The role of historical factors and natural selection in the evolution of breeding systems of Oxalis alpina in the Sonoran desert ‘Sky Islands’

J. Pérez-Alquicira*, F. E. Molina-Freaner†, D. Piñero*, S. G. Weller‡, E. Martínez-Meyer§, J. Rozas‖ & C. A. Domínguez*

*Departmento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, México Distrito Federal, México
†Departmento de Ecología de la Biodiversidad, Instituto de Ecología, Universidad Nacional Autónoma de México, Hermosillo, Sonora, México
‡Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, USA
§Departamento de Zoología, Instituto de Biología, Universidad Nacional Autónoma de México, México Distrito Federal, México
‖Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain

Abstract

Pleistocene climatic oscillations are known to influence the patterns of genetic diversity and the distribution of traits that are the target of selection. Here, we combine phylogeographical and ecological niche modelling (ENM) approaches to explore the influence of historical factors (Pleistocene climatic shifts) and natural selection on the evolution of distylous (two floral morphs) from tristyly (three floral morphs) of Oxalis alpina in the Sky Islands of the Sonoran Desert. Molecular data and ENM indicate that historical factors have had a strong influence on the genetic structure and the geographical distribution of reproductive systems of O. alpina. Moreover, genetic results suggest the possibility that distylous populations do not represent a monophyletic group. We propose that the combined effects of natural selection and genetic drift have influenced the tristyly–distylly transition.

Keywords:
distylous;
ecological niche modelling;
genetic structure;
historical factors;
natural selection;
phylogeography;
pleistocene;
Sky Islands;
tristyly.

Introduction

Plant reproductive systems exhibit a dramatic diversity that has attracted the attention of many naturalists and evolutionary biologists since Sprengel’s times (Sprengel, 1793). This fascination has motivated many investigations into the selective forces responsible for the adaptation and the diversification of reproductive systems among flowering plants (Castillo et al., 2002; Pérez-Barrales et al., 2006; Sakai et al., 2006; Hodgins & Barrett, 2008). Whereas these studies have emphasized the role of microevolutionary forces within populations, few have explored the influence of historical factors such as past migration, fragmentation and demographic history associated with events like climate oscillations. These events have produced shifts in the distributional range of plant species, and they are associated with founder events or genetic bottlenecks that affect the evolutionary trajectory of populations (Hewitt, 2000, 2004) including the evolutionary dynamics of plant reproductive systems (Dorken & Barrett, 2004; Hodgins & Barrett, 2007). Species inhabiting regions that have been subject to strong climatic oscillations and that exhibit variation among populations in their reproductive system provide the opportunity to evaluate the relative importance of historical vs. microevolutionary processes for the evolution of plant reproduction (Eckert & Barrett, 1995; Dorken & Barrett, 2004; Hodgins & Barrett, 2007).

The Sky Islands from northwestern Mexico and southwestern USA are particularly interesting for studies of the effects of isolation and climatic fluctuation on evolutionary processes (Weller, 1979, 1986; Barber, 1999; Knowles, 2000; Masta, 2000; Downie, 2004; DeChaine & Martin, 2006; McCormack et al., 2008; Moreno-Letelier & Piñero, 2009). These isolated mountain ranges have the northernmost extent of Madrean evergreen woodland, which becomes contiguous as it extends south into the Sierra Madre Occidental in northern Mexico (Brown & Lowe, 1980; Brown, 1982). Coniferous forest species in this region are restricted to...
isolated mountaintop refugia surrounded by desert scrublands. The current distribution of coniferous forests in the Sky Islands is thought to be a consequence of the climate changes that characterized the end of the Pleistocene period (Van Devender, 1990a,b; Metcalfe et al., 2000; Thompson & Anderson, 2000; Holmgren et al., 2003). During the last glacial maximum (LGM), this region experienced an average decrease in temperature of 6 °C and an increased precipitation regime (Metcalfe, 2006). Analyses of macrofossils and pollen records from the present Sonoran andChihuahuan deserts indicate that during the LGM, the landscape was characterized by extensive areas of woodland vegetation (including different species of Pinus and Juniperus) that reached lower latitudes and altitudes and maintained greater connectivity among populations through gene flow (Van Devender, 1990a,b; Metcalfe et al., 2000; Thompson & Anderson, 2000; Holmgren et al., 2003). The transition between the Pleistocene (11 000 years ago) and the Holocene epochs was characterized by climate changes that produced drier and warmer conditions and promoted a northward and upward range shift in cool-adapted species (Van Devender, 1990a,b; Metcalfe et al., 2000). These processes, in turn, resulted in fragmentation and isolation of populations that were eventually restricted to the few mountaintops with the climate conditions necessary for the survival of these species.

Climatic oscillations and the range shifts they produced are expected to have left measurable traces in the genetic structure of populations. For instance, if latitudinal changes in range distribution produced repeated events of colonization and extinction from southern populations serving as refugia, southern populations should harbour higher levels of genetic variance, whereas northern ones should be genetically impoverished because of recent colonization (Petit et al., 1997; Hewitt, 2000, 2004; Cuevas et al., 2006). Additionally, altitudinal displacements may produce vicariant events restricting gene flow among disjunct populations, and thus allowing for different microevolutionary trajectories (Knowles, 2000; Masta, 2000).

Oxalis alpina (Rose) Knuth (section Ionoxalis, Oxalidaceae) is a heterostylous perennial herb inhabiting evergreen Madrean woodlands (Denton, 1973). Heterostyly is a reproductive system characterized by the presence of two (distyly) or three (tristyly) floral morphs differing in the relative length of their reproductive whors. Tristyly, for example, has three floral morphs, the short-styled, mid-styled and long-styled morphs, which differ in the length of the styles as well as the positions of anther whors. Each morph has two stamen whors located at complementary positions within the flower, mid- and short stamens for the long-styled morph, long and short stamens for the mid-styled morph and mid- and long stamens for the short-styled morph (Fig. 1). This polymorphism is associated with an incompatibility system preventing seed production after crosses between stamens and styles located at different positions (illegitimate crosses, sensu Darwin, 1877). Oxalis alpina exhibits remarkable modifications among Sky Islands populations, indicating that its reproductive system is evolving from tristyly to distyly. Tristylos populations range from isoplethic tristyly (equal frequency of the three morphs) to populations in which the mid-styled morph is very rare and the incompatibility system has been completely modified to resemble distylos populations (Weller et al., 2007). Cytogeographical evidence indicates that distylous species of section Ionoxalis in southern Mexico are highly diverse and typically diploid, whereas distylous species range further north and generally have higher chromosome numbers, a pattern suggesting that distyly is the derived condition in section Ionoxalis (Weller & Denton, 1976). A previous study in O. alpina indicated there is a negative relationship between the extent of incompatibility modification in the short- and long-styled morphs and the frequency of the mid-styled morph, suggesting that incompatibility modifications are responsible for the loss of the mid-styled morph in this species (Weller et al., 2007). The derived nature of distyly in O. alpina is similar to the pattern found in section Ionoxalis. In addition to incompatibility modifications that result in the loss of the mid-styled morph, the evolution of distyly is characterized by morphological floral changes that follow a trend towards the establishment of a distylous phenotype (Sosenski et al., 2010). Therefore, current evidence suggests that microevolutionary forces such as natural selection acting on the incompatibility system and flower morphology are responsible for the evolution of distyly in this species (Weller et al., 2007; Sosenski et al., 2010). However, because the Sky Islands have been subject to marked
climatic oscillations since the LGM, it is possible that genetic drift associated with historical colonizations has also played a role in the evolution of distyly.

The presence of populations representing different steps in the tristyly–distyly transition offers the opportunity to explore the relative importance of historical vs. selective processes in this evolutionary shift. Evidence of several independent evolutionary origins of distyly would support a selection-driven interpretation because the repeated evolution of distyly through the loss of the mid-styled morph is unlikely to be explained solely by genetic drift and founder events. If historical factors have also played a role on the evolution of _O. alpina_ in the Sky Islands, we predict this influence should be expressed as a significant relationship between latitude and breeding system distribution and the magnitude of genetic variability. Southern populations should maintain the ancestral breeding system, whereas distyly, the derived breeding system, should be more common in recently colonized northern populations. Likewise, because southern populations acted as refugia during glacial maxima, these populations should harbour the highest levels of genetic variability.

In this article, we used a combination of phylogeographical and ecological niche modelling methods to analyse the role of historical factors and natural selection on the evolutionary transition from tristyly to distyly in _O. alpina_ populations from the Madrean Sky Islands. We addressed the following questions: (i) Have climatic oscillations left a genetic signature among _O. alpina_ populations? (ii) If so, is this effect also apparent in breeding system distribution? (iii) Do genetic diversity and the genetic structure differ among tristylos and distylos populations? (iv) Does ecological niche modelling support the conclusions derived from genetic studies about the impact of historical factors on the genetic structure of _O. alpina_?

### Materials and methods

**Study species**

_Oxalis alpina_ is a herbaceous perennial plant found in temperate coniferous forest located at high elevations in the mountain ranges from southwestern United States to Guatemala (Denton, 1973). _Oxalis alpina_ is mainly pollinated by solitary bees (_Heterosaurus bakeri_ and _Heterosaurus neomexicanus_; Weller, 1981; C.A. Domínguez, personal observation); other flower visitors are dipterans (primarily Syrphidae), wasps and lepidoptera (Pieridae and Hesperiidae; C.A. Domínguez, personal observation).

The chromosome number of _O. alpina_ varies geographically, but in the Sky Islands region, those populations that have been sampled are uniformly tetraploid, suggesting they are monophyletic (Weller & Denton, 1976).

**Sampling**

From 2001 to 2004, bulbs were collected from fourteen tristylos and six distylos populations (Table 1) located at elevations above 1900 m in isolated mountain ranges.

### Table 1 Breeding system and chloroplast genetic variation for 20 populations of _Oxalis alpina_ from the Sky Islands region.

<table>
<thead>
<tr>
<th>Population</th>
<th>Geog. coord. (latitude; longitude)</th>
<th>RS</th>
<th>S : M : L (%)</th>
<th>IS</th>
<th>n</th>
<th>H</th>
<th>h</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santa Catalina</td>
<td>32.4279; 110.7552</td>
<td>D</td>
<td>57 : 0 : 50</td>
<td>C</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pinalero</td>
<td>32.6459; 109.8508</td>
<td>D</td>
<td>53 : 0 : 47</td>
<td>?</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Santa Rita</td>
<td>31.6772; 110.7452</td>
<td>D</td>
<td>53 : 0 : 47</td>
<td>?</td>
<td>11</td>
<td>3</td>
<td>0.34</td>
<td>0.0004</td>
</tr>
<tr>
<td>Pinery Chiricahua</td>
<td>31.9328; 109.2718</td>
<td>D</td>
<td>50 : 0 : 50</td>
<td>?</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sierra Ancha</td>
<td>33.8404; 110.9556</td>
<td>D</td>
<td>9 : 0 : 91</td>
<td>?</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pinal</td>
<td>33.2996; 110.8415</td>
<td>D</td>
<td>60 : 0 : 40</td>
<td>?</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Miller Canyon Huachuca</td>
<td>31.4158; 110.2779</td>
<td>T</td>
<td>29 : 28 : 43</td>
<td>P</td>
<td>14</td>
<td>2</td>
<td>0.43</td>
<td>0.0008</td>
</tr>
<tr>
<td>Morse Canyon Chiricahua</td>
<td>31.8262; 109.3286</td>
<td>T</td>
<td>27 : 25 : 48</td>
<td>C</td>
<td>12</td>
<td>3</td>
<td>0.59</td>
<td>0.0009</td>
</tr>
<tr>
<td>San Luis</td>
<td>31.2184; 108.8283</td>
<td>T</td>
<td>46 : 21 : 34</td>
<td>N</td>
<td>11</td>
<td>4</td>
<td>0.60</td>
<td>0.0014</td>
</tr>
<tr>
<td>Los Ajos</td>
<td>30.9453; 109.6635</td>
<td>T</td>
<td>34 : 33 : 33</td>
<td>C</td>
<td>17</td>
<td>5</td>
<td>0.78</td>
<td>0.0021</td>
</tr>
<tr>
<td>La Punca</td>
<td>30.5492; 109.7512</td>
<td>T</td>
<td>22 : 30 : 48</td>
<td>P</td>
<td>16</td>
<td>2</td>
<td>0.12</td>
<td>0.0004</td>
</tr>
<tr>
<td>San Jose</td>
<td>31.2572; 110.9901</td>
<td>T</td>
<td>22 : 24 : 53</td>
<td>P</td>
<td>15</td>
<td>2</td>
<td>0.38</td>
<td>0.0003</td>
</tr>
<tr>
<td>Animas</td>
<td>31.5671; 108.7774</td>
<td>T</td>
<td>27 : 29 : 44</td>
<td>C</td>
<td>11</td>
<td>2</td>
<td>0.53</td>
<td>0.0014</td>
</tr>
<tr>
<td>Maripítica</td>
<td>31.0537; 110.3834</td>
<td>T</td>
<td>34 : 36 : 30</td>
<td>N</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pinos Altos</td>
<td>32.9223; 108.2126</td>
<td>T</td>
<td>38 : 21 : 41</td>
<td>C</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Galiro</td>
<td>32.5170; 110.2639</td>
<td>T</td>
<td>?</td>
<td>?</td>
<td>15</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>White</td>
<td>33.6824; 109.4482</td>
<td>T</td>
<td>15 : 36 : 49</td>
<td>P</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>30.7285; 109.8343</td>
<td>T</td>
<td>33 : 33 : 34</td>
<td>P</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Elenita</td>
<td>31.0461; 110.3827</td>
<td>T</td>
<td>39 : 32 : 29</td>
<td>P</td>
<td>11</td>
<td>2</td>
<td>0.43</td>
<td>0.0015</td>
</tr>
<tr>
<td>Azul</td>
<td>30.7412; 110.5732</td>
<td>T</td>
<td>33 : 34 : 33</td>
<td>N</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

RS, reproductive system; D, distylos and T tristylos; S : M : L indicate short : mid : long morph frequency; IS, incompatibility system; C, complete modification; P, partial modification; N, no modification; ?, information not available; n, sample size; H number of haplotypes; h, haplotype diversity; π, nucleotide diversity.
of Arizona, New Mexico, Sonora and Chihuahua. Bulbs were sampled at intervals of at least 1 m to minimize sampling the same genet. Bulbs were grown in a greenhouse at the Instituto de Ecología, Universidad Nacional Autónoma de México where foliar tissue (two leaves) was collected from 10 to 15 individuals from each population. Leaves were preserved in silica gel until DNA was extracted.

DNA extraction and sequencing

Total DNA was extracted using a modified version of the Doyle & Doyle (1987) procedure. These modifications involved grinding 100 mg of dry leaf tissue and adding 1.5 units of ribonuclease A (Sigma-Aldrich, St Louis, MO, USA) and 2 units of proteinase k (Fermentas, Hanouer, MO, USA). DNA was dissolved in 50 μL of water (UltraPure™ water; Invitrogen, Carlsbad, CA, USA). An intergenic spacer from the chloroplast genome located between the Psba [photosystem Q (B)] and trnH (histidine tRNA) genes was amplified by PCR (polymerase chain reaction) using the following pair of primers: trnH CGC GCA TGG TGG ATT CAC AAT CC (Tate & Simpson, 2003; Shaw et al., 2005) and psbA GTT ATG CAT GAA CTT AGT CCT (Sang et al., 1997; Shaw et al., 2005). The NADH dehydrogenase subunit F (ndhF) gene was also amplified from chloroplast DNA using species-specific primers designed for this study (NDH forward: AAT GGT ACG TTA GGT TTG and NDH reverse: GGT AAT GCC TGC AGC TC). The PCRs were performed in 25 μL final volume containing 20–60 ng of genomic DNA, 2 mM MgCl₂, 0.1 mM dNTPs, 0.2 μM of each primer, 1.5 units of Taq DNA polymerase, 2.5 μL of 10× Buffer (Invitrogen). Amplification included 5 min at 80°C; 30 cycles at 94°C for 30 s, 55°C for 30 s (trnH-psbA) or 60°C (NADH) and 72°C for 2 min; with a final extension for 5 min at 72°C. PCR products were gel-purified with the QIAEX II Gel Extraction kit (Qiagen, Maryland, USA) following manufacturer’s instructions. The sequencing reactions were performed by Macrogen DNA sequencing service (Seoul, Korea; http://dna.macrogen.com).

Data analysis

Sequences were aligned with BioEdit Sequence Alignment Editor (version 7.0.8; Hall, 1999) and later adjusted visually with MacClade (version 4.08; Maddison & Maddison, 2005). Before concatenating the two cpDNA regions (trnH-psbA intergenic fragment and ndhF gene), we first assessed the level of congruence between two fragments using the incongruence length difference test (ILD; Farris et al., 1995) implemented in PAUP* (version 4.10b; Swofford, 2003) as the partition homogeneity test.

Genetic variability was estimated as haplotype diversity (hT) and nucleotide diversity (π; Nei, 1987) using DnaSP (version 5; Librado & Rozas, 2009), whereas total haplotype diversity (hT) and the average diversity within population (hπ) were estimated following methods described by Pons & Petit (1996) using PERMUT (http://www.pierroton.inra.fr/genetics/labo/Software/PermutCpSSR/index.html). To detect potential differences in the patterns of genetic variation in the northern and southern populations of the Sky Islands, we performed a regression analysis between latitude and the two estimators of variability (h, π). The normal distribution of residuals was tested by the Shapiro-Wilk W test. ANOVA was carried out to test differences in the levels of genetic variability between distylos and tristylos populations using JMP (version 5.0; SAS Institute Inc., 2005); for this analysis, we included the genetic variation value for each population. We removed the effect of the latitude because there is an evident pattern in the distribution of reproductive systems, with distylos populations occurring mainly in the northern region of the Sky Islands.

To examine the partitioning of genetic variation within and between populations, as well as between distylos and tristylos populations, we performed a molecular analysis of variance (AMOVA) with 1000 permutations implemented in Arlequin (version 3.11; Excoffier et al., 2005, University of Bern, Switzerland). We calculated other statistics of population differentiation including GST and NST because the comparison of both parameters provides useful information about phylogeographical structure (association between geography and gene genealogy; Pons & Petit, 1996). NST takes into account genetic similarities among haplotypes, and GST is based on the allele frequencies. A significantly higher value of NST indicates that more genetically similar haplotypes are also geographically closer, suggesting the presence of phylogeographical structure. The comparison of GST and NST was performed using PERMUT. We compared the calculated NST values with those obtained using 1000 random permutations of haplotype sequences to test whether NST > GST (Burban et al., 1999). The spatial analysis of molecular variance (SAMOVA) program (SAMOVA, version 1.0; Dupanloup et al., 2002) was used to identify groups of populations based on genetic and geographical distances. The SAMOVA algorithm assigns populations to groups with the constraint that they must be geographically adjacent and genetically homogenous. The main goal of this procedure is to create clusters with the highest among-group divergence and within-group relatedness. To find the maximum level of differentiation among groups (FCT), we ran the program with k = 2–10 (where k = number of groups) and 100 simulated annealing processes for each K group. Exact test of population differentiation based on a Markov Chain procedure were carried out among populations (Arlequin version 3.11; Excoffier et al., 2005). To illustrate the genetic distance between populations, we generated a UPGMA phenogram using Nei’s genetic distances Nei (1972) implemented in TFPGA (version 1.3;
Miller, 1997). Nodal support was estimated using bootstrap analysis with 1000 replicates. To assess the relationship between geographical and genetic distances ($N_{ST}$), we performed a Mantel test (Mantel, 1967) using the software XLSTAT (version 5.1; Addinsoft, Barcelona, Spain). The geographical distances between pairs of populations were calculated using GeoDis (version 2.6; Posada et al., 2000). The relationship among haplotypes was determined using a statistical parsimony network (Templeton et al., 1987, 1992) implemented in TCS (version 1.21; Clement et al., 2000). We included three congeneric species as outgroups, Oxalis latifolia (Rose) Knuth, Oxalis drummondii (Gray) and Oxalis magnifica (Rose) Knuth; all section Ionoxalis, to root the network.

Ecological niche modelling

We used the desktop version 1.1.6 of the Genetic Algorithm for Rule-set Production (GARP) modelling system (Stockwell & Peters, 1999; http://www.nhm.ku.edu/desktopgarp/index.html) to identify the current climatic conditions under which the species maintains populations (i.e. its ecological niche). This method is based on the relationship between georeferenced occurrence data of the species and a set of 19 climatic parameters (drawn from the WorldClim database; Hijmans et al., 2005), representing extreme, seasonal and mean conditions of temperature and precipitation in the area. The niche model was then projected onto the landscape to produce maps interpreted as the potential geographical distribution of the species (Soberón & Peterson, 2005).

To generate the niche-based distribution models, we used a set of 85 unique presence locations of O. alpina across the US and Mexico, from which 60 points were used for training the model and the remaining 25 for validating it using a chi-square analysis (Anderson et al., 2003). GARP is a ‘trial and error’ system that uses a genetic algorithm to characterize the ecological niche of a species and produce a presence/absence map representing the areas where conditions are suitable for that species to establish (i.e. its potential distribution). A full description of the GARP algorithm can be found in Stockwell & Peters (1999).

To produce a potential distribution map for the LGM, we used the desktop version 1.1.6 of the Genetic Algorithm for Rule-set Production (GARP) modelling system (Stockwell & Peters, 1999; http://www.nhm.ku.edu/desktopgarp/index.html) to identify the current climatic conditions under which the species maintains populations (i.e. its ecological niche). This method is based on the relationship between georeferenced occurrence data of the species and a set of 19 climatic parameters (drawn from the WorldClim database; Hijmans et al., 2005), representing extreme, seasonal and mean conditions of temperature and precipitation in the area. The niche model was then projected onto the landscape to produce maps interpreted as the potential geographical distribution of the species (Soberón & Peterson, 2005).

To generate the niche-based distribution models, we used a set of 85 unique presence locations of O. alpina across the US and Mexico, from which 60 points were used for training the model and the remaining 25 for validating it using a chi-square analysis (Anderson et al., 2003). GARP is a ‘trial and error’ system that uses a genetic algorithm to characterize the ecological niche of a species and produce a presence/absence map representing the areas where conditions are suitable for that species to establish (i.e. its potential distribution). A full description of the GARP algorithm can be found in Stockwell & Peters (1999).

To produce a potential distribution map for the LGM, the niche model was also projected onto a parallel geographical dataset representing the climatic conditions prevailing during the LGM according to the General Circulation Model of the Model for Interdisciplinary Research on Climate (MIROC, version 3.2; http://www.ccr.u-tokyo.ac.jp/~hasumi/MIROC/), which was integrated and downscaled by R. Hijmans and made available at the WorldClim website (http://www.worldclim.org/pastdown.htm). Final grids for the present were at 0.01° pixel resolution (~1 km²) and 0.04° (~25 km²) for the LGM.

Given all the random processes involved in the construction of the niche model with GARP, two maps generated using the same datasets are usually somewhat different; thus, we produced 100 individual maps and selected a subset of the ten best models based on their performance according to omission and commission errors, following Anderson et al. (2003). The ten best models were summed together using GIS to obtain a final model with pixel values ranging from 0 to 10, where 0 represents the areas where all models predicted the species to be absent, 1 is a pixel in which one of 10 models predicted the species to be present, and so on, to 10, which represents those areas where all models predicted the presence of the species. This map was then converted to a presence–absence map, selecting as the threshold value those pixels in which all occurrences were correctly predicted, in this case corresponding to values 9 and 10. The same subset of best models was selected and processed identically for the LGM, resulting in one final map for the present and one for the Pleistocene. GARP has been used successfully to produce reliable distribution maps both in the present and in the past (Martínez-Meyer et al., 2004; Martínez-Meyer & Peterson, 2006; Jakob et al., 2009).

Results

Genetic variation

The two cpDNA regions yielded a total of 1104 base pairs (bp), consisting of 348 bp from trnH-psbaA spacer and 756 bp from ndh F gene. We detected seven nucleotide substitutions and only one indel (insertion–deletion) event. The two concatenated DNA fragments produced a total of eight haplotypes, with the haplotype number varying from one to five among populations. Nucleotide diversity ($\pi$) for the total sample was 0.0021; total haplotype diversity ($h_t$) was 0.79 and average diversity within population ($h_s$) was 0.21. The highest level of genetic variation was found in a tristylos population from Sierra Los Ajos, whereas 11 populations showed no genetic variation at all (Table 1). Both nucleotide diversity (average $\pi$ for distylos populations = 0.00006, average $\pi$ for tristylos populations = 0.00064; $t_{18} = -2.71$, $P = 0.015$) and haplotype diversity (average $h$ for distylos populations = 0.057, average $h$ for tristylos populations = 0.278; $t_{18} = -2.299$, $P = 0.034$) showed that tristylos populations maintain higher levels of genetic variation than distylos populations. Regression analyses showed a significant and negative relationship between latitude and both nucleotide and haplotype diversity (Fig. 2: $R^2 = 0.16$, $F_{1,18} = 4.75$, $P = 0.04$; $R^2 = 0.15$, $F_{1,18} = 4.48$, $P = 0.04$, respectively). We also found a significant negative association between latitude and mid-styled morph frequency ($R^2 = 0.28$, $F_{1,15} = 7.35$, $P = 0.02$). Further analysis indicated that distylos populations are located at higher latitudes than tristylos populations ($F_{1,18} = 6.47$, $P = 0.02$).
Because distylos populations are located at the northern range, and there is lower genetic variation at higher latitudes, we evaluated whether the differences in the levels of genetic variation between distylos and tristylos populations were an artefact owing to geographical location. Because residuals from these analyses between latitude and both \( \pi \) and \( h \) represent the amount of genetic variation not accounted for by latitude variation, we compared these values between reproductive systems. In both cases, we obtained nonsignificant results, suggesting that differences in latitude, but not between reproductive systems, account for the observed genetic differences between distylos and tristylos populations (\( \pi, t = -1.15, P = 0.26; h, t = -1.01, P = 0.32 \)).

**Table 2** Analysis of molecular variance (AMOVA) among (a) 20 populations of Oxalis alpina in the Sky Islands region (b) reproductive system and populations within reproductive system (c) tristylos populations (d) distylos populations.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>20</td>
<td>73.40</td>
<td>0.292</td>
<td>74.53*</td>
</tr>
<tr>
<td>Within populations</td>
<td>236</td>
<td>23.60</td>
<td>0.100</td>
<td>25.47*</td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
<td>97.01</td>
<td>0.392</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among groups</td>
<td>1</td>
<td>45.80</td>
<td>0.349</td>
<td>24.48*</td>
</tr>
<tr>
<td>Among populations</td>
<td>19</td>
<td>190.98</td>
<td>0.802</td>
<td>56.21*</td>
</tr>
<tr>
<td>within groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>236</td>
<td>65.02</td>
<td>0.275</td>
<td>19.31*</td>
</tr>
<tr>
<td>Total</td>
<td>256</td>
<td>301.80</td>
<td>1.426</td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>14</td>
<td>48.70</td>
<td>0.269</td>
<td>67.16*</td>
</tr>
<tr>
<td>Within populations</td>
<td>172</td>
<td>22.69</td>
<td>0.131</td>
<td>32.84*</td>
</tr>
<tr>
<td>Total</td>
<td>186</td>
<td>71.40</td>
<td>0.401</td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>5</td>
<td>19.09</td>
<td>0.328</td>
<td>95.85*</td>
</tr>
<tr>
<td>Within populations</td>
<td>64</td>
<td>0.909</td>
<td>0.014</td>
<td>4.15*</td>
</tr>
<tr>
<td>Total</td>
<td>69</td>
<td>20</td>
<td>0.342</td>
<td></td>
</tr>
</tbody>
</table>

*Indicates significant values (\( P < 0.05 \)).

**Genetic structure**

AMOVA showed that almost 75% of the genetic variation was accounted for by differences among populations (Table 2). A hierarchical analysis further indicated that reproductive systems only explained 25% of the genetic variation, whereas more than half of the total genetic variation was accounted for by differences among populations within reproductive systems (Table 2). Results from two independent AMOVA’s aimed to determine the extent of genetic structure within each reproductive system, indicated that distylos populations are more differentiated than tristylos populations (95.85% and 67.16% of among populations variation, respectively; see AMOVA from Table 2).

SAMOVA produced an optimum of five genetically distinct groups (Figs 3, 5): group A (Azul, Buenos Aires, Purica, Ajos and Animas), group B (Elenita, Mariquita and San Jose), group C (Chiricahua Morse, Santa Rita and Huachuca), group D (Galiuro, Pinos Altos, Pinaleño, Chiricahua Pinery, White, Santa Catalina and San Luis) and group E (Pinal and Sierra Ancha). Excepting Los Ajos (group A), populations within groups did not show significant genetic differentiation. Distylos populations were assigned to three different groups (groups C, D and E). Group E included only distylos populations, whereas groups C and D were composed of both distylos and tristylos populations. The UPGMA phenogram of Oxalis alpina populations from the Sky Islands region as a function of latitude (a) haplotype diversity (b) nucleotide diversity. The lines represent the adjustment to the linear regression model.
recovered the same five groups detected by **SAMOVA**, with bootstraps values ranging from 55% to 100%. The lower bootstrap value corresponded to group C (57%), whereas the highest value corresponded to the branch that separate group E from the other four groups (100%; Fig. 3). The UPGMA phenogram detected three additional groups, one associated with southern populations (group A from **SAMOVA**), a second one including the central and northern populations (groups B, C and D from **SAMOVA**), and one group with the northwestern populations (group E from **SAMOVA**).

The Mantel test revealed a significant association between genetic and geographical distance ($r = 0.42$, $P = 0.0001$), suggesting the occurrence of isolation by distance (Fig. 4). Although we found significant genetic structure among populations of *O. alpina* at the Sky Islands region, we did not detect a phylogeographical structure, because $N_{ST}$ was not significantly higher than $G_{ST}$ ($N_{ST} = 0.78$, $G_{ST} = 0.73$, $P = 0.07$).

**Genealogy, frequency and geographical distribution of haplotypes**

The different haplotypes in this network are all separated by one mutational step, except haplotypes H4 and H2, and haplotypes H1 and H6, (Fig. 5). The frequency of haplotypes showed a markedly skewed distribution with a clear dominance of H2 (48%). This haplotype occurred in 13 out of 20 sampled populations. The second most frequent haplotype was H1 (15%), which occurred in six populations from the Sky Islands. Both haplotypes are located in a central position within the network with two mutational steps separating them. H3, in contrast, was the less frequent haplotype (8%) and only occurred in three populations. The Los Ajos population had the only private haplotype (H8), that is, a haplotype occurring in only one population.

Seven out of eight haplotypes were found in tristylos populations, whereas four occurred in distylos populations (Fig. 5). The H1, H2 and H3 haplotypes were shared between distylos and tristylos populations. H4, the haplotype located at the tip of the network two mutational steps away from H2, was found only in the most northern distylos populations (Sierra Ancha and Pinal). According to the outgroup species, the ancestral haplotype is H2. Coalescent theory makes the same prediction, because ancestral haplotypes are expected in the central position of the network (Templeton et al., 1995).
Niche modelling and the putative Pleistocene distribution of *O. alpina*

The niche model for the present distribution of *O. alpina* was highly predictive. All ten best models showed low values of omission error for the extrinsic validation points (0–8%), significantly better than random expectation ($\chi^2$ values > 150.0, $P = 0.0002$). In addition to the statistical performance, a visual inspection also indicates that models predicted regions of suitable habitat under current climate conditions that very precisely correspond with the known distribution of *O. alpina* in the Sky Islands, with no apparent major over prediction (Fig. 6a,b). A robust model for the present indicates that the critical variables determining the current distribution of the species are well characterized, a mandatory condition for projecting a model onto past climatic scenarios (Nogués-Bravo, 2009).

By projecting present-day models onto climate conditions for 18 000 years ago, *O. alpina* was predicted to have had two distinct distributional areas. One relatively small and fragmented area included the western part of central Arizona, ending just east the Sierra Ancha. A second, larger and more continuous distributional area included southeastern Arizona, southwestern New Mexico, northern Chihuahua and a small area in northern Sonora (Fig. 6c). The climate changes occurring since the Pleistocene appear to have produced dramatic shifts in the distributional area of this species (Fig. 6c). The formerly large and continuous populations from southeastern Arizona are now represented by a highly fragmented relictual distribution confined to high elevation mountains with almost no geographical overlap between the two time periods (Fig. 6c). Post-Pleistocene changes were also characterized by a northern distributional expansion involving the colonization of the White and Pinos Altos Mountains, whereas Sierra Ancha and Pinal may have been colonized by the northern refuge located at the western part of central Arizona (Fig. 6c). Finally, the process also included the colonization of high elevation peaks in Sonora, Chihuahua, southeastern Arizona, and southwestern New Mexico; probably from the large populations located in the southeastern and central portions of the Sky Islands distribution of *O. alpina* (Fig. 6c).

**Discussion**

*Oxalis alpina* harbours relatively high levels of genetic variation and has significant genetic structure among populations. In accordance with our expectations, there was a significant and negative relationship between latitude and both the frequency of the mid-styled morph and genetic diversity. These results support the hypothesis that range shifts produced by climatic oscillations influenced the genetic structure of *O. alpina*, but also the
evolution of reproductive systems because distylos populations are mainly located at ranges where founder events were likely to have had a prominent role. Nonetheless, the likelihood of multiple evolutionary transitions to distyly in several populations of the Sky Islands and the reduction of the frequency of the mid-styled morph as a consequence of incompatibility changes between long- and short-styled morphs suggest that selective processes also influenced the evolution of distyly. Once natural selection has reduced the frequency of the mid-styled morph, this morph is especially susceptible to loss during founder events and bottlenecks associated with climatic oscillations. 

*Oxalis alpina* showed significantly higher levels of cpDNA diversity than the average value obtained by Petit *et al.* (2005) for 32 perennial herbs (cpDNA data; $h_T = 0.79$ and $h_T = 0.69$, respectively; 95% confidence interval for a sample of 32 species = 0.6–0.76). This difference was not because of variation in the number of populations sampled in each study because there was not a significant relationship between the number of populations sampled and $h_T$ ($F_{1,27} = 1.43$, $P = 0.25$). Instead,
given that the estimation of $h_T$ was much higher than the average $h_S$, the relatively high level of genetic variation found in *O. alpina* could be better explained as a consequence of a strong genetic structure, but also as a consequence of high effective population sizes at the southern range of the distribution. Haplotype diversity was also high in comparison with two other Oxalis species (Zietsman *et al.*, 2009), *Oxalis tomentosa* ($h_T = 0.352$) and *Oxalis oligophylla* ($h_T = 0.395$), and equivalent to that of *Oxalis purpurea* ($h_T = 0.771$). Nonetheless, because these estimations are based in small sample sizes (2, 1 and 1 population for *O. tomentosa* (L.f.), *O. oligophylla* (Salter) and *O. purpurea* (L.), respectively), we do not know whether the amount of genetic variation of *O. alpina* is really higher than that of other Oxalis species.

Both the geographical distribution of genetic variation and niche modelling analyses suggest that most of the contemporary distribution of this species is a consequence of relatively recent colonization from a large and continuous distribution in the southern and central part of the Sky Islands that harboured high levels of genetic variation. Accordingly, the latitudinal impoverishment of genetic variation could be a consequence of a series of founder events followed by range expansion, especially in the northernmost populations. Although this interpretation holds for most populations, both the Sierra Ancha and Pinal populations seem to have followed a different path and probably originated from the fragmented populations located in the northwestern part of central Arizona. These two populations are fixed for a single haplotype that is at the tip of the network and is not shared by any other population. Hence, the genetic uniqueness and northern position of these populations, along with evidence derived from niche modelling analyses suggesting the presence of isolated *O. alpina* populations at relatively high latitudes during the LGM (18 000 years ago), further suggest these populations resulted from an ancient divergence event.

In addition to latitudinal differences in genetic variation, we found significant genetic structure among *O. alpina* populations. Some of the factors that have contributed to this pattern of genetic variation are the low seed dispersion and the restricted microhabitat of *O. alpina* contributed to this pattern of genetic variation. We found significant genetic structure among populations. Some of the factors that have contributed to this pattern of genetic variation are the low seed dispersion and the restricted microhabitat of *O. alpina*. Although historical factors could have influenced the evolutionary breakdown of distyly and the pattern of distribution of reproductive system of *O. alpina*, the available evidence suggests that distyly has probably evolved on several occasions in the Sky Islands. Genetic and ecological lines of evidence support this conclusion. First, the magnitude of genetic differentiation is 40% higher among distyly populations than tristyly populations, a result indicating that the crossability and fitness of the hybrids among distyly is not consistent with the high level of genetic differentiation found among distyly populations in this study. Second, distyly populations were found in three out of the five genetically homogeneous groups produced by *SAMOVA*. The same five groups were recovered by the UPGMA phenogram. Distyly occurred in two of the groups with distyly, and in these cases, no genetic differences were found between populations with
different reproductive systems, suggesting that distyly in both groups is very recently evolved. According to niche modelling, the distylos populations, Sierra Ancha and Pinal Mts. populations were isolated in a northwest refuge during last glacial maximum. In addition to this result, the occurrence of a haplotype two mutational steps (H4) from the ancestral haplotype (H2) suggests an ancient divergence of these populations. Third, phylodocography studies performed on jumping spiders (Habronattus pigilis, Griswold) and tree frogs (Hyla arenicola, Cope) indicate that during the LGM, there was a barrier to gene flow between the Galapagos, Pinaleño and Santa Catalina Mountains and the Huachuca and Santa Rita Mountains (Barber, 1999; Masía, 2000). Given that we also found the same pattern of genetic differentiation between these mountain ranges, the presence of distyly in both groups suggests distyly has evolved independently in these two complexes of mountains. These different patterns of evidence suggest distyly might have originated on three different occasions, although our results are preliminary and require further support with additional molecular markers.

An integrated approach from phylogeography and ecological niche modelling has provided a powerful tool to study the evolutionary history of *O. alpina* in the Sky Island region. We have found that historical factors (climatic oscillations) have had a major influence on the genetic structure with an impoverishment of genetic variation at northernmost ranges. Our results also indicate that the action of genetic drift alone on the evolution of distyly is less likely because under this hypothesis we would not expect consistent loss of the mid-styled morph in isolated populations. Theoretical models and empirical studies indicate that short-styled morph is more susceptible to loss under the action of genetic drift than the other morphs (Heuch, 1980; Eckert & Barrett, 1992; Husband & Barrett, 1992), which we have not observed in any populations of *O. alpina*. Therefore, the convergent evolution of distyly through the loss of mid-styled morph, the congruence in the floral morphology among distylos populations and the strong negative association between loss of incompatibility reactions and the reduction of mid-styled morph (Weller *et al.*, 2007; Sosenski *et al.*, 2010) suggest natural selection has played an important role in the evolution of this sexual system. In other distylos species of *Oxalis*, the mid-styled morph is always the floral morph that is lost, indicating that the selection model for *O. alpina* may be widespread in section Ionoxalis. However, because distyly in *O. alpina* is mainly distributed in areas where founder events have been stronger, it is possible that genetic drift has also played a role in loss of mid-styled morph. We propose that the combined effects of natural selection against the mid-styled morph and genetic drift resulting in further reductions in the frequency of the mid-styled morph have influenced the tristyly – distyly transition.

Acknowledgments

We extend thanks to Rubén Pérez for technical assistance; Luis Eguiarte and Juan Fornoni for the helpful comments on the manuscript and Michael Hickerson for discussion of ideas. J.P.-A. acknowledges the academic support received from the Doctorado en Ciencias Biomedicas and The Instituto de Ecologia of the Universidad Nacional Autónoma de México. She was supported by PhD fellowships from CONACYT and UNAM-DGEP, and by an international mobility scholarship from CONACYT and the Society of Systematic Biology. Our project was also supported by grants from UC-MEXUS, UNAM (DGAPA IN2178/03) and CONACYT (47858-Q).

References


Evolution of breeding systems of O. alpina


Received 13 January 2010; revised 1 July 2010; accepted 7 July 2010