

ECOLOGICAL AND GEOGRAPHIC MODES OF SPECIES DIVERGENCE IN WILD TOMATOES¹

TAKUYA NAKAZATO^{2,5}, DAN L. WARREN³, AND LEONIE C. MOYLE⁴

²Department of Biology, The University of Memphis, Memphis, Tennessee 38152 USA;

³Population Biology Graduate Group, University of California, Davis, California 95616 USA; and

⁴Department of Biology, Indiana University, Bloomington, Indiana 47405 USA

Understanding the role of geography and ecology in species divergence is central to the study of evolutionary diversification. We used climatic, geographic, and biological data from nine wild Andean tomato species to describe each species' ecological niche and to evaluate the likely ecological and geographical modes of speciation in this clade. Using data from >1000 wild accessions and publicly available data derived from geographic information systems for various environmental variables, we found most species pairs were significantly differentiated for one or more environmental variables. By comparing species' predicted niches generated by species distribution modeling (SDM), we found significant niche differentiation among three of four sister-species pairs, suggesting ecological divergence is consistently associated with recent divergence. In comparison, based on age-range correlation (ARC) analysis, there was no evidence for a predominant geographical (allopatric vs. sympatric) context for speciation in this group. Overall, our results suggest an important role for environmentally mediated differentiation, rather than simply geographical isolation, in species divergence.

Key words: age range; allopatric; *Lycopersicon*; species distribution model; Solanaceae; sympatric; wild tomato.

Adaptation to local environment is a primary force driving morphological evolution and speciation (Schluter, 2000, 2001; Levin, 2004), as demonstrated empirically in many cases including among species of Galápagos finches (Grant and Grant, 2003), Caribbean anoles (Losos et al., 2006) and the cichlid fishes in the African lakes (Kocher, 2004), as well as in classical and contemporary studies in plants (Clausen, 1951; Levin, 2000). Traditionally, determining the factors driving adaptation and speciation requires laborious field measurements of the key environmental variables in natural populations. However, the recent development of remote sensing technologies and rapidly accumulating environmental data derived from geographic information systems (GISs) now provide information on the patterns of terrestrial environmental variation at global and continental scales. In combination with species occurrence data, these tools permit detailed description of the environmental conditions at georeferenced natural population localities for one or more species. Increasingly, these data are being used to complement traditional analyses of adaptation and speciation, to identify environmental differences between populations and species (reviewed in Kozak et al., 2008), and to infer ecological and evolutionary processes from these patterns ("macroecological" analyses, e.g., Brown, 1995).

One example of these new approaches is species distribution modeling (SDM), also known as ecological niche modeling or predictive distribution modeling (Graham et al., 2004; Peterson et al., 2007; Kozak et al., 2008). This technique uses the environmental conditions at a species' known presence (and ab-

sence, if available) localities to build a model of its ecological tolerances, which can then be used to predict the geographic space that fulfills those requirements (Looijen, 1995; Jackson and Overpeck, 2000; Ackerly, 2003). Because these models seek to identify the environmental and climatic features that characterize a species' known distribution (their bioclimate envelope), SDM can provide basic quantitative information about species' apparent habitat preferences. These data can provide a starting point for quantitative analyses of adaptive differentiation (e.g., Nakazato et al., 2008) or to assess likely ecological and evolutionary forces acting to shape species geographical distributions and habitat preferences (e.g., Graham et al., 2004). For example, data from SDM can be used to assess whether species occupy their entire potential distribution; absence from some part of the predicted niche ("overprediction", Anderson et al., 2002b) might indicate the operation of other (unmodeled) factors that act to exclude the species from these locations, such as competition, pollinator limitation, and dispersal limitation. In addition, across a clade, SDM data can be used to evaluate the mean and range of niche space occupied by a set of closely related species, to identify differences between species, and to generate hypotheses about biological factors that might underlie these differences (Kozak et al., 2008; Warren et al., 2008). For example, niche characteristics like size or environmental variability might be related to species-wide characteristics such as genetic variation, mating system, or dispersal modes. Useful applications of SDM include (1) identification of geographic areas where additional populations of rare or economically important species might be found (e.g., Anderson et al., 2002a; Ferrier et al., 2002), (2) prediction of areas where endangered species can be introduced or invasive species can spread (reviewed in Peterson, 2003), and (3) estimation of potential impacts of climate change on species range (reviewed in Pearson and Dawson, 2003).

These SDM approaches have also been used to assess patterns of adaptation and speciation in several contexts. For example, comparison of species distribution models between

¹ Manuscript received 17 July 2009; revision accepted 5 February 2010.

The authors thank the TGRC for the geographic information of the tomato accessions. This work was partly supported by National Science Foundation (Division of Environmental Biology) grant 0532097 to L.C.M.

⁵ Author for correspondence (e-mail: tnkazato@memphis.edu)

sister-species pairs has been used to test whether these pairs are more ecologically divergent than some random expectation (e.g., Peterson et al., 1999; Peterson and Holt, 2003; Graham et al., 2004; Knouft et al., 2006; Warren et al., 2008). These analyses aim to evaluate whether ecological niches (as described by environmental features of species distributions) are conserved or highly divergent among closely related species. Strong niche conservation between sister species supports the importance of phylogenetic inertia in shaping species habitat preferences (that is, higher than average niche similarity between closely related species that is due to relative slow evolution of niche differences). Conversely, strong niche differentiation among closely related species supports a role for ecological divergence either during speciation or shortly thereafter. In all analyses of this kind, the “niche” being estimated is not intended to represent all aspects of a species’ biology; SDMs are necessarily limited by the choice of variables and the geographic scale of the data that is used to produce them (Pearson et al., 2006; Kozak et al., 2008).

Georeferenced data on species occurrences can also be used to address the relative importance of allopatric and sympatric modes of speciation, with or without accompanying environmental information. Allopatric speciation has long been considered the dominant mode of speciation (e.g., Poulton, 1889–1907; Gulick, 1905; Jordan, 1905, 1908; Grinnell, 1914; Mayr, 1959, 1963); however, recent theoretical and empirical studies suggest that sympatric speciation might be more common than traditionally recognized (reviewed in Coyne and Orr, 2004). With species occurrence information, the relative prevalence of allopatric and sympatric speciation can be assessed by comparing the observed geographic range size and degree of overlap of species pairs (Lynch, 1989; Chesser and Zink, 1994). If speciation is predominantly allopatric, most closely related species pairs will have non-overlapping geographical ranges, whereas if speciation is predominantly sympatric geographic ranges of sister taxa will overlap substantially. Because species ranges can change stochastically after speciation (thereby obscuring the geographical context during speciation), age-range correlation (ARC; Barraclough et al., 1998; Berlocher, 1998; Barraclough and Vogler, 2000; Berlocher and Feder, 2002; Fitzpatrick and Turelli, 2006) approaches have been developed to take into account both species ranges and phylogenetic relationships when assessing the likely geographical context of speciation. Under this approach, if sympatric speciation is the dominant process, range overlap should be ~100% between young species pairs but decrease among older pairs, due to post-speciation range changes. If allopatric speciation is the dominant process, range overlap between young species pairs should increase from ~0% to random association as species pairs become more divergent.

In this study, we use climatic, geographic, and biological data for nine wild tomato species to describe each species’ bioclimate envelope using SDM and to evaluate the likely ecological and geographical modes of speciation in this group. The center of diversification of wild tomato species is from Ecuador to northern Chile, along the Andean mountains and the adjacent Pacific coast. The recent rapid uplift of the Andes since the Eocene (~55 million years ago [Ma], Campbell, 1975; Stibane, 1975) has created a dramatic altitudinal cline (sea level to >6500 m a.s.l.) and correspondingly diverse climate zones ranging from temperate desert along the western coast (<5 mm of annual precipitation), through seasonable highlands in the Andes, to tropical rainforests in the Amazon basin (>4 m annual precipitation), all within 500 km (Young et al., 2002). Despite recent

divergence of the tomato group (~7 Ma; Nesbitt and Tanksley, 2002), each species appears to display a characteristic geographical distribution pattern and habitat preference across this environmentally diverse region (Rick, 1973, 1978, 1979; Taylor, 1986; Smith and Peralta, 2002). Species are also morphologically diverse, and some traits are likely adaptive responses to local habitats (e.g., Rick, 1973, 1976, 1978; Patterson et al., 1978; Vallejos, 1979; Bloom et al., 2004; Nakazato et al., 2008), suggesting that abiotic ecological conditions could play an important role in phenotypic evolution and speciation in this group. Previous field collections have outlined geographic distributions and provide anecdotal observations of habitat preferences; however, species distributions have not been examined systematically to provide quantitative comparisons among species. Similarly, the relative role of ecological and geographic differentiation in tomato speciation has not been evaluated.

Here we conducted quantitative multivariate analyses and constructed SDMs using >1000 georeferenced accessions from nine species and GIS-derived key environmental (climate and soil) variables to estimate each species’ niche. First, we provide a quantitative description and comparison of the bioclimate envelopes (niches) for each species, including an estimate of intraspecific variation. Second, we evaluate whether overall niche features (niche filling, niche breadth) are associated with species-wide characteristics, including geographical range and mating system. Third, we assess the nature and strength of ecological differentiation between four putative sister species pairs, to evaluate quantitatively whether ecology could have played a specific role in species divergence. We focus on sister-species pairs because geographical and environmental differences between them likely reflect recent divergence processes that occurred at or around speciation, rather than post-speciation changes in ecological niches. Finally, based on the estimated geographic species ranges and pairwise genetic distances among species, we conduct ARC analysis to evaluate evidence for a predominant geographic (allopatric vs sympatric) mode of speciation in this group.

MATERIALS AND METHODS

Study species and accessions—Wild and cultivated tomatoes form a small monophyletic clade consisting of 13 closely related species, within the large and diverse family Solanaceae (D’Arcy, 1978; Peralta et al., 2008). All members of the tomato clade are diploids ($2N = 24$; Peralta and Spooner, 2001; Nesbitt and Tanksley, 2002) that share a high degree of genomic synteny (Chetelat and Ji, 2007) and are to some degree intercrossable (Rick, 1979). Recent taxonomic revision has nested the tomato clade (formerly the genus *Lycopersicon*) within the genus *Solanum* (now *Solanum* sect. *Lycopersicon*), based on molecular genetic characters (Peralta and Spooner, 2001; Spooner et al., 2005). We use the revised nomenclature here. All but three wild species in *Solanum* sect. *Lycopersicon* are included in this study (Fig. 1). *Solanum cheesmaniae* and *S. galapagense*, which are endemic to the Galápagos Islands, were not included because detailed environmental data were not available for these islands. In addition, *S. huaylasense*, a species that has been newly split from *S. peruvianum* s.l. (Peralta et al., 2008), was excluded from the current study because it has very few populations (eight) available for analyses and a phylogenetic placement that is not yet fully resolved. Geographic coordinates of the studied accessions were obtained from the Tomato Genetics Resource Center (TGRC) at the University of California, Davis. When only collection site information for a given accession (e.g., province and town name) was available, its coordinates were inferred from the NGA GEOnet Names Server (<http://earth-info.nga.mil/gns/html/index.html>). In cases where accession location records were incomplete or could not be unambiguously verified, these records were removed from the analysis (<20 records total). These geographical data are the product of over five decades of field sampling by multiple investigators, so are likely to be a good approximation of actual species’ current range distributions.

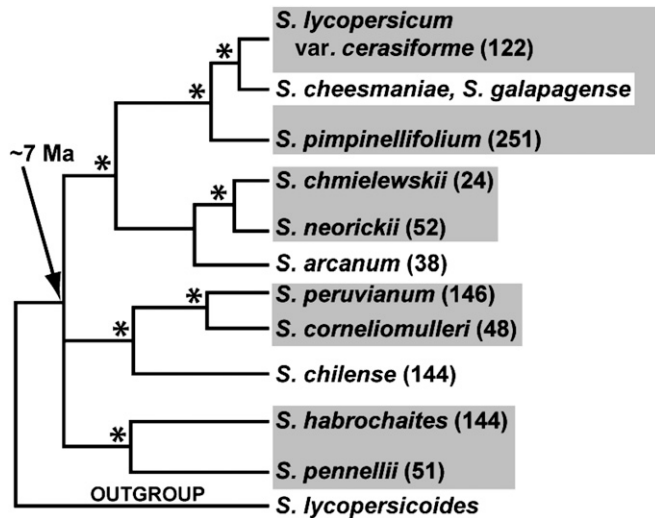


Fig. 1. General cladogram of wild tomato species, modified from Spooner et al. (2005), which is a single consensus Fitch tree based on AFLPs (Spooner et al., 2005), *GBSSI* (Peralta and Spooner, 2001), cpDNA restriction sites (Palmer and Zamir, 1982), and ITS (Marshall et al., 2001). The cladogram has been modified to include *S. arcanum*, *S. corneliomulleri*, *S. galapagense*, and *S. neorickii* in light of Peralta et al. (2008). Asterisks indicate branches that are supported in molecular analyses of the group (see Materials and Methods). For each species, the number of populations analyzed in the current SDM study is shown. Sister-species pairs analyzed in this study are shown in gray boxes. *Solanum cheesmaniae*, *S. galapagense*, and *S. huaylense* were not included in this study due to lack of detailed geographic or environmental information for these species (see Methods).

Environmental variables and GIS analyses—Of numerous environmental variables that are publicly available, we chose 10 climate and soil variables for our analyses (Table 1). These variables are not highly correlated (pairwise $r < 0.7$ based on all sample locations) and are most likely important for species distributions based on prior knowledge of the species' biology (Nakazato et al., 2008; T. Nakazato, unpublished data). Climate data were handled and analyzed in the program ArcGIS 9.2 (ESRI, Redlands, California, USA). Mean normalized difference vegetation index (NDVI; a measure of the density of vegetation and therefore a proxy for biotic competitive environment) was obtained by averaging the monthly NDVI values. Sun exposure (SunEx) data (shape files) were converted to the raster format by interpolation using the Natural Neighbor function of ArcGIS. All environmental layers were converted to the cell size of 30×30 arc-seconds for analysis to ensure all environmental variables were analyzed at the same resolution. Prior to multivariate analyses, environmental values at the occurrence localities were extracted using ArcGIS.

Species distribution modeling (SDM)—We constructed SDMs for each species using the program Maxent (Phillips et al., 2006; <http://www.cs.princeton.edu/~schapire/maxent/>), which generates a continuous binomial probability distribution representing habitat suitability. Maxent was run for each species using the default setting, except that 50% of the occurrence localities were used for testing the model performance. We focused our SDM analyses on the native range of tomato species in South America, north of -26° and west of -65° .

Maxent performed reasonably well in generating predicted niches. The area under the receiver operating characteristic (ROC) curve (AUC) was above 0.95 for all species, and the threshold of equal sensitivity (the fraction of all unsuitable areas that are predicted to be unsuitable, or a true negative rate) and specificity (the fraction of all suitable areas that are predicted to be suitable, or a true positive rate) was relatively small for each species (range: 7.85–22.56%), indicating that the modeling performed reasonably well. The binomial probabilities for all species at this threshold were significantly better ($<10^{-17}$) predictions than the random expectation.

Species geographical, environmental, and niche evaluations—A rough estimate of the geographical range of each species was produced by creating a buffered area (20 km radius) around each known occurrence locality and merging overlapping areas using the Buffer function of ArcGIS. We did not use the more common minimum convex polygon or local convex hull approaches because they do not capture the relatively patchy nature of the distribution of the tomato species' local populations (R. Chetlat, TGRC, personal communication) and therefore would likely overestimate species ranges. Species range overlaps were estimated based on the map pixels predicted as presences for both species divided by the number of pixels predicted as presences for the species with a smaller range (i.e., 100% overlap if a species range is enclosed in its sister species range) as in previous studies (Lynch, 1989; Chesser and Zink, 1994; Barraclough and Vogler, 2000; Berlocher and Feder, 2002; Fitzpatrick and Turelli, 2006).

To quantitatively assess environmental heterogeneity within each species, we performed a TwoStep clustering analysis in SPSS 15.0 (SPSS, Chicago, Illinois, USA), which estimates the number of ecotypic clusters within species and their membership based on population values for each of our 10 climatic and environmental variables. This analysis follows two steps: (1) sequential clustering of each entry into previously formed subclusters using a modified cluster feature (CF) tree, and (2) grouping the subclusters determined in (1) into clusters that maximize the Bayesian information criterion (BIC).

On the basis of the environmental values extracted at the occurrence localities for each species, we used two approaches to test whether the observed bioclimate features differ significantly between species. First, we used one-way analyses of variance (ANOVAs) on each of our 10 environmental variables to test for significant species effects. Second, conventional multivariate analysis of variance (MANOVA; Wilk's λ) was conducted on all environmental variables simultaneously for each species pair to evaluate whether the observed environmental conditions differed significantly between each possible species pair. Statistical analyses were conducted using SPSS.

For each species, we calculated the degree of "niche filling" (the proportion of the space predicted to suit the species' ecological requirements that is actually known to be occupied by each species) by comparing the geographical range of suitable habitat from SDM to its known geographical range estimated

TABLE 1. GIS-derived environmental variables used in the study.

Environmental variable	Abbreviation	Resolution	Source
Mean annual precipitation (mm)	AnnPrec	30×30 s	WORLDCLIM (http://worldclim.org)
Mean annual temperature (C)	AnnTemp	30×30 s	WORLDCLIM (http://worldclim.org)
Precipitation seasonality	PrecSeason	30×30 s	WORLDCLIM (http://worldclim.org)
Temperature seasonality	TempSeason	30×30 s	WORLDCLIM (http://worldclim.org)
Sun exposure (%) ^a	SunEx	30×30 min	Digital atlas of central and South America (http://webgis.wr.usgs.gov/globalgis)
Mean normalized difference vegetation index	NDVI	2.5×2.5 min	Digital atlas of central and South America (http://webgis.wr.usgs.gov/globalgis)
Topsoil pH	SoilPH	5×5 min	FAO, GeoNetwork (http://www.fao.org/geonetwork)
Soil moisture storage capacity (mm/m)	SoilMoist	5×5 min	FAO, GeoNetwork (http://www.fao.org/geonetwork)
Effective soil depth (cm)	SoilDepth	5×5 min	FAO, GeoNetwork (http://www.fao.org/geonetwork)
Topsoil carbon/nitrogen ratio	C/NRatio	5×5 min	FAO, GeoNetwork (http://www.fao.org/geonetwork)

^a listed as Cloud Cover in the data source

from empirical occurrence records (as described earlier). This test requires binary predictions of habitat suitability instead of the continuous output produced by Maxent, so “suitable habitat” was defined based on sensitivity-specificity sum maximization approach, which has been shown to perform well for this purpose (Liu et al., 2005). Niche filling was estimated as the number of pixels in the “suitable habitat” geographic region that were occupied by a known species occurrence. In addition, for each species, we calculated an estimate of “niche breadth” by applying a traditional measure of niche breadth (inverse concentration; Levin, 1968) to the Maxent predictions of habitat suitability for each species.

Niche overlap of sister species was estimated using two alternative measures of similarity (Schoener's D and I). These metrics, that are similar in construction but differ in their assumptions about the interpretation of SDM suitability scores (Warren et al., 2008), do not require a threshold for predicted presence, a procedure that can be problematic when making interspecific comparisons. Associations between species-wide biological, geographical, and niche characteristics were assessed with standard correlations in the program JMP 5.0 (SAS Institute, Cary, North Carolina, USA).

Sister-species comparisons—To assess the degree and nature of differentiation among the closest species pairs, we examined geographic distributions and ecological niches in more detail in four pairs of species: (1) *S. chmielewskii*–*S. neorickii*, (2) *S. habrochaites*–*S. pennellii*, (3) *S. lycopersicum* var. *cerasiforme*–*S. pimpinellifolium*, and (4) *S. peruvianum* s.s.–*S. corneliomulleri* (Table 3). Taxonomic resolution has proven to be challenging in *Solanum* sect. *Lycopersicon*, making the selection of sister species challenging. However, we focused on species pairs whose close relationships are relatively well supported in one or more previous systematic analyses. Pair 1 is well supported by most phylogenetic studies that include both these species (Peralta et al., 2008). Pair 2 is a well-supported clade (98% bootstrap support) in the strict consensus tree from an analysis that combined one morphological and five molecular data sets (Spooner et al., 2005). This pair is likely more distantly related than our other species pairs, but neither species is consistently grouped with any other wild tomato species. Pair 4 includes two taxa that were traditionally considered members of a single species (*S. peruvianum* s.l.), but were split from each other in the recent taxonomic revision (Peralta et al., 2008). The same revision also differentiated two additional species (*S. huaylasense* and *S. arcanum*; Peralta et al., 2008); however, most molecular analyses place populations of these additional species separately from *S. peruvianum* and *S. corneliomulleri*. Accessions from this pair generally group together, in many cases, with intercalated molecular relationships (Peralta et al., 2008). These data indicate that this pair is at a very early stage of divergence, potentially even continuing to exchange genes.

Despite these potential ambiguities in sister status, our primary goal is to assess factors associated with divergence between closely related species (or taxa); given the current taxonomic data, our selection of species pairs appears to be reasonable for this purpose. We estimate that the four putative sister pairs vary in divergence time from <1 to ~4.7 Myr based on inference from the divergence of sequences studied in (Peralta and Spooner, 2001; Nesbitt and Tanksley, 2002, see Appendix S1 in Supplemental Data with the online version of this article). For each species pair, we determined (1) species ecological differentiation based on multivariate and principle component analyses (PCA) of environmental factors and (2) species differentiation in ecological niche using several tests of niche differentiation based on SDMs.

MANOVA and PCA—As outlined earlier, MANOVAs were conducted on the environmental variables extracted at the occurrence localities of each species to test whether the observed environmental conditions differ significantly between each species pair. In addition, for sister pairs, the relative contributions of environmental variables to the discrimination of species were determined by discriminant function analysis (DFA) implemented in SPSS and expressed by the function structure matrix based on this DFA. Species differences were visualized using three-dimensional plots of the first three principal components from the PCA analysis.

SDM-based tests of niche differentiation/conservatism—Two recently developed tests were used to assess whether ecological niches are conserved between sister species: niche identity tests and background similarity tests (Warren et al., 2008). These tests serve to complement the MANOVA in that they allow the testing of different null hypotheses and use more traditional metrics of niche similarity (D and I), making their outputs more directly interpretable in light of classical studies of niche differences. Niche identity tests were conducted by comparing the similarity of sister species' actual environmental niche models to

a distribution of niche similarities obtained from pairs of pseudoniches constructed based on randomly reshuffled occurrence points for the two species, according to Warren et al. (2008). This approach tests whether the observed niche similarity differs significantly from that expected under the null hypothesis that the two species are ecologically equivalent. Background similarity tests were conducted by comparing the predicted niche of a focal species with a set of pseudoniches modeled based on randomly sampling its sister species' geographic range, according to Warren et al. (2008). This method tests whether sister species are more or less ecologically divergent than would be expected given the differences in the local environmental backgrounds of the regions where they occur. Significant results can suggest that the ecological niches of sister species are either more different or more similar than expected, so this is treated as a two-tailed test. Niche breadth, niche overlap, and equivalency and background tests were conducted using ENMTools (available at website <http://enmtools.com>; Warren et al., 2008).

Estimation of genetic distances between species pairs and ARC—Pairwise species genetic distances were estimated based on observed silent substitutions (synonymous substitutions in coding regions, in addition to noncoding substitutions) at five unlinked nuclear loci (*fw2.2*, *orf44*, *Adh2*, *TG10*, *TG11*) analyzed in Nesbitt and Tanksley (2002). Because *S. chmielewskii* and *S. arcanum* were not included in this previous study, we estimated the missing pairwise genetic distances by interpolating the missing values from a regression of the genetic distances among species common to the Nesbitt and Tanksley (2002) study and a phylogenetic analysis based on the *GBSSI* locus (Peralta and Spooner, 2001) that included these two species (online Appendix S1). *Solanum corneliomulleri* comparisons were excluded as there is no data on genetic distance for this newly described species. ARC based on the estimated pairwise genetic distances and species geographical range overlaps (see above) was tested by Mantel test (Mantel and Valand, 1970) using the *zt*-program (Bonnet and Van de Peer, 2002). Note that removing the interpolated data had no effect on our substantive results for this analysis (data not shown).

RESULTS

Environmental variation among and within tomato species—Consistent with field observations and descriptions of the species habitats, we found substantial environmental variation among tomato species. All 10 environmental variables showed a significant species effect (Welch's ANOVA; $P < 0.001$; Appendix S2, see online Supplemental Data). MANOVA of the environmental variables between all species pairs were significant after Bonferroni correction (Wilks' λ ; $P < 0.0001$; see online Appendix S3), except between *S. peruvianum* and *S. corneliomulleri* ($P = 0.095$), indicating that most species have unique ecological preferences. Differences between *S. lycopersicum* var. *cerasiforme* and *S. chilense* capture many aspects of the environmental extremes observed across the nine species (Fig. 2). For example, *S. l. cerasiforme* occurs in regions with warm and wet environments, little seasonal variation in temperature and precipitation, and relatively poor acidic soil, while *S. chilense* occurs in high-altitude temperate deserts with high temperature and precipitation seasonality and relatively rich soil. The most highly variable factors among species included annual mean temperature, altitude, and NDVI (Fig. 2). In contrast, with the exception of *S. l. cerasiforme*, tomato species are generally distributed in drier environments that receive on average <1 m of precipitation a year (Fig. 2).

Among the 10 tomato species, there were different apparent levels of intraspecific climatic variation, ranging from no detectable substructure to several distinguishable within-species climatic clusters (Table 2). The number of within-species clusters was consistently correlated with species range size (Spearman's $\rho = 0.905$, $P = 0.0003$), that is, a larger range size was more likely to encompass >1 climatic cluster within species. Climatic clusters were often clearly geographically, altitudinally,

or latitudinally structured. For example, two clusters were identified for *S. neorickii*, largely corresponding to the northern and southern populations of the species, and differentiated by seasonality variables and sun exposure (online Appendices S3, S4). In some cases, detected clusters did not have obvious geographical structure. For example, three clusters were identified for *S. habrochaites* (online Appendix S5) of which one cluster was strongly restricted to northern Peru, while the other two showed little geographic structure. Temperature seasonality often differentiated within-species clusters, reflecting latitudinal and/or altitudinal segregation, as well as soil depth and precipitation-related variables such as annual precipitation and precipitation seasonality (online Appendix S4).

Species distribution modeling—Based on Maxent jackknife scores, the climate variables, particularly annual precipitation and annual temperature, generally contributed more to species distribution (i.e., showed a greater association between environmental variation and species presence) than did soil variables (Fig. 3). Annual precipitation explained more than 60% of the species distribution for species found in semiarid or arid environments, particularly *S. chilense* and *S. pennellii*, and to a lesser degree for *S. peruvianum*, suggesting that aridity is a major determinant of habitat suitability. In contrast, annual temperature had a high contribution (53%) to the distribution of *S. l. cerasiforme*, possibly indicative of a preference for warmer environments. Annual temperature and temperature seasonality had relatively high contributions to the highland species *S. chmielewskii* and *S. neorickii*, consistent with their occurrence in cool and seasonally stable environments.

Species differed substantially in their estimated niche breadth and in the degree of observed niche filling (i.e., the proportion of their potential distribution—estimated from SDM—occupied according to species occurrence records; Table 2). For example, *S. l. cerasiforme*, *S. habrochaites*, and *S. peruvianum*

had very broad predicted niches, whereas *S. arcanum* and especially *S. chmielewskii* had estimated niche breadths 5–10 times smaller. Niche breadth was significantly associated with species range size (Spearman's $\rho = 0.794$, $P = 0.006$), which could result simply because a smaller range will, on average, span a smaller subset of environmental variation than a larger area. Conversely, it may indicate that species are narrowly distributed because they have a restricted set of environmental tolerances compared to more widely distributed species.

The degree of estimated niche filling varied from 8 to 52%. For *S. chilense*, *S. pennellii*, and *S. pimpinellifolium*, the predicted distributions matched fairly well with the actual distributions of the species (i.e., known geographical occurrences). In these cases, our models are likely capturing many of the important ecological determinants of these species' ranges. In comparison, the distributions of *S. arcanum*, *S. chmielewskii*, *S. neorickii*, and especially *S. l. cerasiforme* were predicted to be much wider than their observed ranges. Besides uncertainties involved in the definition of the actual and predicted habitats of these species, the discrepancy between them may suggest that some critical factors determining the distributions of these species are missing in our models. Arguably, dispersal limitation is unlikely the primary cause of the discrepancy because the areas of unoccupied suitable habitat are largely adjacent to areas that are known to be occupied, although we cannot rule out non-equilibrium explanations. Potential factors include competition with other plant species, pollinator limitation, and other microenvironmental factors and local geographic features, including rivers. In addition, in the case of *S. l. cerasiforme*, accessions of this species can often be found close to human settlements, and some populations might act as human commensals. In these cases, regional climate might be a poorer predictor of habitat preferences in comparison to local microenvironmental conditions, leading to inaccuracies (and most likely overprediction) in predicted geographical distributions (see also Nakazato et al.,

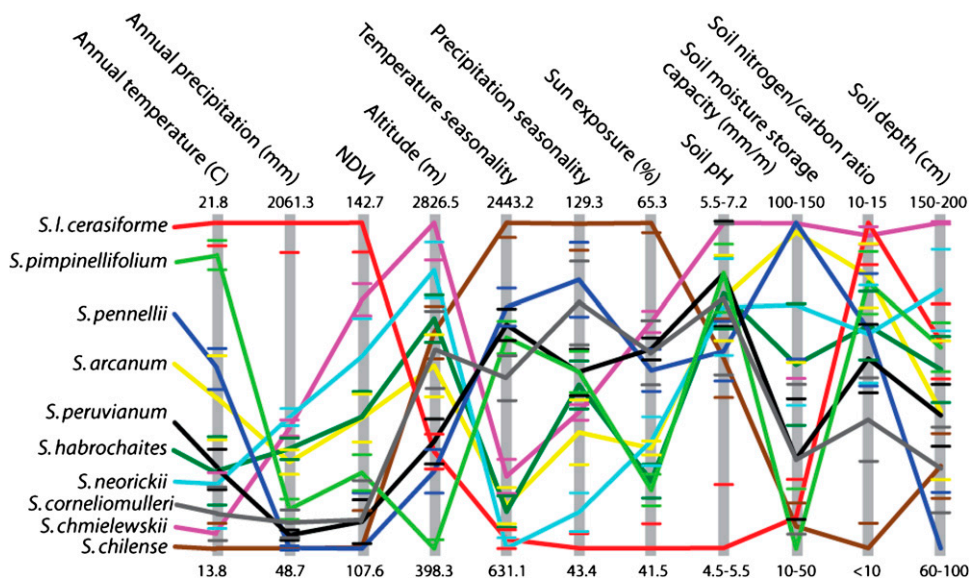


Fig. 2. Environmental variation among tomato species (*Solanum*). Different colors represent different species; vertical gray bars represent the 10 environmental variables studied plus altitude, with population mean values for each species indicated where each species line crosses the vertical bar. Numbers indicate the maximum and minimum of the mean species values for each environmental variable. Short horizontal lines on the gray bars show the 95% confidence intervals of the environmental values for species with respective colors (lines above and below the maximum and minimum values are omitted). NDVI, normalized difference vegetation index; *S. l. cerasiforme*, *Solanum lycopersicum* var. *cerasiforme*.

TABLE 2. Species biological and niche characteristics

Species	Species range size (km ²)	No. intraspecific clusters	Niche breadth ($\times 10^{-3}$)	% Niche filling	Mating system	Geographic location ^a
<i>S. arcanum</i>	23488.32	1	64.15	19.1	Outcross	Mid-altitude regions of northern Peru
<i>S. lycopersicum cerasiforme</i>	94570.76	3	419.55	11.7	Self	Low-altitude edges of Amazon basin
<i>S. corneliomulleri</i>	39747.09	2	342.02	24.4	Outcross	Mid-altitude regions of central to southern Peru
<i>S. chilense</i>	62401.6	2	170.55	31.5	Outcross	High-altitude Altiplano of northern Chile
<i>S. chmielewskii</i>	17616.24	1	28.02	7.7	Self	High-altitude tropic-alpine southern Peruvian Andes
<i>S. habrochaites</i>	88213.64	3	259.30	25.2	Outcross	Mid to high-altitude regions of northern Peru
<i>S. neorickii</i>	33540.86	2	120.51	19.8	Self	High-altitude tropic-alpine Peruvian Andes
<i>S. pennellii</i>	33565.8	1	88.21	27.0	Outcross	Coastal Peru
<i>S. peruvianum</i>	80961.02	3	388.80	22.4	Outcross	Low to mid-elevations in Peru and northern Chile
<i>S. pimpinellifolium</i>	128408.3	3	181.47	51.6	Self	Low-elevations in Ecuador and northern Chile

^a Within S. America only; note that *S. l. cerasiforme* is found in many other tropical and subtropical regions of the world.

2008). More detailed studies of the individual species and their habitats would be helpful in resolving these discrepancies.

Apart from the significant associations detected between species geographic range size, number of within-species climatic clusters, and niche breadth, we found little evidence for consistent associations among species-wide characteristics. There were no significant relationships detected between mating system and either geographical range size, niche breadth, or niche filling (data not shown) regardless of whether mating system was defined as self-compatible vs. self-incompatible or as three categories (obligate selfing, facultative selfing, obligate outcrossing; Table 2). Sample size was significantly related to species range size, number of intraspecific clusters, and estimated species niche filling ($P < 0.01$ for each), likely due to the fact that it is easier to obtain more samples from more widely distributed species, and that these are also more likely to encompass intraspecific climatic variation. Niche breadth was apparently unrelated to niche filling or range size (Spearman's $\rho < 0.5$, $P > 0.1$).

Comparison of sister-species pairs: Species ranges and range overlap—Sister species pairs varied from those having fairly similar range size to more than 2-fold differences (Table 3). Spatial arrangement and overlap also varied from substantial regions of sympatry to almost complete geographical separation (Table 3). For example, despite their relatively recent divergence, *S. l. cerasiforme* and *S. pimpinellifolium* populations are almost completely allopatric and clearly separated by the Andean cordillera; their ranges overlap slightly (8% of population occurrences) in Ecuador and northern Peru. Conversely, *S. chmielewskii* and *S. neorickii* are largely sympatric, where the former's range overlaps substantially with the latter. Both occur together in high-altitude tropic-alpine habitats; however, *S. neorickii* occurs widely in the Peruvian Andes, while *S. chmielewskii* occurs in restricted areas of the southern Peruvian Andes. Overall, our results suggest that the contemporary degree of geographical range overlap between closely related taxa can vary widely.

Ecological differentiation between species: PCA and DFA—As suggested by the analyses that included all species pairs (see earlier), all four sister-species pairs differed significantly ecologically based on multivariate analyses of climatic data extracted from population occurrences (Wilk's λ , $P < 0.001$; Table 3, online Appendix S3), with the exception of *S. peruvianum* and *S. corneliomulleri*. However, even this species pair—the most closely related in our comparisons—was significantly differentiated with respect to some individual climate variables. Temperature seasonality differentiated all four pairs (marginally for *S. peruvianum* and *S. corneliomulleri*). In contrast, NDVI and annual precipitation contributed most to ecological differences between *S. pennellii* and *S. habrochaites* and between *S. l. cerasiforme* and *S. pimpinellifolium*, whereas *S. chmielewskii* and *S. neorickii* were differentiated by sun exposure and precipitation seasonality, and *S. corneliomulleri* and *S. peruvianum* by precipitation seasonality and annual temperature (Table 4).

PCA generally clearly separated sister species in ecological space (Fig. 4), again with the exception of *S. peruvianum* and *S. corneliomulleri*. In most cases, one member of the species pair appeared more ecologically "restricted". For example, *S. pennellii* appears to have a narrower ecological space compared to *S. habrochaites* (Fig. 4D); its habitats are clearly warmer, drier,

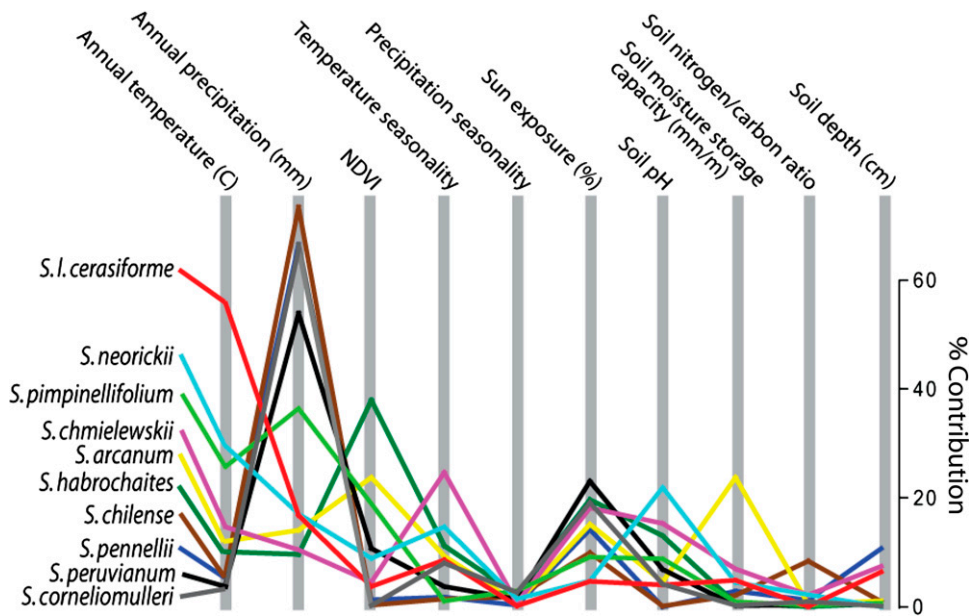


Fig. 3. Contributions of each environmental variable to SDM predictions for wild tomato species. Different colors represent different species; vertical gray bars represent the 10 environmental variables studied. Percentage contributions of each environmental variable to SDM predictions for each species are indicated where each species line crosses the vertical bar, as indicated by the scale for percentage contribution given on the right hand side.

and more seasonable than those of *S. habrochaites*. Note that the relative “sizes” of the ecological niches are also reflected in the number of detected within-species ecological clusters (see above; Table 2); in each sister pair, the species with the smaller range had fewer intraspecific clusters.

Ecological differentiation between species: SDM-derived potential distributions—We evaluated the size and distribution of the area predicted to be suitable for each species and the degree of predicted niche overlap for each species pair. For two of four species pairs, the degree of overlap was generally small in comparison to the size of the predicted distribution for each species (Fig. 5). Among *S. habrochaites* and *S. pennellii*, SDM predicted largely distinct niches for the two species, but with a small region of overlap in the mid-elevation of the Andes (Fig. 5D). Similarly, among *S. pimpinellifolium* and *S. l. cerasiforme*, there is some predicted niche overlap in the northern range of *S. pimpinellifolium*, where the two species partially overlap geographically, although habitat suitability for *S. l. cerasiforme* is relatively low in these areas (Fig. 5A). In comparison, there was substantial overlap between *S. chmielewskii* and *S. neorickii* (Fig. 5B) and even more between *S. corneliomulleri* and *S. peruvianum* (Fig. 5C). These are our youngest species pairs. For *S. chmielewskii* and *S. neorickii*, the estimated niche filling for both species is relatively low (Table 3).

Ecological differentiation between species: Niche identity and background tests—With the exception of the *S. corneliomulleri*–*S. peruvianum* sister pair, randomization tests of niche identity indicated that species in each sister pair are not ecologically equivalent ($P < 0.01$), regardless of the measure of similarity used (Shoener’s D or I). Although these differences could reflect evolutionary divergence of the species analyzed, they might simply reflect the fact that populations of these species are frequently allopatric and thus are exposed to different envi-

ronmental backgrounds. In comparison, our failure to reject a hypothesis of niche equivalency/identity for *S. corneliomulleri*–*S. peruvianum* is consistent with the large degree of geographical overlap between these species.

Results of background tests were more complex but also supported ecological differentiation between two of our species pairs (Table 5, online Appendix S6). We established an a priori definition of “background” as any grid cell within 20 km of known occurrences. For two comparisons, background tests indicated that sisters were less similar than expected based on characteristics of habitat in the region surrounding one of the two species, but not in the reciprocal comparison. For example, background tests showed that *S. l. cerasiforme* is less similar to *S. pimpinellifolium* than it is to the local environmental background against which *S. pimpinellifolium* occurs, although the reciprocal comparison is not significant. In contrast, one comparison shows opposite results for reciprocal comparisons; *S. habrochaites* and *S. pennellii* are more similar than expected given the environmental background surrounding known occurrences of *S. habrochaites*, but less so than expected given the environmental background surrounding occurrences of *S. pennellii*. While counterintuitive, this result is likely to be driven by the differences in the heterogeneity of the environmental background for the two species (Appendix S6, panel E). That is, because the environmental background of *S. habrochaites* appears to be more heterogeneous, the known locations of *S. pennellii* occurrence look more similar to known locations of *S. habrochaites* than to this heterogeneous background, even though their overall similarity is low. For *S. corneliomulleri*–*S. peruvianum*, both reciprocal comparisons showed that these species are more similar than expected based on the available habitat (Table 5), consistent with our failure to detect many differences in ecological preferences between these species.

Although we chose an a priori buffer diameter of 20 km, we also examined the sensitivity of our analyses to the definition of

TABLE 3. Characteristics of *Solanum* sister-species pairs

Species pair	Divergence (Myr)	Range size difference	Species with wider range	% Overlap	DFA (Wilks' λ)	DFA (P)
<i>S. lycopersicum</i> cerasiforme– <i>S. pimpinellifolium</i>	~1.3	1.4×	SPm	8.27	0.275	<0.001
<i>S. chmielewskii</i> – <i>S. neorickii</i>	~1.6	1.9×	SN	36.92	0.446	<0.001
<i>S. corneliomulleri</i> – <i>S. peruvianum</i>	<1.0	2.0×	SPv	60.26	0.907	0.095
<i>S. habrochaites</i> – <i>S. pennellii</i>	~4.7	2.6×	SH	29.66	0.328	<0.001

Note: DFA, Discriminant function analysis

background by conducting background tests with buffer diameters of 2 and 200 km, with all other execution details the same as for 20 km (Appendix S6, panels A–D). For contrast, we also compared each species with its own background. While some comparisons were insensitive to the definition of background (*S. l. cerasiforme* vs. *S. pimpinellifolium* background and *S. corneliomulleri* vs. *S. peruvianum* background), others showed statistically significant differences at some buffer sizes but not others. Only one comparison showed a shift in the directional signal with a changing definition of background (*habrochaites* vs. *pennellii* background), which showed that sisters were significantly more similar at the 2 km scale, but significantly less similar at the 20 and 200 km scales. In both species, comparing points from their background to their own observed locations shows that projected suitability of habitat for both *S. habrochaites* and *S. pennellii* falls off relatively rapidly with distance from known occurrence points. This combination of results might indicate that the two species choose somewhat similar habitat at a fine scale, but that at broader scales there is available habitat that is more suitable for *S. habrochaites* than are the locations where *S. pennellii* is found. This suggests that the niche is conserved to some extent (i.e., given the limited set of conditions in a very narrow geographic area, the species choose similar habitat), but that significant change between species has also occurred (i.e., when more heterogeneous habitat is available, the two species will occupy different areas). Interpretation of these background tests is further discussed in online Appendix S6.

In contrast to the three other species pairs, the background test was not significant in either direction when comparing *S. chmielewskii* and *S. neorickii* at the 20 km scale, and was only significant in one direction at any scale. This is consistent with the large overlap in predicted suitable habitat and in geographical range between these two species (Fig. 5). This result suggests that the ecological divergence between this pair of species is no greater than is expected due to the general environmental differences between the areas in which they occur, rather than a

result of local habitat selection. This does not necessarily indicate that changes in climate preferences have not occurred; rather it indicates that a certain amount of change in the niche is inevitable given the regional differences in environmental variables and that the observed divergence between species is within the bounds expected due to that cause alone. Alternatively, as outlined earlier, this result might be due to the fact that niche filling for these species is relatively low, suggesting either relatively less “precise” predicted niches for these species or other biological or ecological factors act to exclude these species from substantial areas of suitable habitat.

Age-range correlation—We evaluated the correlation between pairwise genetic distances and species range overlap to examine whether the patterns of range overlap change over evolutionary time and to assess the role of geography in speciation. ARC among tomato species was not significant (Mantel test; $r = -0.007$, two-tailed $P = 0.934$; Fig. 6), suggesting that a single geographic mode of speciation (allopatric or sympatric) does not predominate among wild tomato species. Closely related species pairs are most often have the majority of their population locations in non-overlapping geographical regions, and the majority of species pairs showed <50% overlap regardless of genetic divergence. Nonetheless, species range overlaps vary from 0–73.9%, and two closely related species pairs (*S. neorickii* and *S. chmielewskii*; *S. corneliomulleri* and *S. peruvianum*) are substantially sympatric. These observations suggest that substantial or complete geographical separation is not essential for species divergence in this group.

DISCUSSION

Ecological differences among wild tomato species—Our SDM predicted potential distributions of tomato species with high confidence, suggesting that their geographic distributions are substantially influenced by the abiotic environmental conditions

TABLE 4. Structure matrix of discriminant function analysis of environmental variables between sister species of *Solanum*. Values corresponding to the three most significant variables are in boldface.

Variable	<i>S. lycopersicum</i> cerasiforme– <i>S. pimpinellifolium</i>	<i>S. chmielewskii</i> – <i>S. neorickii</i>	<i>S. corneliomulleri</i> – <i>S. peruvianum</i>	<i>S. habrochaites</i> – <i>S. pennellii</i>
TempSeason	-0.524	0.796	0.491	0.757
SunEx	-0.134	0.680	-0.018	0.379
AnnPrec	0.713	-0.170	-0.153	-0.562
PrecSeason	-0.384	0.722	-0.725	0.215
NDVI	0.538	0.209	0.128	-0.683
SoilPH	-0.289	0.237	0.213	-0.092
C/NRatio	0.048	0.284	0.280	0.008
AnnTemp	0.086	-0.130	0.588	0.170
SoilDepth	0.015	0.233	0.273	-0.276
SoilMoist	0.032	0.096	0.098	0.108

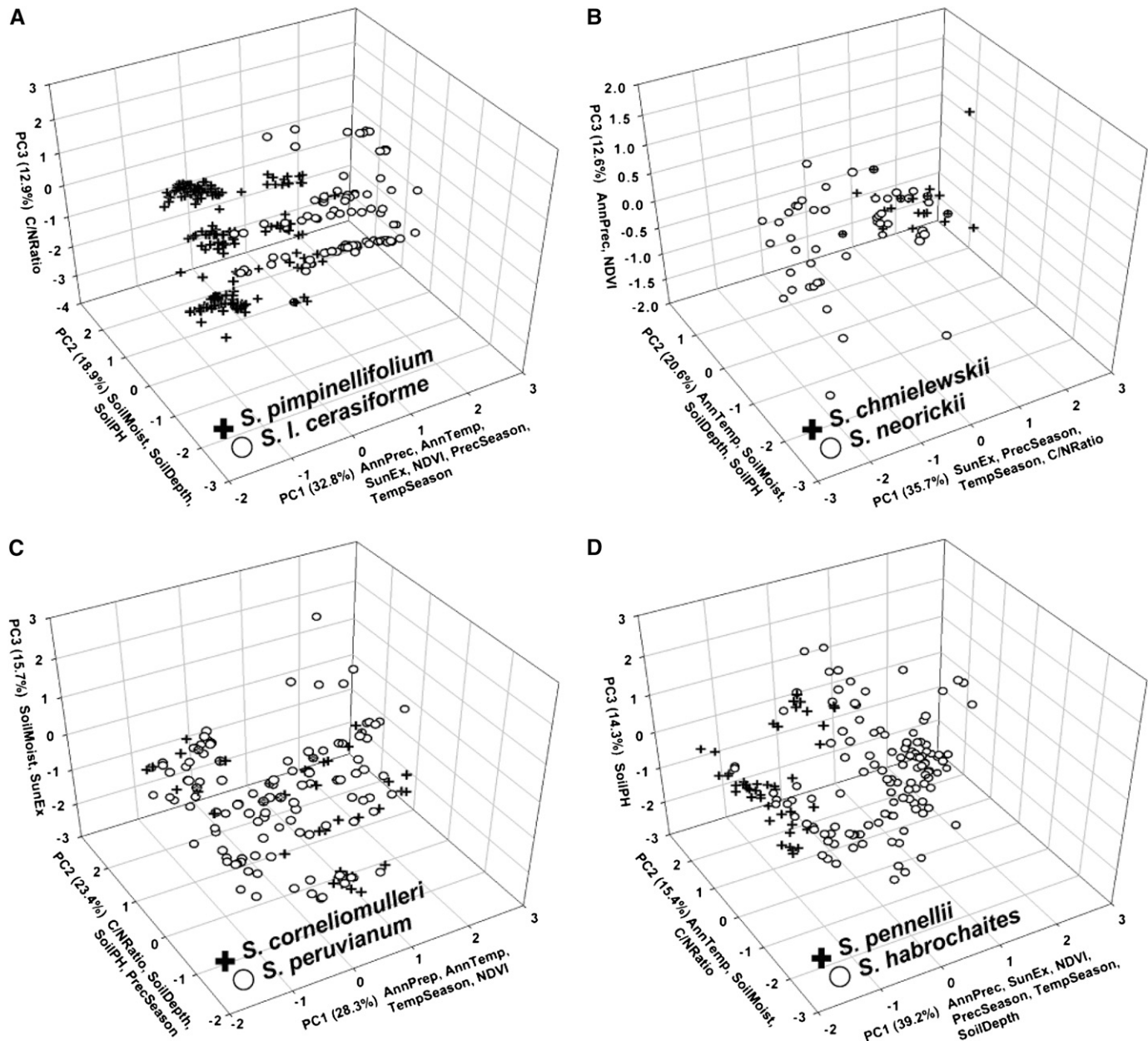


Fig. 4. Distribution of population accessions for four pairs of sister species, in terms of the three most explanatory principle components of environmental variation in each species pair. The PCA was based on the 10 environmental variables used in the study. Percentage variance explained and environmental variables that contributed most to each component are indicated on the three PC axes for each pair.

on which we focused. Climatic variables in general were more highly predictive of species distribution than soil variables (Fig. 3). Soil characters often vary on scales much smaller than GIS mapping resolution (e.g., John et al., 2007); accordingly, soil factors critical in determining species' habitat preferences might vary on a scale smaller than that distinguishable by our GIS-based analyses. Regardless, the strong differences observed between species in one or more abiotic climatic factors do indicate that these factors are relevant and likely important to species distributional and habitat differences in nature.

Many of the patterns of environmental variation and the niche characteristics we have identified are consistent with prior morphological and physiological observations in wild tomato

species. For example, *S. pennellii* has cactus-like features including succulent leaves and very shallow, spreading roots, which are likely to be adaptive in dry environments. In addition, whereas plants generally have more stomata on the bottom of leaves, *S. pennellii* leaves have an equal or larger number of stomata on their upper surface—a trait suggested to be an adaptation to absorb moisture from occasional fog during winter months (Rick, 1973). Our analysis confirms that this species is found in relatively arid coastal environments and in those with periodic fog and shows its predicted distribution is most heavily influenced by mean annual precipitation and sun exposure (Fig. 3). *Solanum chilense* also shows phenotypic and physiological adaptations to dry environments, although these putatively

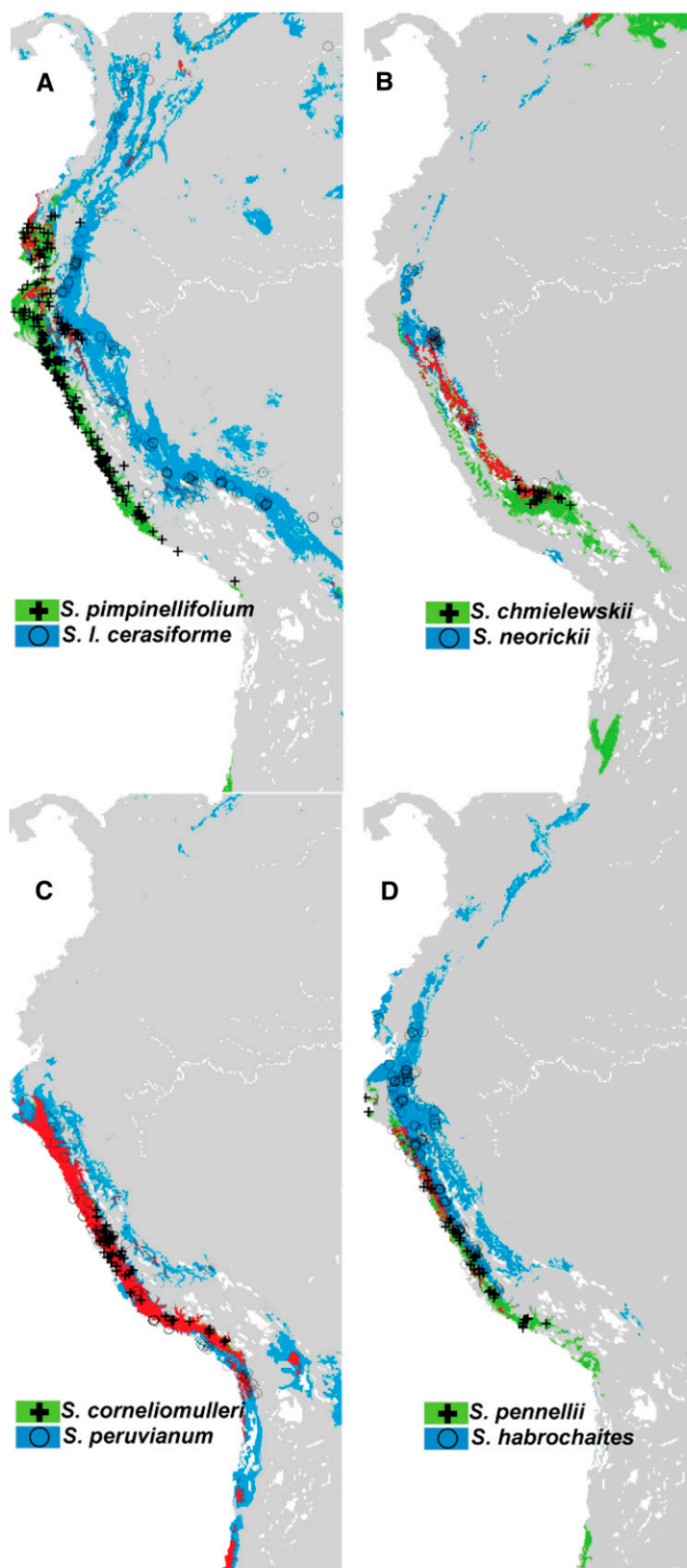


Fig. 5. Geographic distribution of population locations and SDM-derived potential distributions for the four putative sister-species pairs. Observed species locations are given by symbols. The spatial distribution of potentially suitable habitat, based on SDMs at the threshold of equal sensitivity and specificity [see Materials and Methods, *Species distribution modeling (SDM)*], is given by blue and green shading for each species pair. Regions of overlap are shown in red.

TABLE 5. Results of background similarity tests.

Species for observed distribution	Species for background	Schoener's <i>D</i>	<i>I</i>	More or less similar than expected?
<i>S. lycopersicum cerasiforme</i>	<i>S. pimpinellifolium</i>	0.142	0.410	Less ($P < 0.01$)
<i>S. pimpinellifolium</i>	<i>S. l. cerasiforme</i>	0.142	0.410	NS
<i>S. chmielewskii</i>	<i>S. neorickii</i>	0.256	0.521	NS
<i>S. neorickii</i>	<i>S. chmielewskii</i>	0.256	0.521	NS
<i>S. corneliomulleri</i>	<i>S. peruvianum</i>	0.751	0.805	More ($P < 0.01$)
<i>S. peruvianum</i>	<i>S. corneliomulleri</i>	0.751	0.805	More ($P < 0.01$)
<i>S. habrochaites</i>	<i>S. pennellii</i>	0.143	0.424	Less ($P < 0.01$)
<i>S. pennellii</i>	<i>S. habrochaites</i>	0.143	0.424	More ($P < 0.01$)

adaptive traits appear to be quite different from those of *S. pennellii*. For example, *S. chilense* develops extremely deep roots, possibly to absorb water stored in deep soil from occasional flooding (Rick, 1973). As well as demonstrating expected trait-climate matching that is characteristic of adaptation, these anecdotal comparisons suggest that each species' adaptation to specific environmental conditions might be accomplished by acquisition of unique morphological or physiological innovations, a hypothesis that can be tested with direct physiological and phenotypic analysis of each species (e.g., Nakazato et al., 2008).

Finally, the relative importance of each environmental variable in contributing to each species' predicted distribution can provide insight into the critical conditions acting on these species and how many of these are common among many or all species. If species respond and adapt to different environmental variables (e.g., *S. pennellii* to dry environments and *S. l. cerasiforme* to warm environments), different environmental variables should contribute to determining their distributions more or less heavily. Our results indicate that annual precipitation and temperature are significant factors in determining the distribution of most species (Fig. 3); however, other environmental factors are more idiosyncratically influential. For example, NDVI is quite influential for *S. habrochaites*, as is temperature seasonality for *S. chmielewskii*, suggesting that different species can be shaped by different suites of environmental variables (Fig. 3). Similarly, different climatic variables were most

influential in differentiating each sister species pair, although temperature and precipitation variables, especially seasonality in these factors, were often important (Table 4). Overall, our GIS-based approach provides a valuable starting point for generating and testing hypothesis about which specific environmental conditions are driving differentiation within and among particular species. It is notable that many of our influential environmental variables, including temperature and precipitation, vary strongly with altitude. High rates of Andean biological diversification are thought to have been driven by changes accompanying the recent uplift of the Andes, both via vicariance processes and through the modification and creation of new environmental conditions (Richardson et al., 2001; Young et al., 2002; Antonelli et al., 2009 and references therein). The occurrence of many of our study species at mid- and higher-elevation sites in the Andes, suggests that lineage divergence might often have been driven by shifts into new ecological niches created by the rapid Andean uplift, as observed in other groups. Future detailed work will be valuable in assessing the relative importance of specific abiotic factors in determining species habitat characteristics, their adaptive phenotypic responses, and their possible connection with geological and environmental changes in the region.

Modes of differentiation and speciation—Our observed patterns of geographic distribution indicate that each species has a unique range that often includes a large allopatric component. Nonetheless, lack of significant ARC (Fig. 6) suggests that speciation in this group frequently involves more complex scenarios than can be described by simple geographic context (allopatry vs. sympatry). In comparison, our bioclimate results strongly support the involvement of ecological differentiation in species differences, whether they have extensive or little range overlap (Tables 3, 5). Moreover, significant ecological niche divergence between more distant species pairs indicates that niches remain differentiated long after speciation, regardless of geographical context. Note that these findings do not exclude a contributing role for allopatry in species divergence. Rather, in combination they suggest that geographical isolation per se is less important in driving species divergence than environmental differences, including those differences associated with different geographical regions. Although the present analyses cannot tell us whether ecological niche shifts directly caused speciation events, it is clear that ecology is closely linked to persistent species differentiation. Interestingly, our most closely related species pairs tended to show the least amount of ecological divergence, particularly the newly described species pair *S. corneliomulleri* and *S. peruvianum* s.s., consistent with their substantial geographical overlap and possibly ongoing gene flow (see Materials and Methods). Nonetheless,

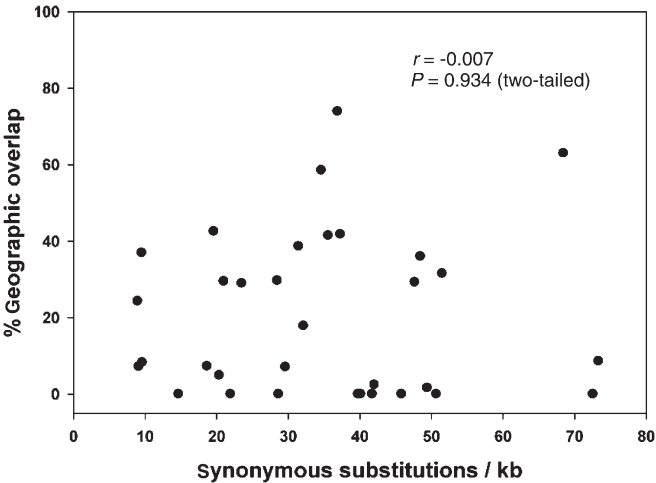


Fig. 6. Correlation between pairwise genetic distances and species range overlap (age–range correlation). The genetic distances are based on observed silent substitutions at five unlinked nuclear loci in Nesbitt and Tanksley (2002) and the *GBSSI* locus (Peralta and Spooner, 2001).

even in this case, several environmental factors distinguish this pair (Table 4).

Based on the substantial ecological differentiation we detected, it could be argued that ecological niches are generally not conserved among tomato species. Several previous studies have suggested that ecological niches can be quite labile among closely related species (e.g., Graham et al., 2004; Ackerly et al., 2006; Knauft et al., 2006), although others have not (e.g., Peterson et al., 1999; Prinzing et al., 2001). In many cases, however, whether ecological niches are considered significantly conserved can depend on the statistical criteria used as well as the temporal, spatial, and ecological scale of the analysis (Warren et al., 2008). Furthermore, as noted by Wiens and Graham (2005, p. 520), “Species will always inhabit environments that bear some similarity to those of their close relatives, [and] to some extent, niches will always be conserved.” In fact, despite significant ecological differentiation among tomato species, semiaridity of their habitats is largely conserved when ecological niches of *Solanum* as a whole are considered, with an exception of *S. l. cerasiforme* in wet, tropical environments. Given this, wild tomatoes could either be considered predominantly niche conserved or niche differentiated, depending on the particular environmental factor under consideration (for similar findings, see e.g., Evans et al., 2009). This pluralism seems appropriate as there is no reason to expect a priori that all ecological tolerances will evolve at the same rate. In addition, the differences in rates of evolution between ecological axes themselves might help to illuminate the physiological mechanisms underlying ecological divergence, or the developmental constraints that prohibit it.

Applications of GIS-based environmental characterization and SDM—In addition to assessing modes of speciation, SDM provides quantitative data on the environmental conditions that characterize a species’ range that can be useful in identifying populations most likely to exhibit adaptive traits of interest. For example, adaptations for drought tolerance might be most expected in populations where native rainfall is highly restricted. These predictions have at least two applications. First, they can identify populations of interest when studying ecological adaptation within and between wild species (e.g., Nakazato et al., 2008). Second, in close relatives of economically important species, they can provide an additional quantitative approach for identifying wild germplasm that might provide useful natural genetic variation. Because of the close relationship among all species in the clade and the vastly reduced genetic variation within the cultivated tomato, wild tomato species have been the source of multiple agriculturally valuable traits for crop improvement, including disease resistance and vitamin content in fruits (Zamir, 2008). Allelic variation from the wild species is introduced into domesticated tomato by means of selective genetic introgression (Grandillo et al., 2007; Zamir, 2008). Plant breeders have frequently used climatic or environmental factors to guide the choice of wild germplasm to evaluate for this purpose (Atherton and Rudich, 1986); however, to date, there has been no detailed quantification of abiotic environmental factors across known collections of wild tomato species. Conversely, predictions of habitat suitability can be used as geographic guides for future collecting efforts as they identify a priori geographical locations that are likely (or unlikely) to be suitable habitats for a species of interest. In this way, GIS environmental characterization can provide quantitative tools to evaluate useful germplasm in existing collections, and SDM can provide an assessment of promising areas for future germplasm collections.

Finally, quantitative GIS-based analyses can potentially provide additional insight into the basis of taxonomic differences. For example, although our analyses did not aim to address correlates of differentiation among newly described species within *S. peruvianum* s.l. (Peralta et al., 2008), we did find evidence that these species are differentiated according to one or more climatic variables. GIS-based environmental characterization and accompanying SDM might be an additional piece of data to consider when evaluating taxonomic status and distinctiveness in this and other systems in the future.

LITERATURE CITED

- ACKERLY, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences* 164: S165–S184.
- ACKERLY, D. D., D. W. SCHWILK, AND C. O. WEBB. 2006. Niche evolution and adaptive radiation: Testing the order of trait divergence. *Ecology* 87 (supplement 7): S50–S61.
- ANDERSON, R. P., M. GÓMEZ-LAVERDE, AND A. T. PETERSON. 2002a. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. *Global Ecology and Biogeography* 11: 131–141.
- ANDERSON, R. P., A. T. PETERSON, AND M. GÓMEZ-LAVERDE. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* 98: 3–16.
- ANTONELLI, A., J. A. A. NYLANDER, C. PERSSON, AND I. SANMARTÍN. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences, USA* 106: 9749–9754.
- ATHERTON, J. G., AND J. RUDICH. 1986. The tomato crop: A scientific basis for improvement. Chapman and Hall, London, UK.
- BARRACLOUGH, T. G., AND A. P. VOGLER. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155: 419–434.
- BARRACLOUGH, T. G., A. P. VOGLER, AND P. H. HARVEY. 1998. Revealing the factors that promote speciation. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 353: 241–249.
- BERLOCHER, S. H. 1998. Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence? In D. J. Howard and S. H. Berlocher [eds.], *Endless forms: Species and speciation*, 99–113. Oxford University Press, New York, New York, USA.
- BERLOCHER, S. H., AND J. L. FEDER. 2002. Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology* 47: 773–815.
- BLOOM, A. J., M. A. ZWIENIECKI, J. B. PASSIOURA, L. B. RANDALL, N. M. HOLBROOK, AND D. A. ST. CLAIR. 2004. Water relations under root chilling in a sensitive and tolerant tomato species. *Plant, Cell & Environment* 27: 971–979.
- BONNET, E., AND Y. VAN DE PEER. 2002. *zt*: A software tool for simple and partial Mantel tests. *Journal of Statistical Software* 7: 1–12.
- BROWN, J. H. 1995. *Macroecology*. University of Chicago Press, Chicago, Illinois, USA.
- CAMPBELL, C. J. 1975. Ecuador. In R. W. Fairbridge [ed.], *The encyclopedia of world regional geology*, 261–270. Dowden Hutchinson & Ross, Stroudsburg, Pennsylvania, USA.
- CHESSER, R. T., AND R. M. ZINK. 1994. Modes of speciation in birds: A test of Lynch’s method. *Evolution* 48: 490–497.
- CHETELAT, R. T., AND Y. JI. 2007. Cytogenetics and evolution. In M. K. Razdan and A. K. Mattoo [eds.], *Genetic improvement of solanaceous crops*, 77–112. Science Publishers, Enfield, New Hampshire, USA.
- CLAUSEN, J. 1951. *Stages in the evolution of plant species*. Cornell University Press, Ithaca, New York, USA.
- COYNE, J. A., AND H. A. ORR. 2004. *Speciation*. Sinauer, Sunderland, Massachusetts, USA.
- D’ARCY, W. G. 1978. New name for the cherry tomato (*Lycopersicon*—*Solanaceae*). *Annals of the Missouri Botanical Garden* 65: 771–772.

- EVANS, M. E. K., S. A. SMITH, R. S. FLYNN, AND M. J. DONOGHUE. 2009. Climate, niche evolution, and diversification of the "bird-cage" evening primroses (*Oenothera*, sections *Anogra* and *Kleinia*). *American Naturalist* 173: 225–240.
- FERRIER, S., G. WATSON, J. PEARCE, AND M. DRIELSMAN. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation* 11: 2275–2307.
- FITZPATRICK, B. M., AND M. TURELLI. 2006. The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution* 60: 601–615.
- GRAHAM, C. H., S. R. RON, J. C. SANTOS, C. J. SCHNEIDER, AND C. MORITZ. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58: 1781–1793.
- GRANDILLO, S., S. D. TANKSLEY, AND D. ZAMIR. 2007. Exploitation of natural biodiversity through genomics. In R. K. Varshney and R. Tuberosa [eds.], *Genomics-assisted crop improvement: Genomics approaches and platforms*, 121–150. Springer, Dordrecht, Netherlands.
- GRANT, B. R., AND P. R. GRANT. 2003. What Darwin's finches can teach us about the evolutionary origin and regulation of biodiversity. *Bioscience* 53: 965–975.
- GRINNELL, J. 1914. The Colorado River as a hindrance to the dispersal of species. *University of California Publications in Zoology* 12: 100–107.
- GULICK, J. T. 1905. *Evolution, racial and habitudinal*. Carnegie Institution, Washington, D.C., USA.
- JACKSON, S. T., AND J. T. OVERPECK. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26: 194–220.
- JOHN, R., J. W. DALLING, K. E. HARMS, J. B. YAVITT, R. F. STALLARD, M. MIRABELLO, S. P. HUBBELL, ET AL. 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences, USA* 104: 864–869.
- JORDAN, D. S. 1905. The origin of species through isolation. *Science* 22: 545–562.
- JORDAN, D. S. 1908. The law of geminate species. *American Naturalist* 42: 73–80.
- KNOUFF, J. H., J. B. LOSOS, R. E. GLOR, AND J. J. KOLBE. 2006. Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology* 87: S29–S38.
- KOCHER, T. D. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. *Nature Reviews. Genetics* 5: 288–298.
- KOZAK, K. H., C. H. GRAHAM, AND J. J. WIENS. 2008. Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution* 23: 141–148.
- LEVIN, D. A. 2000. *The origin, expansion, and demise of plant species*. Oxford University Press, Oxford, UK.
- LEVIN, D. A. 2004. The ecological transition in speciation. *New Phytologist* 161: 91–96.
- LEVIN, R. 1968. *Evolution in changing environments: Some theoretical explorations*. Princeton University Press, Princeton, New Jersey, USA.
- LIU, C. R., P. M. BERRY, T. P. DAWSON, AND R. G. PEARSON. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385–393.
- LOOIJEN, R. C. 1995. On the distinction between habitat and niche, and some implications for species' differentiation. *Poznań Studies in the Philosophy of the Sciences and the Humanities* 45: 87–108.
- LOSOS, J. B., R. E. GLOR, J. J. KOLBE, AND K. NICHOLSON. 2006. Adaptation, speciation, and convergence: A hierarchical analysis of adaptive radiation in Caribbean *Anolis* lizards. *Annals of the Missouri Botanical Garden* 93: 24–33.
- LYNCH, J. D. 1989. The gauge of speciation: On the frequencies of modes of speciation. In D. Otte and J. A. Endler [eds.], *Speciation and its consequences*, 527–553. Sinauer, Sunderland, Massachusetts, USA.
- MANTEL, N., AND R. S. VALAND. 1970. A technique of nonparametric multivariate analysis. *Biometrics* 26: 547–558.
- MARSHALL, J. A., S. KNAPP, M. R. DAVEY, J. B. POWER, E. C. COCKING, M. D. BENNETT, AND A. V. COX. 2001. Molecular systematics of *Solanum* section *Lycopersicum* (*Lycopersicon*) using the nuclear ITS rDNA region. *Theoretical and Applied Genetics* 103: 1216–1222.
- MAYR, E. 1959. Isolation as an evolutionary factor. *Proceedings of the American Philosophical Society* 103: 221–230.
- MAYR, E. 1963. *Animal species and evolution*. Belknap, Cambridge, UK.
- NAKAZATO, T., M. BOGONOVICH, AND L. C. MOYLE. 2008. Environmental factors predict adaptive phenotypic differentiation between two wild Andean tomatoes. *Evolution* 62: 774–792.
- NESBITT, T. C., AND S. D. TANKSLEY. 2002. Comparative sequencing in the genus *Lycopersicon*: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* 162: 365–379.
- PALMER, J. D., AND D. ZAMIR. 1982. Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. *Proceedings of the National Academy of Sciences, USA* 79: 5006–5010.
- PATTERSON, B. D., R. PAULL, AND R. M. SMILLIE. 1978. Chilling resistance in *Lycopersicon hirsutum* Humb. and Bonpl.: A wild tomato with a wide altitudinal distribution. *Australian Journal of Plant Physiology* 5: 609–617.
- PEARSON, R. G., AND T. P. DAWSON. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography* 12: 361–371.
- PEARSON, R. G., W. THUILLER, M. B. ARAÚJO, E. MARTINEZ-MEYER, L. BROTONS, C. MCCLEAN, L. MILES, P. SEGURADO, ET AL. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704–1711.
- PERALTA, I. E., AND D. M. SPOONER. 2001. Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (*Solanum* L. section *Lycopersicon* [Mill.] Wettst. subsection *Lycopersicon*). *American Journal of Botany* 88: 1888–1902.
- PERALTA, I. E., D. M. SPOONER, AND S. KNAPP. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Systematic Botany Monographs* 84: 1–186.
- PETERSON, A. T. 2003. Predicting the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* 78: 419–433.
- PETERSON, A. T., AND R. D. HOLT. 2003. Niche differentiation in Mexican birds: Using point occurrences to detect ecological innovation. *Ecology Letters* 6: 774–782.
- PETERSON, A. T., M. PAPES, AND M. EATON. 2007. Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* 30: 550–560.
- PETERSON, A. T., J. SOBERÓN, AND V. SÁNCHEZ-CORDERO. 1999. Conservatism of ecological niches in evolutionary time. *Science* 285: 1265–1267.
- PHILLIPS, S. J., R. P. ANDERSON, AND R. E. SCHAPIRE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- POULTON, E. B. 1889–1907. *Essays on evolution*. Clarendon Press, Oxford, UK.
- PRINZING, A., W. DURKA, S. KLOTZ, AND R. BRANDL. 2001. The niche of higher plants: Evidence for phylogenetic conservatism. *Proceedings of the Royal Society of London, B, Biological Sciences* 268: 2383–2389.
- RICHARDSON, J. E., R. T. PENNINGTON, T. D. PENNINGTON, AND P. M. HOLLINGSWORTH. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293: 2242–2245.
- RICK, C. M. 1973. Potential genetic resources in tomato species: Clues from observations in native habitats. In A. M. Srb [ed.], *Genes, enzymes, and populations*, 255–269. Plenum Press, New York, New York, USA.
- RICK, C. M. 1976. Natural variability in wild species of *Lycopersicon* and its bearing on tomato breeding. *Agraria* 30: 249–259.
- RICK, C. M. 1978. Tomato germplasm resources. 1st International Symposium on Tropical Tomato, Shanhua, Taiwan, 214–224. Asian Vegetable Research and Development Center, Shanhua, Taiwan.
- RICK, C. M. 1979. Biosystematic studies in *Lycopersicon* and closely related species of *Solanum*. In J. G. Hawkes, R. N. Lester, and A. D. Skelding [eds.], *The biology and taxonomy of the Solanaceae*, 667–679. Linnean Society of London, Academic Press, London, UK.
- SCHLUTER, D. 2000. *The ecology of adaptive radiation*. Oxford University Press, Oxford, UK.

- SCHLUTER, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16: 372–380.
- SMITH, S. D., AND I. E. PERALTA. 2002. Ecogeographic surveys as tools for analyzing potential reproductive isolating mechanisms: An example using *Solanum juglandifolium* Dunal, *S. ochranthum* Dunal, *S. lycopersicoides* Dunal, and *S. sitiens* I. M. Johnston. *Taxon* 51: 341–349.
- SPOONER, D. M., I. E. PERALTA, AND S. KNAPP. 2005. Comparison of AFLPs with other markers for phylogenetic inference in wild tomatoes [*Solanum* L. section *Lycopersicon* (Mill.) Wettst.]. *Taxon* 54: 43–61.
- STIBANE, F. R. 1975. Colombia. In R. W. Fairbridge [ed.], The encyclopedia of world regional geology, 245–250. Dowden Hutchinson & Ross, Stroudsburg, Pennsylvania, USA.
- TAYLOR, I. B. 1986. Biosystematics of the tomato. In J. G. Atherton and J. Rudich [eds.], The tomato crop: A scientific basis for improvement, 1–34. Chapman and Hall, London, UK.
- VALLEJOS, C. E. 1979. Genetic diversity of plants for response to low temperatures and its potential use in crop plants. In J. M. Lyons, D. Graham, and J. K. Raison [eds.], Low temperature stress in crop plants: The role of the membrane, 473–489. Academic Press, New York, New York, USA.
- WARREN, D. L., R. E. GLOR, AND M. TURELLI. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 62: 2868–2883.
- WIENS, J. J., AND C. H. GRAHAM. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics* 36: 519–539.
- YOUNG, K. R., C. U. ULLOA, J. L. LUTEYN, AND S. KNAPP. 2002. Plant evolution and endemism in Andean South America: An introduction. *Botanical Review* 68: 4–21.
- ZAMIR, D. 2008. Plant breeders go back to nature. *Nature Genetics* 40: 269–270.