North–South Colonization Associated with Local Adaptation of the Wild Tomato Species *Solanum chilense*

Katharina B. Böndel,*[1] Hilde Lainer,[1] Tetyana Nosenko,[1] Mamadou Mboup,[2] Aurélien Tellier,*[3] and Wolfgang Stephan[1]

[1]Section of Evolutionary Biology, Department of Biology II, Ludwig-Maximilians University of Munich, Planegg-Martinsried, Germany

[2]DuPont de Nemours, European Research & Development Center for Crop Protection, Nambsheim, France

[3]Section of Population Genetics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Freising, Germany

***Corresponding authors:** E-mail: boendel@bio.lmu.de; tellier@wzw.tum.de.

Associate editor: Hideki Innan

Abstract

After colonization population sizes may vary across the species range depending on environmental conditions and following colonizations. An interesting question is whether local adaptation occurs more frequently in large ancestral populations or in small derived populations. A higher number of new mutations and a lower effect of genetic drift should favor selection in large populations, whereas small derived populations may require an initial local adaptation event to facilitate the colonization of new habitats. Wild tomatoes are native to a broad range of different habitats characterized by variable abiotic conditions in South America, and represent an ideal system to study this interplay between demography and natural selection. Population genetic analyses and statistical inference of past demography were conducted on pooled-sequencing data from 30 genes (8,080 single nucleotide polymorphisms) from an extensive sampling of 23 *Solanum chilense* populations over Chile and Peru. We reveal first a north–south colonization associated with relaxed purifying selection in the south as shown by a decrease of genetic variation and an increasing proportion of nonsynonymous polymorphism from north to south, and population substructure with at least four genetic groups. Second, we uncover a dual picture of adaptation consisting of 1) a decreasing proportion of adaptive amino acid substitutions from north to south suggesting that adaptation is favored in large populations, whereas 2) signatures of local adaptation predominantly occur in the smaller populations from the marginal ranges in the south.

Key words: **demography, local adaptation, plant population genetics, positive selection, wild tomato.**

Article

Introduction

Demography and adaptation are important interacting factors that determine the evolution of plant species. Plant species are often substructured into populations or demes connected by variable migration (metapopulations; Husband and Barrett 1996). Following the colonization of new habitats, population sizes tend to decrease from the ancestral to the derived ranges of the species distribution. At the level of populations, local adaptation can best be studied for organisms with restricted migration (Lange et al. 1990). As plants are sessile organisms, they also have to cope with both biotic and abiotic stresses. Therefore, adaptations to local environmental conditions are essential to ensure survival of the species. An interesting unanswered question is under which conditions local adaptation occurs in such a metapopulation system and to what extent population size and migration strength influence local adaptation. In large populations more new mutations arise each generation and, as the effect of genetic drift is low, selection is expected to increase the frequency of advantageous mutations over time. In small populations, however, the effect of genetic drift is strong and can therefore counteract selection and eliminate possible advantageous mutations. This classic population genetics expectation suggests that local adaptation should occur more frequently in large than in small populations (e.g., Willi et al. 2006; Hough et al. 2013). A survey study of reciprocal transplant experiments revealed that local adaptation was indeed more common in large populations (Leimu and Fischer 2008). However, recently derived populations may also exhibit local genetic adaptation to novel conditions. This is puzzling because founding populations are usually small and derived from a larger ancestral population, thus their adaptive potential is a priori limited (Willi et al. 2006). In the new habitat, founding populations are confronted with different environmental conditions and therefore different selection pressures. The establishment of a derived founding population should thus require an initial local adaptation event to the new habitat in which selection acts on available standing genetic variation (Innan and Kim 2008). A key for local adaptation may thus be whether a new advantageous mutation arises in a derived population, or whether an existing, previously neutral mutation becomes advantageous due to new or changing environmental conditions. Furthermore, migration can not only accelerate the spread of beneficial mutations across populations but also prevent the fixation of beneficial mutations by introducing

other variants to the population (migration generates maladaptation; Kirkpatrick and Barton 1997).

Wild tomatoes (*Solanum* section *Lycopersicon*) are a good model system to study plant demography and local adaptation. They represent a young group of 13 species (Peralta et al. 2008), which diversified about 3 Ma (Särkinen et al. 2013). They are distributed from Ecuador to northern Chile with two endemic species on the Galápagos Islands (Moyle 2008; Peralta et al. 2008). Many wild tomato species show spatial structuring with populations scattered throughout their species distribution and exhibiting extinction and recolonization events (e.g., Peralta et al. 2008; Chetelat et al. 2009). The geographic distribution of the various species is determined by the abiotic and biotic conditions they encounter (Nakazato et al. 2010). Signatures of adaptation to abiotic and also biotic stresses were already detected in the two green-fruited outcrossing sister species *S. peruvianum* and *S. chilense* (Xia et al. 2010; Fischer et al. 2011; Hörger et al. 2012). Significant patterns of isolation by distance (IBD) were observed in some wild tomato species including *S. pimpinellifoium*, *S. lycopersicum*, and *S. peruvianum* (Nakazato and Housworth 2011; Nakazato et al. 2012) and in two related nightshade species, *S. lycopersicoides* and *S. sitiens*, which occur in sympatry with some wild tomato species in Chile (Chetelat et al. 2009; Albrecht et al. 2010). Furthermore, it was hypothesized that some of the wild tomato species, including *S. chilense*, might be derived from *S. peruvianum* (Baudry et al. 2001). This suggests that colonization of new habitats from an ancestral species is a likely scenario for several wild tomato species. Previous studies on genetic variation and demography revealed that the between-species nucleotide variation depends strongly on the type of mating system, and self-incompatible species exhibit significantly higher silent nucleotide diversity than self-compatible species (Baudry et al. 2001; Roselius et al. 2005). In addition, the outcrossing sister species *S. peruvianum* and *S. chilense* present very high levels of genetic diversity, which are explained by the existence of long-term seed banks (Tellier, Laurent, et al. 2011).

In this study, we focused on the wild tomato species *S. chilense*. This species is native to southern Peru and northern Chile. It exhibits a patchy population distribution across desert, high altitude and coastal regions (Peralta et al. 2008; Chetelat et al. 2009). Therefore, different populations encounter different abiotic stresses including drought, cold, salinity, and also combinations of these factors. In comparison to other wild tomato species, *S. chilense* can grow under the driest and coldest conditions (Moyle 2008). This makes *S. chilense* an ideal candidate species to study demography and the local adaptation associated with it. In fact, signatures of local adaptation possibly associated with either drought or cold stress were found in abiotic stress-related genes (Xia et al. 2010; Fischer et al. 2011; Mboup et al. 2012). Different levels of genetic variation were also observed between populations indicating different demographic histories of individual populations (Arunyawat et al. 2007).

We specifically addressed two objectives: 1) To infer the demographic history of the species, that is, the genetic relationships between populations and their migration or colonization pattern; and 2) to reveal where local adaptation occurs. To accomplish these goals, we sequenced multiple genes (>42 kb) from an extensive sampling of 23 populations using a pooling approach of 25 diploid plants per population. Pooling provides a cost-effective approach to study a high number of populations by neglecting individual information. This method was successfully applied to detect selection in animals and plants (e.g., Obbard et al. 2009; Turner et al. 2010). We applied a "relaxed" candidate gene approach, that is, like in the "classic" candidate gene approach our gene set comprises genes that may be under selection (the candidate genes) and genes that may evolve neutrally (the reference genes). To study past demography and spatial structuring, we used all genes together to approximate the genome average. When studying selection acting on the candidate genes, we also explored the possibility of selection to occur at reference genes. We used 1) standard population genetic methods including the proportion of adaptive amino acid substitutions and the distribution of fitness effects (DFE) to characterize the population nucleotide diversity, 2) maximum likelihood and distance-based methods to quantify the relationships between populations, and 3) outlier methods to identify single nucleotide polymorphisms (SNPs) under selection. Our analyses revealed four groups of populations and a strong north–south pattern, which is in accordance with colonization from north to south. Outlier methods identified SNPs that are highly differentiated over the species range. Five of these SNPs are in the transcription factors *AREB1* and *JERF3*, which are involved in the abiotic stress response. These genes are highly differentiated at the coast and in the southern mountain range indicating local adaptation to the southern environments.

Results

We sequenced 30 genes from 23 populations of the wild tomato species *S. chilense* using a pooling approach (25 plants per population). The 30 genes consist of genes involved in the abiotic stress response and presumably neutral reference genes (see supplementary materials and methods and tables S2 and S3, Supplementary Material online). On average, 10,082,038 (6,328,429–14,601,371) paired-end reads were obtained per population and 91.96% (81.40–97.41 %) were successfully mapped to reference sequences (see supplementary materials and methods and table S7, Supplementary Material online). Number of reads and number of mapped reads are highly correlated ($R^2 = 0.981$, Spearman's rank correlation $\rho = 0.983$, P value < 0.001). No correlations with nucleotide variation (as measured by θ_W and π; Watterson 1975; Nei and Li 1979) were found (number of reads with θ_W: $R^2 = 0.002$, $\rho = 0.025$, $P = 0.907$; number of reads with π: $R^2 = 0.014$, $\rho = 0.061$, $P = 0.777$; number of mapped reads with θ_W: $R^2 = 0.007$, $\rho = 0.063$, $P = 0.768$; number of mapped reads with π: $R^2 = 0.029$, $\rho = 0.106$, $P = 0.621$).

To evaluate the repeatability of our approach, we sequenced one population sample twice and found significant correlations between the repetitions for the population genetic statistics θ_W, π, and Tajima's D (θ_W: $R^2 = 0.786$, $\rho = 0.864$,

P value < 0.001; π: $R^2 = 0.982$, $\rho = 0.976$, P value < 0.001; Tajima's D: $R^2 = 0.804$, $\rho = 0.880$, P value < 0.001; number of segregating sites, S: $R^2 = 0.982$, $\rho = 0.982$, P value < 0.001).

The number of SNPs per population varied between 672 and 1,914 and per gene between 72 and 688 (supplementary table S8, Supplementary Material online), as a function of the gene length ($R^2 = 0.827$, $\rho = 0.891$, P value < 0.001).

Population Genetic Analyses Reveal a Clinal Pattern

We assessed the genetic variation of each population by calculating the Watterson estimator (θ_W) and the nucleotide diversity (π) for all sites and silent sites, averaging over two gene sets to disentangle selection from demographic effects: 1) The 14 reference genes and 2) all 30 genes (supplementary table S9, Supplementary Material online). The average population values of θ_W and π are in the ranges of previously reported values for *S. chilense* populations (Städler et al. 2008; Mboup et al. 2012). The overall pattern between the populations is similar for both gene sets and also for all sites and silent sites. Genetic variation differs greatly between the populations. The northern populations have three (θ_W) to five (π) times higher genetic variation than the populations in the south. This results in a significant decrease of genetic variation from north to south (fig. 1*A* and supplementary figs. S2 and S3, Supplementary Material online). The test statistic Tajima's D was also calculated for each population for both gene sets (supplementary table S9, Supplementary Material online). Again the populations differ greatly in their values. Tajima's D values range from above 0 for some populations from the central region to values below −1 in the southern range leading to a significant decrease from north to south (fig. 1*B* and supplementary fig. S4, Supplementary Material online). This cline in genetic variation and Tajima's D values indicates a north-to-south colonization scenario.

For both estimators of genetic variation and for Tajima's D, this pattern is observed for both gene sets and for all sites and silent sites, therefore giving comparable results. This implies that a moderate number of genes are sufficient to analyze the demographic history of a population or species even if some of the genes are possibly under selection, especially in *S. chilense* where nucleotide diversity is high, generating a large numbers of SNPs.

As the seeds were not collected in the wild, but from populations outbred for a few generations at the Tomato Genetics Resource Centre (supplementary table S1, Supplementary Material online), breeding procedures could influence the observed genetic variation. However, we do not find any correlation between genetic variation and the number of generations ex situ (θ_W: $R^2 = 0.016$, $\rho = 0.015$, $P = 0.9456$; π: $R^2 = 0.020$, $\rho = 0.231$, $P = 0.2892$).

Population Differentiation, IBD, and Population Clustering

To infer the genetic differentiation between each population pair, we calculated average F_{ST} values over the 14 reference genes and over all 30 genes for each population pair (supplementary table S10, Supplementary Material online). Overall F_{ST} values increase with increasing geographic distance leading to a significant pattern of IBD (fig. 1*D* and supplementary fig. S6A, Supplementary Material online). Interestingly, high F_{ST} values are observed not only between populations from north and south but also between populations from the coast in the south and populations from high altitudes in the south. This is remarkable as the geographic distance between the coast and the high altitudes in the south is approximately 200–300 km whereas the distance between the southern and northern populations is in the range of 800–1,200 km. The analysis of the genetic differentiation indicates that the 23 populations can be clustered into four groups with high between-group and low within-group F_{ST} values: A northern group (NG: LA1930, LA3784), located in the northern range of the species; a southern low altitude group (SLG: LA2750, LA2932, LA4108, LA4107), located in the southern coastal range; a southern high altitude group (SHG: LA4332, LA4118, LA4119, LA2880), located in the southern mountain range; and a central group (CG: All remaining populations; fig. 2 and supplementary table S11, Supplementary Material online). Although the pattern of IBD is observed for the whole sample, it is not observed for the CG populations. A trend for IBD, however, is observed within SLG and SHG (supplementary fig. S6, Supplementary Material online). The underlying structure of the data set is further characterized by a higher relative number of private than shared SNPs for the groups (supplementary fig. S7, Supplementary Material online). The relative number of private SNPs ranges from 53 in SHG to 373.5 in NG. The relative number of shared SNPs is lower ranging from 1.7 (shared between NG, SHG, and SLG) to 14.9 (shared between CG and SHG).

This population grouping is supported by the following additional methods that were applied to further investigate the genetic relationships between the *S. chilense* populations. TreeMix infers population splits and admixture events from SNP frequency data (Pickrell and Pritchard 2012). The TreeMix topology reflects the geographic distribution of the populations and the four groups: The CG populations split from NG, and SHG and SLG split from CG. Also the most northern populations of CG (LA0456, LA1958, LA3111) split first from NG. The TreeMix topology including a pooled population sample from *S. chilense*'s sister species *S. peruvianum* shows that the NG populations are the closest to *S. peruvianum* (supplementary fig. S8, Supplementary Material online). All gene and reference gene consensus phylogenies also imply that the NG populations are early derived populations, as they form a highly supported sister clade to all other *S. chilense* populations (fig. 2*C* and supplementary fig. S9, Supplementary Material online). SHG and SLG are highly supported clades nested within the populations of CG, but do not form sister clades as might be expected from the TreeMix topology. They also do not form sister clades in the genetic distance networks (supplementary fig. S10, Supplementary Material online).

According to the F_{ST} values (supplementary table S10, Supplementary Material online), LA4332 is as close to populations from CG as it is to the other three populations from SHG and could therefore be grouped within CG as well. The

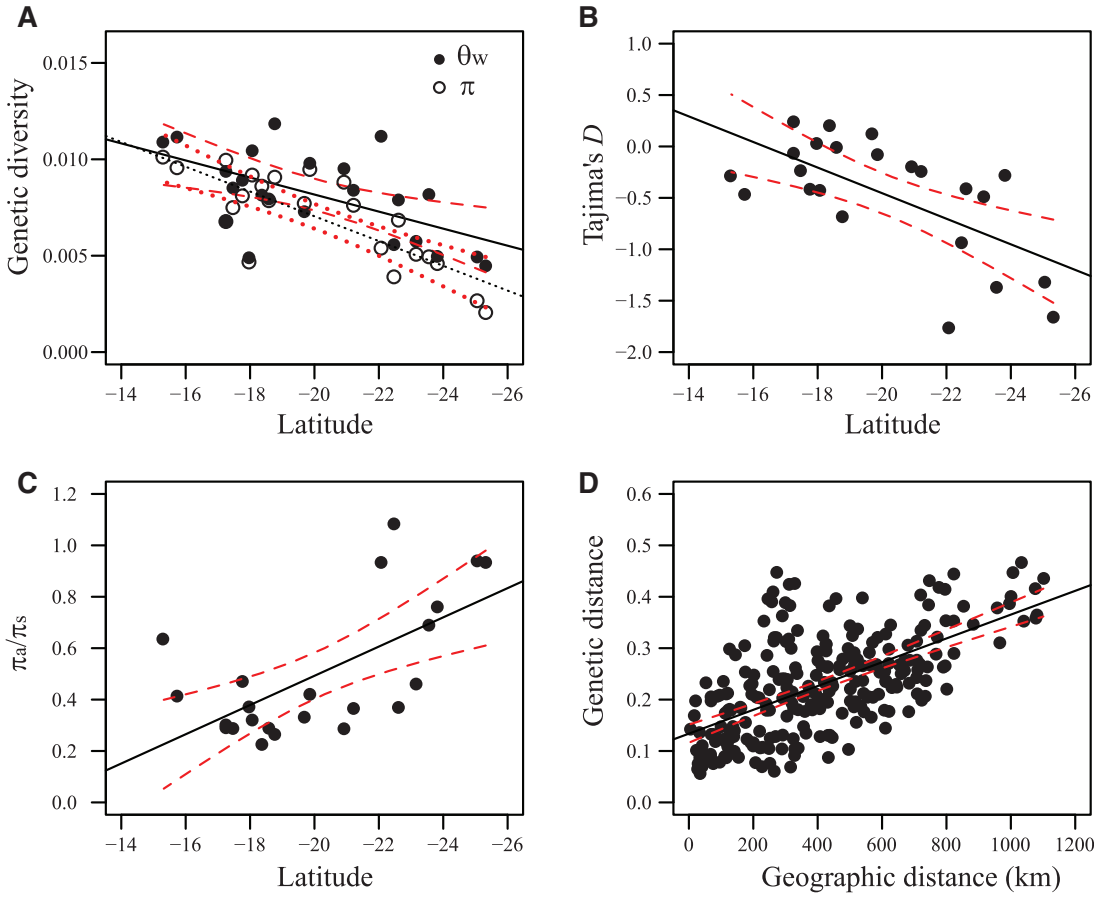


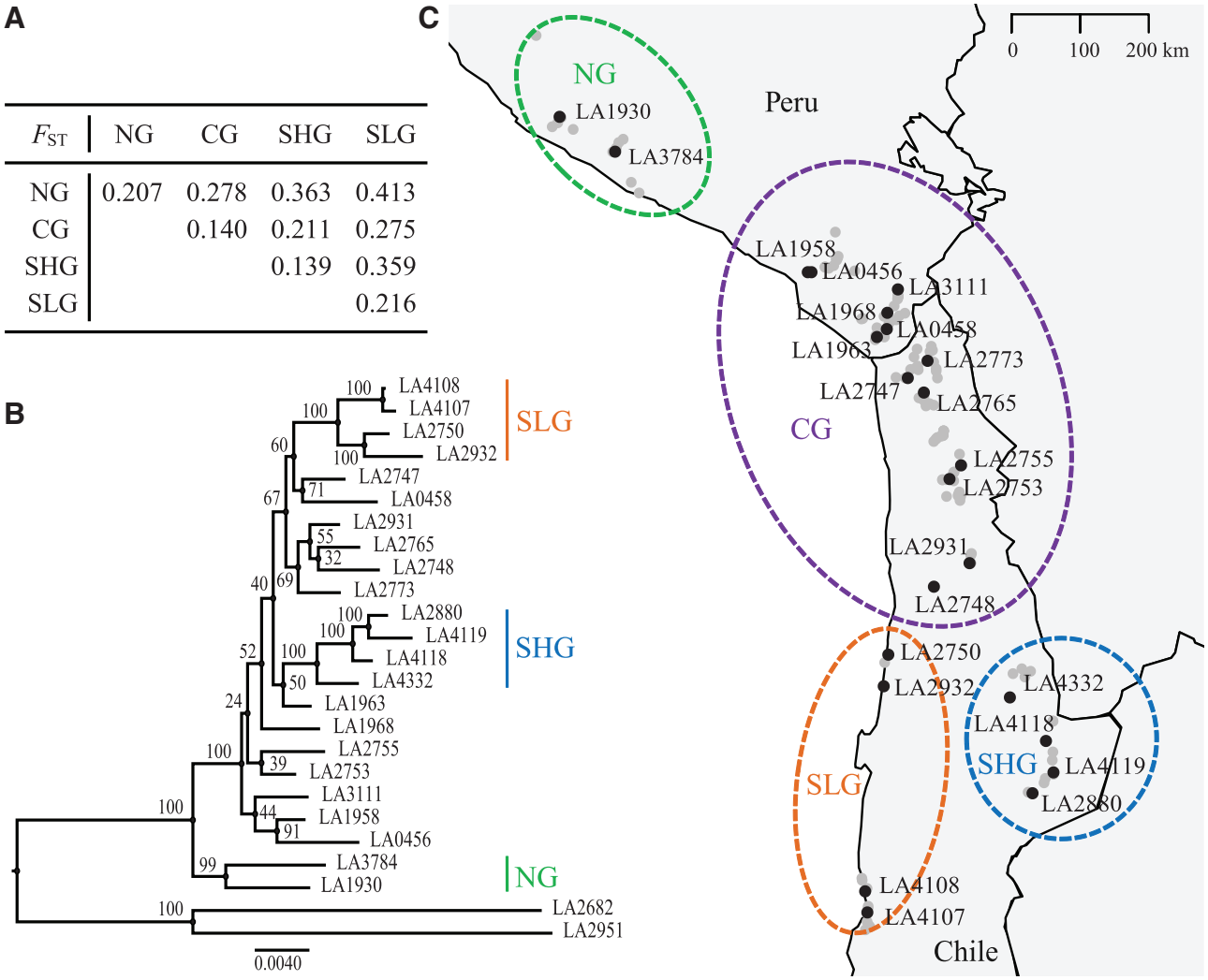
Fig. 1. Correlations between genetic data and latitude. (*A*) Correlation between genetic diversity and latitude: Watterson estimator (θ_W; filled circles; solid line; $R^2 = 0.336$, $\rho = 0.492$, $P = 0.0182$; dashed line: 0.95 confidence interval) and nucleotide diversity (π; open circles; dotted line; $R^2 = 0.639$, $\rho = 0.712$, $P = 0.0002$; dotted line: 0.95 confidence interval). (*B*) Correlation between Tajima's *D* values and latitude ($R^2 = 0.410$, $\rho = 0.525$, $P = 0.0112$; dashed line: 0.95 confidence interval). (*C*) Correlation between π_a/π_s ratio and latitude ($R^2 = 0.0.410$, $\rho = -0.513$, $P = 0.0122$; dashed line: 0.95 confidence interval). θ_W, π, π_a/π_s, and Tajima's *D* values were averaged over all genes. (*D*) Genetic distance (F_{ST}; averaged over all genes) between populations plotted against geographic distance reveals significant pattern of IBD ($R^2 = 0.382$, Mantel test *P* value < 0.001; dashed line: 0.95 confidence interval). See also supplementary table S9 and figures S2–S6, Supplementary Material online.

TreeMix analyses, the genetic distance networks and the phylogenies, however, support the placement of LA4332 in SHG.

These results indicate a likely scenario where *S. chilense* originates in the north, possibly from *S. peruvianum*, with an early split of NG and CG. The two southern groups, SHG and SLG, appear to have diverged independently from CG and colonized different habitats.

Adaptive Evolution and DFE

We assessed the proportion of adaptive amino acid substitutions, α, with the methods of Bierne and Eyre-Walker (2004; hereafter α_{BEW}) and Eyre-Walker and Keightley (2009; hereafter α_{EWK}). The latter takes the effect of slightly deleterious mutations into account by simultaneously estimating the DFE of new deleterious mutations. Slightly deleterious mutations contribute to polymorphism but not to divergence and therefore can lead to an underestimation of α (Charlesworth and Eyre-Walker 2008). In fact, the α_{EWK} values are higher than the α_{BEW} values for most populations. The majority of the northern/central populations (53 %) has positive α_{EWK} values, four even have an α_{EWK} value higher than 25% (LA1963: $\alpha_{EWK} = 0.277$, LA1968: $\alpha_{EWK} = 0.273$, LA2748: $\alpha_{EWK} = 0.353$, LA2773: $\alpha_{EWK} = 0.262$), whereas only one of the southern populations has a positive α_{EWK} value (LA4332: $\alpha_{EWK} = 0.194$). Both α estimates decrease from north to south (supplementary fig. S5*A*, Supplementary Material online), and the decreasing trend is stronger for α_{BEW}. This implies that the southern populations have more slightly deleterious mutations, that the adaptive potential may be higher in the northern/central populations, and that both contribute to the observed cline. The DFE shows that the proportions of weakly deleterious (Ns 1–10) and deleterious (Ns 10–100) new mutations increase from north to south, whereas the proportion of strongly deleterious (Ns > 100) new mutations decreases (fig. 3). This is in agreement with α_{BEW}. We further observe an increase of π_a/π_s from north to south (fig. 1*C*) due to a steeper decrease of π_s than π_a (supplementary fig. S5*B*, Supplementary Material online). In addition, we also show that the southern populations tend to have more genes with low *P* values for the McDonald–Kreitman (MK) test due to



F_{ST}	NG	CG	SHG	SLG
NG	0.207	0.278	0.363	0.413
CG		0.140	0.211	0.275
SHG			0.139	0.359
SLG				0.216

FIG. 2. Population substructure of the data set. (*A*) Mean within- and between-group F_{ST} values for all genes. (*B*) Phylogenetic tree from concatenated consensus sequences with *Solanum ochranthum* (LA2682) and *S. lycopersicoides* (LA2951) as outgroup species. (*C*) Map with distribution of all sampled *S. chilense* populations by the TGRC (gray), the 23 *S. chilense* populations of this study (black), and the four population groups (dashed circles). Population groups are indicated: NG, northern group; CG, central group; SLG, southern low altitude group; SHG, southern high altitude group. See also supplementary table S11 and figures S7–S10, Supplementary Material online.

an excess of nonsynonymous low-frequency polymorphisms (supplementary fig. S5C, Supplementary Material online). These findings suggest relaxed purifying selection in the south, likely due to the stronger effect of genetic drift.

Signatures of Selection

We applied a correlation-based (Bayenv2; Günther and Coop 2013) and a frequency-based (BayeScan; Foll and Gaggiotti 2008; Riebler et al. 2008) method to detect outlier SNPs in our data set. Bayenv2 detected 142 outlier SNPs among the 5,760 SNPs that were used in this analysis (supplementary table S12, Supplementary Material online). Most of them were detected as outliers for correlations including either latitude or longitude or both. These SNP frequency patterns follow the geographic distribution of the populations with highest differentiation between northern and southern populations. Four SNPs correlate with temperature and either longitude or altitude, differentiating thus between low and high altitude populations. No correlation between any of the SNPs and precipitation (as measured as annual precipitation and precipitation of the wettest quarter) was detected.

BayeScan detected 37 outlier SNPs for positive selection based on the F_{ST} values (supplementary table S13, Supplementary Material online). The majority of these SNPs differentiates either populations from the northern range or SLG populations from the remaining populations. Two SNPs differentiate the SHG populations from populations from the other three groups.

We compared the results from the two methods and found 29 SNPs that were identified by both (table 1 and supplementary table S14, Supplementary Material online). Twenty-two of them differentiate the NG populations from the other three groups or populations from NG and some from CG from the remaining populations. Interestingly, 15 of these SNPs are in the gene *CT189*, which encodes a 40S ribosomal protein S19. Three SNPs (SNP_168, SNP_205 in *AREB1*, and SNP_4383 in *CT251*) and two SNPs (SNP_1070 and

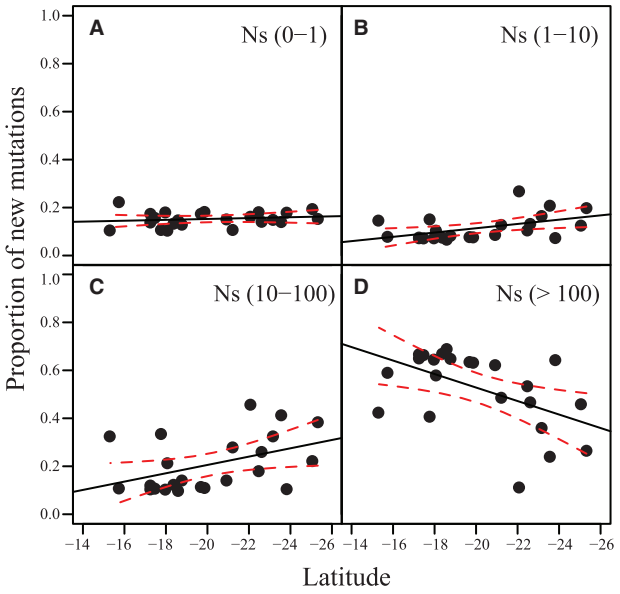


Fig. 3. DFE of new mutations for the 23 *Solanum chilense* populations plotted against latitude with 0.95 confidence intervals. (*A*) Proportion of effectively neutral mutations, Ns (0–1) with $R^2 = 0.030$, $\rho = -0.216$, $P = 0.323$; (*B*) proportion of weakly deleterious mutations, Ns (1–10) with $R^2 = 0.242$, $\rho = -0.464$, $P = 0.027$; (*C*) proportion of deleterious mutations, Ns (10–100) with $R^2 = 0.196$, $\rho = -0.425$, $P = 0.043$; and (*D*) proportion of strongly deleterious mutations, Ns (> 100) with $R^2 = 0.268$, $\rho = 0.488$, $P = 0.019$.

SNP_1244 in *JERF3*) are in high frequency in the SLG and SHG populations, respectively, and one SNP (SNP_276 in *AREB1*) is in high frequency in the SLG and SHA populations. Interestingly, most of the high frequency SNPs in SHG and/or SLG populations are also present in lower frequency in at least one of the other populations suggesting local adaptation from standing genetic variation in the derived southern populations. One SNP (SNP_3441 in *CT166*) is either fixed or in high frequency in the NG and SLG populations. Nineteen of the SNPs identified by both programs are in intronic regions, nine are synonymous, and one is nonsynonymous (SNP_4753 in *GBSSI*) implying that protein-coding positions are not the primary target of selection.

Discussion

In this study, we sequenced multiple genes of a large population sample of the wild tomato species *S. chilense* and investigated the demographic history and patterns of natural selection and adaptation. In the following sections, we discuss the main findings in detail.

Clinal Pattern Indicates Colonization from North to South

Population genetic analyses revealed great differences between populations and significant north–south correlations for most statistics (fig. 1 and supplementary table S9 and figs. S2–S4, Supplementary Material online). This north–south pattern suggests that *S. chilense* originated in the northern range of its current distribution where it co-occurs with its sister species *S. peruvianum* (e.g., Moyle 2008; Städler et al. 2008) and migrated to the south. This scenario is also in agreement with the hypothesis that *S. chilense* may be derived from *S. peruvianum* (Baudry et al. 2001). This colonization resulted in three (θ_W) to five (π) times lower genetic variation in southern than in northern/central populations. Note that the southern *S. chilense* populations, despite the weak bottleneck, still have relatively high levels of genetic variation (mean $\theta_W = 0.0066$, mean $\pi = 0.0044$; supplementary table S9, Supplementary Material online), showing that there was no strong founder event in these populations and pointing to the potential role of seed banks in maintaining genetic diversity (Tellier, Laurent, et al. 2011). Most populations have negative Tajima's *D* values indicating that they are in the process of population expansion. Due to the more recent colonization these signatures are stronger in the south. However, some high altitude populations from the central region have elevated Tajima's *D* values suggesting that they are less expanding possibly due to geological and geographical constraints in the mountains.

Furthermore, a significant pattern of IBD was observed for the 23 *S. chilense* populations (fig. 1*D*). This is also in agreement with a north–south colonization. This was also reported for other wild tomato species including *S. peruvianum*, *S. pimpinellifolium* and *S. lycopersicum* (Nakazato and Housworth 2011; Nakazato et al. 2012) and in the related Solanaceae species *S. lycopersicoides* and *S. sitiens* (Albrecht et al. 2010), which occur in sympatry with *S. chilense* in northern Chile (Peralta et al. 2008; Chetelat et al. 2009). This may indicate that IBD is a common feature of wild tomatoes and related Solanaceae species in western South America.

Seed banks contribute to the observed genetic variation in *S. chilense* (Tellier, Laurent, et al. 2011). The lower genetic variation in the southern range of the species distribution indicates shorter seed banks in the south. This could be explained by the populations in the south being younger either because the southern range was colonized later or because generation times are longer in the south or both. The more extreme environmental conditions in the south could account for longer generation times. The coastal region of Chile is greatly affected by the El Niño's southern oscillations, which provide heavy rainfall. Some *Solanum* populations were reported to be present only during El Niño years (Chetelat et al. 2009). The El Niño strongly affects plant growth, reproduction, and seed banks in coastal Chile (Holmgren et al. 2001; Gutiérrez and Meserve 2003). Reproduction and/or seed germination may only be possible during El Niño years leading therefore to longer generation times. However, mainly coastal populations should be affected by the El Niño and not the high altitude populations (Houston 2006; Chetelat et al. 2009). Therefore, it is unlikely that the El Niño caused elongated generation times, and thus lower genetic variation, in both regions.

Clustering of Populations

Our analyses revealed that the 23 *S. chilense* populations of our sample can be clustered in four groups (CG, NG, SHG, and SLG; fig. 2). On average, the within-group F_{ST} values are lower

Table 1. Summary of Outlier SNPs Detected Jointly by Bayenv2 and BayeScan.

SNP No.	Gene	Type	Bayenv2		BayeScan		
			Environmental Variable	Bayes Factor	Alpha	q Value	F_{ST}
168	*AREB1*	Intron	Latitude	4.04	1.226	0.02454	0.731
205	*AREB1*	Intron	Latitude	6.26	0.812	0.01994	0.664
276	*AREB1*	Intron	Latitude	7.46	0.658	0.02624	0.633
1070	*JERF3*	S	Longitude	6.24	1.263	0.00108	0.747
			T_{min}	3.35			
1234	*JERF3*	Intron	Longitude	4.60	1.029	0.03499	0.698
1239	*JERF3*	S	Longitude	4.92	1.312	0.00765	0.751
1244	*JERF3*	S	Longitude	4.93	0.910	0.02978	0.680
			T_{min}	3.01			
1736	*pLC30-15*	S	Longitude	3.25	1.048	0.03568	0.700
2441	*CT021*	Intron	Latitude	6.09	0.847	0.01172	0.672
			Longitude	6.73			
3441	*CT166*	Intron	Longitude	5.76	0.674	0.01661	0.638
			Altitude	4.05			
			T_{mean}	3.31			
			T_{min}	3.75			
3659	*CT189*	S	Latitude	6.28	1.054	0.00745	0.710
			Longitude	12.71			
3662	*CT189*	S	Latitude	7.11	1.194	0.00241	0.735
			Longitude	14.02			
3663	*CT189*	S	Latitude	4.59	1.341	0.00043	0.760
			Longitude	14.48			
3672	*CT189*	Intron	Latitude	5.01	1.340	0.00027	0.760
			Longitude	12.55			
3682	*CT189*	Intron	Latitude	4.70	1.356	0.00049	0.762
			Longitude	11.31			
3685	*CT189*	Intron	Latitude	6.42	0.775	0.02509	0.656
			Longitude	9.52			
3686	*CT189*	Intron	Latitude	6.01	0.773	0.02798	0.656
			Longitude	10.78			
3766	*CT189*	Intron	Latitude	6.13	1.203	0.00193	0.737
			Longitude	13.77			
3783	*CT189*	Intron	Latitude	6.15	1.193	0.00234	0.735
			Longitude	12.43			
3789	*CT189*	Intron	Latitude	7.68	0.816	0.01848	0.665
			Longitude	10.25			
3801	*CT189*	Intron	Latitude	7.73	0.860	0.01464	0.674
			Longitude	12.48			
3810	*CT189*	Intron	Latitude	6.53	1.215	0.00139	0.739
			Longitude	11.94			
3812	*CT189*	Intron	Latitude	6.67	0.932	0.00848	0.688
			Longitude	10.79			
3816	*CT189*	Intron	Latitude	7.14	0.909	0.01035	0.684
			Longitude	13.44			
3822	*CT189*	Intron	Latitude	4.45	0.884	0.02624	0.676
			Longitude	10.49			
4383	*CT251*	S	Latitude	3.48	0.816	0.01144	0.666
4753	*GBSSI*	NS (Lys/Thr)	Longitude	6.34	1.396	0.00275	0.766
4777	*GBSSI*	Intron	Longitude	4.44	1.221	0.00705	0.737
4785	*GBSSI*	S	Longitude	4.34	1.197	0.00986	0.733

NOTE.—S, synonymous; NS, nonsynonymous; T_{min}, minimum temperature of the coldest month; T_{mean}, mean annual temperature.

than the between-group F_{ST} values. This suggests that gene flow between populations from the same group is higher than between populations from different groups. This is also supported by the finding that IBD is present in the whole sample but absent from CG (supplementary fig. S6, Supplementary Material online). The groups differ in their degree of genetic variation following the north–south colonization: Populations of NG and CG tend to have more genetic

variation than the populations of SHG and SLG (fig. 1*A* and supplementary table S9, Supplementary Material online). The phylogenies show that SHG and SLG are derived from CG and that NG is a sister group to all other *S. chilense* populations (fig. 2*B* and supplementary fig. S9, Supplementary Material online). This is also confirmed by the TreeMix analyses and the genetic distance networks (supplementary figs. S8 and S10, Supplementary Material online).

CG comprises the populations from the central region of the species distribution (approximately 16°30′S–21°30′S; fig. 2*C*). Two groups are in the southern range of the species distribution and are according to the phylogenetic and TreeMix analyses derived from CG. SHG includes the four populations from the high altitude region around Calama, San Pedro de Atacama and the nearby Salar de Atacama (LA4332, LA4118, LA4119, LA2880). This group is also morphologically distinctive, for example, the leaf segments are broader (Chetelat et al. 2009). However, it has to be noted that LA4332 is genetically close to the other three populations of this group and also to many populations from CG (supplementary table S10, Supplementary Material online). This suggests that LA4332 is a "link" between SHG and CG. The same conclusion was reached by the Tomato Genetics Resource Center (TGRC) collectors (Chetelat et al. 2009). A recent admixture or hybridization event has been hypothesized for an *S. chilense* population in 50-km distance to LA4332 (Baudry et al. 2001). This might indicate that the *S. chilense* populations in this area represent "links" between populations from the central part and populations from the Salar de Atacama region or that this area was colonized by at least two different colonization waves from the central region. Nevertheless, LA4332 grouped clearly with the other three SHG populations in the phylogenetic tree confirming our classification of LA4332 as belonging to SHG (fig. 2*B*).

The second from CG-derived group in the south (SLG) includes the four populations from the coast near Tocopilla (LA2750, LA2932) and Taltal (LA4107, LA4108). However, F_{ST} values between Tocopilla and Taltal are in the range of the observed between group F_{ST} values. Therefore, the four SLG populations most likely form two subgroups. A further interesting aspect is that the genetic differentiation between SHG and SLG is similar to the genetic differentiation between these two groups and NG. This is surprising as SHG and SLG are geographically much closer to each other. This finding could be explained by the climatic differences between the two southern habitats. The SLG populations at the coast receive most rainfall during the winter season, whereas the SHG populations at the high altitudes receive most rainfall during the summer season (Houston 2006). This factor could lead to a temporal barrier to gene flow. Additionally, the desert conditions separating the two groups could prevent pollen and seed dispersal between the groups. Analyses also show that SHG and SLG are derived from CG (fig. 2*B*). However, they do not form sister clades in the phylogenies or genetic distance networks although TreeMix suggests that they are derived from the same ancestral population (supplementary figs. S8–S10, Supplementary Material online). This disagreement may be explained by old split times between the groups. Simulation studies will be required to estimate split times between the groups.

The third group is represented by only two populations of our sample, LA1930 and LA3784, located north of CG. Interestingly, the two NG populations form a sister clade to the rest of the *S. chilense* populations in the phylogeny of the consensus sequences indicating that they are not derived from CG (fig. 2*B*). In the TreeMix topology, they are closest to the sister species *S. peruvianum* (supplementary fig. S8*B*, Supplementary Material online). This could indicate that the northern *S. chilense* populations are an early diverged group of *S. chilense* and/or that they are partly admixed with other wild tomato species, that is, that interspecies gene flow is still occurring. The split between *S. chilense* and its sister species *S. peruvianum* was estimated to be relatively recent (<1 Ma; Städler et al. 2008; Särkinen et al. 2013). Therefore, it is also possible that speciation is not complete yet and that gene flow continues in regions where both species co-occur. Speciation with gene flow has actually been reported for the species pair *S. peruvianum* and *S. chilense* (Städler et al. 2005, 2008). This further suggests that this group most likely coincides with the origin of the species from its sister species *S. peruvianum* as suggested before (Baudry et al. 2001). This is also supported by the high genetic variation observed. Alternatively, incomplete lineage sorting could cause this pattern. Ancestral alleles that were lost during the first split (NG and CG) in the majority of the *S. chilense* populations could still be present in the northern populations. In fact, a *trans*-species polymorphism at the *CBF2* locus is shared between a northern *S. chilense* population and an *S. peruvianum* population (Mboup et al. 2012). Seed banks could further contribute to the observed pattern. The large seed banks in the north could maintain seeds from prespeciation times that carry ancestral alleles.

Colonization Associated with a Shift in Adaptive Evolution and Intensity of Selection

The proportion of adaptive amino acid substitutions, α, is lower in the south than in the north (supplementary fig. S5*A*, Supplementary Material online). When estimated with the method of Bierne and Eyre-Walker (2004), α is negative for all populations whereas several populations have positive α values with the method of Eyre-Walker and Keightley (2009). We found four populations with a relative high potential of adaptive evolution ($\alpha_{EWK} > 25$ %). The difference between the two methods implies that slightly deleterious mutations are present in all populations and that the southern populations have more slightly deleterious mutations than the northern ones (see also DFE), possibly contributing to the observed cline. Positive α values are correlated with genetic variation in the *S. chilense* data, implying the potential for adaptive evolution predominantly in large populations (supplementary fig. S11, Supplementary Material online). This would support the classic population genetics expectation. In fact, nine of the ten populations with positive α values are from NG and CG. A positive correlation between adaptive evolution and effective population size has been previously

suggested (Gossmann et al. 2012). Overall our data imply that adaptive evolution is low in *S. chilense*, especially in the southern range, which is consistent with the little evidence of adaptive evolution found in many plant species (but see *Helianthus*; Gossmann et al. 2010) and a previous study in wild tomatoes (Tellier, Fischer, et al. 2011), due to the spatial structuring of populations and also the seed banks which slow down selection and increase the time for beneficial mutations to get fixed.

Furthermore, a shift in the strength of selection and genetic drift is associated with the colonization. The DFE shows an increase in the proportion of (weakly) deleterious and a decrease in the proportion of strongly deleterious new mutations (fig. 3), suggesting that purifying selection is in the south less efficient than in the north. This shift may be due to the more recent colonization through a bottleneck and the associated effects of genetic drift. Accordingly, the southern populations have a higher proportion of nonsynonymous genetic variation as indicated by the π_a/π_s and lower MK test *P* values due to an excess of nonsynonymous (low-frequency) polymorphisms (supplementary fig. S5, Supplementary Material online). It may also be possible that in the southern range environments these mutations are not deleterious.

Signatures of Selection

Methods to detect outlier SNPs, that is, SNPs under selection, are frequently used in plant evolutionary genetics (e.g., Alberto et al. 2013; Cullingham et al. 2014). We applied a correlation-based and a frequency-based method (Foll and Gaggiotti 2008; Günther and Coop 2013). Note that as our scenario involves a metapopulation expansion and nonequilibrium population structure, we cannot exclude that the BayeScan results could reflect the underlying demographic history rather than selective events. In fact, strong bottlenecks in a data set can cause false positives (Foll and Gaggiotti 2008). However, as a reasonable number of outlier SNPs was detected by both methods, selection is more likely.

The two methods detected a high number of outlier SNPs; 29 outliers were detected by both. Most of these SNPs are either intronic or synonymous. These positions may be involved in alternative splicing or codon usage bias. A similar trend was also found previously (Cullingham et al. 2014). As linkage disequilibrium (LD) is notoriously low in *S. chilense*, it is not expected that most of these SNPs are in linkage with advantageous nonsynonymous mutations (Arunyawat et al. 2007).

Two outlier SNPs are in the transcription factor *JERF3* and may be associated with adaptation of the SHG populations to the high altitudes in the south. This gene is upregulated under salinity and low-temperature stress in the cultivated tomato and can enhance salt and freezing tolerance in transgenic tobacco (Wang et al. 2004; Wu et al. 2008). *JERF3* may therefore be involved in adaptation to these stresses in the SHG populations. Three outlier SNPs are in the transcription factor *AREB1*. This gene may be the key transcription factor in the abscisic acid-dependent abiotic stress pathway and has been shown to be involved in different abiotic stress responses including drought and salinity in tomato (Yáñez et al. 2009; Orellana et al. 2010). As these SNPs are in high frequency in the SLG populations, they may indicate that *AREB1* is involved in adaptation to high soil salinity at the coast. A closer look at these high-frequency SNPs in SHG and/or SLG in *JERF3* and *AREB1* revealed that they are also present in lower frequencies in at least one of the other populations (supplementary table S14, Supplementary Material online). This fact suggests that these SNPs represent neutral standing genetic variation that became advantageous, and thus adaptive, during the colonization of the new habitats in the south.

Interestingly, we could not detect any outlier SNP differentiating between populations of CG although the CG populations occur in different environments. As the CG populations represent a genetically homogeneous group, as shown by the absence of IBD, recurrent gene flow slows down local adaptation.

Outlier SNPs were detected not only in genes related to abiotic stress but also in the reference genes. This could indicate that some of these genes are also involved in environmentally important pathways. Interestingly, 15 outlier SNPs, possibly in LD (supplementary table S14, Supplementary Material online), were detected by both methods in the 40S ribosomal protein S19 encoding reference gene *CT189*. Further analyses of this gene will be required to explain the observed pattern.

Conclusions

In our study, we reconstructed the evolutionary history of the wild tomato species *S. chilense*. Our analyses revealed a strong demographic signature with a north–south pattern present in most statistics and four distinct population groups. Together, these results suggest that the origin of *S. chilense* lies in the north of its current distribution and that it colonized its present range from north to south. During the colonization, the CG split from the NG and the two southern groups split from the CG likely through weak bottleneck events. Furthermore, the colonization was accompanied by a shift in the strength of purifying selection and in the proportion of adaptive amino acid substitutions suggesting a higher potential of adaptive evolution in the northern/central range. We identified five SNPs in two abiotic stress-related transcription factors, *AREB1* and *JERF3*, which could be associated with adaptation to the abiotic environments in the southern range of the species distribution.

Our results may point toward two other interesting aspects regarding the evolutionary history of *S. chilense* that may be of interest for future studies. First, we observed populations at the southern marginal ranges with decreased genetic variation when all sites were considered, but increased nonsynonymous genetic variation. Such a pattern has been observed in selfing compared with outcrossing species (Slotte et al. 2013; Zuellig et al. 2014). Transitions from outcrossing to selfing in populations from the marginal ranges of their species distribution are not uncommon in wild tomatoes (Rick et al. 1979; Rick and Tanksley 1981; Graham et al. 2003) and therefore such a transition may also occur in *S. chilense*. Second, we observed an increase in seed weight and size

from north to south (supplementary fig. S12, Supplementary Material online). As seed size is determined by major quantitative trait loci in tomato (Doganlar et al. 2000), it is unlikely that genetic drift alone would be responsible for this increase by acting upon single mutations in single genes. This suggests that the north–south colonization of *S. chilense* may have been accompanied by selection favoring larger seeds. Larger seeds may be advantageous for *S. chilense* in the more extreme environments, as has been observed in other species (Manga and Yadav 1995; Pluess et al. 2005). Larger seeds are usually associated with shorter seed banks, because they remain at near-surface depth in the soil (Fenner and Thompson 2005). Shorter seed banks may therefore be an adaptation in southern populations of *S. chilense*.

Materials and Methods

Detailed methods are described in the supplementary materials and methods, Supplementary Material online.

Sequencing Approach

We sequenced 30 genes of 23 *Solanum chilense* populations in a pooling approach (25 diploid individuals of each population). *Solanum chilense* is a diploid ($2n = 24$), self-incompatible perennial plant, native to a multitude of different habitats in South America ranging from mesic areas in southern Peru to the hyperarid areas in the Atacama Desert and the high altitudes in the Andean Mountains in northern Chile (Moyle 2008; Peralta et al. 2008; Chetelat et al. 2009). The 23 populations are distributed over the entire species and habitat range (fig. 2*A* and supplementary table S1, Supplementary Material online). For outgroup comparisons, one population of each *S. ochranthum* and *S. lycopersicoides* was chosen (supplementary table S1, Supplementary Material online) and the 30 genes were sequenced as described in Fischer et al. (2011). Both species are close to the wild tomato clade with estimated split times of around 5.59 and 5.95 Ma, respectively (Särkinen et al. 2013). Seeds were obtained from the TGRC (UC Davis; http://tgrc.ucdavis.edu, last accessed August 10, 2015) and plants were grown as described in Fischer et al. (2013).

The 30 genes were selected based on previous publications: 14 genes serve as reference genes and are evenly distributed over the entire genome, whereas 16 genes are involved in the abiotic stress response and serve as candidate genes for selection (supplementary tables S2–S5, Supplementary Material online). Genomic DNA was isolated from each plant using the DNeasy Plant Mini Kit (Qiagen). Genes were amplified by polymerase chain reaction (PCR) using the Phusion High-Fidelity DNA Polymerase kit (Finnzymes, distributed by New England BioLabs, Inc.). PCR products were purified with the MinElute Gel Extraction Kit (Qiagen). We applied a three-step pooling approach: 1) Mixing of five individuals (DNA) to prepools before PCR amplification, 2) mixing of five prepools (PCR product) to pools before PCR product purification, and 3) mixing of all 30 purified PCR product pools of each population (supplementary fig. S1, Supplementary Material online). DNA library construction and high-throughput sequencing (paired-end, 100 bp) on a Genome Sequencer Illumina HiSeq2000 were performed by the GATC Biotech AG (Konstanz, Germany). The population LA1968 was sequenced twice starting from the five prepools to evaluate the repeatability of the sequencing approach.

Data Assembly and Analyses

The *S. chilense* short reads were mapped with Stampy v1.0.20 (Lunter and Goodson 2011) against reference sequences. The SAMtools program (Li et al. 2009) was used to generate pileup files per population. All positions with SNP information were extracted from the pileup files and transformed into a table with SNP information for each position. We conducted all data analyses based on these SNP tables. Positions with a sequencing depth below 2,000, a base quality, and/or a mapping quality below 20 were masked and excluded from all analyses. Mutations present in less than 1% of the reads were considered as sequencing errors and therefore excluded from all analyses. Indel polymorphisms were also excluded.

We calculated the summary statistics, including the Watterson estimator, θ_W (Watterson 1975), the nucleotide diversity π (Nei and Li 1979), and the neutrality test statistic Tajima's *D* (Tajima 1989), for each gene in each population for all sites and for silent sites. We also assessed π for synonymous and nonsynonymous sites and applied statistics based on the ratios of synonymous and nonsynonymous polymorphism and divergence to *S. ochranthum*: The MK test (McDonald and Kreitman 1991), the proportion of adaptive substitutions α, and the DFE. The DoFE software was used to calculate α according to Bierne and Eyre-Walker (2004) and jointly α and the DFE according to Eyre-Walker and Keightley (2009). We tested for correlations with geographic parameters by performing regressions and calculating the Spearman's rank correlation ρ in R (R Development Core Team 2005).

We assessed the pairwise genetic differentiation between each population pair with the F_{ST} statistics (Hudson et al. 1992). We further tested for IBD in *S. chilense* by plotting the pairwise F_{ST} values against the geographic distance and performing regressions and a Mantel test in R (Mantel 1967; R Development Core Team 2005). Geographic distances were inferred from the latitudinal and longitudinal decimals from the TGRC (http://tgrc.ucdavis.edu, last accessed August 10, 2015).

We used TreeMix to investigate the relationships between the populations (Pickrell and Pritchard 2012). As population data of *S. chilense*'s sister species *S. peruvianum* are available for seven of the reference genes (Arunyawat et al. 2007), we also ran TreeMix on this combined data set to distinguish between ancestral and derived *S. chilense* populations. In addition to TreeMix, we also constructed genetic distance networks from the F_{ST} values in MEGA v5.2.2 (Tamura et al. 2003) and phylogenetic trees from concatenated consensus sequence alignments using RaxML v7.2.7 (Stamatakis et al. 2005). For the latter, *S. ochranthum* and *S. lycopersicoides* were included as outgroup species.

To identify signatures of selection and local adaptation in our data set, we applied two methods that detect outlier

SNPs and take the underlying demography into account. Bayenv2.0 tests for correlations between SNP frequency patterns and environmental variables (Günther and Coop 2013). We used geographic and climatic data as environmental variables (supplementary table S6, Supplementary Material online). BayeScan v2.1 detects outlier SNPs based on F_{ST} statistics (Foll and Gaggiotti 2008).

Supplementary Material

Supplementary materials and methods, references, tables S1-S14, and figures S1-S12 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank G. Kumpfmüller for excellent technical assistance with the outgroup sequencing, Armin Scheben for help with the π_a, π_s, and MK test analyses, Pablo Duchén for help with the data processing, the Munich evolutionary biology group for valuable discussions, and the handling editor and two anonymous reviewers for their comments. They are also grateful to the Tomato Genetics Resource Center for providing the seeds and Roger Chetelat for additional information. This work was supported by the Deutsche Forschungsgemeinschaft (grants STE 325/13 and STE 325/15 to W.S.).

References

Alberto FJ, Derory J, Boury C, Frigerio JM, Zimmermann NE, Kremer A. 2013. Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195:495–512.

Albrecht E, Escobar M, Chetelat RT. 2010. Genetic diversity and population structure in the tomato-like nightshades *Solanum lycopersicoides* and *S. sitiens*. *Ann Bot.* 105:535–554.

Arunyawat U, Stephan W, Städler T. 2007. Using multilocus sequence data to assess population structure, natural selection, and linkage disequilibrium in wild tomatoes. *Mol Biol Evol.* 24:2310–2322.

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

Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol Biol Evol.* 21:1350–1360.

Charlesworth J, Eyre-Walker A. 2008. The McDonald-Kreitman test and slightly deleterious mutations. *Mol Biol Evol.* 25:1007–1015.

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.

Cullingham CI, Cooke JE, Coltman DW. 2014. Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytol.* 204:215–229.

Doganlar S, Frary A, Tanksley SD. 2000. The genetic basis of seed-weight variation: tomato as a model system. *Theor Appl Genet.* 100:1267–1273.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol.* 26:2097–2108.

Fenner M, Thompson K. 2005. The ecology of seeds. Cambridge (United Kingdom): Cambridge University Press.

Fischer I, Camus-Kulandaivelu L, Allal F, Stephan W. 2011. Adaptation to drought in two wild tomato species: the evolution of the *Asr* gene family. *New Phytol.* 190:1032–1044.

Fischer I, Steige KA, Stephan W, Mboup M. 2013. Sequence evolution and expression regulation of stress-responsive genes in natural populations of wild tomato. *PLoS One* 8:e78182.

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993.

Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol.* 4:658–667.

Gossmann TI, Song BH, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol.* 27:1822–1832.

Graham EB, Shannon SM, Petersen JP, Chetelat RT. 2003. A self-compatible population of *Lycopersicon peruvianum* collected from N. Chile. *TGC* 53:22–24.

Günther T, Coop G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* 195:205–220.

Gutiérrez JR, Meserve PL. 2003. El Nino effects on soil seed bank dynamics in north-central Chile. *Oecologia* 134:511–517.

Holmgren M, Scheffer M, Ezcurra E, Gutiérrez JR, Mohren GM. 2001. El Nino effects on the dynamics of terrestrial ecosystems. *Trends Ecol Evol.* 16:89–94.

Hörger AC, Ilyas M, Stephan W, Tellier A, van der Hoorn RA, Rose LE. 2012. Balancing selection at the tomato *RCR3* guardee gene family maintains variation in strength of pathogen defense. *PLoS Genet.* 8:e1002813.

Hough J, Williamson RJ, Wright SI. 2013. Patterns of selection in plant genomes. *Annu Rev Ecol Evol Syst.* 44:31–49.

Houston J. 2006. Variability of precipitation in the Atacama Desert: Its causes and hydrological impact. *Int J Climatol.* 26:2181–2198.

Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.

Husband BC, Barrett SCH. 1996. A metapopulation perspective in plant population biology. *J Ecol.* 84:461–469.

Innan H, Kim Y. 2008. Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* 179:1713–1720.

Kirkpatrick M, Barton NH. 1997. Evolution of a species' range. *Am Nat.* 150:1–23.

Lange BW, Langley CH, Stephan W. 1990. Molecular evolution of *Drosophila* metallothionein genes. *Genetics* 126:921–932.

Leimu R, Fischer M. 2008. A meta-analysis of local adaptation in plants. *PLoS One* 3:e4010.

Li H. Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 21:936–939.

Manga VK, Yadav OP. 1995. Effect of seed size on developmental traits and ability to tolerate drought in pearl millet. *J Arid Environ.* 29:169–172.

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

Mboup M, Fischer I, Lainer H, Stephan W. 2012. Trans-species polymorphism and allele-specific expression in the *CBF* gene family of wild tomatoes. *Mol Biol Evol.* 29:3641–3652.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.

Moyle LC. 2008. Ecological and evolutionary genomics in the wild tomatoes (*Solanum* sect. *Lycopersicon*). *Evolution* 62:2995–3013.

Nakazato T, Franklin RA, Kirk BC, Housworth EA. 2012. Population structure, demographic history, and evolutionary patterns of a green-fruited tomato, *Solanum peruvianum* (Solanaceae), revealed by spatial genetics analyses. *Am J Bot.* 99:1207–1216.

Nakazato T, Housworth EA. 2011. Spatial genetics of wild tomato species reveals roles of the Andean geography on demographic history. *Am J Bot.* 98:88–98.

Nakazato T, Warren DL, Moyle LC. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *Am J Bot.* 97:680–693.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 76:5269–5273.

Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 5:e1000698.

Orellana S, Yáñez M, Espinoza A, Verdugo I, González E, Ruiz-Lara S, Casaretto JA. 2010. The transcription factor SlAREB1 confers drought, salt stress tolerance and regulates biotic and abiotic stress-related genes in tomato. *Plant Cell Environ.* 33:2191–2208.

Peralta IE, Spooner A, Knapp S. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicon*, sect. *Juglandifolia*, sect. *Lycopersicoides*; Solanaceae). *Syst Bot Monogr.* 84:1–186.

Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967.

Pluess AR, Schütz W, Stöcklin J. 2005. Seed weight increases with altitude in the Swiss Alps between related species but not among populations of individual species. *Oecologia* 144:55–61.

R Development Core Team. 2005. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for statistical Computing (editor).

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 Syst Evol.* 132:279–298.

Rick CM, Tanksley SD. 1981. Genetic variation in *Solanum pennellii*: comparisons with two other sympatric tomato species. *Plant Syst Evol.* 139:11–45.

Riebler A, Held L, Stephan W. 2008. Bayesian variable selection for detecting adaptive genomic differences among populations. *Genetics* 178:1817–1829.

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

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol Biol.* 13:214.

Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F, Guo YL, Steige K, Platts AE, Escobar JS, Newman LK, et al. 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet.* 45:831–835.

Städler T, Arunyawat U, Stephan W. 2008. Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics* 178:339–350.

Städler T, Roselius K, Stephan W. 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* 59:1268–1279.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tamura T, Hara K, Yamaguchi Y, Koizumi N, Sano H. 2003. Osmotic stress tolerance of transgenic tobacco expressing a gene encoding a membrane-located receptor-like protein from tobacco plants. *Plant Physiol.* 131:454–462.

Tellier A, Fischer I, Merino C, Xia H, Camus-Kulandaivelu L, Städler T, Stephan W. 2011. Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity (Edinb)* 107:189–199.

Tellier A, Laurent SJY, Lainer H, Pavlidis P, Stephan W. 2011. Inference of seed bank parameters in two wild tomato species using ecological and genetic data. *Proc Natl Acad Sci U S A.* 108:17052–17057.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet.* 42:260–263.

Wang H, Huang Z, Chen Q, Zhang Z, Zhang H, Wu Y, Huang D, Huang R. 2004. Ectopic overexpression of tomato *JERF3* in tobacco activates downstream gene expression and enhances salt tolerance. *Plant Mol Biol.* 55:183–192.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.

Willi Y, Van Buskirk J, Hoffmann AA. 2006. Limits to the adaptive potential of small populations. *Annu Rev Ecol Evol Syst.* 37:433–458.

Wu L, Zhang Z, Zhang H, Wang XC, Huang R. 2008. Transcriptional modulation of ethylene response factor protein JERF3 in the oxidative stress response enhances tolerance of tobacco seedlings to salt, drought, and freezing. *Plant Physiol.* 148:1953–1963.

Xia H, Camus-Kulandaivelu L, Stephan W, Tellier A, Zhang Z. 2010. Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. *Mol Ecol.* 19:4144–4154.

Yáñez M, Cáceres S, Orellana S, Bastías A, Verdugo I, Ruiz-Lara S, Casaretto JA. 2009. An abiotic stress-responsive bZIP transcription factor from wild and cultivated tomatoes regulates stress-related genes. *Plant Cell Rep.* 28:1497–1507.

Zuellig MP, Kenney AM, Sweigart AL. 2014. Evolutionary genetics of plant adaptation: insights from new model systems. *Curr Opin Plant Biol.* 18C:44–50.