Local and temporal adaptation to climatic change in a wild tomato species via selective sweeps

Kai Wei *, Gustavo A Silva-Arias *, Aurélien Tellier

Professorship for Population Genetics, Department of Life Science Systems, School of Life Sciences, Technical University of Munich, Liesel-Beckmann Strasse 2, 85354 Freising, Germany

* these authors contributed equally

Corresponding author:

Aurélien Tellier tellier@wzw.tum.de

Keywords:

Local adaptation, selective sweeps, age of sweeps, last glacial maximum, gene network

Summary

Positive Darwinian selection is the driving force underpinning local (in space) and temporal (in time) adaptation, and leaves footprints of selective sweeps at the underlying major genes. These two adaptive processes are classically considered independently, so that most genomic selection scans uncover only recent sweeps underpinning spatial adaptation. However, understanding if these adaptive processes are intermingled and share common genetic bases is crucial to predict a species' evolutionary potential, for example in response to changing environmental conditions. We use whole genome data from six populations across three different habitats of the wild tomato species *Solanum chilense*, to 1) infer the past demographic history, and 2) search for genes under strong positive selection. We then correlate the demographic history, allele frequencies in space and time, the age of selection events and the reconstructed historical ecological distribution of the species over five main climatic periods spanning 150,000 years. We find evidence for several selective sweeps targeting regulatory networks involved in root hair development in low altitude, and response to photoperiod and vernalization in high altitude populations. These sweeps occur in a concerted fashion in a given regulatory gene network and only at particular time periods, thereby underpinning temporal local adaptation. These genes under positive selection provide subsequently the basis for spatial local adaptation to novel habitats when new ecological niches become available. Our results reveal the importance of jointly studying spatial and temporal adaptations in species during habitat colonization.

Introduction

Positive Darwinian selection is a key process driving macro and micro-evolution and underlying local adaptation when species expand their range and colonize novel habitats. Local adaptation encompasses thus the spatial component of adaptation to a new habitat, but also a temporal component when climatic/abiotic conditions vary in time after colonization[1]. The latter is understood here at the scale of hundreds up to thousands of years during which climatic conditions are changing due to glacial/inter-glacial periods or medium to long-term changes in abiotic and biotic environmental conditions (the so-called moving environmental optimum). Population genetics theory investigates the characteristics and polymorphism footprints of local adaptation while separating the two processes: spatial adaptation to a local optimum based on Fisher's geometric model of adaptation[2] and temporal adaptation to a moving optimum[1,3,4]. Under both models and large enough population sizes, it is typically expected that bigger steps of adaptation occur first at sites with strong selective coefficients, possibly generating selective sweeps[2–4]. Selective sweeps are polymorphism patterns (footprints) in the genome due to the rapid fixation of advantageous alleles and the associated hitch-hiking effect[5,6]. However, there is so far no complete theory investigating the more realistic scenario combining both local spatial adaptation to newly colonized habitats and further temporal adaptation to temporally variable climatic conditions and the expected patterns of selection.

We identify therefore a number of unanswered theoretical and empirical questions regarding the joint effect of spatial and temporal local adaptation. How many genes underlie adaptation to new habitats and/or changing habitat in time? How many genes underpinning temporal or spatial adaptation show strong selection (selective sweeps)? Are the same genes involved in both adaptive processes? Do these genes belong to one or few gene pathways/regulatory networks? What is the timing of adaptation and can selection at a given gene be linked to past changes in climatic conditions? This lack of knowledge regarding the underpinnings of local

adaptation represents a major bottleneck to predict/assess the species potential for adaptability and resilience in the face of the current global climatic change. Indeed, despite numerous studies highlighting genes or even SNPs underlying spatial climatic adaptation[7,8], future predictions of the strength of selection and change in allele frequencies in response to the predicted climate change trends remain fairly speculative.

To date, numerous studies uncovering the genetic bases of spatial adaptation, have been conducted in few model organisms such as humans, invertebrates (*Caenorhabditis elegans, Drosophila melanogaster*) or plants (*Arabidopsis* spp., *Arabis* sp., *Capsella* spp.). Using genome scan methods based on the selective sweep/hitchhiking model[5,6], Genome Wide Association studies (GWAs) and functional studies, the crucial role of changes in genes/proteins (coding regions) and/or gene expression and the relevant gene network/pathways has been investigated (review in [9,10]). Interestingly, most model study systems exhibit spatial local adaptation due to recent selective sweeps being dated less than 25,000 years old[11–13] which are associated with strong bottlenecks during colonization events (*e.g.* colonization of Europe in humans and *Drosophila*) or due to recent recolonization after glacial period, bottlenecks in glacial refugia and selfing (in *Arabidopsis thaliana*, *Capsella* sp. and *Arabis* sp.). The observation of recent sweeps, and lack of older selection events, is explained by the lack of power to detect loci under strong positive local and temporal selection due to three factors. First, there is a confounding effect of past demographic history of the populations[14] especially under strong bottlenecks often characterizing the colonization of new habitats[15]. Disentangling the two types of signatures, neutral demography and selection, is possible by inferring the past demography over the whole genome and search for outlier portions of the genome under selection. However, under strong demographic bottlenecks, it is difficult to carefully control the statistical rate of false positive sweeps and sweep signatures pre-dating the bottleneck are blurred[12–14]. Second, strong selection at few genes is

only the tip of the iceberg, as adaptation likely occurs for traits which are polygenic with selection acting over many genes. Different number of major genes with selective sweeps signatures can be observed depending on the distribution of selection coefficients, the efficiency of selection (a function of effective population size and recombination rate), the architecture of the traits, place of genes in gene networks/pathways, and gene pleiotropy[16,17]. While theoretical predictions exist, a general understanding of the genetic architecture of local adaptation is still missing (but see [16]). Third, under temporal and polygenic selection model, strong selection can only occur if population sizes are large enough[3,14]. Furthermore, theory predicts that selective sweeps can be observed and detected reliably only for ca. 0.1*Ne* generations after selection, where *Ne* is the effective population size[5]. As a result, in species undergoing habitat range expansion with strong bottleneck during colonization and life history traits decreasing effective population size (*e.g.* selfing), we expect 1) only few sweeps to occur under polygenic selection, and 2) only recent sweeps to be detected using polymorphism data. Therefore, in order to reveal the underpinnings of long-term spatial and temporal local adaptation due to strong selection (selective sweeps), the study of alternative species with larger effective population sizes during colonization events is required.

Solanum chilense is an outcrossing wild tomato species found in southern Peru and northern Chile in mesic to very arid habitats[18]. The ancestral species range is found around the Peruvian and Chilean border mainly along the (mid-altitude) 'pre-cordillera' region (800 - 2000 m altitude) characterized as marginal desert habitat. *S. chilense* colonized independently two different southern, but arid, isolated regions around the Atacama desert at different time points[19,20]: an early divergence (older than 50,000 ya) with the colonization of coastal habitats (in Lomas formations), and a more recent lineage divergence (less than 20,000 ya) restricted to highland altitudes (above 2,400m) of the Andean plateau. Signatures of natural selection (positive or balancing) at few genes involved in abiotic stress (drought, cold, salt) adaptation

were found when scanning few candidate genes for biotic and abiotic stresses[19–24]. In the present study, we obtain full genome sequence data for 30 diploid and highly heterozygous plants from six populations (five plants per population) representing the three main habitats of the species (Figure 1A): the central group (area of origin at low to high altitude), south-coastal group, and south-highland group. These habitats differ in their current climatic conditions: the south-highland group is different from the central group, while the south-coastal appears as only marginally different from the environment prevailing in the central group (Figure 1B). Our aims are first to infer accurately the past demographic history of the species colonization as well as to reconstruct the past climatic history of the species' range. Second, we conduct genome scans for genes under selection in these populations and assign functions and gene network position to these candidate genes. We then link climatic and genetic data at candidate genes to highlight the relevance of few key gene regulatory networks (pathways) for local spatial adaptation. Finally, we connect past species niche reconstruction and age estimates of the selective sweeps in order to reveal the history of the temporal local adaptation to ever changing climatic conditions.

Results

Past demography of *S. chilense* is influenced by colonization events and climatic variations

We sequence whole genomes of 30 heterozygous plants from *Solanum chilense* from six populations (LA3111, LA1963, LA2931, LA2932, LA4107, LA4330) (Figure 1A, Table S1). All reads are aligned to the reference genome assembly of *S. pennellii* and all 30 *S. chilense* individuals show high-quality sequence and mapping scores with more than 97% of mapping paired reads, individual genome coverage ranging between 16 to 24 reads per base, and >70% genome coverage per sample (Data S1A). After SNP calling and stringent filtering, a total of 34,109,217 SNPs are

identified across all samples (Table S2) for a genome size estimated approximately to 914Mb[25]. Phylogenetic and principal component analysis (PCA) analyses support population structuring into three genetic groups (Figure 1C; Figure S1C; confirming the results in [19]): a central group (LA1963, LA3111, LA2931), the south-highland group (LA4330), and the south-coast group (LA2932 and LA4107). Moreover, population structure analyses support that the two south-coast populations constitute independent groups (best K=4; Figure 1D; Figure S1C-E). Only the individuals of the population LA2931 (the southernmost of the central group) display small admixed ancestry coefficients (< 5%) with the south-highland group (Figure 1D). There is no significant correlation between genetic (pairwise Nei's distance) and geographical distance (Pearson test, r = 0.35, P = 0.20; Figure S1A-B).

As indicated by nucleotide diversity and differentiation statistics (Figure S2; Table S3 and S4), we confirm the previous results obtained using a reduced number of genetic markers: *S. chilense* has colonized independently the coastal and highland southern-habitats, and the species area of origin is likely the mid- altitude of the central region[19,20]. Using multiple sequentially Markovian coalescent analyses, we estimate historical changes in effective population size (N_e, Figure 2A), divergence and potential post-divergence gene flow (Figure S2), and construct a consensus demographic model (Figure 2B; Data S1B). These estimates are compared with the reconstructed past climatic variation highlighting five Marine isotope stages (MIS) climatic periods[26,27]. The two south-coast populations found in Lomas habitats show early divergence consistent with the admixture analysis (during the Last Inter-Glacial period, MIS5) likely from the lowland area of the central group (LA1963). The colonization of the highland likely occurred later, first in the central group region (LA3111, LA2931) between the last inter-glacial and Last Glacial Maximum (LGM) periods (ca. 75-130 kya, MIS3-4) and then with further colonization of southern highlands (from 30 kya, MIS1-2, LA4330). All populations show a reduction of diversity matching with the estimated time of the LGM characterized as a cold and

dry period and supported by a contraction of the suitable habitats to a narrow strip in lower altitudes, and a subsequent expansion thereafter (Figure 2C). Indeed, the local habitat at the current location of LA2931 and LA4330 was likely unsuitable for the establishment of the most southern highland population until 15 kya (after the LGM, *i.e.* during MIS1-2). The lower genetic diversity of the south populations (and estimated *Ne*) is thus due to the mild bottleneck processes during the southward expansion (Figure 2A; Figure S2; Table S3). Both south-coast populations show consistent signals of long-term history of colonization, subsequent isolation with negligible gene flow, and possible local specialization to sparsely suitable Lomas habitats along the coast (Figure 2B and 2C, Figure S2).

The divergence between the central group populations (during MIS3-4) occurs after the colonization of the coastal habitat (Figure S3), but before the colonization of the south-highland (LA4330). Moreover, strong post-divergence gene-flow and low differentiation are found in the central group, especially among the pairs LA1963-LA3111 and LA3111-LA2931 (Figure S2), consistent with their geographical and environmental proximity (Figure 1B) and the range contraction during the LGM (MIS2 in Figure 2). The colonization of high-altitude regions in the central group is thus accompanied by high levels of gene flow despite these populations ranging across a large altitudinal gradient (2500m of altitude difference between LA1963 and LA3111 or LA2931). The divergence history results in the south-coast and south-highland populations to be fairly isolated from one another (as separated by the Atacama desert) leading to the suggestion of an incipient speciation process (Figure 2B; Figure S2)[28]. We want to assess if and how many selective sweeps underpin such adaptation to novel habitats (colonization of the coastal and highlands) and to temporally variable climatic conditions (past climatic conditions).

Selective sweeps underpin local adaptation

To detect candidate genes under positive selection we scan the genomes for selective sweep signals using both an SFS-based method (site-frequency spectrum based) SweeD, and an LD-based method (linkage disequilibrium based) OmegaPlus. To disentangle outlier regions putatively under selection from the neutral expected distribution, we set conservative threshold values for the SweeD and OmegaPlus statistics obtained from coalescent simulations with varying rates of recombination based on the results of the demographic inference (Table 1; Figure S3). The overlapping genomic regions under selection between the two methods are then considered as the candidate regions under selection. In total, we find 2,921 regions with SweeD (mean size 212,858bp +/- 3,938) and 13,106 with OmegaPlus (mean size 59,618 bp +/- 521) across all six populations, yielding a total of 520 overlapping regions (mean size 41,082 bp +/- 1,618). Within those regions we find a total of 799 protein-coding candidate genes detected as being likely under positive selection (Data S1C). Among them, the largest number of candidate genes (354) is detected in LA4330 (Table 1), likely because the population 1) has been established recently (Figure 2B), 2) has a lower population size and thus higher amount of linkage disequilibrium so that selective sweeps are easier to detect[29] (Figure S4C), and 3) its habitat is very different from the rest of the species range (Figure 1B). In LA4107, we find 61 candidate genes and about 100 candidate genes were detected in each of the other four populations (Table 1). However, only a few candidate genes are shared among different populations, with the central and south-highland populations sharing a small number of candidate genes, while almost none are shared between the two south-coastal populations (Figure S3C). This lack of common candidate genes among populations is likely due to 1) the high nucleotide diversity and high effective population sizes generating new variants across many genes which are then differentially picked up by selection across different populations, and 2) the relatively old inter-population divergence and timing of local adaptation.

Our candidate regions exhibit typical characteristics of positively selected regions when compared to the genome-wide statistics, and do not show any peculiar pattern of recombination (or lack thereof) which could invalidate our scan results (see SI Text 2; Figure S4A-D; Table S3 and S4). Furthermore, we find an overlap between our candidate genes under selective sweep and genes exhibiting signals of positive selection in previous studies in *S. chilense*. It is noteworthy that these previous studies were based on few chosen genes, different plants, different populations and different sample sizes than ours. Among our candidate genes, we indeed find three genes (*JERF*3, *TPP* and *CT*189) involved in abiotic stress tolerance such as salt, drought or cold[19] as well as three NLRs (nucleotide binding leucine rich repeat, SOLCI006592800, SOLCI001535800, SOLCI005342400) possibly linked to resistance to pathogens[20]. We also find that two of the seven most up-regulated genes under cold conditions in a transcriptomic study of *S. chilense*[21] do appear in our selection scan in high altitude populations: *CBF*3 (Solyc03g026270) in LA2931, and *CBF*1 (Solyc03g026280) in LA4330. These overlapping results indicate that our genome-wide selective sweep scan generalizes the previous studies of selection in *S. chilense* and support the functional relevance of our candidate genes.

We use the function integrated in the program McSwan to detect sweeps and compute their age, that is the time of the appearance of the selected allele (Figure 3). Conservatively, we select only sweeps detected with McSwan, SweeD and OmegaPlus (Table 1). A total of 112 sweep regions (mean size 22,171 bp +/- 3,240) overlap with our candidate regions from SweeD and OmegaPlus, containing 175 genes and with a mean sweep age of ca. 61,000 years. The ages of sweeps range from as early as 150 kya to 5 kya, and only the highland populations exhibit sweeps more recent than 15 kya, consistent with recent (re)colonization of higher altitudes (Figure 3). Contrary to previous observations in other species (see introduction), we find remarkably old sweep ages. We explain this result by the large effective population sizes (per population in Figure 2, but also species wide[19,30]), mild

bottlenecks during colonization (Figure 2A) and the presence of strong seed banks[30]. Persistent seed banks indeed decrease the effect of genetic drift[31] and allow to retain for longer periods of time the footprints of selection[32]. We then further investigate the function of our candidate genes.

Gene regulatory networks underlying local adaptation in *S. chilense*

A Gene Ontology (GO) enrichment analysis of the 799 candidate genes reveals significantly different and common biological processes across populations. We find common GO categories to all populations such as basic cell metabolism, immune response, specific organ development and response to external stimuli (Figure S5). Most interesting, however, are four GO categories specifically found in several populations and link to local differences in habitats: (i) root hair cell differentiation functions are enriched in 15 candidate genes, only in the three coastal populations (LA1963, LA2932 and LA4107); (ii) response to circadian rhythm, photoperiodicity and flowering time are enriched in 12 candidate genes in two high-altitude (LA3111 and LA4330) and a south-coast (LA2932) populations; (iii) vernalization response is enriched in eight candidate genes in the three high-altitude populations (LA2931, LA3111, LA4330), and (iv) protein lipidation is enriched in seven candidate genes in the south-highland population (LA4330). Based on the wealth of available data in cultivated tomato and *S. pennellii* (with a well annotated genome) as well as gene ontology in *A. thaliana*, we further study the gene regulatory networks to which the candidate genes belong.

For adaptation to high-altitude conditions, 15 candidate genes are found interconnected in a flowering gene network, which is itself sub-divided into two sub-networks related to flowering, photoperiod and vernalization control pathways (Figure 4A; Data S1D). Photoperiod responsive genes can sense changes in sunlight and affect the circadian rhythm to regulate plant flowering[33,34], while vernalization genes regulate the flowering and germination through long-term low temperature [35–37].

These two sub-networks are connected through several key genes, some of which appearing as candidate genes entailing local adaptation in our populations: FL FLOWERING LOCUS C (FLC or AGL25), FLOWERING LOCUS T (FT) and AGAMOUS-LIKE genes (AGL) (Figure 4A and 4B). These key genes are essential regulators acting on the flowering regulation pathway[38–41]. Remarkably, some candidate genes in the recently diverged south-highland population (LA4330) aggregate into an independent network involved in circadian rhythm regulation, connected to the photoperiod network by JUMONJI DOMAIN CONTAINING 5 (JMJD5) also a candidate gene in LA3111 (Figure 4A). In the central-highland population (LA3111) several other candidate genes of the photoperiod network also regulate circadian rhythm and flowering time. Interestingly, the three high-altitude populations (LA3111, LA2931, and LA4330) present candidate genes of the AGAMOUS-LIKE (AGL) gene family in the vernalization network (Figure 4A). We also note that the network of protein lipidation genes appears to be related to the synthesis of fatty acids in the south-highland population (Figure 4D; Data S1D). We speculate here that this latter adaptation may be related to adaptation to lowest-temperature stress of LA4330 (Data S1E)[42,43]. Adaptation to high altitude involves the regulation of the flowering, including photoperiod and vernalization pathways, but through different genes in different populations, while cold stress and its consequence (adaptation in lipidation pathway) may be relevant for adaptation to the highest altitudes.

Regarding adaptation to coastal conditions, we find 11 candidate genes related to root development and cellular homeostasis functions clustered in a single network (Figure 4C; Data S1D). We speculate that the drought and water shortage typical of the coastal conditions (Data S1E) would promote the differentiation and extension of plant roots[44,45]. The cell WALL ASSOCIATED LINASE 4 (WAK4), a candidate gene identified LA4107, is shown to act as linker of signal from the cell wall to plasma membrane and thus serve a vital role in lateral root development[46,47]. It is also

possible that the soft soil at the seaside is also beneficial to the extension of root hairs to increase the water absorption area[48,49]. In addition to root development, we find also genes involved in cell homeostasis (Figure 4C; Data S1D), which would be critical for the coastal drought and salinity conditions to maintain stability of intracellular environment in the coastal habitats[50,51].

Candidate genes show genotype-environment associations to local climatic conditions

Our candidate loci are hypothesized to be responsible for adaptation to local climatic conditions, so we test for genotype-environment association (GEA) using redundancy analysis (RDA). RDA tests for correlation between allele frequencies and climatic variables and can be used 1) to search more finely for specific selective variants within the genomic sweep regions, and 2) reveal selective gradients defined by multiple environmental variables[52,53]. We here perform first a “present day” RDA using 144,713 SNPs from all candidate regions and 63 climatic variables representing current (present) conditions for temperature, precipitation, solar radiation, and wind (Data S1E). We find that the two-first RDA axes are significant (ANOVA's $p < 0.001$) and retain most (38% and 21%) of the putative adaptive genetic variance identified in the genome scans in all populations (Figure 5A). Table S5 and S6 summarize outlier SNPs in different RDA models and their correlation with climatic variables. In concordance to the PCAs of both climatic and genomic variation (Figures 1B and 1C), the two main RDA axes cluster the individuals into three groups corresponding to the main geographical regions (central, south-highland, and south-coastal) supporting that those axes synthetize the principal selective pressures for spatial local adaptation along the species distribution (Figure 5A). RDA1 represents the differentiation of the two south-coast populations in correlation with higher precipitation of coldest quarter (Bio19) and annual variation of solar radiation (CV_R). There are 480 and 501 SNPs strongly associated with these two climatic variables, respectively (Table S6). RDA2 summarizes a climatic gradient differentiating the

south-highland population manly driven by annual potential evapotranspiration (annualPET) and temperature annual range (Bio7) with 1,184 and 372 strongly associated SNPs, respectively (Table S6).

Further RDA analyses based on gene variants of the GO categories circadian rhythm-photoperiodism, vernalization, root-hair differentiation, and protein lipidation highlight combinations of climatic variables and genetic variants related to local spatial adaptation (Figure S6A, 6C, 6E and 6G). These analyses show that about 40% of the variation is explained by the two main RDA axes. Climatic variables representing temperature variability through the year such as temperature seasonality (Bio4) and temperature annual range (Bio7) are consistently correlated with adaptive variation of the south-highland population (Figure 5A; Figure S6). A total of 68 SNPs within candidate genes of the population LA4330 are strongly associated to these two variables (Bio4, Bio7) in three of the RDA based on the GO categories (circadian rhythm-photoperiodism, vernalization, and protein lipidation; Data S1F). The RDA based on the root-hair differentiation GO category shows a strong differentiation between lowland and highland populations, with stronger correlation for the southern populations (Figure S6E). This differentiation correlates with atmosphere water vapor availability variables such as annual minimum water vapor pressure (ann_Vmin) and annual actual evapotranspiration (ann_AET) variables for the coastal populations.

Considering that the sweep age estimations indicate that a fraction of the adaptive variation currently observed in *S. chilense* may have originated from the influence of past climate change processes (Figure 3), we implement an “LGM” RDA using 37 climate variables projected to the Last Glacial Maximum conditions (Figure 5B; Figure S6B, 6D, 6F and 6H; Data S1E). This analysis implies the assumption that the species did not shift its geographical range following the late Pleistocene climatic oscillations, but rather underwent a shift in its niche while retaining large portions of

its distribution[54,55]. This process is especially relevant for the highland regions where past distribution projections show no suitable habitat, while population persistence being likely since the demographic reconstruction indicates 1) only mild population size reductions (Figure 2C), and 2) pre-LGM divergence of the highland populations (LA3111 and LA2931 during MIS3-4 climatic periods, Figure 2B). Therefore, we expect that this analysis can uncover additional genomic variation selected in response to temporal climatic changes and underlying the niche expansion towards highly differentiated conditions in the south-highland region (Figure 1B). RDA analyses using LGM climatic variables capture a smaller proportion of the genetic variability in the first two constrained axes (30%) compared to that using the current climatic variables. We find that some outlier/candidate SNPs appear not to show any correlation with the past climatic variables, but about 30% outlier SNPs are identified in additional genomic regions correlated with past climatic variables and not with the current variables (Table S5 and S6). The analyses also emphasize the differentiation of central, south-highland, and south-coast populations. Indeed, we find a different RDA for vernalization candidates using the two sets of (present and past) climatic variables. For example, the central populations LA3111 and LA2931 are separated in the past RDA of vernalization using LGM climatic variables indicating warmer climate after LGM may drive gene flow among central populations as seen in the current RDA (Figure 2B; Figure S6C and S6D; Figure S3). The past RDA of LGM climatic variables unveils that high-altitude populations, especially LA4330, have SNPs correlating with temperature (*i.e.* annual mean minimum temperature; ann_mTmin, and temperature annual range; Bio7) whereas coastal populations SNPs do correlate with precipitation and potential evapotranspiration of the coldest and driest seasons (Figure 5B).

Discussion

By correlating selective sweeps at candidate genes to current and past climatic data, we demonstrate that local spatial adaptation to novel habitats and temporal

adaptation to changing conditions are intermingled. Nevertheless, our results show that it is possible to separate the underpinnings of the spatial and temporal adaptation in a species with large effective population size using niche modeling and past climatic reconstruction. As a word of caution, we note that our discussion focuses on four main GO categories which can be reliably associated with physiological traits likely underlying adaptation (though pinpointing the regulatory or coding SNPs under selection was not possible with our sample sizes), while functional information on many candidate genes in Figures 3 and S5 is still lacking to provide a complete picture. Yet, we are confident that our candidate genes under selection are functionally relevant, as demonstrated by the overlap with previous studies focusing on few genes involved in abiotic and biotic stress tolerance and transcriptomic data under cold conditions[19–21].

For each of the five past climatic periods depicted in Figures 2 and 3, we show different temporally restricted bursts of sweeps for either the south-highland or south-coastal groups of populations. When overlaying specifically the age of the four gene network categories of candidate genes (photoperiod, vernalization, protein lipidation, and root hair differentiation), we find that the sweeps within a given category/gene network are found at the time of a given particular climatic period (Figure 3). This demonstrates that selective sweeps occur following temporal adaptation in response to variable climatic conditions, as suggested under the adaptation to a moving optimum theory[1]. We observe that during the colder climate phases (MIS2 at 30-15kya), the suitable areas of the species decreased to a narrow belt along lower altitude areas (Figure 2C). The persistence of populations at higher altitude would have been possible by niche expansion driven by selection at genes involved in the vernalization gene networks (Figure 3). We also observe a correlation between the age of the sweeps and our “LGM” RDA analysis based on the past climatic data. Therefore, we suggest the photoperiod and protein lipidation adaptations to have occurred before the LGM (Figure 3) and to be instrumental pre-

adaptations favoring subsequently the colonization of new habitats in the southern highlands (after the LGM). During this most recent colonization event, sweeps occurred at other genes, including the complete photoperiod gene sub-network in LA4330 (Figure 4A). The mentioned pre-adaptation sweeps could have arisen in the ancestral populations from the highland-central region and by gene flow have facilitated the recent expansion of the species' geographic range[56]. We do not detect these sweeps in the central-highland populations probably because these haplotypes can be broken by the combined effect of recombination (more efficient in central populations with higher N_e) and gene flow. In addition, adaptations in the photoperiod and protein lipidation gene networks in the highland populations are adaptations older than the LGM, presumably due to earlier colonization of higher altitude habitats under the warmer climate of the period MIS3-4 (Figure 3). Therefore, we suggest that temporal adaptation does constrain or facilitate spatial local adaptation at later time steps. The respective contributions of spatial and temporal can be disentangled by comparing the significant genes (and SNPs) found in the past and present RDA analyses.

The *S. chilense* lineage likely originated from coastal habitats in southern Peru. This explains the early divergence and southward colonization process, accompanied by habitat fragmentation and contraction, which yields two highly isolated populations (Figure 2C). The coastal colonization seems to involve fewer sweeps than the adaptation to higher altitudes, for example a burst of selective sweeps in genes related to root anatomical traits during the LGM period (Figures 3, 4C, S6E-F). However, some of the adaptive genomic signals in the coastal populations could be blurred due to the older timing or stronger drift (due to habitat fragmentation along the coast), or be incomplete or partial sweeps which we do not detect with our current scans.

We tentatively suggest some answers to the questions highlighted in the introduction on the basis of our empirical results. We find between 60 and 350

selective sweeps per population, but contrary to our naïve expectations, sweeps have a large variance in age. Indeed, our initial assumption was that most sweeps in the south populations would be recent and occurring during only the recent spatial local adaptation to a new habitat. On the contrary, the resulting picture of local adaptation we observe is that of a complex process involving several discrete adaptive steps, each due to temporally restricted burst of selective sweeps within a given gene network. This adaptive process can be traced back using a combination of sweep age estimates, RDA analyses and past and present climatic models. We suggest that several sweeps do occur concomitantly in a given gene pathway/network at a given time period, possibly as a temporal adaptation to a moving environmental optimum (our climatic periods). Thereafter, these sweeps in a given gene (sub-)network may lay the foundations, in other words, provide exaptation[57,58], for colonizing new habitats at future climatic periods when new ecological niches/habitats become available. The enabled future spatial adaptation seems to involve selection at genes in different gene networks (possibly improving upon and fine-tuning the previous phenotypic adaptation). Note that we focus here on selective sweeps resulting from strong positive selection as we cannot assess in our data the occurrence of weaker positive or polygenic selection. It appears that few gene networks seem to be under strong selection at any climatic period, as predicted under polygenic selection model of adaptation to a moving optimum[1,3,4,16]. We speculate further, with the aim to broaden the current theoretical framework, that strong selection underlying temporal adaptation to a moving optimum would occur at one (or few) phenotypic trait, while local spatial adaptation to a new optimum requires strong selection at several traits.

To our knowledge, this is the first study to attempt to disentangle the complex spatial and temporal processes involved in colonization to new highly stressful environments such as hyper-dry high-mountain regions. In the wild tomato *Solanum chilense*, the recruitment of multiple genes belonging to gene networks responsible

for regulating the circadian cycle and flowering time is evident. This divergent selection process arises in response to recent climate change processes. The detectability of this polygenic pattern of selection is favored by the recent divergence of the populations involved, persistence of high genetic variability due to mild bottleneck during the expansion process and presence of seed banks, as well as the use of a combination of genome scans and availability of past climatic data. We demonstrate the power to use past climatic reconstruction with scans for selection to detect the underpinnings of local and temporal adaptation, when old selective sweeps can be recovered. Specifically, adaptation in space and time occurs over short time windows during which few genes of a given gene network are strongly selected upon, consistent with a polygenic model of adaptation to a moving optimum.

Acknowledgments

AT acknowledges funding from DFG (Deutscheforschungsgemeinschaft) Grant Number: 317616126 (TE809/7-1); KW was funded by the Chinese Scholarship Council, and GSA funded by the Technical University of Munich. We thank the Tomato Genetics Resource Center (TGRC) of the University of California, Davis for generously providing us with the seeds of the accession included in this study. We thank Daniela Scheikl for assistance with plant work and Christine Würmser for the Illumina sequencing.

Authors contributions

Conceptualization, GSA and AT; Software, Formal analysis and Investigation, KW and GSA, Writing KW, GSA and AT; Funding Acquisition, AT; Supervision, GSA and AT.

Declaration of interests

The authors declare no competing interests

Resource availability

Lead contact

Further information on materials, datasets, and protocols should be directed to and will be fulfilled by the Lead Contact, Aurélien Tellier (tellier@wzw.tum.de).

Data and code availability

The raw pair-end sequencing genomic data used in all the analyses can be accessed at the European Nucleotide Archive (ENA) project accession PRJEB47577.

All code used in this study and other previously published genomic data is available at the sources referenced in the key resource table.

Figure legends

Figure 1. Geographic distribution and population structure of *S. chilense.* (A) Map with distribution of all *S. chilense* populations by the TGRC, the six *S. chilense* populations in this study (black circles), and the four population groups (colorful circles). (B) The PCA of all populations using 63 current climatic variables (Data S1E). Population structure using SNP data based on (C) PCA and (D) Admixture analysis (optimal K value is 4, Figure S1D).

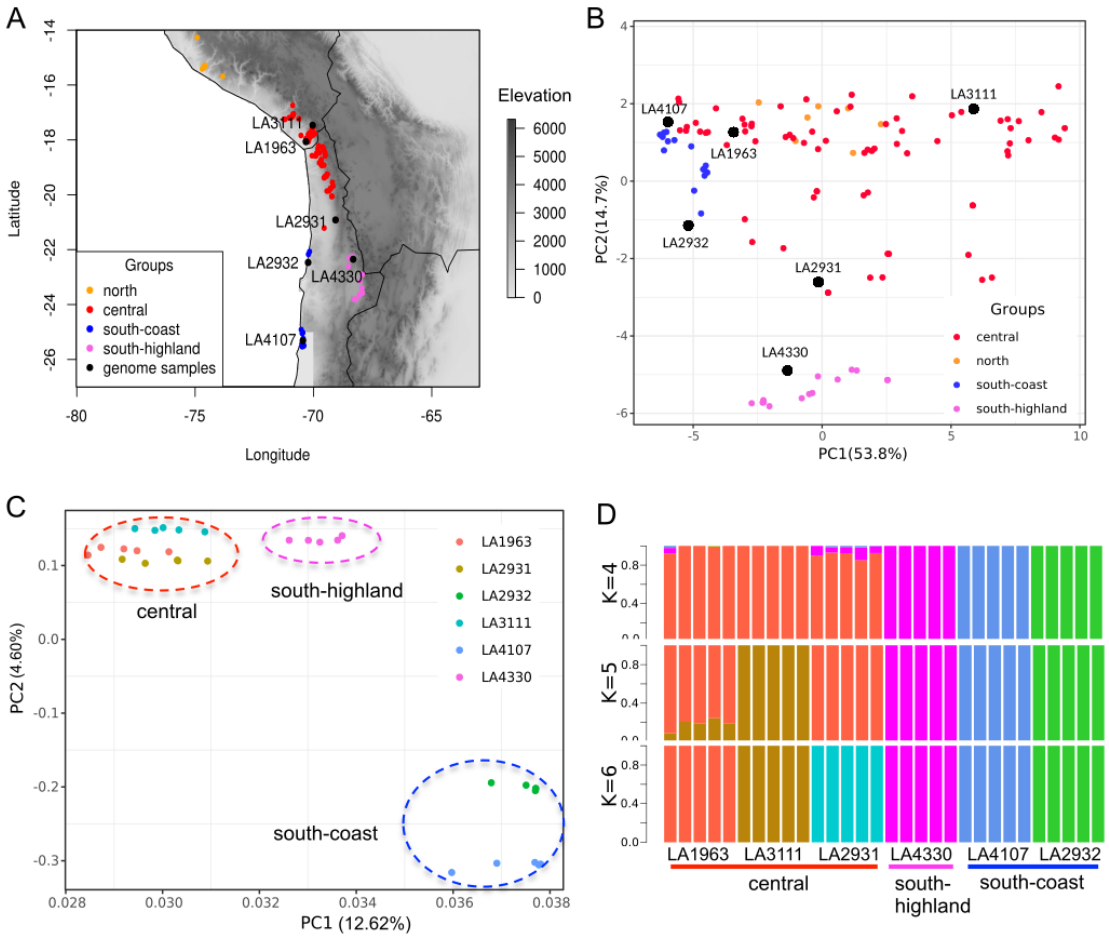


Figure 2. Demographic history and species distribution model of *S. chilense* for current and Last Glacial Maximum (LGM) climate conditions (A) The estimation of historical patterns of effective population size (N_e) for 10 pairwise genome comparisons per population using the MSMC model. (B) Interpreted demographic scenario for the six samples populations of *S. chilense* including likely estimations of effective population size, divergence times and gene-flow. The width of the boxes represents the relative effective population size, arrows represent the migration between population pairs. Grey background boxes indicate five Marine isotope stages (MIS) climatic periods. (C) Overlay of the reconstruction of the distribution model for *S. chilense* using current climatic variables (red) and LGM past climatic variables (blue). Darker color of the gradient indicates higher suitable habitat for a given climatic period.

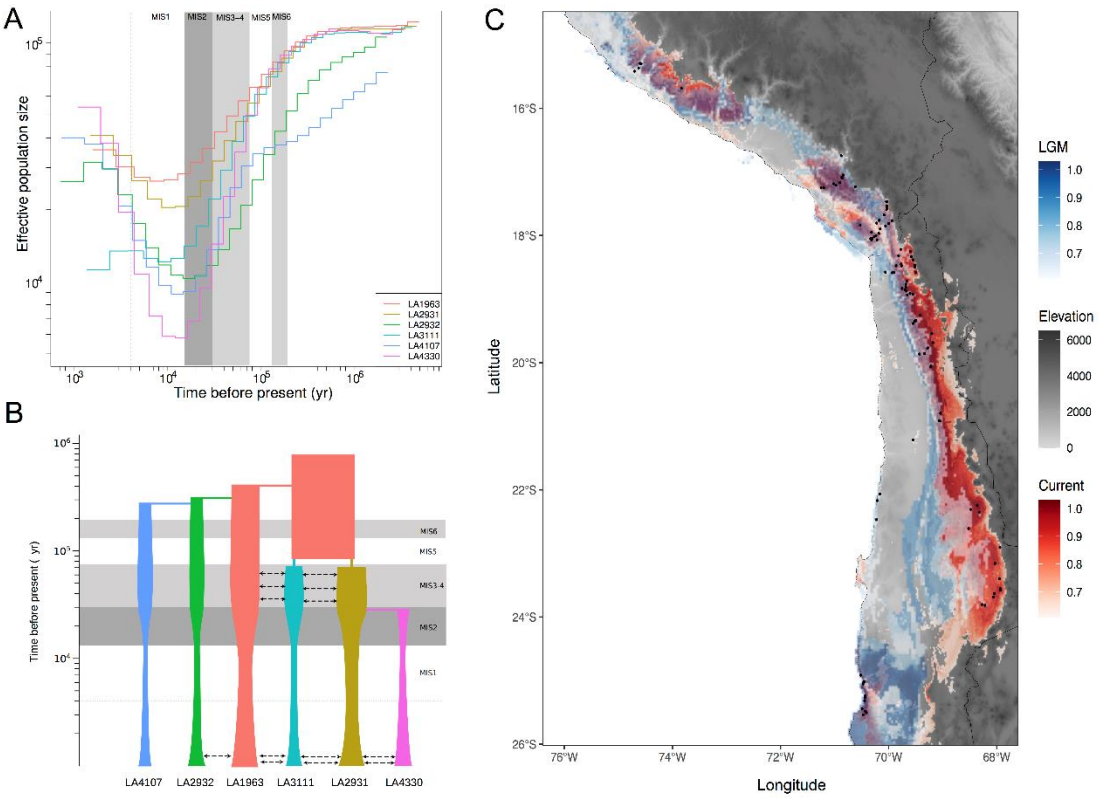


Figure 3. Distribution of sweep age across the five MIS climatic periods.

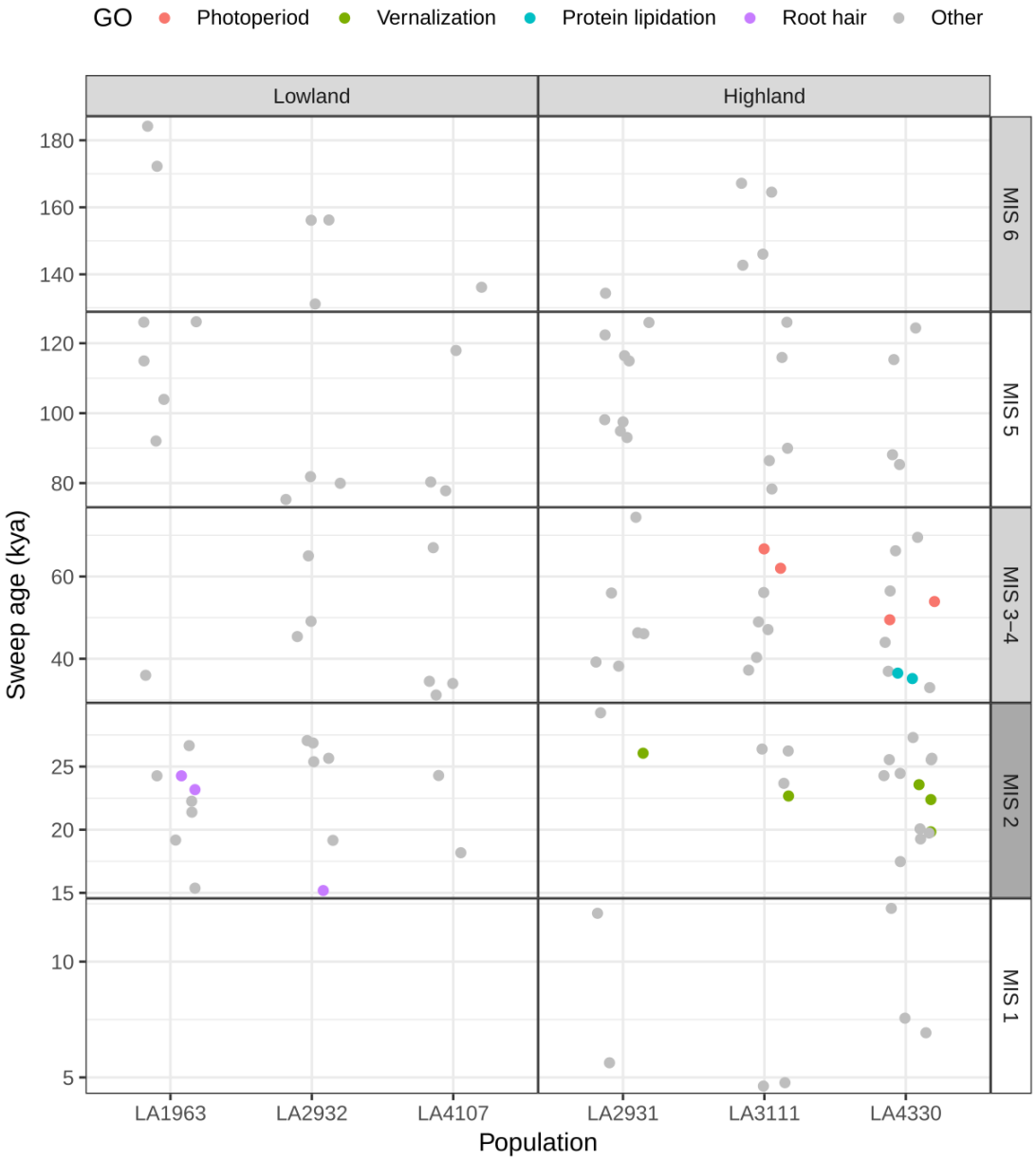
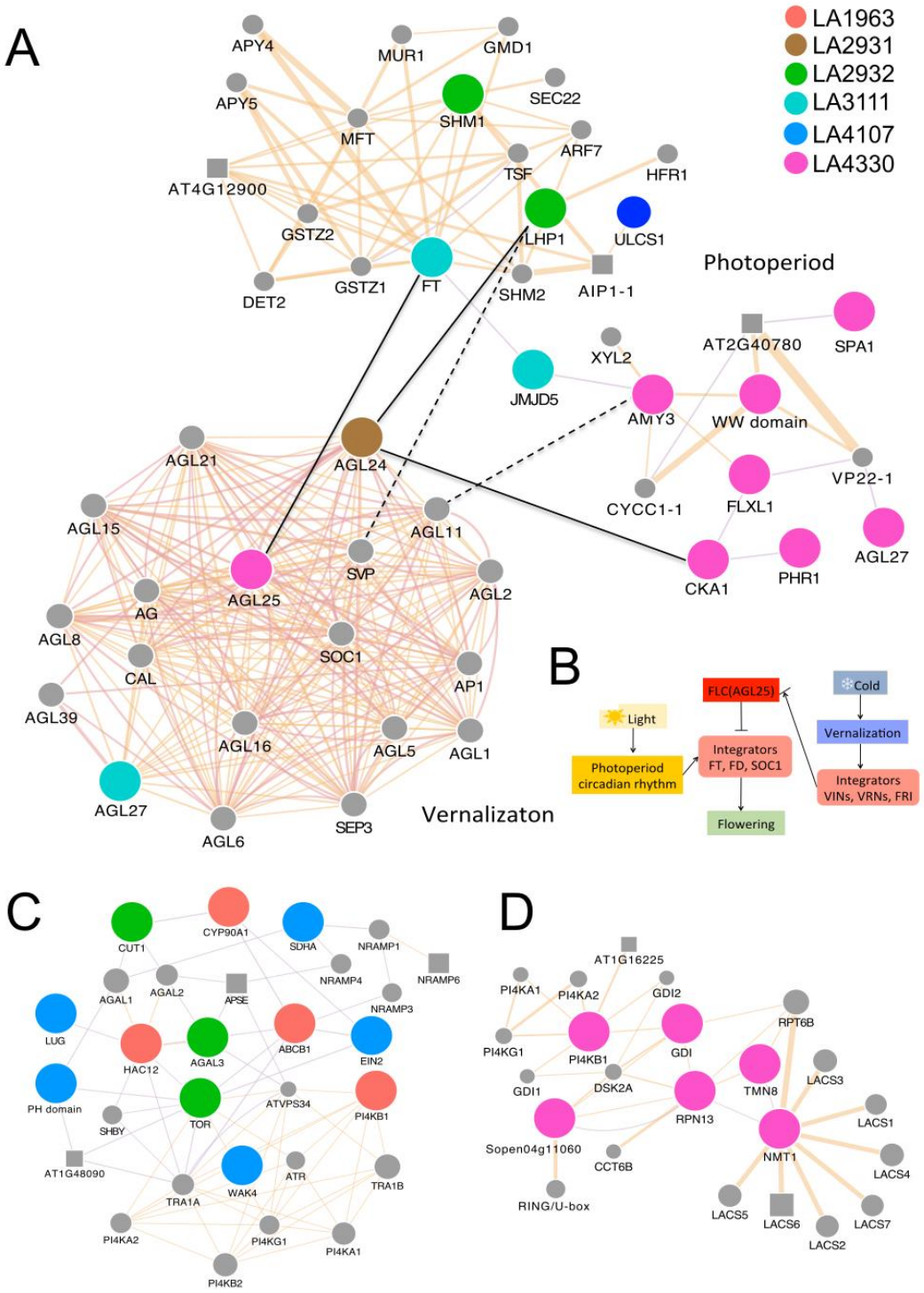


Figure 4. Interaction genetic networks of candidate genes. (A) The network of flowering regulation involved two sub-networks, photoperiod and vernalization pathways, for regulation of flowering. (B) The schematic diagram of flowering regulation involved photoperiod and vernalization is adapted from [36]. "┤" indicates repressive effects on gene expression; "→" indicates promotive effects on gene expression. (C) The network of root development and cell homeostasis. (D) The networks of protein lipidation. Connections represent gene interactions based on physical interactions, informatics predictions and co-expression analyses. Connection thickness is proportional to weighted value of the connected genes. The black lines connected two sub-networks, genes under selection were connected by solid lines and other genes were connected by dashed line. Node colors correspond to genes were detected the different populations in genome scans. Gray circles, not detected in genome scan, but present in *S. chilense*; gray squares: not present in *S. chilense*.


A
LA1963
LA2931
LA2932
LA3111
LA4107
LA4330
APY4
MUR1
GMD1
APY5
SEC22
SHM1
MFT
ARF7
TSF
HFR1
AT4G12900
GSTZ2
LHP1
ULCS1
FT
GSTZ1
SHM2
AIP1-1
DET2
Photoperiod
AT2G40780
SPA1
XYL2
JMJD5
AMY3
WW domain
AGL21
AGL24
VP22-1
CYCC1-1
FLXL1
AGL15
AGL11
AGL27
SVP
CKA1
PHR1
AGL25
AGL2
AG
AGL8
SOC1
CAL
AP1
AGL39
AGL16
AGL5
AGL1
AGL27
SEP3
Vernalizaton
AGL6
B
Light
FLC(AGL25)
Cold
Vernalization
Photoperiod circadian rhythm
Integrators FT, FD, SOC1
Integrators VINs, VRNs, FRI
Flowering
C
CUT1
CYP90A1
SDHA
NRAMP1
NRAMP4
NRAMP6
AGAL1
AGAL2
APSE
NRAMP3
LUG
HAC12
AGAL3
ABCB1
EIN2
PH domain
TOR
ATVPS34
PI4KB1
SHBY
AT1G48090
WAK4
ATR
TRA1B
TRA1A
PI4KA2
PI4KG1
PI4KA1
PI4KB2
D
AT1G16225
PI4KA1
PI4KA2
GDI2
RPT6B
PI4KG1
PI4KB1
GDI
GDI1
DSK2A
TMN8
LACS3
RPN13
LACS1
Sopen04g11060
CCT6B
NMT1
LACS4
RING/U-box
LACS5
LACS6
LACS7
LACS2

561

Figure 5. Redundancy analysis (RDA) ordination bi-plots between the climatic variables, populations and the genetic variants in all candidate sweeps. (A) RDA using current climatic variables. (B) RDA using LGM climatic variables. Arrows indicated the direction and magnitude of variables correlated with the populations. The abbreviations of climatic variables are shown in Data S1E.

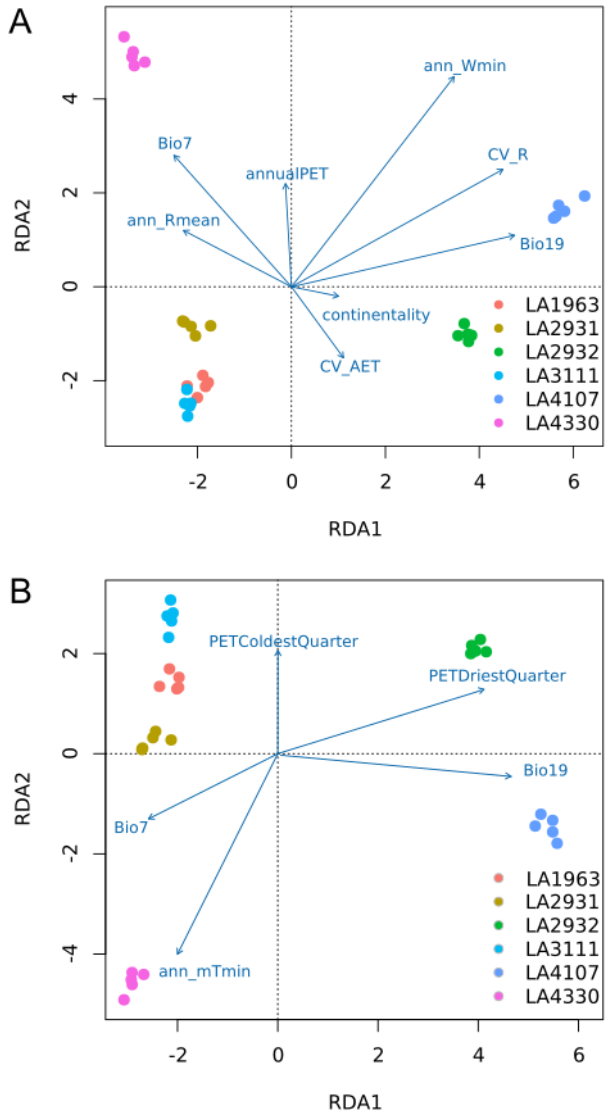


Table 1. Summary of genome scan results and estimation of sweep age.

Populations	Genome Scans				Sweeps age			
	N_{SweeD}	$N_{OmegaPlus}$	$N_{overlaps1}$	N_{genes1}	N_{McSwan}	$N_{overlaps2}$	N_{genes2}	Age_{mean} (kyr)
LA1963	385	2,474	98	86	267	16 39	14	71±55
LA2931	517	2,268	109	125	355	19	26	62±41
LA2932	374	1,717	46	101	302	15	29	65±48
LA3111	663	2,307	105	107	377	22	22	60±51
LA4107	203	2,047	37	61	194	10	10	63±40
LA4330	779	2,293	125	354	438	39	74	41±30

N_{SweeD}, number of outlier regions from SweeD; $N_{OmegaPlus}$, number of outlier regions from OmegaPlus; $N_{overlaps1}$, number of overlapping regions between SweeD and OmegaPlus; N_{genes1}, number of candidate genes in overlaps1, and all candidate genes show in Data S1C; N_{McSwan}, number of outlier regions from McSwan; $N_{overlaps2}$, number of overlapping regions between McSwan and overlaps1; N_{genes2}, number of genes in overlaps2; Age_{mean}, mean age of overlaps2.

STAR Methods

Sample collection, sequencing and bioinformatics

Plants were grown from individual seeds obtained from the Tomato Genetics Resource Center (TGRC, University of California, Davis, USA; http://tgrc.ucdavis.edu). We sampled five plants from accessions LA1963, LA3111, LA2931, LA4330, LA2932, LA4107; Table S1) which were grown in individual pots in standard glasshouse conditions at the Dürnast plant research facilities of the TUM School of Life Sciences. Genomic DNA was extracted using the DNA extraction kit from Quiagen and sequenced on an Illumina HiSeq 2500 with standard library size of 300 bp (Eurofins Genomics, Germany). The 30 *S. chilense* whole genome sequencing data are available on ENA in BioProject PRJEB47577.

Quality control of the raw reads was performed with FastQC (v0.11.6)[59]. Base calls with insufficient quality or showing adapter contamination were removed using Trimmomatic (v0.36) using default parameters[60]. The high-quality clean reads were mapped to the *Solanum pennellii* reference genome[61] available from Solanaceae Genomics Network using Burrows-Wheeler Alignment tool (v0.7.16) with default settings [62] and afterward sorted with Samtools (v1.5) [63]. The raw alignments were then processed to add read groups, mark duplicates and fix mates using GATK (v3.3) suite[64] and PicardTools (v2.10.6)[65]. Variant calling was performed using the HaplotypeCaller tool of GATK with default parameters for each sample and individual genomic variant files were then combined into a variant matrix with the GenotypeGVCFs tool. Finally, after filtering, the variants were annotated using the snpEff software (v4.3) with default settings[66]. For each SNP, the annotation were identified based on the gene annotation of the *S.pennellii* reference. For details on commands and parameters used see Supplementary SI text 1.

Population genetics analyses

For all population genetics analyses we use *S. pennellii* population LA716 as outgroup. The phylogenetic tree was constructed by the maximum likelihood (ML) method using SNPhylo pipeline (v20160204) based on all SNPs[67]. The reliability of each branch was evaluated by bootstrapping with 1,000 replicates and the tree was visualized with iTOL (v5.0)[68]. A principal component analysis (PCA) was performed for the whole SNP dataset using GCTA (v1.91.4) [69]. The inference of population structure was performed using the program ADMIXTURE (v1.3.0)[70]. Nine scenarios (ranging from K = 2 to K = 10) were assessed for genetic clustering.

Population genetics statistics, nucleotide diversity (π), Tajima's D and F_{ST}, were calculated with VCFtools (v1.15)[71] for each population (or pairs of populations) over 100 kb sliding non-overlapping windows. Watterson's estimator of theta (θ_w) and Fay and Wu's H were calculated using ANGSD (v0.921)[72]. The linkage disequilibrium (LD) levels were calculated per population as the genotype correlation coefficient (r^2) between two loci (within the same chromosome) using VCFtools with a maximum distance of 1,000 kb. LD decay was then assessed for all pairs of SNPs within 1,000 kb distance using PopLDdecay (v3.27)[73].

Inference of demographic history

The demographic analysis was conducted using the Multiple Sequentially Markovian Coalescent method (MSMC2) with phased VCF files and 40 hidden states[74]. The cross-coalescence analysis was performed for each pairwise comparison of genomes between pairs of populations to estimate the population separation history and the migration rate with MSMC-IM[75]. The phased data were generated with SHAPEIT (v2.17)[76]. We assumed generation time of 5 years and a rate of 1×10^{-8} mutations per generation to scale time in the result plots.

Modeling present and past species distribution

We reconstructed and visualized the environmental space occupied by *S. chilense* extracting the environmental conditions at the occurrence points and summarize them by PCA[77]. The environmental data include 63 climatic layers obtained from three public databases WorldClim2[78], ENVIREM[79], and the Consultative Group on International Agricultural Research (CGIAR)[80] (Data S1E). The PCA was performed by the *prcomp* function in R (v4.0.3)[81].

We then performed an ensemble modeling framework[82] using the BIOMOD package[83,84] in R including all known localities (from TGRC database) covering the full extant geographic range of *S. chilense* using eight modeling algorithms, five cross-validation replicates, and ten pseudo-absence sampling sets, therefore completing a total of 400 models. The environmental predictors include five variables selected from those used in the climatic PCA. Variable selection was implemented by keeping low correlated predictors. We assessed the predictive performance of the models through the AUC and TSS statistics. Consensus niche models were obtained using a TSS-weighted average method to account for the predictive power of each fitted model. Models with low predictive power (TSS < 0.7) were discarded. All fitted suitability models were then projected to infer the distribution of the *S. chilense* suitable habitat under current and past conditions during the Last Glacial Maximum (LGM; ~21 Kya).

Genome-wide selection scans

We identified selective sweeps using biallelic SNPs by SweeD (v3.3.1)[85] and OmegaPlus (v2.3.0)[86]. The CLR statistics in SweeD were calculated with default parameters except for using a resolution of 10 kb intervals within each chromosome. OmegaPlus statistics (ω) were computed at 10 kb intervals. We specified a minimum window of 10 kb and a maximum window of 100 kb to be used for computing LD

values between SNPs, respectively. Outlier CLR and ω statistics indicative of a selective sweep are defined by comparison to the genome-wide distribution values and the thresholds are defined by coalescent simulations of the the inferred demographic history. We used the coalescent simulator SCRM (v1.7.3)[87] to generate 10,000 neutral datasets of 10 Mb based on the demographic history of each population. To reduce false positives, we set up simulations with varying recombination rate every 100 kb within each 10 Mb simulated block. We then extracted the overlap regions of the two methods using the genomic coordinates and those regions were regarded as high confident selective sweep regions. The age of sweeps was estimated using McSwan[88] (see the supplementary methods).

GO enrichment analysis and gene networks

Due to the lack of a complete gene function annotation database, we performed a BLASTX against the NCBI database of non-redundant proteins (nr) screened for green plants (e-value cutoff was 10^{-6}) and used Blast2GO to assign GO terms for each gene identified in the genome scan analysis[89,90] as well as performed a blast to the *A. thaliana* dataset TAIR10 separately to remove redundant terms[91]. The false discovery rates (FDR) were calculated to estimate the extent to which genes were enriched in given GO categories. P-values smaller than 0.05 were used as cutoff for a significant level of correlation. For each of the genes enriched in specific biological processes, we retrieved the interacting gene neighbors using GeneMANIA (v3.5.2)[92]. We generated aggregate interaction networks in GeneMANIA, based on physical interactions, predicted and co-expression. Finally, we performed hierarchical clustering and manually optimized the weighted value cutoff for displaying the gene network. The functions of networks were extracted form GeneMANIA outputs with FDR cutoff 0.05 (Data S1D).

Redundancy analysis (RDA) analysis

The RDA was then performed using the *rda* function from the vegan package implemented in R[93], modeling genotypes as a function of the same climatic predictor variables used for the niche reconstruction analyses, and producing constrained axes and representative predictors. The significance of RDA constrained axes was assessed using the *anova.cca* function and significant axes were then used to identify candidate loci ($P < 0.001$). Candidate loci were identified using 2.5 standard deviation as cutoff (two-tailed p-value = 0.012).

Supplementary information

Available as a single pdf with additional figures, tables and methods.

References

1. Polechová, J., Barton, N., and Marion, G. (2009). Species' range: Adaptation in space and time. The American Naturalist *174*, E186–E204.

2. Chevin, L.-M., Martin, G., and Lenormand, T. (2010). Fisher's model and the genomics of adaptation: Restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution *64*, 3213–3231.

3. Jain, K., and Stephan, W. (2017). Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics *206*, 389–406.

4. Matuszewski, S., Hermisson, J., and Kopp, M. (2014). Fisher's geometric model with a moving optimum. Evolution *68*, 2571–2588.

5. Kim, Y., and Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics *160*, 765.

6. Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. *23*, 23–35.

7. Exposito-Alonso, M., Burbano, H.A., Bossdorf, O., Nielsen, R., and Weigel, D. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. Nature *573*, 126–129.

8. Aguirre-Liguori, J.A., Ramírez-Barahona, S., and Gaut, B.S. (2021). The

evolutionary genomics of species' responses to climate change. Nat Ecol Evol, 1–11.

9. Savolainen, O., Lascoux, M., and Merilä, J. (2013). Ecological genomics of local adaptation. Nature Reviews Genetics *14*, 807–820.

10. Fagny, M., and Austerlitz, F. (2021). Polygenic adaptation: Integrating population genetics and gene regulatory networks. Trends in Genetics *37*, 631–638.

11. Zou, Y.-P., Hou, X.-H., Wu, Q., Chen, J.-F., Li, Z.-W., Han, T.-S., Niu, X.-M., Yang, L., Xu, Y.-C., Zhang, J., et al. (2017). Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. Genome Biology *18*, 239.

12. Schrider, D.R., and Kern, A.D. (2017). Soft sweeps are the dominant mode of adaptation in the human genome. Molecular Biology and Evolution *34*, 1863–1877.

13. Garud, N.R., Messer, P.W., and Petrov, D.A. (2021). Detection of hard and soft selective sweeps from *Drosophila melanogaster* population genomic data. PLOS Genetics *17*, e1009373.

14. Stephan, W. (2016). Signatures of positive selection: From selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Molecular Ecology *25*, 79–88.

15. Li, H., and Stephan, W. (2006). Inferring the demographic history and rate of adaptive substitution in *Drosophila*. PLOS Genetics *2*, e166.

16. Jain, K., and Stephan, W. (2017). Modes of rapid polygenic adaptation. Molecular Biology and Evolution *34*, 3169–3175.

17. Barghi, N., Hermisson, J., and Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. Nat Rev Genet *21*, 769–781.

18. Nakazato, T., Warren, D.L., and Moyle, L.C. (2010). Ecological and geographic modes of species divergence in wild tomatoes. American Journal of Botany *97*, 680–693.

19. Böndel, K.B., Lainer, H., Nosenko, T., Mboup, M., Tellier, A., and Stephan, W. (2015). North–south colonization associated with local adaptation of the wild tomato species *Solanum chilense*. Molecular Biology and Evolution *32*, 2932–2943.

20. Stam, R., Silva-Arias, G.A., and Tellier, A. (2019). Subsets of NLR genes

show differential signatures of adaptation during colonization of new habitats. New Phytologist *224*, 367–379.

21. Nosenko, T., Böndel, K.B., Kumpfmüller, G., and Stephan, W. (2016). Adaptation to low temperatures in the wild tomato species *Solanum chilense*. Molecular Ecology *25*, 2853–2869.

22. Xia, H., Camus-Kulandaivelu, L., Stephan, W., Tellier, A., and Zhang, Z. (2010). Nucleotide diversity patterns of local adaptation at drought-related candidate genes in wild tomatoes. Molecular Ecology *19*, 4144–4154.

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

24. Böndel, K.B., Nosenko, T., and Stephan, W. (2018). Signatures of natural selection in abiotic stress-responsive genes of *Solanum chilense*. Royal Society Open Science *5*, 171198.

25. Stam, R., Nosenko, T., Hörger, A.C., Stephan, W., Seidel, M., Kuhn, J.M.M., Haberer, G., and Tellier, A. (2019). The *de novo* reference genome and transcriptome assemblies of the wild tomato species *Solanum chilense* highlights birth and death of NLR genes between tomato species. G3 *9*, 3933–3941.

26. Lisiecki, L.E., and Raymo, M.E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic δ18O records. Paleoceanography *20*.

27. Ritter, B., Wennrich, V., Medialdea, A., Brill, D., King, G., Schneiderwind, S., Niemann, K., Fernández-Galego, E., Diederich, J., Rolf, C., et al. (2019). Climatic fluctuations in the hyperarid core of the Atacama Desert during the past 215 ka. Sci Rep *9*, 5270.

28. Raduski, A.R., and Igić, B. (2021). Biosystematic studies on the status of *Solanum chilense*. American Journal of Botany *108*, 520–537.

29. Koropoulis, A., Alachiotis, N., and Pavlidis, P. (2020). Detecting positive selection in populations using genetic data. In (Humana, New York, NY), pp. 87–123.

30. Tellier, A., Laurent, S.J.Y., Lainer, H., Pavlidis, P., and Stephan, W. (2011). Inference of seed bank parameters in two wild tomato species using ecological and genetic data. Proceedings of the National Academy of Sciences *108*, 17052–17057.

31. Tellier, A. (2019). Persistent seed banking as eco-evolutionary determinant of plant nucleotide diversity: novel population genetics insights. New Phytologist *221*, 725–730.

32. Živković, D., and Tellier, A. (2018). All but sleeping? Consequences of soil seed banks on neutral and selective diversity in plant species. In Mathematical Modelling in Plant Biology, R. J. Morris, ed. (Springer International Publishing), pp. 195–212.

33. Song, Y.H., Shim, J.S., Kinmonth-Schultz, H.A., and Imaizumi, T. (2015). Photoperiodic flowering: time measurement mechanisms in leaves. Annual review of plant biology *66*, 441–464.

34. Johansson, M., and Staiger, D. (2015). Time to flower: interplay between photoperiod and the circadian clock. Journal of experimental botany *66*, 719–730.

35. Guo, X., Liu, D., and Chong, K. (2018). Cold signaling in plants: Insights into mechanisms and regulation. Journal of integrative plant biology *60*, 745–756.

36. Xu, S., and Chong, K. (2018). Remembering winter through vernalisation. Nature Plants *4*, 997–1009.

37. Iida, H., and Mähönen, A.P. (2020). Growth-mediated sensing of long-term cold in plants. Nature *583*, 690–691.

38. Michaels, S.D., and Amasino, R.M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. The Plant Cell *11*, 949–956.

39. Sheldon, C.C., Rouse, D.T., Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (2000). The molecular basis of vernalization: the central role of FLOWERING LOCUS C (FLC). Proceedings of the National Academy of Sciences *97*, 3753–3758.

40. Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of Florigen: FLOWERING LOCUS T moves center stage. Annual Review of Plant Biology *59*, 573–594.

41. Putterill, J., and Varkonyi-Gasic, E. (2016). FT and florigen long-distance flowering control in plants. Current Opinion in Plant Biology *33*, 77–82.

42. Jiang, H., Zhang, X., Chen, X., Aramsangtienchai, P., Tong, Z., and Lin, H. (2018). Protein lipidation: occurrence, mechanisms, biological functions,

and enabling technologies. Chemical reviews *118*, 919–988.

43. Maksimov, E.G., Mironov, K.S., Trofimova, M.S., Nechaeva, N.L., Todorenko, D.A., Klementiev, K.E., Tsoraev, G. V, Tyutyaev, E. V, Zorina, A.A., and Feduraev, P. V (2017). Membrane fluidity controls redox-regulated cold stress responses in cyanobacteria. Photosynthesis research *133*, 215–223.

44. Li, X., Cai, W., Liu, Y., Li, H., Fu, L., Liu, Z., Xu, L., Liu, H., Xu, T., and Xiong, Y. (2017). Differential TOR activation and cell proliferation in *Arabidopsis* root and shoot apexes. PNAS *114*, 2765–2770.

45. Xiong, Y., McCormack, M., Li, L., Hall, Q., Xiang, C., and Sheen, J. (2013). Glucose–TOR signalling reprograms the transcriptome and activates meristems. Nature *496*, 181–186.

46. He, Z.-H., Cheeseman, I., He, D., and Kohorn, B.D. (1999). A cluster of five cell wall-associated receptor kinase genes, Wak1–5, are expressed in specific organs of *Arabidopsis*. Plant molecular biology *39*, 1189–1196.

47. Lally, D., Ingmire, P., Tong, H.-Y., and He, Z.-H. (2001). Antisense expression of a cell wall–associated protein kinase, WAK4, inhibits cell elongation and alters morphology. The Plant Cell *13*, 1317–1332.

48. White, R.G., and Kirkegaard, J.A. (2010). The distribution and abundance of wheat roots in a dense, structured subsoil–implications for water uptake. Plant, cell & environment *33*, 133–148.

49. Bengough, A.G., Loades, K., and McKenzie, B.M. (2016). Root hairs aid soil penetration by anchoring the root surface to pore walls. Journal of Experimental Botany *67*, 1071–1078.

50. Forni, C., Duca, D., and Glick, B.R. (2017). Mechanisms of plant response to salt and drought stress and their alteration by rhizobacteria. Plant and Soil *410*, 335–356.

51. Zhao, C., Zhang, H., Song, C., Zhu, J.-K., and Shabala, S. (2020). Mechanisms of plant responses and adaptation to soil salinity. The innovation *1*, 100017.

52. Capblancq, T., Luu, K., Blum, M.G.B., and Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. Molecular Ecology Resources *18*, 1223–1233.

53. Forester, B.R., Lasky, J.R., Wagner, H.H., and Urban, D.L. (2018).

Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. Molecular Ecology *27*, 2215–2233.

54. Jezkova, T., Olah-Hemmings, V., and Riddle, B.R. (2011). Niche shifting in response to warming climate after the Last Glacial Maximum: Inference from genetic data and niche assessments in the chisel-toothed kangaroo rat (*Dipodomys microps*). Global Change Biology *17*, 3486–3502.

55. Sillero, N., Reis, M., Vieira, C.P., Vieira, J., and Morales-Hojas, R. (2014). Niche evolution and thermal adaptation in the temperate species *Drosophila americana*. Journal of Evolutionary Biology *27*, 1549–1561.

56. Kottler, E.J., Dickman, E.E., Sexton, J.P., Emery, N.C., and Franks, S.J. (2021). Draining the swamping hypothesis: Little evidence that gene flow reduces fitness at range edges. Trends in Ecology & Evolution *36*, 533–544.

57. Gould, S.J., and Vrba, E.S. (1982). Exaptation—a missing term in the science of form. Paleobiology *8*, 4–15.

58. Larson, G., Stephens, P.A., Tehrani, J.J., and Layton, R.H. (2013). Exapting exaptation. Trends in Ecology & Evolution *28*, 497–498.

59. Andrews, C. (2010). FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

60. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

61. Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., et al. (2014). The genome of the stress-tolerant wild tomato species *Solanum pennellii*. Nature Genetics *46*, 1034–1038.

62. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics *25*, 1754–1760.

63. Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

64. Auwera, G.A.V. der, Carneiro, M.O., Hartl, C., Poplin, R., Angel, G. del, Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastq data to high-confidence variant calls: The Genome

Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics *43*, 11.10.1-11.10.33.

65. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

66. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly *6*, 80–92.

67. Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics *15*, 162.

68. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Research *49*, W293–W296.

69. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics *88*, 76–82.

70. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research *19*, 1655–1664.

71. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., and Sherry, S.T. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

72. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC bioinformatics *15*, 1–13.

73. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics *35*.

74. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nature genetics *46*, 919–925.

75. Wang, K., Mathieson, I., O'Connell, J., and Schiffels, S. (2020). Tracking human population structure through time from whole genome sequences.

PLoS genetics *16*, e1008552.

76. Delaneau, O., Marchini, J., and Zagury, J.-F. (2012). A linear complexity phasing method for thousands of genomes. Nature methods *9*, 179–181.

77. Legendre, P., and Legendre, L. (2012). Numerical ecology Third English edition. (Elsevier).

78. Fick, S.E., and Hijmans, R.J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology *37*, 4302–4315.

79. Title, P.O., and Bemmels, J.B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography *41*, 291–307.

80. Trabucco, A., and Zomer, R.J. (2018). Global aridity index and potential evapotranspiration (ET0) climate database v2. CGIAR Consort Spat Inf.

81. R Core Team (2020). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

82. Araujo, M., and New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution *22*, 42–47.

83. Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M.B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. Ecography *32*, 369–373.

84. Thuiller, W., Georges, D., and Engler, R. (2014). BIOMOD2: Ensemble platform for species distribution modeling. R package version 3.1-62/r677.

85. Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. Molecular Biology and Evolution *30*, 2224–2234.

86. Alachiotis, N., Stamatakis, A., and Pavlidis, P. (2012). OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics *28*, 2274–2275.

87. Staab, P.R., Zhu, S., Metzler, D., and Lunter, G. (2015). scrm: Efficiently simulating long sequences using the approximated coalescent with recombination. Bioinformatics *31*, 1680–1682.

88. Tournebize, R., Poncet, V., Jakobsson, M., Vigouroux, Y., and Manel, S. (2019). McSwan: A joint site frequency spectrum method to detect and

date selective sweeps across multiple population genomes. Molecular Ecology Resources *19*, 283–295.

89. Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

90. Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. International journal of plant genomics *2008*.

91. Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. genesis *53*, 474–485.

92. Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., and Lopes, C.T. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research *38*, W214–W220.

93. Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O’Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2015). VEGAN: Community ecology package. R package version 2.3-0.