Annals of Botany **105**: 535–554, 2010
doi:10.1093/aob/mcq009, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

Genetic diversity and population structure in the tomato-like nightshades *Solanum lycopersicoides* and *S. sitiens*

Elena Albrecht[1,†], Miguel Escobar[2] and Roger T. Chetelat[1,*]

[1]*Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA and* [2]*Universidad de Chile, Depto. Producción Agrícola, Universidad de Chile, Casilla 1004, Santiago, Chile*
* *For correspondence. E-mail trchetelat@ucdavis.edu*
[†]*Present address: Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands.*

Received: 9 October 2009 Returned for revision: 24 November 2009 Accepted: 10 December 2009 Published electronically: 13 February 2010

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

• *Background and Aims* Two closely related, wild tomato-like nightshade species, *Solanum lycopersicoides* and *Solanum sitiens*, inhabit a small area within the Atacama Desert region of Peru and Chile. Each species possesses unique traits, including abiotic and biotic stress tolerances, and can be hybridized with cultivated tomato. Conservation and utilization of these tomato relatives would benefit from an understanding of genetic diversity and relationships within and between populations.
• *Methods* Levels of genetic diversity and population genetic structure were investigated by genotyping representative accessions of each species with a set of simple sequence repeat (SSR) and allozyme markers.
• *Key Results* As expected for self-incompatible species, populations of *S. lycopersicoides* and *S. sitiens* were relatively diverse, but contained less diversity than the wild tomato *Solanum chilense*, a related allogamous species native to this region. Populations of *S. lycopersicoides* were slightly more diverse than populations of *S. sitiens* according to SSRs, but the opposite trend was found with allozymes. A higher coefficient of inbreeding was noted in *S. sitiens.* A pattern of isolation by distance was evident in both species, consistent with the highly fragmented nature of the populations *in situ*. The populations of each taxon showed strong geographical structure, with evidence for three major groups, corresponding to the northern, central and southern elements of their respective distributions.
• *Conclusions* This information should be useful for optimizing regeneration strategies, for sampling of the populations for genes of interest, and for guiding future *in situ* conservation efforts.

Key words: Microsatellites, allozymes, biogeography, *Solanum lycopersicoides*, *Solanum sitiens*, *Solanum chilense*, genetic diversity, tomato, *Solanum* section *Lycopersicoides*.

INTRODUCTION

The processes of domestication and artificial selection contributed to the depletion of genetic diversity in many crop plants, including tomato (Tanksley and McCouch, 1997). Without genetic variation, breeding efforts remain ineffective and crops may lack important traits such disease and insect resistances (Tanksley and McCouch, 1997). Wild germplasm constitutes a potentially valuable resource for crop improvement (Zamir, 2001). In the Solanaceae, crops such as potato, tomato and pepper have greatly benefited from the use of wild relatives in breeding programmes. Virtually all of the disease resistances in modern tomato varieties originated in related wild species (Rick and Chetelat, 1995).

The wild relatives of the cultivated tomato are a diverse group of species native to the Andean region of western South America. According to the most recent taxonomic treatment (Peralta *et al.*, 2008), there are 13 species within *Solanum* section *Lycopersicon*, considered the ‘tomato clade’, and two species each in *Solanum* sect. *Juglandifolia* and sect. *Lycopersicoides*. The present study focuses on the latter clade, which includes the sister taxa *Solanum lycopersicoides* and *Solanum sitiens*.

In nature, *S. lycopersicoides* and *S. sitiens* consist of small, highly fragmented populations found only in the Atacama Desert region of southern Peru and northern Chile (Chetelat *et al.*, 2009). Both species grow as short-lived perennials on arid, rocky slopes and dry washes. Both species are rare. *S. lycopersicoides* is found in only a few drainages on the western slopes of the main Andean cordillera near the Peru–Chile border. It thrives at altitudes of up to approx. 3700 m, higher than any other tomato relative, where it is exposed to chilling and sub-zero temperatures. It is usually found on the cooler, less arid south-facing slopes. In contrast, *S. sitiens* grows in some of the driest parts of the hyper-arid Atacama Desert of northern Chile, the driest desert on earth, with annual rainfall well below 10 mm in some areas (Houston, 2006). Its distribution is restricted to a small part of the Atacama, approx. 230 km from north to south and 2500 to 3500 m elevation, on the slopes of the Cordillera Domeyko and other minor cordillera between the coastal and the principal Andean ranges (Rick, 1988; Smith and Peralta, 2002; Peralta *et al.*, 2008; Chetelat *et al.*, 2009). This zone is so dry that only a handful of other perennial plants, mainly herbaceous xerophytes, such as *Calandrinia crassifolia*, *Adesmia atacamensis* and *Nolana* spp. (Rick, 1988; http://tgrc.ucdavis.edu), are found in association with *S. sitiens.*

Besides their extreme abiotic stress tolerances, *S. lycopersicoides* and *S. sitiens* harbour disease resistances

© The Author 2010. Published by Oxford University Press on behalf of the Annals of Botany Company. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

and other traits that are of potential value for tomato improvement (Rick, 1988; Chetelat *et al.*, 1997). With the aid of embryo rescue, both species have been hybridized with cultivated tomato (Chetelat *et al.*, 1997; Pertuze *et al.*, 2003). A library of introgression lines each containing one to several chromosome segments from *S. lycopersicoides* in the genetic background of cultivated tomato was recently developed (Canady *et al.*, 2005). These lines provide a convenient tool for locating genes or quantitative trait loci (QTL) for disease resistance, abiotic stress tolerances or other economic traits. QTL conferring resistance to *Botrytis cinerea*, the causal pathogen of grey mould disease, were mapped using these lines (Davis *et al.*, 2009). Exploitation of traits from *S. sitiens* by similar means can be anticipated in the future.

Natural populations of both species are threatened by various factors in the native region and have suffered declines in recent decades (Chetelat *et al.*, 2009). Grazing by livestock and agricultural development pose a serious threat to *S. lycopersicoides*, whereas *S. sitiens* is vulnerable to disturbance and habitat loss associated with mining activities. Thus, an understanding of how genetic diversity is distributed among and within natural populations of these species is needed, both from a practical standpoint (e.g. for planning germplasm conservation efforts, or for searching for desired traits), as well as to address more fundamental questions related to processes of local adaptation and speciation.

Genetic diversity in the wild tomato species has been studied using various marker techniques, including allozymes (Rick and Fobes, 1975; Rick *et al.*, 1976, 1977, 1979; Rick and Tanksley, 1981; Rick and Holle, 1990; Breto *et al.*, 1993), restriction fragment length polymorphism (RFLP) (Miller and Tanksley, 1990), random amplification of polymorphic DNA (RAPD) (Egashira *et al.*, 2000; Ercolano *et al.*, 2005), amplified fragment length polymorphism (AFLP) (Nuez *et al.*, 2004; Nakazato *et al.*, 2008), microsatellites (Alvarez *et al.*, 2001) and DNA sequences (Caicedo and Schaal, 2004; Roselius *et al.*, 2005). A consensus finding of these analyses is that self-incompatible species as a whole are vastly more diverse than self-compatible taxa. Both *S. lycopersicoides* and *S. sitiens* are self-incompatible, obligate outcrossers (Rick, 1988), and thus can be expected to harbour substantial genetic variation both within and between populations.

Allozymes, the first molecular markers used for studies of natural variation, are limited by the number of loci, low levels of polymorphism compared with DNA-based markers due to their lower effective mutation rates, and lack of neutrality, as proteins are targets of natural selection (e.g. Karl and Avise, 1992; Dhuyvetter *et al.*, 2004). Simple sequence repeat (SSR) markers are much more abundant and polymorphic (Morgante and Olivieri, 1993). Compared with some other marker technologies (AFLPs, RAPDs), SSRs provide co-dominant expression, high 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 (Bindler *et al.*, 2007), neutrality to selection (reviewed in Li *et al.*, 2002) and cost-effectiveness (Loridon *et al.*, 2005). For tomato, a large number of SSR loci with known map locations are available (http://www.sgn.cornell.edu), most of which are derived from expressed sequence tag (EST) sequences, thus avoiding over-representation of centromeric and telomeric regions (Frary *et al.*, 2005).

The goals of the current project were to quantify the amounts of genetic diversity in *S. lycopersicoides* and *S. sitiens* at the protein and DNA levels, and to characterize the genetic structure of populations within each species. The results are discussed in terms of the present status of these tomato relatives, their likely demographic histories and possible conservation strategies.

MATERIALS AND METHODS

Plant materials

Fourteen *S. lycopersicoides* and seven *S. sitiens* populations (Table 1; Fig. 1) were selected for this study based on three criteria: (1) geographical distribution (the aim was to sample the entire range of each), (2) population and sample size (prioritizing relatively large populations for which seed from multiple individuals was available) and (3) quality of the collection site information (i.e. precise geographical coordinates were desired). One population of *Solanum chilense* (LA2773 from Zapahuira, Tarapaca, Chile) was included to provide a reference with a wild tomato. This species was chosen because it has a similar ecology and mating system, and its geographical distribution overlaps those of *S. lycopersicoides* (often sympatric) and *S. sitiens* (Fig. 1). Samples of *Solanum lycopersicum* ('VF36'), *Solanum pennellii* (LA0716) and F_1 *S. lycopersicum* × *S. lycopersicoides* (LA3857) served as marker controls. Seeds and collection data were obtained from the C. M. Rick Tomato Genetics Resource Center (TGRC, University of Davis). For each source population, approx. 25 plants were genotyped, generally composed of five full- or half-sib progeny from five individual mother plants (Table 1; Supplementary Data Tables S1 and S2, available online). For populations collected from fewer than five seed-bearing plants *in situ* (Table 1), the 25 genotyped plants were derived from as many source plants as possible. To avoid genetic changes associated with artificial selection, drift and/or inbreeding, plants were grown from seed collected *in situ* wherever possible.

To promote germination, seed were treated with 2·5 % sodium hypochlorite for 40 min, rinsed under running water for 15 min and incubated in a germination chamber (25 °C, 12-h photoperiod). Those that failed to germinate within 10 d were nicked to facilitate radicle emergence. At the cotyledon stage, seedlings were transplanted into flats filled with a soil mix optimized for desert plants (http://tgrc.ucdavis.edu/soils.html). Small plantlets were grown in 4-inch pots of the same soil. Because of their sensitivity to soil pathogens, shoots of *S. sitiens* were grafted onto F_1 *S. lycopersicum* × *S. pennellii* rootstocks. All plants were grown at the Plant Sciences greenhouse facility at UC Davis.

Marker analysis

Allozymes. Plant shoot sample preparation, horizontal slab gel electrophoresis and staining procedures were carried out as described in Chetelat *et al.* (1997). Eight enzyme systems

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

TABLE 1. *S. lycopersicoides and S. sitiens accessions used in the diversity analysis*

No.	Accession number	Population name	Department or Region	Country	Latitude (°S)	Longitude (°W)	Altitude (m)	Population size	Sample size	Year of collection
	S. lycopersicoides									
1.	LA4018	Aricota #1	Tacna	Perú	−17·333	−70·250	2888	many	1	1988
2.	LA2387	Aricota #2	Tacna	Perú	−17·355	−70·313	2852	NA	4	1981
3.	LA1964	Chupapalca	Tacna	Perú	−17·761	−69·912	3459	50	25	1979
4.	LA1966	Palca	Tacna	Perú	−17·767	−69·950	3134	60	15–20	1979
5.	LA2781	Putre #1	Tarapacá	Chile	−18·199	−69·540	3736	large	10	1986
6.	LA2777	Putre #2	Tarapacá	Chile	−18·203	−69·564	3462	large	9	1986
7.	LA2776	Perquejeque	Tarapacá	Chile	−18·210	−69·596	3115	very large	12	1986
8.	LA2772	Zapahuira	Tarapacá	Chile	−18·271	−69·580	3416	many	6	1986
9.	LA4320	Lluta	Tarapacá	Chile	−18·318	−69·805	1509	small	7	2005
10.	LA4130	Pachica	Tarapacá	Chile	−18·908	−69·604	2672	>20	1	2001
11.	LA4131	Esquina	Tarapacá	Chile	−18·926	−69·551	2341	>18	7	2001
12.	LA4126	Nama	Tarapacá	Chile	−19·287	−69·396	3156	>50	11	2001
13.	LA4123	Camiña	Tarapacá	Chile	−19·306	−69·421	2599	>50	19	2001
14.	LA2730	Moquella	Tarapacá	Chile	−19·404	−69·600	1719	NA	7	1985
	S. sitiens									
15.	LA4116	Paqui	Antofagasta	Chile	−22·159	−68·782	2935	>100	8	2001
16.	LA4114	Carbonatera	Antofagasta	Chile	−22·191	−68·757	2736	>35	10	2001
17.	LA4113	Cere	Antofagasta	Chile	−22·235	−68·762	2652	>20	12	2001
18.	LA4112	Limón Verde	Antofagasta	Chile	−22·617	−68·948	2780	>47	21	2001
19.	LA4331	Quimal	Antofagasta	Chile	−22·969	−68·821	3074	15–20	9	2005
20.	LA4110/4111	San Juan	Antofagasta	Chile	−23·098	−69·033	2718	>60	42	2001
21.	LA4105	Escondida	Antofagasta	Chile	−24·211	−69·241	2618	>20	19	2001

Passport data were obtained from http://tgrc.ucdavis.edu. More detailed collection information is available on the website. 'Population size' is the number of individuals observed at the site at the time of collection, and 'Sample size' is the number of plants from which seeds were obtained. The geographical locations are shown in Fig. 1.

yielded 14 polymorphic loci of a total of 16 (Table 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 were recorded as migration differences (in mm) relative to the tomato (*S. lycopersicum*) reference on the same gels.

Microsatellites. Marker information including primer sequences of tomato EST-derived SSRs were obtained from the SOL Genomics Network database (http://www.sgn.cornell.edu; Frary *et al.*, 2005). Selection criteria were (1) even genomic distribution and the avoidance of an over-representation of centromeric or telomeric regions, (2) amplification quality, (3) frequency of polymorphisms and (4) ease of scoring (i.e. sufficient spatial separation of alleles on the polyacrylamide gel). DNA was extracted from young leaf tissue as described by Fulton *et al.* (1995). Thermocycling reactions were according to Frary *et al.* (2005): after an initial denaturation for 5 min at 94 °C, 40 cycles of amplification were run, each consisting of 30 s denaturation at 94 °C, 45 s annealing at 55 °C and 45 s extension at 72 °C followed by a final extension of 72 °C for 10 min. Samples were resolved on 5·2 % polyacrylamide using a LiCor 4200 sequencing apparatus to detect fluorescently labelled fragments. Fluorescent dyes (IR-700 or IR-800) were incorporated into PCR amplicons by the tailed primer method via a labelled M13 primer (TTTCCCAGTCACGACGTT; MWG-Biotech) that was added at 0·05 μg μL^{-1} to the PCR reaction mix. A total of 15 of 37 tested markers were used for the population genotyping (Table 2). Alleles were recorded according to their fragment sizes (in base pairs). Rare alleles were validated by repeated SSR genotyping.

Not all plants were fully genotyped with both sets of markers. Three populations were analysed with SSRs but not with allozymes. In addition, some plants died before genotyping could be completed. Data analyses were based on the full datasets for each marker type, as well as a reduced, 'common' set of fully genotyped individuals to use for comparisons between marker types and species.

Statistical tests

Prior to the computational analyses, data were tested for agreement with Hardy–Weinberg (HW) expectations to avoid the inclusion of loci that did not exhibit neutral genetic variation, which would violate the assumptions of most test statistics. A Markov-chain algorithm (Guo and Thompson, 1992) was implemented to test for deviations from HW equilibrium using the 'exact Hardy–Weinberg test' (Haldane, 1954; Weir, 1990; Guo and Thompson, 1992). The single- and multiple-samples version of the score test ('U test'; Raymond and Rousset, 1995) was employed to test for heterozygote deficiency or excess. HW tests were performed in GENEPOP online (http://genepop.curtin.edu.au; Raymond and Rousset, 1995). Multiple statistical tests were adjusted by the Bonferroni correction.

For each locus and population, the number of alleles (k), the number of private alleles (i.e. unique to a population), allele frequencies, percentage polymorphic sites (P), allelic richness per locus and sample (R_S) and total samples (R_T), Nei's gene diversity (H_E), Wright's inbreeding coefficient per population

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

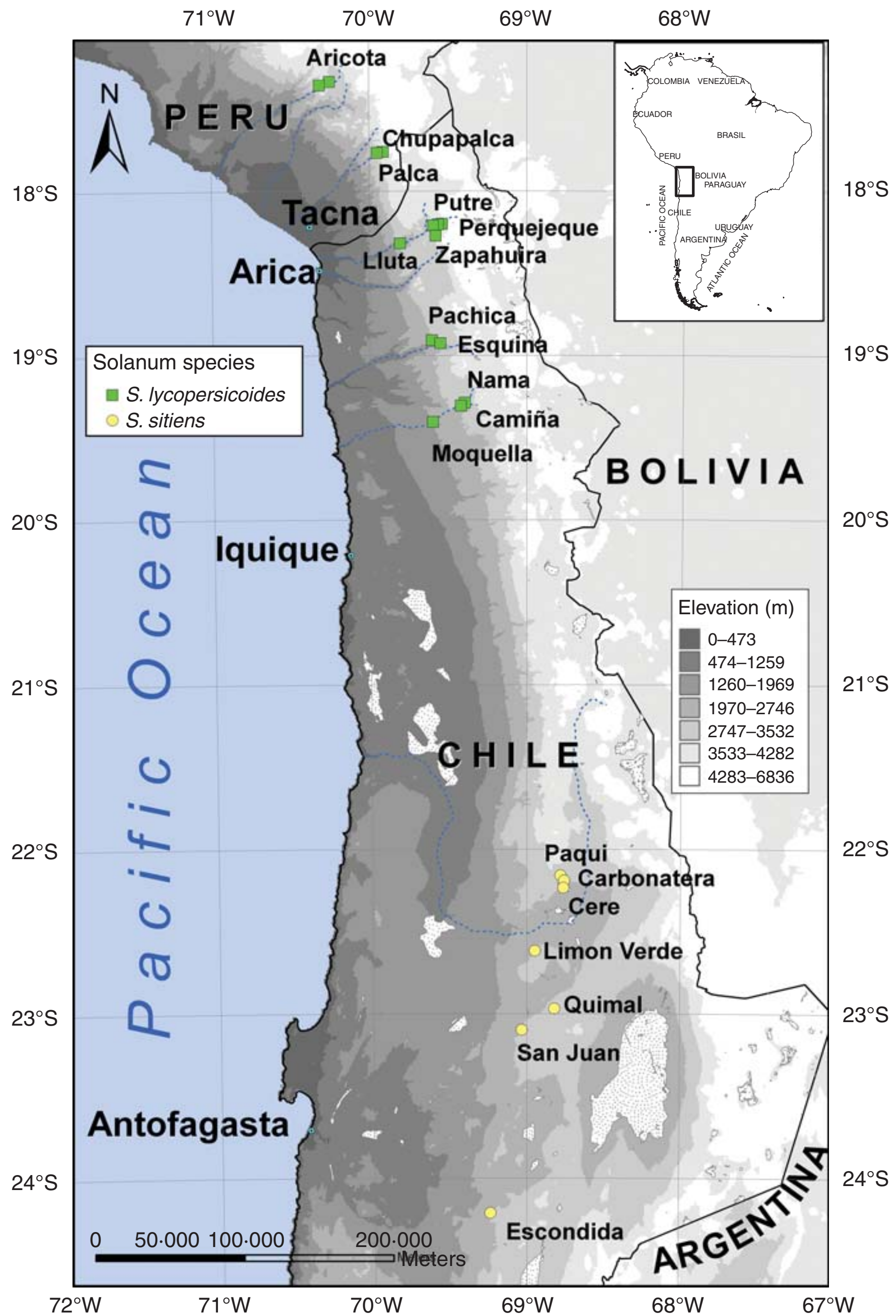

FIG. 1. Map showing the geographical locations of *S. lycopersicoides* and *S. sitiens* populations used in this study. Each population is named after the collection site (see Table 1).

(F_{IS}) and the population pair-wise F_{ST} were computed in FSTAT ver. 2.9.3 (Goudet, 2001). To obtain a measure independent of sample size, standardizations were realized via an adaptation of the rarefaction method (El Mousadik and Petit, 1996). All estimates were based on the smallest number of individuals genotyped at a given locus within a sample. For comparisons among the three species, the population averages for allele numbers and estimates of the main descriptive parameters were recalculated from a common set of markers. These included seven allozyme loci and 14 SSR loci (the 'common' dataset; Table 2). The inbreeding coefficient per population (F_{IS}) was estimated according to Nei (1987).

The inbreeding coefficient per locus and the population pair-wise F_{ST} were calculated as Weir and Cockerham's (1984) unbiased estimator θ. Significant deviations of the inbreeding coefficient per population (F_{IS}) from the null hypothesis of panmixia (i.e., F_{IS}) 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 pair-wise 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

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

Table 2. *List of allozyme and microsatellite markers used*

	Marker name	Chr	Position (cM)	LOD score	*N*, *S. lycopersicoides*	*N*, *S. sitiens*
Allozymes						
1.	*Aco-1*	12	3·7	–	–	150
2.	***Aco-2***	7	38·6	–	296	150
3.	***6-Pgdh-2***	12	47·1	>3	208	149
4.	***6-Pgdh-3***	5	26·6	–	272	146
5.	***Idh-1***	1	0	–	279	146
6.	***Adh-2***	6	34·8	low	285	124
7.	*Fdh-1*	2	–	–	297	–
8.	***Got-2***	7	42·4	–	298	142
9.	***Pgi-1***	12	39·1	low	298	148
10.	***Pgm-1***	3	25	–	–	147
11.	*Pgm-2*	4	40	low	298	150
	Mean				281·2	145·2
	Total				298	150
Microsatellites						
1.	**SSR125**	2	106·6	2	315	154
2.	**SSR15**	8	22·7	2	306	152
3.	**SSR320**	3	158·0	2	317	153
4.	**SSR325**	5	18·5	2	314	154
5.	**SSR341**	1	137·5	2	313	154
6.	**SSR345**	12	72·5	2	311	149
7.	**SSR43**	4	14·0	3	316	155
8.	**SSR50**	2	70·5	2	316	155
9.	SSR578	6	44·0	≥3	317	–
10.	**SSR599**	9	103·0	3	315	155
11.	**SSR74**	10	74·0	2	315	155
12.	**SSR76**	11	38·0	2	316	155
13.	**SSR80**	11	20·0	2	316	155
14.	**SSR85**	10	55·0	2	316	153
15.	**SSR98**	1	31·8	≥3	317	155
	Mean				314·7	153·7
	Total				317	155

Chromosomal positions of allozyme loci are from Tanksley *et al.* (1992) and SSRs are from the tomato-EXPEN 2000 map, available at http://www.sgn.cornell.edu. N = number of individuals genotyped. Gene symbols in bold type indicate common markers used in the analysis of *S. lycopersicoides*, *S. sitiens* and the reference species *S. chilense*. LOD scores represent the statistical significance of linkage for each marker locus. A LOD=2 corresponds to $P \leq 0{\cdot}01$, LOD=3 to $P \leq 0{\cdot}001$.

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. Isolation by distance (IBD) and other matrix-based correlations were tested with the Mantel test in ARLEQUIN using 10 000 permutations (Mantel, 1967; Smouse *et al.*, 1986). For IBD tests, transformations of Weir and Cockerham's (1984) unbiased estimator θ via the formula $F_{ST}/(1 - F_{ST})$ and of geographical distances into log-transformed distances were obtained in GENEPOP, as described by Rousset (1997). A single negative pair-wise 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 non-parametric Mann–Whitney test were used to assess differences between allozyme and microsatellite diversity estimates. In the reduced, common dataset, 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 pair-wise F_{ST} (θ) matrices. Spearman's rank test was used to evaluate correlations of various diversity estimates and allele frequencies with geographical 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 the parameter of Barton and Slatkin (1986). Wright's traditional equation, $Nm \equiv (1 - F_{ST})/4F_{ST}$ (Wright, 1951), was included as a comparison. Populations were investigated for signatures of recent bottlenecks using BOTTLENECK ver. 1.2.02 (Piry *et al.*, 1999) with 10 000 iterative runs assuming the infinite alleles model (IAM) via two detection methods, in the absence of historical population information or reference population data. Significant heterozygosity excess was tested using the Wilcoxon sign-rank test as it shows robustness with few (fewer than 20) polymorphic loci (Cornuet and Luikart, 1996; Luikart, 1997; Piry *et al.*, 1999).

In order to quantify the partitioning of genetic variability among populations, global analyses of molecular variance (AMOVA) were conducted in ARLEQUIN ver. 3.01 (Weir and Cockerham, 1984; Excoffier *et al.*, 2005). Significance tests were based on non-parametric permutations (16 000) on the covariance components associated with the structural levels (within and among individuals, within and among

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

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). Tree reconstruction was based on the UPGMA (Sokal and Sneath, 1963) and the neighbour-joining (Saitou and Nei, 1987) methods. Bootstrapping was carried out over 10 000 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). Principal component analyses (PCA) were conducted on Weir and Cockerham's pair-wise θ using 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 burn in and run length of 10 000 steps each was used. The admixture and correlation models were chosen to analyse 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 $\Pr(X|K)$ and its second-order rate of change (ΔK) as described by Pritchard *et al.* (2000) and Evanno *et al.* (2005), respectively. GIS maps were generated using SIG ArcReader v. 9·3.1 and ArcMap (ESRI, Redlands, CA, USA).

RESULTS

Hardy-Weinberg tests

Thirteen of 16 tested allozyme loci were polymorphic in *S. lycopersicoides*. Four of the polymorphic loci were excluded from further analyses: *6-Pgdh-1* did not fulfil the criteria of >1 mm band separation, and *Aco-1*, *Got-3* and *Pgm-1* showed significant ($P < 0{\cdot}042$) deviations from HW expectations in the form of heterozygote deficiencies in >50 % of the populations. This left nine loci that were informative and met the above criteria (Table 2). In *S. sitiens*, 13 of 16 loci were polymorphic. *6-Pgdh-1* was discarded for the same reason as in *S. lycopersicoides. Adh-1* and *Got-3* showed significant ($P < 0{\cdot}008$) heterozygote deficiencies, leaving a total of ten informative loci for this species. After removal of marker loci that showed scoring uncertainties or deviated from HW equilibrium, only two populations, both *S. sitiens*, demonstrated a significant departure from HW expectations for any of the remaining marker loci. The population from Cere was deviant by both tests, exhibiting a significant excess of heterozygotes ($P < 0{\cdot}006$). The Paqui population was deviant according to the 'exact Hardy-Weinberg test', and showed both heterozygote excess and deficiency, depending on the locus ($P < 0{\cdot}005$).

Amplicon sizes of SSR loci did not vary in a strict stepwise fashion, and therefore the IAM (Kimura and Crow, 1964) rather than the stepwise mutation model (SSM; Ohta and Kimura, 1973) was assumed to best describe the microsatellite mutation mode. For eight of the 37 SSR loci tested, no amplicon was obtained, another five were monomorphic and nine exhibited ambiguous banding patterns in one or both of the species, leaving a total of 15 informative markers for the *S. lycopersicoides* analyses and 14 for the *S. sitiens* analyses (Table 2). These SSR loci all showed compliance with HW expectations. The multi-sample version of the 'U-test' detected one *S. lycopersicoides* population (Palca; $P < 0{\cdot}003$) and one *S. sitiens* population (Escondida; $P < 0{\cdot}004$) with heterozygote deficiencies.

Allelic diversity

Fewer alleles were detected with allozymes than with SSRs. A total of 25 alleles (four private) were found at nine allozyme loci among 298 individuals in 12 populations of *S. lycopersicoides* (Table S1, available online). The number of alleles per locus ranged from two to five (mean = 2·8). Ninety SSR alleles (13 private) were detected at 15 loci in 317 individuals of 14 populations. The number of alleles per locus ranged from two to 12 (mean = 6·0). In *S. sitiens*, the difference between marker systems was less pronounced. A total of 31 alleles (four private) were detected at ten allozyme loci among 150 individuals in six *S. sitiens* populations, and there were 2–4 (mean = 3·1) alleles per locus (Table S2). The SSRs detected 60 alleles (eight private) at 14 loci among 155 individuals in seven *S. sitiens* populations. There were 2–8 alleles per SSR locus (mean = 4·3). Analysis of a subset of individuals fully genotyped with a common set of allozymes and SSRs supported these trends in the relative allelic diversity of the two marker systems.

Sixty-one SSR alleles were found among the 23 individuals of the single *S. chilense* population genotyped for comparison purposes. Of the 98 alleles scored at the 14 common loci, 38 were unique to *S. lycopersicoides*, ten to *S. sitiens*. The two species shared 58·0 and 83·3 % of their alleles, respectively. *S. lycopersicoides* shared 36·4 % of its alleles with *S. chilense*, and 45·0 % with *S. sitiens*; conversely, of the alleles found in *S. chilense*, 52·5 % were shared by *S. lycopersicoides* and 44·3 % by *S. sitiens*.

Population-level diversity

In *S. lycopersicoides*, diversity estimates based on SSRs were two to four times as large as those based on allozymes (Table S1, Fig. 2). The proportion of polymorphic sites (P) at allozymes was 42·6 %, with the highest levels (55·6 %) observed in the Aricota #2, Camiña and Moquella populations, and the lowest (22·2 %) in Esquina. At SSR loci the mean P was substantially higher (90·5 %), with the lowest rates (73·3 %) in Pachica and the highest (100 %) in Zapahuira, Lluta, Nama and Moquella. Average allelic richness (R_S) at allozymes over populations was 1·26–1·63 (mean = 1·42). At SSR loci calculations of allelic richness were based on a minimum sample size of 17 diploid individuals. The Pachica population was excluded from the dataset in order to avoid a downward bias caused by its lower sample size ($n = 11$ vs. mean $n = 23{\cdot}5$ in the remaining

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

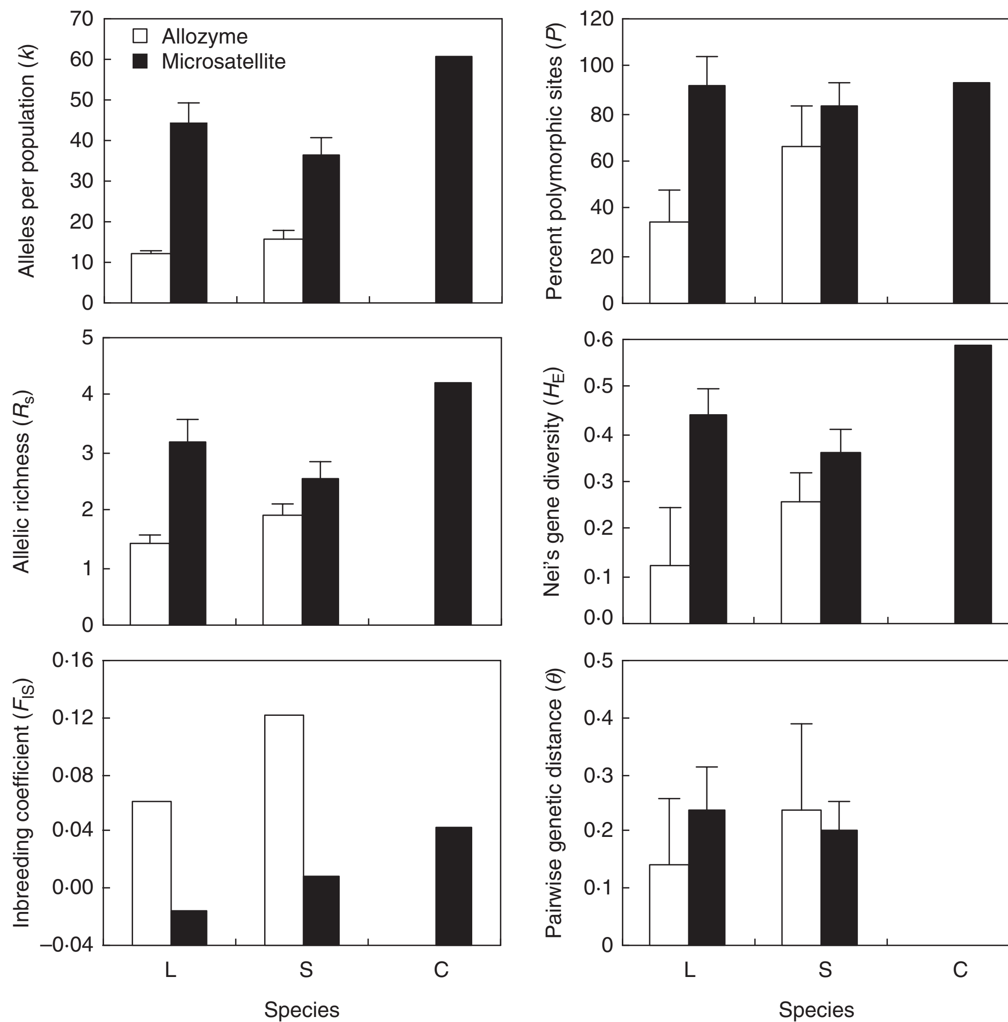

FIG. 2. Comparison of population genetic diversity in *S. lycopersicoides* (L), *S. sitiens* (S) and *S. chilense* (C). Values are based on a common set of markers (Table 2) applied to 12–13 *S. lycopersicoides* populations (one additional population was genotyped with SSRs only), 6–7 *S. sitiens* populations and one *S. chilense* population (SSRs only). Bars represent allozyme data and microsatellite data, as indicated. Values are the averages (+s.d.) of all tested populations within each species.

populations). SSR R_S ranged from 2·58 (Aricota #2) to 3·49 (Zapahuira) with an average of 3·01. In both marker systems the observed heterozygosity (H_O) was slightly below the expected heterozygosity (H_E), indicating a tendency towards inbreeding. The expected heterozygosity (H_E) averaged 0·117 at allozymes. Among populations the genetic diversity estimates were lowest in Moquella ($H_E = 0{\cdot}052$) and highest in Palca ($H_E = 0{\cdot}171$). H_O ranged from 0·050 in Esquina to 0·194 in Palca. With SSRs, the average gene diversity across loci and populations (H_E) was 0·422. Values for H_O ranged from 0·344 (Aricota #1) to 0·486 (Moquella), and averaged 0·433. Estimates among populations ranged from 0·360 in Aricota #1 to 0·530 in Lluta. The level of inbreeding within populations of *S. lycopersicoides* was near zero for SSR loci. The SSR data indicated the highest level of inbreeding in Palca ($F_{IS} = 0{\cdot}149$; significantly different from zero at $P <$ 0·0033), and the lowest in Aricota #2 ($F_{IS} = -0{\cdot}102$). The two types of marker data were significantly different in *S. lycopersicoides*: all but the F_{IS} and the F_{IT} per locus and the F_{IS} per population revealed significant differences ($P < 0{\cdot}05$) in the subset of fully genotyped individuals.

In *S. sitiens*, diversity estimates based on SSRs were generally higher than those based on allozymes (Table S2, Fig. 2). Seventy per cent of allozyme loci were polymorphic in *S. sitiens*, ranging from 100 % in the Paqui population to 60 % in the Carbonatera, San Juan and Escondida populations. Eighty-four per cent of all sites were polymorphic at the SSR loci, ranging from 71·4 % in Escondida to 100 % in Limón Verde. Average allelic richness (R_S) at allozyme loci ranged from 1·64 (Escondida) to 2·40 (Paqui) with a mean of 2·05 (base number = 14 individuals). At SSR loci, R_S ranged from 2·16 (Carbonatera) to 3·07 (Limón Verde), with a mean of 2·53 (base number = 17 individuals). The average gene diversity at allozyme loci was lowest in Escondida ($H_E = 0{\cdot}163$), and highest in Paqui ($H_E = 0{\cdot}390$). Observed heterozygosity ranged among populations from 0·145 (Escondida) to 0·383 (Carbonatera). At SSR loci, the mean H_E was lowest in Carbonatera ($H_E = 0{\cdot}289$) and highest in Limón Verde (0·452). H_O was lowest in Carbonatera ($H_O =$ 0·289) and highest in Limón Verde ($H_O = 0{\cdot}452$). In both

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

marker systems the observed heterozygosity (H_O) was lower than the expected heterozygosity (H_E), and the discrepancy was larger than in *S. lycopersicoides*, indicating higher levels of inbreeding in *S. sitiens*. This was also reflected in F_{IS} statistics (inbreeding coefficient per population), which were higher in *S. sitiens*. F_{IS} values at allozyme loci ranged from 0·019 in Paqui to 0·121 in Cere, although none was significantly different from zero. At SSR loci, all populations showed F_{IS} values close to zero, except Escondida ($F_{IS} = 0{\cdot}128$, not significant). A comparison of the two types of markers revealed significant differences only in the k, R_S and F_{IS} statistics.

Comparison among species. Estimates based on a common set of markers indicated that the selected *S. chilense* population was genetically more diverse than the average *S. lycopersicoides* or *S. sitiens* population (Fig. 2). Of the two nightshades, *S. sitiens* was more diverse according to allozyme data, and *S. lycopersicoides* more diverse based on SSRs.

Signatures of recent population bottlenecks

There was relatively little sign of historical genetic bottlenecks in most populations. Only one *S. lycopersicoides* (Aricota #2) and one *S. sitiens* population (Limón Verde) were highlighted by both statistical methods, the former with SSRs only, the latter with allozymes only. In addition, a significant excess of heterozygosity ($P < 0{\cdot}003$) was detected by SSRs in the *S. lycopersicoides* population from Chupapalca (LA1964), which was accompanied by a mode-shift distortion. And the *S. sitiens* population from Escondida (LA4105) also exhibited a significant excess of heterozygosity ($P < 0{\cdot}004$). There was little evidence of a correlation between census or sample size at the time of collection and diversity estimates in the sample. For instance, the Escondida and Limón Verde populations were both relatively large (Table 1).

Correlations between genetic and geographical distances

Among *S. lycopersicoides* populations, geographical distances were smallest between Putre #2 and Perquejeque (2·5 km apart), and largest between Aricota #2 and Moquella (240 km), with a mean of 99 km between any two populations (Fig. 1). Altitudinal distances varied substantially among populations, ranging from just 3 m between Chupapalca and Putre #2, to 2227 m between Putre #1 and Lluta, with a mean of 744 m between populations. At allozyme loci, Weir and Cockerham's pair-wise genetic distance was greatest ($\theta = 0{\cdot}532$) between populations Putre #1 and Esquina, and smallest ($\theta = -0{\cdot}001$) between the neighbouring populations at Putre #1 and Zapahuira, with a mean of 0·169. Most values were significant at the adjusted 5 % nominal level. The average pair-wise θ was highest for the population Esquina ($\theta = 0{\cdot}345$) and lowest for one of the central populations (Perquejeque; $\theta = 0{\cdot}088$). The correlation between allozyme genetic and geographical distance was low but significant ($r = 0{\cdot}223$, $P < 0{\cdot}05$). The correlation between genetic and altitudinal distance was stronger ($r = 0{\cdot}381$, $P < 0{\cdot}05$). At SSR loci, θ measures ranged from 0·060 for Nama and Camiña (neighbouring sites) to 0·464 for Aricota #1 and Esquina (located far apart), with a mean of 0·261. The northernmost population, Aricota #1, displayed the highest average θ (0·342), which was almost twice as high as the lowest average value (0·199) in Perquejeque at the centre of the distribution range, indicating isolation by distance. All estimates were highly significant ($P < 0{\cdot}001$). In contrast to the allozyme analyses, the Mantel test detected IBD at SSR loci. A highly significant positive correlation ($r = 0{\cdot}680$, $P < 0{\cdot}001$) was observed between genetic variation and geographical distance. No correlation was observed between genetic and altitudinal distance measures.

Geographical distances among *S. sitiens* populations were similar to those of *S. lycopersicoides*, ranging from 4·3 km between Paqui and Carbonatera to 233 km between Paqui and Escondida, with a mean distance of 95 km. Overall altitudinal distances among *S. sitiens* populations were much less than among *S. lycopersicoides*, ranging from 18 m between Carbonatera and San Juan to 456 m between Quimal and Escondida, with a mean distance of 163 m. At allozyme loci, pair-wise θ values ranged from 0·070 for neighboring populations Paqui and Carbonatera to 0·535 between distant Carbonatera and Escondida (mean $\theta = 0{\cdot}280$). The average pair-wise θ was highest for the southernmost population, at Escondida ($\theta = 0{\cdot}476$), and lowest for Limón Verde ($\theta = 0{\cdot}225$) at the centre of the distribution, consistent with expectations for IBD. The correlation between allozyme genetic distance and geographical distance was much more pronounced ($r = 0{\cdot}785$, $P < 0{\cdot}001$) than in *S. lycopersicoides*. There was no correlation between genetic and altitudinal distances. SSR data showed similar trends. F_{ST} values were substantially higher for the southernmost population at Escondida, providing more evidence that this accession is marginal and partially inbred. The Limón Verde again showed the greatest similarity to the other populations, providing further support for IBD. The Mantel test revealed a high correlation between genetic and geographical distance ($r = 0{\cdot}847$, $P < 0{\cdot}001$), but there was no correlation between genetic and altitudinal distances.

Gene flow

The private allele method estimated a very high migration rate in *S. lycopersicoides* based on allozyme data: 2·54 migrants on average per population, after correction for population size, or 1·23 according to Wright's formula. At SSR loci the average number of migrants was estimated at 0·72 after correction for population size, or 0·71 according to Wright's formula. In contrast, the estimated gene flow in *S. sitiens* was low in both marker systems: 0·36–0·64 migrants per population by allozymes, or 0·64–0·98 by SSR data.

Correlations between geographical location and genetic diversity

Latitude. In *S. lycopersicoides* the parameters P (polymorphism rate) and θ (pair-wise genetic distances) were correlated with latitude (the former with SSR data only), with both diversity parameters increasing towards the south. In *S. sitiens*, the allozyme data showed stronger signals of geographical clines. 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, the correlations indicated a decrease of genetic diversity and an increase of genetic distance from north to south.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

Longitude. Parameters P and θ were correlated with longitude in both *S. lycopersicoides* and *S. sitiens*.

Elevation. In *S. lycopersicoides*, P increased with elevation 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 by distance. In *S. lycopersicoides*, the average geographical distance was positively correlated with P (allozyme data) and θ (SSR data), and 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 geographical distance, i.e. towards the margins of the distribution, while θ increased with average geographical distance in both datasets.

Correlations between geographical parameters and allele frequencies

Most of the correlations between allele frequencies and geographical parameters were evident with latitude, followed by longitude and to a lesser extent with elevation (data not shown). The highest number of significant correlations was observed with the *S. lycopersicoides* SSR dataset, in which 32 % of the alleles showed a trend with latitude, 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 were observed at allozyme loci: 23 % showed trends with latitude, 10 % with longitude and 20 % with elevation.

Population genetic structure

AMOVA analysis. AMOVA tests revealed that most of the genetic variation was found within populations, with less among populations, for both species and for both marker types. In *S. lycopersicoides*, the majority of genetic variation, 75 and 88 % for SSRs and allozymes, respectively, was partitioned within populations, with a relatively small fraction partitioned among populations (Table S3). The proportion of genetic variation among individuals within populations (F_{IS}) was zero or nearly so. The F_{ST} and F_{IT} values were highly significant ($P < 0{\cdot}0005$). For *S. sitiens*, 78 and 66 % of SSR and allozyme variation, respectively, was within populations (Table S4). Results were highly significant ($P < 0{\cdot}001$).

Phylogenetic analysis. Three genetic distances (Reynolds' distance, Cavalli-Sforza's chord distance and Nei's genetic distance) in combination with two tree construction methods (UPGMA and neighbour-joining) were compared for each of the four datasets (i.e. a total of six dendrograms per dataset) 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 geographical groupings. The best results were obtained with Nei's genetic distances, except for the *S. lycopersicoides* SSR dataset, for which Cavalli-Sforza's chord distances generated the most likely results. Topologies generated by the UPGMA method were supported by higher bootstrap values and showed better congruence with geographical distribution than trees derived by the neighbour-joining approach.

For *S. lycopersicoides*, the phylogenetic tree derived from SSR data was highly supported and showed population groupings in concordance with geographical patterns, i.e. reflecting the major river drainages (Fig. 3). A divide into 'Peruvian' (cluster A) and 'Chilean' (clusters B and C) populations was evident in the hierarchy; the latter was further partitioned into a 'central' (cluster B) and 'southern' group (cluster C). The allozyme dendrogram showed low bootstrap support and a topology inconsistent with geographical distribution (data not shown). In *S. sitiens*, dendrograms based on either marker type identified clusters that corresponded well with geographical origins of the populations (Fig. 4). Escondida clearly emerged as an 'outlying' population (branch C), in agreement with its geographical isolation from the other accessions, and its relatively distinctive morphology (data not shown). The remainder of the populations clustered into a 'northern' group composed of three populations in close geographical proximity (cluster A), and a 'central' group composed of three populations that were more dispersed (cluster B). Similar results were obtained by a PCA of the data (Figs S1 and S2).

STRUCTURE analysis. The results of STRUCTURE analyses also corresponded well with the phylogenetic groupings. In *S. lycopersicoides*, the allozyme data failed to reveal any structural pattern (data not shown), and therefore computations were carried out on the SSR data only (Fig. 5). Assignment patterns from SSR data were generated by three runs each from $K = 1$ to $K = 15$ and analysed to determine cluster relationships. Populations and geographical groupings were largely reflected in the resulting images, with the following exceptions and specificities. Unambiguous assignments were made at $K = 2$ and $K = 3$, resulting in a 'northern' grouping with populations from Peru (Lago Aricota and Palca), and a 'southern' group of populations from Chile, the latter being subdivided at $K = 3$ into groups representing the Putre and Camarones/Camiña accessions. The northern group (Lago Aricota and Palca) retained its integrity up to $K = 12$. The two accessions from Lago Aricota (LA4018 and LA2387), which are the northernmost for *S. lycopersicoides*, consistently formed one group up to $K = 15$, demonstrating distinctness from the other populations. The ΔK method identified three 'true' populations ($K = 3$).

For *S. sitiens*, the STRUCTURE analysis at $K = 2$ revealed a 'northern' group (Paqui, Carbonatera and Cere) and a 'central' group composed of the remaining populations (Fig. 6). The southernmost Escondida population emerged as distinct from $K = 3$. Two of the three populations in the northern grouping (Paqui and Carbonatera) 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. The single, isolated southernmost population at Escondida showed an allele frequency pattern that set it apart from the remainder of populations. The distribution of $\Pr(X|K)$ based on SSR data indicated the presence of seven distinct groups; the estimator

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

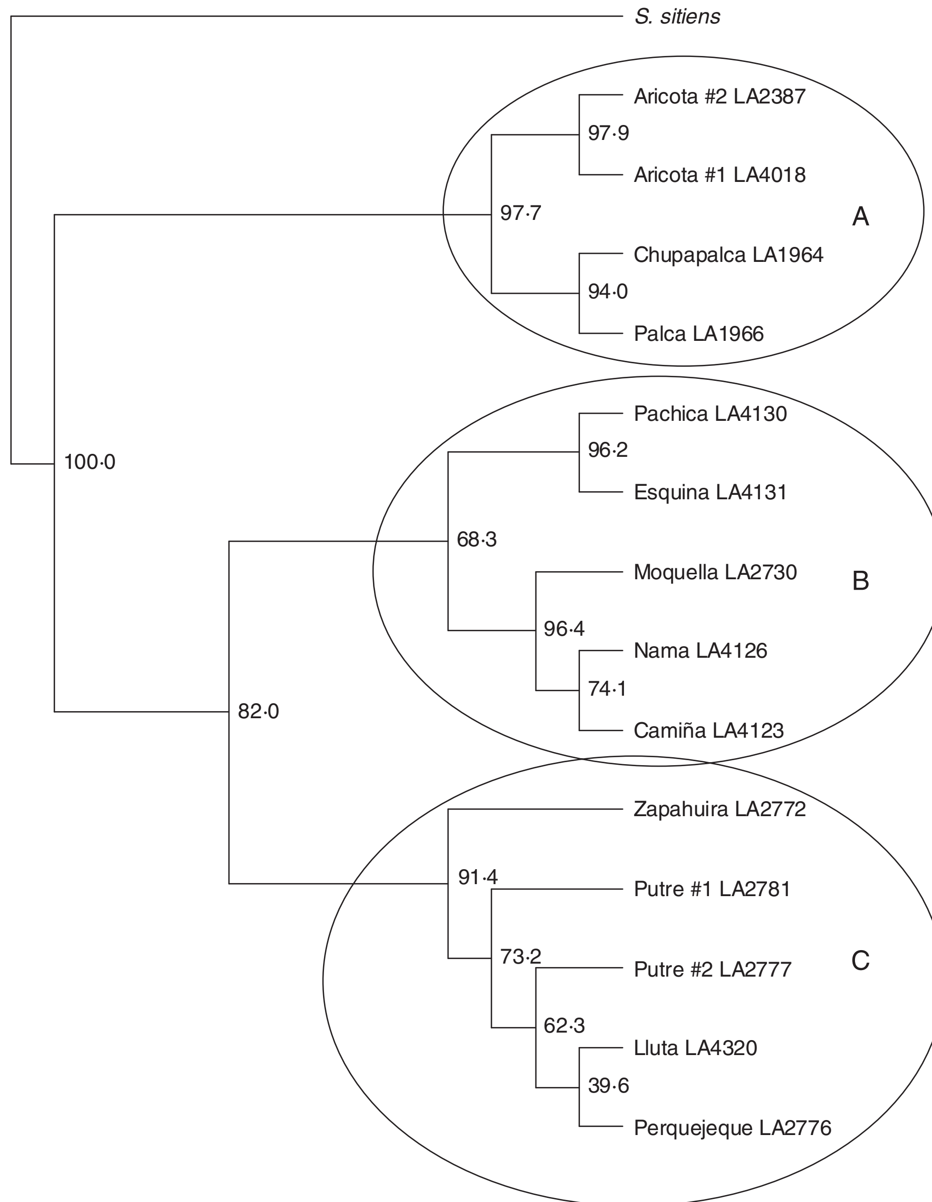

FIG. 3. Microsatellite phylogeny of *S. lycopersicoides* accessions based on Cavalli-Sforza's chord distance. Numbers at nodes represent the frequency (expressed as a percentage) at which a node occurred among 10 000 replications. Geographical groups are circled. A = northern, B = central and C = southern groups.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

ΔK returned $K = 3$ as the number of real populations. Allozyme simulations were in good agreement with those from SSR data with minor exceptions (data not shown).

Genetic diversity within 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. In *S. lycopersicoides*, a greater amount of variation at SSR loci was observed among groups than among populations within groups (18 vs. 13 %) as expected for 'true' population clusters. The corresponding fixation indices were highly significant ($P < 0{\cdot}001$). When considering three groups vs. one total group, a lower portion of the genetic variation was found to be partitioned among populations (16, 16 and 16 % vs. 27 %). Similar trends were seen in *S. sitiens*: 19 % of the SSR variation and 25 % of allozyme variation was present among groups, versus only 7 and 10 %, respectively, among populations within groups, and the fixation indices were highly significant ($P < 0{\cdot}001$). The fraction of genetic variation among populations within groups was substantially lower than among populations of the species as a whole: 7 % in the north and 9 % in the central vs. 22 % in the total species for SSRs (allozyme data were similar). Variability within the southern group could not be computed as it was composed of just one accession.

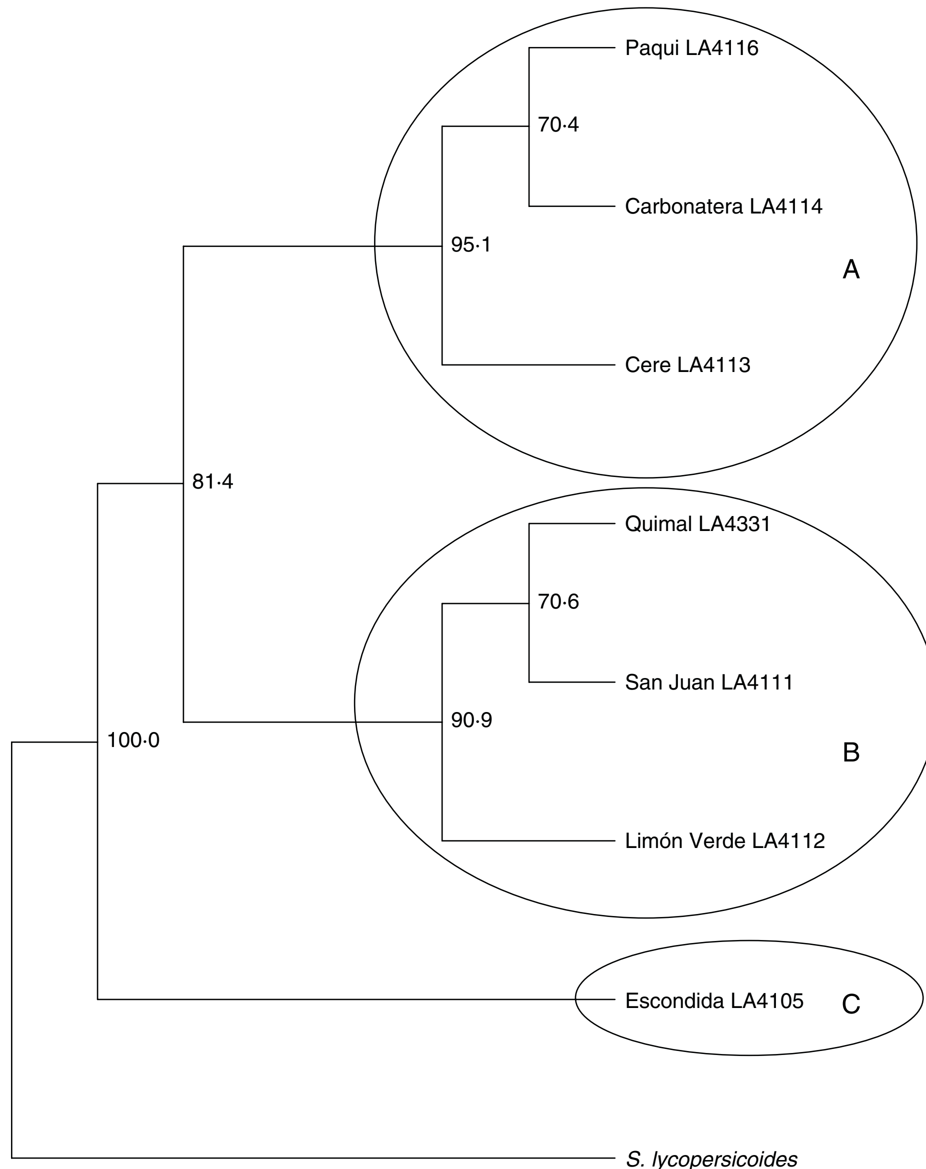

FIG. 4. Microsatellite phylogeny of *S. sitiens* accessions based on Nei's genetic distance. Numbers at nodes represent the frequency (expressed as a percentage) at which a node occurred among 10 000 replications. Geographical groups are circled. A = northern, B = central and C = southern groups.

Levels of genetic diversity were compared across the population clusters in each species. For *S. lycopersicoides*, the differences between the geographical groups were small at SSR loci (Fig. 7). Levels of inbreeding were highest in the northern group. Average pair-wise genetic distances within groups (θ) was lowest in the northern group and increased towards the south. The relatively high θ value in the south suggested a substantial amount of structuring within that group relative to the others, consistent with the relatively wide geographical separation between the two drainages that comprise this group. Considered separately, the pair-wise distance estimates fell below those of the two other regions ($\theta = 0{\cdot}104$ at Camarones and $\theta = 0{\cdot}091$ at Camiña). Finally, the average global pair-wise genetic distance (θ) was calculated for each group (i.e. the average genetic distance of the populations within a group to all other populations), but the differences were not significant. For *S. sitiens*, overall diversity was highest in the central group of accessions (Fig. 8). The polymorphism rate, allele richness and gene diversity of SSRs were all highest in the centre of the distribution ($P = 90{\cdot}5\,\%$, $R_S = 2{\cdot}72$, $H_E = 0{\cdot}391$). The highest number of private alleles (three) and the highest inbreeding coefficients were seen in the southern region, represented by the single Escondida population.

DISCUSSION

Contrasting patterns of allozyme and microsatellite variation

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 al.*,

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

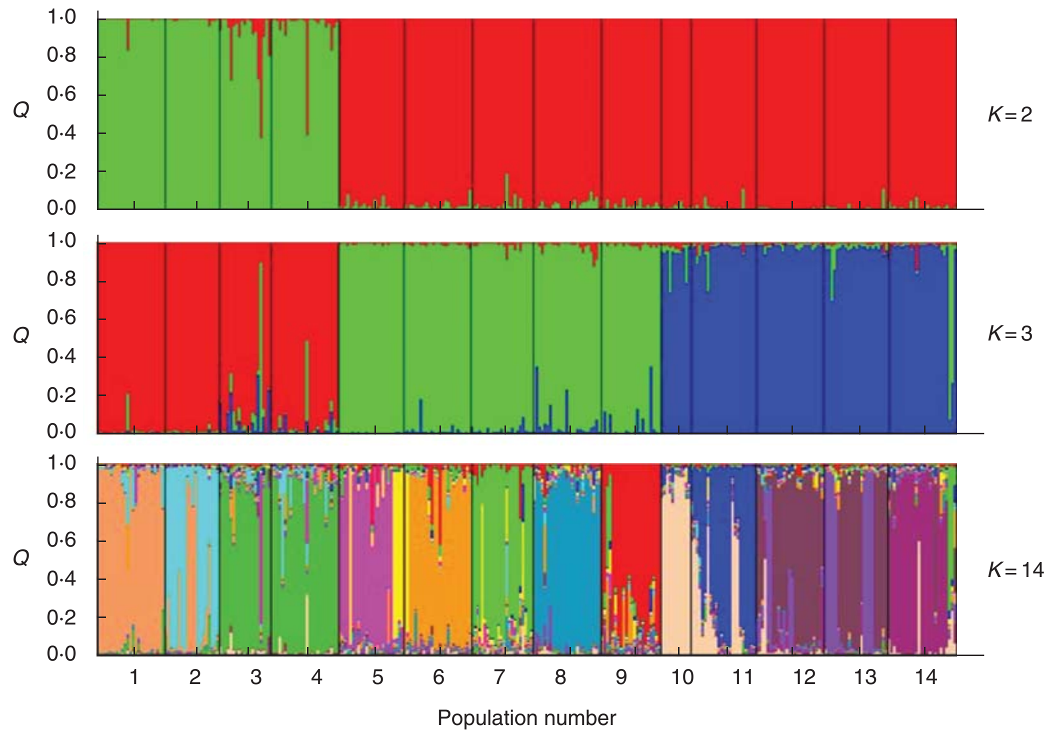

FIG. 5. STRUCTURE analysis of *S. lycopersicoides* populations based on SSR data. Bars represent the membership coefficients (Q) of individual plants (298 total in 14 populations), based on SSR allele frequencies, using K values for two, three or 14 groups. Numbers on the horizontal axes correspond to the population numbers in Table 1.

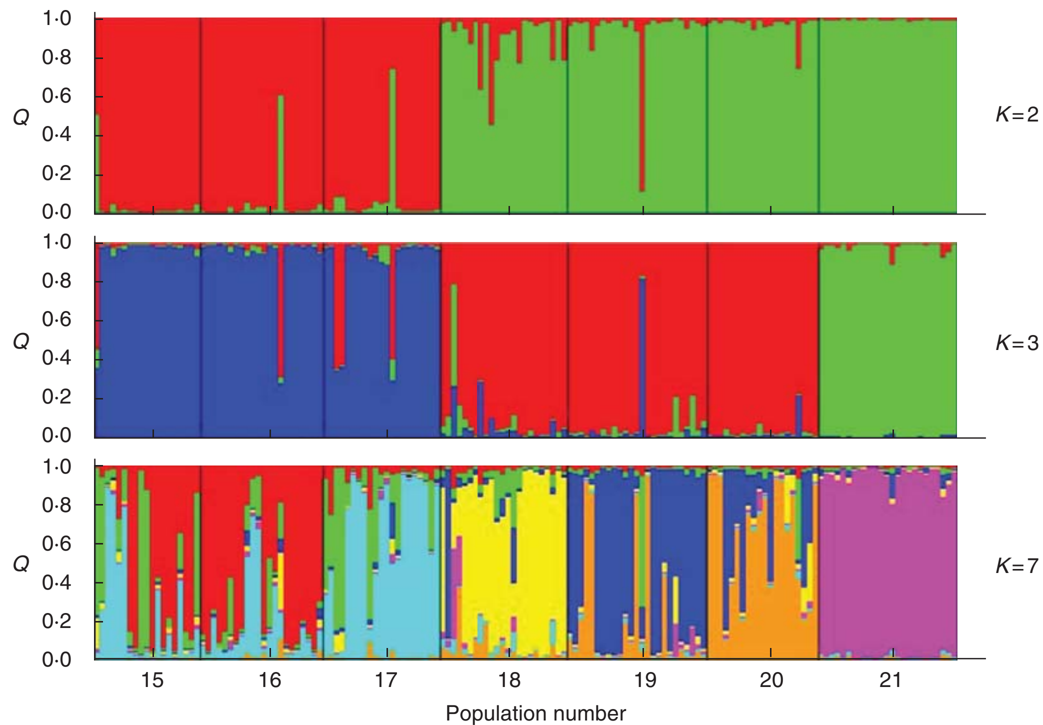

FIG. 6. STRUCTURE analysis of *S. sitiens* populations based on SSR data. Bars represent the membership coefficients (Q) of individual plants (155 total in seven populations), based on SSR allele frequencies, using K values for two, three or seven groups. Numbers on the horizontal axes correspond to the population numbers in Table 1.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

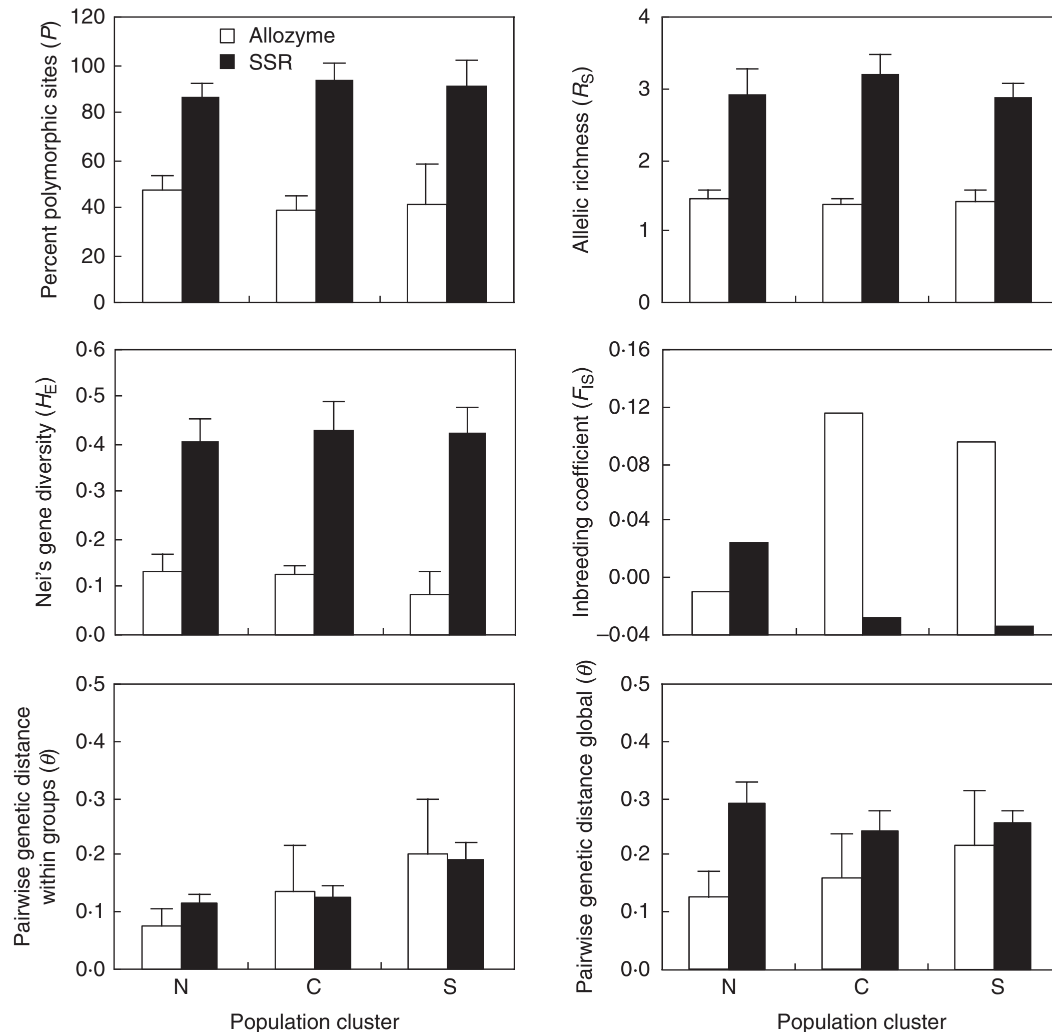

FIG. 7. Levels of genetic diversity within population clusters of *S. lycopersicoides*. Values represent the averages (+s.d.) of populations within each group, based on allozymes (stippled bars) or SSRs (striped). N = northern, C = central and S = southern cluster.

2004), and consistent with the mutation rates, which are several orders of magnitude higher for microsatellites (approx. 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 marker systems was most pronounced in *S. lycopersicoides*. Estimates of descriptive diversity parameters at allozyme loci were typically two to three times lower than those at SSR loci in this species. In *S. sitiens*, diversity estimates for allozymes were higher, nearly as high as for SSRs. As expected, inbreeding levels were higher at allozyme loci than at SSRs, the difference being greatest in *S. lycopersicoides.*

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 behaviour was evident in *S. lycopersicoides* and *S. sitiens*. However, geographical patterns of diversity and diversity structure among rice populations were in good agreement between the two marker types (albeit with a greater resolution at SSR loci) (Gao *et al.*, 2002). In contrast, the two marker systems showed little, if any, congruence in *S. lycopersicoides.* Particularly striking was that no population structure could be identified with allozymes in *S. lycopersicoides*, whereas strong geographical structure was detectable with SSRs. Furthermore, the allozyme phylogeny for *S. lycopersicoides* was poorly supported and corresponded neither to geographical relationships nor to the clusters suggested by the SSR dataset. The simplest explanation is the lower statistical power of allozymes in the present analysis: fewer loci in combination with lower levels of polymorphism reduced the amount of informative sites to one-third in *S. lycopersicoides* and to two-thirds in *S. sitiens* relative to that of the SSR markers used.

In spite of these inconsistencies, matrices of pair-wise F_{ST} values of the two marker systems were significantly correlated in both species, although more weakly in *S. lycopersicoides* than in *S. sitiens*. Graham (2005) found a strong correlation between genetic distances estimated from allozymes and SSRs in *S. chilense*. The cause for the discrepancy in *S. lycopersicoides* is unknown, but probably reflects the lower information content of allozymes in this study. Alternatively, purifying selection might be operating at some

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

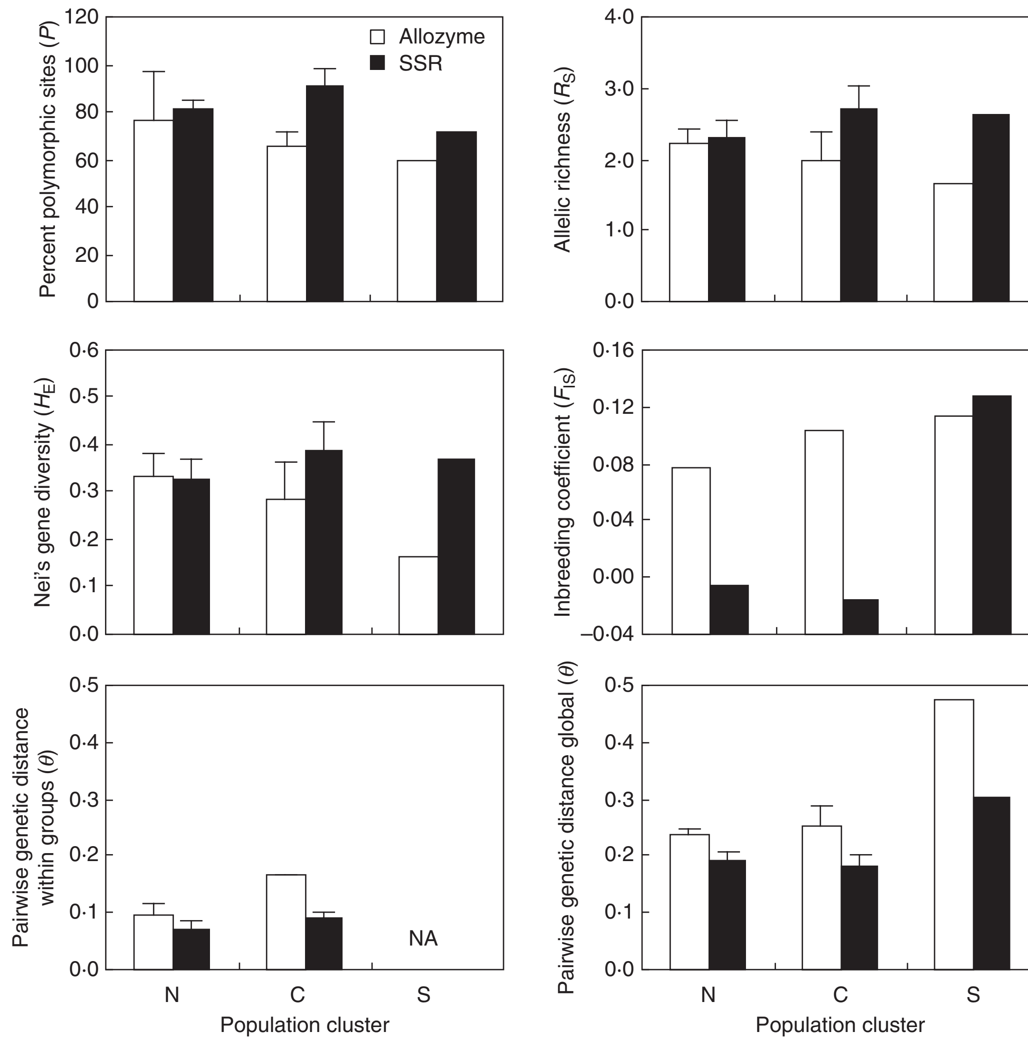

FIG. 8. Levels of genetic diversity within population clusters of *S. sitiens*. Values represent the averages (+s.d.) of populations within each group, based on allozymes (stippled bars) or SSRs (striped). N = northern, C = central and S = southern cluster.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

of the allozyme loci in *S. lycopersicoides*. Non-neutral behaviour of allozymes has been reported many times (e.g. Rick *et al*., 1977; Karl and Avise 1992; Dhuyvetter *et al*., 2004). Although several polymorphic allozyme loci were excluded from the analysis in *S. lycopersicoides* because of significant heterozygote deficiencies, the same was true of *S. sitiens*. Effects of selection therefore seem an unlikely explanation for the results.

In both species, the two marker systems identified different populations as departing from HW equilibrium or as 'bottlenecked', complicating general conclusions. With respect to HW 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 a deficiency of rare alleles. The 'allele deficiency' is dependent on (1) the elapsed time following the bottleneck, (2) the mutation rate and (3) the sample size in terms of genes (Maruyama and Fuerst, 1985). Factor (1) is equal for both datasets while (2) and (3) are elevated for microsatellites. High mutation rates erase the footprint of a bottleneck faster, and thus allozymes should be able to reach further in the past than SSRs to detect bottleneck events (Cornuet and Luikart, 1996). Conversely, microsatellites may be more useful for uncovering recent bottleneck events due to their greater sensitivity (Cornuet and Luikart, 1996). Hence, in the present study the two marker systems not only varied quantitatively, i.e. in their level of sensitivity, but also qualitatively, with regard to the diversity picture they presented. The disagreement was foremost species-specific, and not locus-specific, which could be the result of differences in demographic histories and/or life-history traits between the two species. 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 report neutral genetic diversity correctly in *S. lycopersicoides*.

Therefore, conclusions based on microsatellite variation are judged to be more robust.

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* according to microsatellite loci. Levels of inbreeding were near zero in both species, although *S. sitiens* populations were more inbred on average than *S. lycopersicoides* populations. The result may stem from the slightly smaller census sizes in terms of both number of populations and number of individuals. The extent of population fragmentation appears more severe in *S. sitiens*, which also may contribute to inbreeding. *Solanum sitiens* populations tend to be isolated from other populations by areas totally lacking in plants, due to the extreme aridity of the Atacama Desert. Pollinating insects were not observed on *S. sitiens* flowers during several visits by plant collectors, suggesting they might be relatively scarce, thus restricting gene flow between populations (Chetelat *et al.*, 2009). Under these circumstances, levels of inbreeding are likely to rise, leading to a loss of diversity. This interpretation is consistent with the higher level of divergence between populations seen in *S. sitiens* compared with *S. lycopersicoides.* There was indication (based on k, R_S and H_E, but not P) that both species harbour substantially less genetic variability than the related species *S. chilense*, which is sympatric with *S. lycopersicoides* and grows near 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 to 25 °S) and elevation (sea level to $>$3500 m), and hence presumably possesses adaptation to a wider range of environments. Population sizes for *S. chilense* also tend to be larger than for either *S. sitiens* or *S. lycopersicoides.* This wider geographical distribution is accompanied by a greater demographic representation with respect to both numbers of populations and average population sizes (http://tgrc.ucdavis.edu). The true discrepancy in diversity between the species may be greater than reported herein. The diversity estimates for *S. chilense* might be biased downwards because only one population was tested, and loci were selected according to polymorphism content in *S. lycopersicoides* and *S. sitiens*. However, a recent study (Graham, 2005) reported similar or slightly lower overall estimates of genetic diversity in a broad survey of *S. chilense* populations. The estimates for *S. chilense* accession LA2773 were slightly below the species average in Graham's study, suggesting that the relative genetic depletion of *S. lycopersicoides* and *S. sitiens* may be even more significant than the present results indicate.

Other estimates for gene diversity in *S. chilense* (Alvarez *et al.*, 2001) are in agreement with the present data. The gene diversity of *S. lycopersicoides* reported herein was lower than their estimates for *S. peruvianum* and *S. chilense*, but higher than their estimate for *S. pennellii*. The estimate here of average gene diversity for *S. sitiens* was similar to that of accessions of *S. arcanum* (formerly a subspecies of *Lycopersicon peruvianum* from northern Peru), but higher than reported for *S. pimpinellifolium*, *S. habrochaites* or *S. pennellii* (Alvarez *et al.*, 2001). Substantially lower levels of diversity were reported in populations of *S. cheesmanii*, *S. pimpinellifolium* and *S. galapagense* (Nuez *et al.*, 2004), which is not surprising given that all three are self-compatible and inbreeding, unlike the species studied herein which are all self-incompatible. Several independent diversity studies using different genotyping technologies all found *S. peruvianum* to be the most diverse tomato species, followed by *S. chilense*, while the self-compatible species were depleted of genetic variation (Miller and Tanksley, 1990; Breto *et al.*, 1993; Egashira *et al.*, 2000; Baudry *et al.*, 2001; Nuez *et al.*, 2004; Nakazato *et al.*, 2008). *S. pennellii* is mostly self-incompatible, widespread and harbours more genetic diversity than other outcrossing tomato species, such as *S. habrochaites* and *S. pimpinellifolium* (Rick and Tanksley, 1981). These results imply that *S. lycopersicoides* and *S. sitiens* are more diverse than some wild tomatoes, despite having more limited geographical ranges.

Genetic diversity among populations and geographical trends

Solanum lycopersicoides populations are found in five major geographical areas or river drainages: from north to south, Lago Aricota, Palca, Putre, Camarones and Camiña. There was no evidence for a clear geographical cline in diversity levels. The most diverse populations were spread across the north–south geographical distribution. However, geographical clines were evident for individual allele frequencies, of which mainly P and θ were affected. These two were significantly higher at more southern latitudes. In the absence of selection, geographical clines portray the combined effects of mutation and drift. The observed diversity pattern does not provide clear support for a south/north or east/west expansion. However, the high incidence of allele frequencies that exhibited trends with latitude (one-third of the SSR alleles) or longitude (one-quarter) is consistent with migration and/or progressive fragmentation along these two geographical axes. Geographical structuring was less clear cut at allozyme loci, possibly reflecting high levels of ancient genetic variation at the protein level that pre-date population expansion and/or fragmentation. Populations of *S. pimpinellifolium* were shown to contain nuclear intron divergence in discordance with geographical patterns, presumably as relicts of a pre-colonization phase (Caicedo and Schaal, 2004). Yet, significant IBD was detected in *S. pimpinellifolium* using AFLPs, and weak IBD was detected in *S. lycopersicum* var. *cerasiforme* populations in certain areas (Nakazato *et al.*, 2008). Compared with these two tomato species, the correlation between genotypes and geography was almost twice as high in *S. lycopersicoides* and *S. sitiens*. Populations of *S. lycopersicoides* typically grow within river valleys on the west side of the Andes, with relatively arid mountain ridges separating adjacent valleys. Although little is known about the actual pollen or seed dispersal mechanisms, it is likely that these ridges pose a substantial barrier to gene flow between 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 Mya, a time when the genus *Solanum* presumably diverged from the most recent ancestor (Wikstrom *et al.*, 2001). The Andes in that part of South America were half their present height at 10·4 Mya (Gregory-Wodzicki, 2000). It is therefore

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

conceivable that dispersal across mountain ridges was less impeded in the early days of *S. lycopersicoides*, which might explain the signature at the protein level. Given the preference of this species for relatively high elevations, it seems likely that elevation played a role in shaping genetic diversity in *S. lycopersicoides*. Although a few allele frequencies showed correlations with altitude, surprisingly, overall correlations with pair-wise genetic distances (θ) were not significant. Altitudinal distances between populations within drainages are often large, yet populations are likely to retain some connectivity through seed or pollen dispersal.

Levels of inbreeding were generally low in *S. lycopersicoides*. The most inbred populations were medium-sized ones at the time of sampling that might have undergone bottleneck events in the recent past. At allozyme loci, levels of inbreeding were highest in Putre #1, surprising because this is a very large population, though at the highest distribution point of all tomato populations (nearly 3800 m). The result may be indicative of selection pressures operating on allozyme genes, or may reflect an ancient founder event (in line with the marginal location) that is no longer evident at SSR loci. Only one *S. lycopersicoides* population (Aricota #2, at the northernmost location) showed a clear signal of a recent bottleneck event. Levels of genetic diversity were slightly below average in that population.

Populations of *S. sitiens* are relatively few and scattered, and most geographical locations were represented by a single collection. Levels of genetic variability in populations of *S. sitiens* did not show clear geographical trends. The most diverse populations were Limon Verde and Paqui in the central and northern parts of the distribution; the least diverse was Carbonatera, also from the north. However, the present investigation was limited by the small number of available populations (seven), 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 geographical 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, and correlations were often stronger than seen in *S. lycopersicoides*. Values of P increased towards the east, with elevation, and with geographical distance between populations. Genetic distances (θ) showed spatial patterns in both datasets, increasing towards the south, west and with geographical isolation, in congruence with the finding of significant isolation by distance.

IBD was more pronounced in *S. sitiens* than in *S. lycopersicoides*, possibly because genetic drift is stronger in the smaller *S. sitiens* populations and/or because gene flow is restricted by intervening stretches of desert that separate populations. Mechanisms and extent of pollen and seed dispersal are unknown (Chetelat *et al.*, 2009). Buzz pollination by bees as in the tomatoes seems the most likely mode but has not yet been confirmed due to the absence of pollinating insects at the time of collection. Striking differences in anther morphology between these two species and members of the tomato clade, including poricidal dehiscence, lack of a sterile tip, lack of fusion between anthers and white not yellow colour, suggest they may be less specialized for bee pollination. It seems likely that the hyper-arid conditions and lack of vegetation (food sources) in this region would limit the number of animal species capable of providing pollen and seed dispersal services for *S. sitiens*.

The proportion of allele frequencies in *S. sitiens* that were correlated with geographical features was only half as large at SSR loci, but twice as large at allozyme loci, compared with *S. lycopersicoides*. This emphasizes that the signature of successive colonization events is more pronounced in *S. lycopersicoides*, or, conversely, 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, 20 % of the allozyme alleles, showed correlations with altitude, although the altitudinal range was less than one-quarter (<500 m) of that shown by *S. lycopersicoides* populations. In contrast, not a single SSR allele was correlated with elevation. Census population size was positively correlated with diversity measures P and θ. The Cere population in the north and the remote, southernmost population Escondida deviated from HW equilibrium with a heterozygote excess and deficiency, respectively. Both were relatively small populations (approx. 20 individuals at the time of collection), and hence they may have been pushed into disequilibrium via genetic drift. Both carried signatures of recent genetic bottlenecks at the SSR level and were the most inbred populations at the protein level. In addition, Escondida was the only population to show substantial levels of inbreeding at SSR loci. Plants of this population grown *ex situ* failed to set spontaneous fruit without manual cross-pollination, and thus are presumed to be self-incompatible, like the other populations. However, self-incompatibility would not prevent all matings between related individuals, and thus inbreeding could occur, for example if plant numbers are reduced or if some plants are flowering when pollinators visit.

Population genetic structure

As expected for outcrossing species, the majority (up to 88 %) of genetic diversity in *S. lycopersicoides* and *S. sitiens* was partitioned within populations. In *S. chilense*, another outcrossing species, only half of the total variation was present within populations (Graham, 2005). The geographical range of this species is much broader than that of *S. lycopersicoides* or *S. sitiens*, and structuring is evident among regional groups, each of which is comparable in size to the entire distribution of *S. lycopersicoides* and *S. sitiens*. Within these regions, however, the degree of variability among populations was similar to that observed in *S. lycopersicoides* and *S. sitiens*.

The results also indicated that gene flow among populations was limited in both species. Estimated numbers of migrants per generation were mostly below one. If fewer than one migrant per generation is exchanged, populations are expected to diverge over time by means of drift (Allendorf, 1983), and may eventually develop into new species. The process will be slower in large populations because of the lesser impact of drift. Thus, the present results suggest populations of both species are diverging, and that those of *S. sitiens* are doing so faster. However, as the amount of gene flow appeared to be principally governed by geographical distance and geographical barriers, it is expected to underlie local variation. For example, in *S. lycopersicoides* gene flow within drainages is expected to be far greater than gene flow between drainages.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

Although gene flow is difficult to measure directly (Avise, 2004), these spatial patterns were also supported by observations here of highly significant IBD.

Population clusters revealed by phylogenetic, principal component and STRUCTURE analyses agreed, and corresponded to predicted geographical groups, for the most part. Populations of both species segregated into three major clusters along a north–south axis. The number of major and minor clusters according to SSR data was confirmed via probabilistic derivations. 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 drainages). External branches clustered according to drainages, with only a few exceptions. The genetic similarity of the Perquejeque and Lluta populations, but not Putre, was surprising given the relative distances between the three. One explanation may be that the low-elevation Lluta population has greater gene flow than Perquejeque, possibly through seed dispersal downriver.

A primarily latitudinal structuring was in line with the high incidence of allele frequencies showing a geographical 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, lower elevations toward the Pacific are associated with environmental conditions inadequate for this species, presumably via climatic effects, as rainfall declines exponentially with decreasing elevation (Houston, 2006).

Almost one-fifth of the microsatellite variation in *S. lycopersicoides* was present among the three clusters, only slightly more than among populations within clusters. 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 this, the overall level of genetic diversity was higher in the central group. The northern group was fairly inbred, while an increasing excess of outcrossing was evident towards the south. The central region also appeared to be the richest in terms of populations per spatial unit, and census sizes tended to be larger than those in the other areas. Finally, the central populations showed the greatest altitudinal range in this region.

Geographical diversity may foster unique adaptations. Indeed, features of 'internal endemism' were evident among populations. The most genetically 'distinct' populations were Zapahuira and Putre #1, the latter located at the highest elevation. They appeared to have been somewhat isolated for a longer period as they showed only minimal signs of shared ancestry with other populations. The average pair-wise genetic distance within regions confirmed the homogeneity observed in STRUCTURE analysis among populations in the north but a considerable amount of substructure in the south, 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 Putre and Rio Lluta. Although speculative, the data point to this area as the centre of origin, from which colonization could have occurred northward and southward.

In *S. sitiens* three major groups (northern, central and southern) were also evident from the marker data. Cluster analysis revealed a principal divide between the southernmost location (Escondida) and all the other populations. This was confirmed by STRUCTURE analysis at allozyme loci, whereas, at SSRs, surprisingly, the 'northern' area was identified as separate from the other two regions, possibly reflecting the weaker geographical signal at the microsatellite level. Structuring among *S. sitiens* populations occurred primarily across a latitudinal scale, in line with the finding that most allelic trends were observed along a north/south axis, presumably influenced by topographical, climatic and/or biological factors.

Several aspects of the geographical distribution of *S. sitiens* suggest a link with local precipitation patterns. The species is found within a relatively narrow altitudinal range (Peralta *et al.*, 2008; Chetelat *et al.*, 2009). *S. sitiens* is not found east of the Salar de Atacama, or on the slopes of the main Andean cordillera where rainfall is more abundant, or west of the Cordillera Domeyko and Cerros de Paqui mountain ranges, where rainfall drops with elevation to near zero along the coast. Interestingly, the 'northern' cluster of populations are all located on E- or SE-facing slopes, whereas those of the 'central' and 'southern' groups are all found on W- or NW-facing slopes.

Population groupings were slightly more pronounced with allozymes than with SSR markers, but both indicated that a much larger share of total diversity is partitioned among population groups than among populations within groups. The hierarchy among populations in the north showed a close relationship between the Carbonatera, Cere and Paqui populations, which are all located within a short distance of one another. The higher degree of homogeneity within the northern group relative to other regions was confirmed by a lower average pair-wise genetic distance. These trends are probably explained by the geographical separation of the northern populations, which are found in the Cerros de Paqui range, and populations to the south, located in the Cordillera Domeyko, with the Rio Loa valley between them.

The southernmost population, Escondida, was clearly the most isolated genetically and physically, and exhibited the lowest incidence of admixture. Separated by a distance of approx. 120 km from its nearest neighbouring population (San Juan), it has presumably undergone little genetic exchange with other populations for many generations. Its relatively small population size (20 or so individuals in 2001) is likely to result in further genetic divergence through drift. Escondida was the only *S. sitiens* population that showed substantial levels of inbreeding. Thus, the relatively depleted genetic diversity of that population makes sense.

The small number of *S. sitiens* populations and their observed patterns of genetic diversity made it difficult to pinpoint one area as the centre of diversity for this species. Genetic variation, especially at allozymes, declined progressively in the three southern populations. The southern elements of this species may therefore have originated via founder events from the central/northern part of the present distribution. Geographical trends were weaker at microsatellites,

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

but the greatest genetic variation was also displayed in the central region (Limon Verde). Colonizations may therefore have taken place further back in the past, so that the signature at SSR loci has been eroded by mutations. However, in view of the current demographic picture and continuing population decline (see below), it seems likely that fragmentation has played a major role in forming this species' genetic structure, confounding historical events.

Implications for conservation strategies

Recent field observations suggest that both *S. lycopersicoides* and *S. sitiens* are threatened or endangered by a variety of factors, making conservation efforts a high priority (Chetelat *et al.*, 2009). Preservation strategies should be aimed at capturing the largest amount of genetic diversity at the lowest possible cost, and at preserving unique ecotypes or subspecies. By providing insights into the partitioning of genetic diversity, the present study may help to meet these goals.

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. lycopersicoides* and *S. sitiens*, but the strong pattern of isolation by distance in combination with the small number of populations would justify broad conservation efforts aimed at preserving as many populations as possible. Compared with more abundant and widely distributed tomato species (e.g. *S. chilense*) a relatively large share of genetic diversity is found within a small number of individuals. Hence, further reductions in population sizes would lead to an accelerated genetic erosion in the two nightshade species.

In *S. lycopersicoides*, sampling should include each drainage, with a focus on the central area around Putre. The more homogeneous genetic makeup of the populations at the northern distribution edge would allow the maintenance of fewer accessions from that region. The most diverse populations, Zapahuira and Lluta, can be considered good representatives for the central region, Palca for the northern region and Moquella for the southern region. Considering their limited numbers and fragmented distribution, most populations of *S. sitiens* analysed in the present study should form part of a conservation program. In the northern region, priority should be given to the most diverse population, Paqui.

In addition, conservation programmes should give special attention to populations with idiosyncratic features. These may be spotted as geographical outliers, by a large genetic distance to other populations or a high occurrence of private alleles. Of greatest interest in *S. lycopersicoides* is Putre #1, located at the highest elevation of any tomato relative and displaying great genetic distance. Aricota #1 also shows great genetic distance from the others and has unique morphological features (Chetelat *et al.*, 2009). Zapahuira and Pachica contain several private alleles. Among the *S. sitiens* populations, Escondida is unique and isolated from the other populations, and harbours the highest number of private alleles. Due to its small population size and proximity to two large surface mining operations (Mina Escondida and Mina Zaldivar), it is considered particularly vulnerable (Chetelat *et al.*, 2009). Finally, Limon Verde is genetically distant from the others and rich in private alleles.

Substantial *ex situ* conservation efforts have been undertaken over the past decades. Samples of all known *S. lycopersicoides* and *S. sitiens* populations, collected during five trips in the native region or obtained by donations, are maintained by the C.M. Rick Tomato Genetics Resource Center (TGRC) at Davis. Future efforts should therefore be directed at *in situ* conservation, such as the establishment of reserves and provision of ecological buffers.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following figures and tables. Figure S1: Principal components analysis of genetic relationships in *S. lycopersicoides*. Figure S2: Principal components analysis of genetic relationships in *S. sitiens*. Table S1: Genetic diversity in populations of *S. lycopersicoides*. Table S2: Genetic diversity in populations of *S. sitiens*. Table S3: Analysis of molecular variance (AMOVA) and global levels of inbreeding in *S. lycopersicoides*. Table S4: Analysis of molecular variance (AMOVA) and global levels of inbreeding in *S. sitiens*.

ACKNOWLEDGEMENTS

We gratefully acknowledge the following individuals who participated in collecting the wild *Solanum* accessions used in this study: Charles Rick, Miguel Holle, Andres Contreras, Rudolf Thomann, Ricardo Pertuzé, Luis Faúndez, Pedro Leon, Elaine Graham, Carl Jones, David Spooner, Carlos Ochoa and Alberto Salas. We also thank Katie Smith and the crew of the C.M. Rick Tomato Genetics Resource Center (TGRC) for maintaining plants, and Rebecca Benitez for her valuable assistance in the laboratory and greenhouse.

LITERATURE CITED

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

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**: 1283–1292.

Avise JC. 2004. *Molecular markers, natural history, and evolution*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

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

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

Bindler G, van der Hoeven R, Gunduz I, *et al*. 2007. A microsatellite marker based linkage map of tobacco. *Theoretical and Applied Genetics* **114**: 341–349.

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

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

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

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

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

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

Chetelat RT, Pertuzé RA, Faúndez L, Graham EB, Jones CM. 2009. Distribution, ecology and reproductive biology of wild tomatoes and related nightshades from the Atacama Desert region of northern Chile. *Euphytica* **167**: 77–93.

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.

Davis J, Yu D, Evans W, Gokirmak T, Chetelat RT, Stotz H. 2009. Mapping of loci from *Solanum lycopersicoides* conferring resistance or susceptibility to *Botrytis cinerea* in tomato. *Theoretical and Applied Genetics* **119**: 305–314.

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**: 1065–1074.

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**: 23–31.

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.

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.

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

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**: 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.

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**: 574–578.

Felsenstein J. 2005. PHYLIP (phylogeny inference package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

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, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter* **13**: 207–209.

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**: 173–180.

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.

Goudet J. 2001. *FSTAT a program to estimate and test gene diversities and fixation indices* (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

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

Graham EB. 2005. *Genetic diversity and crossing relationships of Lycopersicon chilense*. PhD thesis, University of California, Davis, USA.

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

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

Haldane JBS. 1954. An exact test for randomness of mating. *Journal of Genetics* **52**: 631–635.

Houston J. 2006. Variability of precipitation in the Atacama Desert: its causes and hydrological impact. *International Journal Climatology* **26**: 2181–2198.

Kahler AL, Allard RW, Miller RD. 1984. Mutation rates for enzyme and morphological loci in barley (*Hordeum vulgare*). *Genetics* **106**: 729–734.

Karl SA, Avise JC. 1992. Balancing selection at allozyme loci in oysters. Implications from nuclear RFLPs. *Science* **256**: 100–102.

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

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**: 2453–2465.

Loridon K, McPhee K, Morin J, *et al*. 2005. Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). *Theoretical and Applied Genetics* **111**: 1022–1031.

Luikart G. 1997. *Usefulness of molecular markers for detecting population bottlenecks and monitoring genetic change*. PhD thesis, University of Montana, Missoula, USA.

Mantel N. 1967. The detection of disease clustering and generalized regression approach. *Cancer Research* **27**: 209–220.

Maruyama T, Fuerst PA. 1985. Population bottlenecks and non equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**: 675–689.

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

Morgante M, Olivieri AM. 1993. PCR-amplified microsatellites as markers in plant genetics. *Plant Journal* **3**: 175–182.

Nakazato T, Bogonovich M, Moyle LC. 2008. Environmental factors predict adaptive phenotypic differentiation within and between two wild Andean tomatoes. *Evolution* **62**: 774–792.

Nei M. 1972. Genetic distance between populations. *American Naturalist* **106**: 283–292.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Nuez F, Prohens J, Blanca JM. 2004. Relationships, origin, and diversity of Galapagos tomatoes: implications for the conservation of natural populations. *American Journal of Botany* **91**: 86–99.

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.

Page RDM. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357–358.

Peralta IE, Spooner DM, Knapp S. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Systematic Botany Monographs* **84**: 1–186.

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

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**: 502–503.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

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. 1988. Tomato-like nightshades: affinities, autecology, and breeders opportunities. *Economic Botany* **42**: 145–154.

Rick CM, Chetelat RT. 1995. Utilization of related wild species for tomato improvement. *Acta Horticulturae* **412**: 21–38.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010

Rick CM, Fobes JF. 1975. Allozyme variation in the cultivated tomato and closely related species. *Torrey Botanical Club Bulletin* **102**: 376–384.

Rick CM, Holle M. 1990. Andean *Lycopersicon esculentum* var. *cerasiforme*. Genetic variation and its evolutionary significance. *Economic Botany* **44**: 69–78.

Rick CM, Tanksley SD. 1981. Genetic variation in *Solanum pennellii*. Comparisons with two other sympatric tomato species. *Plant Systematics and Evolution* **139**: 11–46.

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**: 55–68.

Rick CM, Fobes JF, Holle M. 1977. Genetic variation in *Lycopersicon pimpinellifolium*. Evidence of evolutionary change in mating systems. *Plant Systematics and Evolution* **127**: 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**: 279–298.

Roselius K, Stephan W, Stadler T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* **171**: 753–763.

Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.

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**: 341–349.

Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**: 627–632.

Sokal RR, Sneath PHA. 1963. *Principles of numerical taxonomy*. San Francisco: WH Freeman.

Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**: 1063–1066.

Tanksley SD, Ganal MW, Prince JP, *et al*. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**: 1141–1160.

Vigouroux Y, Jaqueth JS, Matsuoka Y, *et al*. 2002. Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution* **19**: 1251–1260.

Weber JL, Wong C. 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* **2**: 1123–1128.

Weir BS. 1990. *Genetic data analysis*. Sunderland, MA: Sinauer Publications.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wikstrom N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: Calibrating the family tree. *Proceedings of the Royal Society Biological Sciences Series B* **268**: 2211–2220.

Wright S. 1951. The genetical structure of populations. *Annals of Eugenics* **15**: 323–354.

Zamir D. 2001. Improving plant breeding with exotic genetic libraries. *Nature Reviews Genetics* **2**: 983–989.

Downloaded from http://aob.oxfordjournals.org at Universidad de Chile on June 10, 2010