Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing

Guangpeng Ren¹,²,³, Rubén G. Mateo¹,⁴, Jianquan Liu³, Tomasz Suchan¹, Nadir Alvarez¹, Antoine Guisan¹,⁴, Elena Conti⁵ and Nicolas Salamin¹,²

¹Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland; ²Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland; ³State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou, 730000 Gansu, China; ⁴Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, 1015 Lausanne, Switzerland; ⁵Department of Systematic and Evolutionary Botany and Boranic Garden, University of Zurich, Zollikkerstrasse 107, 8008 Zurich, Switzerland

Author for correspondence: Nicolas Salamin
Tel: +41 21 692 4154
Email: nicolas.salamin@unil.ch

Received: 4 July 2016
Accepted: 23 August 2016

New Phytologist (2016)
doi: 10.1111/nph.14221

Key words: demography, Himalayas, isolation by distance, phylogeography, population structure, Quaternary climatic changes, restriction site-associated DNA sequencing (RADseq).

Summary

- The effects of Quaternary climatic oscillations on the demography of organisms vary across regions and continents. In taxa distributed in Europe and North America, several paradigms regarding the distribution of refugia have been identified. By contrast, less is known about the processes that shaped the species’ spatial genetic structure in areas such as the Himalayas, which is considered a biodiversity hotspot. Here, we investigated the phylogeographic structure and population dynamics of *Primula tibetica* by combining genomic phylogeography and species distribution models (SDMs).
- Genomic data were obtained for 293 samples of *P. tibetica* using restriction site-associated DNA sequencing (RADseq). Ensemble SDMs were carried out to predict potential present and past distribution ranges.
- Four distinct lineages were identified. Approximate Bayesian computation analyses showed that each of them have experienced both expansions and bottlenecks since their divergence, which occurred during or across the Quaternary glacial cycles. The two lineages at both edges of the distribution were found to be more vulnerable and responded in different ways to past climatic changes.
- These results illustrate how past climatic changes affected the demographic history of Himalayan organisms. Our findings highlight the significance of combining genomic approaches with environmental data when evaluating the effects of past climatic changes.

Introduction

Biodiversity hotspots that harbor extremely high species richness are often associated with mountains (Myers et al., 2000). The origin and evolution of biodiversity in mountainous areas are highly dependent on historical orogenesis and associated climatic changes (Hoorn et al., 2010; Favre et al., 2014; Liu et al., 2014; Wen et al., 2014). The alteration of topography and past climatic changes associated with mountain uplifts can cause fragmentation of species distributions, which can lead to reduced gene flow between isolated populations. This process initiates allopatric divergence that can ultimately drive populations towards speciation (Mayr, 1963; Rice & Hostert, 1993). It has recently been proposed that mountain uplift can also result in divergence and speciation in the face of gene flow across a continuous altitudinal gradient (Filatov et al., 2016). In this context, climatic oscillations during the Quaternary could have reinforced allopatric divergence and driven intraspecific divergence as well as local adaptation (Davis & Shaw, 2001; Hewitt, 2004; Li et al., 2013; Liu et al., 2013; Schorr et al., 2013), as populations experienced repeated cycles of retreat to refugia and postglacial expansions (Abbott et al., 2000; Avise, 2000; Petit et al., 2003). The demographic changes involved in these range shifts affected the spatial patterns of genetic variation within and among populations (Hewitt, 2004). However, the detailed processes involved are still poorly understood in most species.

The Himalayas, especially their core region (i.e. the Qinghai-Tibet Plateau, QTP), comprise one of the key high-altitude biodiversity hotspots in the world (Myers et al., 2000). The uplift of the QTP created a large altitudinal gradient across the region spanning from 500 to 8848 m (Wu, 1987). The eastern Himalayas are associated with deep valleys and characterized mainly by a warm and wet climate (Liu et al., 2013; Fig. 1). By contrast, the central and western Himalayas are characterized by a cold and dry climate because of high mountains forming the southern ridge of the Himalayas (six mountain summits exceed...
8000 m; Favre et al., 2014) and the high average altitude (> 4000 m). The geological events created large and profound ecological heterogeneity (Li et al., 1995; Shi et al., 1998; Yin & Harrison, 2000), which potentially led to divergent selection and adaptation associated with different ecological niches that created numerous endemic species (Wu, 1987; Favre et al., 2014; Liu et al., 2014). It is also proposed that these geological events have provided opportunities for species to migrate out of the region (Liu et al., 2006; Jia et al., 2012; Zhou et al., 2013; Wen et al., 2014; Ren et al., 2015). Although the region is assumed to be particularly vulnerable to climatic changes (Zheng, 1996; Yao et al., 2007), the pattern and extent of glaciation during the Quaternary and their effects on the evolutionary history of species within the Himalayas have not yet been fully examined, especially based on population genomic data.

By contrast, large-scale phylogeographic studies based mainly on a few plastid DNA regions have been conducted on species occurring in the QTP (e.g. Zhang et al., 2005; Meng et al., 2007; Yang et al., 2008; Wang et al., 2009; Shimono et al., 2010; Qiu et al., 2011; Li et al., 2013; Liu et al., 2013). The existence of a deep divergence between the Himalayan populations and those occurring in other regions of the plateau was already inferred, and extensive private haplotypes have been found in the Himalayan populations (e.g. Opgeethoven et al., 2010; Wang et al., 2010b; Jia et al., 2011), implying that multiple plant refugia probably existed in the Himalayas. However, these studies were unable to detect the detailed effects of past climatic changes on the demographic history of the studied organisms. Next-generation sequencing (NGS) methods (Davey et al., 2011), such as restricted site-associated DNA sequencing (RADseq; Peterson et al., 2012), which have been shown to be highly effective in tracing postglacial recolonization and reconstructing detailed demographic histories of species (e.g. Emerson et al., 2010; Lanier et al., 2015), could provide opportunities to better understand the effects of past climatic changes in driving speciation and evolution of alpine organisms in the Himalayas.

In this study, we focus on *Primula tibetica* (Primulaceae), one of the most widely distributed alpine plant species in the Himalayas (Hu & Kelso, 1996; Richards, 2003). *P. tibetica* is an insect-pollinated (mostly by bees), heterostylos, herbaceous, perennial plant that occurs in diverse habitats at elevations ranging from 2600 to 5000 m (Hu & Kelso, 1996). Its scape is sometimes hidden among the leaves or can be as long as 13 cm. *P. tibetica* is an outcrossing small herb of variable height (2–13 cm) that disperses its seeds largely by gravity and usually grows in wet meadows or along hill streams (Hu & Kelso, 1996; Richards, 2003). Previous biogeographic analyses have indicated that *P. tibetica* originated in the Himalayas after the recent QTP uplift (i.e. 3.4–1.6 million yr ago (Ma); Ren et al., 2015) and subsequent climatic oscillations during the Quaternary are likely to have played important roles in its intraspecific divergence and demographic history. This herbaceous species hence represents an ideal candidate to evaluate the effects of past climatic changes on a species’ evolutionary history in the Himalayas. We use an integrative approach combining genomic phylogeography with niche modeling (e.g. Schott et al., 2012) to elucidate the divergence and demographic history of *P. tibetica*. The aims of our study are to identify the phylogeographic pattern of this species in the Himalayas and the factors that triggered its intraspecific divergence; to reconstruct a detailed demographic history of *P. tibetica*; and to combine species distribution models with approximate Bayesian computation (ABC) modeling to evaluate the effects of Quaternary climatic changes on its demographic history. This study represents the first RADseq analysis of a plant species occurring in the QTP and contributes to a better understanding of the role played by Quaternary climatic changes on the present-day distributions of organisms in mountain ranges.

**Materials and Methods**

**Sampling, RAD library preparation and sequencing**

We sampled a total of 61 populations (10–40 individuals for each population) of *Primula tibetica* Watt in Tibet using the distribution described in Flora of China (Hu & Kelso, 1996) as a reference to include all the relevant regions for the species. All materials were dried and stored in silica gel in the field. We selected 16 populations for the genetic study (Fig. 1; Supporting
Information Table S1) that were representative of both the geographical distribution and the diversity of ecological niches of *P. tibetica*. We estimated the latter by extracting the 19 bioclimatic variables of WorldClim (http://www.worldclim.org/current) from the occurrences of the individuals sampled in the 61 populations. We did a principal component analysis (PCA) using the prcomp function in the stats package of R (R Core Team, 2012) and identified the 16 populations based on the PC1 and PC2 axes (which explained nearly 80% of the variance; Fig. S1). Fifteen to 20 individuals were chosen from each population, which gave us a total of 293 individuals that were processed with RADseq. The leaf tissues were ground to dust using an electric tissue homogenizer. Total genomic DNA was then isolated using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer’s instructions. The extracted DNA was further cleaned with phenol-chloroform to remove salts or inhibitors that may reduce the activity of restriction enzymes.

The cleaned genomic DNA was individually barcoded and processed into three libraries using a double-digestion restriction fragment-based procedure following a modified protocol listed in the Supporting Information of Mastretta-Yanes et al. (2015). Briefly, the DNA was double-digested with EcoRI and MseI restriction enzymes, followed by the ligation of Illumina adapter sequences and unique 8 bp barcodes that differed by at least three bases. Ligation products were purified with AMPure XP beads (Beckman Coulter, Brea, CA, USA) and amplified by Phusion High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) with 12 cycles. The amplified products were pooled among samples and size-selected between 300 and 500 bp using AMPure XP beads with bead sample ratios of 0.8 and 0.2 modified from a protocol in https://www.neb.com/protocols/1/01/size-selection-e6270. The libraries were sequenced using single-end reads 100 bp in length in three lanes of Illumina HiSeq2500 according to the manufacturer’s instructions.

**Processing of Illumina data**

Single-end Illumina reads were processed into RAD-tags using the STACKS-1.30 software pipeline (Catchen et al., 2011, 2013) based on its ease of use, features and performance (Davey et al., 2013). Initially, samples were demultiplexed with process_radtags. Reads with an average Phred score of at least 30 and an unambiguous barcode and restriction cut site were retained. All reads were trimmed to 60 bp in length. The raw data were deposited in GenBank (accession no. PRJNA339808). Next, the USTACKS program was used for the de novo assembly of raw reads into RAD-tags. We used all 293 samples to build a catalog in CSTACKS and matched each sample against the catalog to identify alleles in SSTACKS. The execution of these components was accomplished using the denovo_map.pl script with the following settings: minimum number of reads to create a stack, $m=3$; maximum distance allowed between stacks, $M=2$; maximum number of mismatches allowed between loci, $n=2$; -f flag to remove or break up highly repetitive RAD-tags during the USTACKS component and upper bound of error rate, $\varepsilon=0.1$. A conservative bound was preferred over the unbounded model because the latter has been shown to underestimate heterozygotes (Catchen et al., 2013). We used RXSTACKS to further filter the data to increase quality, correct single nucleotide polymorphism (SNP) calls and remove haplotypes that were in excess. The RXSTACKS used the output from the denovo_map.pl script as input combined with the following filters: -conf_filter -conf_lim 0.25 -prune_haplo -model_type bounded -bound_high 0.1 -lnl_dist -10.0. After RXSTACKS, CSTACKS and SSTACKS were run again with the same setting as before to rebuild the catalogue of reads. To test the sensitivity of our results to different sets of parameters, we further processed our RAD data with two other parameter settings: using the same settings as earlier except for $M=3$ and $n=3$, and trim the reads to 90 bp in length ($M=3$, $n=3$ and 90 bp); and $M=5$, $n=3$ and 90 bp. The results of the population structure analyses based on the three datasets were

![](image)

**Fig. 2** Distribution of individuals of *Primula tibetica* along principal component (PC) scores (PC1, 20% vs PC2, 9.4%; PC1 vs PC3, 6.6%) of genetic variation based on the analysis of single nucleotide polymorphism (SNP) dataset; individuals are color-coded according to their population identities (see Fig. 1).
qualitatively similar (Figs 2, S2), and we only presented results from our analyses based on the dataset generated by $M=2$, $n=2$, given the increased number of assembled loci (3509 vs 2822 vs 2031).

We filtered the catalog of reads using the POPULATIONS module to produce datasets for downstream population genetic analyses. We first retained RAD-tags with a minimum stacks depth, $m=3$. Polymorphic RAD loci that were present in at least 50% of the individuals of each population and in all 16 populations were retained. Potential homologs were excluded by removing loci showing heterozygosity > 0.5 within samples (Hohenlohe et al., 2011). We further filtered our dataset with a minor allele frequency (MAF) > 0.01 and kept only biallelic SNPs to comply with the assumptions of the current methods for analyzing SNP data. Population genetic statistics, including nucleotide diversity ($\pi$), Wright's $F$-statistic ($F_{ST}$) and observed heterozygosity ($H_{obs}$) were calculated using the POPULATIONS program in the STACKS pipeline (Holsinger & Weir, 2009; Catchen et al., 2013). Pairwise $F_{ST}$ values were calculated among populations in GENODIVE v.2.0b27 (Meirmans & Van Tienderen, 2004), and significance was determined using $1 \times 10^4$ permutations.

Characterization of population genetic structure

We first identified population genetic structure using the Bayesian method implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000). SNPs located at the same locus are physically linked and cannot be handled by STRUCTURE. We thus filtered out linked SNPs using the ~write_single_snp option in the POPULATIONS script. Analyses were performed under the ‘Admixture model’ and the ‘Correlated allele frequency model’ with $K$-values ranging from 1 to 18. Ten independent runs were performed for each value of $K$ using $1 \times 10^5$ generations for the burnin and $2 \times 10^5$ generations for the sampling. The optimal $K$ was chosen using the delta-$K$ method of Evanno et al. (2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). The coefficient for cluster membership of each individual was averaged across the 10 independent runs using CLUMPP (Jakobsson & Rosenberg, 2007) and plotted using DISTRUCT (Rosenberg, 2004). We further performed a PCA to visualize the major axes of variation of the population genetics using the adegenet package (gPCA function; Jombart, 2008) in R. Finally, we estimated a maximum-likelihood phylogeny of the 16 populations from unlinked SNPs with a GTR+G model using PHYML 3.0 (Guindon et al., 2010). Primula matsumi and Primula fasciculata were used as outgroups. Nodal support was estimated using 1000 bootstrap replicates.

Relationships between genetic differentiation and geography

The first two components of the PCA performed on the genetic data and the geographic coordinates (latitude and longitude) of the 16 populations were used in a Procrustes analysis using the R package VEGAN (Oksanen et al., 2013). This analysis minimizes the sum of squared Euclidean distances between two sets of points by rotating one set of points to match the other, while preserving the relative distances among all points within the map (Wang et al., 2012). The similarity of the two maps is quantified using the Procrustes similarity statistic $b = \sqrt{1 - D}$, where $D$ is the minimum sum of the squared Euclidean distances between the two maps, scaled to range between 0 and 1 (Wang et al., 2010a, 2012). We used the PROTEST function in VEGAN to test, using $1 \times 10^5$ permutations, the probability of observing a similarity statistic higher than the observed $b$ if no geographic pattern is assumed (Wang et al., 2012). We also tested for the presence of isolation by distance (IBD) by comparing pairwise $F_{ST}$ values and Euclidean geographic distances among populations within and among groups that were identified by the PCA and STRUCTURE analyses. We further tested the significance of the relationship between geographic and genetic distance within groups with a Mantel test in the package VEGAN using $1 \times 10^5$ permutations.

Estimates of historical demography

To decipher the historical demography of $P. tibetica$, we estimated divergence times, admixture and changes in population size among different population groups using ABC. We pooled the population samples into four ‘groups’ (eastern group E, R01–R02; central-eastern group CE, R03; central group C, R04–R12; western group W, R13–R16) for the ABC simulations based on the first two axes of the PCA that captured the main characteristics in population histories (Fig. 2). We tested three competing scenarios using DIY-ABC v.2.1.0 (Cornuet et al., 2010, 2014) based on the results from STRUCTURE and the phylogenetic tree (Fig. S3). In all scenarios, groups E and C diverged first and group W originated from group C. The scenarios modeled the possible hypotheses about the origin of the group CE, which can arise from either groups E or C, or be the result of an admixture between the two groups (Fig. S4a). We selected for these analyses a single SNP per locus, and the SNPs further had to be present in at least 70% of the individuals from each group and in all four groups. The simulated SNP dataset was generated following the algorithm proposed by Hudson (2002). We further chose MAF = 0.01 to increase the mean amount of genetic variation of both the observed and simulated datasets and to reduce the proportion of loci that may correspond to sequencing errors. We gave each scenario a uniform prior probability (Table S2) and selected all summary statistics to generate a reference table containing $3 \times 10^6$ simulated datasets (on average $10^6$ per scenario). We used 1% of the simulated datasets closest to the observed data to estimate the relative posterior probabilities for each scenario via logistic and posterior distribution of historical demographic parameters according to the most likely scenario (Cornuet et al., 2010). The time parameters are estimated in generations and converted into years by multiplying generation time, which was set to 1 yr for $P. tibetica$. Although there is no information of generation time for $P. tibetica$, filed observations are coherent with this assumption and other studies on related species of Primula have also used a generation time of 1 yr to study demographic history of $P. obconica$ (Yan et al., 2012). In addition, we also
considered the substructure (R11–R12) identified by the PCA and Structure as a fifth group for ABC modeling (Fig. S5). However, simulations based on five groups were not stable enough to provide a convincing outcome compared with the ABC modeling with four groups, which could further indicate that these two populations do not form a homogeneous cluster (see Notes S1 for a full description of the ABC modeling with five groups).

Finally, DIY-ABC was used to investigate changes in population sizes of the four groups in the recent past. We first selected only one SNP per locus and used two thresholds (i.e. SNP had to be present in at least 70% vs 80% of the individuals in each group) to generate datasets for each group. We then did PCA based on these datasets, and the two thresholds resulted in similar structure patterns for each group (Fig. S6). We used the datasets generated based on the 80% threshold for these ABC analyses, because they have fewer missing data and it saved computational time. We tested the following scenarios of demographic changes: continuous expansion since divergence; recent expansion; expansion followed by shrinkage; and expansion followed by shrinkage and a new expansion event (Fig. S7a; Wang et al., 2016). We used the same strategy as earlier to choose the most likely scenario and estimate the parameters of interest.

Species distribution models

An ensemble of species distribution models (SDMs; Guisan & Zimmermann, 2000) was generated for P. tibetica using three different techniques: generalized linear model, gradient boosting machine and random forests, as implemented in the R package biomod2 (Thuiller et al., 2009; see Methods S1 for similar results; Fig. S8). A total of 58 species occurrences obtained directly from the filed collections were used as presence data to calibrate the models. We used the 19 bioclimatic variables of Worldclim (http://www.worldclim.org; Hijmans et al., 2005) as environmental predictors. To avoid multicollinearity, we ran a Pearson correlation analysis to eliminate one of the variables in each pair with a correlation value higher than 0.8 (Dormann et al., 2013). A set of seven variables was finally used to carry out the SDM (Methods S1). For a proper evaluation, models were calibrated on 70% of the data and evaluated on the remaining 30% using area under the curve (AUC) and true skill statistic (TSS) statistics (Allouche et al., 2006). This sampling procedure was replicated 10 times. The potential distribution was considered as a consensus across statistical techniques (Mateo et al., 2012) and their contribution to the ensemble was proportional to their AUC values. The consensus model was converted to a binary model (presence/absence) applying three different threshold criteria (Methods S1): thresholds that allow a maximum of 5% or 10% of omission error (i.e. omission error is the percentage of the real presence predicted as absences in the model; Fielding & Bell, 1997), and the threshold maximizing the AUC statistic. The consensus model was then projected onto different past climatic periods using the data available in the Worldclim dataset: the last interglacial (LIG; 0.12–0.14 Ma), the Last Glacial Maximum (LGM; 0.022 Ma), and the mid-Holocene (MH; 0.006 Ma). For the MH and LGM we employed three different general circulation models (GCMs; Earth-system climatic models coupling the ocean, the atmosphere and the land surface; CCSM4, MIROC-ESM, MPI-ESM-P available from http://cmip-pcmdi.llnl.gov/cmip5/processed on www.worldclim.org). Only one GCM is available for the LIG period.

Results

Sequence data quality and processing

We sequenced 293 individuals of P. tibetica using three lanes of Illumina that produced a total of > 730 million reads. Over 560 million reads passed our quality controls and over 460 million reads were used in the assembly of the RAD-tags (Table S3). We obtained 3509 RAD loci containing 8930 SNPs that could be used for population genetics analyses. The dataset was used to estimate historical scenarios of P. tibetica containing 4882 single-SNP loci. Finally, four datasets containing 8579, 5401, 7777, and 10 431 single-SNP loci were used to estimate the changes in population sizes of groups E, C, CE and W, respectively.

Population structure

The first two axes of PCA identified four genetic groups and explained 20% and 9.4% of the total variation, respectively (Fig. 2). The first axis, PC1, showed considerable correspondence between the genetic data and the east–west geographic axis. The two eastern populations (R01, R02) and four western populations (R13–R16) formed two separate groups (groups E and W) that were located on the two extreme sides of the distribution. One central-eastern population (R03) and the rest of populations (R04–R12) were further isolated from the groups E and W by the second axis of the PCA (PC2; Fig. 2) and formed two other groups (groups CE and C), respectively. The third axis of the PCA (PC3; 6.6% of the total variation) showed a substructure within group C, with four populations separating gradually from the rest of five populations following the increase of geographic distance (Figs 1, 2). This pattern of population structure was also supported by the Structure analysis, which best explained the data with \( K = 4 \) (Fig. 3). Looking at intermediate \( K \)-values, the analyses showed that, at \( K = 2 \) (the second most probable number of genetic clusters; Fig. S9), group E first diverged from the rest of the populations (Fig. 3), which was also evident in the phylogenetic tree (Fig. S3). Group CE was always represented by admixed populations between the groups E and C at any values of \( K \) between 2 and 4 (Fig. 3). By contrast, the substructure (R09–R12) within group C identified by the PC3 was not always represented by admixed populations in Structure from \( K = 2 \) to \( K = 4 \). Moreover, the populations comprising this substructure were not clustered together along the PC3. We therefore did not include this substructure when performing ABC modeling (see more details in Notes S1 for the reason not including the substructure in ABC analyses). Finally, the Procrustes analysis identified a significant similarity score between the populations in
genetic PC space and their actual geographic locations ($t_0 = 0.815$, $P < 10^{-5}$). A graphical examination of the rotated genetic coordinates (Fig. 4) showed that individuals of *P. tibetica* were more genetically similar within each group than would be expected given the geographic distance among populations.

Genetic diversity and IBD

The average within-population genetic diversity $\pi$ ranged from 0.0011 to 0.0044, when considering all genetic positions, including those not polymorphic anywhere in the dataset (Table 1). Group E exhibited the lowest genetic diversity, which was three times lower than the diversity measured in groups CE and C, or two times lower than that of group W. The same pattern was also suggested by other standard measures of genetic diversity (e.g. observed heterozygosities; Table 1).

Differentiation among populations was significant, with $F_{ST}$ values ranging from 0.032 to 0.807 (Table S4). Genetic distances between populations of different groups increased with geographical distances > 200 km, but populations among groups located at smaller geographical distances displayed high genetic divergence (Fig. 5a). The genetic distance between populations of the same group was, however, always smaller than the distances among groups, which is congruent with the strong genetic structure observed in *P. tibetica* (see earlier). Furthermore, genetic distances increased with larger geographic distances among populations within groups (Fig. 5a), which was consistent with the significant pattern of IBD when performing a Mantel test among populations of group C ($r = 0.51$, $P = 0.016$; Fig. 5b).

Although genetic distances among populations of group W were small (Table S4), we found a strong effect of IBD on population differentiation of this group ($r = 0.99$, $P = 0.042$; Fig. 5b).

Estimates of historical demography

Approximate Bayesian computation modeling of the demographic history of *P. tibetica* indicated that the scenario depicting an origin of group CE as a result of admixture between groups C and E provided the best fit to our RADseq data (Fig. S4b), with posterior probabilities significantly higher than the other
scenarios (0.816, 95% credible interval: 0.797–0.834; Table S5). Modeling the changes in population size for each group recovered complicated demographic histories for the four groups of populations. Analyses for groups E and CE supported a scenario of ‘expansion–shrinkage–expansion’, while the two other groups were better modeled by a scenario of ‘expansion–shrinkage’ (Fig. S7; Table S5).

We estimated the divergence time and the population sizes as well as the timing and extent of these changes for the four groups. Group C was found to be the ancestral population of *P. tibetica* and started to expand its distribution c. 1.11 Ma (95% highest posterior density (HPD): 0.53–1.65 Ma; Table S6), followed by a slight bottleneck c. 0.063 Ma (HPD: 0.007–0.136 Ma). Group E diverged from the ancestral populations c. 0.76 Ma (HPD: 0.49–0.96 Ma; Table S7). It started to expand until c. 0.45 Ma (HPD: 0.15–0.92 Ma), before experiencing a severe bottleneck that decreased by c. 25 times its population size c. 0.12 Ma (HPD: 0.063–0.2 Ma). Then it quickly expanded just before LGM c. 0.037 Ma (HPD: 0.011–0.078 Ma) and reached the previous population size. During the first expansion of group E, it came into secondary contact with group C, exchanged genes and resulted in the formation of group CE c. 0.37 Ma (HPD: 0.213–0.525 Ma). Group CE experienced ancient expansion and shrinkage, and a recent expansion during the LGM (Fig. 6). Group W diverged from the ancestral population more recently, c. 0.095 Ma (HPD: 0.037–0.203 Ma), followed by expansion and a slight bottleneck during the LGM.

### Species distribution models

The consensus models were highly accurate with regard to AUC (0.966) and TSS (0.998) values. Current potential distribution based on the three threshold approaches predicted similar results,

<table>
<thead>
<tr>
<th>Genetic cluster</th>
<th>Population</th>
<th>n</th>
<th>Private (%)</th>
<th>Polymorphic (%)</th>
<th>π</th>
<th>H_{obs}</th>
<th>F_{IS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group E</td>
<td>R01</td>
<td>15</td>
<td>0.22</td>
<td>0.40</td>
<td>0.0013</td>
<td>0.0012</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>R02</td>
<td>18</td>
<td>0.19</td>
<td>0.34</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>R03</td>
<td>15</td>
<td>2.58</td>
<td>1.13</td>
<td>0.0037</td>
<td>0.0029</td>
<td>0.0021</td>
</tr>
<tr>
<td>Group CE</td>
<td>R04</td>
<td>12</td>
<td>1.15</td>
<td>1.33</td>
<td>0.0038</td>
<td>0.0028</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>R05</td>
<td>15</td>
<td>1.31</td>
<td>1.50</td>
<td>0.0042</td>
<td>0.0033</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>R06</td>
<td>16</td>
<td>0.64</td>
<td>1.60</td>
<td>0.0041</td>
<td>0.0033</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>R07</td>
<td>17</td>
<td>10.64</td>
<td>1.82</td>
<td>0.0043</td>
<td>0.0035</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>R08</td>
<td>15</td>
<td>0.11</td>
<td>1.80</td>
<td>0.0044</td>
<td>0.0035</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>R09</td>
<td>15</td>
<td>0.36</td>
<td>1.40</td>
<td>0.0038</td>
<td>0.0031</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>R10</td>
<td>13</td>
<td>0.31</td>
<td>1.29</td>
<td>0.0038</td>
<td>0.0031</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>R11</td>
<td>16</td>
<td>0.48</td>
<td>0.67</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>R12</td>
<td>16</td>
<td>2.41</td>
<td>1.08</td>
<td>0.0031</td>
<td>0.0025</td>
<td>0.0019</td>
</tr>
<tr>
<td>Group C</td>
<td>R03</td>
<td>15</td>
<td>2.58</td>
<td>1.13</td>
<td>0.0037</td>
<td>0.0029</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>R04</td>
<td>12</td>
<td>1.15</td>
<td>1.33</td>
<td>0.0038</td>
<td>0.0028</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>R05</td>
<td>15</td>
<td>1.31</td>
<td>1.50</td>
<td>0.0042</td>
<td>0.0033</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>R06</td>
<td>16</td>
<td>0.64</td>
<td>1.60</td>
<td>0.0041</td>
<td>0.0033</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>R07</td>
<td>17</td>
<td>10.64</td>
<td>1.82</td>
<td>0.0043</td>
<td>0.0035</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>R08</td>
<td>15</td>
<td>0.11</td>
<td>1.80</td>
<td>0.0044</td>
<td>0.0035</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>R09</td>
<td>15</td>
<td>0.36</td>
<td>1.40</td>
<td>0.0038</td>
<td>0.0031</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>R10</td>
<td>13</td>
<td>0.31</td>
<td>1.29</td>
<td>0.0038</td>
<td>0.0031</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>R11</td>
<td>16</td>
<td>0.48</td>
<td>0.67</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>R12</td>
<td>16</td>
<td>2.41</td>
<td>1.08</td>
<td>0.0031</td>
<td>0.0025</td>
<td>0.0019</td>
</tr>
<tr>
<td>Group W</td>
<td>R13</td>
<td>18</td>
<td>0.10</td>
<td>1.02</td>
<td>0.0029</td>
<td>0.0026</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>R14</td>
<td>18</td>
<td>0.01</td>
<td>1.09</td>
<td>0.0029</td>
<td>0.0026</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>R15</td>
<td>15</td>
<td>0.00</td>
<td>1.08</td>
<td>0.0030</td>
<td>0.0025</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>R16</td>
<td>13</td>
<td>0.10</td>
<td>0.84</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

Included are the average number of individuals genotyped at each locus (n), the proportion of polymorphic single nucleotide polymorphisms (SNPs) unique to each population (% private), the percentage of SNPs (% polymorphic) in each population, the average nucleotide diversity (π), the average observed heterozygosity per locus (H_{obs}) and the Wright’s inbreeding coefficient (F_{IS}). The total number of DNA sites (polymorphic + invariable) in the RADSeq loci is 210 540. E, eastern group; CE, central-eastern group; C, central group; W, western group.
but the 5% omission error generally yielded a better representation of the actual distributions of the species, and we therefore presented all results based on the 5% omission threshold. The paleoclimatic conditions of LIG predicted large differences in annual mean precipitation in the Himalayas compared with the ones observed at the present, the MH or the LGM (Table S8). Therefore, it was not possible to predict the optimum climatic niche for the species during the LIG in this area considering the only available GCM model (Fig. S10; Methods S1). The predictions to MH conditions based on three GCMs (CCSM4, MIROC and MPI) yielded a continuous and less occupied overall distribution compared with current conditions, but larger distributions than the prediction at the LGM (Figs 7, S10). During the LGM, the three GCMs yielded similar patterns but fragmented palaeodistributions of *P. tibetica* (Figs 7, S10). All three GCMs suggested a main refugium in the central Himalayas and another in the southwestern Himalayas. The incongruence between models at the LGM yielded eastern or western expansions of suitable habitat compared with the predictions for the present and MH.

Discussion

*Primula tibetica* displays a strong geographic structure and we identified four main groups of populations that may represent multiple past refugia for this species in the Himalayas. IBD had an effect on genetic distance among populations within groups but not among groups. Instead, past climatic events were inferred to be the major factors in shaping the large-scale spatial genetic structure into four groups. The divergent times of the four groups based on ABC modeling are dated to \(< 1\) Ma and the divergences are congruent with past glacial and interglacial events, providing support for intraspecific divergence driven by the Quaternary climatic oscillations. The use of genomic data coupled with extended evolutionary modeling allowed us to recover for the first time a detailed demographic history of a plant species native and endemic to the Himalayas. The changes in population sizes that we inferred, combined with species distribution modeling, suggest that the two easternmost and westernmost gene pools were more affected by past climatic changes than the ancestral populations. The response to climatic changes of populations of a species depends on its specific ecological preferences, and the range dynamics identified for this cold-tolerant species during the last glaciation differ from species associated with warmer environments.

Multiple refugia and IBD

The use of genomic data allowed us to identify four distinct groups of populations for *P. tibetica*, which occupy the eastern, central-eastern, central and western areas of the species distribution (Figs 1, 2). These results, as well as the projected habitat at the LGM (Fig. 7c), suggest that multiple potential allopatric refugia existed for this species, probably located in the eastern, central and southwestern Himalayas. Although previous studies have...
found extensive private haplotypes in populations of diverse species and suggested multiple plant refugia in the Himalayas (Opseqnoolth et al., 2010; Wang et al., 2010b; Jia et al., 2011), the clear pattern identified by our genomic-level data was not yet described in the region. For example, Opseqnoolth et al. (2010) found that private haplotypes were evenly spread across the distribution range of a juniper complex, indicating that these junipers maintained multiple glacial (crytic) refugia throughout their current range and underwent only localized postglacial expansions. The use of plastid and nuclear markers, which provide less resolution compared with genomic-level data, may prevent the detection of a clear pattern.

Procrustes analysis shows a high similarity score between the overall rotated genetic space and their geographic locations (\(g_0 = 0.815, \ P < 10^{-3}\); Fig. 4), which is probably a result of the large-scale spatial genetic structure shaped by the refuge-driven vicariance. Long-distance dispersal and gene flow that may disturb this pattern of population structure is unlikely in P. tibetica, because this small herb (2–13 cm) is pollinated mainly by insects (e.g. bees) and disperses its seeds largely by gravity (Richards, 2003). Its poor ability to disperse, associated with the extreme altitudinal gradient present in the Himalayas, has probably caused fragmentation, reduced gene flow and further reinforced the genetic structure (Liu et al., 2014; Wen et al., 2014). IBD plays a minor role in the large-scale pattern of population structure in P. tibetica (Fig. 5a). However, at narrow scales, there are IBD effects on the genetic distance of populations within groups (Fig. 5). The decrease of genomic similarities between populations within groups is probably a result of limited dispersal among populations (e.g. Ferchaud et al., 2010; Lanier et al., 2015). However, separating the specific effects of geography and the environment on population structure is difficult (Thorpe et al., 2008; Wang et al., 2013). Our results show some differentiation of the ecological niches of the population (Fig. S1), but finer-scale analyses are needed to identify and quantify the importance of these variables (e.g. Lexer et al., 2014).

Quaternary climatic oscillations trigger intraspecific divergence in Primula tibetica

The genomic data presented here provide clear evidence that intraspecific divergence in P. tibetica was mainly driven by Quaternary climatic oscillations. The effects of Quaternary climatic oscillations on the distribution patterns and phylogeographic structure of species in the mid- to high-latitude regions of Europe and North America (Comes & Kadereit, 1998; Abbott et al., 2000; Avise, 2000; Hewitt, 2004; Anderson et al., 2006; Emerson et al., 2010), and in high-altitude areas (Qiu et al., 2011; Liu et al., 2014; Wen et al., 2014; Sun et al., 2015) have already been described. However, no studies yet exist for the Himalayas, and our analysis therefore provided a unique opportunity to uncover the detailed Quaternary demographic history of high-altitude populations and to better understand the processes playing a role in their distribution in this region.

The time frame of the first divergence between the eastern and central populations (groups E and C; Fig. 6) is congruent with the largest Naynayxungla glaciation in the QTP. This event reached its maximum between 0.8 and 0.5 Ma with an ice sheet covering an area five to seven times larger than its current range (Shi, 2002; Zheng et al., 2002). Such an extensive ice sheet could have caused fragmentation of ancestral populations and triggered the earliest divergence into two groups. The formation of the admixed central-eastern population (group CE) was dated to c. 0.37 Ma (HPD: 0.213–0.525 Ma) and coincides with the old expansion of group E (Fig. 6). During this period, the glaciation became progressively less extensive, but a cold climate is thought to have prevailed in the QTP until 0.17 Ma (Shi, 2002). The old expansion of group E may have been favored by such a cold climate, eventually resulting in a secondary contact with group C and the formation of group CE. Group W diverged from group C most likely during the LIG period when the climate was warmer (Thompson et al., 1997; Shi et al., 1998; Zheng et al., 2002) and may have allowed the ancestral populations to colonize the western high-altitude region.

Demographic history of Primula tibetica

Our analyses of the demographic history of each group of populations show that all have experienced ancient expansions followed by bottlenecks (Fig. 6). The western, central and central-eastern groups of populations that occur at high altitudes have experienced only slight bottlenecks during the last glaciation (Fig. 6), a period that started from 0.07 Ma and continued until the end of the LGM in the QTP (0.01 Ma; Thompson et al., 1997; Zheng et al., 2002). Our ABC modeling of changes in population sizes shows that populations comprising group C experienced the most ancient expansion c. 1.11 Ma (HPD: 0.53–1.65 Ma), which indicates that the origin of this species probably occurred in the central Himalayas (Fig. 6). The time estimated for the most ancient expansion of this species is congruent with the divergent time from its two closely related species obtained from previous phylogenetic study (1.19 Ma; HPD: 0.51–2.13 Ma; Ren et al., 2015). The current populations of group C occur at an average altitude of 4260 m (Table S1) and are thus probably adapted to live in cold environments. Their tolerance to cold might thus have facilitated the persistence of populations at high-altitude glacial refugia during past glaciations (Fig. 7c).

By contrast, the eastern populations (group E), which occur at the lowest altitude (average 2887 m; Table S1), experienced a severe bottleneck during the LIG period, but expanded during the LGM. The unusual demographic history of the eastern populations (group E) can be explained by the warmer climate in this region of the Himalayas, which displays a difference of > 8°C in comparison with the region of the central populations (assuming that current temperature in the Himalayas decreases by 0.64°C per 100 m; Li & Zhang, 2010). The warmer interglacial period could have been detrimental for a cold-adapted species, whereas the population expansion during the LGM corresponds to a period of colder climate more similar to the situation that prevailed for its ancestral populations, but warm enough in the eastern regions to avoid extensive coverage by ice sheets (Shi et al., 1998; Zheng et al., 2002; Owen, 2009). Further evidence...
supporting the reduction of population size during warmer periods comes from the current and MH SDMs which show restricted predicted distributions in the eastern Himalayas (Fig. 7). Nevertheless, the possible recent reduction of population size in the eastern Himalayas detected by SDM is not supported by our genomic data (Figs 6, 7). This period represents a small timescale (i.e. 18 000–25 000 yr), however, and a small density of population sampling (i.e. two populations) of group E may not provide enough information for such a recent reduction.

Finally, the western populations (group W) occur at the highest average altitude (4552 m; Table S1) and expanded during the LIG period before retreating to a southwestern refugium during the LGM (Figs 6, 7). The warm climate during the LIG period may have, in contrast to the eastern populations, facilitated expansion of this group through the opening of new potential habitats. The expansion to high-altitude areas in the western Himalayas during warmer periods is also supported by the comparison of the SDMs between the present and the MH, where more areas were predicted at present than in the MH (temperature is higher at present than the MH; Table S9). During the LGM, this area may have become too cold for this species to persist at such high altitudes as shown in the SDMs (Fig. 7). The two marginal populations that have colonized opposite geographical directions corresponding to very different altitudinal ranges are more vulnerable and respond differently to past climatic changes. Knowing the possible effect of past climatic changes on current populations may thus provide new insights into their future range dynamics in facing ongoing climatic changes and be useful for future management strategies (e.g. Lanier et al., 2015).

Conclusion

We combined genomic information and SDMs to identify the processes driving the phylogeographic structure of a high-altitude plant species over a large area of the Himalayas. Our analyses demonstrate the effects of past climatic changes on the intraspecific divergence of *P. tibetica* and highlight new patterns that are important to understand the current distributions of plant species in the Himalayas. The combination of population genomics and SDMs also provides new insights to predict the impact of future climatic changes on population dynamics. Taken together, we suggest that the central Himalayas was an ancient glacial refugium throughout the Quaternary glaciations in the area. The remaining lineages have persisted in additional refugia with different responses to climatic cooling during the LGM. Our study, taken together with those recently reported for other cold-adapted species that occur in the QTP (e.g. Shimono et al., 2010; Li et al., 2013; Liu et al., 2013), makes it clear that such species have exhibited different range dynamics (i.e. population persistence in high-altitude areas or even expansion) during the last glaciation relative to species associated with warmer environments.

Acknowledgements

We would like to thank Victor Rossier, Pawel Rosikiewicz and Dr Iakov Davydov for their great help in generating and analyzing data. The paper benefited greatly from the comments received from three anonymous reviewers. This work was funded by the University of Lausanne research fund and a grant 31003A_138282 from the Swiss National Science Foundation to N.S., and the China Scholarship Council (award to G.R. for 4 years’ PhD study abroad at the University of Lausanne). We received support for computational work from the Vital-IT facilities from the Swiss Institute of Bioinformatics. J.L. was funded by the National Natural Science Foundation of China (grant no. 3159820011), the Ministry of Science and Technology of the People’s Republic of China (no. 2010DFA34610) and International Collaboration 111 Projects of China. R.G.M. was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (ACONITE, PIEF-GA-2013-626220). N.A. was funded by the Swiss National Science Foundation (grant no. PP00P3_144870).

Author contributions

G.R., N.S. and E.C. planned and designed the research. G.R. carried out the sampling and the laboratory work, performed the molecular analysis. T.S. and N.A. participated in the initial test of the laboratory work. R.G.M. performed the SDM analysis. G.R. and N.S. wrote the manuscript with the help of R.G.M., J.L., T.S., N.A., A.G. and E.C.

References


Li JJ, Shi YF, Li BY. 1995. Uplift of the Qinghai-Xizang (Tibet) plateau and global change. Lanzhou, China: Lanzhou University Press.


Qiu YX, Fu CX, Comes HP. 2011. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and
environmental change in the world’s most diverse temperate flora. Molecular Phylogenetics and Evolution 59: 225–244.


Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 The dispersion in environmental space of the position of 16 sampled populations used in our genomic analyses relative to the PC values for 61 Primula tibetica sampling locations used in the SDMs.

Fig. S2 Distribution of individuals along PC scores of genetic variation based on the analysis of SNP dataset generated by two other parameter settings (M = 3 and M = 5) in STACKS.

Fig. S3 Maximum-likelihood estimation of population relationships based on an analysis of the 2092 concatenated, unlinked SNPs: nodal support values were assessed by 1000 bootstrap replicates.

Fig. S4 Alternative demographic scenarios of P. tibetica analyzed by DIY-ABC and posterior probabilities of the three scenarios obtained by logistic regression of 1% of the closest simulated datasets.

Fig. S5 Alternative demographic scenarios of groups W and CW analyzed by DIY-ABC.

Fig. S6 Tests for the genetic structure based on the two thresholds of the four groups for AStc modeling of changes in population sizes.

Fig. S7 Schematic representation of AStc modeling of changes in population size and posterior probabilities obtained by logistic regression of 1% of the closest simulated datasets for the groups W, C, CE and E.

Fig. S8 Habitat suitability predicted by the SDMs for present, MH and LGM using four techniques.
**Fig. S9** $\Delta K$-values identified using **STRUCTURE HARVESTER**.

**Fig. S10** Species distribution models for *Primula tibetica* at present, LIG, MH and LGM climatic conditions using three techniques.

**Fig. S11** Estimates of the prior and posterior distributions of parameters revealed by the DIY-ABC modeling of the best scenario (scenario 2) for the demographic history of *Primula tibetica*.

**Table S1** Locations of 16 populations of *Primula tibetica*

**Table S2** Descriptions of prior settings for all parameters used in DIY-ABC

**Table S3** Summary of genomic data collected for each population

**Table S4** Population pairwise $F_{ST}$ values

**Table S5** Posterior probabilities of modeled scenarios obtained by logistic regression of 1% of the closest simulated datasets for ALL and groups E, CE, C and W

**Table S6** Estimations of posterior distributions of parameters revealed by DIY-ABC for the best scenarios of changes in population sizes of groups E, C, CE and W, respectively

**Table S7** Estimations of posterior distributions of parameters revealed by DIY-ABC for the best scenario of demographic history of *Primula tibetica*

**Table S8** The differences of maximum, minimum and mean values of annual precipitation (bio12) between the current and past climatic conditions under each model for each area

**Table S9** The differences of maximum, minimum and mean values of annual temperature (bio1) between the current and past climatic conditions under each model for each area

**Methods S1** Species distribution models of *Primula tibetica*.

**Notes S1** ABC modeling based on five genetic groups.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.