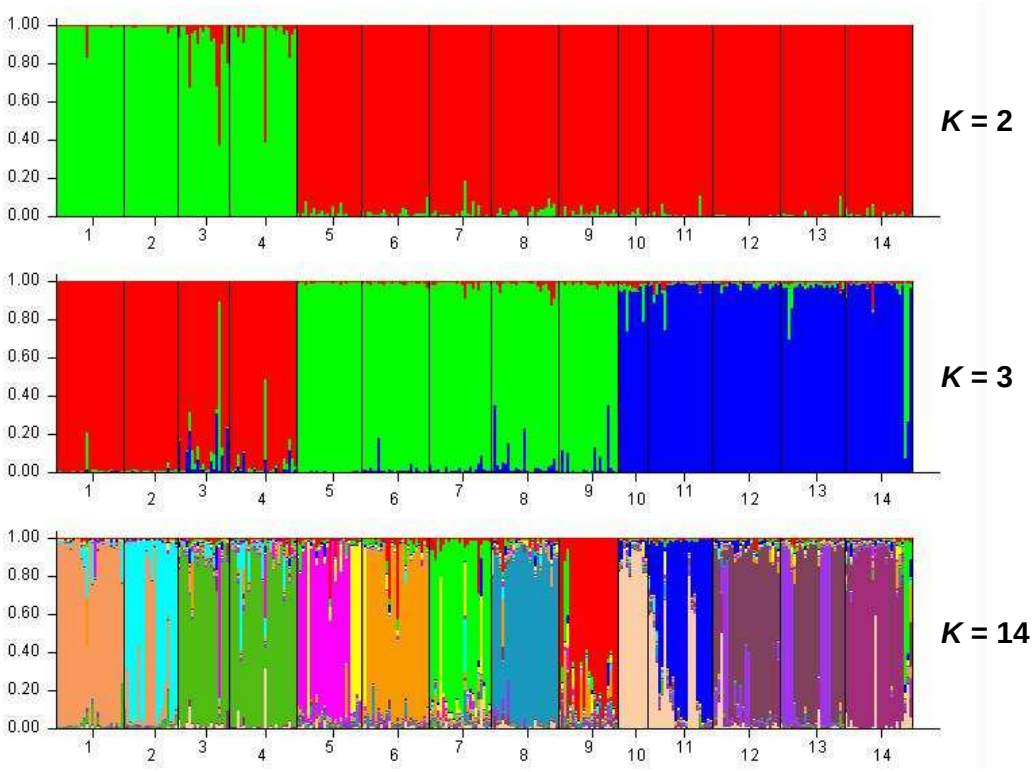


Figure 12. STRUCTURE analysis in *S. lycopersicoides* based on SSR data.

Membership structure of 298 *S. lycopersicoides* plants based on SSR allele frequencies, assuming two (K = 2), three (K = 3) and 14 groups (K = 14). Horizontal numbers correspond to the population numbers in Table 2, membership coefficients (Q) are depicted vertically for each individual.

'Putre' populations (Perquejeque - LA 2776, Putre #2 - LA2777 and Lluta - LA4320) associations were strongest between Perquejeque (LA2776) and Lluta (LA4320) ; Putre #2 (LA2777) subdivided at K = 7. The two southern locations (Camarones: Pachica and Esquina, and Camiña: Nama, Camiña and Moquella) were consistently identified as two separate clusters at $K \geq 7$. At Camarones Pachica (LA4130) appeared as genetic subset of Esquina (LA4131). Within the southernmost group (Camiña) Camiña (LA4123) and Nama (LA4126) were closely connected, i.e. still formed one coherent group (albeit composed of two clusters) at K = 15. At rising K values the first individual populations to contain > 1 K were Zapahuira (LA2772), Putre #2 (LA2777) and Putre #1 (LA2781), indicating high levels of genetic variation in that group.

The 'true' number of clusters was estimated from the posterior probability of the log-likelihood distribution $Pr(X \mid K$; Pritchard et al. 2000) and an estimator derived from the second order rate of change of the likelihood distribution of K (ΔK; Evanno et al. 2005). Using three runs for each scenario (from K = 1 to K = 15) a peak was observed at K = 14, which, although representing the

actual number of populations, would not be identical with the 14 physical populations, because the algorithm detected a higher number of populations among central populations at the expense of lower numbers at the northern/southern margins (data not shown). The second approach (based on ΔK) identified a total of three 'true' populations (K = 3) and a secondary peak at K = 14.

The parameter α measures the degree of admixture and is therefore also an indicator of population structure. At an α value near zero levels of admixture are low, at an α value > 1 most individuals are admixed (Falush et al. 2003). The average α of 0.303 over all runs therefore confirmed that clustal patterns were prevailing in *S. lycopersicoides*.

S. sitiens. Simulations from K = 1 to K = 8, three runs each, were carried out with SSR data. Spatial patterns were evident in the STRUCTURE report. Assuming K = 2 individuals segregated into a 'northern' group (Paqui - LA4116, Carbonatera - LA4114 and Cere - LA4113; Figure 13) and the remainder of populations, from which the southernmost location Mina la Escondida (LA4105) split at K = 3. Even under the assumption of the true population number (seven), two (Paqui -

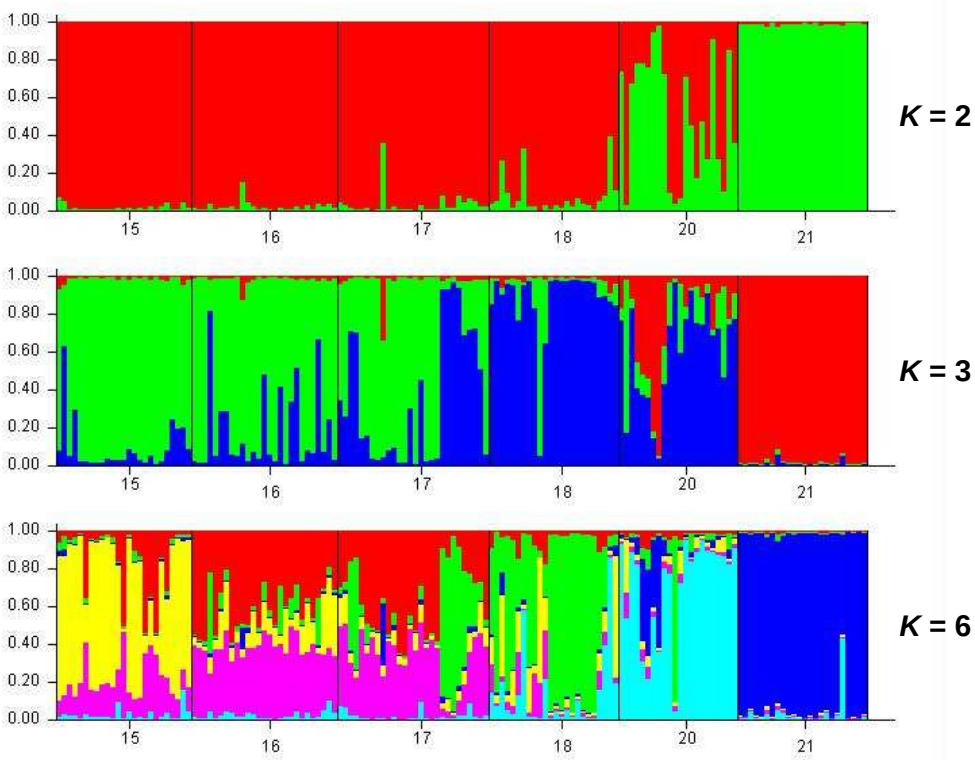


Figure 13. STRUCTURE analysis in *S. sitiens* based on allozyme data.

Membership structure of 150 *S. sitiens* plants based on allozyme allele frequencies, assuming two (K = 2), three (K = 3) and six groups (K = 6). Horizontal numbers correspond to the population numbers in Table 2, membership coefficients (Q) are depicted vertically for each individual.

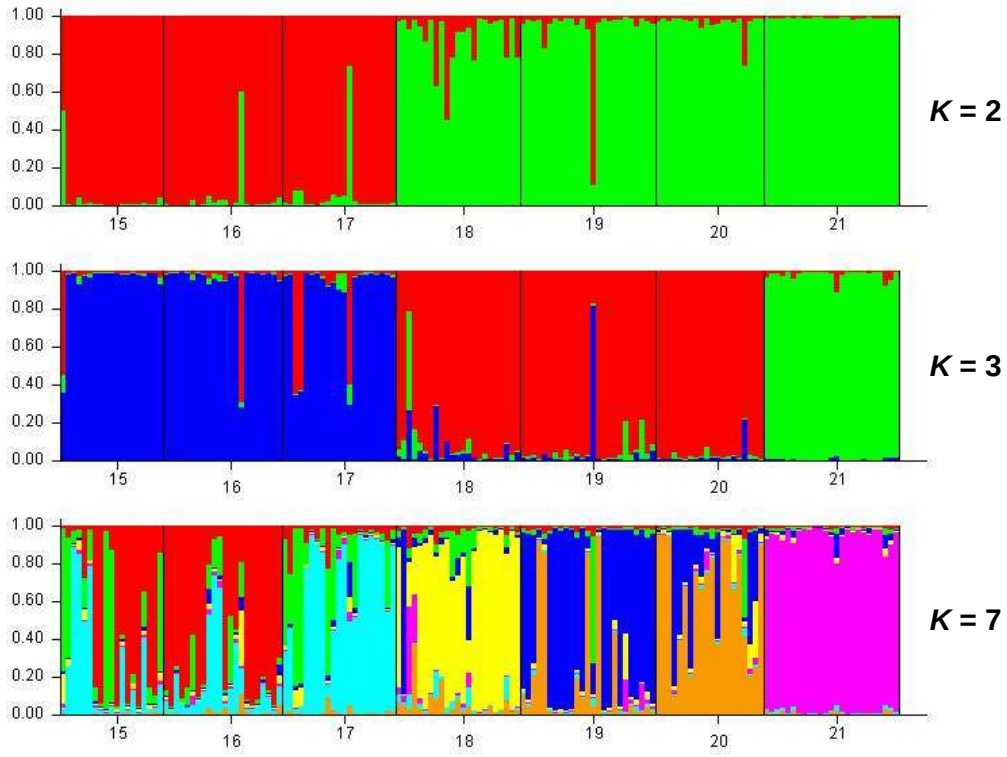


Figure 14. STRUCTURE analysis in *S. sitiens* based on SSR data.

Membership structure of 155 *S. sitiens* plants based on SSR allele frequencies, assuming two (K = 2), three (K = 3) and seven groups (K = 7). Horizontal numbers correspond to the population numbers in Table 2, membership coefficients (Q) are depicted vertically for each individual.

LA4116 and Carbonatera - LA4114) of the three populations at the northern location appeared substantially intermingled up to K = 8 while the rest of the populations were identified as clearly distinct groups.

The central populations showed lower levels of admixture. An association was observed between Quimal (LA4331) and the (geographically distant) San Juan (LA4111). The single, isolated southernmost population at Mina La Escondida (LA4105) showed an allele frequency pattern that set it apart from the remainder of populations.

Singular allozyme simulations were in good agreement with those from SSR data with these exceptions: the principle divide laid between Escondida (LA4105; Figure 14) and the remainder of populations, the closest relationship in the northern group was observed between Carbonatera (LA4114) and Cere (LA4113), and Cere (LA4113), the most southern population within the northern group, showed some admixture from the next population further south, LA4112. Both marker systems showed the lowest levels of admixture for Escondida (LA4105), followed by San Juan (LA4111) at allozyme loci and Limón Verde (LA4112) at microsatellite loci. The distribution

of Pr(X | *K*) based on SSR data indicated the presence of seven distinct groups, the estimator Δ*K* returned *K* = 3 as the number of real populations, but as in *S. lycopersicoides* a secondary peak was evident at *K* = 7. The average α over all runs was 0.036 (data not shown).

3.2.4.4 AMOVA within and among population substructures

Amova analyses were conducted against the backdrop of the three major population clusters, and revealed significant differences among the predicted groups in both species.

S. lycopersicoides. Because of the ambiguities associated with the *S. lycopersicoides* allozyme dataset individual grouping analyses are presented in greater detail on the basis of the SSR dataset (which were in congruence with geographic distributions) only. At allozyme loci only 4.2 % of the total variation was present among groups, 12.1 % among populations within groups (Appendix 12A). Both fixation indices were significantly different from zero, indicating structuring. However, results suggested that the groupings did not correspond to the allozyme genetic clusters. While the northern (9.2 %) and the central group (7.2 %) showed a smaller portion of the variation partitioned among populations compared to the species' total (15.3 %), the amount was much larger in the southern group (23.0 %).

At SSR loci a greater amount of variation was observed among groups than among populations within groups (17.6 % vs. 12.9 %) as expected for 'true' population clusters. The corresponding fixation indices were highly significant (*P* = 0). In all three groups a lower portion of the genetic variation was partitioned among populations compared to in the total species (15.5 %, 15.9 % and 15.5 % vs. 26.9 %).

As a comparison, only 18.7 % of the variation was partitioned among and 10.7 % within the five principle geographic groups (1 - 5) at SSRs, indicating that structuring was hardly more pronounced among five vs. the three groups. At allozyme loci more genetic variation resided among five groups vs. three (11.3 %) and less within each group (5.7 %), suggesting that the five-partite grouping was more accurate in that case.

S. sitiens. As much as a quarter (25.2 %) of the allozyme variation was distributed among groups and only 10 % among populations within groups, supporting the predicted clustering (Appendix 12B). The two fixation indices were highly significant (*P* = 0). Percent genetic variation among populations within groups was substantially lower than among populations of the entire species (10.4 % in the north and 16.7 % in the south vs. 30.2 % in the total species). The southern 'group' was composed of just one species, Escondida (LA4105), therefore no value is available.

Nineteen percent of the SSR variation was present among groups and only 6.5 % among populations within groups, in congruence with the structural pattern, and the fixation indices were highly significant (P = 0). The portion of among-population variation was reduced to almost one third compared to in the total species (7.1 % and 9.0 % vs. 21.5 %).

3.2.4.5 Genetic diversity within population substructures

To characterize the three main regions (north, center and south) more specifically, population groups were investigated for levels of genetic diversity.

S. lycopersicoides. At allozyme loci overall genetic diversity was highest in the northern group (P = 47.2 %, R_S = 1.47, H_E = 0.133), at SSR loci in the central group (P = 93.3 %, R_S = 3.20, H_E = 0.429; Figure 15). The number of private alleles was highest in the south for the allozyme loci (3 private alleles), and in the center at SSR loci (7 alleles). At allozyme loci the northern cluster demonstrated a slight excess in outcrossing (F_{IS} = -0.009), while the other two groups showed a substantial amount of inbreeding (F_{IS} = 0.115 and F_{IS} = 0.095).
The picture was reversed at microsatellite loci: only the northern group was inbred, and only slightly so (F_{IS} = 0.024), while the other two were more outcrossed (F_{IS} = -0.028 and F_{IS} = -0.034).
The average pairwise genetic distance among populations within each group (θ) was lowest in the northern region and rose towards the south according to both marker systems (θ = 0.077, 0.137 and 0.202 at allozyme loci and θ = 0.117, 0.127 and 0.192 at SSR loci). The high value in the south suggested a substantial amount of substructuring in that group relative to the others. Indeed, when the two locations in the southern group were considered separately, pairwise distance estimates fell below those of the two other regions (θ = 0.104 at Camarones and θ = 0.091 at Camiña). Therefore, the division of these two subgroups would be the next (third) split in the hierarchy of the *S. lycopersicoides* dendrogram.
Finally, the average global pairwise genetic distance (θ) was calculated for each group (i.e. the average genetic distance of the populations within a group to all other populations). At allozyme loci values descended from north to south (θ = 0.128 in the north, 0.161 in the center and 0.218 in the south), at SSR loci these estimates decreased from the center (θ = 0.241 in the center, 0.256 in the south and 0.293 in the north), indicating that the central location were the most ‘distinct’.

S. sitiens. Allozymes identified the northern cluster of accessions as the most diverse (P = 76.7 %, R_S = 2.24, H_E = 0.332; Figure 16), and two private alleles each were present in the northern and southern cluster. On the other hand, microsatellites indicated the highest levels of diversity in

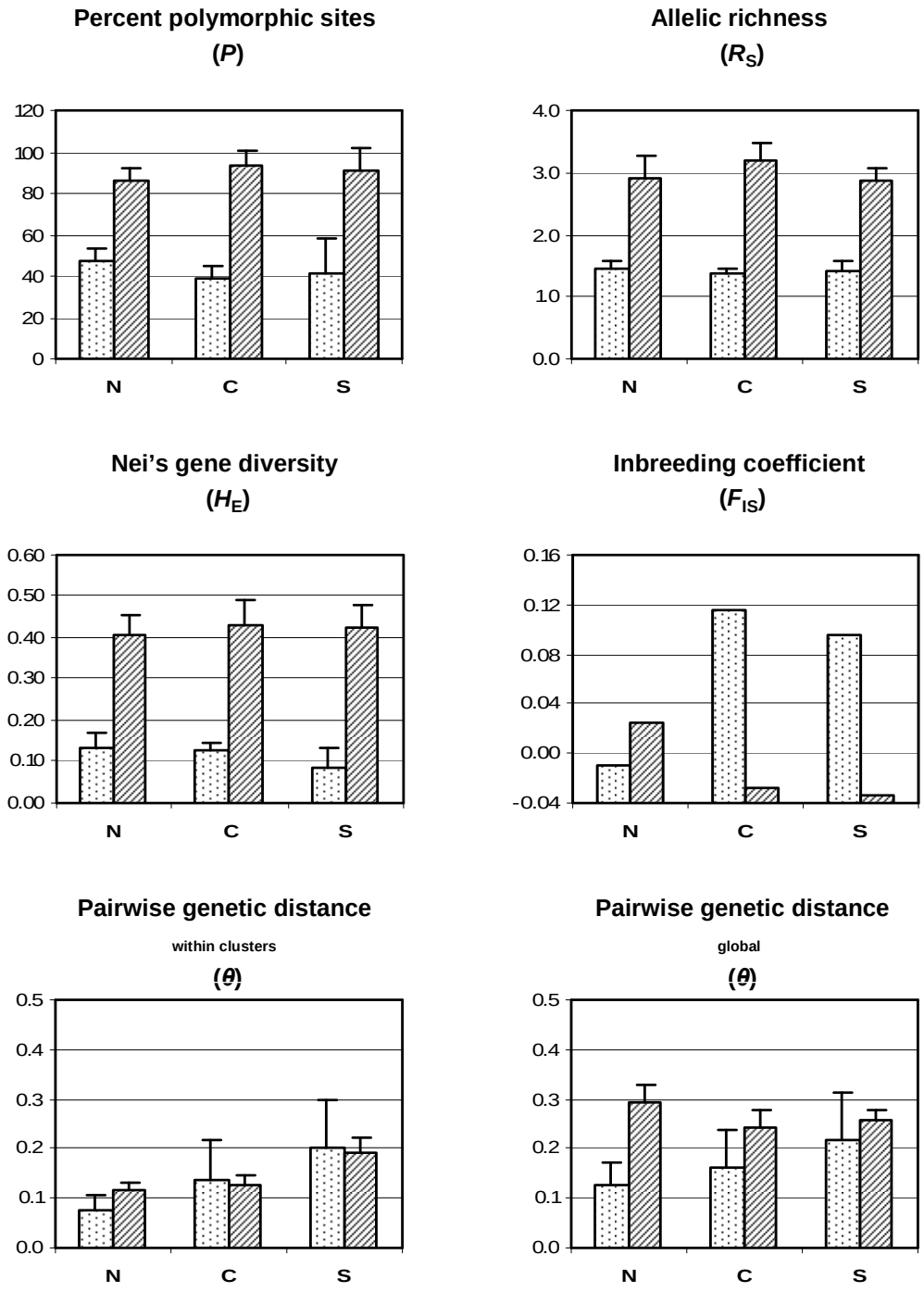


Figure 15. Levels of genetic diversity within population clusters of *S. lycopersicoides*.

P = percentage of polymorphic sites; R_S = allelic richness per population; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient per population; θ = average pairwise genetic distance; N = northern cluster; C = central cluster; S = southern cluster. Dotted areas represent allozyme data, striped areas microsatellite data.

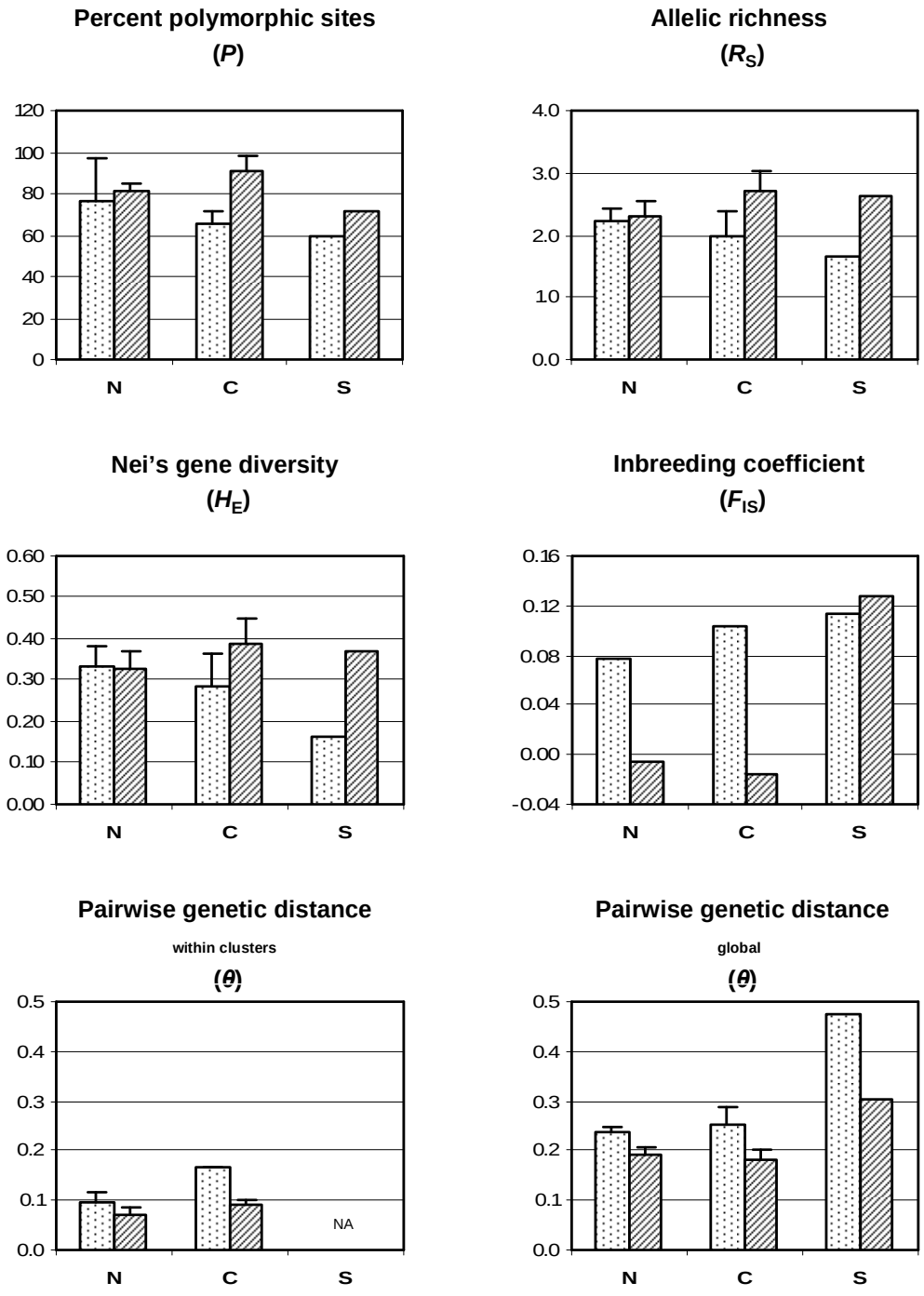


Figure 16. Levels of genetic diversity within population clusters of *S. sitiens*.

P = percentage of polymorphic sites; R_S = allelic richness per population; H_E = expected heterozygosity; F_{IS} = inbreeding coefficient per population; θ = average pairwise genetic distance; N = northern cluster; C = central cluster; S = southern cluster. Dotted areas represent allozyme data, striped areas microsatellite data.

the central accessions (P = 90.5 %, R_S = 2.72, H_E = 0.391) with the most private alleles (3) in the southern region which contained just one population (Escondida – LA4105).

The degree of inbreeding was considerably higher in the southern population (F_{IS} = 0.113 for allozymes and F_{IS} = 0.128 for SSRs), declined towards the north at allozyme loci (F_{IS} = 0.104 in the center and F_{IS} = 0.076 in the north), while it was approximately zero in those two regions at microsatellite loci (F_{IS} = -0.016 and -0.005).

The pairwise genetic distance among populations within groups was higher in the central region (θ = 0.166 at allozyme loci and θ = 0.089 at SSR loci) than in the northern region (θ = 0.098 and 0.073, respectively). The global pairwise genetic distance was similar for the center and the south, but substantially higher for the southern population, almost twice as high at allozyme loci (θ = 0.476 vs. 0.250 in the center and 0.235 in the north at allozyme loci, and θ = 0.304 vs. 0.182 in the center and 0.191 in the north at SSR loci).

4. Discussion

4.1 Comparative mapping analysis

The principal aim of the underlying experiment was to construct a comparative genetic linkage map for *S. ochranthum* and *S. juglandifolium* in order to assess the level of synteny with the closely related genome of tomato (*S. lycopersicum*), and other *Solanaceae* species.

The first part of the discussion is focused on characteristic features of the *Juglandifolium* linkage map, which are analyzed in the light of phylogenetic relationships and genetic reproductive mechanisms. In the second part of the discussion genome comparisons are made among *Juglandifolium* and other solanaceous species. Hotspots for rearrangements are identified, and a hypothetical evolution scenario is presented. Finally, practical implications from the results for tomato breeding are outlined.

4.1.1 Genetic linkage map for *Solanum* sect. *Juglandifolium*

An F_2 *S. ochranthum* × *S. juglandifolium* genetic linkage map was constructed with 132 markers (73 % CAPS, 13 % microsatellites and 14 % RFLPs). All of these were single-copy loci on the tomato map to reduce the likelihood of comparison of paralogous sites. Duplications in the *S. ochr.* or *S. jugl.* genomes cannot be ruled out, however.

Analysis of linkage identified 12 linkage groups, spanning a total of 761 cM. The linkage groups, however, did not correspond perfectly to the 12 tomato chromosomes. Tomato chromosome 1 was represented by two linkage groups that could not be linked (probably due to insufficient marker density in this region) and tomato chromosome 8 and 12 were fused into one large linkage group owing to a reciprocal translocation in one of the parental species. Disintegration of pseudolinkage resulted in two pairs of balanced chromosomes, one carrying the tomato configuration and the other the reciprocal whole-arm translocation between chromosome 8 and 12. Therefore, under assumption of both scenarios the present study proposes two putative maps for sect. *Juglandifolium*, one for *S. juglandifolium*, one for *S. ochranthum*, each covering 790 cM.

4.1.2 Species relationships within sect. *Juglandifolium*

The two sister taxa *S. ochranthum* and *S. juglandifolium* resemble each other closely both morphologically and with regard to habitat requirements (Rick 1988). It was therefore expected to find the close association reflected on the genetic level. Surprisingly, the results in this study provide evidence of substantial genetic divergence between them, including the aforementioned translocation, as well as reduced recombination.

4.1.2.1 Crossing barriers

Hybridizations between *S. ochranthum* and *S. juglandifolium* are successful only with deployment of special techniques (embryo rescue) to overcome postzygotic reproductive barriers that exist between the two species. Crossing barriers may arise from:

1) An imbalanced endosperm balance number (EBN) ratio: As hybrid plants can be obtained via implementation of embryo rescue techniques, the main difficulties appear to reside in the endosperm. Seed abortion due to failure of normal endosperm development is the principal cause for postfertilization barriers (Wann and Johnson 1963). In the absence of other reproductive barriers the maternal to paternal EBN ratio in the endosperm, normally 2:1, will determine the failure or success of a cross (Johnston et al. 1980). The mechanism secures species barriers in sympatry and has been shown to account for the separation of the two crossing complexes within sect. *Lycopersicon* (Rick 1979; Ehlenfeldt and Hanneman 1992; Moyle and Graham 2005).

2) Structural differences among the crossing parents: Chromosomal rearrangements can lead to sexual isolation (White 1978). The presence of rearrangements was indicated by linkage analysis (outlined in greater detail below) and supported by chromosome pairing behavior. Less than two thirds of the chromosomes paired as bivalents during meiosis of the F_1 hybrid. The observations were in striking contrast to those in F_1 *S. sitiens* × *S. lycopersicoides*, where bivalent pairing was almost complete and crosses are readily obtained in either direction without the aid of embryo culture (Rick 1979; Pertuzé and Chetelat 2002). In addition, F_1 pollen fertility in the *S. ochranthum* × *S. juglandifolium* hybrid (39 %) was only half of that of the *S. sitiens* × *S. lycopersicoides* hybrid (92 % and 69 % in both reciprocal crosses). Translocation heterozygosity is known to cause semisterility (Burnham 1962), therefore the observed low pollen fecundity is consistent with our evidence of structural differences between the parental genomes.

4.1.2.2 Sequence divergence

Restricted fragment length polymorphisms are caused by base substitutions or insertion and deletions, (of e.g. repetitive sequences; Dvorak and Akhunov 2005) and therefore represent a measure of divergence at the DNA sequence level. The herein observed RFLP polymorphism rate was 21 % (or 24 % when only enzymes in common to the experiment by Pertuzé et al. (2002) were considered), slightly lower than that reported for *S. lycopersicoides* and *S. sitiens* (27 %; Pertuzé et al. 2002) and suggested a closer relationship between *S. ochranthum* and *S. juglandifolium*. In comparison, the RFLP polymorphism rate was 63 % between *S. lycopersicum* and *S. pennellii* and 80 % between *S. lycopersicum* and *S. lycopersicoides*. The two species are unilaterally compatible with cultivated tomato; crosses succeed only when the wild species is used as pollen parent

(Chetelat et al. 2000). The relative difficulty in hybridizing *S. ochranthum* and *S. juglandifolium* is therefore surprising. However, crossability clearly does not vary in direct proportion to sequence relatedness. For example, cultivated tomato is more easily hybridized with *S. pennellii*, the basal taxon in the *Lycopersicon* clade, than with *S. peruvianum*, with which it shares a closer relationship based on molecular (allozymes, Breto et al. 1993; SSRs, Alvarez et al. 2001; ITS sequence data, Marshall et al. 2001; AFLPs, Spooner et al. 2005) and morphological phylogenies (Peralta et al. 2005).

Comparisons of granule-bound starch synthase (*waxy* gene) sequence data, generated by Peralta and Spooner (2001), and available at NCBI (http://www.ncbi.nlm.nih.gov), were in agreement with divergence estimates from RFLP polymorphism rates: A slightly higher level of sequence homology was revealed for *S. ochranthum*/*S. juglandifolium* (99.7 %) compared to *S. sitiens*/*S. lycopersicoides* (98.5 %), which was nearly equal to that of *S. lycopersicum*/*S. pennellii* (98.5 %).

4.1.2.3 Recombination suppression

The herein reported map (790 cM) was substantially shorter than the tomato reference map tomato-EXPEN 2000 (42 %) and the F_2 *S. sitiens* × *S. lycopersicoides* map (33 %; Pertuzé and Chetelat 2002). Map length compression can be caused by a variety of factors.

a) In areas of low marker density double cross-over events may escape detection. Although marker density was ten times lower compared to the reference map (6.0 vs. 0.6 cM/marker), it is unlikely that it is causative of the disparity in map lengths. Marker density was similar to that in F_2 *S. sitiens* × *S. lycopersicoides* (Pertuzé and Chetelat 2002) where the genome-wide map reduction amounted to only a few percentage points. In addition, if double cross-over events were accountable for the map size compression, larger intervals should be more severely affected; a prediction not consistent with the empirical data presented herein. Conversely, higher density maps may be artificially inflated by scoring errors. However, as the reference map tomato-EXPEN 2000 is similar in length to tomato maps of lower densities, this possibility can also be excluded as the principal cause for the length differences.

b) Markers tend to cluster around centromeres (Tanksley et al. 1992). Enhanced clustering of markers at putative centromere positions was observed on two thirds of all chromosomes (chr. 2, chr. 4, chr. 5, chr. 6, chr. 7, chr. 8, chr. 10 and chr. 12). However, the effect on the over-all map length should be marginal because centromeric regions only represent a small portion of the total map units in the genome.

c) It is well-established that sequence divergence among the parents leads to reduced recombination in wider crosses (Rick 1969; Bonierbale 1988; Gebhardt et al. 1991; Burke et al.

2004). This option for *Julgandifolium* is discussed below. Special attention should be directed to areas where recombination was eliminated completely, because this may be diagnostic for inversion heterozygosity, which rarely produces viable recombinants (Livingstone and Rieseberg 2003). The short arm of chromosome J10 was reduced to just 0.8 cM. The residual recombination may represent a genotyping artifact or J10S may be rearranged among the parental species. Similar situations were found on J1S and J3S (87 % and 98 % reduction, respectively), but as chromosome 1 is subtelocentric, it is difficult to determine whether the observed shrinkage – deduced from just three markers - goes beyond the average genome-wide reduction. Likewise, the size reduction on J3S was based on only two markers.

d) Map length reduction may be a by-product of segregation distortion, e.g. when recombinant gametes are eliminated by selection, and actual crossover rates may be more normal (Rick 1969). Hence, segregation distortion may be responsible for the map length compressions around *sd1.1*, *sd2.2* and *sd5.1*, and sequence divergence may not be greater in these areas with respect to other regions in the genome. However, as sequence divergence itself can also be causative of segregation distortion (Grant 1975; Zamir and Tadmor 1986), the causal link as outlined above may only apply to areas that are under the control of segregation distorter loci.

e) Map size reductions may also reflect differences in gene or DNA content. Given the conservativeness in gene content in related species (outlined in greater detail below) this explanation does not seem very likely.

f) The suppression could be the work of genes that control recombination frequency, as have been identified in petunia (*Petunia hybrida*), another *Solanaceae* (Maizonnier et al. 1984). The petunia genetic map is ~ ten-fold smaller than that of tomato (Strommer et al. 2002) while its haploid genome content is larger (1200 Mb vs. 950 Mb; Arumuganathan and Earle 1991).

g) Finally, recombination could be reduced in one gamete, probably the male, in the present mapping cross. The complete lack of male recombination would cause 50 % map shrinkage, similar to the observed. The phenomenon, first reported for *Drosophila*, has been detected among hybrid progeny of tomato and *S. pennellii* (de Vicente and Tanksley 1991) and tomato and *S. peruvianum* (van Ooijen et al. 1994). Hence, this explanation may be particularly attractive to explain the present map size reduction. Other factors, however, e.g. structural and sequence divergence, segregation distortion and/or recombination modifiers may also have contributed to the outcome in *S. ochranthum* × *S. juglandifolium* .

4.1.2.4 Chromosomal restructuring

Linkage analysis provided evidence that the two species within sect. *Juglandifolium* are separated by a reciprocal whole-arm translocation between chromosome 8 and 12. A reciprocal translocation in one of the parents of a mapping population will lead to pseudolinkage between markers close to the interchange breakpoints (Burnham 1991). On the *S. ochranthum* × *S. juglandifolium* linkage map two markers (SSR15 and C2_At42740), located on tomato chromosome 8 and 12, were connected by strong linkage. Both markers map near the centromere in tomato, a region prominent for chromosome breakage in many species (Tanksley et al. 1992; Moore et al. 1997). In addition, semisterility and irregular chromosome pairing during the F_1 meiosis, two characteristic features of translocation heterozygosity (Burnham 1962), were also observed in the present study. Future analysis may determine which of the species is the carrier of the translocation.
Further analyses are needed to investigate whether the severe map size reduction on the short arm of chromosome 10 is caused by structural rearrangements, e.g. an inversion among the parental species. Likewise, the presence of whole-arm inversions in *Juglandifolium* with respect to tomato on chromosome 3S, 6S and 8S requires substantiation.
In view of the high degree of collinearity, the genomes of *S. ochranthum* and *S. juglandifolium* can be considered homeologous.

4.1.2.5 Speciation within Sect. *Juglandifolium*

The data suggest that the genomes of *S. ochranthum* and *S. juglandifolium* are primarily differentiated by a reciprocal whole-arm translocation during a relatively recent divergence (i.e. potentially even more recent than the *S. sitiens*/*S. lycopersicoides* divergence) that did not allow the accumulation of many mutations, thereby preserving a high level of homology at the DNA sequence level (as indicated by RFLP polymorphism rates and sequence divergence at the GBSSI locus). A close relationship would also be in agreement with morphological similarities between the two species, however, other lines of evidence (recombination suppression, segregation distortion, crossing barriers) suggest a more distant relationship, but these may also be caused by factors that do not reflect phylogenetic distance.
Doganlar et al. (2002a) reported that at least over larger evolutionary times in the *Solanaceae* family the number of chromosomal rearrangements that differentiate lineages corresponds to the amount of change on the nucleotide sequences level.
Unless position effects, which are assumed to be rare in higher plants (Burnham 1962), play a role, as long as the gene content is preserved, rearrangements generally do not result in apparent phenotypic changes (Lande 1979). Therefore the amount of chromosomal rearrangements cannot be

taken as a measure for morphological diversity and, conversely, the amount of phenotypic evolution is not a good predictor for the amount of genomic evolution (Dobzhansky 1972; Carr 1977). Hence, the strong morphological resemblance of the two *Juglandifolium* species is not in conflict with structural heterogeneities and, by extension the existing crossing barriers.

Assuming sympatric or parapatric conditions at the time the two lineages commenced to diverge (as the two species are overlapping in their present-day distribution; Rick 1988; Smith and Peralta 2002; http://tgrc.ucdavis.edu), there are several possibilities how a speciation scenario may have played out in *Juglandifolium*. For speciation to occur in sympatry it is a prerequisite that reproductive barriers be erected between the diverging lineages (Dobzhansky 1937) in order to avoid subsequent merging of the incipient species with their progenitor. One way to achieve reproductive isolation is via chromosomal rearrangements (White 1978). There are many accounts in the scientific literature where chromosomal rearrangements have been identified as causal factors in speciation processes of plants and animals (Livingstone and Rieseberg 2004). Likewise, rearrangements have been shown to be more common among sympatric versus allopatric species (Noor et al. 2001), suggesting that they may be advantageous under these conditions, possibly owing to their isolating effect. The situation, however, is paradoxical because the stronger the negative fitness effect of a chromosomal rearrangement, the greater is its potential to confer reproductive isolation, but at the same time the smaller is its chance for fixation (i.e. the strongest barriers have the least chance to be fixed). This dilemma has been subject of much debate and sparked the development of several models (Livingstone and Rieseberg 2003; Burke 2004).

It has often been proposed that rearrangements are deleterious in the heterozygous state, but not in the homozygous state (Levin 2002). Some theories stress that the underdominant fitness effect of rearrangement heterozygotes is instrumental in the speciation process through the erection of crossing barriers (Dobzhansky 1937). In self-incompatible species chromosomal rearrangements with underdominant effects may require genetic drift to be brought to fixation, making them only likely to occur in populations with small effective population sizes (Lande 1979; Lagercrantz 1999; Burke et al. 2004). Most known *S. ochranthum* and *S. juglandifolium* populations are small in size, and their geographic distribution is highly fragmented (Roger Chetelat, pers. comm.; http://tgrc.ucdavis.edu; Smith and Peralta 2002), therefore it is likely that genetic drift constitutes an important evolutionary force in these species.

Alternatively, rearrangements may become established because their carriers are spared from competition with their progenitors, as a consequence of either reproductive isolation or differential survival, e.g. by means of special adaptations (Levin 2000; Burke 2004).

Other models predict that chromosomal rearrangements have no or just a small effect on fitness, and that reproductive barriers are built up gradually in the presence of gene flow (Burke et al. 2004; Livingstone and Rieseberg 2003). One of these models proposes that single rearrangements may only be associated with weak underdominance, thereby allowing the carrier's survival within a population, but that a combination of several rearrangements will eventually cause reproductive isolation (Burke et al. 2004). Alternatively, weakly underdominant rearrangements may be involved in speciation in sym-/parapatry primarily through their effect on recombination (Rieseberg 2001; Noor et al. 2001). Sheltered from gene flow, isolating factors such as Dobzhansky-Mueller incompatibilities have the chance to accumulate in the rearranged region until eventually the reproductive barrier is complete (Navarro and Barton 2003a).

Empirical evidence that rearrangements may persist within the original population prior to splitting into a new lineage was provided by interspecific sequence analyses that revealed greater divergence in rearranged regions compared to collinear regions (Navarro and Barton 2003b; Schaeffer et al. 2003). Among potato and tomato Livingstone and Rieseberg (2003) observed increased sequence differentiation on 10L relative to non-rearranged chromosome locations. They concluded that the rearrangement may have facilitated the origination of the tomato lineage in parapatry.

In summary, based on the evidence of overlapping distributions, translocation heterozygosity, and postzygotic reproductive barriers, any of the above models could potentially be invoked to explain the speciation process in *Juglandifolium*. However, as the 'recombination-model' resolves the "dilemma" associated with reproductive isolation in sympatry without requiring the presence of other conditions (e.g. genetic drift or accumulation of several rearrangements), and because it is supported by empirical evidence, it is deemed the most probable. Sequence analyses within critical genomic regions in *S. ochranthum* and *S. juglandifolium* could elucidate the possibility of evolution in sym~/parapatry via chromosomal rearrangements for the two species. In case of sympatric speciation, one might hypothesize further that the split of *S. ochranthum* and *S. juglandifolium* from a common ancestor occurred in an area they cohabit, which are currently parts of Ecuador.

4.1.3 Segregation distortion

Significant departures from the expected 1:2:1 Medelian ratio were observed in over one third of the loci in F_2 *S. ochranthum* × *S. juglandifolium*, affecting nine out of twelve chromosomes.

Non-Mendelian segregation is a widespread phenomena among a diverse range of taxa including fungal, plant and animal systems (Taylor and Ingvarsson 2003). Forms and mechanisms through which segregation distortion is achieved can be manifold but show common features. In general, segregation distorter loci, (also known as 'selfish genetic elements', 'non-Mendelian element' or

'outlaw genes' control the preferential transmission of a genotypic class, thereby disobeying Mendel's laws (Dawkins 1982; Taylor and Ingvarsson 2003). Segregation distorter loci can cause speciation (Hurst and Werren 2001) and are therefore widely regarded as an evolutionary force (Sandler and Novitski 1957; Taylor and Ingvarsson 2003).

Distorted segregation has often been observed in interspecific crosses between crop plants and their wild relatives, including tomato (Zamir and Tadmor 1986; Wendel et al. 1987; Bonierbale et al. 1988). The extent of aberrant segregation generally increases with the level of divergence between parental species (Grant 1975; Zamir and Tadmor 1986). In general, segregation patterns observed in intraspecific crosses of *S. lycopersicum* are in compliance with Mendelian laws (Rick 1948) although there are also examples of genes that do not transmit in Mendelian fashion, due to selection in various stages of development. In a cross between *S. lycopersicum* and its close relative *S. pimpinellifolium,* 8 % of loci showed skewed segregation (Grandillo and Tanksley 1996), whereas 51 % and 69 % of the loci deviated from expected Mendelian ratios in crosses between *S. lycopersicum* and *S. cheesmaniae* and *S. lycopersicum* and *S. neorickii*, respectively (Paterson et al. 1988, 1991). In crosses with the more distantly related *S. pennellii* and *S. chilense*, 60 % and 80 % of markers showed skewed segregation (Zamir and Tadmor 1986).

The herein reported extent of non-Mendelian segregation (32 %) was also similar to that observed in a cross between *S. sitiens* × *S. lycopersicoides* (24 %; Pertuzé et al. 2002). However, a direct comparison, i.e. inferring a slightly greater divergence for the pair *S. ochranthum*/*S. juglandifolium* may be inappropriate because a) estimates were derived from different marker systems, and b) interspecific hybrids and F_2 progeny were relatively difficult to obtain in the present study, suggesting stronger selection pressures.

A total of thirteen putative segregation distorter loci were identified, similar to the fourteen detected in the genomic background of *S. sitiens* × *S. lycopersicoides* (Pertuzé et al. 2002). This number is probably a conservative estimate because effects of minor distorter loci would be hidden by those of major distorter loci in neighboring areas of the chromosome. Distorter loci may also go undetected because they are driven to fixation before they have a chance of being observed (Taylor and Ingvarsson 2003). Zamir and Tadmor (1986) observed a positive correlation between the decrease of donor genome content and the decline of segregation distortion in interspecific backcrosses, concluding that a high level of homozygosity characterizes loci associated with the regulation of reproduction at the intraspecific level. Distorter loci may also go unnoticed because their effects are deleterious and prompt the selection for modifier genes at secondary loci that suppress them, a situation also known as 'genetic conflict' (Taylor and Ingvarsson 2003).

Sterility factors causing skewed segregation were identified in several crop species. Pollen killers that abort pollen carrying the non-driving allele have been reported in tobacco, wheat and rice (Cameron and Moav 1957; Loegering and Sears 1963; Sano 1983). So-called gamete eliminators render only those gametes dysfunctional which contain the alternate allele in the heterozygous sporophytic parent. Gamete promoters operate in the reverse fashion. Both have been observed in tomato (Rick 1966; Pelham 1968). Common features include incomplete penetrance (Rick 1966) and expression variation among different genetic backgrounds (Loegering and Sears 1963). Sano (1990) showed that pollen killers may become gamete eliminators through interaction with modifiers, thereby causing hybrid sterility, which poses a major problem for the introgression of distantly related germplasm in breeding programs. Hence, the accumulation of sterility factors and modifier genes plays a central role in the establishment of reproductive barriers and, by inference, speciation.

Distorter loci may exert selection in the pre-zygotic phase during gametophyte development and the pollination process or at the post-zygotic stage from seed and embryo formation, germination and plant growth (which included embryo culture in the present study) until the moment plant material is harvested (Grant 1975). In *S. ochranthum* × *S. juglandifolium* selection operated in both phases with preference for both allelic types (i.e. the *ochranthum* allele or the *juglandifolium* allele) and on all three possible genotypes (both parental homozygotes or the heterozygote).

4.1.3.1 Conservation of segregation distorters among closely related species

For several of the segregation distorter loci detected in this study, counterparts with analogous trends were also found in the genomic background of other tomato species, suggesting that they represent true distorter loci that are of importance for fertility regulation and that their positions and functions have been conserved across tomato taxa.

Sd2.2 mapped in close proximity to a segregation distorter locus in BC_1 *S. lycopersicum* × *S. lycopersicoides* (Chetelat et al. 2000). Regions of distorted segregation on the long arm of chromosome 5 were also identified in both crosses (Chetelat et al. 2000).

Most prominent was the great amount of similarities among genome-wide patterns of segregation distortion in this cross and that of the closely related sister-taxa pair *S. sitiens* × *S. lycopersicoides* (Pertuzé and Chetelat 2002), especially with respect to deviation patterns on chromosome 4 and 7. Several segregation distorters were identified on both chromosomes in F_2 *S. sitiens* × *S. lycopersicoides* but only the most significant ones were also detected in the present study. The segregation distorter locus at the distal end of chromosome 4 (*sd4.1*) appeared to be identical to the one reported in *S. sitiens* x *S. lycopersicoides*, albeit its action was attenuated in *S. ochranthum* × *S.*

juglandifolium. Rick (1966) reported the presence of a gamete eliminator (*Ge*) mapping to the centromeric region of chromosome 4 on the tomato classical map, approximately at the same position as *sd4.1*. Both location and mode of action called for the identity of the herein reported *sd7.1* and the segregation distorter locus in the centromeric region on chromosome 7 in *S. sitiens* × *S. lycopersicoides*.

On chromosome 8 the same marker (TG510) that detected one of the two distorter loci that promote the heterozygous class (*sd8.2*) also detected a locus of similar action in *S. sitiens* × *S. lycopersicoides*, pointing to the presence of a gamete promoter factor at this locus. Strong deviations on chromosome 9 were also reported for *S. sitiens* × *S. lycpersicoides*. The herein detected *sd9.1* mapped to the same location as *sd9.2* in *S. sitiens* × *S. lycopersicoides*, thus suggesting locus identity. The mode of action, however, showed some differences. A negative selection against one homozygous class was more severe in *S. sitiens* × *S. lycopersicoides*. Also, in the latter the heterozygote was the primarily favored genotype whereas in the present study the alternative homozygote was as abundant as the heterozygote. Therefore, present data suggest gene action at the pre-fertilization stage, rather than at the post-fertilization stage at this locus.

The locations of the two segregation distorters observed herein (*sd9.1* and *sd9.2*) correspond to the two gamete promoters at the centromeric region and the distal end of chromosome 9 reported by Pelham (1968) and Fulton et al. (1997) in crosses between *S. peruvianum* and cultivated tomato. Strong segregation distortion in favor of the wild allele along the entire chromosome 9 was also observed in a cross between *S. lycopersicoides* and tomato (Chetelat et al. 2000).

In conclusion, *S. ochranthum* and *S. juglandifolium* appear to be slightly less divergent than *S. sitiens* and *S. lycopersicoides* according to RFLP polymorphism rate and GBSSI data. On the other hand, the extent of recombination suppression (and maybe also of segregation distortion) may exceed the expectations for such a close relationship, but could be explained by alternative causes, which are outlined above (such as lack of recombination in one gamete). Likewise, the closer morphological resemblance of the *Juglandifolium* species compared to that of the *S. sitiens*/*S. lycopersicoides* pair is suggestive of a more recent divergence, while the stronger crossing barriers could indicate a more distant relationship of *S. ochranthum* and *S. juglandifolium*. However, morphological differences or crossing relationships, when taken alone as indicators for species divergence, can be misleading. For example, genetic differentiation is not correlated with reproductive compatibility in *S. chilense*, where adjacent groups show stronger barriers than distant groups (Graham 2005). (Reproductive barriers in that species might be reinforced among adjacent types and relaxed among allopatric types.) Similarly, *S. pennellii* as well as the species in *Solanum* subsect. *Lycopersicodes* are less reproductively isolated from tomato than the more distantly related

S. peruvianum and the *Juglandifolium* species, respectively (Rick 1979; Pertuzé et al. 2002, 2003; Spooner et al. 2005).

4.1.4 Genome comparisons with tomato and other *Solanaceae*

4.1.4.1 Genome content and recombination suppression

The herein reported map for *S. ochranthum* and *S. juglandifolium* (790 cM) encompassed only 58 % and 67 % of the mapping units in tomato-EXPEN 2000 (1363 cM) and F_2 *S. sitiens* and *S. lycopersicoides* (1192 cM), respectively (http://www.sgn.cornell.edu; Pertuzé et al. 2002). Although the over-all net reduction of the latter relative to the reference map amounted to only a few percentage points, variation was large across chromosomes, ranging from over 24 % size reduction (chromosome 8) to almost 13 % size expansion (chromosome 11). Interestingly, most potato genetic maps show a substantial size reduction in comparison to tomato genetic maps (Bonierbale et al. 1988; Gebhardt et al. 1991; Tanksley et al. 1992), e.g. the Potato-TXB map spans 684 cM (Tanksley et al. 1992), and the one developed by Bonierbale (1988) 606 cM. However, Bonierbale et al. (1988) and Gebhardt (1991) attributed the size discrepancies to sequence divergence among the mapping parents rather than to a generally lower level of recombination in potato. Several lines of evidence (segregation distortion, crossing barriers) suggested structural differences between the two species. However, the phylogenetic distance within *Juglandifolium* does not seem larger than that of the parental species of the other crosses: divergence estimates from RFLP polymorphism rates were three times higher among parental accessions of the tomato map (63 %) compared to the *Juglandifolium* map (21 %; Chetelat et al. 2000), and GBSSI sequence data (Peralta and Spooner 2001) indicated lower levels of sequence divergence among the *Juglandifolium* species compared to the species pairs *S. sitiens*/*S. lycopersicoides* and *S. lycopersicum*/*S. pennellii*. Map sizes in eggplant (*Solanum melongena*) and pepper (*Capsicum annuum*), on the other hand, are very similar to those of tomato (1480 cM and 1246 cM, respectively; Livingstone et al. 1999; Doganlar 2002a). The genome content of potato differs only slightly from that of tomato and eggplant (C = 0.88 vs. C = 1.03 and C = 0.98, respectively), whereas the pepper genome is four times larger (C = 4.0; Bennett and Leitch 2004), indicating that map size is a poor indicator for genome size. As the extra genome content in pepper is presumed to be primarily composed of retrotransposons (Livingstone et al. 1999), and recombination is mainly restricted to homologous genes (Thurieaux 1977) the map size reduction in *Juglandifolium* could also be the result of a lower gene content. The gene content of the two species is unknown. However, gene content has been quite conserved even over larger evolutionary distances in the *Solanaceae* (Livingstone et al. 1999), making this explanation less likely.

Therefore, recombination suppression in *Juglandifolium* may be caused by a combination of factors (which were stated earlier), including reduced male recombination, recombination modifiers, sequence divergence and/or segregation distortion.

Besides the genome-wide map size reduction, relative recombination rates (i.e. stretches delimitated by common markers) were similar in F_2 *S. ochranthum* × *S. juglandifolium* and the tomato map, reflecting the high level of synteny between the genomes. This finding was not surprising given that Doganlar (2002a) detected a significant positive correlation of recombination frequencies even among the far more distantly related eggplant and tomato.

4.1.4.2 Collinearity with tomato

In accordance with expectations the comparative mapping experiment revealed a high level of synteny between the genomes of sect. *Juglandifolium* and tomato (*S. lycopersicum*), i.e. gene order was almost perfect. Highly preserved gene orders within rearranged blocks are a common finding even in broader comparisons such as those of tomato, potato, eggplant and pepper (Bonierbale et al. 1988; Tanksley et al. 1992; Livingstone et al. 1999; Doganlar et al. 2002a). The retention of large intact blocks is not only a feature of *Solanaceae* genomes, but, as many reports indicate, a widespread characteristic of plant genome evolution. (Paterson et al. 2000 and references therein).

4.1.4.3 Chromosomal restructuring

A maximum of seven chromosomal rearrangements was reported in this study to explain the genome differentiation between tomato and *Juglandifolium*. One of the *Juglandifolia* species (i.e. either *S. ochranthum* or *S. juglandifolium*[4]) is presumed to carry a reciprocal whole-arm translocation among chromosome 8 and 12 relative to tomato. In addition, in six cases marker order deviated from that of tomato. Two loci mapped to different chromosomes (T0308, in tomato on chromosome 10L, was positioned at the distal end of J3S and TG581, in tomato on chromosome 6L, was placed on J4S in *Juglandifolia*), and one locus appeared translocated within J10L. Flipped positions of adjacent loci were suggestive of whole-arm inversions: on J3S, J6S and J8S. These erosions of colinearity are based on single-marker evidence only, and may hence represent artifacts, caused by, e.g., limited mapping power (resulting from the small population size), reduced recombination in the F_2 *S. ochranthum* × *S. juglandifolium* population, mapping of secondary or duplicated loci and/or genotyping errors. For substantiation additional cytological or mapping data is required.

[4] Although considered less likely, the possibility that the translocation is a unique feature of the parental accession and does not represent the species as a whole cannot be discounted.

A recent FISH analysis showed the inversion of 6S among the genomes of tomato and potato (Bai et al. 2007). If 6S in *Juglandifolium* is confirmed to share the potato configuration, then this inversion will represent the most recent (known) rearrangement that occurred in the tomato lineage. Interestingly, *S. peruvianum* bears an inverted stretch of ~ 300 kb on 6S which is associated with nematode resistance (Seah et al. 2004), indicating that the inversion – or parts of it – may also be present at 'lower' branches inside the tomato clade.

Also noteworthy is the severe recombination suppression on J10S (96 % map size reduction) which may be indicative of a putative whole-arm inversion, as mentioned above. An even stronger effect was seen on J3S, albeit solely based on the distance among two loci.

Still, the number of the herein observed rearrangements is likely to be an underestimate of the true extent of chromosomal repatterning that differentiate the genomes of the two lineages. With an average marker density of 6 cM mapping resolution is relatively coarse, and smaller rearrangements are unlikely to be detected. For example, an inversion of the top of chromosome 7 among tomato/*S. pennellii* became evident only via fine-mapping and FISH analysis (van der Knaap et al. 2004).

Also, as chromosomal rearrangements often pose an obstruction to interfertility – albeit they do not unavoidably cause incompatibility (Levin 2002) – crossing behavior among the two and between *S. ochranthum* and tomato are well in line with structural differences.

The number of putative rearrangements in the *Juglandifolium* lineage is surprising in light of the relative structural conservativeness that characterizes the rest of the tomato clade. All genomes within sect. *Lycopersicon* are essentially uniform, differentiated only by small rearrangements and gene substitutions, and are considered homologous (Rick 1979; Tanksley et al. 1992; Paran et al. 1995; Grandillo and Tanksley 1996; Bernacchi and Tanksley 1997; Fulton et al. 1997; van der Knaap et al. 2004). A total of just six paracentric whole-arm inversions separate the genomes of potato and tomato (Tanksley et al. 1992; Bai et al. 2007). Four or five of these occurred in the tomato lineage: The inversions on tomato chromosome 6S, 9S, 10L and 11S are likely derived in tomato because the potato configuration is shared by pepper and eggplant (9S and 10L) or eggplant alone (11S; the state of 6S is unknown for pepper). Evidence suggests that the inversion of 6S occurred after the split of *Juglandifolium*/*Lycopersicon* clades (Seah et al. 2004; the present study). However, the configuration in subsect. *Lycopersicoides* is currently unknown (Pertuzé et al. 2002). The paracentric inversion of tomato 12S, on the other hand, is shared by tomato and eggplant, indicating that this represents the ancestral state and that a rearrangement occurred in the potato lineage. The paracentric inversion on tomato 5S, however, could have occurred in either the potato or the tomato ancestors because neither of the two states is shared by eggplant or pepper (Tanksley et al. 1992; Livingstone et al. 1999; Doganlar 2002a).

Subsect. *Lycopersicoides* is separated from tomato by a single paracentric whole-arm inversion. S10L shares the potato configuration and is therefore assumed to represent a more recent rearrangement in the basal ranks of the tomato clade (Pertuzé et al. 2002). Eggplant (*S. melongena*) is the most distantly related *Solanum* species, vis-à-vis tomato, for which a comparative genetic map exists; it differs from tomato by a total of 28 rearrangements (23 paracentric inversion and five translocations; Doganlar et al. 2002a). The pepper lineage (*C. annuum*) in genus *Capsicum* has undergone extensive restructuring, presumably due to its high content of transposable elements. Pepper differs from tomato by a total of 22 breaks, composed of eight paracentric, two pericentric inversions, five translocations as well as various forms of dis~/ or associations, some of which are multiply nested (Livingstone et al. 1999).

Hence, given the evolutionary distance, the number of putative rearrangements reported herein for *Juglandifolium* seems high. However, evolutionary rates may vary greatly even among lineages within the same family, and the number of changes cannot be equalized with evolutionary divergence time, as has been shown in the grasses (Gale and Devos 1998).

Also noteworthy are the rearrangement classes observed in sect. *Juglandifolium*. Paracentric inversions and small translocations are common in the tomato and potato lineages (Bonierbale et al. 1988; Tanksley et al. 1992; Pertuzé et al. 2002; Bai et al. 2007), and *S. etuberosum* contains a number of rearrangements compared to potato (Perez et al. 1999). However, there are few large, whole-arm translocations of the type reported herein. Translocations do appear to have played an important role during the evolution of other *Solanaceae* species such as eggplant and pepper (Livingstone et. al. 1999; Doganlar et al. 2002a). In tomato it was shown that translocations induced by irradiation are stable as homozygotes and transmit to the next generation (Gill et al. 1980).

Frequencies with which rearrangement types occur seem to be largely conserved even across remotely related taxa. The leading role of paracentric inversions appears to be a widespread phenomena in both plant and animal systems (Ranz et al. 2001; Doganlar et al; 2002a) The tomato/potato lineages are differentiated exclusively by paracentric inversions (Tanksley et al. 1992; Bai et al. 2007). Of the inversions that occurred during the tomato/eggplant and tomato/pepper divergence 100 % and 83 % are paracentric, respectively (Livingstone et al. 1999; Doganlar et al. 2002a). Among structural rearrangements paracentric inversions are suspected to convey the least selective disadvantage; in *Drosophila*, e.g., a mechanism during female meiosis disposes unbalanced recombination products of paracentric inversions into polar bodies (Navarro and Barton 2003a). In contrast, pericentric inversions appear to be extremely rare and are associated with a stronger selective disadvantage than other rearrangement types (Burnham 1962), although they do not produce a higher degree of semisterility than reciprocal translocations (Navarro and

Ruiz 1997). In the heterozygous state translocations cause semisterility and are more detrimental than inversions (Burnham 1962), nonetheless they appear at a frequency intermediate to that of para- and pericentric inversions in the pepper/eggplant/potato/tomato divergence (Doganlar et al. 2002a). The pepper species *C. annuum* and *C. chinense* are differentiated by a reciprocal translocation Livingstone et al. (1999).

Rearrangement break points are not randomly distributed throughout genomes. Centromeric and telomeric regions have often been reported to be prominent spots for chromosomal breakage and fusion, e.g. in dicots such as the *Solanaceae* and *Brassicaceae* (Tanksley et al. 1992; Lagercrantz 1998) as well as in monocots (Moore et al. 1997). Within the *Solanaceae*, however, disruption patterns vary among lineages. Centromeric breakpoints are not more common than non-centromeric breakpoints in the eggplant/potato/tomato divergence, whereas all rearrangements separating potato and tomato are centromeric (Tanksley et al. 1992; Doganlar et al. 2002a). The heterochromatin surrounding centromeres has been associated with chromosome breakage (Khush and Rick 1963; Roberts 1965; Gill et al. 1980). Inversions may be triggered by homologous recombination between repetitive sequences within the heterochromatin of pericentromeric regions and of telomeres (Tanksley et al. 1992), and inverted chromosome arms are subsequently capped with new telomeric repeats to reestablish their stability (Yu and Blackburn 1991). Traces of the original telomeric repeats are maintained at proximal positions in the form of interstitial telomeric repeats (Presting et al. 1996). If homologous recombination is functional in the creation of rearrangements, then regions that harbor repeats should serve as hotspots for structural changes. The increased flexibility seen in the transposable element-rich genomes of pepper and *Drosophila* are in support of this notion (Engels and Preston 1984; Robbins 1989; Livingstone et al. 1999).

Observations in *Juglandifolium* are in agreement with these findings; the reciprocal translocation, three out of four of the observed single-marker translocations, and three stretches of near zero recombination (on J1S, J3S and J10S) involved or were delimited by centromeric positions. In a comparison among *Brassicaceae* genomes Lagercrantz (1998) observed that single transposed, "deviant" loci do not represent – as it may seem at first sight – a random disturbance of collinearity but often collocate with junction points of conserved blocks. Findings herein were in agreement with that prediction: The two interchromosomally transposed loci both derive from centromeric regions in tomato, and one of them assumed a telomeric position in *Juglandifolium* (T0308).

In conlusion, evolution in *Juglandifolium* may be considered atypical for the tomato lineage, but not for the *Solanaceae* family as a whole, both with respect to number and type of rearrangements. Differences may reflect genome-specific abilities to fix rearrangements (Devos and Gale 2000) or external conditions during speciation: The tomato species are assumed to have evolved primarily

through geographic isolation and adaptation (Peralta and Spooner 2005), consistent with the overall colinearity of species in the *Lycopersicon* clade and the absence (among some species combinations) of strong crossing barriers, (Some species do show strong reproductive barriers, particularly in cases where geographic ranges overlap; e.g. *S. peruvianum* with *S. hirsutum, S. pennellii*, and *S. pimpinellifolium*). In contrast, *S. ochranthum* and *S. juglandifolium* most likely originated via speciation in sym~/parapatry, which is often associated with higher occurrences of chromosomal rearrangements (White 1978; Noor et al. 2001).

4.1.4.4 Hotspots for rearrangements in *Juglandifolium* and other *Solanaceae* genomes

Chromosome 1 The *S. ochr.* × *S. jugl.* map is split into two linkage groups around the putative centromeric region, presumably due to a lack of marker saturation. In pepper, that same area marks the breakage point of a translocation with chromosome 8. Genome order is remarkably well preserved in all six genomes.

Chromosome 2 Marker content on chromosome 2 is well preserved among all species, although eggplant and pepper show some areas of disturbed marker order with respect to tomato, *Juglandifolium, Lycopersicoides* and potato.

Chromosome 3 J3S was inverted, severely truncated and/or suppressed with respect to tomato. This may represent an artifact or may at least be exaggerated due to marker scarcity. However, there was also evidence for further chromosomal restructuring in this area as T0308, a locus in proximity of the centromere on tomato 10L, mapped to the distal point on J3S. Both eggplant and pepper carry several small inversions and translocations with regions corresponding to tomato chromosome 5 (eggplant) and 9 (pepper) in this region, leading Doganlar (2002a) to conclude that this represented an unstable region during *Solanaceae* genome evolution.

Chromosome 4 A locus from tomato 6L (TG581) was translocated to J4S. Eggplant and subsect. *Lycopersicoides* both show translocated areas between corresponding tomato chromosomes 4S and 10, while 4S in pepper is associated with tomato 5L, indicating that homeologous tomato 4S is a region prone to rearrangements in the *Solanaceae*.

Chromosome 5 Chromosome 5 is conserved between *Juglandifolium* and tomato, while translocation events are evident in the other species.

Chromosome 6 Eggplant and potato 6S are inverted relative to tomato, and *Juglandifolium* appears to share the former configuration. However, marker coverage is poor on this short arm of J6, reflecting its relative length: chromosome 6 is an acrocentric chromosome with the shortest short arm of the set, except for telocentric chromosome 2 (Sherman and Stack 1995). Therefore further evidence (e.g. FISH) is required to prove the state of this arm. Due to the absence of 6S

markers on both comparative maps the configuration in pepper and subsect. *Lycopersicoides* is unknown. *Lycopersicoides* carries a locus from tomato chromosome 1 in the centromeric region of chromosome 6, indicating that this might be an area of enhanced flexibility. In the middle of J6L two adjacent loci have assumed switched positions in *Juglandifolium* relative to tomato. This may indicate a transposition event, or that the marker order in one of the maps is erroneous.

Chromosome 7 Marker content and order are well conserved among tomato, *Juglandifolium*, *Lycopersicoides*, potato, eggplant and – albeit to a lesser degree – pepper. The top of chromosome 7 is inverted among tomato vs. *S. pennellii*.

Chromosome 8 Either the entire J8S or just the distal portion may be inverted relative to tomato. Again, a final judgment cannot be made as chromosome 8 is also acrocentric and the only markers that are present are closely spaced. In addition, tomato chromosome 8 and 12 have undergone a reciprocal whole-arm translocation in one of the *Juglandifolium* species. The translocation breakpoint seems to correspond to the one in the translocation between chromosome 1 and 8 in pepper. Interestingly, in *Lycopersicoides* a locus from the distal end of tomato 12L is transposed to the distal end of 8S.

Gene order within rearranged segments has been remarkably well preserved during the evolution of all lineages (Doganlar et al. 2002a).

Chromosome 9 Chromosome 9 is largely conserved among tomato, *Juglandifolium* and *Lycopersicoides*, and preservation of gene order within translocated segments is also apparent among the other species.

Chromosome 10 One of the *Juglandifolium* species harbors a putative inversion on J10S. A severe map length compression was also observed in the same region in pepper. Evidence for a hotspot of restructuring was found in a region near the centromere on 10L in *Juglandifolium*. Interestingly, the spot on J10L that corresponds to the location of T0308 in tomato (in *Juglandifolium* T0308 is positioned on J3S, in tomato on 10L) shows some perturbation of gene order and demarcates the end of a stretch of severe map compression on J10. This area also colocalizes with the breakage point of the paracentric inversion in tomato relative to *Lycopersicoides*, eggplant and pepper. Eggplant chromosome 10 is composed of tracts that are homeologous to three different tomato chromosomes as a result of two independent translocation events. Also noteworthy, three loci in that area in tomato are scattered on 4S, the centromeric region of chromosome 7 and 9L in *Lycopersicoides*. This region therefore appears to be an ancient hotspot for structural rearrangements in the *Solanaceae*.

Chromosome 11 Tomato, *Juglandifolium* and *Lycopersicoides* are collinear with respect to chromosome 11, while the other *Solanaceae* species show interchromosomal rearrangements.

Chromosome 12 Marker order between J12 and tomato chromosome 12 was perfectly conserved, but one *Juglandifolium* species, as mentioned earlier, carries a reciprocal whole-arm translocation with chromosome 8. Corresponding regions of tomato chromosome 12 have undergone multiple rearrangements in both the eggplant and the pepper lineage. Eggplant chromosome 5 is a fusion product of homeologous regions on tomato 5L and 12L, whereas a portion of tomato 12S maps onto eggplant 10S. In pepper homeologous regions of tomato chromosome 12S form part of chromosome 9 and 12. Also, several loci from the distal end of tomato 12L are scattered across the genome in *Lycopersicoides*, adding further evidence that this region represents a rearrangement hotspot in the evolution of the *Solanaceae*.

4.1.4.5 Phylogenetic relationships

The configuration of J10L is the same as in tomato, unlike subsect. *Lycopersicoides* which contains the ancestral arrangement, found in all above mentioned *Solanaceae*. This result clearly places sect. *Juglandifolium* as closest outgroup to the tomatoes (sect. *Lycopersicon*) and subsect. *Lycopersicoides* as basal to both. The result is consistent with phylogenies derived from GBSSI sequence and AFLP data (Peralta and Spooner 2001; Spooner et al. 2005). The fact that reproductive barriers with sect. *Lycopersicon* are more pronounced via-a-vis sect. *Juglandifolium* than subsect. *Lycopersicoides* had supported the assumption that subsect. *Lycopersicoides* was more closely related to the tomatoes (Rick 1979). This interpretation was also consistent, with the more tomato-like morphology of subsect. *Lycopersicoides*, as well as their similar ecology and distribution (Rick 1988). The intriguing question which of the species in sect. *Lycopersicon* is the closest to the two *Juglandifolium* species unfortunately remains unanswered. A single combined phylogenetic tree based on AFLP, GBSSI, cpDNA, ITS sequence and morphological data supports *S. habrochaites* and *S. pennellii* as one clade that forms a basal polytomy with southern accessions of *S. peruvianum* and *S. chilense* (Spooner et al. 2005). Although *S. peruvianum* and *S. pennellii* overlap in their latitudinal distribution range with *S. ochranthum*, they inhabit drier areas than the *Juglandifolium* species (http://tgrc.ucdavis.edu). *S. habrochaites*, on the other hand, is not only largely sympatric with both *S. ochranthum* and *S. juglandifolium*, it is also the species in sect. *Lycopersicon* that displays the strongest similarities with sect. *Juglandifolium* morphologically and with respect to habitat adaptations. It is unknown, however, whether morphological resemblances are caused by convergent evolution or common ancestry. *S. habrochaites* is represented from Peru to tropical Ecuador where it overlaps with the putative center of diversity of *S. ochranthum* and *S. juglandifolium*. Furthermore, it is the only species in subsect. *Lycopersicon* that has been collected together with a *Juglandifolium* species, namely with *S. ochranthum* in Leimebamba, Peru (Smith

and Peralta 2002; http://tgrc.ucdavis.edu). On the other hand, the flowers of *S. pennellii* suggest a closer affinity to the ancestral *Solanum* state: lack of sterile anther tip, unfused anthers, and the presence of terminal pore (Peralta and Spooner 2005).

4.1.4.6 Putative evolution scenario for *Juglandifolium*

Peralta and Spooner (2005) proposed an intriguing speciation scenario for the tomatoes: the tomato ancestor might have covered vast areas in central Peru. Climatic changes, foremost the drastic increase in aridity along the southern Peruvian coast during the Holocene aided the selection for adapted forms (Arroyo et al. 1988), thereby accelerating speciation within the tomatoes.

Interestingly, the two most basal groups in the tomato clade, sect. *Juglandifolium* and subsect. *Lycopersicoides*, occupy areas that flank the present-day distribution range of the wild tomatoes (sect. *Lycopersicon*) at their northernmost and southernmost edge, respectively. As no fossil record exists for the tomatoes, the age of the lineage can only been estimated from sequence data. The genus *Solanum* presumably diverged from the most recent ancestor ~ 12 myr (Wilkstrom et al. 2001). Doganlar (2002a) suggested a chromosomal evolution rate of 0.002 rearrangements/Mb/myr for *Solanum*. Assuming constant divergence rates and a genome size of 950Mb (Arumuganthan et al. 1991), tomato may have split from a common ancestor with potato ~ 1.6 myr (under assumption of 3 rearrangements in the tomato lineage) to ~ 2 myr (4 rearrangements), thus at the end of the Pliocene. The increase in aridity in coastal regions of Peru and Chile began ~ 15 myr (Gregory-Wodzicki 2000), therefore desert conditions were probably already well-established at the time of the tomato divergence. According to these estimates sect. *Lycopersicon* and subsect. *Lycopersicoides* separated from a common ancestor only ~ 0.5 myr, i.e. at mid-Pleistocene, making it likely that the cool conditions during that period had an effect on speciation of present-day tomato wild species.

The area in northern Peru which represents the transitional region between the humid, tropical north and the arid south is of an exceptional species richness and it is also the center of diversity of some *Lycopersicon* species such as *S. pimpinellifolium* and *S. habrochaites* (Rick et al. 1979). Hence, the tomato ancestor may have originated in that region and subsequently undergone an expansion, primarily directed southward. Alternatively, the split from the potato lineage and evolution of the present-day tomato genome, including the last paracentric inversions, may have occurred in an area further south in Peru where species-richness in potato is greater (Hijmans and Spooner 2001). The common ancestor of *S. sitiens* and *S. lycopersicoides* may have migrated from there to drier regions in the south while the tomato ancestor spread out northward into more tropical areas. The widespread colonization most likely occurred with an ancestral genome morphologically largely

similar to the current tomato genome (albeit maybe lacking the 6S inversion), i.e. carrying five paracentric inversions. In adaptation of the model advocated by Peralta and Spooner (2005) the unresolved basal polytomy of the tomato clade (Spooner et al. 2005), comprised of *S. habrochaites* (distributed from central Ecuador to central Peru), *S. pennellii* (distributed across Peru and northern Chile) and the *S. peruvianum*/*S. chilense* clade (distributed in southern Peru/northern Chile; http://tgrc.ucdavis.edu) could be explained by the presence of an ancient megapopulation, spread across a vast area from Ecuador to northern Chile, that, under adaptations to local biotic and abiotic specificities, eventually fragmented into the various forms that subsequently developed into distinct species. This scenario suggests that at one point during its evolution the tomato ancestor must have encountered more favorable conditions that allowed for the broad expansion.

4.1.5 Practical implications

Previous attempts at sexual or somatic hybridization to make the *Juglandifolium* group accessible for the transfer of economically relevant genes have been unsuccessful (Rick 1979; Stommel 2001). Hence, *S. ochranthum* and *S. juglandifolium* are currently the only tomato species which are completely isolated from the remainder of tomato species. The results of the present study may give reason to raise new hopes for germplasm introgressions into cultivated tomato. If *S. ochranthum* is the carrier of the translocation between chromosome 8 and 12, then chances for introgressions maybe better with *S. juglandifolium*, a species yet relatively untested for its crossing behavior with most of the tomatoes, including the cultigen.

4.2 Genetic diversity analysis

The present study is, so far as we are aware, the first examination of the status of genetic variability in natural populations of the two tomato-related wild species, *S. lycopersicoides* and *S. sitiens*. The two species are composed of small, fragmented populations across a narrow distribution range in southern Peru/northern Chile. Populations have suffered apparent declines over the recent decades (Chetelat, pers. comm.), raising concern about their threatened or endangered status. They occupy territory that is marginal for wild tomatoes in terms of climatic conditions, i.e. characterized by very low temperatures (*S. lycopersicoides*) or extreme aridity (*S. sitiens*), and are suspected to harbor traits, such as abiotic and biotic stress tolerances, that are of value for tomato breeding (Rick 1988).

Most of the known populations of each species were sampled in the present study, including 14 in *S. lycopersioides* and seven in *S. sitiens*, represented by 333 and 195 plants, respectively. Two

marker systems were implemented to measure genetic variability, one protein-based (allozymes), the other DNA-based (microsatellites), allowing the juxtaposition of genetic variability at the two levels.

The aims of the project were to

1) quantify amounts of genetic diversity in the two tomato related wild species *S. lycopersicoides* and *S. sitiens* at the protein and DNA levels, and to

2) identify the underlying genetic structure of populations of each species in order to cast light on their diversity and status, hypothesize about their demographic histories, and to provide guidelines for conservation strategies.

4.2.1 Differential patterns of allozyme and microsatellite variation

In order to investigate the concordance between the two marker systems among-species comparisons were based on a common set of markers, within-species comparisons on a common set of individuals.

Microsatellite variation exceeded that of allozymes, a frequent observation in studies that include both marker systems (e.g. Estoup et al. 1998; Gao et al. 2002; Dhuyvetter et a. 2004), and consistent with the mutation rates which are several orders of magnitude higher for microsatellites (ca. 10^{-3} to 10^{-5}) than for isozymes (10^{-6} to 10^{-7}) per locus and generation (Kahler et al. 1984; Weber and Wong 1993; Vigouroux et al. 2002).

The discrepancy between the two marker systems was much smaller in *S. sitiens* than in *S. lycopersicoides*. Estimates at allozyme loci were typically two to three times lower than those at DNA loci in *S. lycopersicoides*: Individual allozyme loci contained two to five alleles (*k*) with an average of 2.8 per locus, SSR loci two to 12 alleles, with an average of 5.7[5]. The overall number of polymorphic sites (*P*) was 42 % vs. 92 %, and the mean gene diversity (H_E) was 0.116 vs. 0.421 at allozyme loci and SSR loci, respectively. In *S. sitiens* estimates at the protein level were higher than in the sister species, and at least three quarters of those at the nucleic acid level: Allele numbers at allozyme loci (*k*) ranged from two to four with an average of 3.1 per locus, at SSR loci from four to eight with a mean of 4.1. *P* was 68 % vs. 83 % and the mean H_E 0.285 vs. 0.354.

Conversely, as expected, inbreeding levels were higher at isozyme loci compared to SSRs. Again, the gap was wider in *S. lycopersicoides* where F_{IS} = 0.037 at allozyme loci vs. -0.016 at SSRs. In *S. sitiens* the corresponding F_{IS} values were 0.030 vs. 0013.

[5] The values were estimated from a set of individuals common to both marker analyses.

With the exception of the amount of inbreeding (F_{IS} and the F_{IT} per locus and F_{IS} per population) measures were significantly different in *S. lycopersicoides* between the two marker systems, whereas in *S. sitiens* only the average number of alleles (*k*), the average allelic richness per population (R_S) and the inbreeding coefficient per population (F_{IS}) were significantly distinct.

In populations of wild rice (*Oryza rufipogon*) Gao et al. (2002) observed that most polymorphic allozyme loci showed one allele at a high frequency accompanied by several rare alleles. The same marker behavior was evident in the two tomato species. Although diversity estimates at the two locus types diverged by an even larger factor than the one observed for *S. lycopersicoides* (e.g. the population means were P = 12.7 % vs. 73.3 %, k = 1.2 vs. 3.1 and H_E = 0.030 vs. 0.345 at allozyme vs. SSR loci, respectively) patterns of diversity levels and diversity structure among rice populations were in good agreement between allozyme and SSR loci (with a greater detection resolution at SSR loci; Gao et al. 2002), whereas the two marker systems showed only a rather poor, if any, congruence in *S. lycopersicoides*. Likewise, genetic variation exhibited non-corresponding or even opposite geographic clines in *S. lycopersicoides*, while those in *S. sitiens* were largely mirrored by both marker systems.

Genetic differentiation among populations was more pronounced at allozyme loci than at SSR loci in *S. sitiens*: average pairwise F_{ST} values (θ) were 0.280 and θ = 0.203, respectively. The reverse was true for *S. lycopersicoides* where the average F_{ST} (θ) was only 0.169 for allozymes and 0.261 for SSRs. Among populations of wild rice allozyme differentiation was weaker than SSR differentiation (Gao et al. 2002). Similarly, no population structure could be identified at allozyme loci in *S. lycopersicoides* (STRUCTURE analysis). The phylogeny was poorly supported and corresponded neither to geographic relationships nor to the clusters reported by microsatellites. As mutation rates are higher at microsatellite loci these results may indicate that populations of *S. lycopersicoides* diverged more recently and/or maintain a higher extent of gene flow between them. This hypothesis would be in line with the specific environmental conditions that are more beneficial to gene flow in *S. lycopersicoides*: the environment is moister, rich in other plant species and presumably also pollinators and seed vectors. In addition, a greater historic connectivity (i.e. a more recent divergence) is easier to imagine for *S. lycopersicoides*, where recent population fragmentation has occurred as a consequence of human activities (agriculture, e.g. herding of animals, Chetelat, pers. comm.). Fragmentation in *S. sitiens* is probably primarily governed by the harsh climatic conditions and may therefore have been ongoing for a longer period of time. (According to Gregory-Wodzicki (2000) the increase in aridity in coastal regions of Peru and Chile began ~ 15 myr.). Still, divergence time in *S. sitiens* may be underestimated relative to *S. lycopersicoides*. Generation times are longer in *S. sitiens*. Plants have the capacity to resprout from

roots as a mechanism to survive in harsh conditions. Individual plants are often ancient and almost tree-like, while those of *S. lycopersicoides* are short-lived herbs.

In spite of these discongruences matrices of pairwise F_{ST} values of the two marker systems were significantly ($P < 0.01$) correlated in both species, even in *S. lycopersicoides*, albeit much more weakly ($r = 0.363$) than in *S. sitiens* ($r = 0.916$). Graham (2005), who used nearly the same allozyme systems but different microsatellites loci, found a strong and highly significant ($p < 0.001$) correlation ($r = 0.984$) between the respective genetic distances.

The cause for the discrepany in *S. lycopersicoides* could not be identified. Scoring errors may have caused the differences among the two sister species. However, the discordances remained evident when only the subset of common markers (i.e. makers used in the analyses of both species) was considered: just three (*Aco-1*, *Idh-1*, *Pgm-2*) out of the eight loci behaved similarly in *S. lycopersicoides* and *S. sitiens*. Loci could behave idiosyncratically among species, although Rick (1981) reported similarities at allozyme loci among *S. pennellii*, *S. habrochaites* (formerly *S. hirsutum*) and *S. pimpinellifolium*. Considered 'sister taxa', *S. lycopersicoides* and *S. sitiens* are thought to have diverged from a common ancestor only fairly recently (Pertuzé et al. 2002), therefore the mere amount of species divergence is unlikely to explain the differences observed between the two species. Alternatively, purifying selection could be operating at several of the allozyme loci in *S. lycopersicoides*. Non-neutral behavior of allozymes has been reported many times (e.g. Rick et al. 1977; Karl and Avise 1992; Dhuyvetter et a. 2004). This is not surprising as they are proteins, and many of them are involved in fundamental metabolic pathways. Three out of 13 polymorphic allozyme loci were excluded from the analysis in *S. lycopersicoides* because of significant heterozygote deficiencies. In contrast, none of the 15 SSRs needed to be removed from the dataset. The fact that two of the three deviating loci (*Adh-1* and *Got-3*) also displayed heterozygote deficiencies in *S. sitiens* is in support of selection pressures at some allozyme loci, however, it does not explain the differences between the two species. In addition, if selection was responsible for the incongruence between the two marker systems in *S. lycopersicoides*, it would have escaped detection by Hardy-Weinberg tests in all of the locus/population combinations, which may be unlikely.

In both species marker systems failed to identify populations as departing from Hardy-Weinberg equilibrium or as 'bottlenecked', complicating general conclusions. With respect to Hardy-Weinberg disequilibria, the more sensitive microsatellite assay detected larger numbers of deviating populations but smaller numbers of deviating loci. However, observations were consistent with the expectation in that bottleneck signatures at microsatellite data were primarily characterized by an excess of heterozygosity (Cornuet and Luikart 1996), while those detected with allozyme data

showed an underrepresentation of rare alleles. The 'allele deficiency' is dependent on a) the elapsed time following the bottleneck, b) the mutation rate and c) the sample size in terms of genes (Maruyama and Fuerst 1985). Factor a) is equal for both datasets while b) and c) are elevated for microsatellites. High mutation rates erase the footprint of a bottleneck faster, thus allozymes should be able to reach further in the past than SSRs to report bottleneck events (Cornuet and Luikart 1996). Conversely, microsatellites may be more useful for detecting recent bottleneck events (Cornuet and Luikart 1996) due to their higher sensitivity.

Hence, in the present study the two marker systems did not only vary quantitatively, i.e. in their level of sensitivity, but also qualitatively, with regard to the diversity picture they presented. The degree of congruence was primarily species-specific and to a lesser extent locus-dependent, and could therefore reflect different demographic histories/life history traits.

Although a statistical proof is lacking, several lines of evidence (foremost the poor performance in identifying genetic clusters) gave reason to question the validity of the allozyme loci to correctly report neutral genetic diversity in *S. lycopersicoides*. Interpretations were therefore concentrated on microsatellite variation. Finally, it should be noted that the allozyme dataset possessed a lower statistical power in the present analysis: fewer loci in combination with lower polymorphism levels reduced the amount of informative sites to one third in *S. lycopersicoides* and two thirds in *S. sitiens* relative to that of microsatellites. This was taken into account by prioritizing microsatellite data for final conclusions.

4.2.2 Genetic diversity

4.2.2.1 Global levels of genetic diversity

Populations of *S. lycopersicoides* appeared slightly more diverse (measured as *k* per population, *P*, R_S and H_E) than those of *S. sitiens* (*k* = 44.3 vs. 36.3; *P* = 92 % vs. 84 %, R_S = 3.21 vs. 2.53 and H_E = 0.442 vs. 360, respectively) according to microsatellite loci. Levels of inbreeding were near zero in both species, however, *S. sitiens* populations were more inbred on average than *S. lycopersicoides* populations at SSR loci (F_{IS} = 0.009 vs. F_{IS} = -0.016) and even at allozyme loci (F_{IS} = 0.121 vs. F_{IS} = 0.062, respectively), which were otherwise a lot less diverse in *S. lycopersicoides*. The result may stem from higher levels of biparental inbreeding within populations of *S. sitiens*, in line with the slightly smaller census sizes in terms of both number of populations and number of individuals (Rough estimates for individuals per population were 57 and 44 for *S. lycopersicoides* and *S. sitiens*, respectively.) The extent of population fragmentation appears more severe in the latter species, which may also be causative of the observed lower levels of diversity. *S. sitiens* populations tend to be isolated from other populations by areas totally lacking in plants, due to the

extreme aridity of the Atacama desert. The hostile conditions may hamper gene exchange. In fact, pollinators have never been observed at collection sites (Chetelat, pers. comm.). Under those circumstances levels of inbreeding are likely to rise, promoting further genetic loss via the effects of inbreeding depression and genetic drift (Ellstrand and Elam 1993; Lowe et al. 2004). This interpretation is consistent with the higher level of divergence between populations seen in *S. sitiens* compared to *S. lycopersicoides.*

There was indication (based on *k*, R_S and H_E, but not *P*), that both species harbor substantially less genetic variability compared to the related species *S. chilense* (e.g. H_E = 0.442 in *S. lycopersicoides* and 0.360 *S. sitiens* vs. 0.585 in *S. chilense*), which cohabits some sites with *S. lycopersicoides* and grows in proximity to some *S. sitiens* populations (though never at the same site). Relative to *S. lycopersicoides* and *S. sitiens, S. chilense* displays a much wider distribution in terms of both latitude (from 15° S to 25° S) and elevation (from sea level to > 3500 m), hence presumably possessing adaptation to a wider range of environments. Population sizes for *S. chilense* also tend to be larger than those of either *S. sitiens* or *S. lycopersicoides.* This greater geographical expansion is accompanied by a greater demographic representation with respect to both numbers of populations and average population sizes (http://tgrc.ucdavis.edu). The true diversity discrepancy may be greater than reported herein. *S. chilense* estimates are likely to be biased downwards because 1) only one population was tested, and 2) loci were selected according to polymorphism content in the two other *Solanum* species only. However, a recent study by Graham (2005) based on nine allozyme and four SSR loci reported similar or slightly lower overall estimates of genetic diversity in 33 *S. chilense* populations. For allozymes *P* = 0.5 in *S. chilense* vs. 34 % and 67 % in *S. lycopersicoides* and *S. sitiens*, respectively, and *k* = 1.7 vs. 1.5 and 2.0. For SSRs *P* = 0.9 vs. 92.4 % and 83.7 %, *k* = 2.3 vs. 6.3 and 2.6. For both marker datasets combined H_E = 0.142 vs. 0.122 and 0.256 (allozymes) or 0.442 and 0.360 (SSRs). The estimates for LA2773, of which the identical DNA samples were used in the present study as a reference, were slightly below the *S. chilense* species average in Graham's study, suggesting that the genetic depletion of *S. lycopersicoides* and *S. sitiens* relative to the entire species of *S. chilense* may be greater than indicated herein. Alvarez et al.'s (2001) observations for gene diversity in *S. chilense* (H_E = 0.517) based on 17 microsatellite loci and three populations (five plants each) were in congruence with the present (H_E = 0.585). (Gene diversity in Graham's study was computed from the combined data of allozymes and SSRs and is therefore not directly comparable with the estimates herein.) The herein reported gene diversity of *S. lycopersicoides* (H_E = 0.442) was 22 and 15 % lower than Alvarez et al.'s estimates for *S. peruvianum* (H_E = 0.569) and *S. chilense*, respectively, and almost twice as high as that of *S. pennellii* (H_E = 0.240). The lower gene diversity estimate for *S. sitiens* (H_E = 0.360) was on the

same level with northern accessions of the former *L. peruvianum* (including *L. peruvianum* var. *humifusum*, now renamed to *S. arcanum*) in Alvarez et al.'s study (H_E = 0.363), twice as high as that of *S. pimpinellifolium* (H_E = 0.197) and *S. habrochaites* (H_E = 0.170), 50 % higher than *S. pennellii* and 38 % lower than *S. peruvianum*. (All estimates were based on 15 - 35 individuals in 3 - 7 populations.) However, an early allozyme study by Rick and Tanksley (1981) showed a *P* estimate in northern populations of *S. pennellii* (*P* = 65 %) similar to that reported herein for *S. sitiens* (*P* = 67 %), and a slightly higher average *k* per locus (*k* = 2.54). The *k* in *S. sitiens* (*k* = 1.96 resembled more that of the inbred *S. pennellii* populations at the southern distribution (*k* = 2.08) in their study.

Various genetic diversity analyses that utilized different marker approaches such as allozymes (Breto et al. 1993), RFLPs (Miller and Tanksley 1990), RAPDs (Egashira et al. 2000) and DNA sequences (Baudry et al. 2001) showed *S. peruvianum* as the most diverse species of the tomato clade, followed by *S. chilense*, while the self-compatible species were more genetically depleted. *S. pennellii* is mostly self-incompatible, widespread and harbors more genetic diversity than other outcrossing tomato species, such as *S. habrochaites* and *S. pimpinellifolium* (Rick 1981). Thus, surprisingly in view of the substantially lower representation and narrow distribution, *S. lycopersicoides* and *S. sitiens* appear to still be more diverse than most of the tomatoes, presumably because these have lost a significant amount of genetic diversity during speciation events from a common ancestor with *S. peruvianum* as well as by loss of self-incompatibility. It should be noted, however, that comparisons across studies have to be viewed with caution: Apart from differences in the numbers of loci analyzed, such comparability may be obscured by intrinsic locus specificities, differences in sampling strategies, marker systems and choice of descriptive parameters for the documentation.

4.2.2.2 Levels of genetic diversity among populations within species and geographic trends

S. lycopersicoides. *S. lycopersicoides* populations are found at five major geographic localities (from north to south): 1) Lago Aricota, 2) Palca, 3) Putre, 4) Camarones and 5) Camiña (Figure 17).

Neither the combined (measured by *P*, R_S and H_E) allozyme nor the microsatellite genetic variability in populations of *S. lycopersicoides* exhibited a clear geographic cline. The most genetically diverse populations were Aricota #2 (LA1966) in the north, Zapahuira (LA2772) and Lluta (LA4320) in the center, and Moquella (LA2730) in the southern part of the distribution. Less diverse were Putre #1 (LA2781; central location and highest elevational point) and Aricota #2

(LA2387; northern location). Allozyme data did not mirror the picture presented at SSR loci. Exceptions were Palca (LA1966) which was together with Camiña (LA4123; southern distribution end) the most diverse population, as well as Putre #1 (LA2781; central distribution) which was together with Esquina (LA4131; southern distribution) among the least diverse populations.

However, geographic clines were evident for individual allele frequencies and – to a lesser extent – diversity estimates, of which mainly P and θ were affected. These two were significantly larger at more southern latitudes (θ only at SSR loci). Furthermore, P increased with western latitude and lower altitude at allozymes, but exhibited the opposite trend at SSRs. Genetic distance (θ) was larger towards the east and at lower elevations at allozymes. The population's degree of geographic 'isolation' (i.e. the average distance to the remainder of populations) displayed a positive correlation with P (at allozymes) and θ (at SSRs), but a negative with k at SSRs.

In the absence of selection, geographic clines portray the combined effects of mutation and drift. From the observed pattern a south/north or east/west expansion could hardly be invoked. However, the high incidence of allele frequencies that exhibited trends with latitude (one third of the SSR alleles) and also longitude (one quarter) together with a highly significant (r = 0.680) pattern of isolation by distance were suggestive of processes of migration and/or progressive fragmentation along these two geographical axes. Correlations between allele frequencies and latitude have been observed in many species (Moran et al. 1989; Davis and Shaw 2001), foremost in widespread woody species as a consequence of Quaternary climate changes. Likewise, IBD is a feature commonly found in plant species (e.g. England et al. 2002; Honnay et al. 2007). A pattern of IBD also characterized genetic diversity in population of *S. chilense* (Graham 2005), a species that is overlapping in distribution range with *S. lycopersicoides*. Geographic structuring expressed as IBD was weak in *S. lycopersicoides* (r = 0.223), albeit significant ($P < 0.05$) at allozyme loci and may reflect high levels of ancient genetic variation at the protein level that predate population expansion and/or fragmentation. Populations of *S. pimpinellifolium* were shown to contain allelic distributions in discordance with geographic patterns, presumably relicts of a pre-colonization phase (Caicedo and Schaal 2004). *S. lycopersiciodes* populations typically grow along the drainages of the Andean cordillera which are separated by mountain ridges. Although little is known about the actual pollen or seed dispersal mechanisms it is likely that the mountains pose a substantial if not insurmountable barrier to gene flow among drainages. Due to a lack of fossil data the age of the tomato clade is not known with any certainty, but is estimated to be less than 12 my, a time when the genus *Solanum*

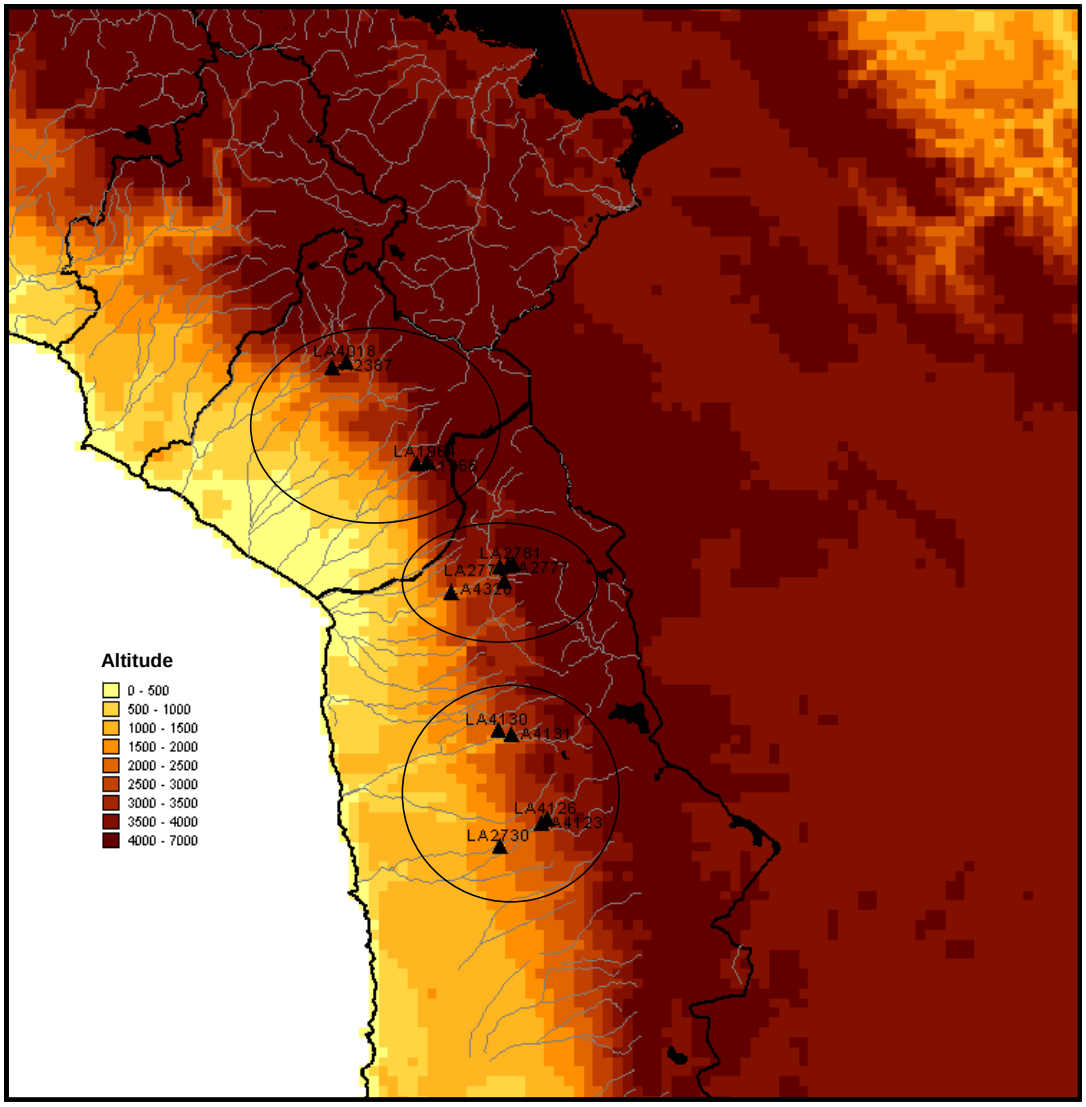


Figure 17. Three major population clusters in *S. lycopersicoides* set against the elevational scale.

Altitude is presented in meters [m].

presumably diverged from the most recent ancestor (Wikstrom et al. 2001). The Andes in that part of South America were half their present elevation at 10.4 myr (Gregory-Wodzicki 2000). It is therefore conceivable that dispersal across mountain ridges was less impeded in the early days of the species, thereby leading to the signature observed at the allozyme level.

Given the high altitudinal range in that species (over 2,220 m), the finding that elevation hardly played a role in shaping genetic diversity in *S. lycopersicoides* are counterintuitive. Although a few allele frequencies showed correlations with altitude, overall correlations with pairwise genetic distances (θ) were non-significant at SSR loci. Elevational distances between populations within

drainages are often large, and yet populations may be able to retain their connectivity through gene flow. Alternatively, populations within the same drainage may have diverged only recently, resulting in a diminished genetic distance between populations that are separated by great altitudinal distances relative to populations at equal elevations but in different drainages. Interestingly, at allozyme loci pairwise genetic distances (θ) were significantly (albeit weakly) correlated to elevation (r = 0.381), and that in spite of the fact that the frequency of only a single allele was correlated with elevation. If allozyme variation reflects a species' more remote history, then populations may have lost connectivity over altitudinal distances long enough ago so that high mutation rates at SSR loci have had enough time to erase the signature.

The central distribution area around the Nevados de Putre was the most population-rich and also held several of the largest populations. Surprisingly, the census population size hardly influenced levels of diversity. Only estimates of P at allozyme and θ at SSR loci exhibited positive correlations. However, population size estimates were rather imprecise and may not well reflect the effective population size (N_e), which is known to have a large impact on levels of genetic diversity (Avise 2004). Many reports state an increase of genetic diversity with population size in plant species (e.g. van Rossum et al. 2004; Prentice et al. 2005), yet it is not uncommon to find no such correlation (Leimu and Mutikainen 2005; Honnay et al. 2007).

None of the populations departed from Hardy-Weinberg equilibrium at allozyme loci, but at SSR loci three populations (Palca - LA1966, Pachica - LA4130 and Lluta - LA4320) showed heterozygote deficiencies. Deviations are usually the result of non-random mating, but may also arise from drift (especially in small populations), natural selection (homozygote advantage), demographic expansions/reductions and theoretically even from mutations (Lowe et al. 2004). The departure in Palca (LA1966) may be the result of non-assortive mating. Drift due to the small population size (estimated 20 individuals at the time of sampling) could be causative for the deviation in Lluta (LA4320), whereas that in Pachica (LA4130) may represent an artifact caused by the exceptionally low sample size (11 individuals vs. 23.5 on average in the remainder of populations). In agreement with this hypothesis, Pachica (LA4130) was also one of the least diverse populations, while Palca (LA1966) and Lluta (LA4320) exhibited high levels of genetic diversity.

Levels of inbreeding were low in *S. lycopersicoides*. Palca (LA1966) was the most inbred population (F_{IS} = 0.149) at SSR loci, followed by its neighboring population Chupapalca (LA1964), at the second location from the north (Palca). Both were medium-sized populations at the time of sampling that might have undergone bottleneck events in the recent past as indicated by non-significant signs. In contrast, Palca (LA1966) exhibited the lowest level of inbreeding at the allozyme level (F_{IS} = -0.136) and was among the most diverse populations at both marker levels.

At allozyme loci levels of inbreeding were highest in Putre #1 (LA2781; F_{IS} = 0.242), a very large population at the highest distribution point of all tomato populations (3,800 m). The result may be indicative of selection pressures operating on allozyme loci, or may reflect an ancient founder event (which would also be in line with the marginal location) that is no longer evident at SSR loci.
Only one *S. lycopersicoides* population (Aricota #2 - LA2387 at the northernmost location) showed a clear signal (i.e. in form of both a significant excess of heterozygosity and a mode-shift distortion) of a recent bottleneck event, and that was based on SSR data only. Levels of genetic diversity were slightly below average in that population. Bottlenecks may occur in form of founder events or as a consequence of habitat fragmentation. Population bottlenecks are relevant for genetic preservation in that they reduce the genetic diversity, and thereby the capacity for evolutionary adaptations. Conversely, the amount of inbreeding is enhanced by a bottleneck event, and the fixation of moderately deleterious alleles becomes more likely (reviewed in Cornuet and Luikart 1996; Luikart et al. 1998). Thus, bottlenecks can push a threatened species into extinction and should be viewed with caution in genetic diversity analyses.

S. sitiens. Five geographic locations can be identified in the distribution of *S. sitiens*: 1) a 'northern' group (north of Chuquicamata), 2) Aguada Limón Verde, 3) Cerro Quimal, 4) Mina San Juan and 5) Mina La Escondida (Figure 18). It should be noted that populations in *S. sitiens* are very fragmented and scattered, therefore often only a single population is found at a given site (e.g. locations 2 - 5). Similar to *S. lycopersicoides* levels of genetic variability in populations of *S.sitiens* did not show specific geographic trends. The most diverse populations were Limón Verde (LA4112) and Paqui (LA4116) in the central/northern part of the distribution, the least diverse was Carbonatera (LA4114), which was also located in the north. However, the present investigation was based on a very small number of populations (7), which hampered the analysis due to chance effects and low statistical power.
Allozyme and microsatellite data showed similar trends, although allozyme variation was more strongly connected to geographic features: all descriptive parameters were affected by latitude at the protein level, declining towards the south. All other trends were restricted to *P* and θ, similar to the observations in *S. lycopersicoides*, although correlations were often stronger than in the latter species. *P* increased towards the east (at allozyme loci only) and with elevation and with increasing average geographic distance between populations (at both locus types). Genetic distance (θ) showed spatial patterns in both datasets, increasing towards the south, west and with geographic isolation, in congruence with the finding of significant isolation by distance (r = 0.848 and r = 0.785 at SSR and allozyme loci, respectively). IBD was more pronounced in *S. sitiens* than *S. lycopersicoides*,

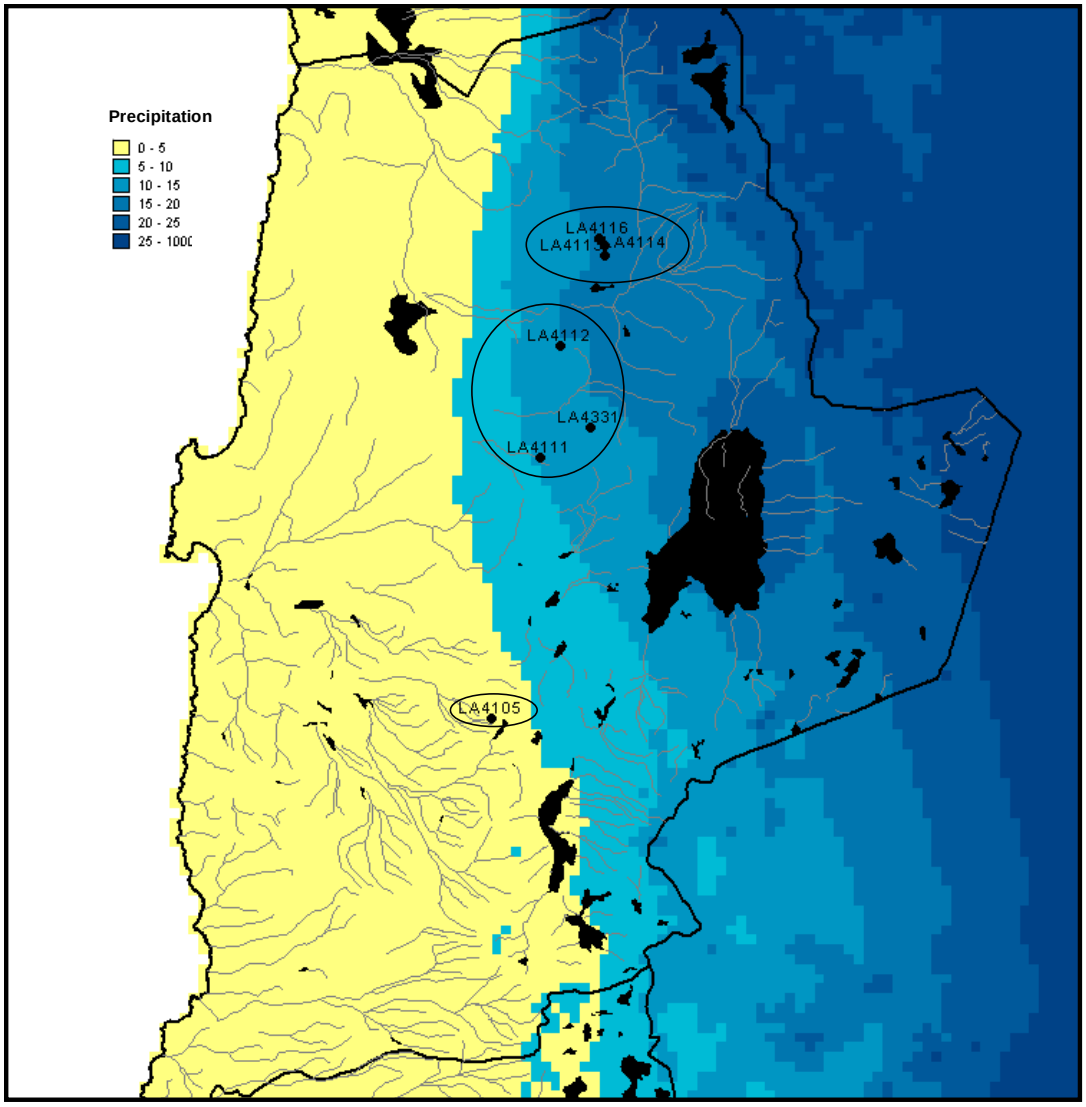


Figure 18. Three major clusters of *S. sitiens* population and regional precipitation rates.
Precipitation is presented in millimeters [mm].

possibly because genetic drift is stronger in the smaller *S. sitiens* populations and/or gene flow is more severely hampered across desert stretches that separate individual plants and populations. Mechanisms and extent of pollen and seed dispersal are unknown. Buzz pollination by bees as in the rest of the tomatoes seems the most likely mode but has not yet been confirmed due to a notable absence of pollinating insects at the collection sites. In addition, the anther architecture is different in these two species (anthers are not connected by interlocking marginal hairs) which could be the result of coevolution with other pollinator species. In any case, it is easily conceivable that both mobility and abundance of any animal involved in pollen/seed dispersal will suffer from the hostile environmental conditions and lack of vegetation (food sources).

The proportion of allele frequencies correlated with geographic features was only half as large at SSR loci, but twice as large at allozyme loci compared to that in *S. lycopersicoides*. This finding emphasizes that the signature of successive colonization events was much more pronounced in *S. lycopersicoides* than *S. sitiens*, or, reversely, the role of genetic drift for population differentiation was larger in *S. sitiens*. Interestingly, it was only in *S. sitiens* that a substantial number of alleles, namely 20 % of the allozyme alleles, showed correlations with altitude, although the altitudinal range was less than one quarter (less than 500 m) of that shown by *S. lycopersicoides*. However, not a single SSR allele was correlated with elevation, and isolation by altitudinal distance failed to be detected with either marker system. Census sizes varied substantially, from an estimated 17 to 100 individuals among populations in *S. sitiens*, but did not follow an apparent geographic pattern. As in the sister species, census population size was positively correlated with P and θ.

Population Cere (LA4113; according to allozyme data) in the north and the remote, southernmost population Escondida (LA4105; according to SSR data) deviated from Hardy-Weinberg equilibrium with a heterozygote excess and deficiency, respectively. Both were amongst the populations with the smallest number of individuals (just 20), hence they may have been pushed into disequilibrium via genetic drift. Both carried signatures of a recent genetic bottlenecks at the SSR level (although those in LA4113 were not significant) and were the most inbred populations at the protein level (F_{IS} = 0.121 and 0.113, respectively). In addition, Escondida (LA4105) was the only population to show substantial levels of inbreeding (F_{IS} = 0.128) at the SSR loci.

Signs of bottlenecks were observed in almost all *S. sitiens* populations (exceptions were Carbonatera (LA4114) and Quimal (LA4331), in consistence with population fragmentation (Schaal and Leverich 1996).

4.2.3 Population genetic structure

4.2.3.1 Genetic partitioning

Typical for outcrossing species and regardless of the marker system, the major portion of genetic diversity resided within populations of both species. A quarter (26.9 %) of the genetic diversity present at SSR loci in *S. lycopersicoides* was partitioned among populations. The proportion was substantially lower at allozyme loci (15.3 %), reflecting their weak power to differentiate between populations in that species. In *S. sitiens* at SSR loci a fraction similar to that in the sister species was found among populations (21.5 %), but twice the amount, i.e. almost one third (30.2 %) of the genetic variation at allozyme loci.

4.2.3.2 Gene flow

This result also indicated that gene flow among populations was limited in both species. Estimated numbers of migrants per generation were below one in both species with the exception of allozyme estimates in *S. lycopersicoides*; somewhat smaller in *S. sitiens* (0.36 at allozyme loci, 0.64 at SSRs) than in *S. lycopersicoides* (0.72 at SSR loci). If less than one migrant per generation is exchanged, populations are expected to diverge over time by means of drift (Allendorf 1983), and may one day potentially develop into new species. The process will be more slowly in larger populations because of the smaller relative impact of drift. Hence, according to these results, populations of both species are diverging, but those of *S. sitiens* might do so faster.

However, as the amount of gene flow is principally governed by geographic distance and geographic barriers, it is expected to underlie great local variation. For example, in *S. lycopersicoides* gene flow within drainages will be far greater than gene flow among drainages.

It should be emphasized that gene flow is particularly difficult to measure directly (Avise 2004) and caution is warranted in the interpretation of migration estimates. Statistical models usually rest on unrealistic assumptions (e.g. the infinite island model, constant equilibrium states over time, absence of mutation and selection; Whitlock and McCauley 1999) and cannot distinguish between gene flow, random genetic drift and mutation. Although Barton and Slatkin's method (1986) is expected to be more accurate because it accounts for varying sample sizes, both statistic approaches only present gross estimates of the actual amount of gene flow.

Spatial restrictions in gene flow were also evidenced by highly significant patterns of IBD. However, both historical and current demographic events are reflected in the outcome. Considering the distinct temporal scales of the two marker types, populations of *S. lycopersicoides* may have been more connected in the remote past, possibly in the form of a mega-population, that started to split up into smaller groups of populations during more recent times. Those of *S. sitiens*, on the other hand, may even have gained some (albeit little) connectivity in the recent past.

4.2.3.3 Population genetic clusters

S. lycopersicoides. Population clusters revealed by phylogenetic analysis, principal component analysis (PCA) and STRUCTURE analysis were largely in agreement and corresponded to geographic locations with the exception of those derived from allozyme data in *S. lycopersicoides*. The structural pattern at allozyme loci in *S. lycopersicoides* was weak and 'aberrant', i.e. populations from various geographic locations mixed into the same genetic groups. These clusters were therefore not in compliance with the population groupings described below, and showed an imbalanced partitioning of genetic variation. Informativeness of the allozyme results

were somewhat questionable as outlined earlier (i.e. they may be influenced by selection pressures), and the following discussion is therefore based on microsatellite results only (unless indicated otherwise).

Populations of both species segregated into three major groups along a north/south axis. The number of major and minor clusters according to SSR data was confirmed via probabilistic derivations and supported by analysis of molecular variance (AMOVA). SSR cluster analysis in *S. lycopersicoides* revealed a major split between the northern 'Peruvian' and the central/southern 'Chilean' populations and a secondary divide within the latter group separating Putre from the two locations further south (Camarones and Camiña). External branches clustered according to drainages with only a few surprising exceptions: In the center the strong association of Perquejeque (LA2776) with Lluta (LA4320) instead of Putre #2 (LA2777) was unexpected given the spatial relationships of the three populations. However, Lluta (LA4320) may have originated from Perquejeque (LA2776), possibly through seed that traveled down the drainage towards the Pacific. The geographic scenario was similar at the southernmost distribution point (Camiña), where a close relationship between Nama (LA4126) and Camiña (LA4123) relative to the more distant Moquella (LA2730) was evident both in the dendrogram as well as the STRUCTURE analysis.

The primarily latitudinal structuring was in line with the high incidence of allele frequencies showing a geographic cline and suggested a distribution and/or fragmentation along a north/south axis. The topological specificities in that area enforce an elongated distribution. The Andean crest to the east forms a physical barrier to population expansions. In the west lowering elevations toward the Pacific are associated with environmental conditions inadequate for this species, either directly through climatic effects or indirectly by means of stronger competition and/or predator pressures.

Almost one fifth (17.3 %) of the microsatellite genetic variation in *S. lycopersicoides* was present among the three clusters, more than among populations within clusters (12.9 %). STRUCTURE analysis clearly showed the central region exhibiting the greatest amount of differentiation. Population groups at the northern and southern end of the distribution range were more homogeneous. In line with that finding, the overall content of genetic diversity, measured as percent polymorphic sites (P), allelic richness (R_S) and Nei's gene diversity (H_E), was higher in that group. The northern group was fairly inbred (F_{IS} = 0.024), while an increasing excess of outcrossing was evident towards the south (F_{IS} = -0.028 in the central and -0.034 in the southern group). Allozyme results did not reflect those trends.

The central region appeared to be the most population-rich per spatial unit, and populations tended to be larger than those in the other areas (64 vs. 53 in the north and 47 in the south). Populations showed the greatest elevational diversity in the territory around the Nevados de Putre, representing

both extremes of the species' altitudinal distribution (from 1509 m to 3736 m), spanning a total of 2227 m (vs. 1437 m in the south and a mere 607 m in the north). Geographic diversity may foster unique adaptations. Indeed, features of 'internal endemism' were evident among populations. The most genetically 'distinct' populations were Zapahuira (LA2772) and Putre #1 (LA2781), the latter of which was located at the highest distribution point. They appeared to have been somewhat isolated for some time as they showed only minimal signs of shared ancestry with other populations. A correlation between geographic and genetic diversity has been proposed in a number of studies concerning animal species (Hedrick 1986; Nevo and Shaw 1972; McDonald and Ayala 1974). However, higher levels of genetic diversity may also be the result of larger population sizes (thereby reducing the risk of genetic loss through inbreeding and drift; Lowe et al. 2004), although correlations between census population size and levels of genetic diversity in *S. lycopersicoides* were only weak. Alternatively to environmental diversity/population size, the central location may be the most diverse because it represented the center of origin of the species. Higher levels of diversity at central vs. marginal locations of the distribution range are evident in other tomato species (Rick et al. 1977, 1979).

The average pairwise genetic distance within regions confirmed the homogeneity observed in STRUCTURE analysis among populations in the north but a considerable amount of substructuring in the south (θ = 0.192 vs. 0.127 and 0.117), indicating that the division among the Camarones/Camiña locations would be the next hierarchical split in the *S. lycopersicoides* phylogeny.

In summary, *S. lycopersicoides* displayed the greatest diversity in the area around the Nevados de Putre. Although it remains rather speculative, it could be imagined that this was the species' center of origin from where colonizations occurred towards the north and the south, presumably during a time before the Andean uplift had reached the present dimensions.

S. sitiens. In *S. sitiens* three major clusters (a northern, a central and a southern) were evident both with allozyme and with microsatellite data. According to the dendrogram the principal divide was between the southernmost location (Mina la Escondida) and the remainder of populations. This was confirmed by STRUCTURE analysis at allozyme loci, whereas, at SSRs, surprisingly, the 'northern area' was identified as separate from the other regions – maybe reflecting the weaker geographic signal at the microsatellite level (recent connectivity is unlikely as explanation).

Structuring in *S. sitiens* occurred primarily across latitudinal scales, in line with the finding that most allelic trends were observed along a north/south axis. Similar to *S. lycopersicoides* the elongated distribution may have been shaped by topographical, climatic and biological specificities.

To the east the species is limited by the Andean mountain range, to the south-east by the increasingly extreme aridity of the Atacama desert, while western territories at lower elevations no longer fulfill the species' aridity requirements.

Groupings were more pronounced at the protein level: one quarter (25.2 %) of the allozyme variation was partitioned among population groups and 10.4 % among populations within groups. One fifth (19.0 %; similar to the amount in *S. lycopersicoides*) of the microsatellite genetic variation was present among population groups, three times more than within clusters (6.5 %).

The hierarchy among populations in the north was different between the allozyme and the SSR dendrogram and confirmed by the respective STRUCTURE analyses: the former revealed a close relationship between Carbonatera (LA4114) and Cere (LA4113), the latter between Paqui (LA4116) and Carbonatera (LA4114). However, the three were located at a very short distance from each other (LA4116 was only 9 km away from LA4113 and LA4114 halfway between the two), and are likely to be connected via ongoing gene flow, which confounds ancestral relationships. The higher degree of homogeneity within that group relative to other regions was confirmed by a lower average pairwise genetic distance. Paqui (LA4116) harbored the most genetic diversity and may have given rise to the other two populations, but one should take into account that the present analysis only represented a snapshot, and that demographic fluctuations may lead to transient alterations in the genetic picture.

The southernmost population Escondida (LA4105) was clearly the most isolated population, exhibiting the lowest incidence of admixture. Separated by a distance of 126 km from the next known existing population (LA4111), it has probably been barred from genetic exchange with other *S. sitiens* populations for quite some time. The average pairwise genetic distance in that population was ca. 50 % (at SSR loci) to > 100 % (at allozyme loci) higher compared to the remainder of populations. However, genetic distance at allozyme loci reached only two thirds of the maximum value found for *S. chilense* populations (F_{ST} = 0.76; Graham 2005), another tomato species adapted to dry conditions – albeit with a much wider distribution range. Its extremely small current population size (consisting of an estimated 25 individuals) will accelerate genetic divergence further through the impact of drift. Although the potential number of mutations is lowered by small population sizes, genetic changes (introduced via mutations of gene flow) have a greater chance to become established (Handel 1983). Escondida (LA4105) was the only *S. sitiens* population that showed substantial levels of inbreeding at SSR loci (F_{IS} = 0.128). Given these circumstances, one would suspect to find severely reduced levels of genetic diversity in that population. Indeed, allozyme estimates were clearly smaller than in the rest of the populations, but the trend was less

pronounced at the DNA level where individual diversity estimates even surpassed some of those of the populations in the north.

The small number of populations in combination with the genetic diversity pattern made it near to impossible to pinpoint an area as the center of diversity in *S. sitiens*. Genetic variation, especially that at allozyme loci, declined progressively in the three southern populations. Southern establishments may therefore have originated via sequential founder events from the central/northern part of the present distribution. Geographic trends were weaker at microsatellites, and the most genetic variation was displayed in the central population Limón Verde (LA4112). Colonizations may therefore have taken place further back in the past, so that the signature at the SSR sites has largely been eroded by mutations. However, in view of the current demographic picture and the continuing population decline, it seems likely that fragmentation has played a major role in forming the species' genetic structure, thereby confounding historic events.

4.2.4 Implications for conservation strategies

Recent field observations suggest that both *S. lycopersicoides* and *S. sitiens* are endanger of extinction, making conservation efforts a high priority. Preservation strategies should be aimed at capturing the highest amount of genetic diversity at the lowest possible cost. By providing valuable insights in a species' diversity status, population genetic studies can help to meet that goal.

When most of the genetic diversity is partitioned among populations rather than within, a conservation plan should integrate a large number of populations (Ellstrand and Elam 1993). The largest portion of the genetic variation is distributed within populations in *S. lycopericoides* and *S. sitiens*, but the strong pattern of isolation by distance in combination with the small number of populations would justify broad conservation efforts. Compared to other tomato species a relatively high amount of genetic diversity is distributed among a small number of individuals. Hence, the loss of individuals in *S. lycopersicoides* and *S. sitiens* is associated with an over-proportional erosion of genetic diversity.

In *S. lycopersicoides* sampling should be done for each drainage with focus on the central area in proximity to the Nevados de Putre. The more homogeneous genetic makeup of the populations at the southern and especially at the northern distribution edge would allow the maintenance of fewer accessions from that region, if necessary. The most diverse populations, Zapahuira (LA2772) and Lluta (LA4320) can be considered particularly valuable for the central, Palca (LA1966) for the northern and Moquella (LA2730) for the southern representation. Of the severely fragmented species *S. sitiens* all populations analyzed in the present study should form part of a conservation

program. The sample number may be reduced at the northern distribution end, where priority should be given to the most diverse population of that group: Paqui (LA4116).

In addition, conservation programs should give special attention to populations with idiosyncratic features. These may be spotted as geographic outliers, by a large genetic distance to other populations or a high occurrence of private alleles. In that respect of great interest is *S. lycopersicoides* Putre #1 (LA2781), located at the highest elevation point of all known tomato populations (ca. 3800 m) and displaying a great genetic distance, Aricota #1 (LA4018), also of great genetic distance, and Zapahuira (LA2772) and Pachica (LA4130) which contain several private alleles.

Among the *S. sitiens* populations the most unique is certainly Escondida (LA4105), which is probably completely isolated from the remainder of populations. It was the most genetically distant population and also harbored the highest number of private alleles. Due to its small population size it is expected to diverge fast and is particularly vulnerable of going extinct. Genetically distant and rich in private alleles was furthermore Limón Verde (LA4112).

Substantial *ex situ* conservation efforts have been undertaken over the past decades. Samples of largely all existing (and some extinct) *S. lycopersicoides* and *S. sitiens* populations, collected during five collection trips in the region or obtained by donations, are maintained at the Tomato Genetics Resource Center (TGRC). Populations in the wild are declining, therefore it is unlikely that the *ex situ* collection can be improved by further sampling. Future efforts should be directed to *in situ* conservation; the establishment of reserves and provision of ecological buffers. *In situ* programs may be particularly effective for *S. lycopersicoides*, where populations are threatened by grazing of mammals, but may be harder to realize for *S. sitiens*, where the decline seems primarily a consequence of severe aridity.

The results of the current analysis will aid to conserve the greatest variability at the lowest possible effort, both *in situ* and *ex situ*. It will also benefit efforts of tomato germplasm enhancement: knowledge of levels and patterns of genetic diversity in combination with environmental data will make screening for beneficial traits more efficient. Techniques such as Ecotilling (Comai et al. 2004) could be employed on selected material to identify informative polymorphisms.

5. Summary

The objectives of the present study were : a) to determine the degree of macrosynteny of the genomes of *S. ochranthum* and *S. juglandifolium* with that of tomato (*S. lycopersicum*) in a comparative mapping analysis, and b) to examine levels and structure of genetic diversity in wild populations of *S. lycopersicoides* and *S. sitiens* at both the protein and DNA level.

A pseudo-F_2 population comprising 66 plants was generated from an interspecific cross of *S. ochranthum* LA3650 and *S. juglandifolium* LA2788. Single copy markers of intermediate to high confidence (LOD $\geq$ 2) were selected from tomato reference maps to provide for genome coverage at an average mapping distance of ca. 10 cM. A total of 132 markers were included in the final analysis; 96 CAPS, 19 RFLPs and 17 microsatellites. The majority of these (62 %) belonged to the 'Conserved Ortholog Set' I and II, a group of single/low copy loci that are conserved between tomato and Arabidopsis or the euasterid plant species and Arabidopsis, respectively. The remainder of markers were tomato genomic clones ('TG'; 26 %), tomato cDNA clones ('CT'; 2 %) and microsatellites (13 %). Linkage analysis and map construction were conducted in MapMaker version 2.0 for Macintosh (Lander et al. 1987) using the Kosambi mapping function (Kosambi 1944). Linkage groups were assigned at threshold parameters of LOD $\leq$ 4 and a recombination fraction $\geq$ 0.3. The stringency was raised to LOD = 6 in order to resolve a spurious association of two chromosomes.

Significant segregation distortion ($P < 0.05$) was detected at one third (32 %) of the genome. Strongest deviations from Mendelian ratios were observed on chromosome 2, 5 and 9, extending (chromosome 2 and 9) across the entire chromosomal range. A total of 13 putative segregation distorter loci were identified on nine of the twelve chromosomes, several of which may be preserved among related species. All three genotypes were present at any locus and segregation across the total genome matched the expected 1:2:1 ratio. Linkage analysis revealed the presence of 12 linkage groups which corresponded to the 12 tomato chromosomes with two exceptions: tomato chromosome 1 was split in two linkage groups, and tomato chromosome 8 and 12 emerged as a single large linkage group due to 'pseudolinkage' caused by a reciprocal translocation in one of the parental species. The result was in line with the observed reduced pollen viability in the F_1 (38 %) and irregular chromosome pairing.

Total map length was only 790 cM, a 42 % reduction relative to the tomato reference map. The outcome was surprising given that the two are expected to have diverged only recently, and may therefore be the work of recombination modifiers in one sex. Chromosome arms J3S and J10S were most severely stunted (nearly 100 %). Collinearity with the tomato genome was high; apart from the translocation in one of the species a total of only two interchromosomal and four intrachromosomal

sinlge-marker translocations were detected. Three of these, (on 3S, 6S and on 8S), may be indicative of whole-arm paracentric inversions, which require further analyses for confirmation.

S. ochranthum and *S. juglandifolium* overlap in their present-day distribution and may have originated in sym~/parapatry through the establishment of an interchromosomal translocation. The two are the closest outgroup to sect. *Lycopersicon* and yet reproductive barriers to tomato (*S. lycopersicum*) are more pronounced than in subsect. *Lycopersicoides*, which is basal to the two groups. However, crossing relationships are based on hybridization attempts with *S. ochranthum* only. Therefore, if the translocation was functional in the crossing barrier, and if the carrier of the translocation was *S. ochranthum*, then prospects for hybridizations with *S. juglandifolium* might be better.

A total of 11 allozyme and 15 microsatellite markers were implemented to analyze genetic diversity in 14 *S. lycopersicoides* and 7 *S. sitiens* populations, represented by 11 - 29 plants each. Both narrowly endemic species, their entire distributional range was covered by the given sample. Descriptive diversity estimators (number of alleles k, number of private alleles $k_{private}$, percent polymorphic sites P, allelic richness R_S and R_T, Nei's gene diversity (i.e. the expected heterozygosity H_E), observed heterozygosity H_O and the inbreeding coefficient per population F_{IS}) were evaluated and tested for spatial correlations. Matrices of pairwise genetic distances were compared to those of geographic and elevational distances to test the hypothesis of isolation by distance. Partitioning of genetic diversity was determined by analysis of molecular variance (AMOVA). Phylogenies were established via cluster analysis and confirmed by principle component analysis (PCA). Admixture within the total population was detected by STRUCTURE analysis.

As expected, at the protein level levels of genetic variability were lower (e.g. H_E per locus = 0.116 vs. 0.421 in *S. lycopersicoides* and 0.285 vs. 0.354 in *S. sitiens*). Congruence between the two marker systems was low in *S. lycopersicoides*, which may be caused by selection pressures at allozyme loci, scoring errors or, alternatively, may reflect differential historic demographic events. According to microsatellites *S. lycopersicoides* was slightly more diverse than *S. sitiens* (e.g. H_E per population = 0.442 vs. 0.360), in line with its smaller demographic representation. Both species were shown to harbor less genetic variability than *S. peruvianum* or *S. chilense*, but similar amounts as other outcrossing tomato species such as *S. pennellii*, in spite of the relatively narrow distribution of the two species. Populations of both species are diverging, however, *S. sitiens* appeared to have a longer history of fragmentation and may diverge at a faster rate; populations are smaller on average and scattered out as small vegetation islands in the Atacama desert. The present analysis also demonstrated the particular vulnerability of that species; the population decline that has been

observed over the past decades was mirrored by signatures of recent bottleneck events in nearly all populations. A pattern of isolation by distance was evident in both species, and, together with spatial correlations of allele frequencies and genetic diversity, was suggestive of historic population expansions and/or population fragmentation along a north/south axis. Three population clusters (a northern, a central and a southern) could be identified in each species. Putative centers of diversity were the central region around the Nevados de Putre in *S. lycopersicoides* and the northern/central area in *S. sitiens*. The remote, southernmost population Escondida (LA4105) has probably been barred from genetic exchange with the remainder of *S. sitiens* populations for a longer period of time.

To encounter further genetic erosion in the two *Solanum* species, conservation efforts, especially those *in situ*, will need to be intensified. Results from the current analysis may provide valuable guidelines towards that goal.

The four tomato-like nightshades, *S. ochranthum*, *S. juglandifolium*, *S. lycopersicoides* and *S. sitiens* are unique within the tomato clade with repect to their adaptations to specific environmental conditions – characterized by either extreme moisture or extreme cold/drought. They are expected to possess traits that are of great value for modern tomato breeding, which is reliant on the reservoir of genetic diversity found in the related wildspecies to enrich the narrow genetic background of the cultivar. The generation of a *S. ochranthum* × *S. juglandifolium* comparative linkage map will elucidate the accessability of the *Juglandifolium* genome for hybridizations with cultivated tomato. Knowledge about the levels of genetic diversity, its distribution and special features connected with environmental data in populations of *S. lycopersicoides* and *S. sitiens* will enable a more directed and efficient search for specific valuable traits in these two species.

6. Reference list

Afify A (1933) The cytology of the hybrid between *Lycopersicon esculentum* and *L. racemigerum* in relation to its parents. Genetica 15: 225-240.

Allendorf FW (1983) Isolation, gene flow and genetic differentiation among populations. In: Genetics and Conservation. Schonewalk-Cox CM, Chambers SM, MacBryde B, Thomas L (eds.). London. Benjamin/Cummings. pp: 51-65.

Alpers CN Brimhall CH (1988) Middle Miocene climatic change in the Atacama Desert, northern Chile: evidence from supergene mineralization at La Escondida. Geological Society of America Bulletin 100: 1640-1656.

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B (2001) Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theoretical and Applied Genetics 103(8): 1283-1292.

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nature Genetics 13(4): 390-391.

Anderson TJC, Su X-Z, Roddam A, Day KP (2000) Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Molecular Ecology 9(10): 1599-1608.

Arens P, Odinot P, Van Heusden AW, Lindhout P, Vosman B (1995) GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. Genome 38(1): 84-90.

Areshchenkova T, Ganal MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. Genome 42(3): 536-544.

Areshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theoretical and Applied Genetics 104(2-3): 229-235.

Arroyo MTK, Primack R, Armesto J (1982) Community studies in pollination ecology in the high temperate Andes of central Chile 1. Pollination mechanisms and altitudinal variation. American Journal of Botany 69(1): 82-97.

Arroyo MTK, Squeo FA, Armesto JJ, Villagran C (1988) Effects of aridity on plant diversity in the northern Chilean Andes. Results of a natural experiment. Annals of the Missouri Botanical Garden 75(1): 55-78.

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9: 208-218.

Avise JC (2000) Phylogeography. The history and formation of species. Cambridge, MA: Harvard University Press.

Avise JC (2004) Molecular markers, natural history, and evolution. Sinauer Associates, Inc. Publishers Sunderland, Massachusetts.

Awadalla P, Ritland K (1997) Microsatellite variation and evolution in the *Mimulus guttatus* species complex with contrasting mating systems. Molecular Biology and Evolution 14(10): 1023-1034.

Azaiez A, Bouchard EF, Jean M, Belzile FJ (2006) Length, orientation, and plant host influence the mutation frequency in microsatellites. Genome 46(11): 1366-1373.

Bai Y, Szinay D, Lang C, Tang X, van der Vossen E, Bachem C, van Ham R, Klein Lankhorst R, Peters S, Teikema W, Visser R, de Jong H (2007) Comparative genetics facilitated by FISH mapping: chromosome 6S inversion between tomato and potato detected by cross-species multi-colour FISH. Proceedings of the 4th *Solanaceae* Genome Workshop, Jeju, S. Korea, September 9-13, p: 92.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. Molecular Ecology 11(2): 155-165.

Bamberg, JB, Martin MW, Schartner JJ (1994) Elite selections of tuber-bearing *Solanum* species germplasm. Report of Inter-Regional Potato Introduction Station, NRSP-6, USDA.

Barakat A, Carels N, Bernardi G (1997) The distribution of genes in the genomes of *Gramineae*. Proceedings of the National Academy of Sciences of the United States of America 94: 6857-6861.

Barton NH, Slatkin M (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. Heredity 56(3): 409-416.

Baudry E, Kerdelhue C, Innan H, Stephan W (2001) Species and recombination effects on DNA variability in the tomato genus. Genetics 158(4): 1725-1735.

Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. Nature 356(6369): 519-520.

Bennett MD, Leitch IJ (2004) Plant DNA C-values database. http://www.rbgkew.org.uk/cval/homepage.html.

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. Genetics 147: 861-877.

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random complementary DNA sequences. Genetics 112(4): 887-898.

Bindler G, van der Hoeven R, Gunduz I, Plieske J, Ganal M, Rossi L, Gadani F, Donini P (2007) A microsatellite marker based linkage map of tobacco. Theoretical and Applied Genetics 114(2): 341-349.

Bitter G (1912) Solana nova vel minus cognita. III.X. Sectio: *Gonatotrichum* Bitter, nov. sect. Repert. Spec. Nov. Regni Veg. 11: 230-234.

Bohn GW, Tucker CM (1940) Studies on fusarium wilt of the tomato. I. Immunity in *Lycopersicon pimpinellifolium* Mill. and its inheritance in hybrids. Missouri Agricu. Expt. Sta. Res. Bull. 311:1-82.

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095-1103.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32: 314-331.

Boulter D, Peacock D, Guise A, Gleaves JT, Estabrook G (1979) Relationships between the partial amino acid sequences of plastocyanin from members of ten families of flowering plants. Phytochemstistry 18: 603-608.

Boutin SR, Young ND, Olson T, Yu Z-H, Shoemaker RC (1995) Genome conservation among three legume genera detected with DNA markers. Genome 38: 928-937.

Brandes A, Thompson H, Dean C, Heslop-Harrison JS (1997) Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Research 5(4): 238-246.

Breto MP, Asins MJ, Carbonell EA (1993) Genetic variability in *Lycopersicon* species and their genetic relationships. Theoretical and Applied Genetics 86(1): 113-120.

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Molecular and General Genetics 250(1): 39-49.

Budiman MA, Mao L, Wood TC, Wing RA (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Research 10(1): 129-136.

Burke JM, Lai Z, Salmaso M, Nakazato T, Tang S, Heesacker A, Knapp SJ, Rieseberg LH (2004) Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics 167(1): 449-457.

Burnham CR (1962) Discussions in Cytogenetics. Burgess Publishing, Minneapolis.

Burnham CR (1991) Discussions in Cytogenetics. Burgess Publishing Co., Minneapolis.

Buteler MI, Jarret RL, LaBonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. Theoretical and Applied Genetics 99(1-2): 123-132.

Caicedo AL, Purugganan MD (2005) Comparative plant genomics. Frontiers and prospects. Plant Physiology (Rockville) 138(2): 545-547.

Caicedo AL, Schaal BA (2004) Population structure and phylogeography of *Solanum pimpinellifolium* inferred from a nuclear gene. Molecular Ecology 13(7): 1871-1882.

Cameron DR, Moav RM (1957) Inheritance in *Nicotiana tabacum* XXVII. Pollen Killer, an alien genetic locus inducing abortion of microspores not carrying it. Genetics 42: 326–335.

Canady MA, Meglic V, Chetelat RT (2005) A library of *Solanum lycopersicoides* introgression lines in cultivated tomato. Genome 48(4): 685-697.

Carr GD (1977) A cytological conspectus of the genus *Calycadenia* (*Asteraceae*): an example of contrasting modes of evolution. American Journal of Botany 64: 694-703.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Evolution 32: 550-570.

Caviedes C (1973) A climatic profile of the north Chilean desert at latitude 20° S. In: (Amiran DHK, Wilson AW, eds.) Coastal deserts: Their natural and human environments. Tucson, Arizona : University of Arizona Press, pp: 115-121.

Chaw S-M, Chang C, Chen H-L, Li W-L (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. Journal of Molecular Evolution 58(4): 424-441.

Chetelat RT (2006) TGRC Annual Progress Report 2006. C.M. Rick Tomato Genetics Resource Center, Dept. of Plant Sciences, University of California, Davis, CA.

Chetelat RT (2006) TGRC Revised List of Miscellaneous Stocks. C.M. Rick Tomato Genetics Resource Center, Dept. of Plant Sciences, University of California, Davis, CA.

Chetelat RT, Cisneros P, Stamova L, Rick CM (1997) A male-fertile *Lycopersicon esculentum* × *Solanum lycopersicoides* hybrid enableous direct backcrossing to tomato at the diploid level. Euphytica 95: 99-108.

Chetelat RT, Deverna JW, Bennett AB (1995) Introgression into tomato (*Lycopersicon esculentum*) of the *L. chmielewskii* sucrose accumulator gene (*sucr*) controlling fruit sugar composition. Theoretical and Applied Genetics 91(2): 327-333.

Chetelat RT, Ji (2007) Cytogenetics and evolution. In: Razdan MK, Mattoo AK (eds.) Genetic Improvement of Solanaceous Crops. vol 1: Tomato. Science Publ, Enfield, NH, USA, pp: 77-112.

Chetelat RT, Meglic V (2000) Molecular mapping of chromosome segments introgressed from *Solanum lycopersicoides* into cultivated tomato (*Lycopersicon esculentum*). Theoretical and Applied Genetics 100(2): 232-241.

Chetelat RT, Meglic V, Cisneros P (2000) A genetic map of tomato based on BC_1 *Lycopersicon esculentum* × *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154(2): 857-867.

Chetelat RT, Pertuzé RA (2006) Recent wild species collections from Chile. Tomato Genetics Cooperative Report 56: 13-14.

Child A (1990) A Synopsis of *Solanum* Subgenus *Potatoe* G. Don D'arcy Tuberarium Dun. Bitter S.L. Feddes Repertorium 101(5-6): 209-235.

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficienct discovery of DNA polymorphisms in natural populations by Ecotilling. Plant Journal 37(5): 778-786.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144: 2001-2014.

Correll DS (1958) A new species and some nomenclatural changes in *Solanum*, section *Tuberarium*. Madroño 14: 232-238.

Correll DS (1961) A new species and some nomenclatural changes in section *Tuberarium* of *Solanum*. Wrightia 2: 169-197.

Correll DS (1962) The potato and its wild relatives. Texas Research Foundation, Renner, Texas.

D'Arcy WG (1979) The classification of the *Solanaceae*. In: Hawkes JG, Lester RN, Skelding AD (eds.) The biology and taxonomy of the *Solanaceae*. Academic Press, London, pp: 3-47.

D'Arcy WG (1972) *Solanaceae* studies II: Typifications of subdivisions of *Solanum* section *Tuberarium*. Ann. Missouri Bot. Gard. 59: 262-278.

D'Arcy WG (1991) The *Solanaceae* since 1976, with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M, Estrada N (eds.) *Solanaceae* III: taxonomy, chemistry, evolution. Royal Botanic Gardens, Kew, pp: 75-137.

Darwin SC, Knapp S, Peralta IE (2003) Taxonomy of tomatoew in the Galapagos Islands: native and introduced species of *Solanum* sect. *Lycopersicon* (*Solanaceae*). Systematics and Biodiversity 1(1): 29-53.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. Science 292: 673-679.

Dawkins R (1982) The extended phenotype. W. H. Freeman, Oxford and San Francisco.

De Vicente MC, Tanksley SD (1991) Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theoretical and Applied Genetics 83(2): 173-178.

DeVerna JW, Chetelat RT, Rick CM (1987) Cytogenetic, electrophoretic, and morphological analysis of progeny of sesquidiploid *Lycopersicon esculentum* - *Solanum lycopersicoides* hybrids × *L. pennellii*. Biologisches Zentralblatt 106: 417-428.

DeVerna JW, Rick CM, Chetelat RT, Lanini BJ, Alpert KB (1990) Sexual hybridization of *Lycopersicon esculentum* and *Solanum rickii* by means of a sesquidiploid bridging hybrid. Proceedings of the National Academy of Sciences of the United States of America 87: 9496-9490.

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell 12: 637-646.

Dhuyvetter H, Gaublomme E, Desender K (2004) Genetic differentiation and local adaptation in the salt-marsh beetle *Pogonus chalceus*: a comparison between allozyme and microsatellite loci. Molecular Ecology 13(5): 1065-1074.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences of the United States of America 91(8): 3166-3170.

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380: 152-154.

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, Deangelis MM, Levinson DM, Kruglyak L, Goodman N, Copeland NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1996) A comprehensive genetic map of the mouse genome. Nature 380: 149-152.

Dirlewanger E, Graziano E, Joobeur T, Garriga-Calderé F, Cosson P, Howad W, Arús P (2004) Comparative mapping and marker-assisted selection in *Rosaceae* fruit crops. PNAS 101(26): 9891-9896.

Dobzhansky T (1937) Genetics and the origin of species. Columbia Univ. Press, New York.

Dobzhansky T (1972) Species of *Drosophila* - new excitement in an old field. Science 177: 664-669.

Doganlar S, Frary A, Daunay M-C, Lester RN, Tanksley SD (2002a) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the *Solanaceae*. Genetics 161(4): 1697-1711.

Doganlar S, Frary A, Daunay M-C, Lester RN, Tanksley SD (2002b) Conservation of gene function in the *Solanaceae* as revealed by comparative mapping of domestication traits in eggplant. Genetics 161(4): 1713-1726.

Doganlar S, Frary A, Ku H-M, Tanksley SD (2002c) Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589). Genome 45(6): 1189-1202.

Doganlar S, Frary A, Tanksley SD (2000) The genetic basis of seed-weight variation: tomato as a model system. Theoretical and Applied Genetics 100: 1267-1273.

Dunal MF (1816) Solanorum generumque affinium synopsis, Ed. 2. Renaud, Montpellier.

Dunal MF (1852) *Solanaceae*. In: de Candolle A (ed.) Prodromus systematis naturalis regni vegetabilis 13(1): 1-690.

Dvorak J, Akhunov ED (2005) Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploidy evolution in the *Aegilops-Triticum* alliance. Genetics 171: 323-332.

Egashira H, Ishihara H, Takashina T, Imanishi S (2000) Genetic diversity of the '*peruvianum*-complex' (*Lycopersicon peruvianum* (L.) Mill. and *L. chilense* Dun.) revealed by RAPD analysis. Euphytica 116(1): 23-31.

Ehlenfeldt MK, Hanneman RE Jr (1992) Endosperm dosage relationships among *Lycopersicon*-Spp. Theoretical and Applied Genetics 83(3): 367-372.

El Mousadik A, Petit RJ, 1996. High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* L. Skeels) endemic to Morocco. Theoretical and Applied Genetics 92: 832-839.

Ellegren H (2000) Microsatellite mutations in the germline: Implications for evolutionary inference. Trends in Genetics 16(12): 551-558.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: Implications for Plant Conservation. Annual Review of Ecology and Systematics 24: 217-242.

Engels WR, Preston CR (1984) Formation of chromosome rearrangements by P factors in *Drosophila*. Genetics 107(44): 657-678.

England PR, Usher AV, Whelan RJ, Ayre DJ (2002) Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. Molecular Ecology 11: 967-977.

Ercolano MR, Sebastiano A, Monti L, Frusciante L, Barone A (2005) Molecular characterization of *Solanum habrochaites* accessions. Journal of Genetics and Breeding 59(1): 15-20.

Estoup A, Jarne P, Cournet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11: 1591-1604.

Estoup A, Rousset F, Michalakis Y, Cornuet JM, Adriamanga M, Guyomard R (1998) Comparative analysis of microsatellite and allozyme markers: A case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). Molecular Ecology 7(3): 339-353.

Evanno, G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes application to human mitochondrial DNA restriction data. Genetics 131(2): 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7(4): 574-578.

FAO (2002) On-line crop production statistics. (http://apps.fao.org). FAO, Rome, Italy.

FAOSTAT (2005) Food and Agriculture Organization of the United Nations. Statistical Databases (http://faostat.fao.org). FAO, Rome, Italy.

Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GM, Tanksley SD, Giovannoni JJ (2004) Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. The Plant Journal 40(1): 47–59.

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

Feingold S, Lloyd J, Noreno N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.) Theoretical and Applied Genetics 111(3): 456-466.

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Foote T, Roberts TM, Kurata N, Sasaki T, Moore G (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the *Ph1* locus in wheat. Genetics 147(2): 801–807.

Frary A, Presting GG, Tanksley SD (1996) Molecular mapping of the centromeres of tomato chromosomes 7 and 9. Molecular and General Genetics 250(3): 295-304.

Frary A, Fulton TM, Zamir D, Tanksley SD (2004) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *L. pennellii* cross and identification of possible orthologs in the *Solanaceae*. Theoretical and Applied Genetics 108(3): 485-496.

Frary A, Nesbitt T, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science (Washington DC) 289(5476): 85-88.

Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley S (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theoretical and Applied Genetics 111: 291-312.

Fulton TM, Bucheli P, Voirol E, Lopez J, Petiard V, Tanksley SD (2002b) Quantitative trait loci (QTL) affecting sugars, organic acids and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. Euphytica 127(2): 163-177.

Fulton TM, Chunwongse J, and Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Molecular Biology Reporter 13 (3): 207-209.

Fulton TM, Grandillo S, Beck-Bunn T, Fridman E, Frampton A, Lopez J, Petiard V, Uhlig J, Zamir D, Tanksley SD (2000) Advanced backcross QTL analysis of a *Lycopersicon esculentum* × *Lycopersicon parviflorum* cross. Theoretical and Applied Genetics 100(7): 1025-1042.

Fulton TM, Nelson JC, Tanksley SD (1997) Introgression and DNA marker analysis of *Lycopersicon peruvianum*, a wild relative of the cultivated tomato, into L*ycopersicon esculentum*, followed through three successive backcross generations. Theoretical and Applied Genetics 95: 895-902.

Fulton T, van der Hoeven R, Eannetta N, Tanksley S (2002a) Identification, analysis and utilization of a conserved ortholog set (COS) markers for comparative genomics in higher plants. Plant Cell 14(7): 1457-1467.

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 282: 656-659.

Gao LZ, Schaal BA, Zhang CH, Jia JZ, Dong YS (2002) Assessment of population genetic structure in common wild rice *Oryza rufipogon* Griff. using microsatellite and allozyme markers. Theoretical and Applied Genetics 106(1): 173-180.

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. Molecular Biology and Evolution 12: 594-603.

Gavrilenko TA, Barbakar NI, Pavlov AV (1992) Somatic hybridization between *Lycopersicon* and non-tuberous *Solanum* species of the *Etuberosa* series. Plant Science 86: 203-214.

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufmann H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salamini F (1991) RFLP maps of potato and their alignment with the homoeologous tomato genome. Theoretical and Applied Genetics 83(1): 49-57.

Gebhardt C, Walkemeier B, Henselewski H, Barakat A, Delseny M, Stueber K (2003) Comparative mapping between potato (*Solanum tuberosum*) and *Arabidopsis thaliana* a reveals structurally conserved domains and ancient duplications in the potato genome. Plant Journal 34(4): 529-541.

Gidoni D, Fuss E, Burbidge A, Speckmann GJ, James S, Nijkamp D, Mett A, Feiler J, Smoker M, de Vroomen MJ, Leader D, Liharska T, Groenendijk J, Coppoolse E, Smit JJM, Levin I, de Both M, Schuch W, Jones JDG, Taylor IB, Theres K, van Haaren MJJ (2003) Multi-functional T-DNA/*Ds* tomato lines designed for gene cloning and molecular and physical dissection of the tomato genome. Plant Molecular Biology 51(1): 83-98.

Gill BS, Burnham CR, Stringham GR, Stout JT, Weinheimer WH (1980) Cytogenetic analysis of chromosomal translocations in the tomato: preferential breakage in heterochromatin. Canadian Journal of Genetics and Cytology 22: 33-341.

Glaubitz JC (2004) CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Molecular Ecology Notes 4: 309-310.

Goldstein DB, Pollock DD (1997) Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88(5): 335-342.

Goodstal FJ, Kohler GR, Randall LB, Bloom AJ, St Clair D (2005) A major QTL introgressed from wild *Lycopersicon hirsutum* confers chilling tolerance to cultivated tomato (*Lycopersicon esculentum*). Theoretical and Applied Genetics 111(5): 898-905.

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

Goudet J (2001) FSTAT a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995).

Goudet J, Raymond M, Demeeus T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Gould WA (1983) Tomato production, processing and quality. AVI Publ, Westport, CT, USA.

Gradziel T, Robinson RW (1989) *Solanum lycopersicoides* gene introgression to tomato, *Lycopersicon esculentum*, through the systematic avoidance and suppression of breeding barriers. Sexual Plant Reproduction 2: 43-52.

Graham, E (2005) Genetic diversity and crossing relationships of *Lycopersicon chilense*. Ph. D. Thesis. University of California, Davis, USA.

Grandillo S, Tanksley SD (1996a) QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species *Lycopersicon pimpinellifolium*. Theoretical and Applied Genetics 92: 935-951.

Grandillo S, Tanksley SD (1996b) Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L. esculentum* and *L. pimpinellifolium*. Theoretical and Applied Genetics 92: 957-965.

Grant V (1975) Genetics of flowering plants. Columbia University Press, New York, London.

Grant V (1981) Plant Speciation. New York, USA: Colombia University Press.

Gregory-Wodzicki KM (2000) Uplift history of the central and northern Andes: a review. GSA Bulletin 112: 1091-1105.

Gregory-Wodzicki KM (2000) Uplift history of the central and northern Andes: a review. Geological Society of America Bulletin 112: 1091-1105.

Guimarães RL, Chetelat RT, Stotz HU (2004) Resistance to *Botrytis cinerea* in *Solanum lycopersicoides* is dominant in hybrids with tomato, and involves induced hyphal death. European Journal of Plant Pathology 110(1): 13-23.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 48: 361-372.

Guri A, Dunbar LJ, Sink KC (1991) Somatic hybridization between selected *Lycopersicon* and *Solanum* species. Plant Cell Reports 10: 76-80.

Haanstra JPW, Wye C, Verbakel H, Meijer-Dekens F, van den Berg P, Odinot P, van Heusden AW, Tanksley S, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* × *L. pennellii* F_2 populations. Theoretical and Applied Genetics 99(1-2): 254-271.

Haldane JBS (1954) An exact test for randomness of mating. J Genetics 52: 631-635.

Handel SN (1983) Pollination ecology, plant population structure, and gene flow. In: Real L (ed) Pollination Biology. Orlando: Academic Press, pp: 163-211.

Hammerschmidt DE, Franklin MJ (2005) The edible wolf peach. J Lab Clin Med 146: 251-252.

Handley LW, Nickels RL, Cameron MW, Moore PP, Sink KC (1986) Somatic hybrid plants between *Lycopersicon esculentum* and *Solanum lycopersicoides.* Theoretical and Applied Genetics 71: 691-697.

Harfe BD, Jinks-Robertson S (2000) Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae.* Genetics 156(2): 571-578.

Harlan JR, de Wet JMJ (1971) Towards a traditional classification of cultivated plants. Taxon 20: 509–517.

Harris H (1966) Enzyme polymorphisms in man. Proceedings of the Royal Society of London Series B 164: 298-310.

Hart GE (1983) Genetics and evolution of multilocus isozymes in hexaploid wheat. Isozymes Curr Top Biol Med Res 10: 365-380.

Hawkes JG (1990) The potato: Evolution, biodiversity and genetic resources. Belhaven, London.

Hedrick UP, Booth NO (1907) Mendelian characters in tomatoes. Proceedings of the American Society for Horticultural Science 5: 19-24.

Hedrick PW (1986) Genetic polymorphisms in heterogeneous environments: A decade later. Annual Review of Ecological Systems 17: 535-566.

Hijmans RJ, Cruz M, Rojas E, Guarino L (2001) DIVA-GIS, version 1.4. A geographic information system for the management and analysis of genetic resources data. Manual. International Potato Center and International Plant Genetic Resources Institute, Lima, Peru.

Hijmans RJ, Spooner DM (2001) Geographic distribution of wild potato species. American Journal of Botany 88(11): 2101-2112.

Hill WG (1987) DNA fingerprints applied to animal and bird populations. Nature 327: 98-99.

Honnay O, Adriaens D, Coart E, Jacquemyn H, Roldan-Ruiz I (2007) Genetic diversity within and between remnant populations of the endangered calcareous grassland plant *Globularia bisnagarica* L. Conservation Genetics 8: 293-303.

Hossain M, Imanishi S, Matsumoto A (1994) Production of somatic hybrids between tomato (*Lycopersicon esculentum*) and nightshade (*Solanum lycopersicoides*) by electrofusion. Breeding Science 44: 405-412

Huang S, van der Vossen EAG, Kuang H, Vleeshouwers VGAA, Zhang N, Borm TJA, van Eck HJ, Baker B, Jacobsen E Visser, RGF (2005) Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. Plant Journal 42(2): 251-261.

Hurlbert, SH (1971) The nonconcept of species diversity: a critique and alternative parameters. Ecology 52: 577-586.

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proceedings of the National Academy of Sciences of the United States of America 87: 4251-4255.

Hurst GDD, Werren JH (2001) The role of selfish genetic elements in eukayotic evolution. Nature Reviews 2: 597-606.

Ihara N, Takasuga A, Mizoshita K, Takeda H, Sugimoto M, Mizoguchi Y, Hirano T, Itoh T, Watanabe T, Reed KM, Snelling WM, Kappes SM, Beattie CW, Bennett GL, Sugimoto Y (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. Genome Research 14(10A): 1987-1998.

Jacobsen E, Daniel MK, Bergervoet-Van Deelen JEM, Huigen DJ, Ramanna MS (1994) The first second backcross progeny of the intergeneric fusion hybrids of potato and tomato after crossing with potato. Theoretical and Applied Genetics 88(2): 181-186.

Jakupciak JP, Wells RD (2000) Gene conversion (recombination) mediates expansions of CTG•CAG repeats. Journal of Biological Chemistry 275(51): 40003-40013.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends in Ecology and Evolution 11(10): 424-429.

Jeffreys AJ, Wilson V, Thein SL (1985a) Hypervariable "minisatellite" regions in human DNA. Nature 314: 67-73.

Jenkins JA (1948) The origin of the cultivated tomato. Economic Botany 2: 379-392.

Johnson FM, Kanapi CG, Richardson RH, Wheeler MR, Stone WS (1966) An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa. Proceedings of the National Academy of Sciences of the United States of America 56: 119-125.

Johnston I (1929) Papers on the Flora of Northern Chile. Contributions from the Gray Herbarium, Harvard University, vol. 85: 1-171.

Johnston IM (1929) A new chilean plant and some nomenclatorial changes. Revista Chilena de Historia Natural 33: 25-27.

Johnston SA, den Nijs TPN, Peloquin SJ, Hanneman RE Jr (1980) The significance of genic balance to endosperm development in interspecific crosses. Theoretical and Applied Genetics 57: 5-9.

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters. Implications from nuclear RFLPs. Science 256(5053): 100-102.

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends in Genetics 13(2): 74-78.

Khush GS, Rick CM (1963) Meiosis in Hybrids between *Lycopersicon esculentum* and *Solanum pennellii*. Genetica 33: 167-183.

Khush GS, Rick CM (1968) Tomato telotrisomics: origin, identification and use in linkage mapping. Cytologia 33: 137-148.

Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genetic Research 11(3): 247-269.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49: 725-738.

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences of the United States of America 75(6): 2868-2872.

Knapp S, Bohs L, Nee M, Spooner DM (2004) *Solanaceae* - a model for linking genomics with biodiversity. Comparative and Functional Genomics 5(3): 285-291.

Kobayashi RS, Deahl KD, Stommel JR, Sinden SL (1994) Evaluation of *Solanum ochranthum* as a potential source of late blight resistance. Tomato Genetics Cooperative Report 44: 15-16.

Kobayashi RS, Stommel JR, Sinden SL (1996) Somatic hybridization between *Solanum ochranthum* and *Lycopersicon esculentum*. Plant Cell Tissue and Organ Culture 45(1): 73-78.

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

Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB (2004) Comparative mapping in the *Pinaceae*. Genetics 168(1): 447-461.

Ku H-M, Vision T, Liu, J, Tanksley S (2000) Comparing sequenced segments of the tomato and Arabidopsis genomes: Large-scale duplication followed by selective gene loss creates a network of synteny Proceedings of the National Academy of Sciences of the United States of America 97(16): 9121-9126.

Labate JA, Grandillo S, Fulton T, Muños S, Caicedo A, Peralta I, Ji Y, Chetelat RT, Scott JW, Gonzalo MJ, Francis D, Yang W, van der Knaap E, Baldo AM, Smith-White B, Mueller LA, Prince JP, Blanchard NE, Storey DB, Stevens MR, Robbins MD, Wang J-F, Liedl BE, O'Connel MA, Stommel JR, Aoki K, Iijima Y, Slade AJ, Hurst SR, Loeffler D, Steine MN, Vafeados D, McGuire C, Freeman C, Amen A, Goodstal J, Facciotti D, Van Eck J, Causse M (2007) Tomato. In: Kole C (ed.) Genome mapping and molecular breeding in plants: Volume 5 Vegetables. Springer Publishing Co. NY, pp: 1-125.

Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. Genetics 150(3): 1217-1228.

Lande (1979) Effective deme sizes during long-term during long-term evolution estimated from rates of chromosomal rearrangements. Evolution 33: 234-251.

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1(2): 174-181.

Lapointes SL, Tingey WM (1984) Feeding response of the green peach aphid *Myzus persicae homoptera aphididae* to potato glandular trichomes. Journal of Economic Entomology 77(2): 386-389.

Leimu R, Mutikainen P (2005) Population history, mating system, fitness variation in a perennial herb with a fragmented distribution. Conservation Biology 19: 349-356.

Lester RN (1991) Evolutionary relationships of tomato, potato, pepino, and wild species of *Lycopersicon* and *Solanum*. In: Hawkes JG, Lester RN, Nee M, Estrada N (eds.) *Solanaceae* III: taxonomy, chemistry, evolution. Royal Botanic Gardens, Kew, pp: 283-301.

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press.

Levinson G, Gutman GA (1987) Slipped-strand mispairing a major mechanism for DNA sequence evolution. Molecular Biology and Evolution 4(3): 203-221.

Li Y-C, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: Genomic distribution, putative functions and mutational mechanisms: A review. Molecular Ecology 11(12): 2453-2465.

Lin C, Mueller LA, Mc Carthy J, Crouzillat D, Petiard V, Tanksley SD (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. Theoretical and Applied Genetics 112(1): 114-130.

Linnaeus C (1753) Species plantarum. Stockholm

Liu ZW, Biyashev RM, Maroof MAS (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theoretical and Applied Genetics 93(5-6): 869-876.

Livingstone K, Rieseberg L (2003) Chromosomal evolution and speciation: A recombination-based approach. New Phytologist 161(1): 107-112.

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK (1999) Genome mapping in Capsicum and the evolution of genome structure in the *Solanaceae*. Genetics 152(3): 1183-1202.

Loegering WQ, Sears ER (1963). Distorted inheritance of stem-rust resistance of Timstein wheat caused by a pollen-killing gene. Canadian Journal of Genetics and Cytology 5: 65–72.

Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, Lejeune-Henaut I, Burstin J (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). Theoretical and Applied Genetics 111(6): 1022-1031.

Lowe A, Harris S, Ashton P (2004) Ecological genetics: design, analysis and application. Oxford: Blackwell Publishing.

Lugon-Moulin N, Brunner H, Wyttenbach A, Hausser J, Goudet J (1999) Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (*Insectivora*: *Soricidae*). Molecular Ecology 8(3): 419-431.

Luikart G (1997a). Usefulness of molecular markers for detecting population bottlenecks and monitoring genetic change. Ph. D. Thesis. University of Montana, Missoula, USA.

Luikart G, Allendorf FW, Cornuet J-M, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity 89: 238-247.

Luikart G, Cornuet JM (1997b) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology 12(1): 228-237.

MacArthur JW (1926) Linkage studies with the tomato. Genetics 11(4): 387-405.

MacArthur JW (1934) Linkage groups in the tomato. Genetics 29: 123-133.

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping applications in rice genetic and breeding. Plant Molecular Biology 35(1-2): 89-99.

Maizonnier D, Cornu A, Farcy E (1984) Genetic and cytological maps in petunia - a sketch for chromosome Vi. Canadian Journal of Genetics and Cytology 26(6): 657-663.

Malpertuy A, Dujon B, Richard G-F (2003) Analysis of microsatellites in 13 hemiascomycetous yeast species: Mechanisms involved in genome dynamics. Journal of Molecular Evolution 56(6): 730-741.

Mantel N (1967) The detection of disease clustering and generalized regression approach. Cancer Research 27(2): 209-220.

Marshall JA, Knapp S, Davey MR, Power JB, Cocking EC, Bennett MD, Cox AV (2001) Molecular systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) using the nuclear ITS rDNA region. Theoretical and Applied Genetics 103(8): 1216-1222.

Marticorena C, Quezada M. (1977) Notas sobre *Solanum* L. Boletín de la Sociedad de Biología de Concepcion LI: 155-157.

Maskin L, Gudesblat GE, Moreno JE, Carrari FO, Frankel N, Sambade A, Rossi M, Iusem ND (2001) Differential expression of the members of the *Asr* gene family in tomato (*Lycopersicon esculentum*). Plant Science 161(4): 739-746.

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea*: Variability, patterns of mutations, and use for evolutionary studies. Theoretical and Applied Genetics 104(2-3): 436-450.

Matthiolus PA (1544) Di Pedacio Dioscoride Anazarbeo libri cinque della historia, et materia medicinale trodotti in lingua volgare Italiana, Venice.

McClean PE, Hanson MR (1986) Mitochondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics 112(3): 649-668.

McDonald JF, Ayala FJ (1974) Genetic response to environmental heterogeneity. Nature 250: 572-574.

McFadden ES, Sears ER (1946) The origin of *Triticum spelta* and its free-threshing hexaploid relatives. Journal of Heredity 37(81-89): 107-116.

Melchers G, Sacristan MD, Holder AA (1978) Somatic hybrid plants of potato and tomato generated from fused protoplasts. Carlsberg Research Communication 43: 203-218.

Menzel MY (1964) Differential chromosome pairing in allotetraploid *Lycopersicon esculentum - Solanum lycopersicoides*. Genetics 50: 855-862.

Michener CD (1962) An interesting method of pollen collecting by bees from flowers with tubular anthers. Revista Biol. Trop. 10: 167-175.

Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theoretical and Applied Genetics 80(4): 437-448.

Miller P (1754) The Gardener's Dictionary. London.

Ming R, Liu S-C, Bowers JE, Moore PH, Irvine JE, Paterson AH (2002) Construction of a *Saccharum* consensus genetic map from two interspecific crosses. Crop Science 42: 570-583.

Miwa T, Saito Y, Yamatoto Y (1958) Bulletin of the Faculty of Agriculture, University of Miyazabi 4: 153–165.

Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: Breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theoretical and Applied Genetics 100(3-4): 471-479.

Moore G, Roberts M, Aragon-Alcaide L, Foote T (1997) Centromeric sites and cereal chromosome evolution. Chromosoma 105(6): 321-323.

Moran GF, Bell JC, Turnbull JW (1989) A cline in genetic diversity in river she-oak *Casuarina cunninghamiana*. Australian Journal of Botany 37: 169-180.

Moretti A, Laterrot H, Bordat D (1990) Observations of *Solanum ochranthum* and *S. juglandifolium*. Tomato Genetics. Cooperative Report 40: 25.

Morgan TH (1912) Complete linkage in the second chromosome of the male of *Drosophila*. Science 36: 718-720.

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3(1): 175-182.

Moyle LC, Graham EB (2005) Genetics of Hybrid incompatibility between *Lycopersicon esculentum* and *L. hirsutum*. Genetics 169: 355-373.

Mueller LA, Tanksley SD, Giovannoni JJ, van Eck J, Stack S, Choi D, Kim BD, Chen M, Cheng Z, Li C, Ling H, Xue Y, Seymour G, Bishop G, Bryan G, Sharma R, Khurana J, Tyagi A, Chattopadhyay D, Singh NK, Stiekema W, Lindhout P, Jesse T, Klein-Lankhorst R, Bouzayen M, Shibata D, Tabata S, Granell A, Botella MA, Giuliano G, Frusciante L, Causse M, Zamir D (2005) The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). Comparative and Functional Genomics 6(3): 153-158.

Müller CH (1940) A revision of the genus *Lycopersicon*. USDA Misc Publ 382: 1-29.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. Cold Spring Harbor Symposium in Quantitative Biology 51: 263-273.

National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov).

Navarro A, Barton NH (2003a) Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. Evolution 57(3): 447-459.

Navarro A, Barton NH (2003b) Chromosomal speciation and molecular divergence – accelerated evolution in rearranged chromosomes. Science 300: 321-324.

Navarro A, Ruiz A (1997) On the fertility effects of pericentric inversions. Genetics 147(2): 931-933.

Neal JJ, Steffens JC, Tingey WM (1989) Glandular trichomes of *Solanum berthaultii* and resistance to the Colorado potato beetle. Entomologia Experimentalis et Applicata 51(2): 133-140.

Neal, CA, Topoleski LD (1983) Effects of the basal medium on growth of immature tomato embryos in vitro. Journal of the American Society for Horticultural Science 108: 434-438.

Nei M (1972) Genetic distance between populations. *American Naturalist* 106: 283-292.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Nevo E, Beharav A, Meyer RC, Hackett CA, Forster BP, Russell, JR, Powell W (2005) Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution Canyon', Israel. Biological Journal of the Linnean Society 84(2): 205-224.

Nevo E, Shaw CR (1972) Genetic variation in a subterranean mammal, *Spalax ehrenbergi*. Biochemical Genetics 7: 235-241.

Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. PNAS 98(21): 12084-12088.

O'Connell MA, Hanson MR (1986) Regeneration of somatic hybrid plants formed between *Lycopersicon esuclentum* and *Solanum rickii*. Theoretical and Applied Genetics 72: 59-65.

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

Oleszek W, Shannon S, Robinson RW (1986) Steroidal alkaloids of *Solanum lycopersicoides*. Acta Societatis Botanicorum Poloniae 55: 653-657.

Olmstead RG, Palmer JD (1997) Implications for the phylogeny, classification, and biogeography of *Solanum* from cpDNA restriction site variation. Systematic Botany 22(1): 19-29.

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD (1999) Phylogeny and provisional classification of the *Solanaceae* based on chloroplast DNA. In: Nee M, Symon DE, Lester RN, Jessop JP (eds.) Solanaceae IV: advances in biology and utilization. Royal Botanic Gardens, Kew, U.K, pp: 111-137.

Omidiji MO (1979) Crossability relationships between some species of *Solanum*, *Lycopersicon* and *Capsicum* cultivated in *Nigeria*. In: Hawkes JG, Lester RN, Skelding AD (ed.) The Biology and Taxonomy of the *Solanaceae*. Linnean Soc Symp Series No 7, pp: 599-604.

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences 12: 357-358.

Paran I, Goldman I, Tanksley SD, Amir D (1995) Recombinant inbred lines for genetic mapping in tomato. Theoretical and Applied Genetics 90: 542-548.

Paterson AH, Bowers JE, Burow MD, Draye X, Elsik CG, Jiang C-X, Katsar CS, Lan T-H, Lin Y-R, Ming R, Wright RJ (2000) Comparative genomics of plant chromosomes. Plant Cell 12(9): 1523-1539.

Paterson AH, Damon S, Hewitt JD, Zamir D, Rabinowitch HD, Lincoln SE, Lander ES, Tanksley SD (1991) Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. Genetics 127(1): 181-198.

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124: 735-742.

Paterson AH, Lan T-H, Reischmann KP, Chang C, Lin Y-R, Liu S-C, Burow MD, Kowalski SP, Katsar CS, DelMonte TA, Feldmann KA, Schertz KF, Wendel JF (1996) Toward a unified genetic map of higher plants, transcending the monocot−dicot divergence. Nature Genetics 14: 380 – 382.

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

Pelham J (1968) Disturbed segregation of genes on chromosome 9: gamete promoter, *Gp*, a new gene. Tomato Genetics Cooperative Report 18: 27-29.

Peng AH, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Beales J, Fish LJ, Worland AJ, Pelica F, Sudhakar D, Christou P, Snape JW, Gale MD, Harberd NP (1999) 'Green revolution' genes encode mutant gibberellin response modulators. Nature 400: 256-261.

Peralta IE, Knapp S, Spooner DM (2005) New species of wild tomatoes (*Solanum* section *Lycopersicon*: *Solanaceae*) from Northern Peru. Systematic Botany 30(2): 424-434.

Peralta IE, Spooner DM (2001) Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum L*. section *Lycopersicon* (Mill.) Wettst. subsection *Lycopersicon*). American Journal of Botany. 88(10): 1888-1902.

Peralta IE, Knapp S, Spooner DM (2004) Unpublished monograph. (http://www.sgn.cornell.edu).

Peralta I., Spooner DM (2005) Morphological characterization and relationships of wild tomatoes (*Solanum L*. Sect. *Lycopersicon*) - A Festschrift for William G D'Arcy. Monographs in Systematic Botany, Missouri Botanical Garden.

Peralta IA, Spooner DM (2007) History, origin and early cultivation of tomato (*Solanaceae*) In: Razdan MK, Mattoo AK (eds.) Genetic Improvement of Solanaceous Crops. vol 1: Tomato. Science Publ, Enfield, NH, USA, pp: 1-24.

Perez F, Menendez A, Dehal P, Quiros CF (1999) Genomic structural differentiation in *Solanum*: Comparative mapping of the A- and E-genomes. Theoretical and Applied Genetics 98(8): 1183-1193.

Pertuzé RA, Ji Y, Chetelat RT (2002) Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato. Genome 45(6): 1003-1012.

Pertuze RA, Ji Y, Chetelat RT (2003) Transmission and recombination of homeologous *Solanum sitiens* chromosomes in tomato. Theoretical and Applied Genetics 107(8): 1391-1401.

Phills BR, Provvidenti R, Robinson RW (1977a) Reaction of *Solanum lycopersicoides* to viral diseases of the tomato. Tomato Genetics Cooperative Report 27: 18.

Phills BR, Robinson RW, Shail JW (1977b) Evaluation of *Solanum lycopersicoides* for resistance to fungal disease and nematodes. Tomato Genetics Cooperative Report 27: 18-19.

Pillen K, Pineda O, Lewis CB, Tanksley SD (1996) Status of genome mapping tools in the taxon *Solanaceae*. In: Paterson AH (ed.) Genome Mapping in Plants. RG Landes Company, pp: 282-308.

Piry S, Luikart G, Cornuet J (1999) A computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity 90(4): 502-503.

Prentice HC, Lonn M, Rosquist G, Ihse M, Kindstrom M (2006) Gene diversity in a fragmented population of *Briza media*: grassland continuity in a landscape context. Journal of Ecology 94: 87-97.

Presting GG, Frary A, Pillen K, Tanksley SD (1996) Telomere-homologous sequences occur near the centromeres of many tomato chromosomes. Molecular and General Genetics 251(5): 526-531.

Prince JP, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper and comparison of synteny with tomato. Genome 36: 404-417.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2): 945-959.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43: 258-275.

Ranz JM, Casals F, Ruiz A (2001) How malleable is the eukaryotic genome? Extreme rates of chromosomal rearrangement in the genus *Drosophila*. Genome Research 11: 230-239.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248-249.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767-779.

Rick CM (1948) Genetics and development of nine male-sterile tomato mutants. Hilgardia 18: 599-633.

Rick CM (1951) Hybrids between *Lycopersicon esculentum* Mill. and *Solanum lycopersicoides* Dun. Proceedings of the National Academy of Sciences of the United States of America 37: 741-744.

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

Rick CM (1966) Abortion of male and female gametes in the tomato determined by allelic interactions. Genetics 53: 85-96.

Rick C.M. (1967) Exploiting species hybrids for vegetable improvement. Proc. XVII Int. Hort. Congr. 3:217-229.

Rick CM (1969) Controlled Introgression of Chromosomes of *Solanum pennellii* into *Lycopersicon esculentum* - Segregation and recombination. Genetics 62(4): 753-768.

Rick CM (1976) Tomato *Lycopersicon esculentum* (*Solanaceae*). In: Simmonds NW (ed.) Evolution of crop plants. Longman, London, pp: 268-273.

Rick CM (1979) Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In: Hawkes JG, Lester RN, Skelding AD (eds.) The Biology and Taxonomy of the *Solanaceae*. Linnean Soc Symp Series No 7. pp: 667-678.

Rick CM (1982a) The potential of exotic germplasm for tomato improvement. In: Vasil IK, Scowcroft WR, Frey KJ (eds.) Plant improvement and somatic cell genetics. Academic Press, New York, pp: 1-28.

Rick CM (1982b) Genetic Relationships between self incompatibility and floral traits in the tomato species. Biologisches Zentralblatt 101(2): 185-198.

Rick CM (1986a) Germplasm resources in the wild tomato species. Acta Horticulturae 190: 39-47.

Rick CM (1986b) Potential contributions of wide crosses to improvement of processing tomatoes. Hortscience 21: 881.

Rick CM (1986c) Reproductive isolation in the *Lycopersicon peruvianum* complex. In: D'arcy WG (ed.) *Solanaceae*: Biology and Systematics; Second International Symposium, St. Louis, MO, USA, Aug. 3-6, 1983. Xiii+603p. Columbia University Press: New York, N.Y., USA, pp: 477-495.

Rick CM (1988) Tomato-like nightshades: affinities, autecology, and breeders opportunities. Economic Botany 42: 145-154.

Rick CM, Butler L (1956) Cytogenetics of the tomato. Advs in Genet 8: 267-382.

Rick CM, Chetelat RT (1995) Utilization of related wild species for tomato improvement. Acta Horticulturae 412: 21-38.

Rick CM, DeVerna JW, Chetelat RT (1990) Experimental introgression to the cultivated tomato from related wild nightshades. In: Bennett AB, O'Neill SD (eds.) Horticultural biotechnology. New York: Wiley-Liss., pp: 19-30.

Rick CM, DeVerna JW, Chetelat RT, Stevens MA (1986) Meiosis in sesquidiploid hybrids of *Lycopersicon esculentum* and *Solanum lycopersicoides*. Proceedings of the National Academy of Sciences of the United States of America 83: 3580-3583.

Rick CM, Fobes JF (1975a) Allozyme variation in the cultivated tomato and closely related species. Torrey Botanical Club Bulletin 102: 376-384.

Rick CM, Fobes JF (1975b) Allozymes of Galapagos tomatoes. Polymorphism, geographic distribution and affinities. Evolution 29(3): 443-457.

Rick CM, Fobes JF, Holle M (1977) Genetic variation in *Lycopersicon pimpinellifolium*. Evidence of evolutionary change in mating systems. plant systematics and evolution 127(2-3): 139-170.

Rick CM, Fobes JF, Tanksley SD (1979) Evolution of mating systems in *Lycopersicon hirsutum* as deduced from genetic variation in electrophoretic and morphological characters. Plant Systematics and Evolution 132(4): 279-298.

Rick CM, Holle M (1990) Andean *Lycopersicon esculentum* var. *cerasiforme*. Genetic variation and its evolutionary significance. Economic Botany 44(3 suppl): 69-78.

Rick CM, Kesicki E, Fobes JF, Holle M (1976) Genetic and biosystematic studies on two new sibling species of *Lycopersicon* from interandean Peru. Theoretical and Applied Genetics 47(2): 55-68.

Rick CM, Tanksley SD (1981) Genetic Variation in *Solanum pennellii*. Comparisons with two other sympatric tomato species. Plant Systematics and Evolution 139(1-2): 11-46.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. Trends in Ecology and Evolution 16(7): 351-358.

Robbins TP, Carpenter R, Coen ES (1989) A chromosome rearrangement suggests that donor and recipient sites are associated during *Tam3* transposition in *Antirrhinum majus*. EMBO Journal 8: 5-13.

Roberts PA (1965) Difference in the behavior of eu- and heterochromatin: crossing over. Nature 205: 725-759.

Robertson LD, Labate JA (2007) Genetic resources of tomato (*Lycopersicon esculentum* var. *esculentum*) and wild relatives. In: Razdan MK, Mattoo AK (eds.) Genetic Improvement of Solanaceous Crops. vol 1: Tomato. Science Publ, Enfield, NH, USA, pp: 25-75.

Robinson RW, Kowaleski E (1974) Breeding tomatoes for frost tolerance. Proc. 19th Int Hort Congr (Warsaw) 1B: 716.

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 49(4): 2007-2023.

Roselius K, Stephan W, Staedler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. Genetics 171(2): 753-763.

Rousseaux MC, Jones MC, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theoretical and Applied Genetics. 111(7): 1396-1408.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145(4): 1219-1228.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds.) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp: 365-386.

Rydberg PA (1924) The section *Tuberarium* of the genus *Solanum* in Mexico and Central America. Bulletin of the Torrey Botanical Club 51: 145-153.

Sacks EJ, Gerhardt LM, Graham EB, Jacobs J, Thorrup TA, St Clair DA (1997) Variation among 41 genotypes of tomato (*Lycopersicon esculentum* Mill.) for crossability to *L. peruvianum* Mill. Annals Of Botany (London) 80: 469-477.

Saitou N, and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Saliba-Colombani V, Causse M, Gervais L, Philouze J (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43(1): 29-40.

Sandler L, Novitski E (1957) Meiotic drive as an evolutionary force. Am. Nat. 91: 105-110.

Sanmiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274(5288): 765-768.

Sano Y (1983) A new gene controlling sterility in F_1 hybrids of two cultivated rice species. Journal of Heredity 74(6): 435-439.

Sano Y (1990) The genic nature of gamete eliminator in rice. Genetics 125: 183–191.

Schaal BA, Leverich WJ (1996) Molecular variation in isolated plant populations. Plant Species Biology 11(1): 33-40.

Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, Kim K, Nelson JG, Anderson WA (2003) Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. PNAS 100(14): 8319-8324.

Schmidt T, Heslop-Harrison JS (1996) The physical and genomic organization of microsatellites in sugar beet. Proceedings of the National Academy of Sciences of the United States of America 93(16): 8761-8765.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. Molecular Biology and Evolution 15(12): 1751-1760.

Schug MD, Mackay TFC, Aquadro CF (1997) Low mutation rates of microsatellite loci in *Drosophila melanogaster*. Nature Genetics 15(1): 99-102.

Seah S, Yaghoobi J, Ross M, Gleason CA, Williamson VM (2004) The nematode-resistance gene, *Mi-1*, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theoretical and Applied Genetics 108(8): 1635-1642.

SGN Solanaceae Genomics Network (2007; http://www.sgn.cornell.edu).

Shaffer JP (1995) Multiple hypothesis testing. Annual review, Psychology 46: 561-584.

Sherman, J. D. and S. M. Stack (1995) Two-dimensional spreads of synaptonemal complexes from Solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics 141: 683-708.

Smith SD, Peralta IE (2002) Ecogeographic surveys as tools for analyzing potential reproductive isolating mechanisms: An example using *Solanum juglandifolium* Dunal, *S. ochranthum* Dunal, *S. lycopersicoides* Dunal, and *S. sitiens* I. M. Johnston. Taxon 51(2): 341-349.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35(4): 627-632.

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphism among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theoretical and Applied Genetics 94(2): 264-272.

Sokal RR, Sneath PHA (1963) Principles of Numerical Taxonomy. WH Freeman, San Francisco, pp: 230-234.

Solanaceae Source (2007; http://www.nhm.ac.uk/solanaceaesource)

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (*Solanaceae*). American Journal of Botany 80: 676-688.

Spooner DM, Peralta IE, Knapp S (2005) Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes (*Solanum* L. section *Lycopersicon* (Mill.) Wettst.). Taxon 54(1): 43-61.

Städler T, Roselius K, Stephan W (2005) Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. Evolution 59(6): 1268–1279.

Stephan W, Langley CH (1998) DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics 150(4): 1585-1593.

Stommel JR (2001) Barriers for introgression of *Solanum ochranthum* into tomato via somatic hybrids. Journal of the American Society for Horticultural Science 126(5): 587-592.

Stommel JR, Kobayashi RS, Sinden SL (2001) Somatic hybridization between *Lycopersicon esculentum* Mill. (tomato) and *Solanum ochranthum* Dun. Nagata T, Bajaj YPS (eds.) Biotechnology in Agriculture and Forestry. Somatic hybridization in crop improvement II. Series Information: Biotechnology in Agriculture and Forestry. Vol. 49. Springer-Verlag Berlin, pp: 218-232.

Strommer J, Peters J, Zethof J De Keukeleire P, Gerats T (2002) AFLP maps of Petunia hybrida: Building maps when markers cluster. Theoretical and Applied Genetics 105(6-7): 1000-1009.

Symon DE (1981) The Solanaceous genera, *Browallia*, *Capsicum*, *Cestrum*, *Cyphomandra*, *Hyoscyamus*, *Lycopersicon*, *Nierembergia*, *Physalis*, *Petunia*, *Salpichora*, *Withania*, naturalized in Australia. Journal of the Adelaide Botanic Gardens 3: 133-166.

Tabaeizadeh Z, Perennes C, Bergounioux C (1985) Plant Cell Reports 4: 7–11.

Tanksley SD (1984) Linkage relationships and chromosomal locations of enzyme coding genes in pepper *Capsicum annuum*. Chromosoma 89(5): 352-360.

Tanksley SD (1985) Enzyme-coding genes in tomato (*Lycopersicon esculentum*). Isozyme Bulletin 18: 43-45.

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proceedings of the National Academy of Sciences of the United States of America 85(17): 6419–6423.

Tanksley SD, Ganal MW, Prince JP, De Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141-1160.

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V, Lopez J and Beck-Bunn T (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theoretical and Applied Genetics 92: 213-224.

Tanksley SD, Loaiza-Figueroa F (1985) Gametophytic self-incompatibility is controlled by a single major locus on chromosome 1 in *Lycopersicon peruvianum*. Proceedings of the National Academy of Sciences of the United States of America 82: 5093-5096.

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277: 1063.

Tanksley SD, Medina-Filho H, Rick CM (1982) Use of naturally occurring enzyme variation to detect and map genes controlling quatitative traits in an interspecific backcross of tomato *Lycopersicon esculentum*. Heredity 49(1): 11-26.

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theoretical and Applied Genetics 92: 191-203.

Tanksley SD, Rick CM (1980) Isozymic gene linkage map of the tomato *Lycopersicon esculentum*. Applications in genetics and breeding. Theoretical and Applied Genetics 57(4): 161-170.

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39: 277-287.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17: 6463-6471.

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica (Dordrecht) 117(1): 27-35.

Terrell EE, Broome CR, Reveal JL (1983) Proposal to conserve the name of the tomato as *Lycopersicon esculentum* P. Miller and reject the combination *Lycopersicon lycopersicum* (L.) Karsten (*Solanaceae*). Taxon 32(2): 310-314.

Tétényi P (1987) A chemotaxonimic classification of the *Solanaceae*. Annals of the Missouri Botanical Garden 74: 600-608.

TGRC Tomato Genetics Resource Center (2007; http://tgrc.ucdavis.edu).

Thurieaux P (1977) Is recombination confined to structural genes on the eukaryotic chromosome? Nature 268: 460-462.

Tingey WM, Gibson RW (1978) Feeding and mobility of the potato leafhopper impaired by glandular trichomes of *Solanum berthaultii* and *Solanum polyyadenium*. Journal of Economic Entomology 71(6): 856-858.

Tingey WM, Mehlenbacher SA, Laubengayer JE (1981) Occurrence of glandular trichomes in wild *Solanum*-Sp. American Potato Journal 58(1): 81-83.

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Research 10(7): 967-981.

Tuskan and 110 others (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). Science (Washington DC) 313: 1596-1604.

USDA-ERS (2006) Tomatoew at a glance. USDA-Economic Research Service. http://www.ers.usda.gov/Briefing/Tomatoes

Van der Knaap E, Sanyal A, Jackson SA, Tanksley SD (2004) High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics 168(4): 2127-2140.

Van der Knaap E, Tanksley SD (2003) The making of a bell pepper-shaped tomato fruit: Identification of loci controlling fruit morphology in Yellow Stuffer tomato. Theoretical and Applied Genetics 107(1): 139-147.

Van Ooijen JW, Sandbrink JM, Vrielink M, Verkerk R, Zabel P, Lindhout P (1994) An RFLP linkage map of *Lycopersicon peruvianum*. Theoretical and Applied Genetics 89(7-8): 1007-1013.

Van Rossum F, Campos de Sousa S, Triest L (2004) Genetic consequences of habitat fragmentation in an agricultural landscape on the common *Primula veris* and comparison with its rare congener, *P. vulgaris*. Conservation Genetics 5: 231-245.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Molecular Biology and Evolution 19(8): 1251-1260.

Villagran C, Varela J (1990) Palynological evidence for increased aridity on the Central Chilean coast during the Holocene. Quaternary Research (Orlando) 34(2): 198-207.

Von Korff M, Wang H, Leon J, Pillen K (2005) AB-QTL analysis in spring barley. I. Detection of resistance genes against powdery mildew, leaf rust and scald introgressed from wild barley. Theoretical and Applied Genetics 111(3): 583-590.

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407-4414.

Vosman B, Arens P (1997) Molecular characterization of GATA/GACA microsatellite repeats in tomato. Genome 40(1): 25-33.

Vrebalov J, Ruezinsky D, Padmanabhan V, White R, Medrano D, Drake R, Schuch W, Giovannoni J (2002) A MADS-box gene necessary for fruit ripening at the tomato *Ripening-Inhibitor* (*Rin*) Locus. Science 296(5566): 343-346.

Wang YH, Tang X, Cheng Z, Mueller L, Giovannoni JJ, Tanksley D (2006) Euchromatin and pericentromeric heterochromatin: Comparative composition in the tomato genome. Genetics 172: 2529-2540.

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theoretical and Applied Genetics 88(1): 1-6.

Wann EV, Johnson KW (1963) Intergeneric hybridization involving species of *Solanum* and *Lycopersicon*. Botanical Gazette 124: 451-455.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Human Molecular Genetics 2(8): 1123-1128.

Weir BS (1990) Genetic data analysis. Sinauer Publications, Sunderland, MA.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

Wendel JF, Edwards MD, Stuber CW (1987) Evidence for multilocus genetic control of preferential fertilization in maize. Heredity 58: 297-301.

Whalen MD, Anderson GJ (1981) Distribution of gametophytic self-incompatibility and infrageneric classification in *Solanum*. Taxon 30: 761-767.

White MJD (1978) Modes of speciation. W.H. Freeman, San Francisco, CA.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm + 1)$. Heredity 82(2): 117-125.

Wilkstrom N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. Proceedings of the Royal Society Biological Sciences Series B 268(1482): 2211-2220.

Williams JG, Kubelik AR, Livak J, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18: 6531-6535.

Wolf S, Yakir D, Stevens MA, Rudich J (1986) Cold temperature tolerance of wild tomato species. Journal of the American Society for Horticultural Science 111(6): 960-964.

Wolfe LH, Gouy M, Yang Y-W, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proceedings of the National Academy of Sciences of the United States of America 86(16): 6201-6205.

Wright S (1951) The genetical structure of populations. Annuals of Eugenics 15: 323-354.

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single copy, orthologous genes (COSII) for comparative, evolutinonary and systematics studies: A test case in the Euasterid plant clade. Genetics 174(3): 1407-1420.

Xiao F, Tang X, Zhou J-M (2001) Expression of 35S::*Pto* globally activates defense-related genes in tomato plants. Plant Physiology 126(4): 1637-1645.

Yang W, Bai X, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis D (2004) Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. Molecular Breeding 14(1): 21-34.

Yu G, Blackburn EH (1991) Developmentally programmed healing of chromosomes by telomerase in tetrahymena. Cell 67: 823-832.

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nature Reviews Genetics 2: 983-989.

Zamir D, Tadmor Y (1986) Unequal segregation of nuclear genes in plants. Botanical Gazette 147: 355-358.

Zamir D, Tanksley SD (1988) The tomato genome is comprised largely of fast-evolving low copy-number sequences. Molecular and General Genetics 213: 254-261.

Zhao L, Qiu C, Li J, Chai Y, Kai G, Li Z, Sun X, Tang KX (2005) Investigation of disease resistance and cold tolerance of *Solanum lycopersicoides* for tomato improvement. Hortscience 40(1): 43-46.

Zhang LP, Khan A, Nino-Liu D, Foolad MR (2002) A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum* × *Lycopersicon hirsutum* cross. Genome 45(1): 133-146.

Zhong X, Fransz P, van Wennekes E, Zabel P, van Kammen A, de Jong (1996) High-resolution mapping on pachytene chromosomes and extended DNA fibres by fluorescence in-situ hybridisation. Plant Molecular Biology Reporter 14: 232-242.

Zygier S, Chaim AB, Efrati A, Kaluzky G, Borovsky Y, Paran I (2005) QTLs mapping for fruit size and shape in chromosomes 2 and 4 in pepper and a comparison of the pepper QTL map with that of tomato. Theoretical and Applied Genetics 111: 437-445.

7. List of tables

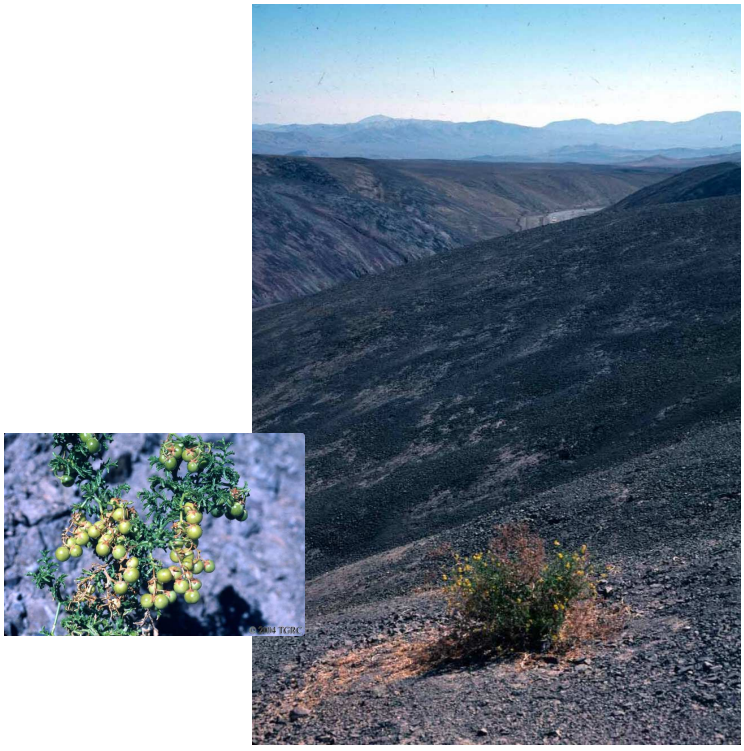
8. List of figures

9. Appendix

Appendix 1A. *S. ochranthum* LA2117 at Tun Tun, Loia, Ecuador and LA0129 (flowers and fruits) at Guaca, Carchi, Ecuador.

Appendix 1B. *S. juglandifolium* LA2118 at San Luca, Loia, Ecuador and LA2788 (flowers and fruits) at Quebrada La Buena, Antioquia, Colombia.

Appendix 1C. *S. lycopersicoides* LA2781 at Desvío a Putre, Tarapacá, Chile (3800 m) and LA2772 (fruits) at Zapahuira, Tarapacá, Chile.



Appendix 1D. *S. sitiens* LA 4105 at Mina La Escondida, Antofagasta, Chile and LA 2885 (fruits) at Caracoles, Antofagasta, Chile.

Marker name	Marker type	Marker technique	RE	F_2 *S. ochr.* × *S. jugl.* Chr	F_2 *S. ochr.* × *S. jugl.* Position (cM)	Tomato-EXPEN 2000 Chr	Tomato-EXPEN 2000 Position (cM)	LOD	Ref
C2_At1g02560	COSII	CAPS/agarose	*Hinf*I	1	66.1	1	115.7	I	1,5
C2_At1g07080	COSII	CAPS/agarose	-	6	0.8	6	3.5	I	1,5
C2_At1g10500	COSII	CAPS/agarose	*Hinf*I	5	50.1	5	76.0	I	1,5
C2_At1g20050	COSII	CAPS/agarose	*Hinf*I	6	69.1	6	101.0	I	1,5
C2_At1g35720	COSII	CAPS/agarose	*Sty*I	4	27.7	4	62.2	I	1,5
C2_At1g48300	COSII	CAPS/agarose	*Rsa*I	12	63.0	12	105.0	F	1,5
C2_At1g63980	COSII	CAPS/agarose	*Bso*BI	8	49.1	8	77.0	I	1,5
C2_At1g67700	COSII	CAPS/agarose	*Hinf*I	5	47.8	5	66.7	I	1,5
C2_At1g71810	COSII	CAPS/agarose	*Msp*I	4	36.5	4	71.5	I	1,5
C2_At1g74520	COSII	CAPS/agarose	*Ban*I	3	39.7	3	113.0	I	1,5
C2_At1g76080	COSII	CAPS/agarose	*Bso*BI	4	77.5	4	137.0	I	1,5
C2_At1g77470	COSII	CAPS/agarose	*Msp*I	6	11.2	6	39.1	I	1,5
C2_At1g78620	COSII	CAPS/agarose	*Alu*I	7	52.1	7	57.0	I	1,5
C2_At1g80360	COSII	CAPS/agarose	*Alu*I	3	51.6	3	145.5	I	1,5
C2_At1g80460	COSII	CAPS/agarose	*Hinf*I	3	43.6	3	129.5	I	1,5
C2_At2g04700	COSII	CAPS/agarose	*Alu*I	2	44.7	2	97.0	I	1,5
C2_At2g06010	COSII	CAPS/agarose	*Alu*I	12	11.9	12	32.7	I	1,5
C2_At2g20390	COSII	CAPS/agarose	*Hinf*I	4	12.6	4	28.0	I	1,5
C2_At2g26590	COSII	CAPS/agarose	*Alu*I	7	12.1	7	15.0	I	1,5
C2_At2g27290	COSII	CAPS/agarose	*Dde*I	11	39.2	11	71.0	I	1,5
C2_At2g28600	COSII	CAPS/agarose	*Msp*I	11	0.0	11	10.5	I	1,5
C2_At2g37240	COSII	CAPS/agarose	*Rsa*I	9	0.0	9	1.0	I	1,5
C2_At2g42810	COSII	CAPS/agarose	*Hinf*I	7	41.5	7	45.0	I	1,5
C2_At2g43360	COSII	CAPS/agarose	*Dde*I	6	53.4	6	67.0	CF	1,5
C2_At2g45620	COSII	CAPS/agarose	*Hae*II	1	50.8	1	69.2	I	1,5
C2_At2g45730	COSII	CAPS/agarose	*Hha*I	4	71.2	4	129.5	F	1,5
C2_At3g02300	COSII	CAPS/agarose	*Dde*I	2	48.6	2	110.2	I	1,5
C2_At3g02420	COSII	CAPS/agarose	*Msp*I	3	2.8	3	72.2	I	1,5
C2_At3g04780	COSII	CAPS/agarose	*Hae*III	1	55.5	1	98.5	I	1,5
C2_At3g12300	COSII	CAPS/agarose	-	1	54.7	1	92.0	I	1,5
C2_At3g14910	COSII	CAPS/agarose	*Alu*I	7	54.6	7	61.0	I	1,5
C2_At3g15290	COSII	CAPS/agarose	-	7	56.1	7	63.5	I	1,5
C2_At3g17000	COSII	CAPS/agarose	*Dde*I	12	68.5	12	115.0	I	1,5
C2_At3g20390	COSII	CAPS/agarose	*Dde*I	10	0.0	10	0.0	I	1,5
C2_At3g24050	COSII	CAPS/agarose	*Hha*I	9	58.6	9	97.0	CF	1,5
C2_At3g25120	COSII	CAPS/agarose	*Hinf*I	6	0.0	6	5.2	I	1,5
C2_At3g47990	COSII	CAPS/agarose	*Mse*I	3	35.0	3	101.5	I	1,5
C2_At3g51010	COSII	CAPS/agarose	*Hae*III	4	9.5	4	19.7	I	1,5
C2_At3g57270	COSII	CAPS/agarose	*Hinf*I	10	29.3	10	53.0	I	1,5
C2_At3g58470	COSII	CAPS/agarose	*Hae*III	10	43.4	10	61.0	I	1,5
C2_At4g03280	COSII	CAPS/agarose	*Msp*I	12	0.8	12	12.5	I	1,5
C2_At4g20410	COSII	CAPS/agarose	*Rsa*I	2	10.0	2	36.9	I	1,5
C2_At4g21580	COSII	CAPS/agarose	*Alu*I	2	22.9	2	68.0	I	1,5
C2_At4g24830	COSII	CAPS/agarose	*Hae*III	5	40.5	5	51.0	I	1,5
C2_At4g25650	COSII	CAPS/agarose	*Rsa*I	4	26.9	4	37.2	I	1,5
C2_At4g30930	COSII	CAPS/agarose	*Msp*I	2	21.3	2	63.1	I	1,5
C2_At4g33250	COSII	CAPS/agarose	*Hinf*I	7	12.1	7	29.0	I	1,5
C2_At4g37280	COSII	CAPS/agarose	*Bso*BI	2	58.0	2	135.0	I	1,5
C2_At5g04590	COSII	CAPS/agarose	*Hae*III	11	29.6	11	56.0	I	1,5
C2_At5g05690	COSII	CAPS/agarose	*Nci*I	6	2.3	6	24.5	I	1,5
C2_At5g06370	COSII	CAPS/agarose	*Hinf*I	1	1.5	1	18.5	I	1,5
C2_At5g06430	COSII	CAPS/agarose	*Dra*I	10	0.0	10	3.2	I	1,5

Marker name	Marker type	Marker technique	RE	F_2 *S. ochr.* × *S. jugl.* Chr	F_2 *S. ochr.* × *S. jugl.* Position (cM)	Tomato-EXPEN 2000 Chr	Tomato-EXPEN 2000 Position (cM)	LOD	Ref
C2_At5g11490	COSII	CAPS/agarose	*Alu*I	8	12.8	8	31.5	I	1,5
C2_At5g13030	COSII	CAPS/agarose	*Bso*BI	1	0.0	1	15.3	I	1,5
C2_At5g16710	COSII	CAPS/agarose	*Dra*I	11	12.6	11	31.4	I	1,5
C2_At5g19690	COSII	CAPS/agarose	*Hae*III	12	8.8	12	27.0	I	1,5
C2_At5g20180	COSII	CAPS/agarose	*Rsa*I	7	12.9	7	6.0	I	1,5
C2_At5g20350	COSII	CAPS/agarose	*Hha*I	7	9.0	7	0.0	I	1,5
C2_At5g25630	COSII	CAPS/agarose	*Rsa*I	8	15.1	8	41.5	F	1,5
C2_At5g25900	COSII	CAPS/agarose	*Alu*I	4	56.9	4	108.5	I	1,5
C2_At5g37360	COSII	CAPS/agarose	*Hae*II	4	27.7	4	56.0	I	1,5
C2_At5g38530	COSII	CAPS/agarose	-	12	48.6	12	93.0	I	1,5
C2_At5g42740	COSII	CAPS/agarose	*Hinf*I	12	19.9	12	54.5	I	1,5
C2_At5g46630	COSII	CAPS/agarose	-	8	1.7	8	0.0	I	1,5
C2_At5g54310	COSII	CAPS/agarose	-	7	59.2	7	75.0	I	1,5
C2_At5g56130	COSII	CAPS/agarose	-	7	73.9	7	108.0	I	1,5
CT068	CT	RFLP	*Eco*RV	8	55.7	8	87.0	F	1,3
CT206	CT	RFLP	*Hind*III	6	67.6	6	92.0	F	1,3
SSR115	SSR	SSR/LiCor	-	5	19.0	5	35.0	F	1,6
SSR15	SSR	SSR/LiCor	-	8	8.1	8	22.7	I	1,6
SSR300	SSR	SSR/agarose	-	3	53.9	3	166.0	F	1,6
SSR318	SSR	SSR/LiCor	-	10	7.9	10	34.5	I	1,6
SSR325	SSR	SSR/LiCor	-	5	12.5	5	18.5	I	1,6
SSR326	SSR	SSR/LiCor	-	6	33.6	6	39.0	I	1,6
SSR34	SSR	SSR/agarose	-	10	1.6	10	25.3	I	1,6
SSR345	SSR	SSR/agarose	-	12	28.5	12	72.5	I	1,6
SSR346	SSR	SSR/LiCor	-	1	65.2	1	125.0	I	1,6
SSR4	SSR	SSR/LiCor	-	10	0.0	10	3.0	I	1,6
SSR43	SSR	SSR/agarose	-	4	4.9	4	14.0	F	1,6
SSR51	SSR	SSR/LiCor	-	1	19.1	1	39.5	F	1,6
SSR565	SSR	SSR/LiCor	-	7	34.4	7	44.2	I	1,6
SSR578	SSR	SSR/LiCor	-	6	19.0	6	44.0	CF	1,6
SSR596	SSR	SSR/agarose	-	10	0.8	10	25.7	I	1,6
SSR599	SSR	SSR/LiCor	-	9	63.4	9	103.0	F	1,6
SSR74	SSR	SSR/agarose	-	10	60.2	10	74.0	l	1,6
T0142	COS	CAPS/agarose	*Hinf*I	11	43.2	11	80.0	F	1,4
T0308	COS	CAPS/agarose	*Msp*I	3	0.0	10	36.5	I	1,4
T0360	COS	CAPS/agarose	*Rsa*I	4	76.7	4	131.0	F	1,4
T0393	COS	CAPS/agarose	*Rsa*I	9	46.9	9	70.0	F	1,4
T0408	COS	CAPS/agarose	*Hha*I	11	4.6	11	26.0	F	1,4
T0724	COS	CAPS/agarose	*Rsa*I	10	63.5	10	86.0	CF	1,4
T0759	COS	CAPS/agarose	*Bso*BI	2	32.4	2	82.0	F	1,4
T0800	COS	CAPS/agarose	*Dde*I	12	68.5	12	118.0	I	1,4
T1286	COS	CAPS/agarose	*Ban*I	3	8.0	3	20.0	I	1,4
T1429	COS	CAPS/agarose	*Hinf*I	3	24.0	3	77.0	I	1,4
T1497	COS	CAPS/agarose	*Hha*I	7	13.7	7	35.0	I	1,4
T1584	COS	CAPS/agarose	*Dde*I	5	57.1	5	108.0	F	1,4
T1682	COS	CAPS/agarose	*Ava*II	10	45.7	10	66.0	I	1,4
T1706	COS	CAPS/agarose	*Rsa*I	2	5.3	2	18.0	F	1,4
T1782	COS	CAPS/agarose	*Rsa*I	1	90.2	1	156.0	F	1,4
TG144	TG	RFLP	*Hae*III	9	39.7	9	*59.2*	I	1,2
TG15	TG	CAPS/agarose	*Dra*I	4	0.0	4	6.5	F	1,2
TG167	TG	RFLP	*Hind*III	2	53.3	2	118.0	I	1,2
TG176	TG	RFLP	*Eco*RV	8	0.0	8	2.0	F	1,2

Marker name	Marker type	Marker technique	RE	F_2 *S. ochr.* × *S. jugl.* Chr	F_2 *S. ochr.* × *S. jugl.* Position (cM)	Tomato-EXPEN 2000 Chr	Tomato-EXPEN 2000 Position (cM)	LOD	Ref
TG180	TG	RFLP	*Xba*I	12	0.0	12	9.0	F	1,2
TG185	TG	RFLP	*Eco*RI	5	60.2	5	119.0	I	1,2
TG233	TG	RFLP	*Eco*RI	10	68.3	10	86.0	F	1,2
TG27	TG	RFLP	*Xba*I	1	93.4	1	165.0	CF	1,2
TG291	TG	CAPS/agarose	*Ava*II	9	23.9	9	*50.4*	I	1,2
TG294	TG	CAPS/agarose	*Rsa*I	8	55.7	8	87.0	I	1,2
TG301	TG	CAPS/agarose	*Alu*I	1	0.0	1	7.0	I	1,2
TG303	TG	CAPS/agarose	*Dde*I	10	0.0	10	11.0	F	1,2
TG31	TG	CAPS/agarose	*Hinf*I	2	2.3	2	4.0	I	1,2
TG33	TG	RFLP	*Xba*I	2	3.3	2	0.0	I	1,2
TG342	TG	CAPS/agarose	*Hae*III	7	0.0	7	*0.0*	I	1,2
TG351	TG	RFLP	*Dra*I	5	54.0	5	102.0	I	1,2
TG365	TG	CAPS/agarose	*Hinf*I	6	39.1	6	50.0	F	1,2
TG393	TG	CAPS/agarose	*Hae*III	11	48.5	11	103.0	F	1,2
TG394	TG	RFLP	*Eco*RI	12	24.6	12	68.0	F	1,2
TG403	TG	CAPS/agarose	*Dde*I	10	48.0	10	71.3	I	1,2
TG460	TG	CAPS/agarose	*Dde*I	1	50.8	1	70.0	F	1,2
TG499	TG	RFLP	*Eco*RV	7	71.6	7	97.0	F	1,2
TG510	TG	RFLP	*Hae*III	8	33.9	8	57.0	F	1,2
TG551	TG	CAPS/agarose	*Hinf*I	9	37.4	9	56.7	I	1,2
TG581	TG	CAPS/agarose	*Sty*I	4	7.2	6	96.0	F	1,2
TG596	TG	RFLP	*Eco*RV	10	0.8	10	*22.0*	I	1,2
TG608	TG	CAPS/agarose	*Hha*I	2	0.0	2	*0.0*	I	1,2
TG623	TG	RFLP	*Eco*RV	5	0.0	5	8.0	F	1,2
TG651	TG	RFLP	*Xba*I	11	3.1	11	18.0	F	1,2
TG71	TG	RFLP	*Eco*RV	1	50.8	1	*67.7*	I	1,2
TG9	TG	RFLP	*Dra*I	9	5.6	9	23.0	F	1,2
U241700	COSII	CAPS/agarose	*Mse*I	10	6.3	10	41.5	I	1

Appendix 2. Markers genotyped in F_2 *S. ochranthum* × *S. juglandifolium*.

Map positions of markers are from tomato-EXPEN 2000. Marker types were TG = tomato genomic probe; COS = conserved ortholog set; SSR = simple sequence repeat. 'Marker technique' indicates the detection mode. RE = restriction enzyme.

Note: Marker positions in italic are derived by extrapolation from adjacent markers with known locations on both tomato-EXPEN 1992 and tomato-EXPEN 2000.

References: 1: SOL Genomics Network (http://www.sgn.cornell.edu); 2: Zamir and Tanksley 1988; Miller and Tanksley 1990; 3: Yu et al. 1991; 4: Fulton et al. 2002a; 5: Wu et al. 2006; 6: Frary et al. 2005.

Chr	cM	Distorter loci	Associated markers	Genotypic segregation					Allele frequencies			
				J/J	J/O	O/O	$\chi 2_{(df = 2)}$	P-value	J	O	$\chi 2_{(df = 1)}$	P-value
1	50.8	*sd1.1*	C2_At2g45620	11	42	10	7.03	0.030	0.508	0.492	0.00	0.987
1	50.8		TG460	12	43	11	6.09	0.048	0.508	0.492	0.00	0.988
2	0.0	sd2.1	TG608	29	34	3	20.55	0.000	0.697	0.303	9.47	0.694
2	2.3		TG31	27	35	4	16.27	0.000	0.674	0.326	7.33	0.727
2	3.3		TG33	22	35	4	11.95	0.003	0.648	0.352	4.74	0.768
2	5.3		T1706	27	35	4	16.27	0.000	0.674	0.326	7.33	0.727
2	10.0		C2_At4g20410	28	33	5	16.03	0.000	0.674	0.326	7.33	0.727
2	21.3	*sd2.2*	C2_At4g30930	34	25	7	25.97	0.000	0.705	0.295	10.24	0.682
2	22.9		C2_At4g21580	31	27	7	19.58	0.000	0.685	0.315	8.14	0.712
2	32.4		T0759	27	33	6	13.36	0.001	0.659	0.341	6.06	0.750
2	44.7		C2_At2g04700	28	30	8	12.67	0.002	0.652	0.348	5.47	0.762
2	48.6		C2_At3g02300	28	29	9	11.91	0.003	0.644	0.356	4.91	0.773
2	53.3		TG167	26	29	11	7.79	0.020	0.614	0.386	2.97	0.820
4	0.0		TG15	6	34	23	9.57	0.008	0.365	0.635	4.06	0.787
4	4.9	*sd4.1*	SSR43	6	36	23	9.65	0.008	0.369	0.631	3.94	0.794
4	7.2		TG581	8	37	21	6.09	0.048	0.402	0.598	2.18	0.844
4	9.5		C2_At3g51010	7	38	21	7.45	0.024	0.394	0.606	2.56	0.832
5	40.5		C2_At4g24830	6	27	30	19.57	0.000	0.310	0.690	8.40	0.703
5	47.8		C2_At1g67700	6	26	34	26.73	0.000	0.288	0.712	11.05	0.671
5	50.1		C2_At1g10500	5	27	34	27.67	0.000	0.280	0.720	11.88	0.660
5	54.0		TG351	4	24	38	NA	0.000	0.242	0.758	16.50	0.606
5	57.1	*sd5.1*	T1584	4	21	40	NA	0.000	0.223	0.777	18.85	0.580
5	60.2		TG185	4	26	36	NA	0.000	0.258	0.742	14.56	0.628
7	12.1		C2_At2g26590	25	29	11	6.78	0.034	0.608	0.392	2.60	0.829
7	12.1	*sd7.1*	C2_At4g33250	26	29	11	7.79	0.020	0.614	0.386	2.97	0.820
7	12.9		C2_At5g20180	25	29	11	6.78	0.034	0.608	0.392	2.60	0.829
7	13.7		T1497	24	29	10	6.62	0.037	0.611	0.389	2.68	0.824
8	1.7	*sd8.1*	C2_At5g46630	14	41	6	9.33	0.009	0.566	0.434	0.80	0.896
8	33.9	*sd8.2*	TG510	11	43	12	6.09	0.048	0.492	0.508	0.00	0.988
9	0.0		C2_At2g37240	8	32	25	8.91	0.012	0.369	0.631	3.94	0.794
9	5.6		TG9	7	30	27	12.75	0.002	0.344	0.656	5.64	0.755
9	23.9	*sd9.1*	TG291	5	30	30	19.62	0.000	0.308	0.692	8.86	0.701
9	37.4		TG551	4	41	21	12.64	0.002	0.371	0.629	3.88	0.797
9	39.7		TG144	4	42	20	12.67	0.002	0.379	0.621	3.41	0.808
9	46.9		T0393	5	39	21	10.48	0.005	0.377	0.623	3.46	0.806
9	58.6	*sd9.2*	C2_At3g24050	5	42	18	10.75	0.005	0.400	0.600	2.22	0.841
9	63.4		SSR599	6	39	20	8.63	0.013	0.392	0.608	2.60	0.829
11	43.2	*sd11.1*	T0142	16	18	21	7.47	0.024	0.455	0.545	0.29	0.928
12	0.8		C2_At4g03280	17	41	8	6.33	0.042	0.568	0.432	0.97	0.892
12	8.8	*sd12.1*	C2_At5g19690	19	38	7	6.75	0.034	0.594	0.406	1.89	0.851
12	24.6	*sd12.2*	TG394	13	45	8	9.48	0.009	0.538	0.462	0.24	0.940
12	28.5		SSR345	15	44	7	9.27	0.010	0.561	0.439	0.74	0.904

Appendix 3. Marker loci showing significant segregation distortion in F_2 *S. ochranthum* × *S. juglandifolium*.

The list includes all markers with significant genotypic segregation distortion (χ^2 values with $P < 0.05$). Values represent the numbers of F_2 individuals in each genotypic class and computed allele frequencies. J = *S. juglandifolium*; O = *S. ochranthum*; χ^2 = goodness-of-fit statistic; df = degrees of freedom.

Locus	*N*	*k*	*P*	R_T	H_T	H_E	H_O	F_{IS}	P-value
Allozymes									
Aco-2	296	3	100	2.89	0.516	0.422	0.411	0.028	0.326
6-Pgdh-2	208	2	25.0	1.39	0.059	0.055	0.043	0.235	0.088
6-Pgdh-3	272	2	16.7	1.08	0.010	0.009	0.009	-0.010	1.000
Idh-1	279	4	33.3	1.44	0.059	0.059	0.067	-0.148	1.000
Adh-2	285	2	33.3	1.28	0.040	0.043	0.040	0.142	0.238
Fdh-1	297	3	100	2.08	0.473	0.372	0.358	0.027	0.430
Got-2	298	2	25.0	1.16	0.021	0.019	0.020	-0.028	1.000
Pgi-1	298	2	8.3	1.08	0.010	0.009	0.009	-0.036	1.000
Pgm-2	298	5	41.7	1.48	0.062	0.063	0.040	0.289	0.004
Mean	281.2	2.8	42.6	1.54	0.139	0.117	0.111	0.055	
SD	(29.1)	(1.1)	(34.0)	(0.59)	(0.203)	(0.161)	(0.157)	(0.140)	
Microsatellites									
SSR125	315	3	64.3	2.55	0.326	0.215	0.230	-0.072	0.859
SSR15	306	12	100	7.88	0.751	0.612	0.593	0.042	0.139
SSR320	317	3	78.6	2.72	0.527	0.268	0.298	-0.118	0.981
SSR325	314	3	100	3.00	0.519	0.354	0.385	-0.080	0.939
SSR341	313	11	92.9	5.40	0.603	0.479	0.478	0.043	0.203
SSR345	311	6	100	5.20	0.752	0.609	0.588	0.026	0.269
SSR43	316	4	85.7	3.07	0.420	0.326	0.354	-0.084	0.950
SSR50	316	12	100	7.31	0.719	0.560	0.545	0.028	0.259
SSR578	317	2	71.4	2.00	0.444	0.140	0.145	0.035	0.415
SSR599	315	4	100	3.53	0.627	0.481	0.483	0.008	0.459
SSR74	315	7	100	5.22	0.691	0.561	0.585	-0.040	0.854
SSR76	316	4	78.6	3.92	0.499	0.335	0.356	-0.056	0.891
SSR80	316	7	100	5.59	0.761	0.612	0.615	-0.002	0.549
SSR85	316	7	100	5.40	0.591	0.484	0.462	0.054	0.127
SSR98	317	5	92.9	3.65	0.298	0.289	0.370	-0.177	1.000
Mean	314.7	6.0	91.0	4.43	0.569	0.422	0.432	-0.026	
SD	(2.9)	(3.3)	(12.2)	(1.743)	(0.153)	(0.156)	(0.143)	(0.069)	

Appendix 4A. Diversity estimates at allozyme and microsatellite loci in populations of *S. lycopersicoides*.

N = number of individuals; *k* = number of alleles per locus; *P* = percentage of polymorphic sites; R_T = total allelic richness; H_T = Nei's total gene diversity; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient per locus; SD = standard deviation. Shaded areas indicate significant values ($P \leq 0.05$).

Locus	*N*	*k*	*P*	R_T	H_T	H_E	H_O	F_{IS}	P-value
Allozymes									
Aco-1	150	3	66.7	2.96	0.584	0.342	0.366	-0.076	0.873
Aco-2	150	3	83.3	2.50	0.489	0.336	0.301	0.109	0.130
6-Pgdh-2	149	4	100	3.32	0.571	0.468	0.393	0.166	0.013
6-Pgdh-3	146	3	100	2.71	0.537	0.420	0.407	0.037	0.367
Idh-1	146	3	33.3	1.49	0.040	0.038	0.028	0.287	0.206
Adh-2	124	4	100	3.75	0.536	0.466	0.313	0.346	0.000
Got-2	142	3	50.0	2.25	0.229	0.105	0.108	-0.012	0.700
Pgi-1	148	2	50.0	1.93	0.154	0.141	0.175	-0.256	1.000
Pgm-1	147	4	100	3.67	0.685	0.488	0.441	0.097	0.084
Pgm-2	150	2	16.7	1.80	0.101	0.074	0.107	-0.455	1.000
Mean	145.2	3.1	70.0	2.64	0.393	0.288	0.264	0.024	0.241
SD	(7.9)	(0.7)	(31.2)	(0.79)	(0.235)	(0.179)	(0.147)	(0.241)	
Microsatellites									
SSR125	153	3	100	2.82	0.538	0.415	0.423	-0.027	0.672
SSR15	154	8	85.7	6.50	0.762	0.533	0.455	0.152	0.005
SSR320	152	2	100	2.00	0.502	0.411	0.394	0.060	0.300
SSR325	153	5	100	4.36	0.626	0.508	0.513	-0.013	0.628
SSR341	154	8	100	6.23	0.491	0.362	0.353	0.030	0.351
SSR345	154	4	100	3.04	0.530	0.444	0.502	-0.149	0.981
SSR43	149	4	57.1	2.89	0.261	0.210	0.212	-0.018	0.631
SSR50	155	6	85.7	4.48	0.614	0.488	0.419	0.159	0.007
SSR599	155	2	85.7	2.00	0.496	0.338	0.334	0.025	0.480
SSR74	155	3	57.1	2.34	0.117	0.109	0.102	0.067	0.455
SSR76	155	4	100	3.69	0.689	0.643	0.692	-0.082	0.936
SSR80	155	4	71.4	2.69	0.190	0.120	0.072	0.418	0.000
SSR85	153	3	28.6	1.81	0.058	0.058	0.048	0.161	0.214
SSR98	155	4	100	2.86	0.397	0.396	0.461	-0.173	0.996
Mean	153.7	4.3	83.7	3.41	0.448	0.360	0.356	0.044	0.150
SD	(1.7)	(1.9)	(22.3)	(1.49)	(0.215)	(0.175)	(0.186)	(0.150)	

Appendix 4B. Diversity estimates at allozyme and microsatellite loci in populations of *S. sitiens*.

N = number of individuals; *k* = number of alleles per locus; *P* = percentage of polymorphic sites; R_T = total allelic richness; H_T = Nei's total gene diversity; H_E = expected heterozygosity; H_O = observed heterozygosity; F_{IS} = inbreeding coefficient per locus; SD = standard deviation. Shaded areas indicate significant values ($P \leq 0.05$).

S. lycopersicoides	**Valid *N***	**Spearman R**	**t(N-2)**	**P-value**
Allozymes				
lat & θ	108	-0.734	-11.14	0.000
long & θ	108	0.636	8.49	0.000
long & *P*	108	-0.493	-5.84	0.000
elev & *P*	108	-0.479	-5.61	0.000
iso & *P*	108	0.465	5.40	0.000
elev & θ	108	-0.441	-5.05	0.000
N_{wild} & *P*	92	-0.334	-3.36	0.001
SSRs				
lat & θ	165	-0.600	-9.58	0.000
long & θ	165	0.564	8.71	0.000
long & *P*	150	0.318	4.08	0.000
iso & θ	165	0.300	4.02	0.000
lat & *P*	150	-0.280	-3.54	0.001
N_{wild} & θ	165	-0.187	-2.43	0.016
iso & *k*	165	-0.158	-2.05	0.042
elev & *P*	150	0.165	2.04	0.043
S. sitiens				
Allozymes				
elev & θ	60	-0.886	-14.53	0.000
lat & θ	60	-0.657	-6.64	0.000
lat & *P*	60	0.617	5.97	0.000
elev & *P*	60	0.617	5.97	0.000
iso & θ	60	0.600	5.71	0.000
iso & *P*	60	-0.494	-4.32	0.000
N_{wild} & θ	60	-0.493	-4.31	0.000
long & θ	60	-0.486	-4.23	0.000
N_{wild} & *P*	60	0.391	3.24	0.002
lat & *k*	60	0.306	2.45	0.017
lat & R_S	60	0.296	2.36	0.022
long & *P*	60	0.278	2.20	0.032
iso & *k*	60	-0.269	-2.13	0.037
SSRs				
iso & θ	98	0.714	10.00	0.000
elev & *P*	98	0.711	9.91	0.000
elev & θ	98	-0.536	-6.22	0.000
lat & θ	98	-0.500	-5.66	0.000
iso & *P*	98	-0.487	-5.46	0.000
N_{wild} & *P*	98	0.444	4.85	0.000
N_{wild} & θ	98	-0.360	-3.79	0.000
long & θ	98	-0.214	-2.15	0.034

Appendix 5. Significant correlations among census population size and diversity estimates, geographic components, and among diversity estimates.in *S. lycopersicoides* and *S. sitiens* at both allozyme and SSR loci.

N = sample number, Spearman R: Spearman's rank correlation coefficient; associated *t*-test and P-values; N_{wild} = census population size; *k* = number of alleles; *P* = percentage of polymorphic sites; R_S = allelic richness; H_E = expected heterozygosity; F_{IS} = the inbreeding coefficient; θ = average pairwise genetic distance; lat = latitude; long = longitude; elev = elevation; iso = isolation.

S. lycopersicoides		**Mean**		**t-value**	**df**	**P-value**	**Valid *N***		**SD**		**F-ratio variances**	**P-value variances**
PL		A	M				A	M	A	M		
	N	269.6	280.1	1.52	22	0.143	9	15	26.95	2.81	91.69	0.000
	k	2.8	5.7	2.75	22	0.012	9	15	1.09	3.08	7.95	0.006
	P	41.7	91.7	5.18	22	0.000	9	15	34.61	11.79	8.63	0.001
	R_T	1.44	4.42	4.86	22	0.000	9	15	0.56	1.78	9.96	0.003
	H_E	0.116	0.421	4.60	22	0.000	9	15	0.16	0.15	1.12	0.810
	F_{IS}	0.037	-0.016	-1.25	22	0.225	9	15	0.14	0.06	5.39	0.006
	F_{ST}	0.088	0.281	3.46	22	0.002	9	15	0.08	0.16	4.11	0.051
	F_{IT}	0.122	0.268	2.07	22	0.050	9	15	0.16	0.17	1.24	0.788
PP												
	N	23.5	23.5	0.00	22	1.000	12	12	2.02	2.02	1.00	1.000
	k	13.8	45.9	23.32	22	0.000	12	12	1.22	4.62	14.46	0.000
	k^{pp}	1.528	3.061	15.79	22	0.000	12	12	0.14	0.31	5.20	0.011
	P	41.7	92.2	14.11	22	0.000	12	12	10.73	6.25	2.95	0.087
	R_S	1.35	2.97	18.39	22	0.000	12	12	0.11	0.28	7.09	0.003
	H_E	0.116	0.421	16.69	22	0.000	12	12	0.04	0.05	1.74	0.370
	F_{IS}	0.073	-0.006	-1.93	22	0.067	12	12	0.12	0.08	2.17	0.214
S. sitiens												
PL		A	M				A	M	A	M		
	N	106.9	109.1	1.60	22	0.124	10	14	5.15	1.03	25.16	0.000
	k	3.1	4.1	1.66	22	0.111	10	14	0.74	1.73	5.50	0.015
	P	68.3	83.3	1.38	22	0.183	10	14	29.87	23.57	1.61	0.424
	R_T	2.22	2.66	1.19	22	0.247	10	14	0.71	1.00	1.99	0.303
	H_E	0.285	0.354	0.94	22	0.360	10	14	0.17	0.18	1.15	0.856
	F_{IS}	0.030	0.013	-0.23	22	0.821	10	14	0.25	0.10	5.87	0.005
	F_{ST}	0.252	0.177	-1.72	22	0.099	10	14	0.13	0.09	2.34	0.160
	F_{IT}	0.283	0.186	-1.42	22	0.168	10	14	0.20	0.13	2.32	0.163
PP												
	N	18.3	18.3	0.00	10	1.000	6	6	5.79	5.79	1.00	1.000
	k	20.3	34.8	5.33	10	0.000	6	6	3.44	5.71	2.74	0.292
	k^{pp}	2.033	2.488	2.09	10	0.063	6	6	0.34	0.41	1.40	0.721
	P	68.3	83.3	1.86	10	0.093	6	6	17.22	9.76	3.12	0.238
	R_S	1.88	2.23	2.47	10	0.033	6	6	0.24	0.25	1.10	0.919
	H_E	0.285	0.354	1.85	10	0.093	6	6	0.07	0.06	1.55	0.645
	F_{IS}	0.091	-0.019	-2.76	10	0.020	6	6	0.09	0.03	11.62	0.018

Appendix 6. *T*-test results of comparisons of diversity estimates per locus and per population based on allozyme and microsatellite data for *S. lycopersicoides* and *S. sitiens*.

Only individuals genotyped with both marker types were considered in the calculations. Significant differences ($P \leq 0.05$) are indicated by shaded boxes. PL = per locus; PP = per population; *N* = number of individuals; df = degrees of freedom; SD = standard deviation; A = allozymes; M = microsatellites; *k* = number of alleles; k^{pp} = number of alleles per locus and population; *P* = percentage of polymorphic sites; R_T = total allelic richness; R_S = allelic richness per population; H_E = expected heterozygosity; F_{IS} = the inbreeding coefficient; F_{ST} = fixation index; F_{IT} = the overall inbreeding coefficient.

S. lycopersicoides	**Rank sum**		**U**	**Z**	**P-value**	**Z adjusted**	**P-value**	**valid *N***		**2* exact P-value**
PL	A	M						A	M	
N	110.5	189.5	65.5	0.12	0.905	0.12	0.903	9	15	0.907
k	65.5	234.5	20.5	2.80	0.005	2.85	0.004	9	15	0.003
P	67.0	233.0	22.0	2.71	0.007	2.82	0.005	9	15	0.005
R_T	49.0	251.0	4.0	3.79	0.000	3.79	0.000	9	15	0.000
H_E	59.0	241.0	14.0	3.19	0.001	3.19	0.001	9	15	0.001
F_{IS}	122.0	178.0	58.0	-0.57	0.571	-0.57	0.571	9	15	0.599
F_{ST}	57.0	243.0	12.0	3.31	0.001	3.31	0.001	9	15	0.000
F_{IT}	86.0	214.0	41.0	1.58	0.114	1.58	0.114	9	15	0.123
PP										
N	150.0	150.0	72.0	0.00	1.000	0.00	1.000	12	12	1.000
k	78.0	222.0	0.0	4.16	0.000	4.18	0.000	12	12	0.000
k^{pp}	78.0	222.0	0.0	4.16	0.000	4.18	0.000	12	12	0.000
P	78.0	222.0	0.0	4.16	0.000	4.16	0.000	12	12	0.000
R_S	78.0	222.0	0.0	4.16	0.000	4.20	0.000	12	12	0.000
H_E	78.0	222.0	0.0	4.16	0.000	4.16	0.000	12	12	0.000
F_{IS}	180.0	120.0	42.0	-1.73	0.083	-1.73	0.083	12	12	0.089
S. sitiens										
PL										
N	105.5	194.5	50.5	1.14	0.254	1.21	0.227	10	14	0.259
k	101.0	199.0	46.0	1.41	0.160	1.47	0.143	10	14	0.172
P	103.0	197.0	48.0	1.29	0.198	1.36	0.173	10	14	0.212
R_T	107.5	192.5	52.5	1.02	0.306	1.03	0.305	10	14	0.312
H_E	111.0	189.0	56.0	0.82	0.412	0.82	0.412	10	14	0.437
F_{IS}	137.0	163.0	58.0	-0.70	0.482	-0.70	0.482	10	14	0.508
F_{ST}	140.0	160.0	55.0	-0.88	0.380	-0.88	0.380	10	14	0.403
F_{IT}	150.0	150.0	45.0	-1.46	0.143	-1.46	0.143	10	14	0.154
PP										
N	39.0	39.0	18.0	0.00	1.000	0.00	1.000	6	6	1.000
k	21.0	57.0	0.0	2.88	0.004	2.89	0.004	6	6	0.002
k^{pp}	28.0	50.0	7.0	1.76	0.078	1.76	0.078	6	6	0.093
P	27.0	51.0	6.0	1.92	0.055	1.92	0.054	6	6	0.065
R_S	26.5	51.5	5.5	2.00	0.045	2.02	0.043	6	6	0.041
H_E	29.0	49.0	8.0	1.60	0.109	1.60	0.109	6	6	0.132
F_{IS}	51.0	27.0	6.0	-1.92	0.055	-1.92	0.055	6	6	0.065

Appendix 7. Nonparametric Mann-Whitney U test results of the comparison of allozyme and microsatellite diversity estimates in *S. lycopersicoides* and *S. sitiens*.

Only individuals genotyped with both marker types were considered in the calculations. Significant differences ($P \leq 0.05$) are indicated by shaded boxes. PL = per locus; PP = per population; U, Z = distribution variate values; 'adjusted' = adjusted for normal approximation (for samples larger than 20); 2*exact P-value = 2*1-sided exact P-value where p is 1 minus the cumulative (one-sided) probability of the respective U statistic; A = allozymes; M = microsatellites. *N* = number of individuals; *k* = number of alleles; k^{pp} = number of alleles per locus and population; *P* = percentage of polymorphic sites; R_T = total allelic richness; R_S = allelic richness per population; H_E = expected heterozygosity; F_{IS} = the inbreeding coefficient; F_{ST} = fixation index; F_{IT} = the overall inbreeding coefficient.

	Among populations				Among individuals				Within individuals				*F*-Statistics					
Locus	**SSD**	**df**	**Va**	**%**	**SSD**	**df**	**Vb**	**%**	**SSD**	**df**	**Vc**	**%**	F_{IS}	**P-value**	F_{ST}	**P-value**	F_{IT}	**P-value**
Aco-2	29.5	11	0.050	19.1	61.9	284	0.006	2.3	61.0	296	0.206	78.6	0.028	0.326	0.191	0.000	0.214	0.000
6-Pgdh-2	2.2	11	0.003	11.2	9.3	286	0.006	20.8	6.0	298	0.020	68.0	0.235	0.088	0.112	0.000	0.320	0.001
6-Pgdh-3	0.1	11	0.000	0.8	1.4	286	0.000	-1.0	1.5	298	0.005	100.3	-0.010	1.000	0.008	0.454	-0.003	1.000
Idh-1	2.2	11	0.004	12.0	6.4	285	-0.004	-13.1	9.0	297	0.030	101.1	-0.148	1.000	0.120	0.000	-0.011	1.000
Adh-2	0.6	11	0.001	3.2	5.7	260	0.003	13.7	4.5	272	0.017	83.1	0.142	0.238	0.032	0.055	0.169	0.101
Fdh-1	22.5	11	0.054	22.2	37.8	196	0.005	2.1	38.0	208	0.183	75.7	0.027	0.430	0.222	0.000	0.243	0.001
Got-2	0.2	11	0.000	2.0	2.7	267	0.000	-2.8	3.0	279	0.011	100.7	-0.028	1.000	0.020	0.129	-0.007	1.000
Pgi-1	0.1	11	0.000	3.4	1.3	286	0.000	-3.5	1.5	298	0.005	100.1	-0.036	1.000	0.034	0.007	-0.001	1.000
Pgm-2	1.2	11	0.002	4.9	10.4	273	0.009	27.5	6.0	285	0.021	67.6	0.289	0.004	0.049	0.001	0.324	0.000
SSR125	33.2	13	0.055	32.7	31.5	301	-0.008	-4.8	38.0	315	0.121	72.2	-0.072	0.859	0.327	0.000	0.278	0.000
SSR15	49.7	13	0.080	21.1	91.4	292	0.013	3.3	88.0	306	0.288	75.5	0.042	0.139	0.211	0.000	0.245	0.000
SSR320	80.8	13	0.135	49.4	37.0	303	-0.016	-6.0	49.0	317	0.155	56.6	-0.118	0.981	0.494	0.000	0.434	0.000
SSR325	53.8	13	0.089	33.4	48.8	300	-0.014	-5.4	60.0	314	0.191	71.9	-0.080	0.939	0.334	0.000	0.281	0.000
SSR341	45.6	13	0.073	23.9	72.9	299	0.010	3.3	70.0	313	0.224	72.9	0.043	0.203	0.239	0.000	0.271	0.000
SSR345	44.0	13	0.069	18.2	95.1	297	0.008	2.1	94.5	311	0.304	79.7	0.026	0.269	0.182	0.000	0.203	0.000
SSR43	28.5	13	0.045	21.2	46.5	302	-0.014	-6.6	57.5	316	0.182	85.3	-0.084	0.950	0.212	0.000	0.147	0.002
SSR50	52.0	13	0.082	22.5	87.9	302	0.008	2.2	87.0	316	0.275	75.3	0.028	0.259	0.225	0.000	0.247	0.000
SSR578	95.9	13	0.162	69.2	22.5	303	0.002	1.1	22.0	317	0.069	29.7	0.035	0.415	0.692	0.000	0.703	0.000
SSR599	52.5	13	0.085	26.5	71.3	301	0.002	0.6	73.5	315	0.233	73.0	0.008	0.459	0.265	0.000	0.270	0.000
SSR74	43.1	13	0.068	19.4	81.6	301	-0.011	-3.2	92.5	315	0.294	83.9	-0.040	0.854	0.194	0.000	0.161	0.000
SSR76	47.9	13	0.078	30.6	50.4	302	-0.010	-3.9	59.0	316	0.187	73.3	-0.056	0.891	0.306	0.000	0.267	0.000
SSR80	52.8	13	0.083	21.6	91.3	302	-0.001	-0.2	96.0	316	0.304	78.6	-0.002	0.549	0.216	0.000	0.214	0.000
SSR85	34.8	13	0.054	17.9	78.2	302	0.013	4.4	73.5	316	0.233	77.7	0.054	0.127	0.179	0.000	0.223	0.000
SSR98	7.5	13	0.010	6.9	34.8	303	-0.025	-16.5	52.0	317	0.164	109.6	-0.177	1.000	0.069	0.000	-0.096	1.000

Appendix 8A. Locus-by-locus AMOVA and *F*-statistics at allozymes and SSRs in *S. lycopersicoides*.

SSD = sum of square deviations; df = degrees of freedom; Va = variance among populations; Vb = variance among individuals; Vc = variance within individuals; % = percentage of variation; F_{IS} = the inbreeding coefficient; F_{ST} = fixation index; F_{IT} = the overall inbreeding coefficient. Shaded boxes indicate significant values ($P \leq 0.004$ for allozymes, $P \leq 0.004$ for SSRs).

	Among populations				Among individuals				Within individuals				*F*-Statistics					
Locus	**SSD**	**df**	**Va**	**%**	**SSD**	**df**	**Vb**	**%**	**SSD**	**df**	**Vc**	**%**	F_{IS}	**P-value**	F_{ST}	**P-value**	F_{IT}	**P-value**
Aco-1	35.3	5	0.138	43.9	23.5	144	-0.013	-4.2	28.5	150	0.190	60.4	-0.076	0.873	0.439	0.000	0.396	0.000
Aco-2	22.7	5	0.087	33.6	27.5	144	0.019	7.2	23.0	150	0.153	59.2	0.109	0.130	0.336	0.000	0.408	0.000
6-Pgdh-2	15.6	5	0.057	19.5	39.6	143	0.040	13.4	29.5	149	0.198	67.1	0.166	0.013	0.195	0.000	0.329	0.000
6-Pgdh-3	18.3	5	0.071	25.2	30.5	140	0.008	2.8	29.5	146	0.202	72.0	0.037	0.367	0.252	0.000	0.280	0.001
Idh-1	0.4	5	0.001	6.3	3.5	140	0.006	26.9	2.0	146	0.014	66.9	0.287	0.206	0.063	0.004	0.331	0.034
Adh-2	9.9	5	0.041	14.8	37.2	118	0.081	29.5	19.0	124	0.153	55.7	0.346	0.000	0.148	0.000	0.443	0.000
Got-2	18.0	5	0.075	59.0	7.0	136	-0.001	-0.5	7.5	142	0.053	41.5	-0.012	0.700	0.590	0.000	0.585	0.000
Pgi-1	3.3	5	0.012	15.5	7.1	142	-0.017	-21.7	12.5	148	0.084	106.1	-0.256	1.000	0.155	0.000	-0.061	1.000
Pgm-1	30.0	5	0.117	32.3	37.9	141	0.024	6.6	32.5	147	0.221	61.1	0.097	0.084	0.323	0.000	0.389	0.000
Pgm-2	4.3	5	0.017	31.3	2.9	144	-0.017	-31.2	8.0	150	0.053	100.0	-0.455	1.000	0.313	0.000	0.000	1.000
SSR125	21.2	6	0.076	27.2	28.9	146	-0.006	-2.0	32.0	153	0.209	74.7	-0.027	0.672	0.272	0.000	0.253	0.001
SSR15	37.7	6	0.136	34.0	44.7	147	0.040	10.0	34.5	154	0.224	56.0	0.152	0.005	0.340	0.000	0.440	0.000
SSR320	13.7	6	0.048	18.5	32.3	145	0.013	4.9	30.0	152	0.197	76.6	0.060	0.300	0.185	0.000	0.234	0.006
SSR325	18.2	6	0.064	19.8	37.2	146	-0.003	-1.0	40.0	153	0.261	81.2	-0.013	0.628	0.198	0.000	0.188	0.002
SSR341	17.9	6	0.063	24.9	28.9	147	0.006	2.3	28.5	154	0.185	72.8	0.030	0.351	0.249	0.000	0.272	0.000
SSR345	14.0	6	0.049	18.0	27.9	147	-0.033	-12.3	39.5	154	0.256	94.3	-0.149	0.981	0.180	0.000	0.057	0.342
SSR43	7.0	6	0.025	18.8	15.2	142	-0.002	-1.4	16.5	149	0.111	82.7	-0.018	0.631	0.188	0.000	0.173	0.029
SSR50	19.6	6	0.067	21.2	42.8	148	0.040	12.5	32.5	155	0.210	66.2	0.159	0.007	0.212	0.000	0.338	0.000
SSR599	26.5	6	0.096	36.7	25.1	148	0.004	1.6	25.0	155	0.161	61.7	0.025	0.480	0.367	0.000	0.383	0.000
SSR74	2.4	6	0.008	13.2	8.2	148	0.003	5.8	7.5	155	0.048	81.0	0.067	0.455	0.132	0.000	0.190	0.030
SSR76	9.6	6	0.029	8.4	43.4	148	-0.026	-7.5	53.5	155	0.345	99.0	-0.082	0.936	0.084	0.000	0.010	0.542
SSR80	9.4	6	0.033	33.4	13.9	148	0.028	27.8	6.0	155	0.039	38.8	0.418	0.000	0.334	0.000	0.612	0.000
SSR85	0.7	6	0.002	6.3	4.6	146	0.004	15.1	3.5	153	0.023	78.6	0.161	0.214	0.063	0.001	0.214	0.063
SSR98	1.9	6	0.004	1.9	23.9	148	-0.034	-17.0	35.5	155	0.229	115.1	-0.173	0.996	0.019	0.159	-0.151	0.993

Appendix 8B. Locus-by-locus AMOVA and *F*-statistics at allozymes and SSRs in *S. sitiens*.

SSD = sum of square deviations; df = degrees of freedom; Va = variance among populations; Vb = variance among individuals; Vc = variance within individuals; % = percentage of variation; F_{IS} = the inbreeding coefficient; F_{ST} = fixation index; F_{IT} = the overall inbreeding coefficient. Shaded boxes indicate significant values ($P \leq 0.005$ at allozymes, $P \leq 0.07$ at SSRs).

A	LA4018	LA2387	LA1964	LA1966	LA2781	LA2777	LA2776	LA2772	LA4320	LA4130	LA4131	LA4126	LA4123	LA2730
LA4018	-	36	571	246	848	574	362	528	1379	216	547	268	289	1169
LA2387	7.2	-	607	282	884	610	1396	564	1343	180	511	304	253	1133
LA1964	59.6	62.1	-	325	277	3	19	43	1950	787	1118	303	860	1740
LA1966	57.7	59.8	4.0	-	602	328	263	282	1625	462	793	22	535	1415
LA2781	122.1	124.6	62.6	64.7	-	274	486	320	2227	1064	1395	580	1137	2017
LA2777	120.9	123.2	61.3	63.4	2.5	-	212	46	1953	790	1121	306	863	1743
LA2776	119.5	121.6	60.0	61.9	6.0	3.5	-	166	1741	578	909	94	651	301
LA2772	126.1	128.0	66.6	68.3	9.0	7.7	7.0	-	1907	744	1075	260	817	1697
LA4320	119.2	119.8	62.9	63.2	30.9	28.5	25.1	24.3	-	1163	832	1647	1090	210
LA4130	187.9	188.3	131.6	132.1	79.1	78.6	77.7	70.9	69.0	-	331	484	73	953
LA4131	191.9	192.4	135.0	135.6	80.8	80.4	79.8	72.9	72.7	6.0	-	815	258	622
LA4126	235.2	235.7	178.2	178.9	121.9	121.9	121.7	114.7	116.1	47.5	43.3	-	557	1437
LA4123	236.2	236.5	179.4	180.0	123.7	123.6	123.3	116.4	117.1	48.3	44.4	3.3	-	880
LA2730	240.2	239.9	185.5	185.7	134.1	133.6	132.8	126.0	122.7	55.1	53.4	25.0	21.8	-

B	LA4116	LA4114	LA4113	LA4112	LA4331	LA4111	LA4105
LA4116	-	199.0	283.0	155.0	139.0	217.0	317.0
LA4114	4.3	-	84.0	44.0	338.0	18.0	118.0
LA4113	8.6	4.9	-	128.0	422.0	66.0	34.0
LA4112	53.7	51.4	46.6	-	294.0	62.0	162.0
LA4331	90.2	86.9	81.9	41.3	-	356.0	456.0
LA4111	107.5	104.8	99.9	54.2	26.0	-	100.0
LA4105	232.9	230.0	225.1	179.7	144.5	125.5	-

Appendix 9. Half-matrices of ***S. lycopersicoides*** (A) and *S. sitiens* (B): geographic distances are depicted below diagonal and elevational distanes above diagonal.

A	LA4018	LA2387	LA1964	LA1966	LA2781	LA2777	LA2776	LA2772	LA4131	LA4126	LA4123	LA2730
LA4018	-	NS	NS	*	*	*	NS	*	*	*	*	*
LA2387	0.039	-	NS	*	NS	NS	NS	NS	*	*	*	*
LA1964	0.119	0.023	-	NS	*	NS	NS	*	*	*	*	NS
LA1966	0.164	0.079	0.036	-	*	*	NS	*	*	*	*	*
LA2781	0.174	0.116	0.243	0.212	-	*	NS	NS	*	*	*	*
LA2777	0.179	0.042	0.056	0.055	0.175	-	NS	*	*	*	NS	*
LA2776	0.077	0.007	0.009	0.025	0.188	0.026	-	*	NS	NS	NS	NS
LA2772	0.143	0.064	0.165	0.150	-0.001	0.103	0.116	-	*	*	*	*
LA4131	0.509	0.372	0.287	0.156	0.532	0.222	0.262	0.441	-	*	*	*
LA4126	0.308	0.153	0.067	0.132	0.417	0.110	0.101	0.325	0.296	-	NS	NS
LA4123	0.133	0.045	0.031	0.077	0.258	0.050	0.030	0.204	0.268	0.061	-	*
LA2730	0.282	0.173	0.082	0.192	0.481	0.199	0.131	0.390	0.452	0.064	0.073	-

M		LA2387	LA1964	LA1966	LA2781	LA2777	LA2776	LA2772	LA4320	LA4130	LA4131	LA4126	LA4123	LA2730
LA4018	-	*	*	*	*	*	*	*	*	*	*	*	*	*
LA2387	0.122	-	*	*	*	*	*	*	*	*	*	*	*	*
LA1964	0.221	0.188	-	*	*	*	*	*	*	*	*	*	*	*
LA1966	0.141	0.148	0.115	-	*	*	*	*	*	*	*	*	*	*
LA2781	0.438	0.395	0.298	0.303	-	*	*	*	*	*	*	*	*	*
LA2777	0.353	0.306	0.302	0.300	0.209	-	*	*	*	*	*	*	*	*
LA2776	0.356	0.280	0.239	0.260	0.159	0.092	-	*	*	*	*	*	*	*
LA2772	0.395	0.306	0.291	0.300	0.188	0.215	0.139	-	*	*	*	*	*	*
LA4320	0.350	0.262	0.251	0.250	0.201	0.134	0.089	0.164	-	*	*	*	*	*
LA4130	0.456	0.380	0.339	0.343	0.305	0.267	0.175	0.231	0.249	-	*	*	*	*
LA4131	0.464	0.400	0.379	0.370	0.345	0.318	0.210	0.220	0.278	0.104	-	*	*	*
LA4126	0.429	0.364	0.310	0.347	0.333	0.257	0.197	0.236	0.251	0.200	0.192	-	*	*
LA4123	0.370	0.323	0.291	0.318	0.286	0.184	0.164	0.197	0.201	0.198	0.190	0.060	-	*
LA2730	0.350	0.304	0.289	0.318	0.359	0.281	0.228	0.236	0.247	0.238	0.220	0.109	0.104	-

Appendix 10A. *S. lycopersicoides* half-matrices of allozyme (A) and SSR (M): pairwise genetic distances (Weir and Cockerham's θ) are shown below diagonal and pairwise significance above diagonal.

* 5 % nominal level after Bonferroni corrections; NS = non-significant. Populations are listed according to their geographic location (from north to south).

A	LA4116	LA4114	LA4113	LA4112	LA4111	LA4105
LA4116	-	*	*	*	*	*
LA4114	0.070	-	*	*	*	*
LA4113	0.144	0.080	-	*	*	*
LA4112	0.174	0.177	0.165	-	*	*
LA4111	0.243	0.286	0.318	0.166	-	*
LA4105	0.512	0.535	0.527	0.444	0.362	-

M	LA4116	LA4114	LA4113	LA4112	LA4331	LA4111	LA4105
LA4116	-	*	*	*	*	*	*
LA4114	0.048	-	*	*	*	*	*
LA4113	0.073	0.098	-	*	*	*	*
LA4112	0.158	0.203	0.146	-	*	*	*
LA4331	0.221	0.243	0.233	0.116	-	*	*
LA4111	0.247	0.281	0.228	0.080	0.072	-	*
LA4105	0.336	0.370	0.339	0.245	0.257	0.278	-

Appendix 10B. *S. sitiens* half-matrices of allozyme (A) and SSR (M): pairwise genetic distances (Weir and Cockerham's θ) are shown below diagonal and pairwise significance above diagonal.

* 5 % nominal level after Bonferroni corrections; NS = non-significant. Populations are listed according to their geographic location (from north to south).

	Valid *N*	Spearman R	t(N-2)	P-value
Allozymes				
elev & Aco-2.18	12	0.832	4.73	0.001
long & Pgm-2.10	12	0.640	2.63	0.025
long & Pgm-2.22	12	0.640	2.63	0.025
lat & Fdh-1.10	12	0.615	2.47	0.033
lat & Fdh-1.85	12	-0.608	-2.42	0.036
SSRs				
lat & SSR345.12	14	0.890	6.77	0.000
lat & SSR345.15	14	0.829	5.13	0.000
lat & SSR85.11	14	-0.806	-4.72	0.000
lat & SSR50.18	14	0.777	4.28	0.001
lat & SSR50.18	14	0.777	4.28	0.001
long & SSR578.11	14	-0.733	-3.74	0.003
long & SSR578.11	14	-0.733	-3.74	0.003
long & SSR578.12	14	0.733	3.74	0.003
long & SSR578.12	14	0.733	3.74	0.003
lat & SSR50.14	14	-0.731	-3.71	0.003
lat & SSR50.14	14	-0.731	-3.71	0.003
long & SSR50.18	14	-0.724	-3.63	0.003
long & SSR50.18	14	-0.724	-3.63	0.003
long & SSR341.11	14	-0.713	-3.52	0.004
lat & SSR43.14	14	0.690	3.30	0.006
lat & SSR43.14	14	0.690	3.30	0.006
lat & SSR341.11	14	0.682	3.23	0.007
long & SSR345.15	14	-0.678	-3.19	0.008
long & SSR320.11	14	-0.674	-3.16	0.008
long & SSR15.13	14	-0.669	-3.12	0.009
long & SSR320.12	14	0.664	3.08	0.010
lat & SSR599.11	14	-0.649	-2.95	0.012
lat & SSR599.11	14	-0.649	-2.95	0.012
lat & SSR578.11	14	0.641	2.89	0.014
lat & SSR578.11	14	0.641	2.89	0.014
lat & SSR578.12	14	-0.641	-2.89	0.014
lat & SSR578.12	14	-0.641	-2.89	0.014
lat & SSR341.22	14	-0.630	-2.81	0.016
lat & SSR345.11	14	-0.627	-2.79	0.016
lat & SSR341.80	14	0.613	2.69	0.020
long & SSR50.21	14	-0.609	-2.66	0.021
long & SSR50.21	14	-0.609	-2.66	0.021
lat & SSR50.21	14	0.602	2.61	0.023
lat & SSR50.21	14	0.602	2.61	0.023
lat & SSR43.13	14	-0.601	-2.61	0.023
lat & SSR43.13	14	-0.601	-2.61	0.023
long & SSR345.12	14	-0.596	-2.57	0.025
long & SSR50.14	14	0.594	2.56	0.025
long & SSR50.14	14	0.594	2.56	0.025
long & SSR50.15	14	-0.590	-2.53	0.026
long & SSR50.15	14	-0.590	-2.53	0.026
long & SSR125.12	14	-0.584	-2.49	0.028
lat & SSR43.11	14	-0.563	-2.36	0.036
lat & SSR43.11	14	-0.563	-2.36	0.036
long & SSR85.11	14	0.562	2.35	0.037
lat & SSR325.14	14	0.545	2.25	0.044
long & SSR325.14	14	-0.545	-2.25	0.044

	Valid *N*	Spearman R	t(N-2)	P-value
SSRs				
long & SSR599.11	14	0.544	2.25	0.044
long & SSR599.11	14	0.544	2.25	0.044

Appendix 11A. Significant correlations between allele frequencies and geographic components at both allozyme and SSR loci in *S. lycopersicoides*.

N = sample number, Spearman R = Spearman's rank correlation coefficient; associated *t*-test and P-values; lat = latitude; long = longitude; elev = elevation.

	Valid *N*	Spearman R	t(N-2)	P-value
Allozymes				
long & Aco-1.15	6	0.986	11.66	0.000
long & Aco-1.22	6	-0.986	-11.66	0.000
elev & 6-Pgdh-3.10	6	-0.943	-5.66	0.005
lat & 6-Pgdh-2.14	6	-0.943	-5.66	0.005
lat & Pgm-1.10	6	0.943	5.66	0.005
lat & 6-Pgdh-2.10	6	0.941	5.57	0.005
lat & 6-Pgdh-2.16	6	0.899	4.10	0.015
elev & 6-Pgdh-3.98	6	0.845	3.16	0.034
elev & Got-2.25	6	-0.845	-3.16	0.034
elev & 6-Pgdh-3.96	6	0.829	2.96	0.042
lat & Pgm-1.12	6	-0.829	-2.96	0.042
elev & Pgi-1.10	6	-0.820	-2.86	0.046
elev & Pgi-1.18	6	0.820	2.86	0.046
long & Aco-2.10	6	0.820	2.86	0.046
lat & Aco-1.15	6	0.812	2.78	0.050
lat & Aco-1.22	6	-0.812	-2.78	0.050
SSRs				
lat & p:15.18	7	0.964	8.06	0.000
long & p:80.14	7	-0.964	-8.06	0.000
lat & p:50.15	7	-0.927	-5.51	0.003
lat & p:80.14	7	-0.927	-5.51	0.003
long & p:50.11	7	0.893	4.43	0.007
lat & p:345.15	7	-0.867	-3.89	0.012
lat & p:50.11	7	0.857	3.72	0.014
long & p:599.11	7	-0.857	-3.72	0.014
long & p:599.13	7	0.857	3.72	0.014
long & p:50.15	7	-0.852	-3.65	0.015
lat & p:599.11	7	-0.821	-3.22	0.023
lat & p:599.13	7	0.821	3.22	0.023
long & p:325.14	7	0.821	3.22	0.023
lat & p:74.16	7	0.802	3.00	0.030
long & p:15.18	7	0.778	2.77	0.039

Appendix 11B. Significant correlations between allele frequencies and geographic components at both allozyme and SSR loci in *S. sitiens*.

N = sample number, Spearman R = Spearman's rank correlation coefficient; associated *t*-test and P-values; lat = latitude; long = longitude; elev = elevation.

	Source of variation	df	SSD	VC	% variation	Inbreeding coefficients		P-value
	Among 3 clusters							
	Allozymes							
T	Among groups	2	15.3	0.021	4.2	F_{CT}	0.042	0.060
	Among populations within groups	9	31.0	0.061	12.1	F_{SC}	0.127	0.000
	Among individuals within populations	286	116.5	-0.015	-3.0	F_{IS}	-0.036	0.812
	Within individuals	298	130.5	0.438	86.7	F_{IT}	0.133	0.000
	Total	595	293.2	0.505				
N	Among populations	3	8.3	0.052	9.2	F_{ST}	0.092	0.000
	Among individuals within populations	85	38.8	-0.058	-10.3	F_{IS}	-0.114	0.956
	Within populations	89	51.0	0.573	101.1	F_{IT}	-0.011	0.695
	Total	177	98.1	0.567				
C	Among populations	3	6.7	0.034	7.2	F_{ST}	0.072	0.000
	Among individuals within populations	103	45.3	0.003	0.6	F_{IS}	0.006	0.432
	Within populations	107	46.5	0.435	92.2	F_{IT}	0.078	0.143
	Total	213	98.6	0.471				
S	Among populations	3	15.9	0.098	23.0	F_{ST}	0.230	0.000
	Among individuals within populations	98	32.4	0.003	0.8	F_{IS}	0.010	0.370
	Within populations	102	33.0	0.324	76.2	F_{IT}	0.238	0.000
	Total	203	81.3	0.425				
	Microsatellites							
T	Among groups	2	393.0	0.794	17.6	FCT	0.176	0.000
	Among populations within groups	11	320.9	0.581	12.9	FSC	0.157	0.000
	Among individuals within populations	303	926.7	-0.068	-1.5	F_{IS}	-0.022	0.920
	Within individuals	317	1012.5	3.194	71.0	F_{IT}	0.290	0.000
	Total	633	2653.2	4.500				
N	Among populations	3	81.3	0.544	15.5	F_{ST}	0.155	0.000
	Among individuals within populations	85	256.9	0.056	1.6	F_{IS}	0.019	0.022
	Within populations	89	259.0	2.910	82.9	F_{IT}	0.171	0.000
	Total	177	597.2	3.510				
C	Among populations	4	127.0	0.603	15.9	F_{ST}	0.159	0.000
	Among individuals within populations	114	350.5	-0.106	-2.8	F_{IS}	-0.033	0.916
	Within populations	119	391.0	3.286	86.9	F_{IT}	0.131	0.000
	Total	237	868.4	3.783				
S	Among populations	4	112.7	0.585	15.5	F_{ST}	0.155	0.000
	Among individuals within populations	104	319.4	-0.127	-3.4	F_{IS}	-0.040	0.924
	Within populations	109	362.5	3.326	87.9	F_{IT}	0.121	0.000
	Total	217	794.6	3.783				

	Source of variation	df	SSD	VC	% variation	Inbreeding coefficients		P-value
	Among 5 clusters							
	Allozymes							
T	Among groups	4	33.3	0.058	11.3	F_{CT}	0.113	0.002
	Among populations within groups	7	13.0	0.029	5.7	F_{SC}	0.064	0.000
	Among individuals within populations	286	116.5	-0.015	-3.0	F_{IS}	-0.036	0.819
	Within individuals	298	130.5	0.438	86.0	F_{IT}	0.140	0.000
	Total	595	293.2	0.509				
	Microsatellites							
T	Among groups	4	493.2	0.828	18.7	F_{CT}	0.187	0.000
	Among populations within groups	9	220.7	0.475	10.7	F_{SC}	0.132	0.000
	Among individuals within populations	303	926.7	-0.068	-1.5	F_{IS}	-0.022	0.921
	Within individuals	317	1012.5	3.194	72.1	F_{IT}	0.279	0.000
	Total	633	2653.2	4.429				

Appendix 12A. Analysis of molecular variance (AMOVA) within and among 3 and 5 population clusters in *S. lycopersicoides*.

T = total population; N,C,S = northern, central and southern cluster, respectively; df = degrees of freedom; SSD = sum of square deviations; VC = variance components; inbreeding coefficients = F_{CT} (among groups), F_{SC} (among populations within groups), F_{ST} (among populations), F_{IS} (within populatios), F_{IT} (within individuals).

	Source of variation	df	SSD	VC	% variation	Inbreeding coefficients		P-value
	Among 3 clusters							
	Allozymes							
T	Among groups	2	116.7	0.524	25.2	F_{CT}	0.252	0.000
	Among populations within groups	3	36.1	0.209	10.0	F_{SC}	0.134	0.000
	Among individuals within populations	144	203.3	0.066	3.2	F_{IS}	0.049	0.064
	Within individuals	150	192.0	1.280	61.6	F_{IT}	0.384	0.000
	Total	299	548.1	2.078				
N	Among populations	2	22.1	0.178	10.4	F_{ST}	0.104	0.000
	Among individuals within populations	77	122.3	0.059	3.5	F_{IS}	0.039	0.174
	Within populations	80	117.5	1.469	86.1	F_{IT}	0.139	0.003
	Total	159	261.9	1.706				
C	Among populations	1	13.9	0.272	16.7	F_{ST}	0.167	0.000
	Among individuals within populations	44	63.8	0.094	5.8	F_{IS}	0.070	0.095
	Within populations	46	58.0	1.261	77.5	F_{IT}	0.225	0.001
	Total	91	135.7	1.627				
S	Among populations	NA	NA	NA	NA	F_{ST}	NA	
	Among individuals within populations	23	17.3	0.033	4.5	F_{IS}	0.045	0.267
	Within populations	24	16.5	0.688	95.5	F_{IT}	NA	
	Total	47	33.8	0.720				
	Microsatellites							
T	Among groups	2	148.4	0.639	19.0	F_{CT}	0.190	0.000
	Among populations within groups	4	48.0	0.219	6.5	F_{SC}	0.081	0.000
	Among individuals within populations	148	373.3	0.021	0.6	F_{IS}	0.008	0.355
	Within individuals	155	384.5	2.481	73.8	F_{IT}	0.262	0.000
	Total	309	954.2	3.359				
N	Among populations	2	18.6	0.171	7.1	F_{ST}	0.071	0.000
	Among individuals within populations	59	130.8	-0.021	-0.9	F_{IS}	-0.009	0.567
	Within populations	62	140.0	2.258	93.7	F_{IT}	0.063	0.125
	Total	123	289.4	2.409				
C	Among populations	2	29.5	0.267	9.0	F_{ST}	0.090	0.000
	Among individuals within populations	65	172.9	-0.053	-1.8	F_{IS}	-0.019	0.696
	Within populations	68	188.0	2.765	92.8	F_{IT}	0.072	0.081
	Total	135	390.3	2.979				
S	Among populations	NA	NA	NA	NA	F_{ST}	NA	
	Among individuals within populations	24	69.6	0.320	12.4	F_{IS}	0.124	0.019
	Within populations	25	56.5	2.260	87.6	F_{IT}	NA	
	Total	49	126.1	2.580				

Appendix 12B. Analysis of molecular variance (AMOVA) within and among 3 population clusters in *S. sitiens*.

T = total population; N,C,S = northern, central and southern cluster, respectively; df = degrees of freedom; SSD = sum of square deviations; VC = variance components; inbreeding coefficients = F_{CT} (among groups), F_{SC} (among populations within groups), F_{ST} (among populations), F_{IS} (within populatios), F_{IT} (within individuals).

10. Abbreviations

AB-QTL	Advanced backcrosss quatitative trait locus strategy
AFLP	Amplified fragment length polymorphism
agar.	Agarose
AMOVA	Analysis of molecular variance
AVRDC	Asian Vegetable Research and Development Center
BC_i	i-th backcross
bp	Base pairs
CAPS	Cleaved amplified polymorphic sequences
cDNA	Complementary DNA
chr.	Chromosome
cM	Centi-Morgan
COS	Conserved ortholog set
cv.	Cultivar variety
DNA	Desoxyribonucleic acid
dNTPs	Deoxyribonucloetide triphosphate
EBN	Endosperm balance number
EST	Expressed sequence tag
F_i	i-th filial generation
FISH	Fluorescent in-situ hybridisation
IBD	Isolation by distance
J	*Solanum* Section *Jjuglandifolium*
kb	Kilobase pairs
(#)L	Long chromosome arm
myr	Million years ago
° N	Northern latitude
NA	Not available/applicable
N_e	Effective population size
Nm	Gene flow
PCA	Polymerase chain reaction/Principal component analysis
pers. comm..	Personally communicated
PGRU	Plant Genetic Resource Unit
QTL	Quantitative ttrait locus
RAPD	Randomly amplified polymorphic DNA
RE	Resctriction enzyme
ref.	Reference
RF	Recombination fraction
RFLPs	Restriction fragment length polymorphism
(#)S	Short chromosome arm
° S	Southern latitude
SCAR	Sequence characterized amplified region
Sd	Segregation distorter locus
sect.	Section (taxonomy)
SSCP	Single-strand conformational polymorphisms
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
subsect.	Subection (taxonomy)
Ta, Tb	Translocated chromosome pairs in *Juglandifolium*
TGRC	Tomato Genetics Resource Center
USDA-ARS	United States Department of Agriculture, Agricultural Research Station

11. Acknowledgements

Among the many people who have contributed to the completion of this thesis I would like to give special thanks to:

- Dr. Roger Chetelat, director of the C. M. Rick Tomato Genetics Resource Center at UCDavis, for giving me the opportunity to conduct research, for his guidance, criticism and patient support.

- PD Dr. Klaus Pillen, my professor in Germany, for mentoring and practical consultation throughout the duration of this project.

- My lab colleagues, foremost Elaine B. Graham, Sheh May Tam, Carl M. Jones, Sompid Samipak and Yuanfu Ji. I am very grateful for invaluable advice, countless favors, encouragement and friendship. Their company over the years has been highly enjoyable and entertaining.

- Rebeca Benitez, the greatest student assistant ever, for her support with data collection and plant maintenance. Her dedication and motivation were unique; it was a pleasure to work with her.

- Katie Smith, the TGRC plant collections manager, for her unparalleled caretaking of the greenhouse plants.

My thanks also go to the remainder of the TGRC crew, and the former plant collections manager Jennifer Peterson, for her help with material collections and the provision of helpful information.

Furthermore, I want to thank Garry W. Pearson, the Plant Sciences greenhouse manager for his excellent work.

I thank the former members of the Chetelat lab, Mike A. Canady and Ricardo A. Pertuzé whose scientific work has greatly contributed to the present study.

Also, I want to thank all my colleagues in the department of Plant Sciences, especially my neighbors in Asmundson Hall, for scientific advice, technical assistance and sharing of equipment.

Last but not least I am grateful to my friends, for their enduring friendship over decades and distances.

I dedicate this work to…

…Manfred, for his unwavering support and friendship, and for generously sharing his home during my years at UCDavis, and

…Werner, for unconditional love.