


Species divergence and maintenance of species cohesion of three closely related *Primula* species in the Qinghai–Tibet Plateau

Guangpeng Ren^{1,2,3}  | Rubén G. Mateo^{4,5} | Antoine Guisan^{4,6} | Elena Conti⁷ | Nicolas Salamin^{2,3}

¹State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou, China

²Department of Computational Biology, Biophore, University of Lausanne, Lausanne, Switzerland

³Swiss Institute of Bioinformatics, Lausanne, Switzerland

⁴Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Switzerland

⁵ETSI de Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Madrid, Spain

⁶Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, Lausanne, Switzerland

⁷Department of Systematic and Evolutionary Botany and Botanic Garden, University of Zurich, Zurich, Switzerland

Correspondence

Nicolas Salamin, Department of Computational Biology, Biophore, University of Lausanne, Lausanne, Switzerland.
Email: nicolas.salamin@unil.ch

Funding information

University of Lausanne research fund; Swiss National Science Foundation, Grant/Award Number: 31003A_138282; China Scholarship Council

Editor: Richard Ree

Abstract

Aim: Understanding the relative roles of geography and ecology in driving speciation, population divergence and maintenance of species cohesion is of great interest to molecular ecology. Closely related species that are parapatrically distributed in mountainous areas provide an ideal model to evaluate these key issues, especially when genomic data are analyzed within a spatially and ecologically explicit context. Here, we used three closely related species of *Primula* that occur in the Himalayas, the Hengduan Mountains and north-east Qinghai–Tibet Plateau (QTP) to examine spatial and ecological effects on interspecific divergence and maintenance of species cohesion.

Location: Himalayas, the Hengduan Mountains and north-east QTP.

Methods: We used genomic data for 770 samples of the three species using double-digest restriction site-associated DNA (ddRAD) sequencing and combined approximate Bayesian computation (ABC) modeling, generalized linear mixed modeling (GLMM) and niche-based species distribution modeling (SDM).

Results: The three species are clearly delimited by the RADseq data. Further ABC modeling indicates that *Primula tibetica* diverged first followed by a later divergence between *Primula nutans* and *Primula fasciculata*. The time frames of the divergences among the three species coincide with the uplifts of the Hengduan Mountains and the northern QTP during the late Miocene and Pliocene followed by a long period of founder events. SDMs indicate that the three species might have survived in different refugia during glaciations and came into secondary contact during the post-glacial expansions but with no significant introgression. Finally, GLMM suggests that both the geographical and ecological factors play roles in population differentiation in *P. fasciculata* and *P. tibetica*, while for *P. nutans*, geography is the major driver of genomic variation. The different roles played by geographical and ecological factors in the three species may have affected the maintenance of species cohesion.

Main conclusion: Our results provide insights of unprecedented details into the start and maintenance of interspecific divergence in the context of changing environments in mountains. Our findings highlight the significance of combining population genomics



with environmental data when evaluating the effects of geography and ecology on interspecific divergence and maintenance of closely related species.

KEYWORDS

climatic changes, closely related species, interspecific divergence, maintenance of species cohesion, population genomics, Qinghai–Tibet Plateau

1 | INTRODUCTION

Understanding the relative roles of geography and ecology in speciation, population divergence, and maintenance of species cohesion is a long-standing goal in molecular ecology and evolutionary biology (Coyne & Orr, 2004; Nosil, 2012). Historically, geographical isolation, leading to reduced gene flow between isolated populations, has been considered a prerequisite for reproductive isolation (i.e., allopatric speciation; Mayr, 1963; Rice & Hostert, 1993). In a mode of speciation that is driven by geography, the range of an ancestral species is divided by a geographical barrier to form two new species (Coyne & Orr, 2004). It has, however, also been proposed that adaptation to different environmental conditions, resulting in separated populations evolving toward different ecological niches, no matter whether geographical isolation is present or not, can also lead to divergence and speciation (Berlocher & Feder, 2002; Nosil, 2008; Rundle & Nosil, 2005). Ecological factors play important roles in keeping new derived species separated. The interplay between these different processes may become very complex in mountainous areas due to their rugged topographic features and the profound ecological heterogeneity created by historical orogenesis and associated climatic changes (Favre et al., 2015; Hoorn et al., 2010). Furthermore, the mechanisms associated with these historical events do not have similar effects on population divergence, speciation and species maintenance across organisms (e.g., Opgenoorth et al., 2010; Ren et al., 2017; Sun et al., 2014; Zhang, Chiang, George, Liu, & Abbott, 2005). Despite their importance in driving the high biodiversity in mountainous areas (Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000), the specific roles played by these mechanisms are still unknown for most species in mountainous areas.

Comparing closely related species at the population level can offer insights into the relative importance of geographical versus ecological segregation (e.g., Abbott et al., 2000; Anacker & Strauss, 2014; Jia et al., 2012), helping to clarify the mechanisms of speciation. The current degrees of range overlap between closely related species can result from continuous contact since divergence or secondary contacts. The latter may obscure the mechanisms responsible for the initial divergence as, for example, in the case of allopatric speciation followed by secondary contact, which may be mistaken for ecological speciation (Barraclough & Vogler, 2000; Fitzpatrick & Turelli, 2006). Furthermore, the evolutionary outcomes can be highly variable when previously differentiated species come into contact. These outcomes can include the complete breakdown of reproductive barriers followed by the extinction of one or more taxa, the reinforcement of reproductive

isolation between related species, the introduction of adaptive variation to species through introgression and occasionally hybrid speciation (Abbott et al., 2013; Coyne & Orr, 2004).

To understand the mechanisms that facilitate these processes, we however need to resolve the phylogenetic relationships between closely related species. This task is usually difficult to achieve with traditional genetic markers (e.g., *rbcl*, *matK*, ITS), especially for genera that harbor high species richness and experience hybridization during their evolution, such as *Primula* L. (Guggisberg, Mansion, & Conti, 2009; Schmidt-Lebuhn, de Vos, Keller, & Conti, 2012) and *Gentiana* L. (Liu, Yan, & Ge, 2016). This problem may be overcome by using many thousands of DNA markers (e.g., Pante et al., 2015; Wagner et al., 2013), for example, using recently developed next-generation sequencing methods such as restriction site associated DNA (RAD) sequencing (Baird et al., 2008). Additionally, population genomics allows for discerning genomic regions that diverge neutrally from those that respond to divergent selection across heterogeneous landscapes (e.g., Lexer et al., 2014), which could provide a more accurate picture of the drivers of divergence compared with traditional neutral marker studies (Nosil, 2012). Here, based on population genomic data, we investigate the relative effects of geographic and ecological factors on interspecific divergence and the maintenance of species cohesion in closely related species. We test using evolutionary modeling the roles played by orogeny and climatic variations on the evolution and distribution of three species of the genus *Primula* (Primulaceae) that is widely and parapatrically distributed in the mountains surrounding the Qinghai–Tibet Plateau (QTP).

The QTP region has experienced drastic habitat changes and harbors extremely rich species diversity and endemism (Wu, 1987), especially the Himalayas and the Hengduan Mountains, which represent two key biodiversity hotspots on Earth (Myers et al., 2000). The historical orogenesis and the associated climatic changes are likely to account for the establishment of high species richness in the region (Wu, 1987). Although the uplifts of the QTP can be dated back as early as 50 Ma, the times and extent of its following uplifts are controversial (reviewed in Renner, 2016). Some evidence have shown that the central part of the QTP (i.e., Tibetan Plateau) has reached 4,000 m since the mid-Eocene (40 Ma; Renner, 2016 and references therein), while others suggest that a more recent uplift has occurred during the late Miocene and Pliocene (2.4–8 Ma) at its eastern and northern edge such as the Hengduan Mountains and northern QTP (Li & Fang, 1999; Mulch & Chamberlain, 2006; Zheng, Powell, An, Zhou, & Dong, 2000). Many previous evolutionary studies that focused on the Hengduan Mountains have indicated that extensive species diversification took



place in the region during the Pliocene (Li et al., 2012; Liu, Gao, Chen, & Lu, 2002; Liu, Wang, Wang, Hideaki, & Abbott, 2006; Wang et al., 2010; Xing & Ree, 2017). However, the effects that the different stages of uplifts had on the divergence pattern in the QTP are still unclear. The three species, *Primula nutans* Georgi, *Primula fasciculata* Balf. f. & Kingdon-Ward, and *Primula tibetica* Watt (section *Armerina*; Richards, 2003), studied here, occurring in different areas of the QTP, that is, the Himalayas, the Hengduan Mountains and the north-east QTP, respectively (Figure 1a), represent a unique opportunity to evaluate, for the first time using population genomic data, whether their initially interspecific divergence may have been triggered by the uplifts of these different mountainous areas. Previous phylogenetic and biogeographic analyses based on several plastid markers indicated a monophyletic clade formed only by the three *Primula* species that might have originated in the Himalayas during or after the Pliocene (De Vos, Hughes, Schneeweiss, Moore, & Conti, 2014; Ren, Conti, & Salamin, 2015). However, their interspecific relationships are still unresolved based on these markers (Ren et al., 2015). Furthermore,

the roles played by historical geological events on their initial inter-specific divergence and the factors that are influencing the current distributions of the species and their maintenance remain unknown.

The three species represent prominent floristic elements of alpine meadows at high altitudes (Hu & Kelso, 1996) in the QTP. All the three species are insect-pollinated, heterostylous, herbaceous, perennial plants and usually grow in wet meadows or along hill streams, but in different areas of the QTP (Hu & Kelso, 1996). *Primula tibetica* and *P. fasciculata* both occur in high altitude between 2,900 and 5,000 m in the Himalayas and the Hengduan Mountains, respectively, these two mountain regions representing two key biodiversity hotspots in the QTP, while *P. nutans* is distributed mainly below 3,800 m in the north-east QTP (Figure 1a). In contrast to the other two species endemic to the QTP, *P. nutans* can also be found in NW China, Central Asia, N Mongolia, N Europe, W&E Siberia, and NW North America (Richards, 2003). Furthermore, there is currently an overlap in the geographical ranges between *P. tibetica* and *P. fasciculata*, and between *P. nutans* and *P. fasciculata*, respectively (Figure 1a). They can occur in sympatry,

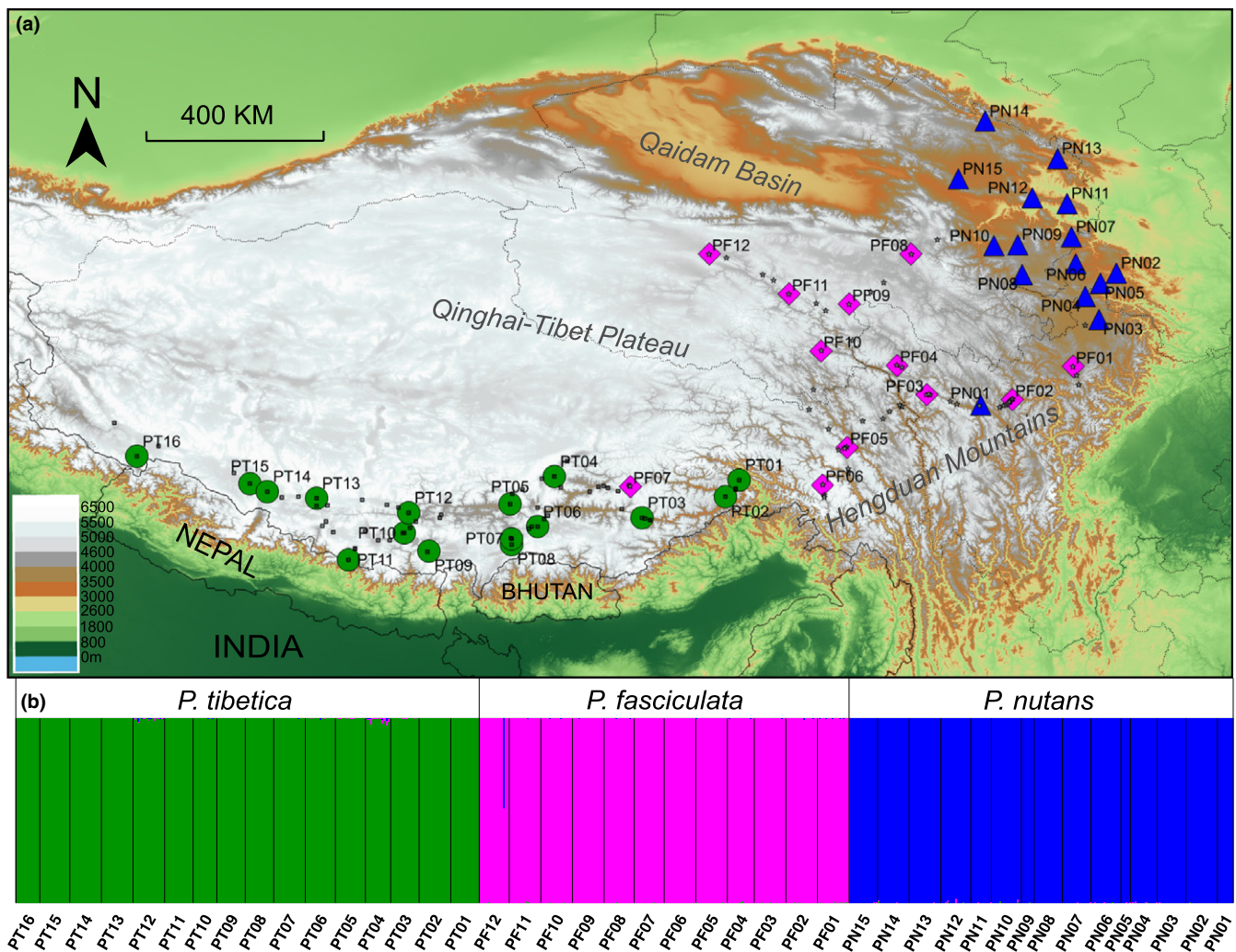


FIGURE 1 (a) Geographic locations of the 43 populations (colour coded) analyzed in the present study. Small grey squares and stars represent sampled populations of *Primula tibetica* and *Primula fasciculata* from fieldwork, respectively. (b) Plots of posterior probabilities for individuals of the three species assigned to K genetic clusters from STRUCTURE analyses for $K = 3$ [Colour figure can be viewed at wileyonlinelibrary.com]

in which case it has been observed that they can flower synchronously and share pollinators, mainly bees (Hu & Kelso, 1996; field observation). The distribution pattern coupled with the use of population genomic data provides a unique opportunity to test the relative roles of geography and ecology in population/species divergence in this region. Specifically, the aims of our study are to: (a) test the interspecific divergence among the three species based on evolutionary modeling (i.e., ABC method) and assess whether the divergences could be congruent with the uplifts of the Hengduan Mountains and northern QTP; (b) combine generalized linear mixed modeling (GLMM) and species distribution modeling (SDM) to better understand the role of geography and ecology in the maintenance of species cohesion.

2 | MATERIALS AND METHODS

2.1 | Sampling, RAD library preparation, sequencing and processing of Illumina data

We selected a total of 43 populations from three closely related species, *P. tibetica*, *P. fasciculata* and *P. nutans*, which were collected from the QTP (Figure 1; Supporting Information Table S1). For *P. tibetica*, we used the same 16 populations sampled in a previous study (Ren et al., 2017). Fifteen populations of *P. nutans* and 12 out of 61 populations of *P. fasciculata*, that is, same strategy as applied for *P. tibetica* to select populations that were representative of both the geographical distribution and the diversity of ecological niches (Ren et al., 2017) were used. In this study, 6 to 20 individuals were chosen from each population, which gave us a total of 770 individuals who were processed with a double-digestion restriction site-associated DNA sequencing (RADseq) following the same protocol as Mastretta-Yanes et al. (2015), which was already used for *P. tibetica* (Ren et al., 2017). The DNA was double-digested with *EcoRI* and *MseI* restriction enzymes. The libraries were sequenced using single-end reads of 100 bp of length.

Single-end Illumina reads were processed into RAD-tags, using the STACK-1.30 software pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Catchen et al., 2011). All reads were trimmed to 90 bp in length. We used all 770 samples to build a catalog and matched each sample against the catalogue to identify alleles. The execution of the *de novo* assembly was accomplished using the *denovo_map.pl* script. Different combination of parameter settings for this script gave similar results as tested in Ren et al. (2017), we therefore only considered the following settings for assembly: minimum number of reads to create a stack (m) = 3; maximum distance allowed between stacks (M) = 3; maximum number of mismatches allowed between loci (n) = 3; $-t$ flag to remove or break up highly repetitive RAD-tags during the *ustacks* component and upper bound of error rate (ϵ) = 0.1. We used *rxstacks* to further filter the data to increase quality, correct 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 -ln_lim -10.0 -ln_dist`. After *rxstacks*, *cstacks*, and *sstacks* were run again with the same setting as before to rebuild the catalog of reads.

We used the same settings as in Ren et al. (2017; $m = 3$, $r = 0.5$, $min_maf = 0.01$, $max_obs_het = 0.5$) to filter the catalog of reads, using the *populations* module to generate three datasets (i.e., one for each species; D1–D3) and one dataset considering all the three species together (D4) for downstream population genetic analyses. We retained polymorphic RAD loci that were only present in all populations for each dataset and scored for each RAD locus only the first SNP if several were present. Pairwise F_{ST} values for different datasets and different genomic fractions (see below) were calculated among populations in GENODIVE 2.0b27 (Meirmans & Van Tienderen, 2004), and significance was determined using 1×10^4 permutations.

2.1.1 | Interspecific divergence

To investigate interspecific divergence and the extent of hybridization/introgression among the three species, we first identified population genetic structure, using a Bayesian method implemented in STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) based on dataset D4. STRUCTURE analysis was performed under the “Admixture model” and the “Correlated allele frequency model” with K -values ranging from 1 to 10. Ten independent runs were performed for each value of K using 5×10^4 generations for the burnin and 2.5×10^5 generations for the sampling. The optimal K was chosen using the delta- K method of Evanno, Regnaut, and Goudet (2005) as implemented in STRUCTURE HARVESTER (Earl & VonHoldt, 2012). The coefficient for cluster membership of each individual was averaged across the ten independent runs using CLUMPP (Jakobsson & Rosenberg, 2007) and plotted using DISTRUCT (Rosenberg, 2004). Secondly, a maximum-likelihood phylogenetic tree was estimated from unlinked SNPs with a GTR + G model, using PHYML 3.0 (Guindon et al., 2010) based on the same dataset. This algorithm is able to estimate phylogenetic trees using standard IUPAC coding for polymorphic nucleotide sites, allowing us to include all unlinked RAD loci within our concatenated dataset. A PHYLIP file was generated by the *populations* module in STACKS using the `-phylip_var` setting. Nodal support of phylogenetic tree was estimated from 1,000 bootstrap replicates.

2.1.2 | Modeling of historical divergence

To further understand the historical divergence of the three species and assess the congruence of the divergence times with the uplifts of the Hengduan Mountain and northern QTP, we used approximate Bayesian computation (ABC) method implemented in DIY-ABC 2.1.0 (Cornuet, Ravigné, & Estoup, 2010; Cornuet et al., 2014). The ABC method, based on coalescent theory, does not need outgroup data and mutation model parameterization for SNP dataset because SNPs are polymorphic and present only two allelic states (Cornuet et al., 2014), which made this method more preferred than SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012) in this study. We subsampled two individuals from each of the 43 populations ten times and pooled individuals into “species” groups based on the results of STRUCTURE and PHYML to generate unbiased estimates of species history and to reduce computational time. We selected for



each subdataset a single SNP per locus and the SNPs had to be present in at least 80% of the individuals from each species and in all the three species. The ten subdatasets were listed in Supporting Information Table S2. We tested seven possible scenarios for the divergence relationships between species: *P. nutans* diverged from the common ancestor first (SC1-SC2), *P. tibetica* diverged from the common ancestor first (SC3-SC4), *P. fasciculata* from the common ancestor first (SC5-SC6) and all three species diverged simultaneously from their ancestor (SC7, Supporting Information Figure S1).

For each subdataset, we gave each scenario a uniform prior probability (Supporting Information Table S3) and selected all summary statistics to generate a reference table containing 7×10^6 simulated datasets (on average 10^6 per scenario). The simulated SNP dataset was generated following the algorithm proposed by Hudson (2002). The parameters defining each scenario (i.e., population sizes, divergence times and times of population size changes) were considered as random variables drawn from prior distributions. For each simulation, DIY-ABC drew a value for each parameter from its prior distribution and performed coalescent simulations to generate a simulated pseudo-observed dataset (POD) with same number of gene copies and loci per population as observed. It then calculated summary statistic for each POD and the observed data. Based on a distance and a tolerance, it decided for each POD whether its summary statistic was sufficiently close to that of the observed data. We used 1% of the simulated datasets closest to the observed data to estimate the relative posterior probabilities for each scenario via a polychotomous weighted logistic regression. Based on the most likely scenario, local linear regression on the 1% closest simulated datasets with a logit transformation was used to estimate the posterior distribution of historical demographic parameters (Cornuet et al., 2010). The time parameters are estimated in generations and converted into years by multiplying generation time, which was set to 1 year for the three species (Ren et al., 2017).

The next step was to evaluate the level of confidence for choosing the best-supported scenario. We evaluated the ability of the ABC analysis to discriminate between the chosen scenarios by analyzing 1,000 simulated datasets with the same number of loci and individuals as each of our subdatasets. We estimated the probabilities of type I and type II error and computed bias and precision for our estimations of divergence times as described in Cornuet et al. (2010). The PODs were produced from the posterior distributions of each parameter, which provide a more relevant estimation of accuracy of parameter estimation in the vicinity of the observed dataset than blindly computing accuracy indicator over the whole prior space (Cornuet et al., 2010). The number of simulated data closest to the observed data was set to 5,000 to have a precise estimation of the accuracy measures. Finally, we performed an evaluation of the fit of each scenario to our datasets by running a model-checking analysis following again Cornuet et al. (2010).

2.1.3 | Species distribution models

The overlap of geographical ranges between the three closely related species was observed during our fieldwork. The current overlap and

potential historical contacts may influence the maintenance of species cohesion if reproductive isolation among them was not complete. To evaluate the potential overlap of distributions of the three species due to past climatic oscillations, an ensemble of species distribution models was generated for *P. nutans* and *P. fasciculata* for the present and the last glacial maximum (LGM) following the same methodology applied for *P. tibetica* (Ren et al., 2017), using three different techniques: generalized linear model, gradient boosting machine and random forests, as implemented in the R package 'biomod2' (Thuiller, Lafourcade, Engler, & Araújo, 2009). A total of 67 and 89 species occurrences obtained from the field collections and herbarium records were used as presence data to calibrate the models for *P. nutans* and *P. fasciculata*, respectively. To avoid potential sampling bias (Syfert, Smith, & Coomes, 2013), we retained only the points that were separated by at least 2.5 arc-minutes from each other (i.e., matching the resolution of the climatic data). These occurrences represented almost all the relevant regions for the two species in the QTP based on the distribution described in Flora of China (Hu & Kelso, 1996) as a reference. We used the WorldClim database (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) for the environmental variables and, to avoid multicollinearity, we used the same ones as for *P. tibetica* (see Ren et al., 2017). The variables included in the models were as follows: isothermality (Bio3), maximum temperature of warmest month (Bio5), minimum temperature of coldest month (Bio6), annual temperature range (Bio7), precipitation seasonality (Bio15), precipitation of the wettest quarter (Bio16), and precipitation of the driest quarter (Bio17). We run ten replicates per method, where each replicate was calibrated on 70% of the data and evaluated on the remaining 30% using the area under the curve (AUC) of a receiver-operating characteristics plot (Swets, 1988) and the true skill statistics (Allouche, Tsoar, & Kadmon, 2006). The averaged and pondered consensus model (the contribution of each replicate was weighted proportionally to their AUC values) was then projected onto LGM with three different general circulation models (GCMs): CCSM4, MIROC-ESM, MPI-ESM-P available from <http://cmip-pcmdi.llnl.gov/cmip5/> processed on WorldClim database. The consensus model was converted into a binary model (presence/absence) by applying a threshold that allow a maximum of 5% of omission error (i.e., omission error is the percentage of the real presence predicted as absences in the model; Fielding & Bell, 1997).

2.1.4 | Outlier detection and GLMM

Geographical and ecological factors may have different effects on the divergences of genomic regions that are under different selective pressure. For example, the divergences of neutral genetic markers are often concordant with a model of isolation by distance, where geographical distance plays the major role (Wright, 1943). Adaptation to ecological niches may cause genomic regions in different populations to become more divergent (positive selection) or similar (negative selection) depending on the direction of selection. To investigate the effects of geographical and ecological factors on driving population divergence and maintaining species cohesion among the three closely related



species, we first performed genomic scans to identify polymorphic loci that could be potentially under balancing (negative) and divergent (positive) selection in each species (datasets D1–D3). We used both *BAYESCAN* 2.1 (Foll & Gaggiotti, 2008) and *LOSITAN* (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008) to compare the results obtained with these two distinct methods. *LOSITAN* (based on the *fdist* F_{ST} outlier methods; Beaumont & Nichols, 1996) uses the island model as a null distribution of F_{ST} , while *BAYESCAN* assumes that populations have diverged independently from a common ancestor. The two methods were shown to have low rates of false positive and false negative errors as suggested using simulations (Narum & Hess, 2011). *BAYESCAN* estimates population-specific F_{ST} coefficients by the Bayesian method described in Beaumont (Beaumont & Balding, 2004) and uses a cut-off based on the mode of the posterior distribution to detect SNPs under selection (Foll & Gaggiotti, 2008). We used a prior odds value of 10, with 1×10^5 iterations and a burn-in of 5×10^4 iterations. We identified loci that were significant outliers at a false discovery rate (FDR) of 0.05. For *LOSITAN*, 100,000 simulations were run on the three datasets separately, with default parameters. Both the “neutral mean F_{ST} ” and “force mean F_{ST} ” options were used. We used a more conservative threshold to identify outliers in *LOSITAN*, which generally identifies greater number of outliers than *BAYESCAN* (e.g., Huang et al., 2017). Loci outside the 99% confidence interval and those at FDR of 0.01 were considered outliers and we only kept the loci that were identified by both methods as outliers. Loci that were identified as balancing and divergent outliers by both methods were further segregated into a negative and a positive outlier dataset, respectively, while the remaining loci (with overlapped outliers removed) comprised the neutral dataset.

The different genomic fractions (i.e., neutral RAD loci and loci under negative and positive selection, respectively) identified by outlier detection for each species were further applied in a GLMM approach to test whether population divergence of these different genomic fractions was driven by geographical or ecological or both factors (e.g., Lexer et al., 2014). The modeling was run in the R package ‘*MCMC_{GLMM}*’ (Hadfield, 2010). Genetic divergence F_{ST} metrics for multilocus nuclear RADseq datasets were used as response variables for each species. We used geospheric distances between populations (“GEO” from here onwards) and pairwise differences of altitude between populations (“ALT” from here onwards) as predictor variables to assess the spatial effects. For environmental predictors, we used two different kinds of variables: (a) 19 WorldClim variables (Hijmans et al., 2005); “CLI” for climatic data from here onwards) and (b) three edaphic variables (Soil-carbon, moisture, PH; “SOL” for edaphic variables from here onwards) obtained from the Center for Sustainability and the Global Environment (<http://nelson.wisc.edu/sage/data-and-models/atlas/maps.php>). The Euclidean distances of these variables between populations for each of the three species were calculated in the R package ‘*ade4*’ (function ‘*dist.quant*’).

We ran 16 different models for each species that resulted from combinations of the four predictor variables: a null model without any predictor; four models with a single predictor variable (GEO, CLI, SOL or ALT); six models with different combinations of two predictors; four models with different combinations of three predictors

and one with all the four predictors. The input files and R script for each species can be found in Data Accessibility. The deviance information criterion (DIC) and associated DIC differences and weights were used to compare all models for each genomic fraction and draw conclusions on the relative roles of different drivers of divergence for each of them. *MCMC_{GLMM}* was initiated with standard priors and run with a burn-in of 2×10^6 followed by 1×10^7 iterations with a thinning interval of 1×10^3 . Chain convergence was confirmed by inspecting trace plots using the R package ‘*CODA*’.

3 | RESULTS

3.1 | Sequence data quality and processing

The average number of sequence reads among the 770 samples of the three species was 2.16 million (2.11–2.23 million) and the average number of reads per sample that were used in the assembly of the RAD-tags was 1.58 million (1.53–1.64 million; Supporting Information Table S4). Datasets D1–D3, containing 2822, 6086, and 12925 single-SNP loci, were used for the outlier detection in *P. tibetica*, *P. fasciculata*, and *P. nutans*, respectively. Dataset D4 including all samples of the three species used for structure analysis and building a phylogenetic tree contained 748 single-SNP loci.

3.2 | Interspecific divergence

Both the *STRUCTURE* and phylogenetic tree suggested a clear species delimitation of the three species (Figure 1b, Supporting Information Figure S2). At $K = 3$ (the best K value chosen by *STRUCTURE HARVESTER*), we detected only one individual of *P. fasciculata* that was significantly introgressed by *P. nutans*. Little gene flow among species was detected.

3.3 | Estimates of historical divergence

Our ABC modeling of the historical divergence of the three species indicated that the scenario involving an initial divergence of *P. tibetica* followed by a later divergence of *P. nutans* and *P. fasciculata* (SC3; Supporting Information Figure S1) provided the best fit to our RADseq data in all of the ten datasets tested (Supporting Information Table S5). The evaluation of the type I error rate (Supporting Information Table S5) showed that 56.4%–70.4% of the datasets simulated with SC3 were correctly identified as being produced by SC3 across the ten subdatasets. A high rate of false negative errors was only observed with SC4 (18.2%–24.1%). Both scenarios (SC3 and SC4) suggested *P. tibetica* diverged first. Estimation of the type II error (i.e., false positive) was also low (Supporting Information Table S5). The only scenario producing significant type II error rate was again SC4, with 20.5%–24.6% of PODs wrongly selected as being generated by SC3 across the 10 subdatasets. Overall, excluding SC4, our analyses displayed a strong power (more than 80%) to discriminate among the tested scenarios. A check of the goodness-of-fit of the distributions of the parameters for the scenarios with

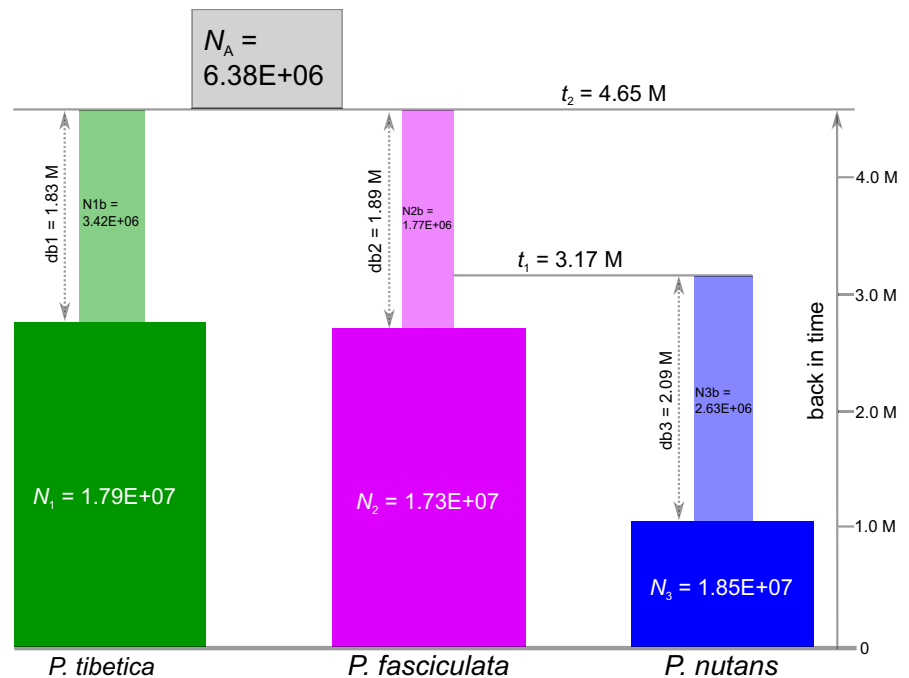


FIGURE 2 Summary of inferred interspecific divergence of the three species. Effective population sizes (N_1 and $N1b$, *Primula tibetica*; N_2 and $N2b$, *Primula fasciculata*; N_3 and $N3b$, *Primula nutans*), times of divergence in years (t_1 , t_2) and durations of founder events (db1–db3) are indicated [Colour figure can be viewed at wileyonlinelibrary.com]

the real dataset further indicated that SC3 was the best-supported scenario (Supporting Information Table S5).

The parameter values were estimated for each dataset based on the best-fit scenario SC3. Results dealing with the estimation of bias and precision on time parameters under SC3 for the ten subdatasets were summarized in Supporting Information Figure S3. The genomic data provided substantial information for the estimation of divergence times. Here we showed only the averaged values based on the ten subdatasets (Supporting Information Table S6). Our ABC modeling suggested that *P. tibetica* diverged first c. 4.65 Ma (95% highest posterior density [HPD]: 1.74–8.80 Ma; Figure 2), and *P. nutans* and *P. fasciculata* originated from a common ancestor c. 3.17 Ma (HPD: 1.58–5.21 Ma). After the divergences, all three species experienced a long period of founder events, and started to expand their population sizes by a factor of five or 10 times at the beginning or middle Quaternary (Supporting Information Table S6).

3.4 | Species distribution models

The results of SDMs showed that the three species occupy mostly different environments (Figure 3a,b). The predicted potential distribution at the current conditions was consistent with current records for each species, with some overlap between *P. tibetica* and *P. fasciculata* (880 pixels), but large overlap between *P. nutans* and *P. fasciculata* (5,520 pixels; Figure 3c). During the LGM, the predictions based on the three GCMs (CCSM4, MIROC and MPI) were mostly consistent, although large mismatches between the estimated distributions were observed for the three species. We only considered here the most likely predictions that were recovered based on all three GCMs (Figure 3). When comparing the predicted distributions during the present and LGM for each species, all three species have

experienced expansions from the LGM to present (Figure 3c). Two main glacial refugia were identified for *P. tibetica*, which were located in the central and southwestern Himalayas. *Primula fasciculata* was predicted to have retreated to eastern QTP after having occupied a much larger region during the LGM. Finally, the prediction for *P. nutans* during the LGM yielded restricted refugia distributed mainly in the north-east QTP and some valleys of the Hengduan Mountains, mostly nested in the main refugium of *P. fasciculata*.

3.5 | Outlier loci and modeling of genetic, spatial, and environmental data

A larger number of outliers were identified by LOSITAN than by BAYESCAN (Supporting Information Table S7), but we only considered the outliers detected by both methods to reduce false positive errors. Outlier detection by both methods identified 106 potentially non-neutral outlier RAD loci in *P. fasciculata*, 60 of which were negative outliers and 46 were positive outliers (Supporting Information Table S7). In *P. nutans*, 61 RAD loci were revealed as outliers, 24 of which were showing negative and 37 were positive. Similarly, 76 RAD loci in *P. tibetica* were identified as outliers, 43 of which were negative and 33 were positive. None of these loci was shared as outlier in all of the three species.

Generalized linear mixed models of genetic divergence for negative, neutral and positive RAD polymorphisms with GEO, CLI, SOI, and ALT as predictor variables revealed complex and different drivers of variation in the genomic data for the three species (Table 1, Supporting Information Table S8). For *P. nutans*, geographical distance (GEO) was the main driver for the variation in negative and neutral RAD datasets, with DIC weights equal to 0.197 and 0.352, respectively. The best predicted models for the divergence of positive outlier RAD loci were GEO+CLI+SOI (DIC weight 0.295), indicating that climatic and edaphic variables were involved in triggering

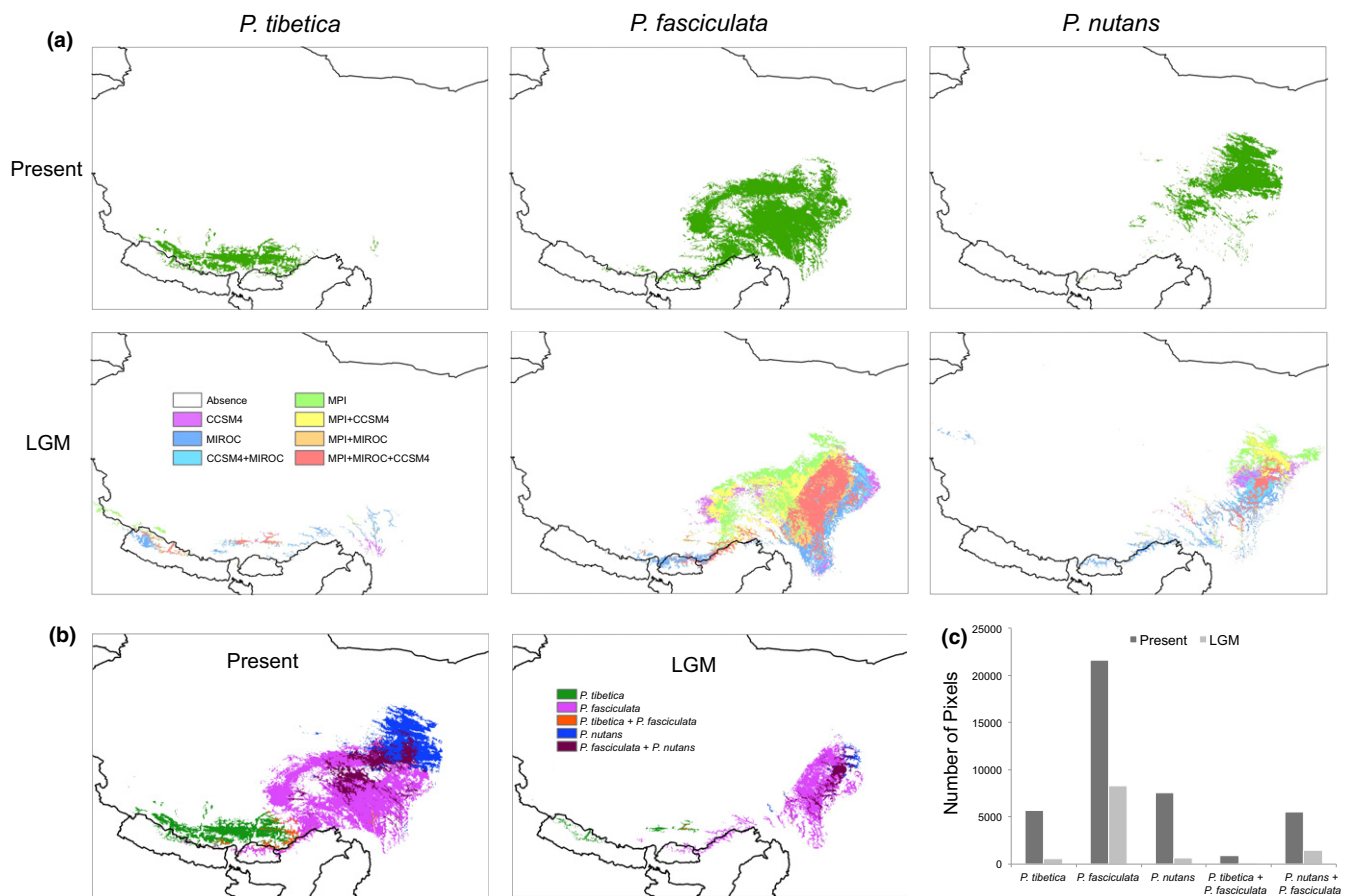


FIGURE 3 (a) Habitat suitability of the three species predicted by species distribution models (SDMs) for present and LGM. SDMs for the LGM are based on three different general circulation models (GCM). SDMs for *Primula tibetica* were modified from Ren et al. (2017). (b) The predicted and the overlapping distributions between the three species for the present and LGM. Only the distributions predicted by all the three GCMs during the LGM (red areas in (a)) were considered here. (c) Number of pixels counted based on (b) [Colour figure can be viewed at wileyonlinelibrary.com]

divergent selection between populations in *P. nutans*. For *P. fasciculata*, GEO+SOI+ALT (DIC weight 0.456) was the best model in driving the variation in negative RAD polymorphisms. The model including all of the four predictors (GEO+CLI+SOI+ALT) was best supported in driving the variation in neutral and positive RAD polymorphisms in this species. The highest DIC weight value (0.861) for the positive dataset indicated that both the spatial and environmental variables have played important roles in divergent selection between populations of this species. The model including all four predictors also best predicted the divergence of positive outlier RAD polymorphisms in *P. tibetica* (DIC weight 0.429). For negative and neutral datasets in *P. tibetica*, the best-supported model was GEO+CLI+ALT (DIC weight 0.444 and 0.553, respectively). Similarly, more variables were involved in driving divergent selection among populations in these species as expected.

4 | DISCUSSION

Our results based on RAD sequencing clearly distinguished the three *Primula* species, in contrast to a previous study based on few

chloroplast and nuclear genes that failed to delimit the relationships among them (Ren et al., 2015). The timeframes of the first divergence of *P. tibetica* from a common ancestor and a later divergence between *P. nutans* and *P. fasciculata* estimated by ABC modeling coincide with the extensive uplift of the Hengduan Mountains and the northern QTP, which occurred during the late Miocene and the Pliocene. The expansion of common ancestral populations due to open cold habitats created by these uplifts of mountains followed by geographic isolation may have driven the initial interspecific divergence. Our study therefore suggests that both geographical and environmental factors have been involved in the initial divergence of populations that led to the separation of the three species. These factors were also important for the maintenance of species cohesion despite potential secondary contact between them due to the Quaternary climatic oscillations.

4.1 | Interspecific divergence in response to the uplifts of mountains

Using a replicated subsampling strategy, all the ABC models support a scenario that *P. tibetica* diverged first from a common ancestor and that a divergence event between *P. nutans* and *P. fasciculata* in a later stage.



TABLE 1 Results of GLMM set up to predict genetic divergence (F_{ST}) between populations of the three species with GEO, CLI, SOI and ALT as predictor variables. Deviance information criterion (DIC), DIC difference to the best-supported model (delta DIC) and DIC weights for each model are detailed shown in Supporting Information Table S8. Here we showed only the values of DIC weights. For each model comparison, the best-supported model is shown in bold italics

Model	<i>Primula nutans</i>			<i>Primula fasciculata</i>			<i>Primula tibetica</i>		
	RAD/ negative	RAD/ neutral	RAD/ positive	RAD/ negative	RAD/ neutral	RAD/ positive	RAD/ negative	RAD/ neutral	RAD/ positive
NULL	0.132	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GEO	0.197	0.352	0.117	0.105	0.000	0.000	0.003	0.000	0.000
CLI	0.129	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOI	0.109	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ALT	0.052	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GEO+CLI	0.057	0.155	0.161	0.080	0.000	0.000	0.036	0.031	0.129
GEO+SOI	0.034	0.146	0.099	0.000	0.000	0.002	0.003	0.000	0.000
GEO+ALT	0.078	0.141	0.052	0.046	0.000	0.000	0.051	0.000	0.000
CLI+SOI	0.045	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
CLI+ALT	0.054	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOI+ALT	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GEO+CLI+SOI	0.011	0.058	0.295	0.207	0.279	0.100	0.040	0.027	0.118
GEO+CLI+ALT	0.022	0.065	0.080	0.031	0.000	0.000	0.444	0.553	0.335
GEO+SOI+ALT	0.014	0.058	0.043	0.456	0.000	0.003	0.039	0.000	0.000
CLI+SOI+ALT	0.019	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000
GEO+CLI+SOI+ALT	0.004	0.024	0.151	0.075	0.721	0.861	0.385	0.388	0.419

Primula tibetica diverged first c. 4.65 Ma (HPD: 1.74–8.80 Ma), whereas *P. nutans* and *P. fasciculata* originated from a common ancestor more recently c. 3.17 Ma (HPD: 1.58–5.21 Ma; Figure 2; Supporting Information Table S6). These divergence times were largely consistent with a period of recent uplift of the eastern and northern QTP during the late Miocene and Pliocene (Li, Shi, & Li, 1995; Mulch & Chamberlain, 2006). In fact, numerous studies of other herb, shrub, and conifer groups that grow in the eastern and northern QTP have demonstrated that intra- or interspecific divergences took place during the Pliocene (Li et al., 2013; Liu et al., 2002, 2013; Xing & Ree, 2017; Xu et al., 2010; Zhou et al., 2012). Furthermore, our sampling of the three species that covers the Himalayas, the Hengduan Mountains and north-east QTP, respectively, provided a unique opportunity to evaluate the effects of the uplift of a particular region surrounding the QTP on species divergence.

The time frames of the interspecific divergence among the three species are congruent with the extensive uplift of the Hengduan Mountains and northern QTP that occurred during the Late Miocene and the Pliocene (8–2.4 Ma; Li & Fang, 1999; Zheng et al., 2000). Given the occurrence of the three species in high altitudes of the QTP region at present and the close relationships among them, their common ancestor may have already adapted to high altitudes (i.e., cold niches), which is also congruent with a previous study on *P. tibetica* (Ren et al., 2017). The open cold habitats created during the uplift of the Hengduan Mountains and northern QTP would probably have allowed the expansion of the common ancestor to the Hengduan Mountains and northern QTP area. However, the occurrence of high mountains separated by deep valleys resulted from this extensive

uplift may have limited gene flow between these expanded populations and the ancestral population, leading to geographical divergence and the origin of *P. tibetica*, which is capable of surviving and reproducing currently at high altitudes in the Himalayas. Similarly, the rapid uplift of the Hengduan Mountains and northern QTP associated with topographic and climatic changes may have further triggered the divergence between *P. nutans* and *P. fasciculata*. Furthermore, the niche differentiation that can be observed in the current SDMs of these three species (Figure 3) would suggest that adaptation to their specific ecological niches have occurred. The long periods with founder events that were identified for all three species followed by a rapid expansion (Figure 2) could have allowed the establishment of different adaptive alleles in the populations (Gavrilets & Boake, 1998), which in turn may have further reinforced the initial interspecific divergence (Barton & Charlesworth, 1984; Templeton, 1980; Weinberg, Starczak, & Jorg, 1992).

It should be noted that the timeframes of divergence estimated with our RADseq data are much older than previous estimates based on few chloroplast markers (1.7–8.8 vs. 0.5–2 Ma; Supporting Information Table S6; Ren et al., 2015). Previous estimates were obtained by secondary calibration that is inherently subject to bias and errors (Sauquet et al., 2012) and should be treated with caution (Ren et al., 2015). The origin of *P. nutans* based on RADseq data in this study is estimated at c. 3.17 Ma (HPD: 1.58–5.21 Ma), which is in agreement with macrofossil evidence that suggests the present-day arctic flora developed c. 3–4 Ma at a time when global temperature decreased sharply (Matthews & Ovenden, 1990; Zachos, Pagani, Sloan, Thomas,



& Billups, 2001). However, this study only focused on the QTP and further studies involving a finer sampling across the entire distribution of *P. nutans* associated with large-scale genomic data should be employed to gain a detailed knowledge of evolutionary history of this species (e.g., Wang et al., 2016).

4.2 | Species maintenance in secondary contact zones

Climatic oscillations during the Quaternary had a dramatic effect on species distribution ranges (Comes & Kadereit, 1998; Hewitt, 2004). The postglacial expansion or retreat to the same refugium may have resulted in secondary contact of previously isolated species, which may cause introgression between species (e.g., Li et al., 2013), or even trigger hybrid speciation if reproductive isolation is incomplete (Abbott et al., 2013; Ma, Szmidt, & Wang, 2006; Rieseberg, 1997; Sun et al., 2014). However, our analyses based on population genomic data indicate no hybridization between the three *Primula* species (Figure 1) despite clear overlap in their geographic distribution and potential secondary contacts, especially between *P. fasciculata* and *P. nutans*, identified based on the niche modeling analysis (Figure 3b,c) or during our field collection (Figure 1). The lack of hybridization or introgression is unexpected between these species because hybridization in the genus *Primula* is common and has been described in multiple studies (Guggisberg et al., 2009; Ma, Tian, Zhang, Wu, & Sun, 2014; Zhu et al., 2009). Furthermore, *P. nutans* (section *Armerina*) can even hybridize with *P. mistassinica* (section *Aleuritia*), a more distantly related species, which resulted in an intersectional allopolyploidization event giving rise to the tetraploid species *P. egaliksensis* (Guggisberg et al., 2009). Hybridization was also a likely explanation for the incongruent relationships of the three species between chloroplast and nuclear trees, but the conclusion may be biased by the use of a single nuclear gene that provides low resolution to infer phylogenetic relationships (Ren et al., 2015). The lack of evidence for nuclear introgression in contact zones based on our population genomic data may suggest complete or nearly complete reproductive isolation between the species. However, the biological characteristics of the three species are not well described and further experimental and field studies are needed to investigate the degree of reproductive isolation among them.

Although there is no clear explanation for the lack of hybridization between the three species in the contact zone, the different drivers of variation in the genomic data observed for the three species (Table 1, Supporting Information Table S8) may provide some insights to explain the maintenance of species cohesion. Our GLMM analyses revealed that the drivers of population divergence in *P. tibetica* and *P. fasciculata* are complex and different from those of *P. nutans* (Table 1). The former two species occur mainly in the Himalayas and Hengduan Mountains, respectively. These regions display extreme elevational gradients within relatively short distances, which lead to profound ecological heterogeneity. Therefore, it is not surprising that both the spatial and environmental variables are involved in population divergence across genomic regions (Table 1). Muñoz-Pajares et al. (2017) also found that both spatial/environmental

variables and historical factors play important roles in shaping patterns of genetic differentiation in a montane herb at different spatial scales. The SDMs suggested that *P. tibetica* and *P. fasciculata* have adapted to their specific ecological niches (i.e., different predicted distributions; Figure 3), and such different ecological niches have influenced their population divergence as indicated by GLMM. The interplay between geographical distance and ecological factors could help the maintenance of species boundaries between them.

By contrast, geographic distance is the predominant mechanism explaining the patterns of divergence and gene flow in the neutral genomic fractions of DNA in *P. nutans*. Geographic distance also received some support for the fraction of SNPs under divergent selection in this species (Table 1). The genetic differentiation among populations is much lower in *P. nutans* than in the other two species (Supporting Information Table S7), indicating higher gene flow among populations in *P. nutans*. The occurrence of this species at lower altitude (average altitude 3,311 m compared with *P. fasciculata*—4,256 m and *P. tibetica*—4,093 m) where topography is less complex (Figure 1), and higher dispersal ability suggested by the much wider distribution occupied by *P. nutans* when compared to the two other species (Richards, 2003) may account for the high gene flow observed in this species. Interspecific gene flow between *P. nutans* and *P. fasciculata* may be restricted because the genomic regions in the latter species were probably linked to local adaptation as indicated by GLMM (Table 1). If hybridization has occurred between them, high gene flow among populations in *P. nutans* and local selection in *P. fasciculata* may have potentially diluted the introgressed alleles (Du, Petit, & Liu, 2009; Petit & Excoffier, 2009; Zhou et al., 2010).

5 | CONCLUSIONS AND PERSPECTIVES

We combined population genomics and SDMs to test the relative roles of geography and ecology in speciation, population divergence and the maintenance of species cohesion over a large area of the QTP. The QTP region represents a prime example for this type of spatial and environmental setting, and the studies of other biodiversity hotspots associated with mountains would be similarly fruitful. Our analyses indicated that the divergence times of three closely related species are congruent with the uplifts of the Hengduan Mountains and northern QTP, which provides evidence for an important role of geographic isolation during orogenesis in triggering the initial interspecific divergence of three cold-adapted species in mountains. When they came into secondary contacts due to past climatic changes, the combination of geographic and ecological segregation would have limited or prevented interspecific gene flow among contact populations, which facilitates the maintenance of their divergence. This nonhybridization pattern in contact zones between closely related species would represent an ideal model to further investigate the mechanisms involved in the development of reproductive barriers in general. Furthermore, knowing the relative roles played by geographic and ecological factors in diversification will help to better understand the processes during the accumulation of biodiversity in this globally unique region.

ACKNOWLEDGEMENTS

This work was funded by the University of Lausanne research fund, the grant 31003A_138282 from the Swiss National Science Foundation to N.S., the China Scholarship Council (award to GR 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. The genomic data was obtained through the Genomic Technology Facility of the University of Lausanne.

DATA ACCESSIBILITY

Raw RAD sequences: GenBank accession no. PRJNA470600. Final datasets used for outlier detection, population genetic analyses and ABC modeling; Input data and scripts for GLMM; Scripts for SDMs: Dryad doi: <https://doi.org/10.5061/dryad.tt8n46q>

ORCID

Guangpeng Ren  <http://orcid.org/0000-0002-4052-9115>

REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Abbott, R., Smith, L., Milne, R. I., Crawford, R. M. M., Wolff, K., & Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. *Science*, 289, 1343–1346. <https://doi.org/10.1126/science.289.5483.1343>
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Anacker, B. L., & Strauss, S. Y. (2014). The geography and ecology of plant speciation: Range overlap and niche divergence in sister species. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132980. <https://doi.org/10.1098/rspb.2013.2980>
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *Bioinformatics*, 9, 323.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3, e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Barracough, T. G., & Vogler, A. P. (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist*, 155, 419–434.
- Barton, N. H., & Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics*, 15, 133–164. <https://doi.org/10.1146/annurev.es.15.110184.001025>
- Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969–980. <https://doi.org/10.1111/j.1365-294X.2004.02125.x>
- Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, 263, 1619–1626. <https://doi.org/10.1098/rspb.1996.0237>
- Berlacher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology*, 47, 773–815. <https://doi.org/10.1146/annurev.ento.47.091201.145312>
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & Roy-Choudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29, 1917–1932.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D.-J. (2011). Stacks: Building and Genotyping loci de novo from short-read sequences. *G3: Genes Genomes, Genetics*, 1, 171–182. <https://doi.org/10.1534/g3.111.000240>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Comes, H. P., & Kadereit, J. W. (1998). The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, 3, 432–438. [https://doi.org/10.1016/S1360-1385\(98\)01327-2](https://doi.org/10.1016/S1360-1385(98)01327-2)
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30, 1187–1189. <https://doi.org/10.1093/bioinformatics/btt763>
- Cornuet, J.-M., Ravigné, V., & Estoup, A. (2010). Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, 11, 401. <https://doi.org/10.1186/1471-2105-11-401>
- Coyne, J., & Orr, H. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.
- De Vos, J. M., Hughes, C. E., Schneeweiss, G. M., Moore, B. R., & Conti, E. (2014). Heterostyly accelerates diversification via reduced extinction in primroses. *Proceedings of the Royal Society B: Biological Sciences*, 281, 1–9.
- Du, F. K., Petit, R. J., & Liu, J. Q. (2009). More introgression with less gene flow: Chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Molecular Ecology*, 18, 1396–1407. <https://doi.org/10.1111/j.1365-294X.2009.04107.x>
- Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Favre, A., Päckert, M., Pauls, S. U., Jähmig, S. C., Uhl, D., Michalak, I., & Muellner-Riehl, A. N. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews*, 90, 236–253. <https://doi.org/10.1111/brv.12107>
- Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Ecological Conservation*, 24, 38–49.
- Fitzpatrick, B. M., & Turelli, M. (2006). The geography of mammalian speciation: Mixed signals from phylogenies and range maps. *Evolution*, 60, 601. <https://doi.org/10.1111/j.0014-3820.2006.tb01140.x>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Gavrilets, S., & Boake, C. R. (1998). On the evolution of premating isolation after a founder event. *The American Naturalist*, 152, 706–716. <https://doi.org/10.1086/286201>
- Guggisberg, A., Mansion, G., & Conti, E. (2009). Disentangling reticulate evolution in an arctic-alpine polyploid complex. *Systematic Biology*, 58, 35–73.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307–321. <https://doi.org/10.1093/sysbio/syq010>

- Hadfield, J. D. (2010). MCMC methods for multiresponse generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33, 1–22.
- Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, 359, 183–195. <https://doi.org/10.1098/rstb.2003.1388>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., ... Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330, 927–931. <https://doi.org/10.1126/science.1194585>
- Hu, C., & Kelso, S. (1996). Primulaceae. In Z. Y. Wu & P. H. Raven (Eds.), *Flora of China* (pp. 99–185). Beijing, China: Science Press.
- Huang, Z., Young, N. D., Reagon, M., Hyma, K. E., Olsen, K. M., Jia, Y., & Caicedo, A. L. (2017). All roads lead to weediness: Patterns of genomic divergence reveal extensive recurrent weedy rice origins from South Asian *Oryza*. *Molecular Ecology*, 26, 3151–3167. <https://doi.org/10.1111/mec.14120>
- Hudson, R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18, 337–338.
- Jakobsson, M., & Rosenber, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jia, D. R., Abbott, R. J., Liu, T. L., Mao, K. S., Bartish, I. V., & Liu, J. Q. (2012). Out of the Qinghai-Tibet Plateau: Evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophae rhamnoides* (Elaeagnaceae). *New Phytologist*, 194, 1123–1133. <https://doi.org/10.1111/j.1469-8137.2012.04115.x>
- Lexer, C., Wüest, R. O., Mangili, S., Heuert, M., Stölting, K. N., Pearman, P. B., ... Bossolini, E. (2014). Genomics of the divergence continuum in an African plant biodiversity hotspot, I: Drivers of population divergence in *Restio capensis* (Restionaceae). *Molecular Ecology*, 23, 4373–4386. <https://doi.org/10.1111/mec.12870>
- Li, L., Abbott, R. J., Liu, B., Sun, Y., Li, L., Zou, J., ... Liu, J. (2013). Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. *Molecular Ecology*, 22, 5237–5255. <https://doi.org/10.1111/mec.12466>
- Li, J., & Fang, X. (1999). Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin*, 44, 2117–2124. <https://doi.org/10.1007/BF03182692>
- Li, J., Shi, Y., & Li, B. (1995). *Uplift of the Qinghai-Xizang (Tibet) plateau and global change*. Lanzhou, China: Lanzhou University Press.
- Li, Z., Zou, J., Mao, K., Lin, K., Li, H., Liu, J., ... Lascoux, M. (2012). Population genetic evidence for complex evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan plateau. *Evolution*, 66, 831–845. <https://doi.org/10.1111/j.1558-5646.2011.01466.x>
- Liu, J. Q., Gao, T. G., Chen, Z. D., & Lu, A. M. (2002). Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Molecular Phylogenetics and Evolution*, 23, 307–325. [https://doi.org/10.1016/S1055-7903\(02\)00039-8](https://doi.org/10.1016/S1055-7903(02)00039-8)
- Liu, J., Möller, M., Provan, J., Gao, L. M., Poudel, R. C., & Li, D. Z. (2013). Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, 199, 1093–1108. <https://doi.org/10.1111/nph.12336>
- Liu, J. Q., Wang, Y. J., Wang, A. L., Hideaki, O., & Abbott, R. J. (2006). Radiation and diversification within the *Ligularia-Cremathodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution*, 38, 31–49. <https://doi.org/10.1016/j.ympev.2005.09.010>
- Liu, J., Yan, H.-F., & Ge, X.-J. (2016). The use of DNA barcoding on recently diverged species in the genus *Gentiana* (Gentianaceae) in China. *PLoS ONE*, 11, e0153008. <https://doi.org/10.1371/journal.pone.0153008>
- Ma, X. F., Szmidi, A. E., & Wang, X. R. (2006). Genetic structure and evolutionary history of a diploid hybrid pine *Pinus densata* inferred from the nucleotide variation at seven gene loci. *Molecular Biology and Evolution*, 23, 807–816. <https://doi.org/10.1093/molbev/msj100>
- Ma, Y. P., Tian, X. L., Zhang, J. L., Wu, Z. K., & Sun, W. B. (2014). Evidence for natural hybridization between *Primula beesiana* and *P. bulleyana*, two heterostylous primroses in NW Yunnan, China. *Journal of Systematics and Evolution*, 52, 500–507. <https://doi.org/10.1111/jse.12077>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15, 28–41. <https://doi.org/10.1111/1755-0998.12291>
- Matthews, J. V. Jr., & Ovenden, L. E. (1990). Late Tertiary plant macrofossils from localities in arctic, sub-arctic North America—a review of the data. *Arctic*, 43, 364–392.
- Mayr, E. (1963). *Animal species and evolution*. Cambridge, UK: Belknap Press of Harvard University Press. <https://doi.org/10.4159/harvard.9780674865327>
- Meirams, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4, 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- Mulch, A., & Chamberlain, C. P. (2006). Earth science: The rise and growth of Tibet. *Nature*, 439, 670–671. <https://doi.org/10.1038/439670a>
- Muñoz-Pajares, A. J., García, C., Abdelaziz, M., Bosch, J., Perfectti, F., & Gómez, J. M. (2017). Drivers of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Molecular Ecology*, 26, 1576–1585. <https://doi.org/10.1111/mec.13971>
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403, 853–858. <https://doi.org/10.1038/35002501>
- Narum, S. R., & Hess, J. E. (2011). Comparison of F ST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194. <https://doi.org/10.1111/j.1755-0998.2011.02987.x>
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17, 2103–2106. <https://doi.org/10.1111/j.1365-294X.2008.03715.x>
- Nosil, P. (2012). *Ecological speciation*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199587100.001.0001>
- Opgenoorth, L., Vendramin, G. G., Mao, K., Miehe, G., Miehe, S., Liepelt, S., ... Ziegenhagen, B. (2010). Tree endurance on the Tibetan Plateau marks the world's highest known tree line of the last glacial maximum. *New Phytologist*, 185, 332–342. <https://doi.org/10.1111/j.1469-8137.2009.03007.x>
- Pante, E., Abdelkrim, J., Viricel, A., France, S. C., Boisselier, M. C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity*, 114, 450–459. <https://doi.org/10.1038/hdy.2014.105>
- Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution*, 24, 386–393. <https://doi.org/10.1016/j.tree.2009.02.011>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Ren, G., Conti, E., & Salamin, N. (2015). Phylogeny and biogeography of *Primula* sect. *Armerina*: Implications for plant evolution under climate change and the uplift of the Qinghai-Tibet Plateau. *BMC Evolutionary Biology*, 15, 161. <https://doi.org/10.1186/s12862-015-0445-7>
- Ren, G., Mateo, G., Liu, J., Suchan, T., Alvarez, N., Guisan, A., ... Salamin, N. (2017). Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing. *New Phytologist*, 213, 1500–1512. <https://doi.org/10.1111/nph.14221>



- Renner, S. S. (2016). Available data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *Journal of Biogeography*, 43, 1479–1487. <https://doi.org/10.1111/jbi.12755>
- Rice, W. R., & Hostert, E. E. (1993). Laboratory experiments on speciation: What have we learned in forty years? *Evolution*, 47, 1637–1653. <https://doi.org/10.1111/j.1558-5646.1993.tb01257.x>
- Richards, J. (2003). *Primula*. Portland, OR: Timber Press.
- Rieseberg, L. H. (1997). Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, 28, 359–389. <https://doi.org/10.1146/annurev.ecolsys.28.1.359>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138.
- Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8, 336–352. <https://doi.org/10.1111/j.1461-0248.2004.00715.x>
- Sauquet, H., Ho, S. Y. W., Gandolfo, M. A., Jordan, G. J., Wilf, P., Cantrill, D. J., ... Udovicic, F. (2012). Testing the impact of calibration on molecular divergence times using a fossil-rich group: The case of *Nothofagus* (Fagales). *Systematic Biology*, 61, 289–313. <https://doi.org/10.1093/sysbio/syr116>
- Schmidt-Lebuhn, A. N., de Vos, J. M., Keller, B., & Conti, E. (2012). Phylogenetic analysis of *Primula* section *Primula* reveals rampant non-monophyly among morphologically distinct species. *Molecular Phylogenetics and Evolution*, 65, 23–34. <https://doi.org/10.1016/j.ympev.2012.05.015>
- Sun, Y., Abbott, R. J., Li, L., Li, L., Zou, J., & Liu, J. (2014). Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: Homoploid hybrid origin and Pleistocene expansion. *Molecular Ecology*, 23, 343–359. <https://doi.org/10.1111/mec.12599>
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.
- Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One*, 8(2), e51518. <https://doi.org/10.1371/journal.pone.0055158>
- Templeton, A. (1980). The theory of speciation via the founder principle. *Genetics*, 94, 1011–1038.
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., ... Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, 22, 787–798. <https://doi.org/10.1111/mec.12023>
- Wang, Q., Liu, J., Allen, G. A., Ma, Y., Yue, W., Marr, K. L., & Abbott, R. J. (2016). Arctic plant origins and early formation of circumpolar distributions: A case study of the mountain sorrel, *Oxyria digyna*. *New Phytologist*, 209, 343–353. <https://doi.org/10.1111/nph.13568>
- Wang, H., Qiong, L., Sun, K., Lu, F., Wang, Y., Song, Z., ... Zhang, W. (2010). Phylogeographic structure of *Hippophae tibetana* (Elaeagnaceae) highlights the highest microrefugia and the rapid uplift of the Qinghai-Tibetan Plateau. *Molecular Ecology*, 19, 2964–2979. <https://doi.org/10.1111/j.1365-294X.2010.04729.x>
- Weinberg, J. R., Starczak, V. R., & Jorg, D. (1992). Evidence for rapid speciation following a founder event in the laboratory. *Evolution*, 46, 1214. <https://doi.org/10.1111/j.1558-5646.1992.tb00629.x>
- Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114–138.
- Wu, C. (1987). *Flora of Tibet*. Beijing, China: Science Press.
- Xing, Y., & Ree, R. H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 3444–3451. <https://doi.org/10.1073/pnas.1616063114>
- Xu, T., Abbott, R. J., Milne, R. I., Mao, K., Du, F. K., Wu, G., ... Liu, J. (2010). Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *BMC Evolutionary Biology*, 10, 194. <https://doi.org/10.1186/1471-2148-10-194>
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, 292, 686–693. <https://doi.org/10.1126/science.1059412>
- Zhang, Q., Chiang, T. Y., George, M., Liu, J. Q., & Abbott, R. J. (2005). Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology*, 14, 3513–3524. <https://doi.org/10.1111/j.1365-294X.2005.02677.x>
- Zheng, H., Powell, C. M., An, Z., Zhou, J., & Dong, G. (2000). Pliocene uplift of the northern Tibetan Plateau. *Geology*, 28, 715. [https://doi.org/10.1130/0091-7613\(2000\)28<715:PUOTNT>2.0.CO;2](https://doi.org/10.1130/0091-7613(2000)28<715:PUOTNT>2.0.CO;2)
- Zhou, Y. F., Abbott, R. J., Jiang, Z. Y., Du, F. K., Milne, R. I., & Liu, J. Q. (2010). Gene flow and species delimitation: A case study of two pine species with overlapping distributions in Southeast China. *Evolution*, 64, 2342–2352.
- Zhou, W.-W., Wen, Y., Fu, J.-Z., Xu, Y.-B., Jin, J.-Q., Ding, L., ... Zhang, Y.-P. (2012). Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai-Tibetan Plateau. *Molecular Ecology*, 21, 960–973. <https://doi.org/10.1111/j.1365-294X.2011.05411.x>
- Zhu, X., Li, Y., Wu, G., Fang, Z., Li, Q., & Liu, J. (2009). Molecular and morphological evidence for natural hybridization between *Primula secundiflora* Franchet and *P. poissonii* Franchet (Primulaceae). *Acta Biologica Cracoviensia Series Botanica*, 51, 29–36.

BIOSKETCH

Guangpeng Ren has finished his PhD study in the Department of Computational Biology at University of Lausanne and is now working as a young investigator in Lanzhou University, China. This paper is part of his thesis, supervised by **Nicolas Salamin**, and focusing on the influence of historical factors in the evolutionary history of *Primula* on Qinghai-Tibet Plateau.

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. R.G.M. performed the SDM analysis. G.R. and N.S. wrote the manuscript with the help of R.G.M., A.G., and E.C.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ren G, Mateo RG, Guisan A, Conti E, Salamin N. Species divergence and maintenance of species cohesion of three closely related *Primula* species in the Qinghai-Tibet Plateau. *J Biogeogr*. 2018;45:2495–2507. <https://doi.org/10.1111/jbi.13415>