Year : 2017

Genetic conséquences of historically geological and climatic events in Qinghai-Tibet Plateau : the Primula section Armerina as a case study

Ren Guangpeng

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Département d'Écologie et d'Évolution

GENETIC CONSEQUENCES OF HISTORICALLY GEOLOGICAL AND CLIMATIC EVENTS IN THE QINGHAI-TIBET PLATEAU : THE *PRIMULA* SECTION *ARMERINA* AS A CASE STUDY

Thèse de doctorat ès sciences de la vie (PhD)
Écologie et Évolution

présentée à la

Faculté de biologie et de médecine de l'Université de Lausanne

par

Guangpeng REN

Master en Ecology et Lanzhou Université, Chine

Jury

Prof. Bengt Kayser, Président
Prof. Nicolas Salamin, Directeur de thèse
Prof. Christian Lexer, expert
Prof. John Pannell, expert

Lausanne 2017

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président· e	Monsieur	Prof. Bengt **Kayser**
Directeur· rice de thèse	Monsieur	Prof. Nicolas **Salamin**
Experts· es	Monsieur	Prof. John **Pannell**
	Monsieur	Prof. Christian **Lexer**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Guangpeng Ren

Master in Ecology of Lanzhou University, China

intitulée

GENETIC CONSEQUENCES OF HISTORICALLY GEOLOGICAL AND CLIMATIC EVENTS IN THE QINGHAI-TIBET PLATEAU: THE *PRIMULA* SECTION *ARMERINA* AS A CASE STUDY

Lausanne, le 22 mai 2017

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Bengt Kayser

To my lovely family

Table of contents

Abstract

Understanding the relative roles of geography and ecology in driving speciation, population divergence and population dynamics is a longstanding goal in ecology and evolutionary biology. This is especially true for organisms inhabiting in mountains that usually harbor extremely high species richness and endemism. In this thesis, I combine macro-evolutionary approaches and population genomic analyses at multiple evolutionary timescales to investigate how historically geological and climatic events affected diversification, speciation and demography of *Primula* section *Armerina* in the Qinghai-Tibet Plateau (QTP). At macro-evolutionary level, I reconstruct phylogenetic trees of the section and discuss the factors accounting for the incongruence between the chloroplast and nuclear trees that mainly occur among closely related species. I demonstrate that the section originated from the Himalayas and the recent uplift of the QTP has probably triggered its diversification. I also show that niche evolution affects biogeographic patterns of three closely related species in this section. Then, I use population genomic data to examine the interspecific divergence and maintenance of species cohesion between these three closely related species. I successfully obtain a clear relationship among the three species and provide a strong evidence for an origin of the three species in the Himalayas. The following divergences between them coincide with the uplift of the Hengduan Mountains and the Northeast QTP, which highlights the important roles played by past geological events in triggering initial interspecific divergence. After a long period of divergence, the three species came into secondary contact triggered by past climatic changes but with no significant hybridization. I further show that spatial and environmental factors may play an important role in the maintenance of species cohesion. In the final two chapters, I investigate demographic histories of two of the three closely related species that are endemic to the QTP and evaluate the effects of the Quaternary climatic oscillations on their demography. I demonstrate that only the ancestral populations of the two species could survive in different refugia during the largest glaciation that occurred in the QTP. Most of the genetic lineages identified by the population genomic analyses diverged from their ancestral populations only after this glaciation event. The last glacial maximum (LGM) had littler effects on these genetic lineages. Moreover, the response to climatic changes of populations of a species depends on its specific ecological preferences. Finally, all the historical, spatial and environmental factors act as drivers of population differentiation in these alpine plant species. Overall, this work contributes to a significant advance to our understanding of the mechanisms of the interplay between geological and ecological factors in driving speciation and evolution in mountains.

Résumé

Comprendre de quelle manière la géographie et l'écologie contribuent au processus de spéciation, à la divergence et la dynamique des populations est un des objectifs majeurs de l'écologie et de la biologie évolutive. Cet intérêt est particulièrement fort pour les régions montagneuses qui présentent une grande diversité d'espèces et un fort taux d'endémisme. Dans cette thèse, j'intègre des approches macro-évolutive avec des analyses de génomique des populations à différentes échelles de temps pour étudier comment les évènements géologiques et climatiques passés ont affecté la diversification, la spéciation et la démographie des espèces de la section Armenia du genre Primula du plateau Tibétain. Au niveau macro-évolutif, j'ai reconstruit l'arbre phylogénétique du groupe et émis des hypothèses pour expliquer l'incongruence observée pour les espèces proches entre les arbres phylogénétiques basés sur des marqueurs nucléaires et chloroplastiques. J'ai démontré que la section Armenia est originaire de l'Himalaya et que la récente orogenèse de cette chaine de montagne a probablement été le facteur déclencheur de sa diversification. J'ai aussi montré que l'évolution de la niche a influencé la biogéographie de trois espèces proches de cette section dont j'ai pu établir avec précision les relations de parenté et leur origine dans l'Himalaya. La divergence entre ces trois espèces coincide avec l'orogenèse des monts Hengduan et le Nord-Est du plateau Tibétain, ce qui démontre l'importance des évènements géologiques passés dans la divergence inter-spécifique. Après une longue période d'isolation les trois espèces, des changements climatiques ont par le passé induit un contact secondaire mais sans que cela entraine une hybridisation significative. J'ai ensuite montré que les facteurs spatiaux et environnementaux pourraient jouer un rôle important dans le maintien de la cohésion des espèces. Dans les deux derniers chapitres, j'étudie l'histoire démographique de deux des trois espèces proches endémiques du plateau Tibétain et j'évalue les effets des oscillations climatiques du Quaternaire sur leur démographie. Je démontre que les populations ancestrales des deux espèces ont pu survivre dans différents refuges au cours de la plus longue glaciation du plateau Tibétain. La plupart des lignées génétiques identifiées par les analyses de génomique des populations ont divergé de ces populations ancestrales suite à cet évènement de glaciation. Le dernier maximum glaciaire n'a eu que peu d'effet sur ces lignées. D'autre part, la réponse des populations d'une espèce aux changements climatiques dépend de ses préférences écologiques spécifiques. Enfin, l'ensemble des facteurs historiques, spatiaux et environnementaux agissent comme des moteurs de la différenciation de ces espèces montagneuses. Dans l'ensemble ce travail contribue de manière significative à la compréhension des méchanismes de l'interaction entre les facteurs géologiques et écologiques dans la formation et l'évolution des espèces au sein des massifs montagneux.

General introduction

Understanding the relative roles of geography and ecology in driving speciation, population divergence and population dynamics is a longstanding goal in ecology and evolutionary biology (Coyne & Orr 2004; Nosil 2012). Speciation is a by-product of various processes between diverging populations, such as allopatric isolation triggered by geological processes (Mayr 1963; Rice & Hostert 1993), development of genomic incompatibilities (Dobzhansky 1936; Muller 1939), hybridization or polyploidy (Soltis & Soltis 2009; Abbott *et al.* 2013) and divergent natural selection (i.e. ecological speciation, Nosil 2012). Although geographical isolation leading to allopatric divergence is generally a key factor in speciation and a common process by which new species arises, ecological speciation has recently been indicated as another common speciation model (Berlocher & Feder 2002; Rundle & Nosil 2005; Nosil 2008; Nosil 2012). The interplay between these different processes may become very complex for organisms inhabiting mountainous areas, which usually harbor high species diversity and endemism, and are identified as biodiversity hotspots (Myers *et al.* 2000). At macro-evolutionary level, the historical orogenesis and associated climatic changes in mountains may have triggered evolutionary radiations (e.g. Liu *et al.* 2006; Hoorn *et al.* 2010). At smaller taxonomic scale, for example between closely related species, the past environmental changes may have caused the shifts of their distributions, leading to secondary contact between previously isolated species, which may result in introgression or even hybrid speciation if the reproductive isolation was incomplete (e.g. Rieseberg 1997; Abbott *et al.* 2013; Sun *et al.* 2014). Finally, demographic changes involved in these range shifts can also affect the spatial patterns of genetic variation within and among populations of a species (Hewitt 2004). Integrating studies at these three scales will advance our understanding of the effects of past geological and climatic changes on evolutionary history of species in mountains.

The temporal and spatial framework at macro-evolutionary level

Evolution and diversification of organisms may be predominantly driven by biotic factors (e.g. competition and predation; Antonelli & Sanmartin, 2011), abiotic factors (such as geological and climatic events), or a combination of both. Biotic factors, as described by the 'Red Queen' model, tend to act over relatively short periods of time and more locally, whereas abiotic factors, as described by the 'Court Jester' model, tend to act over several millions of years and within a climatic zone, a continent, or globally, and generate large-scale patterns (Benton 2009; Favre *et al.* 2015). For mountains that have undergone uplifts over several dozen millions of years, the impact of abiotic processes on the observed patterns likely outweighs that of biotic processes. The effects of geological and associated climatic dynamics on species diversification have received attention of evolutionary biologist for many years. These past events have generally a differential effect on diversification through time and between geographical areas (Benton 2009) and studying this process is one of the

key aspects to identify the factors that promote biodiversity.

The dynamic of species diversification through time can be estimated by molecular clock analysis, such as using Bayesian approaches implemented in BEAST (Drummond & Rambaut 2007; Drummond *et al.* 2012) or MrBayes (Ronquist *et al.* 2012), or by estimating rates of speciation and extinction as they are often vary over time (e.g. Morlon *et al.* 2010; Stadler 2011). The effects of historical orogenesis on diversification can then be evaluated by studying the correlation between the timing of these events and the evolution of the organisms. Firstly, a dated phylogenetic tree can be ideally achieved by sampling a reasonable density of taxa and using multiple reliable fossil evidences, which could allow for a good calibration during the molecular dating analysis (Renner 2005; Albert *et al.* 2009). However, this might represent a challenge, since the fossil record is scarce for most taxonomic groups. In some cases, the sampling of the focal group might be amended by phylogenetically related clades, to be able to include fossils available for the latter. In fact, sampling more outgroup species to include more external fossils was suggested as a better alternative than secondary calibration, because the latter may be inherently subjected to bias and errors, and thus should be interpreted with caution (Blair Hedges & Kumar 2004; Graur & Martin 2004; Sauquet *et al.* 2012). Secondly, studying the temporal variation of diversification rates can address questions such as rapid radiations, massive extinctions or temporal rate variation in general (e.g. Valente *et al.* 2010; Stadler 2011; Condamine *et al.* 2013). Furthermore, it allows comparisons between geographic units, between lineages and with respect to different traits (Valente *et al.* 2010; Linder 2008; Serrano-Serrano *et al.* 2017). Currently, several programmes for the study of diversification rates derived from either extant or fossil data are available, for example MEDUSA (Modeling Evolutionary Diversification Using Stepwise Akaike Information Criterion, Alfaro *et al.* 2009), fossilMEDUSA, BayesRate (Silvestro *et al.* 2011) and PyRate (Silvestro *et al.* 2014).

Another important aspect of studying species diversification is to investigate areas of origin as well as migration routes at the spatial scale. This can be deduced from biogeographic analyses with various analytical methods available to assess the likelihood of alternative biogeographic hypotheses (Lomolino *et al.* 2010), such as dispersal-vicariance analysis (DIVA) and dispersal-extinction-cladogenesis (DEC), with software like RASP (Yu *et al.* 2015), Bayes-DIVA (Nylander *et al.* 2008) and LAGRANGE (Ree *et al.* 2008). All these methods are based on a phylogenetic inference that allows reconstructing the biogeographic history of ancestral clades. The integration of the temporal and spatial framework can provide insights into the patterns and processes of species diversification in response to historical events.

The recent literature has contributed great advances for our understanding of species diversification in response to past geological and climatic changes. For example, Andean uplift has played important

roles in the evolution of Amazonian landscapes and ecosystems, and the accumulation of current biodiversity (reviewed in Hoorn *et al.* 2010). Compared to the Andes, the evolution of biotas in the region of the Qinghai-Tibet Plateau (QTP) remains insufficiently understood despite its outstanding geographic extent (Liu *et al.* 2014; Wen *et al.* 2014; Favre *et al.* 2015). Additionally, insights into the patterns and processes of species diversification have been hampered by the often poorly resolved species phylogenies based on traditional markers. The advance of next-generation sequencing (NGS) approaches brings now great potentials for efficiently sampling entire genomes of species for phylogenetically informative variation (McCormack *et al.* 2013). The NGS approaches have provided unprecedented power and been recently utilized to explore phylogenetic relationships of recently diverged taxa or radiations, including East-African cichlids (Wagner *et al.* 2013), American oaks (Hipp *et al.* 2014), some American sedges (Escudero *et al.* 2014) and *Primula* section *Auricula* (Boucher *et al.* 2016). These new approaches should be employed across a wider range of taxa to better understand the processes occurring during species diversification in mountains.

Interspecific divergence among closely related species

The study of closely related species at a population level can offer insights into relative importance of geographical versus ecological divergence (Abbott *et al.* 2000; Jia *et al.* 2012; Anacker & Strauss 2014) and thus help to understand the mechanisms involved during speciation. To address this issue, we however need to clearly delimit the evolutionary relationships between closely related species. The application of genetic data to build a phylogenetic tree could provide satisfactory results in distantly related taxa, but suffers from a number of issues when dealing with evolutionary relationships at shallow time depths (i.e. closely related species; Nater *et al.* 2015). In such cases, a phylogenetic tree inferred from any given genomic locus (a 'gene' tree) might not unequivocally reflect the true order of speciation events (the 'species' tree), and inconsistent topologies are often obtained across the genome (Maddison 1997; Edwards 2009). Within groups of closely related species, genome-wide variation in gene trees are caused mainly by two biological processes, namely incomplete lineage sorting (ILS) and interspecific gene flow (Degnan & Rosenberg 2009).

ILS occurs when lineages fail to coalesce in the ancestral population of two species (Pamilo & Nei 1988; Maddison & Wiens 1997; Degnan & Salter 2005). Therefore, the probability of ILS depends on both the effective population size (Ne) in the ancestral population of two species, which determines the rate of coalescence of lineages, and the time between two successive speciation events (Hudson 1990; Degnan & Salter 2005). Interspecific gene flow occurs when previously geographically isolated species come into secondary contact usually due to environmental changes. Genes from one species can thus spread into the other one by hybridization and fertility of the hybrid progeny (Abbott *et al.* 2013). The genealogy of these lineages will support a clustering of species that exchanged genes after

their initial split regardless of their evolutionary relationship. Interspecific gene flow can also occur during ecological speciation (Nosil 2012). However, the rate of gene flow varies along the genome, with restricted gene exchange in gene regions linked to reproductive isolation or local adaptation (Wu & Ting 2004; Via & West 2008; Nosil *et al.* 2009), which is called divergent hitchhiking. This restriction of gene flow can extend to the whole genome because of linkage disequilibrium, i.e., genome hitchhiking. Eventually, the barrier to gene flow between species evolves before post speciation divergence (Wu 2001; Feder *et al.* 2012).

The NGS approach is essential to uncover the evolutionary histories of closely related species (e.g. Wagner *et al.* 2013; Pante *et al.* 2015; Meier *et al.* 2017). Given data from thousands of independent loci and multiple individuals per species, the coalescence-based modeling approaches under a given demographic model can reconstruct evolutionary relationships by taking into account both ILS and interspecific gene flow. They are further able to harvest information about effective population size (Ne) in both current and ancestral populations (Hudson 1990; Liu *et al.* 2007; Heled & Drummond 2010). However, the calculation of the likelihood-based function is computationally expensive for genomic level data. Recent advances in approximate methods like approximate Bayesian computation (ABC) offer an elegant way around the problem of solving the complex likelihood function (Beaumont *et al.* 2002; Marjoram *et al.* 2003; Excoffier *et al.* 2013). These approximations allow the investigation of the influence of complex demographic processes on the reconstruction of species trees in closely related species, given the availability of both a large number of independent loci and a population sample from each species. With the software such as DIY-ABC, ABCtoolbox package, *dadi* and fastsimcoal, recent studies have successfully uncovered evolutionary relationships of closely related species in animals, fish and plants (Rubin *et al.* 2012; Nater *et al.* 2015; Wang *et al.* 2016; Meier *et al.* 2017).

Population structure and demographic history

Species tend to occupy heterogeneous environments across their geographic range, and the magnitude and spatial distribution of their genetic variation is expected to vary accordingly (Linhart & Grant 1996; Anderson *et al.* 2011). According to the model of isolation by distance (IBD, Wright 1943), genetic drift may cause populations to become more different from one another at greater geographical distances because gene flow declines among increasingly distant populations. This scenario has been largely supported since its formulation in the 1940s (Dobzhansky & Wright 1943; Imaizumi & Morton 1969; Sharbel *et al.* 2000). Alternatively, a scenario coined ‘isolation by environment’ has been frequently recovered in animals and plants. It proposes that gene flow among populations living in different environments is limited primarily by selection against maladapted migrants (Nosil *et al.* 2009). Furthermore, Quaternary climatic oscillations have been suggested to be a major factor in

shaping large-scale spatial genetic pattern (Hewitt 2000). During the cycles of glacial and interglacial periods, populations repeatedly retreated to one or multiple refugia and recolonized their preferred habitats with potential secondary contact among isolated populations (Abbott *et al.* 2000; Taberlet & Cheddadi 2002). The demographic changes involved in these range shifts affected the spatial patterns of genetic variation within and among populations (Hewitt 2004). Recent empirical studies have put forward that all spatial, environmental and historical factors act as drivers of spatial genetic patterns simultaneously at different spatial scales (Wang *et al.* 2013; Lexer *et al.* 2014; Muñoz-Pajares *et al.* 2017), especially for the organisms inhabiting mountains.

Identifying the genetic structure of a species and the factors that drive it is an important step in understanding of speciation, adaptation and genetic change (Antonovics 1968). Knowing demographic history of a species in response to past environmental changes could shed light on population management because the dynamic spatio-temporal histories of populations can profoundly impact their future evolutionary potential (Lanier *et al.* 2015). Population genomic data provide accurate estimates of genetic structure (Avise 2010; Narum *et al.* 2013), and provide opportunities for estimating demographic hypotheses such as population divergence, constant population size though time, population expansion or bottlenecks, the rate of growth or decline, and migration and changes in population sizes in subdivided populations (e.g. Emerson *et al.* 2010; Bourret *et al.* 2013; Lanier *et al.* 2015). The core of many methods for estimating these demographic parameters is Monte Carlo simulations based on coalescent theory. With the development of genomic sequencing technologies, big data sets from multiple populations will increase the accuracy of the estimation of demographic parameters, for example by using recent developed bioinformatics tools such as approximate Bayesian computation (ABC) modeling (Cornuet *et al.* 2014). Furthermore, a combination of evolutionary modeling and species distribution modeling (SDM) could provide new insights to predict the impact of future climatic changes on population dynamics (e.g. Lanier *et al.* 2015). Additionally, population genomics allow discerning genome 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).

The study area and NGS approaches

The Qinghai-Tibet Plateau (QTP) is the highest and largest plateau in the world with an average altitude of more than 4000 m, and arguably the most prominent topological feature on Earth. The uplifts of the plateau were driven by the collision of the Indian plate with the Eurasian plate, which began at *ca.* 50 million years ago (Ma; Rowley & Currie 2006; Royden *et al.* 2008). Nevertheless, the times of its following uplifts are controversial (as reviewed in Renner et al. 2016). Some scientists

believe that the QTP has reached 4000 m since the mid-Eocene (40 Ma; Renner et al. 2016 and references therein), while others suggest several recent uplifts that occurred since the early Miocene in at least three major periods: 30–23, 15–13, 8–1.6 Ma (Harrison *et al.* 1992; Li *et al.* 1995; Shi *et al.* 1998; Spicer *et al.* 2003). The recent uplift occurred particularly at its eastern and northern edge in the Hengduan Mountains and the Qaidam basin (Li & Fang 1999; Zheng *et al.* 2000; Mulch & Chamberlain 2006). The rise of the QTP created a large altitudinal gradient across the region and modified the global and East Asian climate dramatically (Ruddiman & Kutzbach 1989; Shi *et al.* 1998), triggering and intensifying the Asian monsoon, which in turn profoundly influenced biological processes in the QTP (Li & Fang 1999). At present, about 9,000–12,000 species of vascular plants belonging to *ca.* 1,500 genera occur in this plateau, and at least 20% of these species and *ca.* 50 genera are endemic (Wu 1987). Therefore, the QTP flora comprises several of the important alpine biodiversity hotspots in the world. The fringe of the QTP encompasses parts of four different hotspots of biodiversity (Figure 1), which are listed among the main biodiversity hotspots of the Northern Hemisphere (Myers *et al.* 2000; Myers & Mittermeier 2003; Tang *et al.* 2006) and are assumed to be particularly vulnerable to climate change (Zheng 1996; Yao *et al.* 2007). During the Quaternary, there have been four major glaciations in the QTP, becoming progressively less extensive after the largest Naynayxungla Glaciation, which began *ca.* 1.2 Ma and reached its maximum between 0.8 and 0.5 Ma in the QTP (Shi 2002; Zheng *et al.* 2002). Glacier advances during the last glacial maximum (LGM) were significantly less extensive in this area (Shi *et al.* 1998; Zheng *et al.* 2002; Owen 2009). The uplifts of the QTP and Quaternary climatic oscillations have been widely proposed to facilitate speciation and diversification (Sun 2002), and to have shaped the geographic scale of genetic structure and the recolonization patterns from multiple glacial refugia in the QTP (Qiu *et al.* 2011; Liu *et al.* 2014; Wen *et al.* 2014).

Recent literature has greatly contributed to our understanding of the effects of these past geological and climatic changes on species evolution and diversification in the QTP (reviewed in Qiu *et al.* 2011; Liu *et al.* 2014; Wen *et al.* 2014; Favre *et al.* 2015). However, compared to other areas, such as Europe and North America, the evolution of biotas in the QTP remains insufficiently understood not only because of its remoteness and inaccessibility, but also because of the limited genetic information (i.e. traditional markers) and less powerful bioinformatics tools used in previous studies. So far, there are very few studies that explored how NGS data could help understand the evolution of organisms in the QTP. With the advent of NGS technologies, applications of genomic-level data at different taxonomic scales are particularly needed to provide a comprehensive understanding of evolutionary history of species in this region.

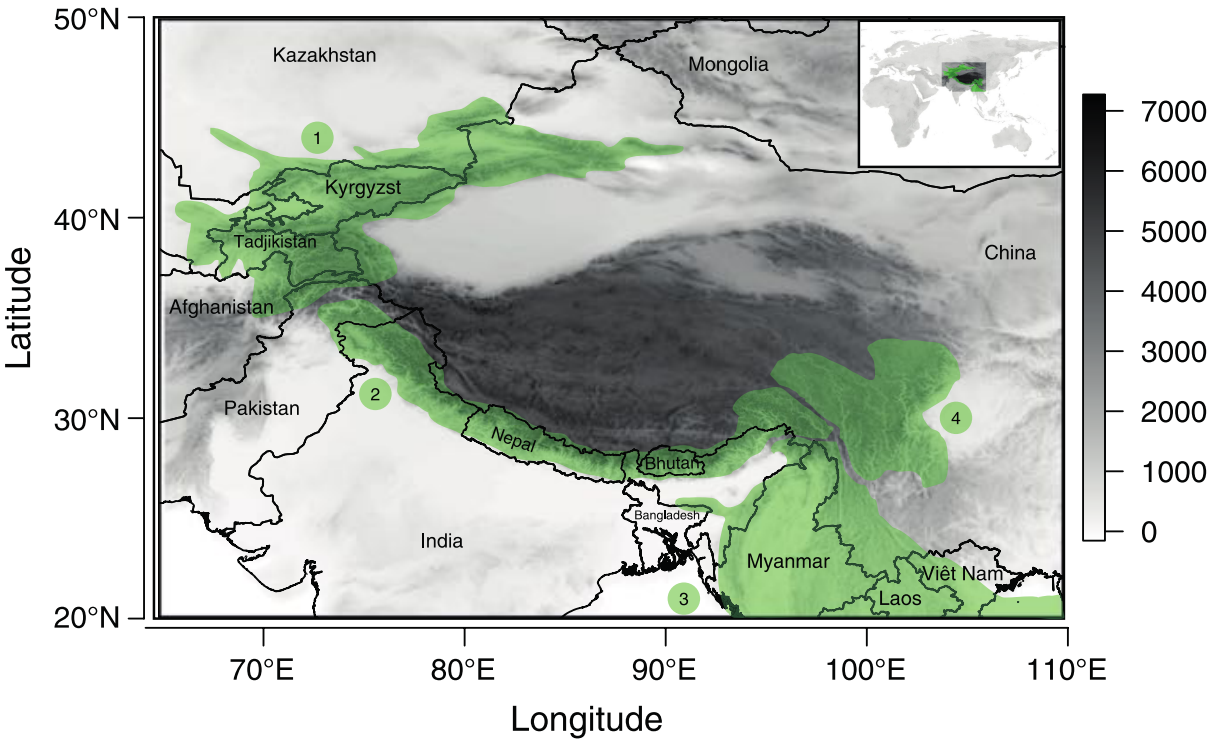


Figure 1. Maps of the region of the Qinghai-Tibet Plateau (QTP; modified from Favre *et al.,* 2015). Four biodiversity hotspots (green areas) surrounding the QTP: (1) mountains of Central Asia; (2) Himalayas; (3) Indo-Burma; (4) Hengduan Mountains.

Currently, the most popular NGS methods include restriction-site associated DNA sequencing (RADseq; Miller *et al.* 2007; Baird *et al.* 2008; Emerson *et al.* 2010; Peterson *et al.* 2012), genotyping-by-sequencing (GBS; Elshire *et al.* 2011), exon capture (DNA enrichment; Hodges *et al.* 2007) and whole genome sequencing (WGS). WGS is more expensive than other methods and rarely used in population genomics (Allendorf *et al.* 2010). RADseq or GBS enables collecting thousands of loci throughout the genome by sequencing short regions adjacent to restriction enzyme cut sites (Miller *et al.* 2007; Baird *et al*. 2008) and is being increasingly used for phylogenetic inference and population genomics (Rubin *et al.* 2012). These methods require no prior genome information and therefore can be used on any species. However, having at least a preliminary draft genome from the study species can improve several aspects of the downstream data analyses, such as the identification of paralogs and of loci that are close together in the genome and likely non-independent due to linkage. The latter can facilitate the identification of linkage groups or genomic regions under divergent or balancing selection (Hohenlohe *et al.* 2010; Andrews & Luikart 2014). The exon capture that requires knowledge of the exon sequences for the study species or a related species to design the capture bait oligonucleotides, is particularly useful for studies focusing on variation in protein-coding genes and the genetic basis of fitness and adaptation, because it provides a gene-targeted approach with SNPs from thousands of genes or the entire exome (Bi *et al.* 2012). The most popular statistical analysis software for RADseq and GBS data are STACKS (Catchen *et al.* 2011, 2013) and PyRAD

(Eaton 2014). The former one is suggested to be more suitable at shallow scales (i.e. population data), whereas the latter one is commonly employed for RADseq studies at deeper phylogenetic studies because of the inclusion of indel variation that improves identification of homology across highly divergent lineages. When a reference genome is available, other genotype calling programmes such as GATK (http://www.broadinstitute.org/gatk/) and SAMtools (Li *et al.* 2009) can also be used.

The Primula study system

Primula L. (Primulaceae) is one of the genera that exhibit high levels of species diversity in the QTP. The group contains ca. 500 species and has a predominantly northern hemispheric distribution, with some representatives in Ethiopia and Southeast Asia and one isolated species in South America (Richards 2003). About 60 % of the species are present in the QTP and its adjacent regions (Hu & Kelso 1996). The floral syndrome 'heterostyly' presented in around 90% of the species in this genus is one of the most remarkable outcrossing mechanisms. Heterostyly is a condition in which populations consist of two floral morphs: 'pins', with anthers in the lower and stigmas in the upper portion of the corolla tube, respectively, and 'thrums' with a reverse arrangement of the sexual organs. The function of heterostyly as a mechanism to promote outcrossing has attracted great attention since Darwin who first elucidated this function in a series of studies of the primrose family (Darwin 1877). The genetic basis of heterostyly in *Primula* is controlled by the S-locus supergene that likely comprises five tightly linked genes, determining style length (the so-called G locus), anther height, pollen size and male and female intra-morph incompatibility (Lewis & Jones 1992). Recent works have built a whole genetic architecture of the S-locus supergene and identified a specific gene CYP734A50 that determines the style-length dimorphism (Huu *et al.* 2016; Li *et al.* 2016). At a macro-evolutionary level, heterostyly was suggested to accelerate species diversification via decreasing extinction rates rather than increasing speciation rates (de Vos *et al*. 2014).

Primula diverged from its closely related genus *Soldanella* ca. 25 Ma (de Vos *et al*. 2014) and has experienced rapid radiation since its origin. As the diversity center of the *Primula*, less attention has been paid to the QTP compared with other regions, such as Europe and North America, where several studies have been conducted to evaluate the effects of the past climatic changes on the genetic structure, hybridization and distribution patterns in multiple species of *Primula* (e.g. Guggisberg *et al*. 2009; Theodoridis *et al*. 2013; Boucher *et al*. 2016; Theodoridis *et al*. 2016). The timeframe of the origin of *Primula* is consistent with the extensive uplifts of the QTP, however, how *Primula* species responded to these historical geological events and the Quaternary climatic oscillation is unclear, especially for the species that are distributed in the QTP. Consequently, the genus *Primula* provides an ideal system to study the mechanisms of the interplay between geological and climatic events in driving speciation and evolution of organisms in the QTP at both the macro- and micro-evolutionary

levels.

The objectives of the thesis

The main goal of this thesis was to investigate the effects of past geological and climatic changes on the evolution and demography of *Primula* species that are distributed in the QTP. To reach this goal, I evaluated the effects of past events within a phylogenetic framework, and used population genomic data to examine interspecific divergence among three closely related species and to reconstruct detailed demographic histories of two species, respectively, at a population level. More specifically:

In **chapter 1**, I investigated the phylogenetic relationships and biogeography of *Primula* section *Armerian* based on five chloroplast and one nuclear genes. The aims were to obtain a detailed and resolved phylogenetic tree for the section, to assess the influence of the uplift of the QTP on its diversification and to illustrate how niche evolution under climatic changes influences biogeographic pattern.

In **chapter 2**, I examined the interspecific divergence and maintenance of species cohesion among three closely related species of *Primula* based on population genomic data (i.e. RADseq data). I chose the three species that are distributed mainly in the Himalayas, the Hengduan Mountains and the Northeastern QTP, respectively, to characterize the interspecific divergence in response to the uplift of the QTP, and to evaluate the effects of past climatic changes and drivers in maintaining species cohesion.

In **chapter 3 and 4**, I focused on *P. tibetica* and *P. fasciculata* to investigate the genetic consequences of Quaternary climatic oscillation in the Himalayas and the Hengduan Mountains, respectively, based on population genomic data. I first identified genetic structure of populations in the two species and the drivers that triggered their intraspecific divergence. Secondly, I used ABC modeling to reconstruct detailed demographic histories of the two species, and combined SDMs with ABC modeling to evaluate the effects of the Quaternary climatic oscillations on their demographic histories. These two chapters represents the first population genomic level studies of alpine plant species occurring in the QTP and contributes to an advance understanding of the role played by Quaternary climatic changes on the present-day distributions of organisms in mountains.

Annex

The annex of this thesis contains another project on which I had the opportunity to work on during the last year of my PhD thesis. In this project, I used 70 whole-genome resequencing data of Cannabis to

estimate its domesticated history and to identify genomic signatures of selection on different types, i.e. hemp and drug. The sampling represents worldwide cultivars and drugs, which allowed me to identify five distinct genetic groups, including two Chinese cultivar groups, European cultivar group, recent selected drug group and feral drugs in South Asia. I have also identified candidate genomic regions that may involve in selection in both the hemp and drug types. Next step will be annotation of these genomic regions to check which function was selected during its domestication.

References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Abbott RJ, Smith LC, Milne RI *et al.* (2000) Molecular Analysis of Plant Migration and Refugia in the Arctic. *Science*, **289**.

Albert EM, San Mauro D, García-París M, Rüber L, Zardoya R (2009) Effect of taxon sampling on recovering the phylogeny of squamate reptiles based on complete mitochondrial genome and nuclear gene sequence data. *Gene*, **441**, 12–21.

Alfaro ME, Santini F, Brock C *et al.* (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 13410–4.

Anacker BL, Strauss SY (2014) The geography and ecology of plant speciation : range overlap and niche divergence in sister species. *Proceedings of the Royal Society B*.

Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics*, **27**, 258–266.

Antonelli A, Sanmartín I (2011) Why are there so many plant species in the Neotropics? *Taxon*, **60**(2), 403–414.

Antonovics J (1968) Evolution in closely adjacent plant populations V. Evolution of self-fertility. *Heredity*, **23**, 219–238.

Avise JC (2010) Perspective: conservation genetics enters the genomics era. *Conservation Genetics*, **11**, 665–669.

Baird N a, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, **3**, e3376.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian Computation in Population Genetics. *Genetics*, **162**.

Benton MJ (2009) The Red Queen and the Court Jester: Species Diversity and the Role of Biotic and Abiotic Factors Through Time. *Science*, **323**.

Berlocher SH, Feder JL (2002) SYMPATRIC SPECIATION IN PHYTOPHAGOUS INSECTS : Moving Beyond Controversy? *Annual Review of Entomology*, **47**, 773–815.

Blair Hedges S, Kumar S (2004) Precision of molecular time estimates. *Trends in Genetics*, **20**, 242–247.

Boucher FC, Casazza G, Szövényi P, Conti E (2016) Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.

Bourret V, Kent MP, Primmer CR *et al.* (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **22**, 532–551.

Condamine FL, Rolland J, Morlon H (2013) Macroevolutionary perspectives to environmental change. *Ecology Letters*, **16**, 72–85.

Cornuet J-M, Pudlo P, Veyssier J *et al.* (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.

Coyne JA, Orr HA (2004) Speciation, Sunderland, Sinauer Associates.

Darwin C (1877) The different forms of flowers on plants of the same species. John Murray.

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, **24**(6), 322–340.

Degnan JH, Salter LA (2005) GENE TREE DISTRIBUTIONS UNDER THE COALESCENT PROCESS. *Evolution*, **59**, 24.

Dobzhansky T (1936) Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila* Pseudoobscura Hybrids. *Genetics*, **21**, 113–35.

Dobzhansky T, Wright S (1943) Genetics of Natural Populations. X. Dispersion Rates in *Drosophila* Pseudoobscura. *Genetics*, **28**, 304–40.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Edwards S V. (2009) IS A NEW AND GENERAL THEORY OF MOLECULAR SYSTEMATICS EMERGING? *Evolution*, **63**, 1–19.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species (L Orban, Ed,). *PLoS ONE*, **6**, e19379.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16196–16200.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust Demographic Inference from Genomic and SNP Data. *PLoS Genetics*, **9**, e1003905.

Favre A, Päckert M, Pauls SU *et al.* (2015) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological reviews*, **90A**, 236–253.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends in Genetics*, **20**, 80–86.

Guggisberg A, Mansion G, Conti E (2009) Disentangling Reticulate Evolution in an Arctic-Alpine Polyploid Complex. *Systematic Biology*, **58**, 55–73.

Hang S (2002) Tethys retreat and Himalayas-Hengduanshan Mountains uplift and their significance on the origin and development of the sino-himalayan elements and alpine flora. *Acta Botanica Yunnanica*, **24**, 273–288.

Harrison TM, Copland P, Kidd WSF, Yin A (1992) Raising Tibet. *Science*, **255**, 1663–1670.

Heled J, Drummond AJ (2010) Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, **27**, 570–580.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B*, 183–195.

Hipp AL, Eaton DAR, Cavender-Bares J *et al.* (2014) A Framework Phylogeny of the American Oak Clade Based on Sequenced RAD Data. *PLoS ONE*, **9**, e93975.

Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010) Amazonia Through Time: Andean Uplift, Climate Change, Landscape Evolution, and Biodiversity. *Science*, **330**.

Hu CM, Kelso S (1996) Primulaceae. In: Wu ZY, Raven PH, eds. *Flora of China* Vol 15. St Louis: Missouri Botanical Garden Press, 99–185. Science Press, Beijng.

Hudson RR (1990) Gene genealogies and the coalescent process. In: Futuyma D.J., Antonovics J.D., editors. Oxford surveys in evolutionary biology. New York: Oxford University Press. p. 1–44.

Huu CN, Kappel C, Keller B *et al.* (2016) Presence versus absence of CYP734A50 underlies the style-length dimorphism in primroses. *eLife*, **5**: e17956.

Imaizumi Y, Morton NE (1969) Isolation by Distance in Japan and Sweden Compared with Other Countries. *Human Heredity*, **19**, 433–443.

Jia DR, Abbott RJ, Liu TL *et al.* (2012) Out of the Qinghai-Tibet Plateau: Evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae). *New Phytologist*, **194**, 1123–1133.

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, **24**, 3688–3705.

Lewis D, Jones DA (1992) The genetics of heterostyly. In *Evolution and function of heterostyly* (pp. 129-150). Springer Berlin Heidelberg.

Lexer C, Wüest RO, Mangili S *et al.* (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.

Li J, Cocker JM, Wright J *et al.* (2016) Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nature Plants*, **2**, 16188.

Li J, Fang X (1999) Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin*, **44**, 2117–2124.

Li JJ, Shi YF, Li BY (1995) Uplift of the Qinghai–Xizang (Tibet) Plateau and Global Change. Lanzhou University Press, Lanzhou.

Linder HP (2008) Plant species radiations: where, when, why? *Philosophical Transactions of the Royal Society*

B: Biological Sciences, **363**, 3097–3105.

Linhart YB, Grant MC (1996) EVOLUTIONARY SIGNIFICANCE OF LOCAL GENETIC DIFFERENTIATION IN PLANTS. *Annual Review of Ecology and Systematics*, **27**, 237–277.

Liu JQ, Duan YW, Hao G, Ge XJ, Sun H (2014) Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution*, **52**, 241–249.

Liu L, Pearl DK, Buckley T (2007) Species Trees from Gene Trees: Reconstructing Bayesian Posterior Distributions of a Species Phylogeny Using Estimated Gene Tree Distributions. *Systematic Biology*, **56**, 504–514.

Liu J-Q, Wang Y-J, Wang A-L, Hideaki O, Abbott RJ (2006) Radiation and diversification within the *Ligularia–Cremanthodium–Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution*, **38**, 31–49.

Lomolino MV, Riddle BR, Whittaker RJ, Brown JH (2010) Biogeography. Fourth Edition. Sinauer Associates, Sunderland.

Maddison WP, Wiens JJ (1997) Gene Trees in Species Trees. *Systematic Biology*, **46**, 523–536.

Marjoram P, Molitor J, Plagnol V, Tavare S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 15324–8.

Mayr E (1963) Animal Species and Evolution. Belknap Press of Harvard University Press, Cambridge.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, **66**, 526–538.

Meier JI, Sousa VC, Marques DA *et al.* (2017) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology*, **26**, 123–141.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, **17**, 240–8.

Morlon H, Potts MD, Plotkin JB, Francois O, Taubenberger JK (2010) Inferring the Dynamics of Diversification: A Coalescent Approach. *PLoS Biology*, **8**, e1000493.

Muller HJ (1939) REVERSIBILITY IN EVOLUTION CONSIDERED FROM THE STANDPOINT OF GENETICS. *Biological Reviews*, **14**, 261–280.

Muñoz-Pajares AJ, García C, Abdelaziz M *et al.* (2017) Drivers of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Molecular Ecology*, doi: 10.1111/mec.13971.

Myers N, Mittermeier RA (2003) Impact and Acceptance of the Hotspots Strategy: Response to Ovadia and to Brummitt and Lughadha. *Conservation Biology*, **17**, 1449–1450.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Nater A, Burri R, Kawakami T, Smeds L, Ellegren H (2015) Resolving evolutionary relationships in closely related species with whole-genome sequencing data. *Systematic Biology*, **64**, 1000–1017.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P (2012) Ecological Speciation. Oxford University Press, Oxford.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Nylander J, Olsson U, Alström P, Sanmartín I, Baker A (2008) Accounting for Phylogenetic Uncertainty in Biogeography: A Bayesian Approach to Dispersal-Vicariance Analysis of the *Thrushes* (Aves: Turdus). *Systematic Biology*, **57**, 257–268.

Owen LA (2009) Latest Pleistocene and Holocene glacier fluctuations in the Himalaya and Tibet. *Quaternary Science Reviews*, **28**, 2150–2164.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **5**, 568–583.

Pante E, Abdelkrim J, Viricel A *et al.* (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–9.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, **7**, e37135.

Ree RH, Smith SA, Baker A (2008) Maximum Likelihood Inference of Geographic Range Evolution by Dispersal, Local Extinction, and Cladogenesis. *Systematic Biology*, **57**, 4–14.

Renner SS (2005) Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science*, **10**, 550–558.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in forty years? *Evolution*, **47**, 1637–1653.

Richards J (2003) Primula. Second Ed. Portland, OR, Timber Press, USA.

Rieseberg LH (1997) Hybrid Origins of Plant Species. *Annual Review of Ecology and Systematics*, 359–389.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539–542.

Rowley DB, Currie BS (2006) Palaeo-altimetry of the late Eocene to Miocene Lunpola basin, central Tibet. *Nature*, **439**, 677–681.

Royden LH, Burchfiel BC, van der Hilst RD (2008) The Geological Evolution of the Tibetan Plateau. *Science*, **321**.

Rubin BER, Ree RH, Moreau CS (2012) Inferring Phylogenies from RAD Sequence Data. *PLoS ONE*, **7**, e33394.

Ruddiman WF, Kutzbach JE (1989) Forcing of late Cenozoic northern hemisphere climate by plateau uplift in southern Asia and the American west. *Journal of Geophysical Research*, **94**, 18409.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Sauquet H, Ho SYW, Gandolfo MA *et al.* (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.

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Serrano-Serrano ML, Rolland J, Clark JL, Salamin N, Perret M (2017) Hummingbird pollination enhanced the diversification of Gesneriaceae in the Neotropics. *Proceedings of the Royal Society B: Biological Sciences*. Accepted

Shi Y (2002) Characteristics of late Quaternary monsoonal glaciation on the Tibetan Plateau and in East Asia. *Quaternary International*, **97**, 79–91.

Shi YF, Li JJ, Li BY (1998) Uplift and Environmental Changes of Qinghai–Tibetan Plateau in the Late Cenozoic. Guangdong Science and Technology Press, Guangzhou.

Silvestro D, Schnitzler J, Liow LH, Antonelli A, Salamin N (2014) Bayesian Estimation of Speciation and Extinction from Incomplete Fossil Occurrence Data. *Systematic Biology*, **63**, 349–367.

Silvestro D, Schnitzler J, Zizka G (2011) A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology*, **11**, 311.

Soltis PS, Soltis DE (2009) The Role of Hybridization in Plant Speciation. *Annual Review of Plant Biology*, **60**, 561–588.

Spicer RA, Harris NBW, Widdowson M *et al.* (2003) Constant elevation of southern Tibet over the past 15 million years. *Nature*, **421**, 622–624.

Stadler T (2011) Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 6187–6192.

Sun Y, Abbott RJ, Li L *et al.* (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.

Taberlet P, Cheddadi R (2002) Quaternary Refugia and Persistence of Biodiversity. *Science*, **297**.

Tang Z, Wang Z, Zheng C, Fang J (2006) Biodiversity in China's mountains. *Frontiers in Ecology and the Environment*, **4**, 347–352.

Theodoridis S, Randin C, Broennimann O, Patsiou T, Conti E (2013) Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia* (R Pearson, Ed,). *Journal of Biogeography*, **40**, 1278–1289.

Theodoridis S, Randin C, Szövényi P, Boucher FC, Patsiou TS, Conti E (2016) How Do Cold-Adapted Plants Respond to Climatic Cycles? Interglacial Expansion Explains Current Distribution and Genomic Diversity in *Primula farinosa* L. in press. Systematic biology.

Valente LM, Reeves G, Schnitzler J *et al.* (2010) Diversification of the African genus *Protea* (Proteaceae) in the Cape biodiversity hotspot and beyond: equal rates in different biomes. *Evolution*, **64**, 745–760.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E (2014) Heterostyly accelerates diversification via reduced extinction in primroses. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 1–9.

Wagner CE, Keller I, Wittwer S *et al.* (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.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

Wang J, Street NR, Scofield DG, Ingvarsson PK (2016) Variation in Linked Selection and Recombination Drive Genomic Divergence during Allopatric Speciation of European and American Aspens. *Molecular Biology and Evolution*, **33**, 1754–1767.

Wen J, Zhang J-Q, Nie Z-L, Zhong Y, Sun H (2014) Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics*, **5**, 4.

Wright S (1943) Isolation by Distance. *Genetics*, **28**, 114–38.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Wu CY (1987) Flora of Tibet. Science Press, Beijing

Wu C-I, Ting C-T (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

Yao T, Pu J, Lu A, Wang Y, Yu W (2007) Recent Glacial Retreat and Its Impact on Hydrological Processes on the Tibetan Plateau, China, and Surrounding Regions. *Arctic, Antarctic, and Alpine Research*, **39**, 642–650.

Yu Y, Harris AJ, Blair C, He X (2015) RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**, 46–49.

Zheng B, Xu Q, Shen Y (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quaternary International*, **97**, 93–101.

Zheng D (1996) The system of physico-geographical regions of the Qinghai-Xizang (Tibet) plateau. *Science in China Series D-Earth Sciences*, **39**, 410–417.

Chapter 1

Phylogeny and biogeography of *Primula* sect. *Armerina*: implications for plant evolution under climate change and the uplift of the Qinghai-Tibet Plateau

From the article: **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. DOI: 10.1186/s12862-015-0445-7

Phylogeny and biogeography of *Primula* sect. *Armerina*: implications for plant evolution under climate change and the uplift of the Qinghai-Tibet Plateau

Guangpeng Ren[1,2,3], Elena Conti[4], Nicolas Salamin[1,2*]

[1] Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[2] Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland

[3] State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou 730000 Gansu, China

[4] Institute for Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland.

Abstract

Background: The historical orogenesis and associated climatic changes of mountain areas have been suggested to partly account for the occurrence of high levels of biodiversity and endemism. However, their effects on dispersal, differentiation and evolution of many groups of plants are still unknown. In this study, we examined the detailed diversification history of *Primula* sect. *Armerina*, and used biogeographic analysis and macro-evolutionary modeling to investigate a series of different questions concerning the evolution of the geographical and ecological distribution of the species in this section.

Results: We sequenced five chloroplast DNA and one nuclear gene for species of *Primula* sect. *Armerina*. Neither chloroplast nor nuclear trees support the monophyly of the section. The major incongruences between the two trees occur among closely related species and may be explained by hybridization. Our dating analyses based on the chloroplast dataset suggest that this section began to diverge from its relatives around 3.55 million years ago, largely coinciding with the last major uplift of the Qinghai-Tibet Plateau (QTP). Biogeographic analysis supports the origin of the section in the Himalayan Mountains and dispersal from the Himalayas to Northeastern QTP, Western QTP and Hengduan Mountains. Furthermore, evolutionary models of ecological niches show that the two *P. fasciculata* clades have significantly different climatic niche optima and rates of niche evolution, indicating niche evolution under climatic changes and further providing evidence for explaining their biogeographic patterns.

Conclusion: Our results support the hypothesis that geologic and climatic events play important roles in driving biological diversification of organisms in the QTP area. The Pliocene uplift of the QTP and following climatic changes most likely promoted both the inter- and intraspecific divergence of *Primula* sect. *Armerina.* This study also illustrates how niche evolution under climatic changes influences biogeographic patterns.

Keywords: *Primula*, incongruence, biogeography, niche evolution, Qinghai-Tibet Plateau

Background

Understanding the processes that shape geographical and ecological distribution of biodiversity is one of the most challenging questions in evolutionary biology and ecology. This is particularly true for regions that have experienced rapid habitat changes and harbor high species diversity. These characteristics are present in many mountainous areas and historical orogenesis has been proposed to play an important role in shaping their current biodiversity [1-3]. The alteration of topography and climatic changes associated with mountain uplifts can cause fragmentation of species distributions, thus limiting gene flow between isolated populations and initiating allopatric divergence and speciation [4-7]. However, extreme environmental changes and fragmented distributions can also lead to the extinction of lineages and species (*e.g.,* [8,9]). The processes occurring during mountain uplifts are therefore complex and we need to better understand the mechanisms that are at play during these events.

The fragmentation of species distributions can be due to the presence of limits on dispersal due, for example, to geographical barriers. Such limitations can induce a reduction in the movement of individuals into new locations and will result in distinct biogeographic patterns in the extant species [10]. However, fragmentation can also occur because of a lower success of establishment of individuals in some areas, which will limit the range of species [11]. This process is primarily set by ecological factors, potentially including both abiotic and biotic variables [10-12]. The dynamics of species range evolution will be constrained by phylogenetic niche conservatism, which is defined as the tendency of species to retain their ancestral ecological niche, thus shaping the geographic ranges of species over time (*e.g.,* [13,14]). However, evidence for rapid shifts in climatic preferences among species also exists [15,16] and macro-evolutionary modeling should be used to characterize the processes driving the evolution of ecological niches [17]. A complete assessment of these processes, coupled with detailed analyses of biogeographic patterns of species distribution, should then be used to help understand the distribution of species diversity [10].

One region that experienced drastic habitat changes and harbors extremely rich species diversity and endemism is the Qinghai-Tibet Plateau (QTP; [18]). While the start of its uplift dates from approximately 50 million years ago (Ma; [19]), the extensive uplifts of the QTP occurred in at least four periods since the early Miocene, specifically between 25-17 Ma, 15-13 Ma, 8-7 Ma, and 3.4-1.6 Ma [9,20-23]. At present, the QTP, with an average altitude of more than 4000 m (a.s.l.), is the highest and one of the most extensive plateaus on Earth [20]. About 9,000 to 12,000 species of vascular plants in ca. 1,500 genera are present in this plateau, and at least 20% of these species and ca. 50 genera are endemic [3,18]. The historical sequence of uplifts of the QTP has been suggested to partly account for the occurrence of high levels of biodiversity and endemism in the region [24]. However, the potential effects of climatic changes during the Quaternary on the diversification and distribution of many

groups of plant species in the QTP are not very well known (see review [2,3,25]).

Primula L. (Primulaceae) is one of the genera that exhibit high levels of species diversity in the QTP. The group, with a predominantly northern hemisphere distribution, contains ca. 500 species. About 60% of the species are present in the QTP and its adjacent regions [26,27]. Although this genus represents an important floristic element of alpine meadows in the region, it remains unclear whether the uplift of the QTP and the following climatic changes affected its diversification and distribution. In this context, a better understanding of the historical biogeography of key floristic elements of the region is an important way to illuminate the evolutionary history of these organisms in space and time. Available studies mainly utilize genus- or family-level phylogenies to elucidate the biogeographic connections between the QTP and neighboring regions [28-32]. However, the presence of a single sample per species hardly provides insights into the biogeographic patterns of species distributions within the QTP. Therefore, sampling multiple individuals per species and focusing on endemic species may help to better understand the mechanisms that were responsible for biogeographic patterns within the QTP.

In this study, we include several samples per species to investigate the historical biogeography of *Primula* sect. *Armerina* Lindley (Primulaceae), which exhibits a typical Sino-Himalayas distribution. According to the most recent global monographic treatment of the genus, *Primula* sect. *Armerina* comprises 14 species [26]. Eight species (*P. fasciculata, P. tibetica, P. conspersa, P. gemmifera, P. zambalensis, P. pumilio. P. pamirica* and *P. involucrata*; Figure 1) are endemic to the QTP, with different geographic distributions [26,27]. Among them, there has been some confusion between *P. tibetica* and *P. fasciculata* because of their morphological similarities at high altitude ([26,27]; field observation). The two species can be easily distinguished when bracts are present. *Primula tibetica* has oblong and pouched bracts, while the bracts of *P. fasciculata* are linear and non-pouched (Figure 1A, D). However, at high altitude, bracts are usually missing in *P. fasciculata* (Figure 1B, C), while in *P. tibetica*, they can also be absent in small individuals with single flower (Figure 1E, F). Both species have wide altitude distributions, ranging from 2900 m to 5000 m [26,27] and the use of molecular data combined with macro-evolutionary modeling may provide useful insights into the dynamics of their range evolution. The four remaining species of this section (*P. iljinskyii, P. chrysostoma, P. knorringiana* and *P. valentinae*) have very restricted areas in regions adjacent to the QTP. *Primula nutans* has the most widespread distribution in the section, including N Europe, W & E Siberia, NW America to N Mongolia, NW China and NW QTP. All species from sect. *Armerina* are considered to be diploid (2n = 18, 20 or 22) [26,27], except *P. egaliksensis*, which is the only tetraploid species (2n = 36, 40) and occurs mainly in North America. It was assigned to sect. *Armerina* based on morphological features [33,34], and might be of hybrid origin between *P. mistassinica* (sect. *Aleuritia*) and *P. nutans* [35-37].

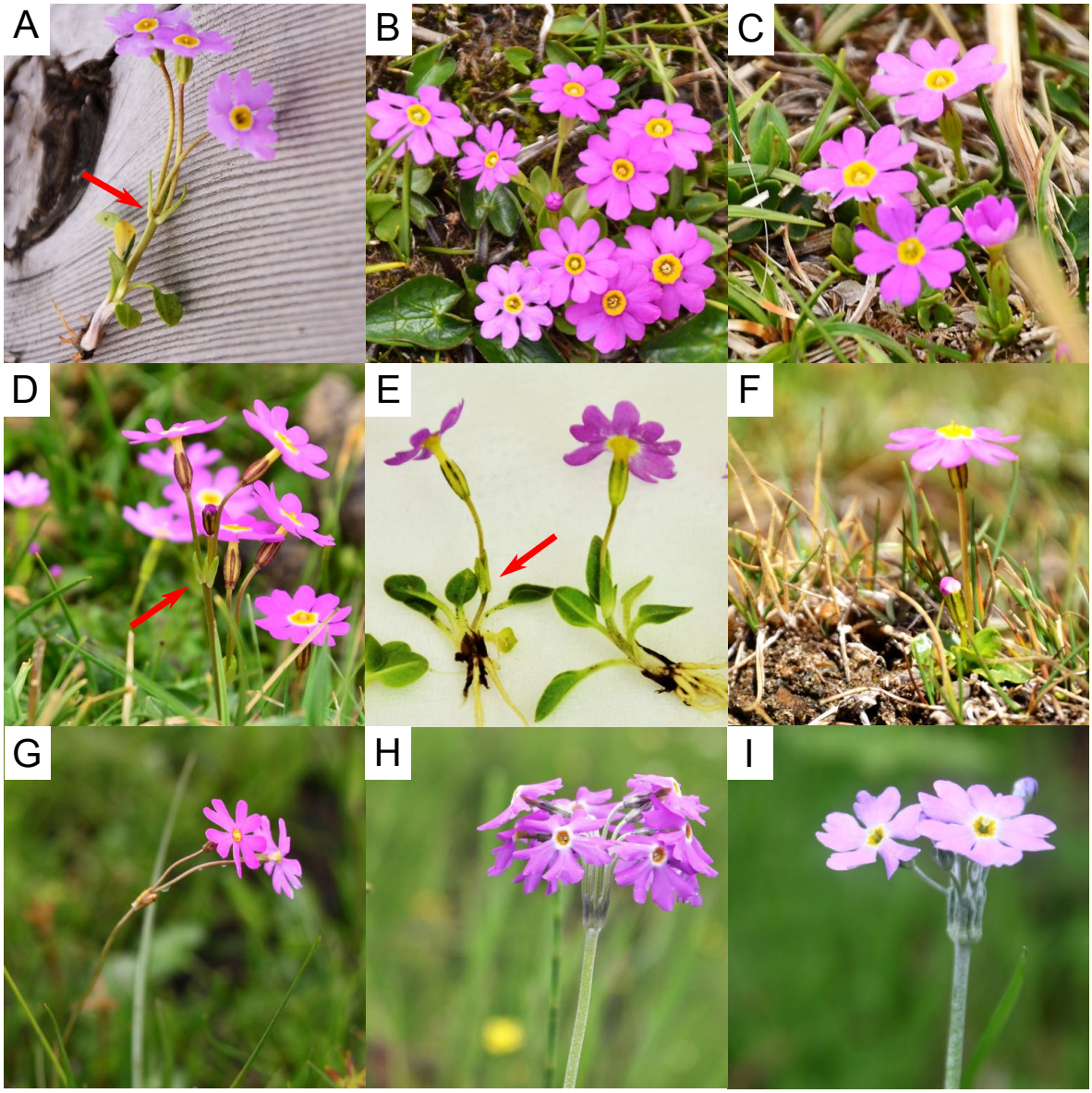


Figure 1. The five species of sect. *Armerina* which showed mainly incongruence between the two trees. (A) *P. fasciculata* with linear and non-pouched bracts, (B) *P. fasciculata* without bracts, (C) one photo of *P. fasciculata* collected from populations of clade F2 (see Results), (D) *P. tibetica* with oblong and pouched bracts at low altitude, (E) and (F) *P. tibetica* with and without bracts at high altitude, respectively, (G) *P. nutans*, (H) *P. gemmifera*, (I) *P. conspersa.* Bracts for *P. fasciculata* and *P. tibetica* are indicated by red arrows. All photos were taken by the first author in the field.

Most species of the *Armerina* section are thus prominent floristic elements of alpine meadows at high altitudes in the QTP and most are endemic to the QTP and its adjacent regions [26,27]. This section of *Primula* hence represents a good candidate to assess the biogeographic history of the QTP and to understand the effects of its uplift and associated climatic changes on the geographical distribution of biodiversity. We analyzed both nuclear and chloroplast DNA sequences of multiple samples per species in the *Armerina* section to reconstruct a comprehensive phylogenetic tree of this group. The aims of our study are to: i) test the inter-specific relationships of sect. *Armerina* to obtain a detailed

and resolved phylogenetic tree for the section; ii) assess whether the diversification of this section was influenced by the uplifts of the QTP; iii) combine biogeographic analyses with macro-evolutionary modeling of ecological niches to better understand the role of dispersal and ecological constraints during the diversification of the three main species in the section (*P. fasciculata, P. nutans* and *P. tibetica*).

Results

Sequence characteristics

Five chloroplast (*mat*K, *rpl*16, *rps*16, *trn*LF and *trn*H-*psb*A) and one nuclear (translin family protein, *tfp*) markers were sequenced in this study for phylogenetic analyses. The *matK* dataset used for phylogenetic analyses comprised 892 characters, 815 of which were constant, 22 variable but parsimony-uninformative, 55 variable and parsimony-informative. The *rpl*16 dataset comprised 1063 characters, 903 of which were constant, 90 variable but parsimony-uninformative, 70 variable and parsimony-informative. The *rps*16 dataset comprised 877 characters, 789 of which were constant, 24 variable but parsimony-uninformative, 64 variable and parsimony-informative. The *trn*LF dataset comprised 968 characters, 840 of which were constant, 54 variable but parsimony-uninformative, 74 variable and parsimony-informative. The *trn*H-*psb*A dataset comprised 629 characters, 512 of which were constant, 46 variable but parsimony-uninformative, 71 variable and parsimony-informative. We combined the five plastid regions for all subsequent analyses, modeling them as five partitions. It was not possible to obtain these sequences for *P. watsonii* and four chloroplast sequences (*mat*K-DQ378314, *rpl*16-DQ378443, *rps*16-FJ786584 and *trn*LF-FJ794215) were downloaded from GenBank for this species.

The aligned nuclear dataset comprised 648 characters, 445 of which were constant, 91 variable but parsimony-uninformative, and 112 variable and parsimony-informative. Despite repeated attempts, the *tfp* sequences for the three samples of *P. tibetica*, as well as the sample of *P. pamirica*, *P. pumilio* and two outgroup species (*P. watsonii* and *P. pinnatifida*) failed to amplify. Two copies were identified in the samples of *P. fasciculata*, *P. conspersa* and *P. egaliksensis* and these clones were added to the sequences obtained directly from PCR in subsequent phylogenetic analyses.

Phylogenetic analyses and molecular dating

The Maximum Likelihood (ML) and Bayesian analyses done on each data set resulted in congruent topologies, but discrepancies were obtained between the two types of markers. The only tetraploid species, *P. egaliksensis,* was included in a well-supported clade with *P. mistassinica* and *P. farinosa* in the chloroplast tree. This result is in agreement with previous studies [35,37,38]. The node subtending the rest of the samples of *Primula* sect. *Armerina* received very low support (posterior probability, PP 0.18, ML 6%) in the choloroplast phylogenetic tree and the relationships between species remained

partly unresolved (Figure 2). Three main clades were inferred in the chloroplast tree. The clade *involucrata* (including *P. involucrata, P. pamirica, P. fasciculata, P. nutans* and *P. tibetica*) and the clade *conspersa* (including *P. conspersa, P. gemmifera and P. zambalensis*) were strongly supported in both ML and Bayesian analyses, while the clade *pumilio* (*P. pumilio*) was not well-supported by ML (74%), but received very high posterior probabilities in the Bayesian analyses (PP 1.0). Overall, well-supported clades (PP > 0.95) in the chloroplast tree grouped sequences from the same species, except for *P. fasciculata*, which was separated into two groups (Figure 2).

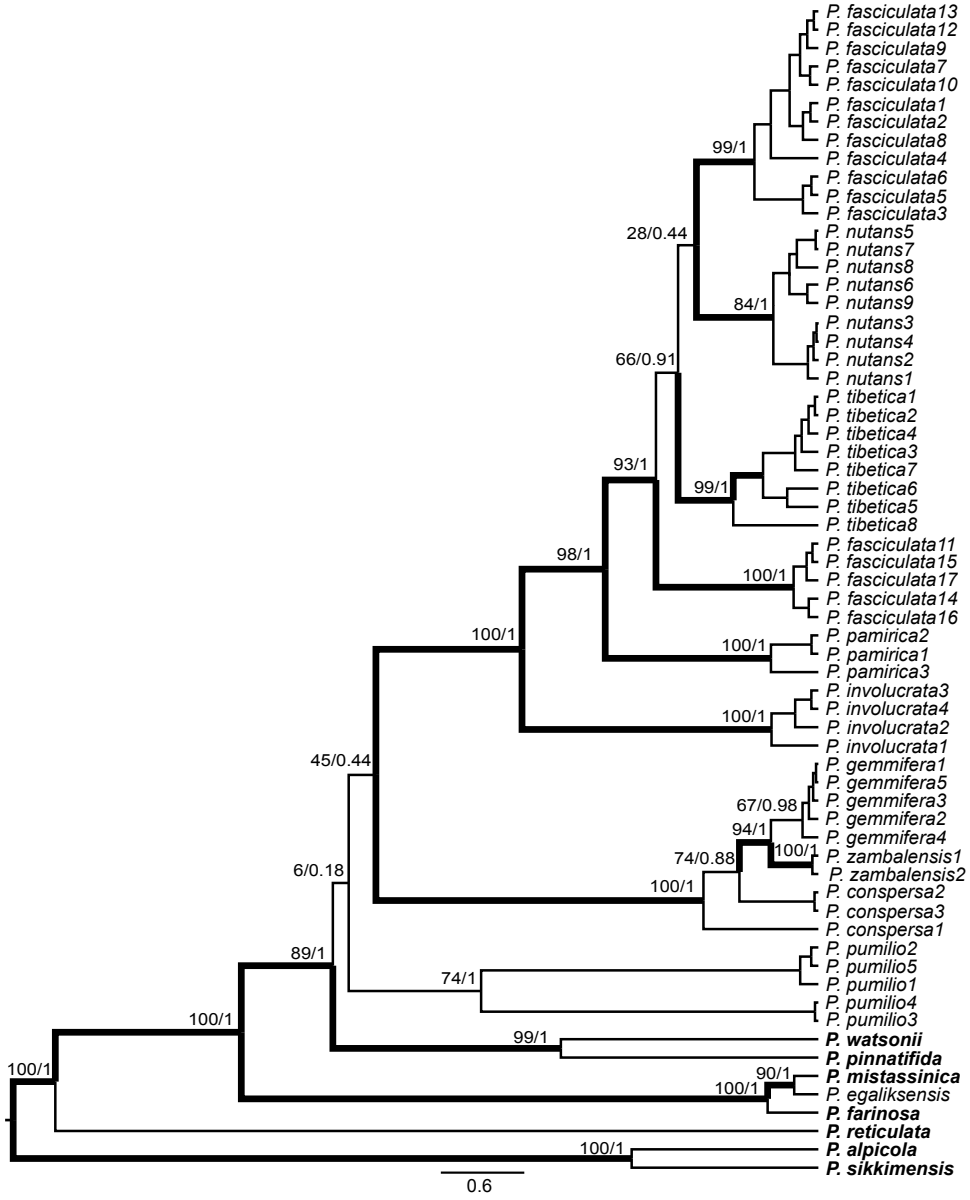


Figure 2. The maximum clade credibility (MCC) tree derived from BEAST analyses of five chloroplast genes. Maximum likelihood (ML) bootstrap values and Bayesian posterior probabilities (PP) are indicated at major nodes. Bootstrap values ≥ 80 and PP ≥ 0.95 are indicated with thicker branches. Outgroup species are shown in bold.

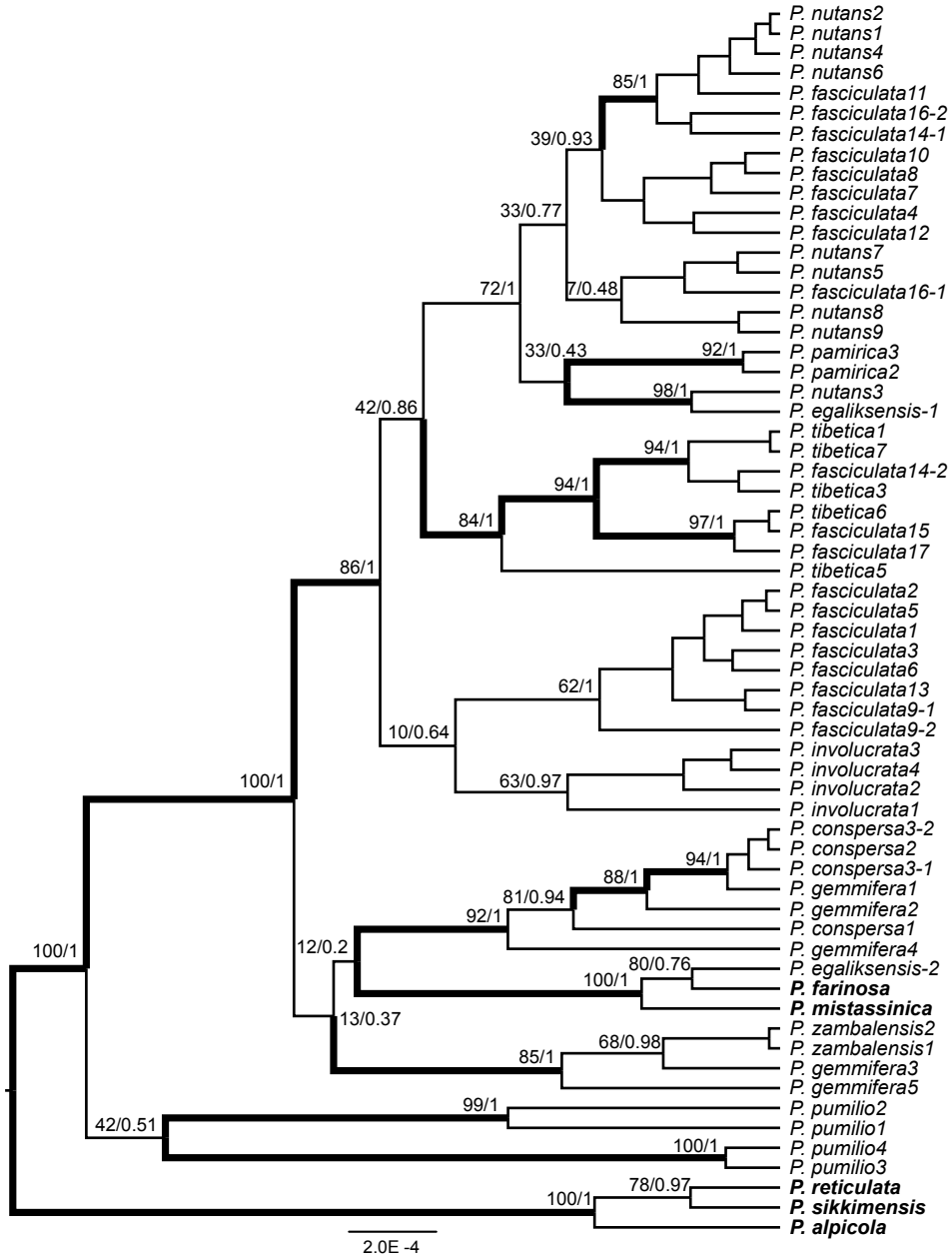


Figure 3. The maximum clade credibility (MCC) tree derived from MrBayes analyses of the nuclear dataset. Maximum likelihood (ML) bootstrap values and Bayesian posterior probabilities (PP) are indicated at major nodes. Bootstrap values ≥ 80 and PP ≥ 0.95 are indicated with thicker branches. Outgroup species are shown in bold. Two nuclear gene copies for some samples are indicated with "-1" or "-2".

In contrast to the plastid dataset, *Primula* sect. *Armerina* and two nested outgroup species received very high node support (PP 1.0, ML 100%) in the nrDNA phylogenetic tree, but the relationships between species were less well supported (Figure 3). Three main clades within the section identified in the chloroplast tree were also inferred in the nuclear tree (Figure 3). The clade *involucrata* was well-supported (PP 1.0, ML 86%), while the clades *conspersa* (except for *P. farinosa, P. mistassinica* and *P. egaliksensis*) and *pumilio* received very weak nodal support in both types of analyses. The relationships within each clade were further incongruent between the trees obtained by the two

datasets. *Primula fasciculata* was divided into three clades in the nrDNA tree (Figure 3). One clade included samples from *P. fasciculata* that cluster with a moderately supported clade representing *P. involucrata*. A second clade included all samples of *P. tibetica* and *P. fasciculata* and one copy of *P. fasciculata*. Finally, the third clade included all samples of *P. nutans* and *P. pamirica*, one copy of *P. egaliksensis* and the remaining samples of *P. fasciculata* (Figure 3). Similarly, *P. gemmifera* separated into two groups, either with *P. zambalensis* or in a clade including all samples of *P. conspersa* (Figure 3). Two copies of *P. egaliksensis* were clustered with either *P. nutans* or *P. mistassinica*, corroborating the hypothesis of the allopolyploid origin of this species [35-37].

Previous dating analyses at the level of the family used low intra-sectional sampling and suggested that sect. *Armerina* diverged from its relatives about 5 Ma [38]. This date is generally congruent with the results of our dating analysis, which indicated that the section (except *P. egaliksensis*) diverged from its two relatives, *P. watsonii* and *P. pinnatifida,* 3.55 Ma (1.76-5.93 Ma, 95% highest probability density, HPD; Fig. 3). Most cladogenetic events in this section occurred during the past 3.4 million years (Figure 4). The crown age of the three closely related species, *P. nutans*, *P. fasciculata* and *P. tibetica*, was about 1.19 Ma (95% HPD: 0.51-2.13 Ma; Figure 4).

Biogeographic inference

Biogeographic analysis based on the chloroplast dataset was reconstructed by Statistical Dispersal–Vicariance Analysis (S-DIVA). Fourteen dispersal and 15 vicariance events for the section were identified in this analysis (Figure 4). The origin of this section was inferred with high confidence in the Himalayan Mountains (B, 91%). We found that one clade (*P. zambalensis*, *P. gemmifera* and *P. conspersa*) colonized the Northeast QTP (C) and subsequently diversified and dispersed to the Hengduan Mountains (A), while *P. pamirica* colonized the Mountains of Central Asia (D). The common ancestral area of *P. fasciculata*, *P. tibetica* and *P. nutans* was inferred to be in the Himalayan Mountains (B, 86%).

Evolution of ecological preferences

We fitted a series of macro-evolutionary models based on 19 bioclimatic variables (*i.e.,* climatic niches) to better understand the biogeographic patterns of three closely related species, *P. fasciculata*, *P. tibetica* and *P. nutans*. We extracted the 19 bioclimatic variables from the sampled localities of the three species (Additional file 3). For *P. nutans*, we used only the samples that were collected in the QTP. The first two axes of the principal component (PC) analysis based on this dataset explained 53.2% and 25.3% of variance, respectively. The first axis (PC1) was strongly and positively correlated with temperature seasonality (BIO4, WorldClim variables) and negatively correlated with temperature in coldest and driest Quarter (BIO6 and BIO9). The second axis (PC2) was correlated strongly and positively with precipitation in coldest and driest Quarter (BIO14, BIO17 and BIO19), and strongly

and negatively with precipitation seasonality and mean diurnal range (BIO2 and BIO15).

We used the values obtained for PC1 and PC2 (Additional file 4) to test for the evolution of the ecological niche in *P. fasciculata*, *P. tibetica* and *P. nutans*. The Brownian motion model was rejected for both PC1 and PC2 in all species sets tested (Additional file 5). For PC1, the best-performing models were OU1 for SET1, SET2 and SET3, and OUMV for SET4. Average AICc weights were 0.46, 0.36, 0.66 and 0.48, respectively (see Additional file 5 for all AICc weights). The OUM was the second-best model for SET1 (Average AICc weights = 0.25). The OUMV, OUMA and OUM models that allow different niche optima for SET2 also received non-negligible AICc weights (0.29, 0.18, 0.12). For PC2, all four sets were best modeled under OUMV (AICc weights 0.97, 0.93, 0.78 and 0.64 respectively; Additional file 5).

The parameters (niche optimum θ, rate of niche evolution σ^2 and strength of selection α) estimated for the three species groups (F1, F2 and NT) from all supported models based on the four group sets were congruent (Additional file 6) and we showed the parameters estimated based on SET2 (Figure 5). We used model averaging to estimate the parameter values for PC1 over the supported models OUMV, OUMA and OUM. The averaged niche optima (θ) across models for group F1, F2 and NT were -0.17, -2.0 and 0.55, respectively (Figure 5). The averaged rate parameter (σ^2) across models for group F2 was two times slower than that for the groups F1 and NT (59 vs. 131 and 112). Finally, the averaged strength of selection estimated across models for the three groups was similar (6.9, 6.3, 6.9). For PC2, model OUMV, which allows for different niche optima and rates of niche evolution among groups, was the only supported model. The optimum values estimated based on this model for the three groups were also different from each other (F1: 0.2, F2: -0.99, NT: -0.33). The group F2 still exhibited the slowest rate of niche evolution (F1: 228, F2: 94, NT: 1723; Figure 5).

Discussion

Non-monophyly of *Primula* sect. *Armerina*

The phylogenetic analyses of *Primula* sect. *Armerina* presented here contain samples of several individuals per species and cover most of the geographic distributions of the species. Neither the chloroplast tree nor the nuclear tree supports the monophyly of sect. *Armerina*. The section and the two outgroup species, *P. watsonii* and *P. pinnatifida* (sect. *Muscarioides*) form a well-supported clade in the chloroplast tree, despite the fact that these two outgroup species are distinguished from sect. *Armerina* by clear morphological traits (*e.g.,* spicate inflorescence *vs.* umbel; [26,27]). Similarly, the two outgroup species, *P. farinosa* and *P. mistassinica* (sect. *Aleuritia*), are grouped with the section and form a well-supported clade in the nrDNA tree. The non-monophyly of sect. *Armerina* is in agreement with previous family-wide analyses [38,39]. Moreover, non-monophyly of sections in genus *Primula* seems pervasive in phylogenetic trees [38,39].

Phylogenetic relationships within the section

The relationships among some of the basal nodes of the section in the nuclear tree are uncertain (Figure 3), which may result from low sequence divergence within the section. The use of a single nuclear gene is thus clearly not sufficient to resolve the relationships within the group, which is a pattern often found also in other lineages (*e.g.,* [40,41]). Multiple nuclear genes or genomic data are therefore needed to resolve the precise relationships between the main clades in this group. However, both phylogenetic trees show three main clades within sect. *Armerina*, which is in agreement with previous phylogenetic studies [39,39] as well as morphological based taxonomy [26,27].

Phylogenetic relationships inferred from the nuclear and chloroplast datasets were incongruent (Figure 2, 3). The tree obtained from the latter is in agreement with morphology-based taxonomy, which contrasts with other studies that showed a better congruence of taxonomy with the trees inferred from nuclear datasets (*e.g.,* [42]). Incongruence between different plant genomic markers is found in numerous studies and can be explained by incomplete lineage sorting, hybridization and introgression [40,42-45]. Introgression represents the transfer of genes between species mediated primarily by backcrossing [46], but it does not seem a likely explanation for the incongruence that we observed. Maternally inherited chloroplast loci with relatively low rates of intraspecific gene flow should be more frequently introgressed [46]. In contrast, biparentally inherited nuclear loci that experience high rates of intraspecific gene flow should enhance species delimitation [46]. We find the opposite pattern in our results (Figure 2 and 3). The chloroplast tree has much clearer species delimitation than the nuclear tree and this pattern seems incompatible with the assumption that the incongruence results from introgression.

Although introgression cannot occur without hybridization, hybridization followed by no backcrossing and introgression could still occur and such phenomenon has been detected in numerous studies (*e.g.,* [47,48]). Natural hybridization in *Primula* is common and has been confirmed by several studies [37,49-51], although, it is currently unclear to what degree species within sect. *Armerina* hybridize with each other. The incongruent placement of *P. egaliksensis* between chloroplast and nuclear gene trees can be explained by hybridization [35-37]. Moreover, our results provide further evidence in support of the hypothesis that *P. egaliksensis* originated from an intersectional allopolyploidization event, which places the two *tfp* copies of *P. egaliksensis* with either *P. nutans* or *P. mistassinica* (Figure 3), confirming previous results by Guggisberg et al. [35-37]. From this perspective, similar incongruence detected for sample number 14 of *P. fasciculata* (two *tfp* copies grouped with either *P. nutans* or *P. tibetica*) may also result from hybridization. Beside hybridization, incomplete lineage sorting is another important explanation for the incongruence between data sets, but the two processes are often difficult to distinguish from each other [52-54]. Although incomplete lineage sorting could

also be involved in the incongruences found in our results, the occurrence of such a process would imply that the origin of the haplotypes of *P. pamirica* preceded the speciation events of the whole clade [52]. Such extensive levels of incomplete lineage sorting may yield gene trees with random patterns of relationships among taxa [55]. The patterns of relationships are, however, non-random in the nuclear tree. The major incongruences result mainly from the division of *P. fasciculata* and *P. gemmifera* in different lineages. We thus consider hybridization as the most likely explanation for the major incongruence between chloroplast and nuclear trees.

However, it should be noted that using a single nuclear gene that provides low resolution of the phylogenetic relationships might not be sufficient to elucidate the reasons of the genealogical incongruences between different genomic markers. Although our results tend to suggest a more probable role of hybridization as the most likely explanation for the major incongruence between two trees, incomplete lineage sorting and introgression cannot be completely excluded. We therefore recognize that gene trees/species trees analyses involving multiple nuclear loci or population genomic approaches would be necessary to clearly discriminate among these possible scenarios.

Biogeographic history

The biogeographic reconstruction based on the chloroplast dataset showed that sect. *Armerina* originated in the Himalayas and subsequently dispersed to the Hengduan Mountains, Northeastern QTP and Western QTP (Figure 4). The lineages involved in these dispersal events further diversified in the Hengduan Mountains, Northeastern QTP and Western QTP, respectively, and gave rise to several of the extant species. Our dating analysis estimates that this section diverged from its closest relatives in the Pliocene about 3.55 Ma (95% HPD: 1.76-5.93 Ma, Figure 4). The timeframe of this event coincides with the recent uplift of the QTP, which occurred between 1.6-3.4 Ma [21,22,56]. A similar time of divergence was also observed in other groups of plants distributed in the QTP [32,57-59]. It has been suggested that the uplifts of the QTP might have limited the spread of many species, but accelerated speciation via vicariance [60]. The timeframe of the uplift also coincides with a period of high climatic oscillations that could have reinforced the processes initiated by the uplifts [57,58].

Vicariance and dispersal triggered by the uplift of the QTP and associated climatic changes are common mechanisms in the diversification of plants in the QTP (*e.g.,* [59,61]), and also in other mountain areas (*e.g.,* [1,62,63]). Based on the S-DIVA analysis, five of the 15 vicariance and seven of the 14 dispersal events account for cladogenetic events, and both events occurred during and after the Pliocene uplift of the QTP (Figure 4). Vicariance and dispersal triggered by the uplift of the QTP and Quaternary climatic oscillations may accelerate the early diversification of sect. *Armerina*, and further shape the biogeographic patterns [59]. Furthermore, ten "vicariance" (for ease of notation, here we still keep the word "vicariance" for the isolation of populations of the same species) and seven

dispersal events are identified within species-specific clades, which might play a role in promoting intraspecific divergence. Extensive inter- and intra-specific divergence took place in the QTP within the Pliocene and Quaternary climatic changes in many groups of plants (*e.g.,* [64-68]). Our analyses together with previous studies thus highlight the importance of the Pliocene uplift of the QTP and Quaternary climatic changes in promoting the diversification of plants in this mountain area.

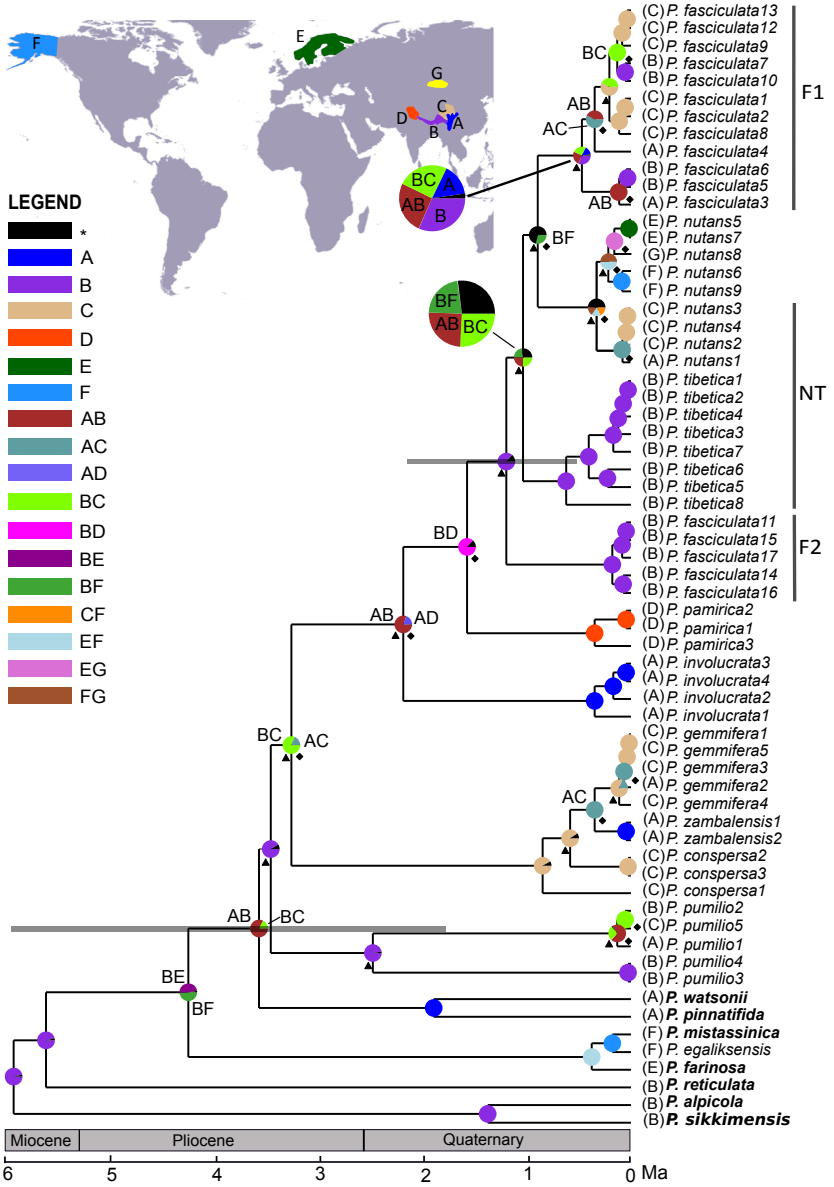


Figure 4. Dispersal–vicariance scenarios for sect. *Armerina* and the outgroup speices based on the chloroplast dataset reconstructed by Statistical Dispersal–Vicariance Analysis (S-DIVA) optimization with the maximum number of area units set to two. Triangle: dispersal event; diamond: vicariance event. Letters denoting area units are indicated on the map. Pie charts at internal nodes represent the marginal probabilities for each alternative ancestral area. Alternative ancestral areas (letters on nodes) are indicated for the major nodes. The grey bars on the nodes represent the 95% highest posterior density intervals of the dates obtained from BEAST analyses. Time scale is shown at the bottom. Three groups (F1, F2 and NT) are used for the evolutionary niche models: groups F1 and F2 are two clades of *P. fasciculata* in the chloroplast tree; group NT includes all samples of *P. tibetica* and samples of *P. nutans* that were only collected from the QTP.

Niche evolution of *P. fasciculata*

The S-DIVA analysis shows different biogeographic patterns for the two *P. fasciculata* clades. One clade (F2; Figure 4) occupies only Northern Tibet, while samples from the other clade (F1) can be found in the Hengduan Mountains, Eastern Tibet and Northeastern QTP. Wiens & Donoghue [69] argued that phylogenetic niche conservatism and niche evolution might be critical in the biogeographic history of many groups. In contrast to most previous studies that have suggested the importance of niche conservatism in setting range limits and creating biogeographic patterns (*e.g.,* [70,71]), niche evolution under climatic changes seems to be the major factor explaining the biogeographic patterns detected here. Although the OU1 model that allows a single niche optimum is the best model along the temperature gradient (PC1), models that allow different niche optima received together higher AICc weights ($AICc_{OUM+OUMV+OUMA}$=0.59). This result suggests that ecological differentiation (*i.e.,* different niche optima) is occurring in this group.

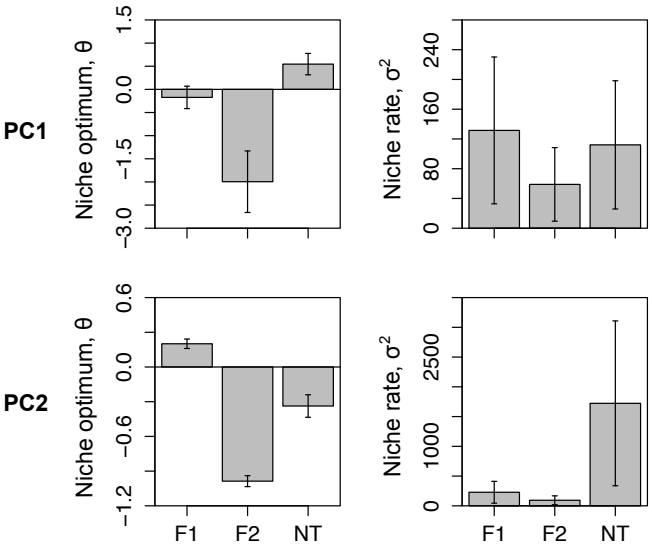


Figure 5. Parameter estimates of models of niche evolution for the three groups (F1, F2 and NT). For PC1, averaged parameters are obtained based on three supported models (OUM, OUMV and OUMA). The averaged strength of selection (α) estimated across models for the three groups is similar and not shown. For PC2, parameter estimates are from the only supported OUMV model (different rates σ^2 and niche optima θ among the three groups).

The two *P. fasciculata* clades and their two closely related species are estimated to diverge from each other during the Quaternary after the uplift of the QTP (Figure 4). Climatic oscillations during the Quaternary had a dramatic effect on species distribution ranges [72]. Many species have repeatedly retreated and expanded their distributions following these climatic oscillations (*e.g.,* [57,58,72,73). In the context of a changing environment, dispersal plays a crucial role in tracking favorable environmental conditions through space [74]. It can also help adaptation of small populations through both demographic and genetic rescue effects [75,76]. Two dispersal events may have provided the opportunities for populations of clade F1 to occupy wide ranges and also invade new habitats and climatic regimes (Figure 4). These events are associated with relatively relaxed niches (*i.e.,* niche optima are not strongly correlated with temperature and precipitation gradient; Figure 5) and fast niche evolution (Figure 5) and these characteristics might have allowed these populations to adapt to the changing environmental conditions [11,77]. In contrast, populations of clade F2 occur at higher

altitudes (average 4600 m) compared to those of clade F1 (average 4200 m). These populations of clade F2 might have been adapted to a colder climate characterized by lower temperature seasonality (*i.e.,* cooler summers) and less precipitation in the coldest Quarter (see results of PCA and Figure 5). Clade F2 displays a lower rate of niche evolution than F1 populations and this lower rate could have limited its dispersal into lower and warmer places. A similar pattern was observed in tropical treefrogs [78], which were unable to extend their ranges further North into temperate regions. Furthermore, recent climatic changes are involved in a shift toward higher elevations in the climatic envelopes of two closely related monkey-flower species in the direction of higher elevations [79]. However, given the harsh environmental conditions, it is plausible that climatic warming in the future might adversely affect the populations of the clade F2 and cause their distributions to shrink [57]. Our results also indicate that contrasting evolutionary processes can occur within closely related lineages, reinforcing the idea that phylogenetic niche conservatism is unlikely to hold at lower spatial scales [80].

While we focus on climatic variables (*i.e.,* temperature and precipitation) to explain the biogeographic patterns detected here, additional ecological factors such as edaphic variables, competition, seed bank and seed number could be involved in creating biogeographic patterns [10,12]. As argued by Hoskin et al. [81], geographic isolation of populations within species and variation in ecological factors are major precursors to cryptic speciation. The ecological differences and biogeographic patterns found between the two *P. fasciculata* clades may have given rise to some degree of differential adaptation to their respective environmental conditions, as also suggested in *Taxus wallochiana* [58]. However, our data is not appropriate to gain a detailed knowledge of the processes at play here and further studies involving a finer sampling of populations associated with large scale genomic data should be employed to better understand the mechanisms involved in the separation of the *P. fasciculata* clades.

Conclusion

Our phylogenetic analyses, based on both chloroplast and nuclear datasets, show non-monophyly of *Primula* sect. *Armerina*, corroborating the results of previous family-level studies [38,39]. The topologies inferred from nuclear gene and concatenated chloroplast datasets are incongruent, which may mainly result from hybridization. This section was suggested to originate in the Himalayas during the Pliocene uplift of the QTP. Subsequent dispersals to the Hengduan Mountains, Northeastern QTP and Western QTP were considered as the consequence of the Pliocene uplift of the QTP and following climatic changes. We further provide a practicable framework for the first time to test the relationship between biogeographic patterns and ecological factors in the QTP area. Our evolutionary models suggest that niche evolution, rather than niche conservatism, seems to explain the biogeographic patterns of the two *P. fasciculata* clades.

Methods

Sampling and extraction

We collected in total 57 samples representing 10 of the 14 species belonging to *Primula* sect. *Armerina* (Additional file 1). We could not obtain plant material for *P. iljinskyii, P. chrysostoma, P. knorringiana* and *P. valentinae*, which have small distributions in Central Asian Mountains and are difficult to obtain due to their geographical locations. Widespread species were collected from different localities across their geographical ranges. For example, *P. nutans* was represented by two samples from N America, two from N Europe, one from NW Mongolia and four from China. Seven outgroup species were sampled based on the large phylogenetic tree of Primulaceae (Additional file 1; [38]). All samples were dried and stored in silica gel after collection, except for *P. pamirica*, which was obtained from Harvard University herbaria. 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 AG, Hombrechtikon, Switzerland) following the manufacturer's instructions.

Amplification and sequencing

Five chloroplast DNA regions and one nuclear gene were sequenced. Three cpDNA loci (*rpl*16 intron; *trn*L-F region, which comprises the *trn*L intron and the *trn*L-*trn*F intergenic spacer; *rps*16 intron) were amplified and sequenced using the published primers [34]. The *mat*K gene and *trn*H-*psb*A intergenic spacers were amplified and sequenced following the protocol described in Li et al. [82]. For the nuclear gene, we designed three pairs of exon-primed-intron-crossing (EPIC) primers based on an *Arabidopsis thaliana* translin family protein locus (*tfp*, AT2G03780) and a *Primula sieboldii* seedling cDNA library (FS228429). Only one pair of primers: *tfp*_e1.F (5'-CGAGAAAGGGTGGTAAAAGC-3') and *tfp*_e1.R (5'-CTGGGGAGTAAGCTCGTCTG-3'), was amplified successfully for sect. *Armerina*. Polymerase chain reactions (PCR) generated double bands and direct sequencing of *tfp*_e1 amplicons produced electropherograms with double peaks and non-complementarity between sequenced strands in the following accessions: *P. fasciculata* (populations 9, 14, 16), *P. conspersa* (population 3) and *P. egaliksensis.* These PCR products were applied on a 1.5% agarose gel, then excised and purified using a QIAquick Gel Extraction Kit (Qiagen, cat. no. 28704). The purified products were subsequently cloned into a pTZ57R/T vector and sequenced. Eight clones were sequenced per band.

All PCR reactions were performed in 25 μL volumes containing 1 × buffer (including 1.5 mM $MgCl_2$), 2 mM $MgCl_2$, 300 μM dNTPs, 0.2 μM of each primer and one unit Taq polymerase (GoTag DNA Polymerase, Promega, Madison, WI, USA). Amplifications were carried out on a thermocycler (Biometra, Goettingen, Germany) using the following conditions: a first cycle at 94°C for 3 min; 36 cycles at 94°C for 40 s, 55 °C for 1 min and 72°C for 1.2 min; a final cycle of 7 min at 72°C. All sequencing reactions used the Big Dye 3.1 Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA, USA), then sequenced on an ABI Prism 3100 genetic analyzer (Applied Biosystems).

DNA sequences were aligned with Geneious 6.1.6 (Biomatters) using MAFFT [83] and revised manually. The nuclear gene data generated from direct sequencing were scanned carefully and edited when necessary to ensure that all double peaks were identified correctly with standard degeneracy codes (*e.g.,* Y means C or T; R means G or A; W means A or T; K means G or T; M means C or A). When double peaks were detected at a site, the site was ascertained as ambiguous only if the weakest signal reached at least 25% of the peak signal strength [84,85]. For individuals that contained multiple clones for the *tfp* gene, we randomly chose a single representative sequence for the phylogenetic analysis if all the clones formed a well-supported clade in a preliminary analysis, while multiple sequences were retained otherwise. All sequences were submitted to GenBank (accessions KT259477-KT259852).

Phylogenetic reconstruction and molecular dating

The five chloroplastic genes were concatenated into a single dataset using SequenceMatrix 1.7.8 [86], while the chloroplast and nuclear datasets were analyzed separately. The GTR+G model of sequence evolution was selected on the basis of the Akaike information criterion (AIC) for all DNA regions as estimated by jModelTest 2.1.4 [87]. Maximum likelihood analyses were done with PhyML (ver. 3.0; [88]) using the BEST algorithm for branch swapping and 10^3 bootstrap replicates to assess node support. We estimated tree topology by Bayesian inference using MrBayes 3.2 [89] with the GTR+G model of evolution and default priors. We unlinked the parameters of the GTR+G model between the five different genes for the analysis of the chloroplast dataset. We repeated the MrBayes analyses three times for each analysis (*i.e.,* chloroplast and nuclear dataset) and each analysis consisted of four chains of 10^7 generations, sampling every 10^3 steps with temperature parameter set to 0.1. We determined convergence by examining trace plots of the log-likelihood values for each parameter in Tracer 1.5.

We used the chloroplast dataset for dating analysis with a secondary calibration strategy, as described in de Vos et al. [38], However, age estimation obtained from this kind of calibration may be inherently subjected to bias and errors [90]. We addressed this concern by comparing our estimated age with previously published ones, but it should be noted that they are estimates that should be treated with caution. Divergence time analysis was performed in BEAST (ver. 1.7; [91]). The fossil record of Primulaceae is too sparse to provide multiple and reliable calibrations within the family [92,93]. The only available fossil that can be used as minimum-age estimate for the split between *Primula* and *Soldanella* is represented by seeds from *Primula riosiae* from the Miocene that are dated at 15.97 Ma (the early-mid Miocene boundary; [94]). Therefore, we performed a completely separate divergence-time analysis from a taxonomically more inclusive sample of six plastid gene regions (*mat*K, *ndh*F, *rbc*L, *trn*L-F, *rps*16 and *rpl*16) available in GenBank (Additional file 2). We included *P. fasciculata*, *P. involucrata*, *P. sikkimensis* and *P. alpicola* in the larger analyses to obtain a root age estimate for *Primula* sect. *Armerina*. The resulting data matrix comprised 7978 aligned sites and 13 species of

Primulaceae, with 8.3% missing data (Additional file 2). Sequence alignment and model specification proceeded as described above, unless otherwise stated. The GTR+G model of sequence evolution was selected by jModelTest 2.1.4 for *rpl*16, *trn*L-F, *ndh*F and *mat*K, GTR+I model for *rps*16 and HKY+G+I model for *rbc*L. A normally distributed prior with a mean of 39.996 Ma and a standard deviation of 11.492 Ma [38] was used to constrain the root of the *Soldanella*/*Androsace* divergence to be within the interval 21.09-58.90 Ma with 95% probability. The calibration point between *Primula* and *Soldanella* based on the fossil of *Primula riosiae* was set to a lognormal prior with an offset of 15.97, a mean of 2.1 and a standard deviation of 0.63. The analyses were run using a random starting tree for 10^8 generations sampling every 10^3 generations under the uncorrelated lognormal relaxed clock model, a birth-death tree prior and the selected models of substitution for different partitions. The analyses were repeated three times to verify convergence by examining the posterior distribution of parameters in Tracer 1.5. After the removal of the burn-in (10 million generations in each analysis, corresponding to 10% of the samples), the inferred age distribution of the node separating the groups containing either *P. fasciculata* and *P. involucrata* or *P. sikkimensis* and *P. alpicola* was estimated in Tracer 1.5.

The age obtained for the *Armerina* section was then used as a calibration point for the root age of the *Armerina* analysis and modeled as a γ prior with a shape of 9.7, a scale of 0.61 and an offset of 1.4. We used similar settings as described above and the samples retained after removal of the burn-in from the three runs were summarized as a maximum clade credibility tree with mean divergence times using TreeAnnotator (part of the BEAST package).

Biogeographic reconstruction

We ran Statistical Dispersal Vicariance Analysis (S-DIVA) using RASP v.2.1 [95,96] to infer the biogeographic history of this section based on the phylogenetic trees constructed only from our concatenated chloroplast dataset. We did not use the *tfp* nuclear dataset since two homologous copies were obtained from some samples, but multiple copies were not present in all species. We defined seven biogeographic regions for the individuals that were collected: A (East Tibet and Hengduan Mountains), B (Himalayas Mountains), C (Northeast QTP), D (Monutains of Central Asia), E (North Europe), F (North America) and G (Mongolian Plateau). Regions A-C were defined according to the biogeographic divisions of China [97], and had been applied in other studies (*e.g.,* [59,98]). Region D was defined based on the distribution area of *P. pamirica*. Regions E-G were defined based on the distribution of *P. nutans* and some outgroup samples used in this study. To account for uncertainties in phylogenetic reconstructions, we randomly chose 20,000 trees from the posterior distribution of trees obtained by BEAST. The number of maximum areas was set to 2 and we estimated the possible ancestral ranges at each node of the selected phylogenetic trees.

Evolution of ecological preferences

Climatic niche is one of the main factors for setting historically biogeographic patterns, especially during drastically climatic changes, such as Quaternary climate oscillations [10-12]. In order to better understand the biogeographic patterns obtained above, we fitted a series of macro-evolutionary models based on 19 bioclimatic variables. We focused on the clades formed by the species *P. fasciculata*, *P. tibetica* and *P. nutans* because they represent the main lineages in the group, and tested whether the evolutionary trajectories of the climatic niches differed among the different clades (F1, F2; Figure 4) obtained for *P. fasciculata* (see results) and its two closely related species *P. nutans* and *P. tibetica* (NT; Figure 4). For this test, we used only the samples of *P. nutans* that were collected in the QTP.

We extracted the 19 bioclimatic variables of WorldClim (http://www.worldclim.org/current; [99]) for all samples of the three groups (F1, F2, NT) using the package raster [100] in R. All the 19 bioclimatic variables were then summarized into principle components using the *prcomp* function in the *stats* package of R [101]. We used the R package OUwie [102] to compare the fit of a series of models (see Table 1 for detailed interpretation for each model) to explain the differences in niche evolution between species inhabiting similar or different biogeographic regions. We tested these models on different sets of groups: (1) F1/F2 *vs.* NT (SET1); (2) F1 *vs.* F2 *vs.* NT (SET2); (3) F1 *vs.* F2/NT (SET3); and (4) F2 *vs.* F1/NT (SET4).

Table 1. Models of niche evolution relevant to different group-sets with their parameters and interpretation, indicating for each model whether the optimal niche value, θ, the intensity of random fluctuations in the evolutionary trajectory, σ^2, and the strength of selection toward the optimal value, α, are modeled with one global parameter or with two or three parameters that are group-specific.

	Parameters			
Model	θ	σ^2	α	Interpretation for models
BM1	Global	Global	-	Evolution is random
BMS	Global	Group-specific	-	Different groups have different rates of niche evolution
OU1	Global	Global	Global	Niche evolution is directed toward an optimal value without being affected by different groups
OUM	Group-specific	Global	Global	Different groups have different optimal values
OUMA	Group-specific	Global	Group-specific	Different groups have different optimal values and strength of selection
OUMV	Group-specific	Group-specific	Global	Different groups have different optimal values and rates of niche evolution
OUMVA	Group-specific	Group-specific	Group-specific	Different groups have different optimal values, strength of selection and rates of niche evolution

Stochastic mapping for all model tests were run 10 times for 100 trees randomly selected from the posterior distribution of trees from the BEAST analysis to account for possible uncertainty in the

estimated values. Model fit was determined using AICc weights calculated from ΔAICc scores [103]. The highest value of AICc weight represents the best model. Finally, we calculated an average AICc weight and lower (2.5%) and upper (97.5%) quantiles of the distributions of AICc weights for each evolutionary niche model.

Additional files

Additional files can be found in https://ndownloader.figshare.com/collections/3635354/versions/1

Abbreviations

QTP: Qinghai-Tibet Plateau; ML: Maximum Likelihood; PP: Posterior Probability; PCA: Principal Component Analysis; EPIC: Exon-primed-intron-crossing; PCR: Polymerase chain reactions; AIC: Akaike information criterion; S-DIVA: Statistical Dispersal Vicariance Analysis; MCC: Maximum Clade Credibility.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NS, GPR and EC conceived the study. GPR carried out the sampling and the lab work, performed the molecular analysis and drafted the manuscript. NS supervised GPR, coordinated the project and helped to draft the manuscript. EC participated in the design and revising of the manuscript. All authors read, reviewed and approved the final manuscript.

Acknowledgements

We would like to thank Prof. JQ Liu for supporting the fieldwork, C. Yang for help in the field, Dr. B. Tian for providing sequences of *Primula zambalensis,* the Harvard University herbaria for providing samples of *P. pamirica* and D. Silvestro, G. Litsios, A. Kostikova & M. Serrano for their great help in analyzing data. We also profited from insightful discussions with M. Nowak. We are grateful for his help in the lab work and the nuclear primers he provided. The paper benefited greatly from the comments received from two anonymous reviewers. This work was funded by the University of Lausanne research fund, the grant 31003A_138282 from the Swiss National Science Foundation to NS and the China Scholarship Council (award to GPR for four year's 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.

References

1. Hoorn C, Wesselingh FP, Ter Steege H, Bermudez MA, Sevink J, et al. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science. 2010;330(6006):927–31.
2. Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, et al. The role of the uplift of the Qinghai-Tibetan plateau for the evolution of Tibetan biotas. Biol Rev. 2014;90(1):236-53.
3. Wen J, Zhang JQ, Nie ZL, Zhong Y, Sun H. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. Front Genet. 2014;5:4.
4. Mayr E. Animal Species and Evolution. Cambridge, Mass: Harvard University Press; 1963.
5. Coyne JA. Genetics and speciation. Nature. 1992;355:511–5.
6. Rice WR, Hostert EE. Laboratory experiments on speciation: what have we learned in 40 years. Evolution. 1993;47:1637–53.
7. Richardson JE, Pennington RT, Pennington TD, Hollingsworth PM. Rapid diversification of a species-rich genus of neotropical rain forest trees. Science. 2001;293:2242–5.
8. Huang WP, Ji HX. Discovery of *Hipparion fauna* in Xizang. Chin Sci Bull. 1979;19:006.
9. Spicer RA, Harris NB, Widdowson WM, Herman AB, Guo S, Valdes PJ, et al. Constant elevation of southern Tibet over the past 15 million years. Nature. 2003;421:622–4.
10. Wiens JJ. The niche, biogeography and species interactions. Philos Trans R Soc B Biol Sci. 2011;366:2336–50.
11. Sexton JP, McIntyre PJ, Angert AL, Rice KJ. Evolution and ecology of species range limits. Ann Rev Ecol Evol Syst. 2009;40:415–36.
12. Lomolino MV, Riddle BR, Brown JH. Biogeography. 3rd ed. Sunderland, MA: Sinauer Associates; 2006.
13. Peterson AT, Soberón J, Sánchez-Cordero V. Conservatism of ecological niches in evolutionary time. Science. 1999;285(5431):1265–7.
14. Wiens JJ, Graham CM. Niche conservatism: integrating evolution, ecology, and conservation biology. Ann Rev Ecol Evol Syst. 2005;36:519–39.
15. Pfenninger M, Nowak C, Magnin F. Intraspecific range dynamics and niche evolution in Candidula land snail species. Biol J Linn Soc. 2007;90(2):303–17.
16. Evans MEK, Smith SA, Flynn RS, Donoghue MJ. Climate, Niche Evolution, and Diversification of the "Bird-Cage" Evening Primroses (*Oenothera*, Sections *Anogra* and *Kleinia*). Am Nat. 2009;173(2):225–40.
17. Kostikova A, Litsios G, Salamin N, Pearman PB. Linking life-history traits, ecology, and niche breadth evolution in North American eriogonoids (Polygonaceae). Am Nat. 2013;182(6):760–74.
18. Wu CY. Flora of Tibet. Beijing: Science Press; 1987.
19. Royden LH, Burchfiel BC, van der Hilst RD. The geological evolution of the Tibetan plateau. Science. 2008;321:1054–8.
20. Harrison TM, Copland P, Kidd WSF, Yin A. Raising Tibet. Science. 1992;255:1663–70.
21. Li JJ, Shi YF, Li BY. Uplift of the Qinghai-Xizang (Tibet) Plateau and Global Change. Lanzhou: Lanzhou University Press; 1995.
22. Shi YF, Tang MC, Ma YZ. The relation of second rising in Qinghai-Tibetan Plateau and Asian monsoon. Sci China Ser D. 1998;28:263–71.

23. Guo ZT, Ruddiman WF, Hao QZ, Wu HB, Qiao YS, Zhu RX, et al. Onset of Asian desertification by 22 Myr ago inferred from loess deposits in China. Nature. 2002;416:159–63.
24. Wu ZY, Wu SG. A proposal for a new floristic kingdom (realm) – the E Asiatic kingdom, its delimitation and characteristics. In: Zhang A, Wu S, editors. Floristic Characteristics and Diversity of East Asian Plants. Beijing: China Higher Education Press; 1996. p. 3–42.
25. Liu JQ, Duan YW, Hao G, Ge XJ, Sun H. Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. J Syst Evol. 2014;52(3):241–9.
26. Richards AJ. Primula. USA, Portland, Oregon: Timber Press; 2003.
27. Hu CM, Kelso S. Primulaceae. In: Wu ZY, Raven PH, editors. Flora of China, Vol 15. Beijing: Science Press; St Louis: Missouri Botanical Garden Press; 1996. p. 99–185.
28. Mao KS, Hao G, Liu JQ, Adams RP, Milne RI. (2010). Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. New Phytol. 2010;188(1):254–72.
29. Mao KS, Milne RI, Zhang LB, Peng YL, Liu JQ, Thomas P, et al. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc Natl Acad Sci USA. 2012;109:7793–8.
30. Sun YS, Wang AL, Wan DS, Wang Q, Liu JQ. Rapid radiation of *Rheum* (Polygonaceae) and parallel evolution of morphological traits. Mol Phylogenet Evol. 2012;63:150–8.
31. Zhang XL, Wang YJ, Ge XJ, Yuan YM, Yang HL, Liu JQ. Molecular phylogeny and biogeography of *Gentiana* sect. *Cruciata* (Gentianaceae) based on four chloroplast DNA datasets. Taxon. 2009;58(3):862–70.
32. Liu JQ, Gao TG, Chen ZD, Lu AM. Molecular phylogeny and biogeography of the Qinghai–Tibetan Plateau endemic *Nannoglottis* (Asteraceae). Mol Phylogenet Evol. 2002;23:307–25.
33. Kelso S. Taxonomy of *Primula* sects. *Aleuritia* and *Armerina* in North America. Rhodora. 1991;93:67–99.
34. Kelso S. The genus *Primula* as a model for evolution in the Alaskan flora. Arct Antarct Alp Res. 1992;24:82–7.
35. Guggisberg A, Mansion G, Kelso S, Conti E. Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid–polyploid species complex of *Primula*. New Phytol. 2006;171(3):617–32.
36. Guggisberg A, Baroux C, Grossniklaus U, Conti E. Genomic origin and organisation of the allopolyploid *Primula egaliksensis* investigated by in situ hybridization. Ann Bot. 2008;101:919–27.
37. Guggisberg A, Mansion G, Conti E. Disentangling reticulate evolution in an arctic–alpine polyploid complex. Syst Biol. 2009;58:55–73.
38. de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. Heterostyly accelerates diversification via reduced extinction in primroses. P Roy Soc B-Biol Sci. 2014;281(1784):20140075.
39. Mast AR, Kelso S, Conti E. Are any primroses (*Primula*) primitively monomorphic? New Phytol. 2006;171(3):605–16.
40. Zhang YX, Zeng CX, Li DZ. Complex evolution in Arundinarieae (Poaceae: Bambusoideae): Incongruence between plastid and nuclear GBSSI gene phylogenies. Mol Phylogenet Evol. 2012;63(3):777–97.

41. Alström P, Höhna S, Gelang M, Eriscon PG, Olsson U. Non-monophyly and intricate morphological evolution within the avian family Cettiidae revealed by multilocus analysis of a taxonomically densely sampled dataset. BMC Evol Biol. 2011;11(1):352.
42. Yu WB, Huang PH, Li DZ, Wang H. Incongruence between Nuclear and Chloroplast DNA Phylogenies in *Pedicularis* Section *Cyathophora* (Orobanchaceae). PLoS ONE. 2013;8(9):e74828.
43. Degnan JH, Rosenberg NA. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol. 2009;24:332–40.
44. Casazza G, Granato L, Minuto L, Conti E. Polyploid evolution and Pleistocene glacial cycles: a case study from the Alpine species *Primula marginata*. BMC Evol Biol. 2012;12(1):56.
45. Schmidt-Lebuhn AN, de Vos JM, Keller B, Conti E. Phylogenetic analysis of *Primula* section *Primula* reveals rampant non-monophyly among morphologically distinct species. Mol Phylogenet Evol. 2012;65:23-34.
46. Petit RJ, Excoffier L. Gene flow and species delimitation. Trends Ecol Evol. 2009;24: 386-93.
47. Hansson B, Tarka M, Dawson DA, Horsburgh GJ. Hybridization but no evidence for backcrossing and introgression in a sympatric population of Great Reed Warblers and Clamorous Reed Warblers. PLoS ONE. 2012;7(2):e31667.
48. Rieseberg LH. Hybrid origins of plant species. Annu Rev Ecol Syst. 1997;28:359–389.
49. Woodell SRJ. Natural hybridization in Britain between *Primula vulgaris* Huds (the primrose) and *P. elatior* (L.) Hill (the oxlip). Watsonia. 1969*;*7*:*115–27.
50. Zhu XF, Li Y, Wu GL, Fang ZD, Li QJ, Liu JQ. Molecular and morphological evidence for natural hybridization between *Primula secundiflora* Franchet and *P. poissonii* Franchet (Primulaceae). Acta Biol Cracov Ser Bot. 2009;51:29–36.
51. Ma YP, Tian XL, Zhang JL, Wu ZK, Sun WB. Evidence for natural hybridization between *Primula beesiana* and *P. bulleyana*, two heterostylous primroses in NW Yunnan, China. J Syst Evol. 2014;52(4):500–7.
52. Buckley TR, Cordeiro M, Marshall DC, Simon C. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada Dugdale*). Syst Biol. 2006;55(3):411–25.
53. Doyle JJ. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst Bot. 1992;17(1):144–63.
54. Joly S, McLenachan PA, Lockhart PJ. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am Nat. 2009;174(2):54–70.
55. Liu L, Pearl DK. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol. 2007;56(3):504–14.
56. Shi YF, Li JJ, Li BY. Uplift of the Qinghai–Tibetan Plateau and East Asia environmental changes during the late Cenozoic. Acta Geogra Sin. 1999;54:10–21.
57. Li L, Abbott RJ, Liu BB, Sun YS, Li LL, Zou JB, et al. Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. Mol Ecol. 2013;22:5237–55.
58. Liu J, Moller M, Provan J, Gao LM, Poudel RM, Li DZ. Geological and ecological factors drive cryptic

speciation of yews in a biodiversity hotspot. New Phytol. 2013;199:1093–108.

59. Zhou Z, Hong D, Niu Y, Li G, Nie Z, Wen J, et al. Phylogenetic and biogeographic analyses of the Sino-Himalayan endemic genus *Cyananthus* (Campanulaceae) and implications for the evolution of its sexual system. Mol Phylogenet Evol. 2013;68(3):482-97.
60. Yang S, Dong H, Lei F. Phylogeography of regional fauna on the Tibetan Plateau: a review. Prog Nat Sci. 2009;19(7):789–99.
61. Zhang JW, Nie ZL, Wen J, Sun H. Molecular phylogeny and biogeography of three closely related genera, *Soroseris*, *Stebbinsia*, and *Syncalathium* (Asteraceae, Cichorieae), endemic to the Tibetan Plateau. SW China Taxon. 2011;60:15–26.
62. Schönswetter P, Tribsch A. Vicariance and dispersal in the alpine perennial *Bupleurum stellatum* L. (Apiaceae). Taxon. 2005;54:725–32.
63. Bryja J, Mikula O, Patzenhauerová H, Oguge NO, Šumbera R, Verheyen E. The role of dispersal and vicariance in the Pleistocene history of an East African mountain rodent, *Praomys delectorum*. J Biogeogr. 2014;41(1):196–208.
64. Xu TT, Abbott RJ, Milne RI, Mao KS, Du FK, Wu GL, et al. Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. BMC Evol Biol. 2010;10(1):194.
65. Wang YJ, Liu JQ, Miehe G. Phylogenetic origins of the Himalayan endemic *Dolomiaea*, *Diplazoptilon* and *Xanthopappus* (Asteraceae: Cardueae) based on three DNA regions. Ann Bot. 2007;99(2):311–22.
66. Li Y, Stocks M, Hemmila S, Källman T, Zhu H, Zhou Y, et al. Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. Mol Biol Evol. 2010;27:1001–14.
67. Liu JQ, Wang YJ, Wang AL, Hideaki O, Abbott RJ. Radiation and diversification within the Ligularia–Cremanthodium–Parasenecio complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. Mol Phylogenet Evol. 2006;38(1):31–49.
68. Jia DR, Abbott RJ, Liu TL, Mao KS, Bartish IV, Liu JQ. Out of the Qinghai–Tibet Plateau: evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae). New Phytol. 2012;194(4):1123–33.
69. Wiens JJ, Donoghue MJ. Historical biogeography, ecology and species richness. Trends Ecol Evol. 2004;19(12):639–44.
70. Stephens PR, Wiens JJ. Bridging the gap between community ecology and historical biogeography: niche conservatism and community structure in emydid turtles. Mol Ecol. 2009;18(22):4664–79.
71. Kozak KH, Wiens JJ. Niche conservatism drives elevational diversity patterns in Appalachian salamanders. Am Nat. 2010;176(1):40–54.
72. Hewitt GM. Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc B Biol Sci. 2004;359(1442):183–95.
73. Ren GP, Abbott RJ, Zhou YF, Zhang LR, Peng YL, Liu JQ. Genetic divergence, range expansion and possible homoploid hybrid speciation among pine species in Northeast China. Heredity. 2012;108(5): 552–62.

74. Polechova J, Barton N, Marion G. Species' range: adaptation in space and time. Am Nat. 2009;174:186– 204.
75. Alleaume - Benharira M, Pen IR, Ronce O. Geographical patterns of adaptation within a species' range: interactions between drift and gene flow. J Evol Biol. 2006;19(1): 203–15.
76. Kawecki TJ. Adaptation to marginal habitats. Ann Rev Ecol Evol Syst. 2008;39: 321–42.
77. Lavergne S, Mouquet N, Thuiller W, Ronce O. Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities. Ann Rev Ecol Evol Syst. 2010;41:321–50.
78. Wiens JJ, Graham CH, Moen DS, Smith SA, Reeder TW. Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. Am Nat. 2006;168(5):579–96.
79. Angert AL. The niche, limits to species' distributions, and spatiotemporal variation in demography across the elevation ranges of two monkeyflowers. Proc Natl Acad Sci USA. 2009;106:19693–8.
80. Kostikova A, Litsios G, Burgy S, Milani L, Pearman PB, Salamin N. Scale - dependent adaptive evolution and morphological convergence to climatic niche in Californian eriogonoids (Polygonaceae). J Biogeogr. 2014;41(7):1326–37.
81. Hoskin CJ, Higgie M, McDonald KR, Moritz C. Reinforcement drives rapid allopatric speciation. Nature. 2005;437:1353–6.
82. Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci USA. 2011;108:19641–6.
83. Katoh, K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008;9:286–98.
84. Fuertes AJ, Rosselló JA, Nieto FG. Molecular evidence for the compilospecies model of reticulate evolution in *Armeria* (Plumbaginaceae). Syst Biol. 1999;48:735–54.
85. Fuertes AJ, Nieto FG. Additive polymorphisms and reticulation in an ITS phylogeny of thrifts (*Armeria*, Plumbaginaceae). Mol Phylogenet Evol. 2003;28:430–47.
86. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics. 2011;27:171–80.
87. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772.
88. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 2009;537:113– 37.
89. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19:1572–74.
90. Sauquet H, Ho SYW, Gandolfo MA, Jordan GJ, Wilf P, Cantrill DJ, et al. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst Biol. 2012;61(2):289–313.
91. Drummond A, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.

92. Schneeweiss GM, Schönswetter P, Kelso S, Niklfeld H. Complex biogeographic patterns in *Androsace* (Primulaceae) anstrd related genera: evidence from phylogenetic analyses of nuclear internal transcribed spacer and plastid trnL-F sequences. Syst Biol. 2004;53:856–76.
93. Boucher FC, Thuiller W, Roquet C, Douzet R, Aubert S, Alvarez N, et al. Reconstructing the origins of high-alpine niches and cushion life form in the genus *Androsace* s.l. (Primulaceae). Evolution. 2012;66:1255–68.
94. Czaja A. Paläokarpologische Untersuchungen von Taphozönosen des Unter- und Mittelmiozäns aus dem Braunkohlentagebau Berzdorf/Oberlausitz (Sachsen). Palaeontogr Abt B. 2003;265:1–148.
95. Yu Y, Harris AJ, He XJ. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Mol Phylogenet Evol. 2010;56:848–50.
96. Yu Y, Harris AJ, He XJ. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol Phylogenet Evol. 2015;87:46-9.
97. Xie Y, MacKinnon J, Li D. Study on biogeographical divisions of China. Biodivers Conserv. 2004;13(7):1391–417.
98. Roquet C, Boucher FC, Thuiller W, Lavergne S. Replicated radiations of the alpine genus *Androsace* (Primulaceae) driven by range expansion and convergent key innovations. J Biogeogr. 2013;40(10):1874–1886.
99. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005;25(15)**:**1965-78.
100. Hijmans R, van Etten J. raster: geographic analysis and modeling with raster data. 2012. http://cran.r-project.org/web/packages/raster/index.html.
101. Venables WN, Ripley BD. *Modern Applied Statistics with S*, Fourth edition. New York: Springer; 2002.
102. Beaulieu JM, Jhwueng DC, Boettiger C, O'Meara BC. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution. 2012;66:2369–83.
103. Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information theoretic approach. New York: Springer-Verlag; 2002.

Chapter 2

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

This chapter was submitted to Journal of Biogeography

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

Guangpeng Ren[1,2,3,4], Rubén G. Mateo[1,5], Antoine Guisan[1,6], Elena Conti[7], Nicolas Salamin[1,2,3*]

[1] Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[2] Department of Computational Biology, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[3] Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland

[4] State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou 730000 Gansu, China

[5] ETSI de Montes, Forestal y del Medio Natural, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain.

[6] Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, 1015 Lausanne, Switzerland

[7] Department of Systematic and Evolutionary Botany and Botanic Garden, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

Abstract

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 parapatricly 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 Northeast Qinghai-Tibet Plateau (QTP) to examine spatial and ecological effects on interspecific divergence and maintenance of species cohesion. We used genomic data for 770 samples of the three species using restriction site-associated DNA (RAD) sequencing and combined approximate Bayesian computation (ABC) modeling, Bayesian generalized linear mixed modeling (GLMM) and niche-based species distribution modeling (SDM). The three species are clearly delimited by the RADseq data. Further ABC modeling indicates that the three species originated in the Himalayas and diverged from each other following the uplifts of the Hengduan Mountains and the Northern QTP during the Pliocene. After a long period of divergence, the three species came into secondary contact triggered by past climatic changes as suggested by the SDMs but with no significant introgression. The three species display complex and different drivers of genomic variation, which provides further insights into the effects of geographical and ecological factors on maintaining species cohesion. Our findings highlight the significance of combining the use of 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, population genomics, Qinghai-Tibet Plateau, secondary contact, speciation, maintenance of species cohesion

Introduction

Understanding the relative roles of geography and ecology in speciation, population divergence and maintenance of species cohesion is a longstanding 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). It has however recently been proposed that adaptation to different environmental conditions, resulting in separated populations evolving different ecological niches, can also lead to divergence and speciation (Berlocher & Feder 2002; Rundle & Nosil 2005; Nosil 2008). 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 (Hoorn *et al.* 2010; Favre *et al.* 2014). Furthermore, the mechanisms associated with these historical events do not have similar effects on population divergence, speciation and species maintenance across organisms (e.g. Zhang *et al.* 2005; Opgenoorth *et al.* 2010; Sun *et al.* 2014; Ren *et al.* 2017). Despite their importance in driving the high biodiversity in mountainous areas (Myers *et al.* 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; Jia *et al.* 2012; Anacker & Strauss 2014), helping to clarify the mechanisms of speciation. To address this issue, one however needs to resolve the phylogenetic relationships between closely related species. This task is usually difficult to achieve with traditional neutral markers, especially for genera that harbor high species richness and experience hybridization during their evolution, such as *Primula* L. (Guggisberg *et al.* 2009; Schmidt-Lebuhn *et al.* 2012) and *Gentiana* L. (Liu *et al.* 2016). This problem may be overcome by using many thousands of DNA markers (e.g. Wagner *et al.* 2013; Pante *et al.* 2015), for example using recently developed next-generation sequencing (NGS) 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, we investigate the effects of geological and climatic factors on population divergence, speciation and the maintenance of species cohesion in three closely related species of the genus *Primula* (Primulaceae): *P. nutans* Georgi, *P. fasciculata* Balf. f. & Kingdon-Ward and *P. tibetica* Watt (section

Armerina). The three species represent prominent floristic elements of alpine meadows at high altitudes (Hu & Kelso 1996) and are widely distributed in the Qinghai-Tibet Plateau (QTP), which experienced drastic habitat changes and harbors extremely rich species diversity and endemism (Wu 1987). 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 million years ago (Ma), the times of its following uplifts are controversial (reviewed in Renner et al. 2016). Some scientists believe that the QTP has reached 4000 m since the mid-Eocene (40 Ma; Renner et al. 2016 and references therein), while others suggest that a more recent uplift has occurred during the late Miocene and Pliocene (2.4-8 Ma), particularly at its eastern and northern edge such as the Hengduan Mountains and the Qaidam basin (Li & Fang 1999; Zheng *et al.* 2000; Mulch & Chamberlain 2006). Many evolutionary studies have indicated that extensive species diversification took place in the region during the Pliocene (Liu *et al.* 2002, 2006; Wang *et al.* 2010; Li *et al.* 2012; Li *et al.* 2013), which seem to support the latter hypothesis. The three species studied here, occurring in the Himalayas, the Hengduan Mountains and the Northeast QTP, respectively (Fig. 1a), represent a unique opportunity and the first time using population genomic data to evaluate whether their divergence may be triggered by such a recent uplift of the eastern and northern edge of the QTP. 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 *et al.* 2014; Ren *et al.* 2015). However, the roles played by historical geological events on the initial interspecific divergence of these three *Primula* species and the factors that are influencing the current distributions of the species and their maintenance remain unknown.

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 2900 and 5000 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 3800 m in the Northeast QTP (Fig. 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 overlap in the geographical ranges between *P. tibetica* and *P. fasciculata*, and between *P. nutans* and *P. fasciculata,* respectively (Fig. 1). This distribution pattern coupled with the use of population genomic data provides an opportunity to investigate the relative roles of geography and ecology in population/species divergence. Specifically, the aims of our study are 1) to characterize interspecific divergence of the three species based on RADseq data by sampling multiple populations that represent most of the distribution in the QTP for each species; 2) to decipher their interspecific divergence, and link their divergence times with historical geological

events; 3) to investigate the effects of climatic changes and identify the drivers maintaining species cohesion.

Materials and methods

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 (Fig. 1; 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* (i.e. same strategy as applied for *P. tibetica* to select populations that were representative of both the geographical distribution and the diversity of ecological niches) were used. Six to 20 individuals were chosen from each population, which gave us a total of 770 individuals that were processed with a double-digestion restriction site-associated DNA sequencing (RADseq) following the same protocol used for *P. tibetica* (Ren *et al.* 2017). The libraries were sequenced using single-end reads of 100 bp of length.

Single-end Illumina reads were processed into RAD-tags using the STACKS-1.30 software pipeline (Catchen *et al.* 2011, 2013). All reads were trimmed to 90 bp in length. We used all 770 samples to build a catalogue 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 --lnl_lim -10.0 --lnl_dist*. After *rxstacks*, *cstacks* and *sstacks* were run again with the same setting as before to rebuild the catalogue 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 catalogue of reads using the *populations* module to generate three data sets (i.e. one for each species; D1 - D3) and one data set considering all the three species (D4) for downstream population genetic analyses. We retained polymorphic RAD loci that were only present in all populations for each data set and scored for each RAD locus only the first SNP if several were present. Pairwise F_{ST} values and analysis of molecular variance (AMOVA) for different data sets and different

genomic fractions (see below) were calculated among populations in GENODIVE v.2.0b27 (Meirmans & Van Tienderen 2004), and significance was determined using 1×10^4 permutations.

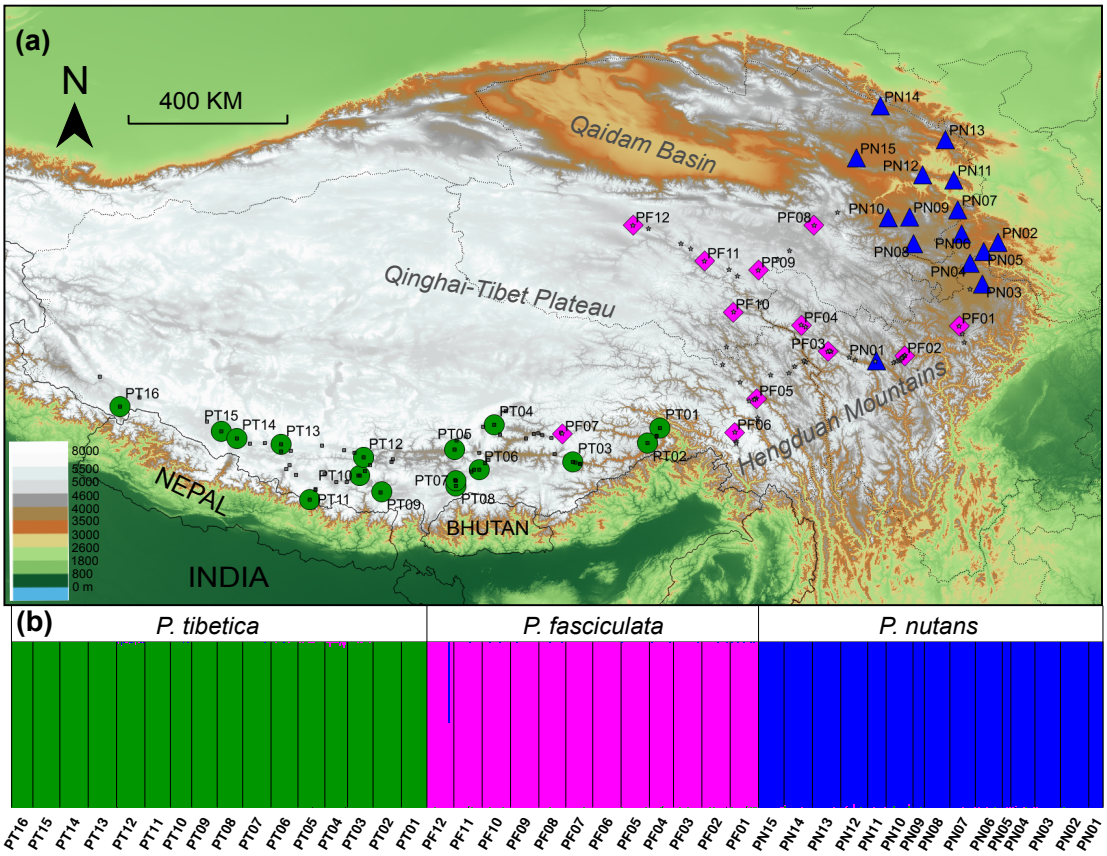


Fig. 1 (a) Geographic locations of the 43 populations analyzed in the present study. Small grey squares and stars represent sampled populations for *P. tibetica* and *P. fasciculata* from fieldwork, respectively. (b) Plots of posterior probabilities for individuals of the three species assigned to *K* genetic clusters from STRUCTURE analyses for $K = 3$.

Outlier detection

Polymorphic loci from the datasets D1 to D3 potentially under balancing and divergent selection for each species were screened for statistical outliers as implemented in BAYESCAN 2.1 (Foll & Gaggiotti 2008). BAYESCAN estimates population-specific F_{ST} coefficients by the Bayesian method described in 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. Loci that were identified as balancing and divergent outliers were segregated into a negative and a positive outlier data set, respectively; the remaining loci (with outliers removed) comprised the neutral data set. We further investigated whether outlier loci in one species were also detected as outlier loci in the other two species.

Interspecific divergence

Dataset D4 was used for population structure analysis using a Bayesian method implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000) and for reconstructing a maximum likelihood tree in PHYML 3.0 (Guindon *et al.* 2010) to investigate interspecific divergence among the three species. 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 *et al.* (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). Nodal support of phylogenetic tree was estimated from 1000 bootstrap replicates in PHYML.

Generalized linear mixed modeling of genomic, spatial and environmental data

A generalized linear mixed modeling (GLMM) approach was performed on the three datasets D1 to D3 to test whether population divergence of different genomic fractions (negative, neutral and positive RAD loci) for each species was driven by isolation by distance (IBD) or isolation by environment (IBE) or both (*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 data sets (negative, neutral and positive RAD loci) 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: 1) 19 WorldClim variables (Hijmans *et al.* 2005; 'CLI' for climatic data from here onwards) and 2) three edaphic variables (Soil- carbon, moisture, PH; 'SOI' 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 sixteen 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 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*.

Modeling of historical divergence

We used approximate Bayesian computation (ABC) method implemented in DIY-ABC v2.1.0 (Cornuet *et al*. 2014) to further decipher the historical divergence of the three species. We subsampled from the D4 dataset two individuals from each of the 43 populations ten times and pooled individuals into 'species' groups to generate unbiased estimates of species history and to reduce computational time. We selected for each sub-dataset 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 sub-datasets were listed in Supporting Information Table S2. We tested seven possible scenarios for the origins and relationships between species: *P. nutans* diverged first (S1-S2), *P. tibetica* diverged first (S3-S4), *P. fasciculata* diverged first (S5-S6) and all three species diverged simultaneously from their ancestor (S7, Fig. S1). We gave each scenario a uniform prior probability (Table S3) and selected all summary statistics to generate a reference table containing 7×10^6 simulated data sets (on average 10^6 per scenario). We used 1% of the simulated data sets 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 one year for the three species (Ren *et al.* 2017).

Species distribution models

An ensemble of species distribution models was generated for *P. nutans* and *P. fasciculata* following the same methodology applied for *P. tibetica* (see 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 *et al.* 2009). A total of 67 and 89 species occurrences obtained from the field collections and herbarium records were used as presences data to calibrate the models for *P. nutans* and *P. fasciculata,* respectively. As environmental variables we used the WorldClim database (Hijmans et al., 2005), and, to avoid multicollinearity, we used the same ones as in Ren *et al.* (2017). 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 (ROC) plot (Swets 1988) and the true skill statistics (TSS; Allouche *et al.* 2006). The averaged and pondered consensus model (the contribution of each replicate was weighted proportionally to their AUC values) was then projected onto the last Maximum Glacial (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),

Results

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; 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.

Table 1 Analysis of molecular variance (AMOVA) estimated for three different genomic fractions: negative, neutral and positive divergent outlier RADseq markers for three species. Number of loci was also shown for each data set.

Genome fraction	No. of loci	Source of variation	df	Variance components	% of variation	F_{ST}
P. nutans						
RAD/negative	36	Within	227	8.38	95.7	
		Among	14	0.38	4.3	0.043*
RAD/neutral	12834	Within	227	906.30	85.0	
		Among	14	160.01	15.0	0.150*
RAD/positive	55	Within	227	9.93	67.6	
		Among	14	4.76	32.4	0.324*
P. fasciculata						
RAD/negative	94	Within	222	23.84	85.6	
		Among	11	101.74	14.4	0.144*
RAD/neutral	5946	Within	222	385.37	69.4	
		Among	11	169.74	30.6	0.306*
RAD/positive	46	Within	222	7.41	36.2	
		Among	11	13.04	63.8	0.638*
P. tibetica						
RAD/negative	54	Within	277	10.67	77.1	
		Among	15	3.17	22.9	0.229*
RAD/neutral	2727	Within	277	191.75	58.6	
		Among	15	135.59	41.4	0.414*
RAD/positive	41	Within	277	5.29	33.4	
		Among	15	10.54	66.6	0.666*

*, $P < 0.001$

Outlier loci

Outlier detection identified 140 potentially non-neutral outlier RAD loci in *P. fasciculata*, 94 of which were in the lower tail (negative outliers) and 46 in the upper tail (positive outliers; Fig. S2; Table 1). In *P. nutans*, 91 RAD loci were revealed as outliers, 36 of which were showing negative and 55 were positive. Similarly, 95 RAD loci in *P. tibetica* were identified as outliers, 54 of which were negative

and 41 were positive. None of these loci was shared as outlier in all of the three species. As expected, AMOVA of negative, neutral and positive RADseq polymorphisms revealed greatly increased among-population variance for the latter ones in the three species. The among-population variances (F_{ST}) were smaller in *P. nutans* than in the other two species (Table 1).

Interspecific divergence

Both the STRUCTURE and phylogenetic tree suggested a clear species delimitation of the three species (Fig. 1b, S3). 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.

Modeling of genetic, spatial and environmental data

Generalized linear mixed models (GLMM) 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 2). For *P. nutans*, geographical distance (GEO) was the main driver for the variation of negative and neutral RAD data sets, with DIC weights equal to 0.396 and 0.353, respectively. The best predicted models for the divergence of positive outlier RAD loci were GEO+CLI (DIC weight 0.199) or GEO+CLI+SOI (DIC weight 0.198), indicating that climatic and edaphic variables were involved in triggering divergent selection between populations in *P. nutans*. GEO alone also received non-negligible support for the divergent selected fraction of genomes in this species (DIC weight 0.168). By contrast, the model including all of the four predictors (GEO+CLI+SOI+ALT) was best supported in driving the variation of negative, neutral and positive RAD polymorphisms in *P. fasciculata*. The highest DIC value (0.861) for the positive data set indicated that both the geographical 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 0.367). For negative and neutral data sets in *P. tibetica*, the best-supported model was GEO+CLI+ALT. Similarly, more variables were involved in driving divergent selection among populations as expected.

Estimates of historical divergence

ABC modeling of historical divergence of the three species indicated that the scenario involving an initial divergence of *P. fasciculata* from *P. tibetica* followed by a later origin of *P. nutans* from *P. fasciculata* (S3; Fig. S1) provided the best fit to our RADseq data in all of the ten data sets (Table S5). The parameter values were estimated for each data set based on the best-fit scenario, and we showed only the averaged values here (Table S6). Our ABC modeling suggested that *P. fasciculata* diverged

from *P. tibetica ca.* 4.65 Ma (95% highest posterior density (HPD): 1.74-8.80 Ma; Fig. 2), and *P. nutans* originated from *P. fasciculata ca.* 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 ten times at the beginning or middle of Pleistocene (Table S6).

Table 2 Results of GLMM set up to predict genetic divergence 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 shown. For each mode comparison, the best-supported model is shown in bold italics.

	RAD/negative			RAD/neutral			RAD/positive		
Model	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight	DIC	Delta DIC	DIC weight
P. nutans									
NULL	-519.0	11.25	0.001	-394.3	37.07	0.000	-137.8	79.56	0.000
GEO	***-530.3***	***0.00***	***0.396***	***-431.3***	***0.00***	***0.353***	-217.0	0.34	0.168
CLI	-523.7	6.57	0.015	-410.6	20.77	0.000	-167.8	49.51	0.000
SOI	-522.2	8.10	0.007	-409.2	22.11	0.000	-185.4	31.92	0.000
ALT	-516.9	13.37	0.000	-392.4	38.90	0.000	-136.2	81.14	0.000
GEO+CLI	-528.2	2.09	0.139	-429.7	1.61	0.158	***-217.3***	***0.00***	***0.199***
GEO+SOI	-528.2	2.12	0.137	-429.6	1.77	0.146	-215.9	1.45	0.096
GEO+ALT	-528.2	2.13	0.137	-429.5	1.87	0.139	-215.5	1.83	0.080
CLI+SOI	-522.0	8.32	0.006	-411.3	20.03	0.000	-185.8	31.55	0.000
CLI+ALT	-275.6	254.7	0.000	-408.5	22.87	0.000	-165.8	51.51	0.000
SOI+ALT	-520.1	10.21	0.002	-407.3	23.99	0.000	-183.7	33.60	0.000
GEO+CLI+SOI	-526.0	4.28	0.047	-427.7	3.60	0.058	-217.3	0.01	0.198
GEO+CLI+ALT	-526.0	4.26	0.047	-428.0	3.36	0.066	-216.1	1.25	0.106
GEO+SOI+ALT	-526.0	4.26	0.047	-427.7	3.65	0.057	-214.4	2.97	0.045
CLI+SOI+ALT	-519.8	10.45	0.002	-409.3	22.07	0.000	-184.0	33.39	0.000
GEO+CLI+SOI+ALT	-523.9	6.40	0.016	-425.9	5.41	0.024	-216.2	1.19	0.109
P. fasciculata									
NULL	-216.3	24.35	0.000	-131.5	57.42	0.000	-19.5	27.38	0.000
GEO	-229.7	10.87	0.002	-153.4	35.59	0.000	-26.0	20.88	0.000
CLI	-214.1	26.54	0.000	-138.8	50.12	0.000	-19.5	27.42	0.000
SOI	-218.7	21.96	0.000	-130.3	58.68	0.000	-20.6	26.27	0.000
ALT	-214.5	26.10	0.000	-133.5	55.48	0.000	-27.8	19.12	0.000
GEO+CLI	-229.0	11.65	0.001	-156.6	32.37	0.000	-26.3	20.63	0.000
GEO+SOI	-239.1	1.52	0.182	-154.4	34.52	0.000	-34.9	12.04	0.002
GEO+ALT	-232.7	7.90	0.007	-152.3	36.63	0.000	-28.0	18.85	0.000
CLI+SOI	-219.6	20.99	0.000	-156.5	32.42	0.000	-32.9	13.99	0.001
CLI+ALT	-212.5	28.09	0.000	-137.7	51.24	0.000	-26.7	20.15	0.000
SOI+ALT	-218.6	22.06	0.000	-133.0	55.94	0.000	-31.2	15.72	0.000
GEO+CLI+SOI	-238.7	1.94	0.147	-187.0	1.92	0.277	-42.6	4.31	0.100
GEO+CLI+ALT	-231.0	9.66	0.003	-157.8	31.21	0.000	-28.0	18.92	0.000
GEO+SOI+ALT	-239.9	0.72	0.270	-148.3	40.69	0.000	-35.5	11.41	0.003
CLI+SOI+ALT	-218.6	22.00	0.000	-157.0	31.94	0.000	-40.4	6.52	0.033
GEO+CLI+SOI+ALT	***-240.6***	***0.00***	***0.388***	***-189.0***	***0.00***	***0.723***	***-46.9***	***0.00***	***0.861***
P. tibetica									
NULL	-265.9	99.33	0.000	-157.5	165.5	0.000	-23.9	133.8	0.000
GEO	-328.3	36.94	0.000	-297.9	25.12	0.000	-129.4	28.26	0.000
CLI	-276.7	88.53	0.000	-176.8	146.2	0.000	-47.2	110.5	0.000
SOI	-264.4	100.8	0.000	-156.7	166.3	0.000	-23.3	134.4	0.000
ALT	-344.7	20.55	0.000	-216.8	106.2	0.000	-43.1	114.5	0.000
GEO+CLI	-333.0	32.22	0.000	-317.2	5.82	0.030	-155.9	1.69	0.157
GEO+SOI	-326.4	38.76	0.000	-297.1	25.83	0.000	-128.4	29.28	0.000
GEO+ALT	-360.5	4.72	0.064	-304.0	19.00	0.000	-129.2	28.46	0.000
CLI+SOI	-275.3	89.93	0.000	-176.3	146.7	0.000	-48.4	109.3	0.000
CLI+ALT	-351.5	13.71	0.001	-232.8	90.13	0.000	-65.5	92.16	0.000
SOI+ALT	-342.5	22.68	0.000	-215.1	107.9	0.000	-41.3	116.3	0.000
GEO+CLI+SOI	-331.2	33.99	0.000	-317.0	5.98	0.027	-155.5	2.10	0.128
GEO+CLI+ALT	***-365.2***	***0.00***	***0.676***	***-323.0***	***0.00***	***0.546***	-157.5	0.11	0.348
GEO+SOI+ALT	-358.4	6.82	0.022	-302.8	20.16	0.000	-128.4	29.21	0.000
CLI+SOI+ALT	-349.4	15.81	0.000	-231.3	91.66	0.000	-64.6	93.01	0.000
GEO+CLI+SOI+ALT	-363.1	2.10	0.237	-322.3	0.64	0.397	***-157.6***	***0.00***	***0.367***

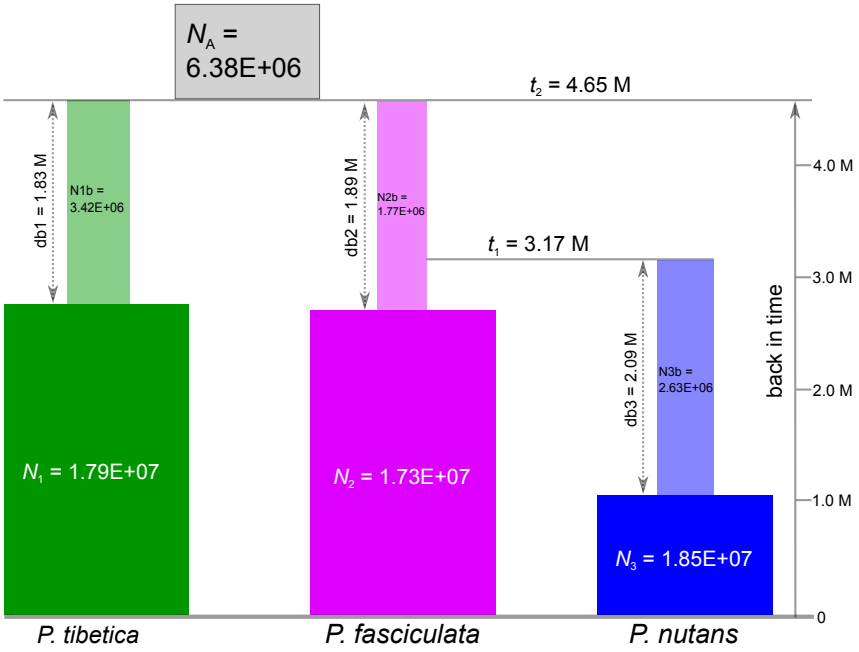


Fig. 2 Summary of inferred interspecific divergence of the three species. Effective population sizes (N1 and N1b, *P. tibetica*; N2 and N2b, *P. fasciculata*; N3 and N3b, *P. nutans*), times of divergence in years (t1, t2) and durations of founder events (db1-db3) are indicated.

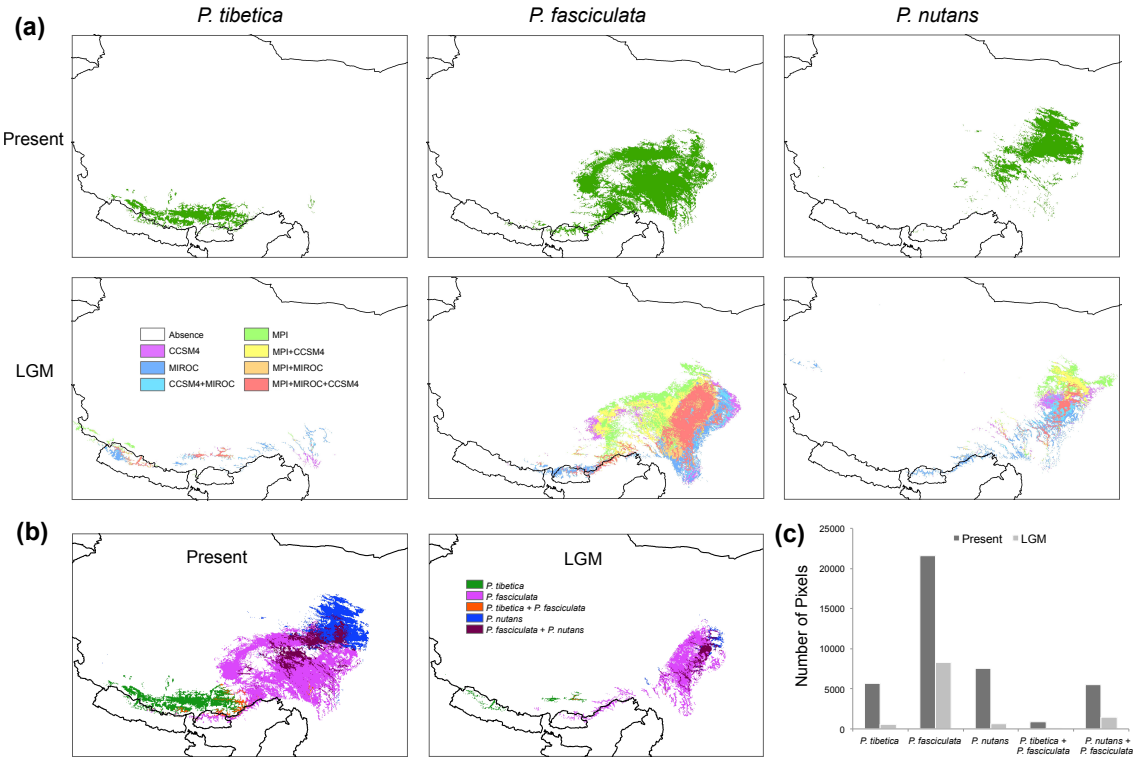


Fig. 3 (a) Habitat suitability of the three species predicted by species distribution models (SDM) for present and LGM. SDMs for the LGM are based on three different general circulation models (GCM). (b) The predicted and the overlapping distributions between the three species for the present and LGM. Only the distributions predicted by all the three GCMs during the LGM (red areas in a) were considered here. (c) The number of pixels counted based on (b).

Species distribution models

The results of SDMs showed that the three species occupy mostly different environments (Fig. 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* (5520 pixles; Fig. 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 (Fig. 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. Two main glacial refugia were identified for *P. tibetica,* which were located in the central and southwestern Himalayas. *P. 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 Northeast QTP and some valleys of the Hengduan Mountains, mostly nested in the main refugium of *P. fasciculata*.

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). Our analyses based on ABC modeling provide strong evidence that the three species originated in the Himalayas. The timeframes of the first divergence of *P. tibetica* and *P. fasciculata* and the later origin of *P. nutans* from *P. fasciculata* coincide with the extensive uplift of the Hengduan Mountains and the Qaidam basin, which occurred during the late Miocene and the Pliocene. These results suggest an important role of geography in driving the initial interspecific divergence. After the divergences, both spatial and environmental factors have been involved in population divergence and to maintain species divergence despite secondary contact triggered by Pleistocene climatic oscillations. Our study thus contributes to a better understanding of how prominent floristic elements of alpine meadows at high altitudes distributed in three most fascinating regions (i.e. Himalayas, Hengduan Mountains and Northeast QTP) in the QTP responded to historical orographic uplift and climatic changes.

Interspecific divergence in response to the uplifts of the QTP

A clear delimitation of the three closely related *Primula* species was recovered based on RADseq data (Fig. 1, S3), which allowed us to estimate accurately their times of divergence. Using a replicated subsampling strategy, all the ABC models support a scenario that *P. fasciculata* diverged from *P. tibetica* first and that *P. nutans* originated from *P. fasciculata* in a later stage. Given the endemic distribution of *P. tibetica* and *P. fasciculata* in the QTP region, these results provide strong evidence

that the three species originated in the Himalayas. *P. fasciculata* was estimated to diverge from *P. tibetica ca.* 4.65 Ma (HPD: 1.74-8.80 Ma), whereas *P. nutans* originated from *P. fasciculata* more recently *ca.* 3.17 Ma (HPD: 1.58-5.21 Ma; Fig. 2; Table S6). These divergence times were largely consistent with a period of recent uplift of the eastern and northern of the QTP during the late Miocene and Pliocene (Li *et al.* 1995; Mulch & Chamberlain 2006). In fact, numerous studies of other herb, shrub, and conifer groups that grow in the QTP have demonstrated that intra- or interspecific divergences took place during the Pliocene (Liu *et al.* 2002; Xu *et al.* 2010; Wang *et al.* 2010; Zhou *et al.* 2012; Li *et al.* 2013; Liu *et al.* 2013). Furthermore, our sampling of the three species that covers the Himalayas, the Hengduan Mountains and Northeast QTP, respectively, provided a unique opportunity to evaluate the effects of the uplift of a particular region in 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 the Qaidam basin that occurred during the Late Miocene and the Pliocene (8-2.4 Ma; Li & Fang 1999; Zheng *et al.* 2000). Elevations as high as today's elevations already occurred in the Himalayas as early as 15-10 Ma (Favre *et al.* 2015; or 40 Ma in Renner *et al.* 2016). The origin of the three species in the Himalayas indicates that the common ancestor may have already adapted to high altitudes (i.e. cold niches), which is congruent with a previous study on *P. tibetica* (Ren *et al.* 2017). If the Hengduan Mountains and the Northeast QTP already reached their current altitudes 40 Ma, the open cold habitats would have probably allowed the expansion of the common ancestor rather than divergence. Instead, in a recent uplift scenario, the extensive uplift of the Hengduan Mountains followed by the occurrence of high mountains separated by deep valleys may have created cold conditions for the geographical isolation and origin of *P. fasciculata*, which is also capable of surviving and reproducing currently at high altitudes in Hengduan Mountains. Similarly, a rapid uplift of the Qaidam basin that occurs nearly at the same time (Zhang *et al.* 2013) may have further triggered the divergence between *P. nutans* and *P. fasciculata* through geographical isolation. Furthermore, the niche differentiation that can be observed in the current SDMs of these three species (Fig. 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 (Fig. 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 (Templeton 1980; Barton 1984; Weinberg *et al.* 1992).

It should be noted that the timeframes of divergence estimated based on our RADseq data are much older than previous estimates based on few chloroplast markers (1.7-8.8 *vs.* 0.5-2 Ma; Table 3; 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). In particular, *P. nutans* occupies now a wide geographical range from Northern Europe eastward to Northeastern

Siberia, Alaska, and central Asian mountains. It is difficult to imagine how this small plant species (10-30 cm tall), which is pollinated mainly by insects and has barochore seeds (Richards 2003), could have dispersed so widely within about one million years. The origin of *P. nutans* based on RADseq data in the present study is estimated at *ca.* 3.17 Ma (HPD: 1.58-5.21 Ma), which is in agreement with macrofossil evidence that suggests the present-day arctic flora developed *ca.* 3-4 Ma at a time when global temperature decreased sharply (Matthews Jr. & Ovenden 1990; Zachos *et al.* 2001). However, the present study only focused on the QTP and further studies involving a finer sampling across the entirely 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).

Species maintenance in secondary contact zones

Climatic oscillations during the Pleistocene 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 (Rieseberg 1997; Ma *et al.* 2006; Abbott *et al.* 2013; Sun *et al.* 2014). However, our analyses based on population genomic data indicate no hybridization between the three *Primula* species (Fig. 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 (Fig. 3b,c) or during our field collection (Fig. 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; Zhu *et al.* 2009; Ma *et al.* 2014). 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 2) 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 2). 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 2). 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 similar pattern found between these two species may suggest adaptation to their specific ecological niches as also shown in the niche modeling (Fig. 3) and could help the maintenance of species boundaries between them (e.g. Zhou *et al.* 2014; Twyford *et al.* 2015).

By contrast, geographic distance is the predominant mechanism explaining the patterns of divergence and gene flow in the negative and neutral genomic fractions of DNA in *P. nutans*. Geographic distance also received non-negligible support for the fraction of SNPs under divergent selection in this species (Table 2). The variation among populations in neutral and selected genomic regions is much lower in *P. nutans* than in the other two species (Table 1), indicating higher gene flow among populations in *P. nutans*. The occurrence of this species at lower altitude (average altitude 3311 m compared with *P. fasciculata* - 4256 m and *P. tibetica* - 4093 m) where topography is less complex (Fig. 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 2; Wu & Ting 2004; Via & West 2008; Nosil *et al.* 2009). 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 *et al.* 2009; Petit & Excoffier 2009; Zhou *et al.* 2010, 2014).

Conclusions

We combined population genomics and SDMs to identify the relative roles of geography and ecology in speciation, population divergence and the maintenance of species cohesion of three high-altitude plant species over a large area of the QTP. Our results highlight the power and importance of the use of population genomic data in delimiting relationship of closely related species that is failed with traditional markers. Our analyses provide clearly evidence for an origin of the three species in the Himalayas and demonstrate the important roles of the uplifts of the Hengduan Mountains and Northern QTP in driving their initial interspecific divergence. The combination of evolutionary modeling on neutral and selected genomic regions and SDMs also provides new insights in maintaining species divergence in secondary contact zones. These findings indicate that a combination

of geography and ecology has played a fundamental role in promoting diversification and evolution of species in mountainous regions such as the QTP.

Acknowledgements

This work was funded by the University of Lausanne research fund, grant 31003A_138282 from the Swiss National Science Foundation, the China Scholarship Council (award to GR for four 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.

Author contributions

G.R., N.S. and E.C. planned and designed the research. G.R. carried out the sampling and the lab 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 can be found in:

https://www.dropbox.com/s/rxgr3fhsj9mq98b/Supporting_information_chapter2.docx?dl=0

References

Abbott RJ, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Abbott RJ, Smith L, Milne RI *et al.* (2000) Molecular Analysis of Plant Migration and Refugia in the Artic. *Science*, **289**, 1343–1346.

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.

Anacker BL, Strauss SY (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.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barton NH, Charlesworth B (1984) Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics*, **15**(1), 133–164.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, **13**, 969–980.

Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, **47**, 773–815.

Catchen JM, Amores A, Hohenlohe P *et al.* (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population

genomics. *Molecular Ecology*, **22**, 3124–3140.

Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, Massachusetts.

Du FK, Petit RJ, Liu JQ (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.

Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources,* **4**, 359–361.

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.

Favre A, Päckert M, Pauls SU *et al.* (2014) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological reviews*, **90**, 236–253.

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Ecological Conservation*, **24**, 38–49.

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.

Gavrilets S, Boake CR (1998) On the evolution of premating isolation after a founder event. *The American naturalist*, **152**, 706–716.

Guggisberg A, Mansion G, Conti E (2009) Disentangling Reticulate Evolution in an Arctic-Alpine Polyploid Complex. *Systematic Biology*, **58**, 35–73.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hadfield JD (2010) MCMC methods for multi-respoinse generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.

Hu CM, Kelso S (1996) Primulaceae. In: Wu ZY, Raven PH, eds. *Flora of China* Vol 15. St Louis: Missouri Botanical Garden Press, 99–185. Science Press, Beijng

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jia DR, Abbott RJ, Liu TL *et al.* (2012) Out of the Qinghai-Tibet Plateau: Evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae). *New Phytologist*, **194**, 1123–1133.

Lexer C, Wüest RO, Mangili S *et al.* (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.

Li L, Abbott RJ, Liu B *et al.* (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.

Li J, Fang X (1999) Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin*, **44**, 2117–2124.

Li JJ, Shi YF, Li BY (1995) Uplift of the Qinghai–Xizang (Tibet) Plateau and Global Change. Lanzhou University Press, Lanzhou.

Liu JQ, Gao TG, Chen ZD, Lu AM (2002) Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Molecular Phylogenetics and Evolution*, **23**, 307–325.

Liu J, Möller M, Provan J *et al.* (2013) Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, **199**, 1093–1108.

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.

Ma XF, Szmidt AE, Wang XR (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.

Ma YP, Tian XL, Zhang JL, Wu ZK, Sun WB (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.

Matthews JJV, Ovenden LE (1990) Late Tertiary Plant Macrofossilsf from Localities in Arctic/Subarctic North America: A Review of the Data. *Arctic*, **43**, 364–392.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Mayr E (1963) Animal Species and Evolution. Belknap Press of Harvard University Press, Cambridge.

Mulch A, Chamberlain CP (2006) Earth science – The rise and growth of Tibet. *Nature,* **439**, 670–671.

Muñoz-Pajares AJ, García C, Abdelaziz M *et al.* (2017) Drivers of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Molecular Ecology*, doi: 10.1111/mec.13971.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.

Nosil P (2012) Ecological Speciation. Oxford University Press, Oxford.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Opgenoorth L, Vendramin GG, Mao K *et al.* (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.

Pante E, Abdelkrim J, Viricel A *et al.* (2015) Use of RAD sequencing for delimiting species. *Heredity*, **114**, 450–9.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology and Evolution*, **24**, 386–393.

Pritchard JK, Stephens M, Donelly 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.

Ren G, Mateo G, Liu J *et al.* (2016) 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.

Renner SS (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.

Rice WR, Hostert EE (1993) Laboratory experiments on speciation: what have we learned in forty years? *Evolution*, **47**, 1637–1653.

Richards J (2003) Primula. Second Ed. Portland, OR, Timber Press, USA.

Rieseberg LH (1997) Hybrid Origins of Plant Species. *Annual Review of Ecology and Systematics*, **28**, 359–389.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Sauquet H, Ho SYW, Gandolfo MA *et al.* (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.

Schmidt-Lebuhn AN, de Vos JM, 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.

Sun Y, Abbott RJ, Li L *et al.* (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.

Swets JA (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285–1293.

Templeton AR (1980) The theory of speciation via the founder principle. *Genetics*, **94**(4), 1011–1038.

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

Twyford AD, Kidner CA, Ennos RA (2015) Maintenance of species boundaries in a Neotropical radiation of *Begonia*. *Molecular Ecology*, **24**, 4982–4993.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

de Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E (2014) Heterostyly accelerates diversification via reduced extinction in primroses. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 1–9.

Wagner CE, Keller I, Wittwer S *et al.* (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.

Wang Q, Liu J, Allen GA *et al.* (2016) Arctic plant origins and early formation of circumarctic distributions: A

case study of the mountain sorrel, *Oxyria digyna*. *New Phytologist*, **209**, 343–353.

Wang H, Qiong L, Sun K *et al.* (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.

Weinberg JR, Starczak VR, Jorg D (1992) Evidence for Rapid Speciation Following a Founder Event in the Laboratory. *Evolution*, **46**, 1214.

Wu CY (1987) Flora of Tibet. Science Press, Beijing

Wu C-I, Ting C-T (2004) Genes and speciation. *Nature Reviews Genetics*, **5**, 114–122.

Xu T, Abbott RJ, Milne RI *et al.* (2010) Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *BMC evolutionary biology*, **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.

Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ (2005) Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology*, **14**, 3513–3524.

Zhang W, Fang X, Song C *et al.* (2013) Late Neogene magnetostratigraphy in the western Qaidam Basin (NE Tibetan Plateau) and its constraints on active tectonic uplift and progressive evolution of growth strata. *Tectonophysics*, **599**, 107–116.

Zheng H, Powell CM, An Z, Zhou J, Dong G (2000) Pliocene uplift of the northern Tibetan Plateau. *Geology*, **28**, 715.

Zhou YF, Abbott RJ, Jiang ZY *et al.* (2010) GENE FLOW AND SPECIES DELIMITATION: A CASE STUDY OF TWO PINE SPECIES WITH OVERLAPPING DISTRIBUTIONS IN SOUTHEAST CHINA. *Evolution*, **64**, 2342–2352.

Zhou WW, Wen Y, Fu J *et al.* (2012) Speciation in the *Rana chensinensis* species complex and its relationship to the uplift of the Qinghai-Tibetan Plateau. *Molecular Ecology*, **21**, 960–973.

Zhou Y, Zhang L, Liu J, Wu G, Savolainen O (2014) Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. *Molecular Ecology*, **23**, 3504–3522.

Zhu XF, Li Y, Wu GL, Fang ZD, Li QJ, Liu JQ (2009). Molecular and morphological evidence for natural hybridization between *Primula secundiflora* Franchet and *P. poissonii* Franche*t* (Primulaceae). *Acta Biologica Cracoviensia Series Botanica*, **51**, 29–36.

Chapter 3

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

From the article: **Ren G**, Mateo G, Liu J *et al.* (2016) 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.

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

Guangpeng Ren[1,2,3], Rubén G. Mateo[1], Jianquan Liu[3], Tomasz Suchan[1], Nadir Alvarez[1], Antoine Guisan[1], Elena Conti[4], Nicolas Salamin[1,2*]

[1] Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[2] Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland

[3] State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou 730000 Gansu, China

[4] Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland.

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 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 a different way 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.

Keywords: demography, Himalayas, Isolation by Distance, phylogeography, population structure, Quaternary climatic changes, RADseq

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 is 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 post-glacial expansions (Abbott, 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 its 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 meters (Wu, 1987). The eastern Himalayas are associated with deep valleys and characterized mainly by 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 meters; Favre *et al.,* 2014) and the high average altitude (more than 4000 meters). 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 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. Opgenoorth *et al.,* 2010; H. Wang *et al.,* 2010; 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 restriction 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). *Primula tibetica* is an insect-pollinated (mostly by bees), heterostylous, herbaceous, perennial plant that occurs in diverse habitats at elevations ranging from 2600 to 5000 meters (Hu & Kelso, 1996). Its scape is sometimes hidden among the leaves or can be as long as 13 centimeters. *Primula tibetica* is an outcrossing small herb of variable height (2-13 centimeters) 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 indicated that *P. tibetica* originated in the Himalayas after the recent QTP uplift (i.e. 3.4-1.6 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. Schorr *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.

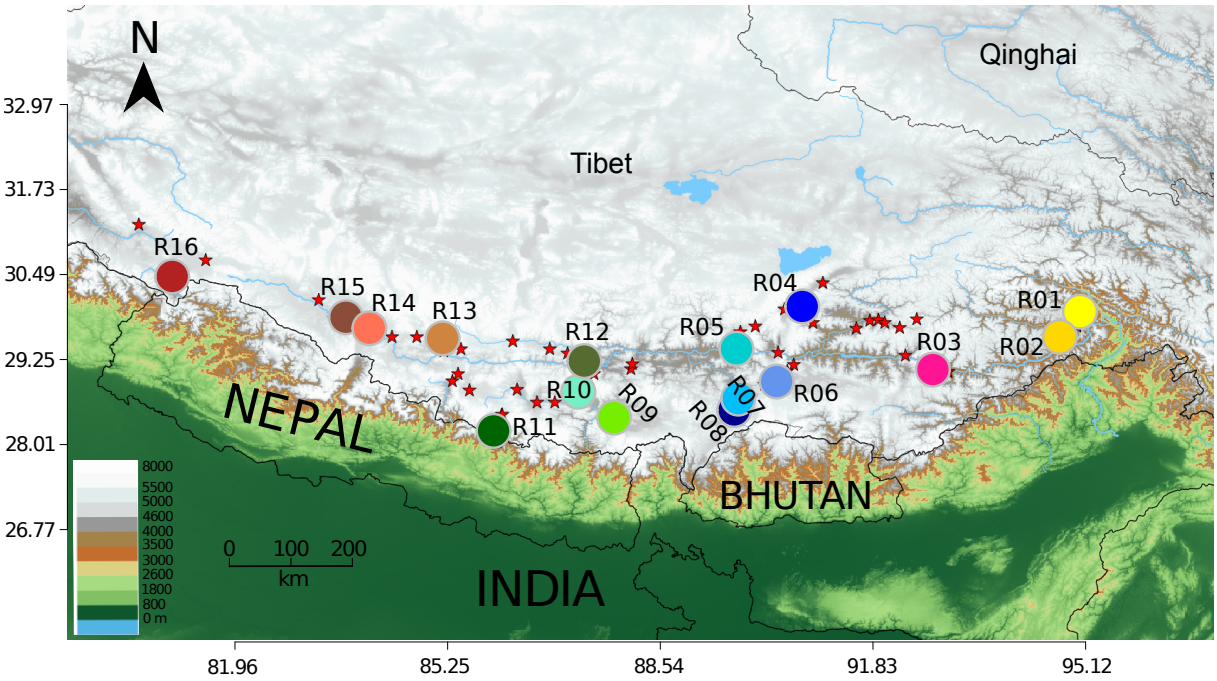


Fig. 1 Sampling locations of all 61 populations of *Primula tibetica* (red stars) and the 16 selected populations (large colored circles) used for genomic analyses in this study.

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 by 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; 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 (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 AG, Hombrechtikon, Switzerland) 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 *Eco*RI and *Mse*I restriction enzymes, followed by the ligation of Illumina adapter sequences and unique 8-base-pair 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 base pairs (bp) using AMPure XP beads with beads:sample ratios 0.8 and 0.2 modified from a protocol in https://www.neb.com/protocols/1/01/01/size-selection-e6270. The libraries were sequenced using single-end reads of 100 bp of 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 catalogue in *cstacks* and matched each sample against the catalogue 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; *-t* flag to remove or break up highly repetitive RAD-tags during the *ustacks* component and upper bound of error rate ε= 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 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_lim -10.0 --lnl_dist*. 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: i) using the same settings as above except for $M = 3$ and $n = 3$, and trim the reads to 90 bp in length ($M = 3$, $n = 3$ and 90bp), and ii) $M = 5$, $n = 3$ and 90bp. The results of the population structure analyses based on the three datasets were qualitatively similar (Fig. 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 catalogue of reads using the *populations* module to produce data sets 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 homeologs were excluded by removing loci showing heterozygosity > 0.5 within samples (Hohenlohe *et al.,* 2011). We further filtered our data set 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 (π), Wright's *F*-statistic (F_{IS}) 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 & Tienderen, 2004), and significance was determined using 1×10^4 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 1×10^5 generations for the burnin and 2×10^5 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, 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). We further performed a principal components analysis (PCA) to visualize the major axes of variation of the population genetics using the *adegenet* package (*glPCA* 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 nutans* and *P. 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 $t_0 = \sqrt{1 - D}$, where *D* is the minimum sum of the squared Euclidean distances between the two maps, scaled to range between 0 and 1 (C. Wang *et al.,*

2010; 2012). We used the *protest* function in *vegan* to test, using 1×10^5 permutations, the probability of observing a similarity statistic higher than the observed t_0 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 1×10^5 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 approximate Bayesian computation (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 data set was generated following the algorithm proposed by Hudson (2002). We further chose MAF = 0.01 to increase the mean level of genetic variation of both the observed and simulated data sets and to reduce the proportion of loci which 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 3×10^6 simulated data sets (on average 10^6 per scenario). We used 1% of the simulated data sets 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 one year 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 one year 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 Note 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 less missing data and it saved computational time. We tested the following scenarios of demographic changes: i) continuous expansion since divergence; ii) recent expansion; iii) expansion followed by shrinkage; iv) expansion followed by shrinkage and a new expansion event (Fig. S7a; Wang *et al.,* 2016). We used the same strategy as above to choose the most likely scenario and estimate the parameters of interest.

Species distribution models

An ensemble of species distribution models (SDM, 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 with MAXENT as a fourth technique, and explanations therein; Fig. S8). A total of 58 species occurrences obtained directly from the filed collections were used as presences 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 AUC and 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 in 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 AUC statistic. The consensus model was then projected onto different past climatic periods using the data available in the Worldclim dataset: 1) the last interglacial (LIG; 0.12-0.14 Ma), 2) the last Maximum Glacial (LGM; 0.022 Ma), and 3) 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 more than 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 3,509 RAD loci containing 8,930 SNPs that could be used for population genetics analyses. The data set was used to estimate historical scenarios of *P. tibetica* containing 4,882 single-SNP loci. Finally, four data sets containing 8,579, 5,401, 7,777, 10,431 single-SNP loci were used to estimate the changes in population sizes of groups E, C, CE and W, respectively.

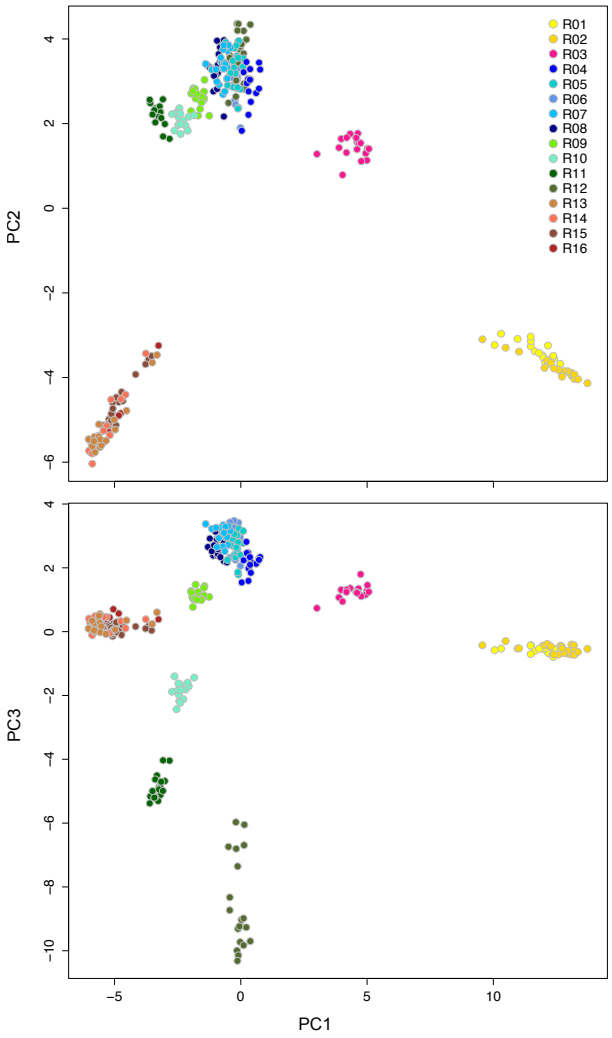


Fig. 2 Distribution of individuals of *Primula tibetica* along principal component (PC) scores (PC1, 20% vs PC2, 9.4%; PC1 vs PC3, 6.6%) of genetic variation based on the analysis of single nucleotide polymorphism (SNP) dataset; individuals are color-coded according to their population identities (see Fig. 1).

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 a large degree of 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 (Fig. 1, 2). This pattern of population structure was also supported by the STRUCTURE analysis, which best explained the data with a K equal to 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 Note 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_value < 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.

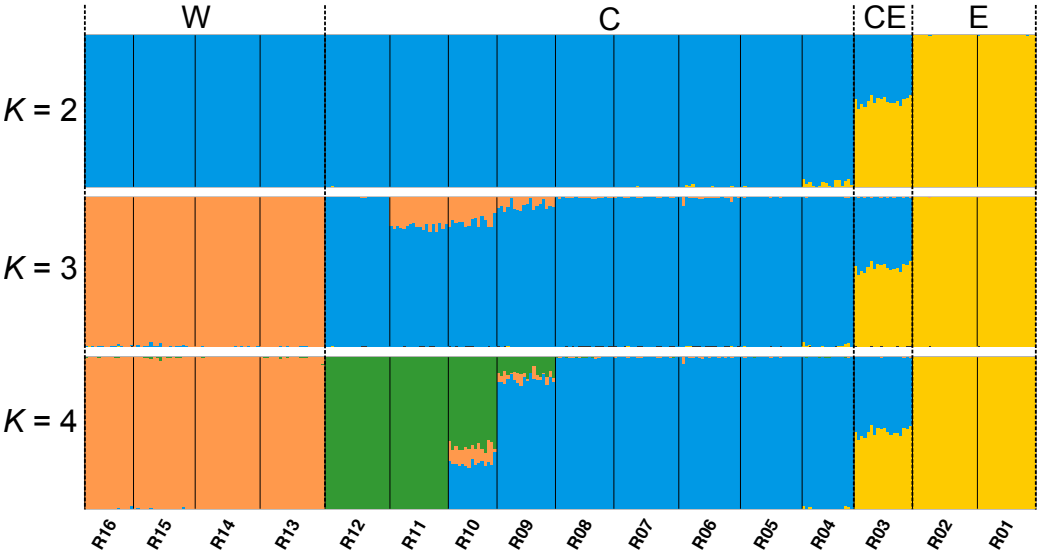


Fig. 3 Plots of posterior probabilities for individuals of *Primula tibetica* assigned to K genetic clusters from STRUCTURE analyses for K = 2–4. Populations are delimited by black lines, with the corresponding population names listed along the bottom of the plot. The four groups are delimited by a dashed black line. E, eastern group; CE, central- eastern group; C, central group; W, western group.

Genetic diversity and Isolation by Distance

The average within-population genetic diversity π ranged from 0.0011 to 0.0044, when considering all genetic positions, including those not polymorphic anywhere in the data set (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).

Table 1 Population summary statistics calculated for the 3509 RADSeq loci. Included are the average number of individuals genotyped at each locus (N), the proportion of polymorphic 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.

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

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 larger than 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 above). 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 mantel test among populations of group C (r = 0.51, *P*_value = 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*_value = 0.042; Fig. 5b).

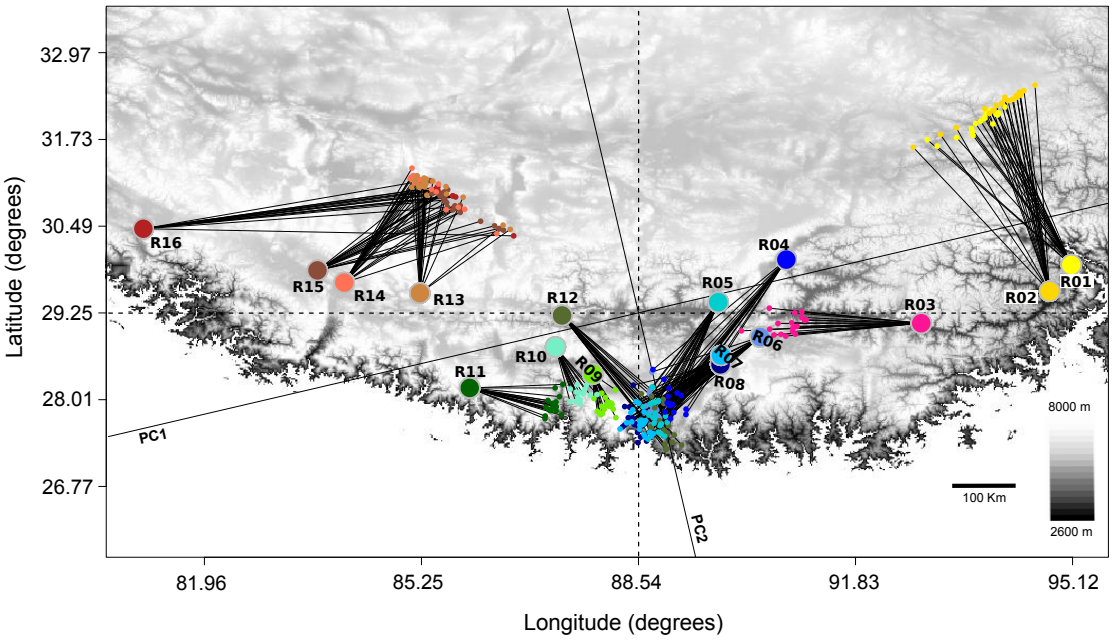


Fig. 4 Procrustes-transformed principal component analysis (PCA) plot of genetic variation with each individual of *Primula tibetica* mapped in PC space (the small circles) relative to the geographic location of populations (the larger circles). Black lines show the orientation of PC1 and PC2 for the genetic data (explaining 20% and 9.4% of the genetic variation, respectively) relative to the geographic longitude and latitudinal axes. The length of the line connecting individuals in PC space to their geographic location represents the extent of the deviation from the expected pattern of genetic variation based on geography.

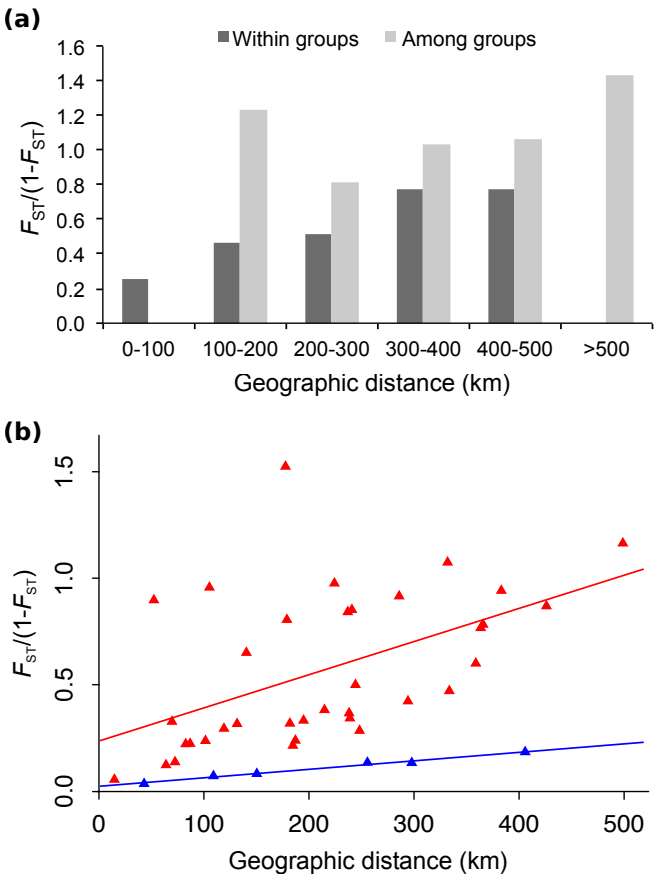


Fig. 5 (a) Averaged pairwise genetic differentiation between populations (FST) within and among genetic clusters for *Primula tibetica* based on six categories of geographic distances. (b) Correlations between pairwise genetic differentiation among populations (FST) within the central group (C; red; r = 0.51, P = 0.016) or the western (W) group (blue; r = 0.99, P = 0.042) and the geographic distance between populations.

Estimates of historical demography

ABC 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).

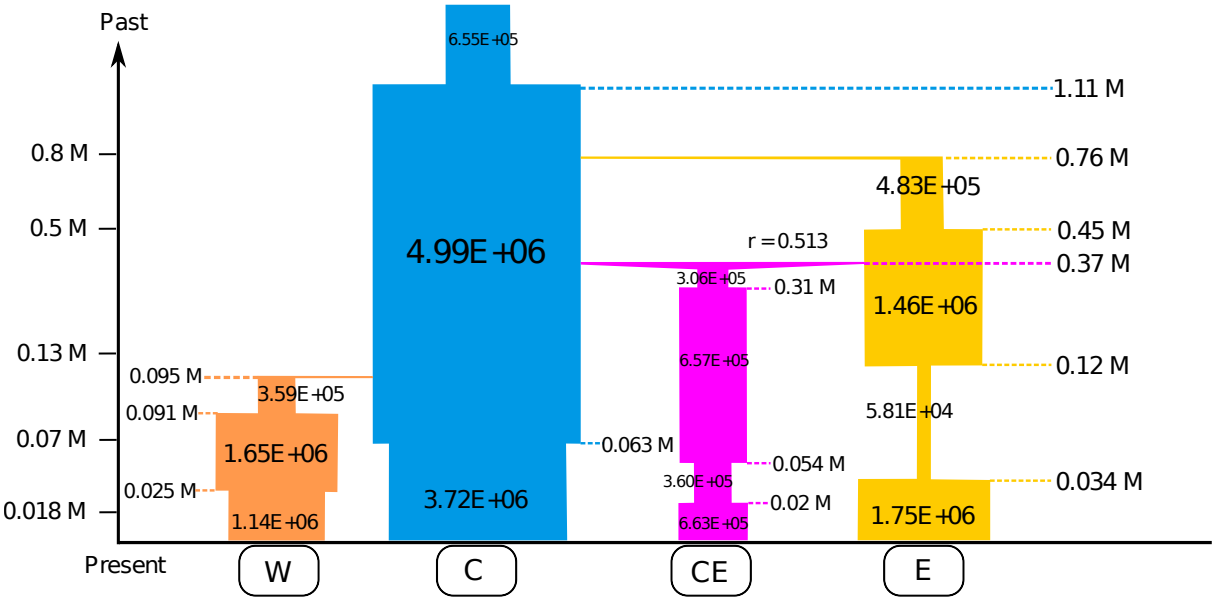


Fig. 6 Summary of inferred demographic history of the four genetic clusters of *Primula tibetica*. Changes in population sizes are integrated into the divergent scenario. Times on the vertical axis represent the glaciation periods that occurred in the Qinghai-Tibet Plateau (QTP) (Zheng *et al.,* 2002). Population sizes are indicated on each cylinder. Times of divergence and changes in population sizes are indicated by horizontal dashed lines. Only the mean values are shown (see Supporting Information Tables S6, S7 for 95% highest posterior density for all values). E, eastern group; CE, central-eastern group; C, central group; W, western group.

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 ca. 1.11 Ma (95% highest posterior density (HPD): 0.53-1.65 Ma; Table S6), followed by a slight bottleneck around 0.063 Ma (HPD: 0.007-0.136 Ma). Group E diverged from the ancestral populations ca. 0.76 Ma (HPD: 0.49-0.96 Ma; Table S7). It started to expand until ca. 0.45 Ma (HPD: 0.15-0.92 Ma), before experiencing a severe bottleneck that decreased by about 25 times its population size around 0.12 Ma (HPD: 0.063-0.2 Ma). Then it quickly expanded just before LGM around 0.037 Ma (HPD: 0.011-0.078 Ma) and reached the previous level of 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 around 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, ca. 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 in regards to AUC (0.996) and TSS (0.998) values. Current potential distribution based on the three threshold approaches predicted similar results, but the 5% omission error yielded generally a better representation of the actual distributions of the species, we therefore presented all results based on the 5% omission threshold. The paleo-climatic conditions of LIG predicted large differences in annual mean precipitation in the Himalayas compared with the ones observed at either 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 to current conditions, but larger distributions than the prediction at the LGM (Fig. 7, S10). During the LGM, the three GCMs yielded similar patterns but fragmented palaeodistributions of *P. tibetica* (Fig. 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.

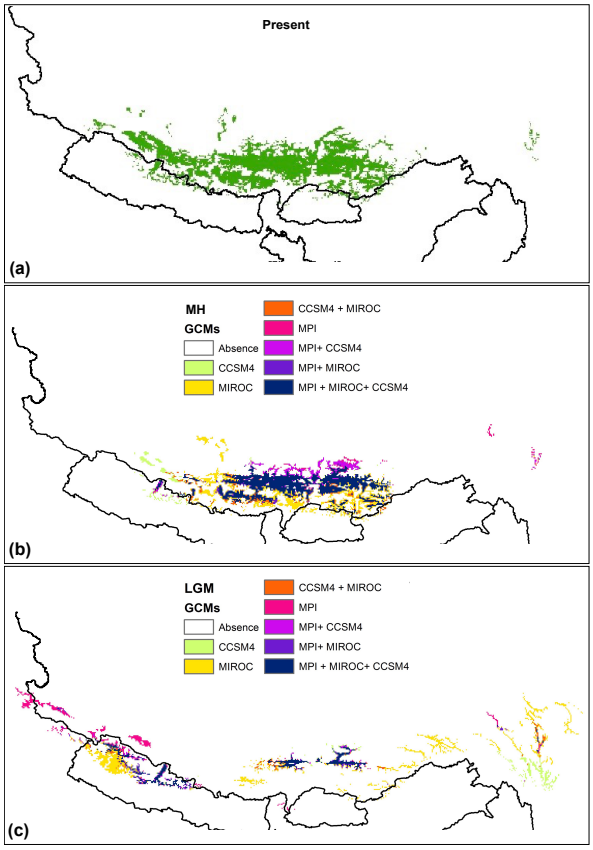


Fig. 7 Habitat suitability of Primula tibetica predicted by species distribution models (SDMs) for present (a), mid-Holocene (MH) (b) and last Maximum Glacial (LGM) (c) using three techniques. SDMs for theMH and LGM are based on three climatic models. GCMs, general circulation models.

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. Isolation by Distance 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 less than 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 Isolation by Distance

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 (Fig. 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, likely 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 (Opgenoorth *et al.,* 2010; H. Wang *et al.,* 2010; Jia *et al.,* 2011), the clear pattern identified by our genomic-level data was not yet described in the region. For example, Opgenoorth *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 (cryptic) 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 ($t_0 = 0.815$, P_value $< 10^{-5}$; Fig. 4), which is likely due to the large-scale spatial genetic structure shaped by the refugium-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 centimeters) 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 likely caused fragmentation, reduced gene flow and further reinforced the

genetic structure (Liu *et al.,* 2014, Wen *et al.,* 2014). Isolation by Distance 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 likely due to 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 populations (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 P. 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 been already 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 timeframe 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 *et al.,* 2002; Zheng *et al.,* 2002). Such 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 ca. 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 still prevailed in the QTP until 0.17 Ma (Shi *et al.,* 2002). The old expansion of group E may have been favored by such 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 last interglacial 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 P. 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 ca. 1.11 Ma (HPD: 0.53-1.65 Ma), which indicates that the origin of this species likely 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 likely 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 last interglacial 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 more than 8°C in comparison with the region of the central populations (assuming that current temperature in the Himalayas decreases by 0.64 °C/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). Evidence supporting the reduction of population size during warmer periods further comes from the current and MH SDMs that 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 (Fig. 6, 7). This period represents however a small timescale (i.e. 18,000-25,000 years) and 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 last interglacial period before retreating to a southwestern refugium during the LGM (Fig. 6, 7). The warm climate during the last interglacial 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 western Himalayas during warmer periods is also supported by the comparison of the SDMs between the present and MH, where more areas were predicted at present than 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 in 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), make clear that such species have exhibited different range dynamics (i.e. population persistence at 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, the grant 31003A_138282 from the Swiss National Science Foundation, the China Scholarship Council (award to GR for four 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. JL was funded by the National Natural Science Foundation of China (grant numbers 3159820011), Ministry of Science and Technology of the People's Republic of China (2010DFA34610) and International Collaboration 111 Projects of China. RGM was funded by a Marie Curie Intra-European Fellowship within the 7th European Community Framework Programme (ACONITE, PIEF-GA-2013-622620). NA was funded by the Swiss National Science Foundation (grant PP00P3_144870).

Author contributions

G.R., N.S. and E.C. planned and designed the research. G.R. carried out the sampling and the lab work, performed the molecular analysis. T.S. and N.A. participated in the initial test of the lab 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.

Supporting Information

Additional supporting information may be found in http://onlinelibrary.wiley.com/store/10.1111/nph.14221/asset/supinfo/nph14221-sup-0001-SupInfo.pdf?v=1&s=38f7d181ee2183d0cfe316dc6f2c790838f2d002

References

Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J. 2000. Molecular analysis of plant migration and refugia in the Arctic. *Science* **289**: 1343–1346.

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.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN. 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences, USA* **103**: 12447–12450.

Avise JC. 2000. *Phylogeography: the History and Formation of Species.* Cambridge, Massachusetts: Harvard University Press.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**: 171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.

Comes HP, Kadereit JW. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* **3**: 432–438.

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, 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.

Cornuet JM, Ravigne 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.

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology* **22**: 3151–3164.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499–510.

Davis MB, Shaw RG. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science* **292**: 673–679.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Marquez JRG, Gruber B, Lafourcade B, Leitao PJ *et al.* 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**: 27–46.

Earl DA, vonHoldt BM. 2012. Structure Harvester: a website and program for visualizing Structure output and

implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences, USA* **107**: 16196–16200.

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.

Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, Muellner-Riehl AN. 2014. The role of the uplift of the Qinghai-Tibetan plateau for the evolution of Tibetan biotas. *Biological Reviews* **90**: 236–253.

Ferchaud A, Lyet A, Cheylan M, Arnal V, Baron JP, Montgelard C, Ursenbacher S. 2010. High genetic differentiation among French populations of the Orsini's viper (*Vipera ursinii ursinii*) based on mitochondrial and microsatellite data: implications for conservation management. *Journal of Heredity* **102**: 67–78.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation* **24**: 38–49.

Filatov DA, Osborne OG, Papadopulos AST. 2016. Demographic history of speciation in a Senecio altitudinal hybrid zone on Mt. Etna. *Molecular ecology* **25**: 2467–2481.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the PhyML 3.0. *Systematic Biology* **59**: 307–321.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. *Ecological modelling* **135**: 147–186.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B* **359**: 183–195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources* **11**: 117–122.

Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST}. *Nature Reviews Genetics* **10**: 639–650.

Hoorn C, Wesselingh FP, ter Steege H, Bermudez MA, Sevink J, Sanmartin I, Sanchez-Meseguer A, Anderson CL, Figueiredo JP, Jaramillo C *et al.* 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.

Hu CM, Kelso S. 1996. Primulaceae. In: Wu ZY, Raven PH, eds. *Flora of China* Vol 15. Beijing: Science Press; St Louis: Missouri Botanical Garden Press, 99–185.

Hudson R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**:1801–1806.

Jia DR, Abbott RJ, Liu TL, Mao KS, Bartish IV, Liu JQ. 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.

Jia DR, Liu TL, Wang LY, Zhou DW, Liu JQ. 2011. Evolutionary history of an alpine shrub *Hippophae tibetana* (Elaeagnaceae): Allopatric divergence and regional expansion. *Biological Journal of the Linnean Society* **102**: 37–50.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24** 1403–1405.

Kiehl JT, Gent PR. 2004. The community climate system model, version 2. *Journal of Climate* **17**: 3666–3682.

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL. 2015. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology* **24**: 3688–3705.

Lexer C, Wüest RO, Mangili S, Heuertz M, Stölting KN, Pearman PB, Forest F, Salamin N, Zimmermann NE, 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.

Li L, Abbott RJ, Liu BB, Sun YS, Li LL, Zou JB, Wang X, Miehe G, Liu JQ. 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.

Li JJ, Shi YF, Li BY. 1995. *Uplift of the Qinghai–Xizang (Tibet) Plateau and Global Change*. Lanzhou: Lanzhou University Press.

Li WH, Zhang YG. 2010. *The vertical climate of Hengduan Mountains region and its impact on forest distribution.* Beijing, China: China Meteorological Press.

Liu JQ, Duan YW, Hao G, Ge XJ, Sun H. 2014. Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematic and Evolution* **52**: 241–249.

Liu JQ, Wang YJ, Wang AL, Hideaki O, Abbott RJ. 2006. Radiation and diversification within the *Ligularia-Cremanthodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution* **38**: 31–49.

Liu J, Moller M, Provan J, Gao LM, Poudel RM, Li DZ. 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist* **199**: 1093–1108.

Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular ecology resources* **15**: 28–41.

Mateo RG, Felicísimo ÁM, Pottier J, Guisan A, Muñoz J. 2012. Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, **7**, e32586.

Mayr E. 1963. *Animal Species and Evolution*. Cambridge: Harvard University Press.

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792–794.

Meng LH, Yang R, Abbott RJ, Miehe G, Hu TH, Liu JQ. 2007. Mitochondrial and chloroplast phylogeography of *Picea crassifolia* Kom. (Pinaceae) in the Qinghai-Tibetan Plateau and adjacent highlands. *Molecular Ecology* **16**: 4128–4137.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for

conservation priorities. *Nature* **403**: 853–858.

Oksanen J, Blanchet FG, Kindt R et al. 2013. vegan: Community Ecology Package. R package version 2.0-7. Available at http://CRAN.R-project.org/package=vegan.

Opgenoorth L, Vendramin GG, Mao KS, Miehe G, Miehe S, Liepelt S, Liu JQ, 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.

Owen LA. 2009. Latest Pleistocene and Holocene glacier fluctuations in the Himalaya and Tibet. *Quaternary Science Reviews* **28**: 2150–2164.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and nonmodel species. *PLoS ONE* **7**: e37135.

Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M *et al.* 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

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.

R Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ren GP, 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.

Richards J. 2003. *Primula* Second Ed. Portland, OR, USA: Timber Press.

Rice WR, Hostert EE. 1993. Laboratory experiments on speciation: what have we learned in 40 years. *Evolution* **47**: 1637–1653.

Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137–138.

Schorr G, Holstein N, Pearman PB, Guisan A, Kadereit JW. 2012. Integrating species distribution models (SDMs) and phylogeography for two species of Alpine *Primula*. *Ecology and Evolution* **2**: 1260–1277.

Schorr G, Pearman PB, Guisan A, Kadereit JW. 2013. Combining palaeodistribution modelling and phylogeographical approaches for identifying glacial refugia in Alpine *Primula*. *Journal of Biogeography* **40**: 1947–1960.

Shi YF. 2002. Characteristics of Late Quaternary monsoonal glaciation on the Tibetan Plateau and in East Asia. *Quaternary International* **97**: 79–91.

Shi YF, Li JJ, Li BY. 1998. *Uplift and Environmental Changes of Qinghai–Tibetan Plateau in the Late Cenozoic.* Guangzhou: Guangdong Science and Technology Press.

Shimono A, Ueno S, Gu S, Zhao X, Tsumura Y, Tang Y. 2010. Range shifts of *Potentilla fruticosa* on the Qinghai–Tibetan Plateau during glacial and interglacial periods revealed by chloroplast DNA sequence variation. *Heredity* **104**: 534–542.

Sun Y, Li L, Li L, Zou J, Liu JQ. 2015. Distributional dynamics and interspecific gene flow in *Picea likiangensis* and *P. wilsonii* triggered by climate change on the Qinghai‐Tibet Plateau. *Journal of Biogeography* **42**, 475–484.

Thompson LO, Yao T, Davis ME, Henderson KA, Mosley-Thompson E, Lin PN. Beer J, Synal HA, Cole-Dai J, Bolzan JF. 1997. Tropical climate instability: The last glacial cycle from a Qinghai-Tibetan ice core. *Science* **276**: 1821–1825.

Thorpe RS, Surget-Groba Y, Johansson H. 2008. The relative importance of ecology and geographic isolation for speciation in anoles. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3071–3081.

Thuiller W, Lafourcade B, Engler R, Araújo MB. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography* **32**: 369–373.

Wang LY, Abbott RJ, Zheng W, Chen P, Wang YJ, Liu JQ. 2009. History and evolution of alpine plants endemic to the Qinghai-Tibetan Plateau: *Aconitum gymnandrum* (Ranunculaceae). *Molecular Ecology* **18**: 709–721.

Wang IJ, Glor RE, Losos JB. 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters* **16**: 175–182.

Wang Q, Liu J, Allen GA, Ma Y, Yue W, Marr KL, Abbott RJ. 2016. Arctic plant origins and early formation of circumarctic distributions: a case study of the mountain sorrel, *Oxyria digyna*. *New Phytologist* **209**: 343–353.

Wang H, Qiong L, Sun K, Lu F, Wang YG, Song ZP, Wu QH, Chen JK, Zhang WJ. 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.

Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, Singleton AB, Rosenberg NA. 2010. Comparing spatial maps of human population-genetic variation using procrustes analysis. *Statistical Applications in Genetics and Molecular Biology* **9**: 13.

Wang C, Zollner S, Rosenberg NA. 2012. A Quantitative Comparison of the Similarity between Genes and Geography in Worldwide Human Populations. *PLoS Genetics* **8**: e1002886.

Wen J, Zhang JQ, Nie ZL, Zhong Y, Sun H. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in genetics* **5**: 4

Wu CY. 1987. *Flora of Tibet.* Beijing: Science Press.

Yan HF, Zhang CY, Wang FY, Hu CM, Ge XJ, Hao G. 2012. Population expanding with the phalanx model and lineages split by environmental heterogeneity: a case study of *Primula obconica* in subtropical China. *PLoS ONE* **7**: e41315.

Yang FS, Li YF, Ding X, Wang XQ. 2008. Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai-Tibetan Plateau and its correlation with the Quaternary climate change. *Molecular Ecology* **17**: 5135–5145.

Yao T, Pu J, Lu A, Wang Y, Yu W. 2007. Recent glacial retreat and its impact on hydrological processes on the Tibetan Plateau, China, and surrounding regions. *Arctic, Antarctic, and Alpine Research* **39**: 642–650.

Yin A, Harrison TM. 2000. Geologic evolution of the Himalayan-Tibetan orogen. *Annual Review of Earth and Planetary Sciences* **28**: 211–280.

Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ. 2005. Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology* **14**: 3513–3524.

Zheng D. 1996. The system of physico-geographical regions of the Qinghai-Xizang (Tibet) plateau. *Science in China Series D-Earth Sciences* 39: 410–417.

Zheng BX, Xu QQ, Shen YP. 2002. The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quaternary International* **97–98**: 93–101.

Zhou Z, Hong D, Niu Y, Li G, Nie Z, Wen J, Sun H. 2013. Phylogenetic and biogeographic analyses of the Sino-Himalayan endemic genus *Cyananthus* (Campanulaceae) and implications for the evolution of its sexual system. *Molecular Phylogenetics and Evolution* **68**: 482–97.

Chapter 4

Striking population genetic structure and demographic history of *Primula fasciculata* in a highly fragmented biodiversity hotspot

This chapter will be submitted to Molecular Ecology soon

Striking population genetic structure and demographic history of *Primula fasciculata* in a highly fragmented biodiversity hotspot

Guangpeng Ren[1,2,3,4], Elena Conti[5], Nicolas Salamin[1,2,3*]

[1] Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[2] Department of Computational Biology, Biophore, University of Lausanne, 1015 Lausanne, Switzerland

[3] Swiss Institute of Bioinformatics, Quartier Sorge, 1015 Lausanne, Switzerland

[4] State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou 730000 Gansu, China

[5] Department of Systematic and Evolutionary Botany and Botanic Garden, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland.

Abstract

Understanding the factors that drive genetic structure of a species and how it responded to past climatic changes is an important first step in modern population management. This is especially true for species inhabiting in highly heterogeneous mountains. Recent advances in population genomics hold great promise to detect cryptic structure and obtain more accurate estimates of demographic parameters, potentially revolutionizing the way genetic data are used to manage wild populations. Here we investigated population structure and demographic history of *Primula fasciculata* that occurs in a highly fragmented biodiversity hotspot, the Hengduan Mountains. We obtained genomic data for 234 samples of the species using restriction site-associated DNA (RAD) sequencing and combined approximate Bayesian computation (ABC) modeling and species distribution modeling (SDM). The structure analyses showed that *P. fasciculata* displays a striking population genetic structure and six genetic lineages were identified. ABC modeling suggested that the current lineages diverged from an ancestral lineage in the eastern Hengduan Mountains after the largest glaciation occurred in the region. Each of them has experienced expansions and/or bottlenecks since their divergence during or across the following Quaternary glacial cycles. By contrast, the last glacial maximum (LGM) had little effects on these genetic lineages. Our study demonstrates the usefulness of population genomics with environmental variables for evaluating the effects of past climatic changes in alpine plant species with shallow population structure.

Keywords: demography, genetic structure, Hengduan Mountains, population genomics, Quaternary climatic changes

Introduction

Plant populations are not randomly arranged assemblages of genotypes but are structured in space and time (Loveless & Hamrich 1984). Because of the limited mobility of plants, their genetic structure

implies spatial structure, where genetic differentiation increases with geographic distance (Wright 1943). Yet, recent empirical studies have put forward that geographic distance by itself fails to fully explain genetic variation in natural systems (e.g. Shafer & Wolf 2013). In fact, geographic, environmental and historical factors have been suggested to act as drivers of spatial genetic patterns simultaneously at different spatial scales (Wang *et al.* 2012, Muñoz-Pajares *et al.* 2016, Ren *et al.* 2017). Identifying the genetic structure of a plant species and the factors that drive it is then an important step not only in understanding speciation, adaptation and genetic change (Antonovics 1968), but also in population management. In the latter case, the dynamic spatio-temporal histories of populations can profoundly impact their future evolutionary potential (e.g. Lanier *et al.* 2015). This is especially true for climate-sensitive species inhabiting highly fragmented environments, such as mountain ranges.

One of the key high-altitude biodiversity hotspots in the world were these processes could be studied are the Hengduan Mountains. They were formed by a recent uplift of the Qinghai-Tibet Plateau (QTP) during the late Miocene and Pliocene (Myers *et al.* 2000; Li & Fang 1999; Zheng *et al.* 2000; Mulch & Chamberlain 2006). The origin and maintenance of the high biodiversity in this region are suggested to result from its specific topographic feature and profound ecological heterogeneity created by the historical orogenesis and associated climatic changes (Wu 1987). Today, the region is characterized by parallel and deep north-south oriented valleys surrounded by high mountain peaks (Fig. 1). The mountains display drastic altitudinal variations ranging from 1000 m to numerous peaks above 6000 m, and the area is particularly vulnerable to climate change (Zheng 1996; Yao *et al.* 2007). With such a complex geological, climatic and ecological diversity, the region has attracted attention of numerous biologists to study the factors affecting species diversification and evolution (reviewed in Qiu *et al.* 2011; Liu *et al.* 2014; Wen *et al.* 2014). Some studies focused on species-level diversification resulted from the uplift of the QTP (e.g. Liu *et al.* 2002, 2006; Ren *et al.* 2015), while others looked at intraspecific levels to investigate the effects of past geological events and Quaternary climatic oscillations on population genetic structure (e.g. Li *et al.* 2013; Liu *et al.* 2013). However, because of the limited genetic information used in previous studies (but see Li *et al.* 2013), a comprehensive understanding of the factors triggering current genetic structure in this region is scarce.

Integrative approaches combining population genomics (e.g. Ren *et al.* 2017) with species distribution modeling (SDMs, Guisan & Zimmermann 2000) have shown excellent results to understand current spatial genetic patterns and the processes behind. Population genomic data can provide accurate estimates of genetic structure (Avise 2010; Narum *et al.* 2013) and increased accuracy when estimating demographic parameters (e.g. Emerson *et al.* 2010; Bourret *et al.* 2013; Larson *et al.* 2013; Lanier *et al.* 2015; Izuno *et al.* 2016; Ren *et al.* 2017), whereas species distribution models allow predicting geographic areas being part of the ecological niche of species at different temporal and

spatial scales. A recent study based on these approaches has contributed a significant advance in understanding of how alpine plant species responded to the Quaternary climatic changes in a adjacent region, the Himalayas (Ren *et al.* 2017). Although the Next-Generation Sequencing (NGS) methods became cost-effective, the application of population genomics on the taxa distributed in the Hengduan Mountains remains rare because of its remoteness and inaccessibility, and consequently, such genomic level studies are particularly needed for this region to provide a better understanding of evolutionary history of species.

Here we focus on *Primula fasciculata* (Primulaceae), one of the most widely distributed alpine plant species in the Hengduan Mountains (Hu & Kelso 1996). It is an insect-pollinated, heterostylous, herbaceous, perennial plant that occurs in diverse habitats at elevations ranging from 2900 to 5000 m. As an outcrossing small herb of variable height (2-10 centimeters), *P. fasciculata* disperses its seeds largely by gravity and usually grows in wet meadows or along hill-streams (Hu & Kelso 1996; Richards 2003). A recent study indicated that *P. fasciculata* originated from its closely related species *P. tibetica* during the Pliocene uplift of the Hengduan Mountains and experienced expansion during the Quaternary (Ren *et al.* submitted). However, the effects of Quaternary climatic oscillations on its intraspecific divergence and demographic history are unknown. Here, we use an integrative approach combining genomic phylogeography with niche modeling to address these issues. The aims of our study are to: i) identify the population structure of this species and understand the factors driving it; and ii) reconstruct a detailed demographic history of *P. fasciculata* and combine species distribution models with ABC modeling to evaluate the effects of Quaternary climatic changes on its demographic history.

Materials and methods

Dataset

The same dataset as the one assembled in Chapter 2 for *P. fasciculata* was used here. It comprises 234 individuals from 12 populations sampled throughout the Hengduan Mountains. We removed the 140 outlier SNPs found in this species (Chapter 2; Ren *et al.* submitted) to obtain estimates of neutral population genetic structure based on 5946 single-SNP loci.

Characterization of population genetic structure

Population genetic structure of *P. fasciculata* was estimated by using the Bayesian method implemented in STRUCTURE 2.3.4 (Pritchard *et al.* 2000) and a principal components analysis (PCA). Structure analyses were performed under the "Admixture model" and the "Correlated allele frequency model" with K-values ranging from 1 to 12. Ten independent runs were performed for each value of K using 1×10^5 generations for the burnin and 2×10^5 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 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). PCA was performed with the *glPCA* function in *adegenet* package (Jombart 2008) in R to visualize the major axes of variation of the population genetics.

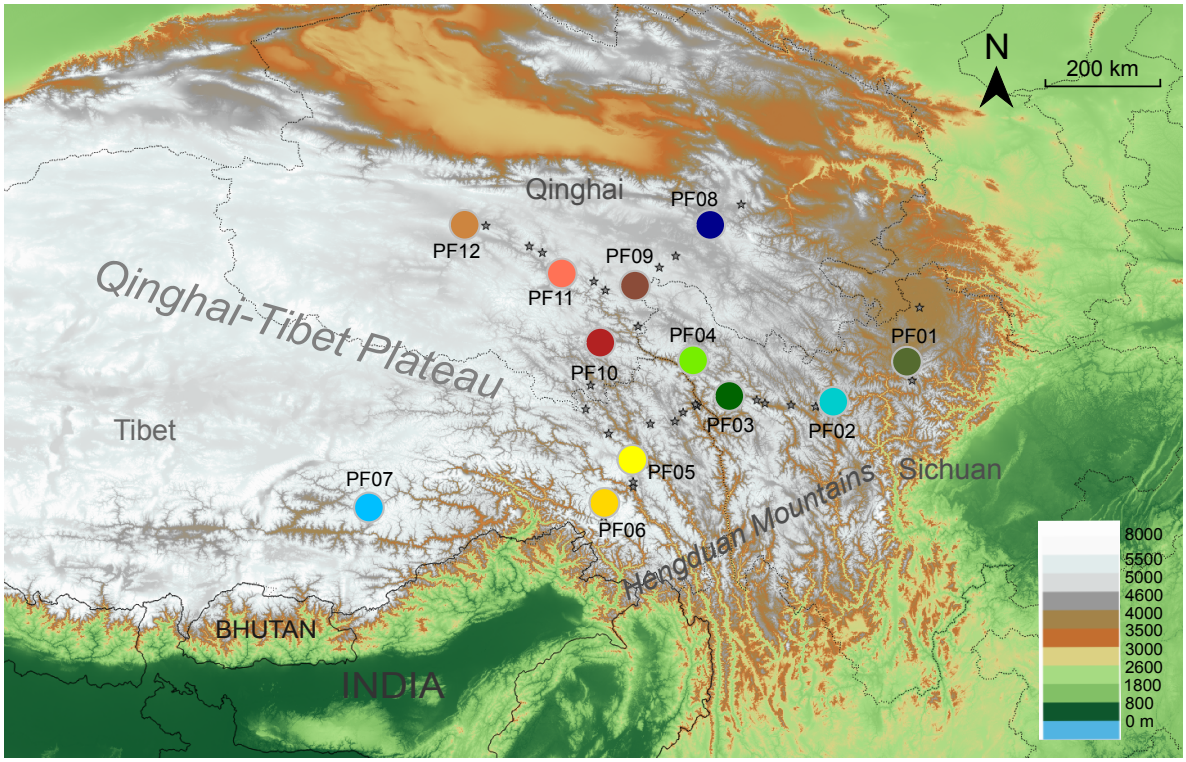


Fig. 1 Sampling locations of all 61 populations of *P. fasciculata* (grey stars) and the 12 selected populations (large colored cycles) used for genomic analyses in this study.

Pairwise F_{ST} values and analysis of molecular variance (AMOVA) among populations were calculated in GENODIVE v.2.0b27 (Meirmans & Tienderen 2004), and significance was determined using 1×10^4 permutations. AMOVA for populations that were further clustered into several groups based on STRUCTURE and PCA results (see Results; Table 1) was applied to evaluate which grouping strategy explains the most percentage of total variance among groups.

The first three components of the PCA performed on the genetic data and the geographic coordinates (latitude and longitude) of the 12 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 t_0 (Wang *et al.* 2010; 2012). We used the *protest*

function in *vegan* to test the probability of observing a similarity statistic higher than the observed t_0 if no geographic pattern is assumed using 1×10^5 permutations (Wang *et al.* 2012).

We used BARRIER v2.2 (Manni *et al.* 2004) to compute the Monmonier's maximum-difference algorithm for identifying biogeographic boundaries or areas exhibiting the largest genetic discontinuities between population pairs based on pairwise genetic distances (F_{ST}). We randomly selected 5,000 loci from the neutral dataset 100 times to generate 100 F_{ST} distance matrices by using *populations* module in the STACTS v1.30 (Catchen *et al.* 2013). The number of barriers was set to vary from 1 to 10, reflecting their descending order of relative importance ('priority') for genetic dispersion (Manni *et al.* 2004). The robustness of the genetic boundaries was assessed by running BARRIER on the 100 F_{ST} distance matrices.

Estimates of historical demography

To decipher the historical demography of *P. fasciculata*, we estimated historical divergence times, admixture and changes in population sizes among different population groups using ABC modeling. We stratified the procedure in three steps (Fig. S1): (i) we investigated the most likely population tree topologies for the three main population lineages (see Results) that were identified by the STRUCTURE and PCA analyses among 13 scenarios describing all possible topologies (Fig. S1); (ii) we selected the tree topology obtained in (i), then split the three main lineages into six groups (see Results; Fig. S1) to estimate their divergence times between two scenarios; (iii) we tested changes in population sizes of each of the six groups in the recent past among four scenarios (Fig. S1; Ren *et al.* 2017). Five individuals that had the least missing data from each of the 12 populations were selected for steps one and two to reduce computational time. For step three, we used those same five individuals for the two groups that contained multiple populations, whereas all individuals were used for all groups that included only one population.

For each step, we tested different scenarios using DIY-ABC v.2.1.0 (Cornuet *et al.* 2010, 2014). We selected for these analyses a single SNP per locus, which had to be present in i) at least 80% of the individuals from each lineage/group and ii) all lineages/groups. We chose MAF = 0.01 to increase the mean level of genetic variation of both the observed and simulated data sets and to reduce the proportion of loci that may correspond to sequencing errors. The datasets used for the ABC modeling and the distributions of prior probabilities for each modeling are summarized in Table S1. We selected all summary statistics to generate a reference table (on average 10^6 datasets per scenario) and used 1% of the simulated data sets closest to the observed data to estimate the relative posterior probabilities for each scenario via logistic regression. Posterior distributions of historical demographic parameters based on the most likely scenario (Cornuet *et al.* 2010) were estimated. The time parameters are

estimated in generations and converted into years by multiplying by the generation time, which was set to one year (Ren *et al.* 2017).

Species distribution models

An ensemble of SDMs (Guisan & Zimmermann 2000) was generated for *P. fasciculata* following the same approach as applied 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). A total of 89 species occurrences were used as presences data to calibrate the models. We used the 19 bioclimatic variables of Worldclim (http://www.worldclim.org, Hijmans *et al.* 2005) as environmental predictors. The potential distributions of 1) the present; 2) the last Maximum Glacial (LGM; 0.022 Ma), and 3) the mid-Holocene (MH; 0.006 Ma) were estimated. 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).

Results

Structuring of population genetic variation

Although the most possible *K* values of STRUCTURE analyses based on the Δ*K* method of Evanno was *K* = 2, the differences of Δ*K* among *K* values were very small (Fig. 2). The Δ*K* of the second most probable *K* value (*K* = 6) differed from the best one by only 2. Other *K* values (3, 4, 5 and 9) also received considerable support. We decided to show all these *K* values in Fig. 2 and combined them with the PCA results to capture the most reasonable lineages for the ABC modeling.

The first two axes of PCA identified three main genetic lineages and explained 13.04% and 7.76% of the total variation, respectively (Fig. 3). The two southwestern populations (PF05, PF06) and four northwestern populations (PF09-PF12) formed two separate lineages (L1 and L2), while the rest of populations form a third lineage (L3). This was in agreement with the STRUCTURE results when *K* =3 (Fig. 2). The third axis of the PCA (PC3; 6.12% of the total variation) showed a separation of population PF07 from lineage L3, which was also shown when *K* = 4 in the STRUCTURE. We identified this population as G1. Looking at *K* values from *K* = 2 to *K* = 6, population PF08 was always represented as an admixed population, which was labeled as G2. The remaining populations of L3 were grouped as G3. The two populations (L1) that diverged from each other in the PCA (Fig. 3) were identified as lineages G4 and G5. The four northwestern populations form the sixth group (G6) evident both in STRUCTURE and PCA. The three main lineages (L1-L3) were used in step one of the ABC modeling to identify the most likely population tree topology, and the six groups (G1-G6) were used in step two of the ABC modeling to estimate the divergence times among these groups.

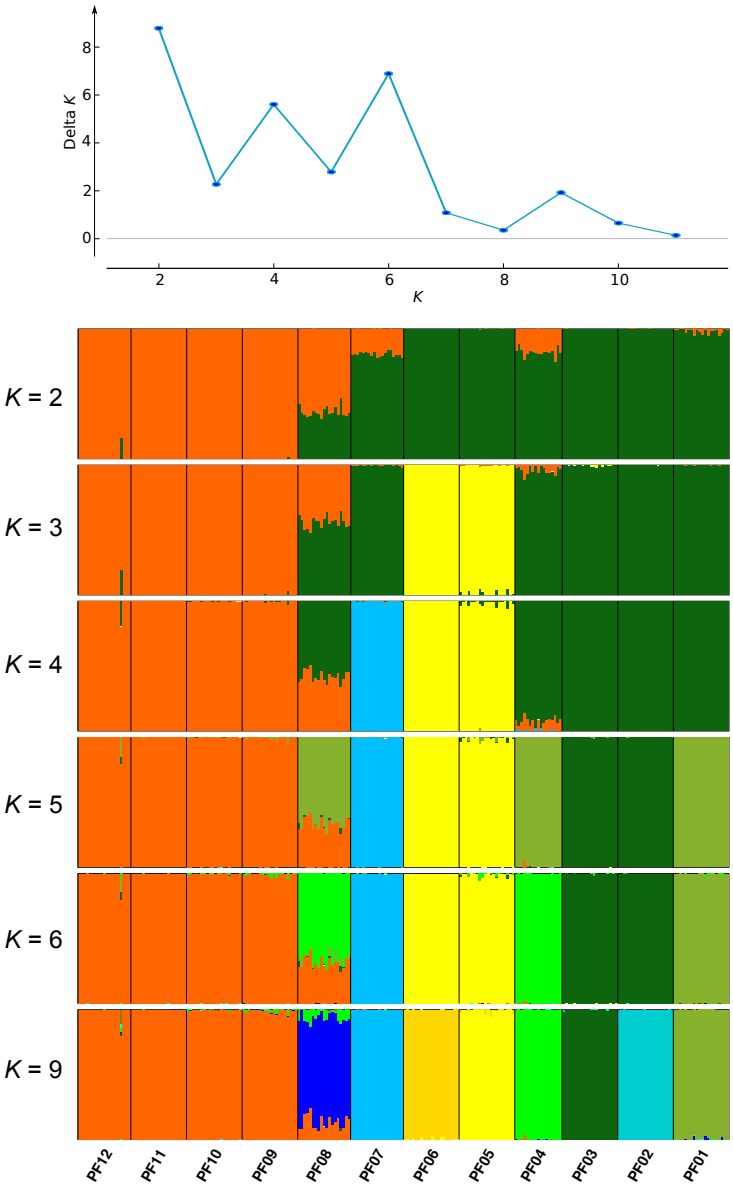


Fig. 2 ΔK values identified using STRUCTURE HARVESTER and plots of posterior probabilities for individuals of *P. fasciculata* assigned to *K* genetic clusters from STRUCTURE analyses for *K* = 2 – 6 and 9. Populations are delimited by black lines, with the corresponding population names listed along the bottom of the plot.

In order to evaluate the reliability of the grouping strategy defined above used for the ABC modeling, we further assigned populations into four, five or seven groups (PF05 and PF06 were separated based on PCA) based on *K*=3, *K*=4 and *K*=5 in the STRUCTURE (Fig. 2), respectively. The strongest signature of population spatial differentiation was obtained (22.3% of total variance; Table 1) when populations were assigned to six groups, which suggest that our grouping strategy for the ABC modeling was reasonable. Differentiation among populations was significant, with F_{ST} values ranging from 0.089 to 0.608 with a mean value of 0.381 (Table S2), which was consistent with AMOVA for the total dataset (F_{ST}=0.306; Table 1).

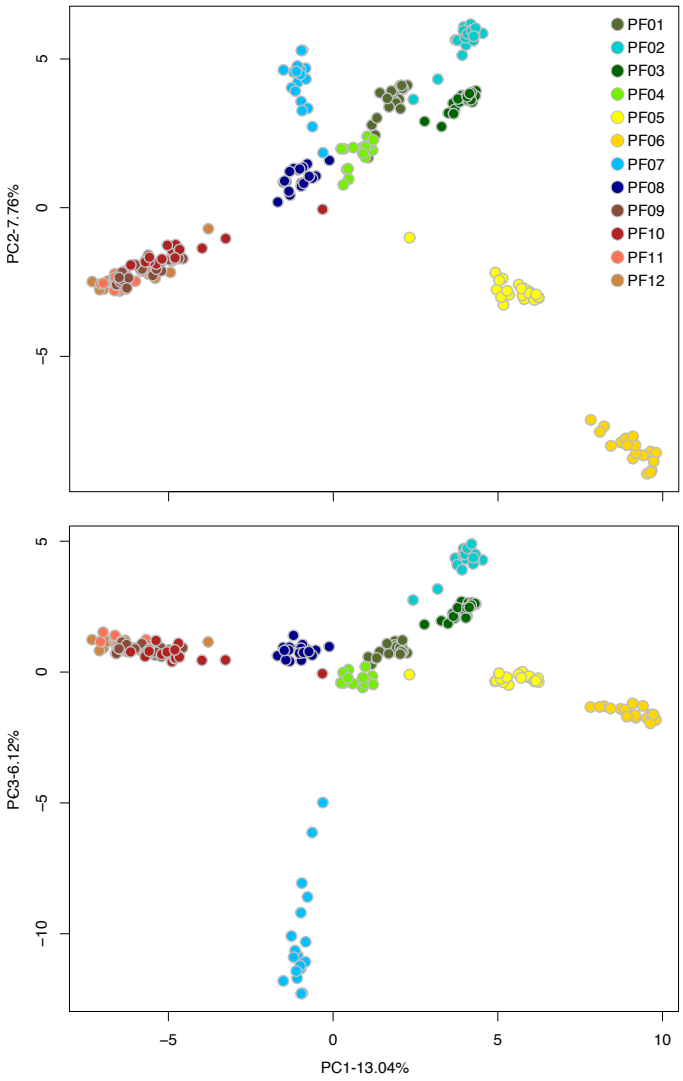


Fig. 3 Distribution of individuals of *P. fasciculata* along PC scores (PC1-13.04% vs. PC2-7.76%; PC1 vs. PC3-6.12%) of genetic variation based on the analysis of SNP dataset; individuals are colour-coded according to their population identities (see Fig. 1).

Procrustes analysis was used to quantify the association between the genetic variation of populations and their geographic locations. The first two PC spaces identified a significant similarity score (t_0 = 0.579, P_value < 10^{-5}), which increased to t_0 = 0.777 when genetic variation in PC1 and PC3 spaces were considered (Fig. 4a). This was caused by the clear separation of the most geographically isolated population PF07 from other populations by the PC3 axis. Individuals from G6 are genetically more similar with each other than would be expected given the geographic distance among the populations forming this group. The general pattern of association with geography for the rest of populations was robust, indicating high level of population divergence. Such level of divergence was also evident in the BARRIER analysis that gave high support to all ten barriers (bootstrap support 100%; Fig. 4b). The presence of such strong barriers between *P. fasciculata* populations indicates an abrupt change in the genetic profile of populations across the species distribution. Although the ranking of the population barriers (Fig. S2) was not in agreement with the STRUCTURE and PCA analyses, the general pattern of spatial genetic structure identified by the BARRIER analysis was consistent with the other analyses.

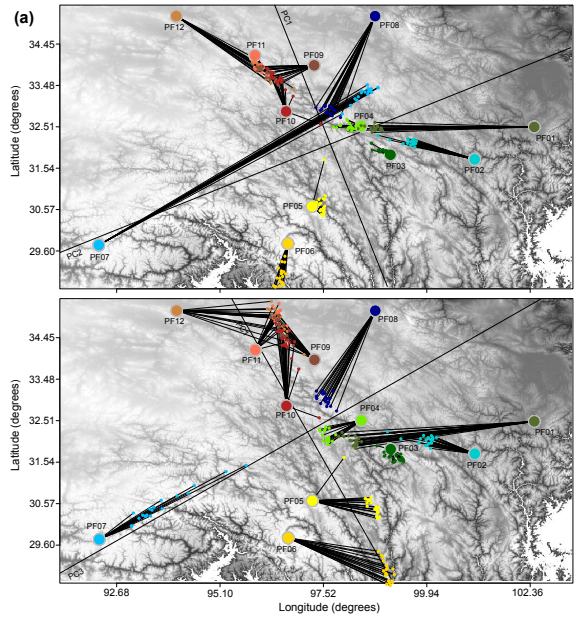
Table 1 AMOVAs for neutral genomic variation based on several groupings of *P. fasciculata* populations.

Grouping/source of variation	df	Variance components	Percentage of total variance	*F*-statistic[a]
Total				
Within populations	222	385.925	69.4	--
Among populations	11	170.004	30.6	F_{ST}=0.306
Three lineages				
Within populations	222	385.246	65.7	F_{ST}=0.343
Among populations	9	108.798	18.6	F_{SC}=0.220
Among lineages	2	91.935	15.7	F_{CT}=0.157
Four groups				
Within populations	222	385.862	66.8	F_{ST} =0.332
Among populations	8	103.728	18.0	F_{SC} =0.212
Among groups	3	88.247	15.3	F_{CT} =0.153
Five groups				
Within populations	222	386.703	66.9	F_{ST} =0.331
Among populations	7	82.003	14.2	F_{SC} =0.175
Among groups	4	109.023	18.9	F_{CT} =0.189
Six groups				
Within populations	222	386.743	66.7	F_{ST} =0.333
Among populations	6	64.083	11.1	F_{SC} =0.142
Among groups	5	128.675	22.2	F_{CT} =0.222
Seven groups				
Within populations	222	385.383	67.7	F_{ST} =0.323
Among populations	5	71.91	12.6	F_{SC} =0.157
Among groups	6	111.893	19.7	F_{CT} =0.197

Abbreviations: AMOVAs, analyses of molecular variance; df = degrees of freedom;
[a] All *F*-values were significant ($P < 0.001$) based on 1000 permutations.

Estimates of historical demography

We used a three-step procedure to estimate the demographic history of *P. fasciculata*. Among the 13 scenarios tested in step one, the scenario depicting an origin of both L1 and L2 from L3, provided the best fit to our data, with posterior probabilities significantly higher than the other scenarios (0.995, 95% credible interval (CI) 0.99, 1.00; Table S3; Fig. S1). According to the main tree topology inferred from step one, the analyses done in step two showed that groups G1, G6 and G4/G5 (i.e. alternative scenarios; Fig. S1) originated from G3, while G2 was formed by admixture between G3 and G6. The scenario where G4 originated from G3 and later G5 diverged from G4 fitted the data much better (0.93, CI: 0.93-0.94; Table S3; Fig. S1). Modeling the changes in population size for each group recovered complicated demographic histories for the six groups of populations. Analyses for groups G3 supported a scenario of "expansion–shrinkage", while groups G2, G4 and G6 were better modeled by a scenario of "expansion–shrinkage–expansion". The other two groups (G1 and G5) were better modeled by a scenario of "recent expansion" (Table S3).



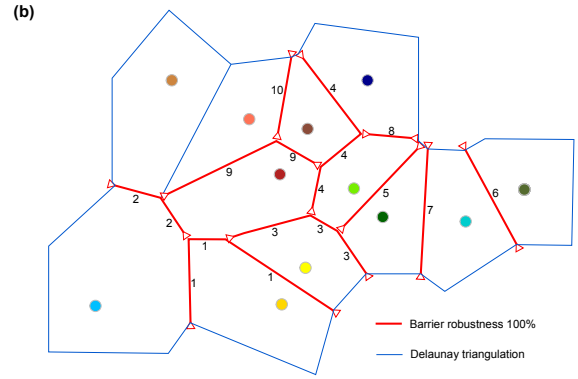


Fig. 4 (a) Procrustes-transformed PCA plot of genetic variation with each individual of *P. fasciculata* mapped in PC space (the small circles) relative to the geographic location of populations (the larger circles). Black lines show the orientation of the genetic space relative to the geographic longitude and latitudinal axes. The length of the line connecting individuals in PC space to their geographic location represents the extent of the deviation from the expected pattern of genetic variation based on geography. (b) Result of BARRIER analysis showing the spatial separation of *P. fasciculata* populations. All the ten barriers (red lines) are highly supported over 100 F_{ST} distance matrixes. Barriers are delimited by small red triangle. Numbers (1-10) represent descending order of relative importance ('priority').

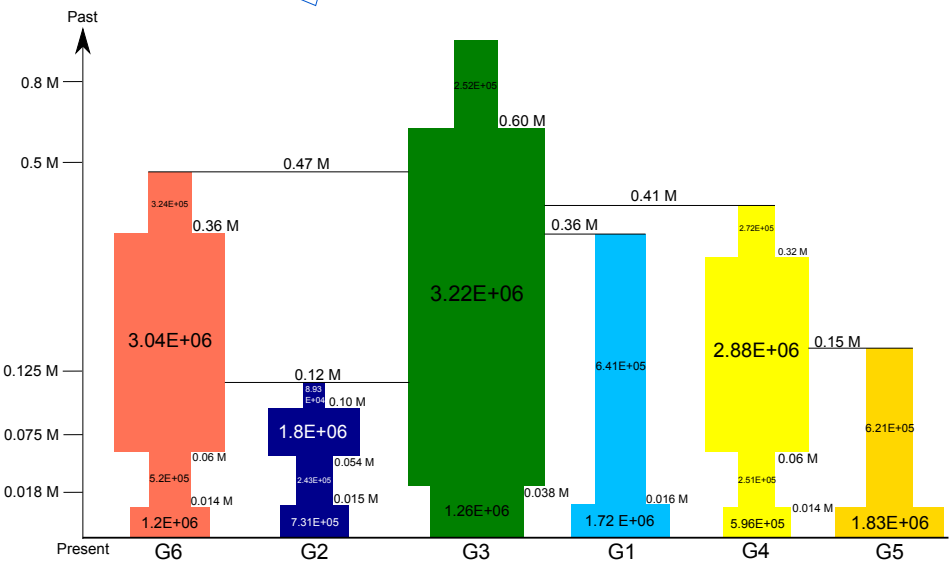


Fig. 5 Summary of inferred demographic history of the six groups of *P. fasciculata*. Changes in population sizes are integrated into the divergent scenario. Times on the vertical axis represent the glaciation periods that occurred in the QTP (Zheng *et al.* 2002). Population sizes are indicated on each square. Times of divergence and changes in population sizes are indicated next to each change in population size. Only the mean values are shown (see Table S4, S5 for 95% credible interval for all values).

We estimated the divergence times and the population sizes as well as the timing and extent of these changes for the six groups. Group G3 was found to be the ancestral population of *P. fasciculata* and started to expand its distribution ca. 0.60 Ma (95% highest posterior density (HPD): 0.27-0.86 Ma; Fig. 5; Table S4), followed by a slight bottleneck around 0.038 Ma (HPD: 0.004-0.075 Ma). G6 diverged from the ancestral populations formed by G3 ca. 0.47 Ma (HPD: 0.38-0.55 Ma; Table S5). It started to expand until ca. 0.36 Ma (HPD: 0.18-0.49 Ma), before experiencing a bottleneck ca. 0.06 Ma (HPD: 0.02-0.09 Ma). Then, it quickly expanded just after the LGM. During the first expansion of this group, it came into secondary contact with the ancestral populations of G3, exchanged genes and resulted in the formation of G2 around 0.12 Ma (HPD: 0.07-0.17 Ma; Table S5). G2 experienced ancient expansion (0.10 Ma) and shrinkage (0.054 Ma) before and during the last glaciation (i.e. 0.015-0.075 Ma), respectively, and a recent expansion after the LGM. G1 diverged from the ancestral populations ca. 0.36 Ma (HPD: 0.23-0.49 Ma) and stayed stable through time before experiencing a recent expansion after the LGM. G4 diverged from G3 ca. 0.41 Ma (HPD: 0.26-0.54 Ma) and started to expand before experiencing a bottleneck during the last glaciation. A recent expansion after the LGM was also detected for this group. G5 was isolated from the ancient expansion of G4 (0.15 Ma, HPD: 0.09-0.21 Ma; Table S5) and experienced a recent expansion after the LGM.

Species distribution models

The predictions to MH conditions based on three GCMs (CCSM4, MIROC and MPI) for this species yielded a continuous and less occupied overall distribution compared to current conditions, but larger distributions than the prediction at the LGM (Fig. 6). During the LGM, *P. fasciculata* was predicted to retreat to eastern QTP and occupied a huge region and to some restricted refugia in the eastern Himalayas based on the three GCMs.

Discussion

Based on population genomic data, we found a striking population genetic structure for *P. fasciculata* in the highly fragmented biodiversity hotspot of the Hengduan Mountains. The patterns of genetic differentiation detected by different structure analyses were congruent, and we identified six groups of populations that capture the main characteristics of the population history of this species. ABC modeling suggested that the divergent times of the six groups are congruent with past glacial and interglacial events, providing support for population divergence driven by the Quaternary climatic oscillations. All six groups have experienced bottlenecks or stayed stable during the last glaciation, while five groups started to expand just after the LGM. These results obtained with genomic data were also supported by the SDM analyses. Taking together with a recent study that investigated factors in driving genomic variation in this species (Ren *et al.* submitted), our results suggest that all the historical factors (i.e. past climatic changes), spatial and environmental variables act as drivers of spatial genetic patterns. This study thus contributes a significant advance to our understanding of how

alpine species were genetically structured and responded to Quaternary climatic oscillations in the Hengduan Mountains.

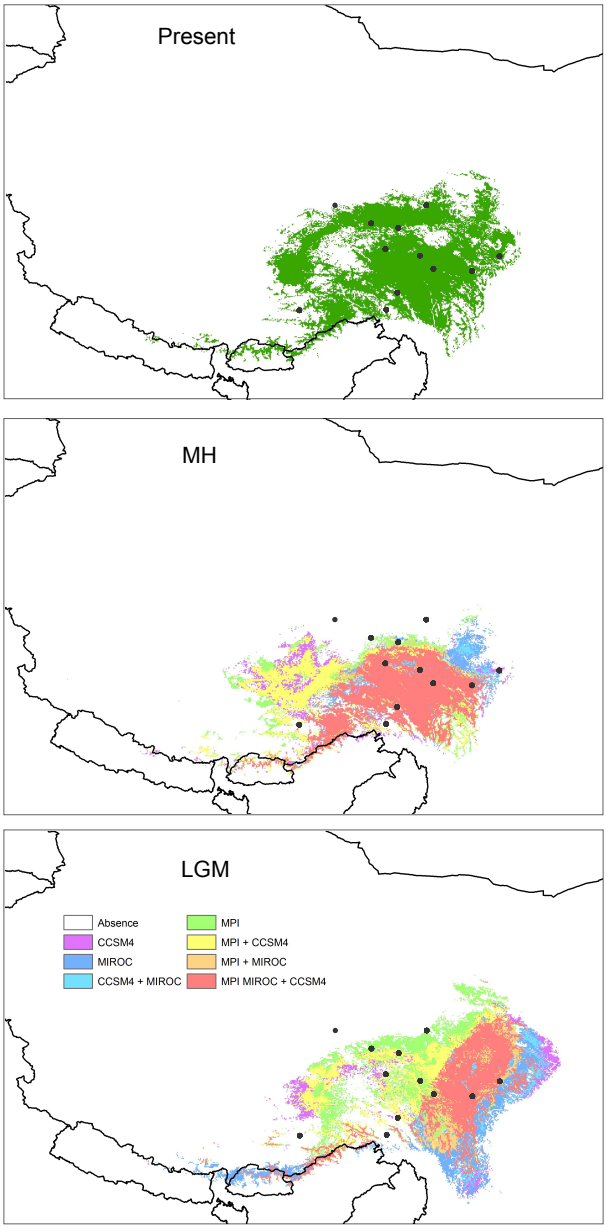


Fig. 6 Habitat suitability of *P. fasciculata* predicted by SDMs for the present, the MH and the LGM. SDMs for the MH and LGM are based on three climatic models. The predicted distributions for the present and the LGM are modified from Ren *et al.* (submitted). The black dots represent the currently geographic locations of the 12 populations used in this study for population genetic analyses.

Spatial patterns of genomic diversity

Our results revealed exceptionally high levels of population divergence across the distribution of *P. fasciculata*, with a F_{ST} value of 0.306 (Table 1). This value is slightly lower than the level of genetic differentiation among populations reported for its closely related species *Primula tibetica* (0.414; Ren *et al.* submitted) and *Bulbophyllum occultum* (0.387; Jaros *et al.* 2016), but it is still within the range usually ascribed for plants with particularly restricted dispersal ability and mainly selfing species. The divergence of populations detected with our neutral genomic markers is thus generally considered 'very high' and translates into <1 migrant per generation under equilibrium conditions (Conner &

Hartl 2004), a value often considered the minimum for maintaining species cohesion. The level of population differentiation found here contrasts with species that are characterized by extensive long-distance gene flow facilitated by the dust-like and wind-dispersed pollen and seeds exhibit rather low genetic differentiation among populations (e.g. orchids, Tremblay *et al.* 2005; epiphytic species, Phillips *et al.* 2012; *Restio capensis*, Lexer *et al.* 2014).

Spatial patterns of nuclear genomic differentiation inferred from STRUCTURE, PCA and BARRIER analyses were largely concordant with each other (Fig. 2, 3, 4b), which suggest a strong correspondence between population differentiation and their geographic locations. The pattern was further supported by the procrustes analysis, which showed a high similarity score between the overall rotated genetic space and their geographic locations (Fig. 4a). The persistent of population divergence may be facilitated by the poor dispersal ability of the species (Richards 2003) and reinforced by the rugged topographic features and profound ecological heterogeneity found in the Hengduan Mountains. Indeed, a recent study has shown that both the spatial (i.e. geographic distance and elevation differences between populations) and environmental (i.e. climatic and edaphic variables) factors acted as drivers of population differentiation in not only the selected but also the neutral genomic regions (Ren *et al.* submitted). Such strong correlation may suggest local adaptation, which may have further reinforced the genetic structure (Savolainen *et al.* 2013; Twyford *et al.* 2015). Furthermore, historical factors (i.e. past climatic oscillations) were inferred to drive the large-scale spatial genetic structure in this species (see below). Similar spatial, environmental and historical factors have been suggested to drive the spatial genetic patterns in a montane pollination-generalist herb (Muñoz-Pajares *et al.* 2016). By contrast, for the plant species *Restio capensis* that occurs in the Cape Floristic Region of South Africa, another biodiversity hotspot in the world, climatic variables were the major drivers of population divergence (Lexer *et al.* 2014). Therefore, drivers of population differentiation may be different and complex in different taxa and areas, and more factors should be considered when evaluating population differentiation of organisms and in particular for those that are distributed in mountainous areas.

Demographic history of P. fasciculata

Quaternary climatic oscillations had a dramatic effect on distribution patterns and phylogeographic structure of species (Comes & Kadereit 1998; Abbott *et al.* 2000; Hewitt 2004), especially for those distributed in high mountains such as the QTP that are assumed to be particularly vulnerable to past climatic changes (Zheng 1996; Yao *et al.* 2007). Despite much effort, we are lacking a detailed demographic history for the species present in this area because of limited genetic information (e.g. Yang *et al.* 2008; Du & Wang 2016; Wan *et al.* 2016; but see Li *et al.* 2013; Shang *et al.* 2015). By contrast, our analysis uncovers a detailed Quaternary demographic history of an alpine species in the

Hengduan Mountains. It corroborates our previous study on *P. tibetica*, which showed similar effects of the different factors in another biodiversity hotspot in the QTP, the Himalayas.

Our ABC modeling of divergence times and changes in population sizes shows that populations included in G3 experienced the most ancient expansions ca. 0.60 Ma (HPD: 0.27-0.86 Ma; Fig. 5) and all other genetic groups originated from G3. This suggests that current populations originated from ancestral populations located in the eastern Hengduan Mountains. The divergence times between the genetic groups and the ancestral populations (Fig. 5) are dated after the largest Naynauxungla glaciation that began ca. 1.2 Ma and reached its maximum between 0.8 and 0.5 Ma in the QTP (Shi 2002; Zheng *et al.* 2002). However, a previous study suggested that *P. fasciculata* originated from its closely related species *P. tibetica* during the Pliocene period (4.65 Ma) and expanded its distributions at the beginning of the Quaternary when the climate became cold (Ren *et al.* submitted). During the period between 4.65Ma and 0.6 Ma, it is unlikely that no population divergence have occurred given the varied topographic features in the region. A more possible explanation would be that extensive extinction of ancestral populations might have occurred during the past environmental changes, most likely during the largest Naynauxungla glaciation, which produced an ice sheet covering an area five to seven times larger than its current range (Shi 2002). Such huge ice sheet and extremely cold climate during the largest glaciation could have caused fragmentation of ancestral populations, lending to isolation and eventual extinction of populations located at high-altitude regions considering the fact that all the current northwestern and southwestern populations occur at more than 4000 m (Table S6). By contrast, the eastern populations, occurring at lower altitude, could have survived in an eastern refugium during the largest glaciation. When the climate became less cold, these populations could have recolonized high-altitude areas again and further gave rise to other genetic lineages triggered by the afterwards glacial and interglacial events (Wang *et al.* 2009; Opgenoorth *et al.* 2010).

The timeframes of the divergence between groups G1, G4 and G6 and the ancestral populations (i.e. G3) are congruent with a period where two other glaciation events and multiple interglacial periods occurred in the QTP (Ou *et al.* 2015). The glaciations during this period became progressively less extensive, but a cold climate prevailed in the QTP until 0.17 Ma (Shi 2002), which may have triggered these divergences. The ABC modeling of changes in populations for each group indicates that both G4 and G6 have experienced ancient expansions while G1 has stayed stable through time until the end of the LGM (Fig. 5). Such different demographic changes may depend on their specific ecological niches (e.g. Ren *et al.* 2017). The current population of G1 occurs at an altitude of 4845 m. The cold climate and less available ecological niches as indicated in the SDMs for this population (Fig. 6) may have prevented the ancient expansion of this group. By contrast, the current populations of the other two groups G4 and G6 occur at lower altitude (4170 m and average 4583 m, respectively). The open new habitats may have facilitated their ancient expansions.

Finally, the remaining two groups G2 and G5 were formed in different ways during the last interglacial when the climate was warm (Fig. 5). It seems that during the ancient expansions of G6 and ancestral populations (G3), the two groups came into secondary contact and resulted in the formation of G2. The divergence between G5 and G4 may due to complex topographic features in this region (Fig. 1). The deep valleys and high mountains may have probably caused fragmentation of the ancient expansion of G4, reduced gene flow between them and reinforced the divergence. Taken together, ice-age cycles pre-dating the LGM have had a much stronger influence on the evolutionary histories of plants in the QTP than previously thought (Qiu *et al.* 2011), especially the largest glaciation period which may have caused massive extinction of ancient populations of plants (see also Ren *et al.* 2017). However, the results of this study, combined with previous studies (Wang *et al.* 2009; Opgenoorth *et al.* 2010; Wang *et al.* 2010; Ren *et al.* 2017), clearly indicate that the alpine species in the QTP could have survived in different refugia at high altitude, conflicting with Renner's opinion that a unique ice-sheet has occurred in the QTP (Renner 2016). Furthermore, all genetic lineages have experienced bottlenecks or stayed stable during the last glaciation and post-glacial expansions, which is also evident in the SDMs (Fig. 6). This result, taken together with those recently reported for other alpine herbs (Hu *et al.* 2016; Wan *et al.* 2016; Ren *et al.* 2017), suggests that alpine plant species survived the last glaciation in multiple refugia in the QTP where most of the diverged lineages were preserved.

Conclusions

Our analysis of population genomic data in a spatially and ecologically explicit context using appropriate analytical tools could identify an accurate genetic structure and uncover a detailed demographic history of an alpine plant species, which could further allow us to demonstrate the effects of past climatic changes on its intraspecific divergence. Knowing these possible effects of past climatic changes on current populations may be useful for predicting their future range dynamics in facing ongoing climatic warming and for future management strategies. The results of this study on *P. fasciculata*, taken together with a recently reported for its closely related species occurs in the Himalayas, *P. tibetic*a (Ren *et al.* 2017), and a study that investigated interspecific divergence between them (Ren *et al.* submitted), put forward that the largest glaciation occurred in the QTP has markedly affected the evolution and demography of these two species, which probably has caused extensive extinction of their ancestral populations. The afterwards divergences are associated with the following climatic oscillations. By contrast, the LGM has little effect on these recent diverged lineages that may likely survive in multiple refugia, as also suggested by other studies in this area (e.g. Wang *et al.* 2009; Opgenoorth *et al.* 2010; Li *et al.* 2013; Hu *et al.* 2016). This response pattern to past climatic changes may be also applicable for other plant species in the QTP that share a preference for cold environments.

Acknowledgements

This work was funded by the University of Lausanne research fund, grant 31003A_138282 from the Swiss National Science Foundation, the China Scholarship Council (award to GR for four 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.

Author contributions

G.R. and N.S. planned and designed the research. G.R. carried out the sampling and the lab work, performed the molecular analysis. G.R. and N.S. wrote the manuscript with the help of E.C.

Supporting information

Additional supporting information can be found in:

https://www.dropbox.com/s/2msk71e16inajii/Supporting_information_chapter4.docx?dl=0

Reference

Abbott R, Smith L, Milne RI *et al.* (2000) Molecular Analysis of Plant Migration and Refugia in the Artic. *Science*, **289**, 1343–1346.

Antonovics J (1968) Evolution in closely adjacent plant populations. V. Evolution of self-fertility. *Heredity*, **23**, 219–38.

Avise JC (2010) Perspective: Conservation genetics enters the genomics era. *Conservation Genetics*, **11**, 665–669.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: An analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Conner JK, Hartl DL (2004) A Primer of Ecological Genetics. Sinauer Associates, Sunderland, Massachusetts.

Cornuet JM, Pudlo P, Veyssier J *et al.* (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.

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.

Du Z-Y, Wang Q-F (2016) Allopatric divergence of *Stuckenia filiformis* (Potamogetonaceae) on the Qinghai-Tibet Plateau and its comparative phylogeography with *S. pectinata* in China. *Scientific reports*, **6**, 20883.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16196–16200.

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.

Favre A, Michalak I, Chen CH *et al.* (2016) Out-of-Tibet: the spatio-temporal evolution of *Gentiana* (Gentianaceae). *Journal of Biogeography*, **43**, 1967–1978.

Fu P-C, Gao Q-B, Zhang F-Q *et al.* (2016) Gene Flow Results in High Genetic Similarity between *Sibiraea* (Rosaceae) Species in the Qinghai-Tibetan Plateau. *Frontiers in Plant Science*, **7**, 1596.

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hu CM, Kelso S (1996) Primulaceae. In: Wu ZY, Raven PH, eds. *Flora of China* Vol 15. St Louis: Missouri Botanical Garden Press, 99–185. Science Press, Beijng.

Hu Q, Peng H, Bi H *et al.* (2016) Genetic homogenization of the nuclear ITS loci across two morphologically distinct gentians in their overlapping distributions in the Qinghai-Tibet Plateau. *Scientific Reports*, **6**, 34244.

Izuno A, Kitayama K, Onoda Y *et al.* (2017) The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae). *Molecular Ecology*, doi: 10.1111/mec.14016.

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jaros U, Fischer GA, Pailler T, Comes HP (2016) Spatial patterns of AFLP diversity in *Bulbophyllum occultum* (Orchidaceae) indicate long-term refugial isolation in Madagascar and long-distance colonization effects in La Réunion. *Heredity*, **116**, 434–446.

Jombart T (2008) Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Lanier HC, Massatti R, He Q, Olson LE, Lacey Knowles L (2015) Colonization from divergent ancestors: Glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, **24**, 3688–3705.

Larson WA, Seeb LW, Everett M V. *et al.* (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, **7**, 355–369.

Lexer C, Wüest RO, Mangili S *et al.* (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.

Li L, Abbott RJ, Liu B *et al.* (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.

Li J, Fang X (1999) Uplift of the Tibetan Plateau and environmental changes. *Chinese Science Bulletin*, **44**, 2117–2124.

Liu JQ, Duan YW, Hao G, Ge XJ, Sun H (2014) Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution*, **52**, 241–249.

Liu JQ, Gao TG, Chen ZD, Lu AM (2002) Molecular phylogeny and biogeography of the Qinghai-Tibet Plateau endemic *Nannoglottis* (Asteraceae). *Molecular Phylogenetics and Evolution*, **23**, 307–325.

Liu J, Möller M, Provan J *et al.* (2013) Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, **199**, 1093–1108.

Liu JQ, Wang YJ, Wang AL, Hideaki O, Abbott RJ (2006) Radiation and diversification within the *Ligularia-Cremanthodium- Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan Plateau. *Molecular Phylogenetics and Evolution*, **38**, 31–49.

Loveless, M. D., & Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annual review of ecology and systematics*, *15*(1), 65–95.

Manni F, Guerard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology*, **76**, 173–190.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Mulch A, Chamberlain CP (2006) Earth science – The rise and growth of Tibet. *Nature,* **439**, 670–671.

Muñoz-Pajares AJ, García C, Abdelaziz M *et al.* (2017) Drivers of genetic differentiation in a generalist insect-pollinated herb across spatial scales. *Molecular Ecology*, doi: 10.1111/mec.13971

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Oksanen J, Blanchet FG, Kindt R, *et al.* (2013) vegan: Community Ecology Package. R package v.2.0-7. URL http://CRAN.R-project.org/ package=vegan.

Opgenoorth L, Vendramin GG, Mao K *et al.* (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.

Ou X, Zhou S, Lai Z, Zeng L (2015) Discussions on quaternary glaciations and their climatic responding in the Qinghai-Tibetan Plateau. *Quaternary Sciences*, **35**(1), 12–28. (in Chinese)

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Phillips RD, Dixon KW, Peakall R (2012) Low population genetic differentiation in the Orchidaceae: Implications for the diversification of the family. *Molecular Ecology*, **21**, 5208–5220.

Pritchard JK, Stephens M, Donelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

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.

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.

Ren G, Mateo G, Liu J *et al.* (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.

Ren G, Mateo G, Liu J, Guisan A, Conti E, Salamin. Species divergence and maintenance of species cohesion of three closely related Primula species in the Qinghai-Tibet Plateau. Submitted.

Renner SS (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.

Richards J (2003) Primula. Second Ed. Portland, OR, Timber Press, USA.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology . *Ecology Letters*, **16**, 940–950.

Shang H-Y, Li Z-H, Dong M *et al.* (2015) Evolutionary origin and demographic history of an ancient conifer (*Juniperus microsperma*) in the Qinghai-Tibetan Plateau. *Scientific Reports*, **5**, 10216.

Shi Y (2002) Characteristics of late Quaternary monsoonal glaciation on the Tibetan Plateau and in East Asia. *Quaternary International*, **97**, 79–91.

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

Tremblay RL, Ackerman JD, Zimmerman JK, Calvo RN (2005) Variation in sexual reproduction in orchids and its evolutionary consequences: A spasmodic journey to diversification. *Biological Journal of the Linnean Society*, **84**, 1–54.

Wan DS, Feng JJ, Jiang DC et al. (2016) The Quaternary evolutionary history, potential distribution dynamics, and conservation implications for a Qinghai-Tibet Plateau endemic herbaceous perennial, *Anisodus tanguticus* (Solanaceae). *Ecology and evolution*, **6**, 1977–1995.

Wang L, Abbott RJ, Zheng W *et al.* (2009) History and evolution of alpine plants endemic to the Qinghai-Tibetan Plateau: *Aconitum gymnandrum* (Ranunculaceae). *Molecular Ecology*, **18**, 709–721.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

Wang H, Qiong L, Sun K *et al.* (2010a) Phylogeographic structure of *Hippophae tibetana* (Elaeagnaceae) highlights the highest microrefugia and the rapid uplift of the Qinghai-Tibetan Plateau. *Molecular Ecology*, **19**, 2964–2979.

Wang C, Szpiech ZA, Degnan JH *et al.* (2010b) Comparing spatial maps of human population-genetic variation using Procrustes analysis. *Statistical applications in genetics and molecular biology*, **9**, Article 13.

Wang C, Zöllner S, Rosenberg NA (2012) A Quantitative Comparison of the Similarity between Genes and Geography in Worldwide Human Populations. *PLoS Genetics*, **8**, e1002886.

Wen J, Zhang JQ, Nie ZL, Zhong Y, Sun H (2014) Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics*, **5**, 1–16.

Wright, S. (1943) Isolation-by-distance. Genetics 28: 1, 14–38.

Wu CY (1987) Flora of Tibet. Science Press, Beijing

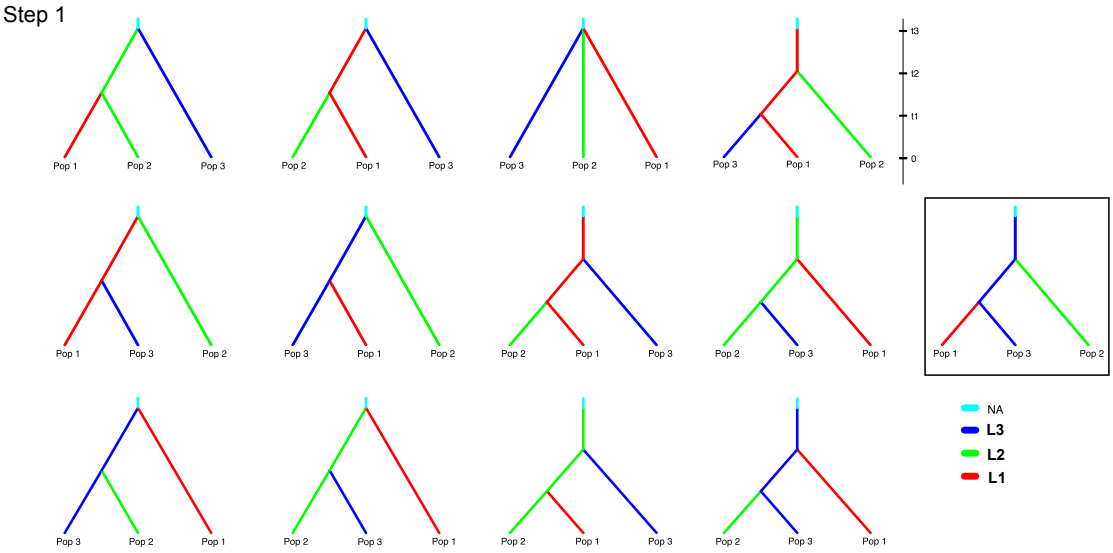
Yang FS, Li YF, Ding X, Wang XQ (2008) Extensive population expansion of Pedicularis longiflora (Orobanchaceae) on the Qinghai-Tibetan Plateau and its correlation with the Quaternary climate change. *Molecular Ecology*, **17**, 5135–5145.

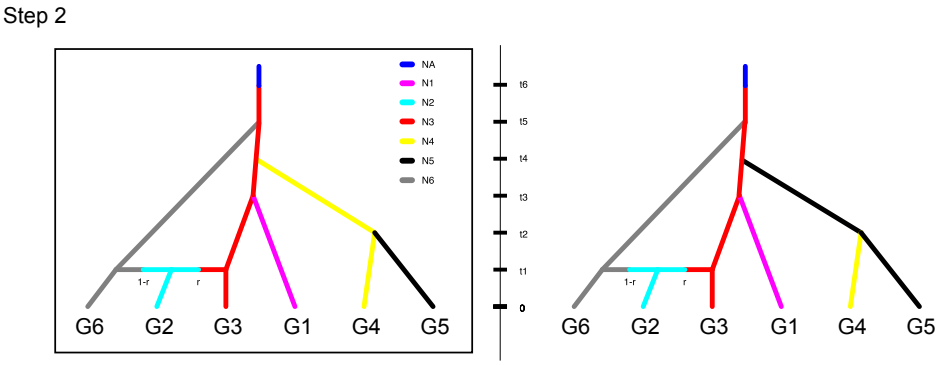
Yao T, Pu J, Lu A, Wang Y, Yu W (2007) Recent Glacial Retreat and Its Impact on Hydrological Processes on the Tibetan Plateau, China, and Surrounding Regions. *Arctic, Antarctic, and Alpine Research*, **39**, 642–650.

Zheng D (1996) The system of physico-geographical regions of the Qinghai-Xizang (Tibet) plateau. *Science in China Series D-Earth Sciences*, **39**, 410–417.

Zheng H, Powell CM, An Z, Zhou J, Dong G (2000) Pliocene uplift of the northern Tibetan Plateau. *Geology*, **28**, 715.

Zheng B, Xu Q, Shen Y (2002) The relationship between climate change and Quaternary glacial cycles on the Qinghai–Tibetan Plateau: review and speculation. *Quaternary International*, **97**, 93–101.





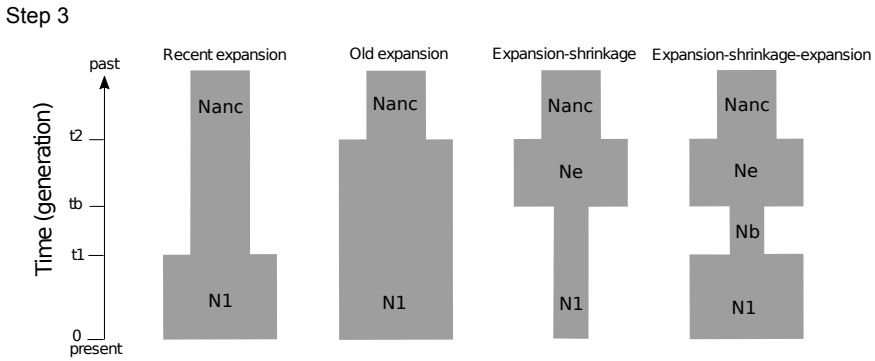


Fig. S1 Alternative demographic scenarios for the three steps analyzed by DIY-ABC. The best-fit scenario was indicated by square in step 1 and 2.

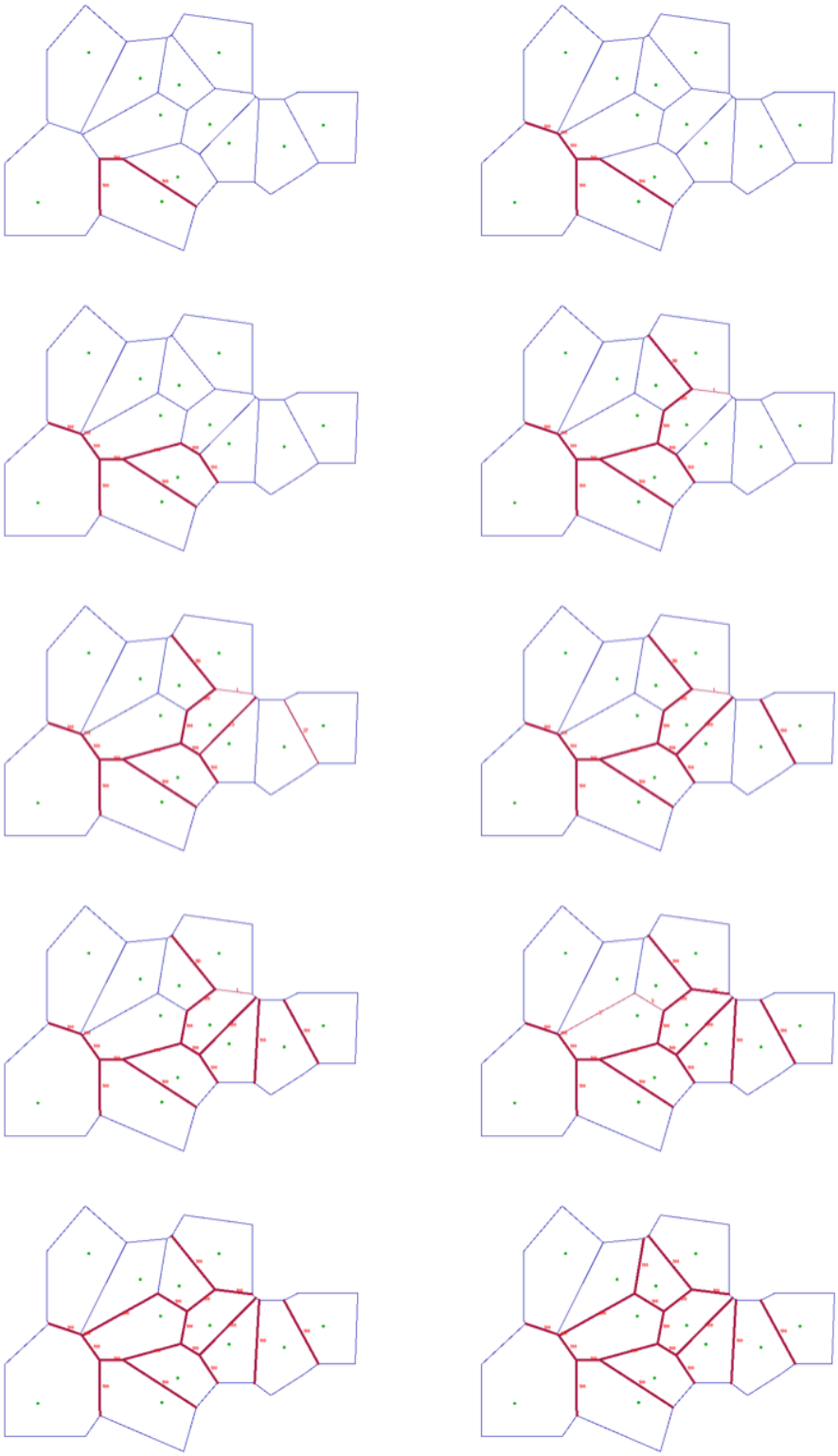


Fig. S2 Result of BARRIER analysis shows the spatial separation of *P. fasciculata* populations for number of barriers from 1 to 10.

Conclusions and perspectives

Understanding the relative roles of geography and ecology in driving species diversification, speciation, population divergence and range dynamics is significantly advanced with an integrative research that connects the macro-evolutionary pattern of biodiversity, interspecific divergence among closely related species, with demographic history of species (Liu *et al.* 2014; Wen *et al.* 2014). The massive amounts of genome level data are transforming the field of population genetics to population genomics, and simultaneously, are revolutionizing our understanding of how natural populations and species evolve (Luikart *et al.* 2003; Charlesworth 2010; Weigel & Nordborg 2015). In my thesis I integrated three taxonomic levels to better understand the evolutionary history of alpine *Primula* species in the QTP. I selected section *Armerica* of *Primula* that mainly occurring in the QTP to investigate the effects of past geological and climatic changes by combining macro-evolutionary approaches and population genomic data at multiple evolutionary timescales. This thesis represents the first population genomic level study in the QTP and contributes to a significant advance of understanding of how plant species responded to historical geological and climatic changes in mountains. Each of these scales has advantages and limitations, and open up the door to many possibilities for future work.

In chapter 1, I have reconstructed detailed phylogenetic trees for *Primula* sect. *Armerina* using both chloroplast and nuclear markers and demonstrated that the section originated from the Himalayas and its diversification timescale matched well with the latest uplift of QTP. I also illustrated the roles that niche evolution has played in shaping biogeographic patterns of three closely related species in this section. Taken all together, my findings in this chapter indicated that both the past geological and climatic events have played important roles in the evolution and distribution of this section, which prompts a question that whether these past events have triggered diversification of the whole genus or across the entire family.

The influence of the historical orogenesis and associated climatic changes on the evolution of organisms in the QTP is frequently investigated at a population level (e.g. Zhang *et al.* 2005; Wang *et al.* 2009; Li *et al.* 2013; Liu *et al.* 2014), but such effects on larger taxonomic level are still elusive in the region. The QTP, one of the world's biodiversity hotspots, is suggested as one of biogeographic source areas in Eurasia. However, only a handful of plant groups have been confirmed with this pattern of origin (Zhang *et al.* 2009; Jia *et al.* 2012; Wen *et al.* 2014). Whether *Primula* or other plant groups that most diversify in the QTP have this pattern of origin, and, if so, how did they dispersed to other regions need further investigation with a biogeographic framework. Furthermore, future work should be conducted to assess whether the uplifts of the QTP have increased the diversification rate in

Primula. More specifically, studies could test if differences in diversification rates occur between QTP lineages and non-QTP lineages, as well as QTP organisms and taxa from other mountain systems versus lowland taxa. As the QTP has experienced at least three extensive uplifts since the Miocene, whether there is correlation between these uplift events and the diversification rates over time is also interesting to test. However, carrying out these evaluations will face the difficulties of obtaining a highly resolved phylogenetic tree and reliable fossil records. A good taxonomic sampling and excellent coverage of the genus distribution range, combined with phylogenomic approaches (e.g. Wagner *et al.* 2013; Escudero *et al.* 2014; Hipp *et al.* 2014; Boucher *et al.* 2016) could be helpful to provide new insights into the evolution of biodiversity hotspots associated with mountain formation (Favre *et al.* 2014).

In chapter 2, I have examined the interspecific divergence between three closely related species of this section based on population genomic data. I successfully obtained a clear relationship among the three species based on population genomic data, which highlight the power and importance of the use of population genomic data in delimiting relationship of closely related species. I also found clear evidence for an origin of the three species in the Himalayas and demonstrated that their initial interspecific divergence may have been driven by the uplifts of the Hengduan Mountains and Northern QTP, pointing to an important role of historical geological events in the speciation process of the genus. Unexpectedly, I found no significant hybridization/introgression between the three species even in secondary contact zones. This is surprising because one of the three species, *P. nutans*, can even hybridize with *P. mistassinica* (section *Aleuritia*), a more distantly related species, and resulted in an intersectional allopolyploidization tetraploid species *P. egaliksensis (Guggisberg et al. 2009)*. I primarily showed that the spatial and environmental variables could have played a role in the maintenance of divergence between them, but further studies are needed to investigate the degree of reproductive isolation among them.

What mechanism facilitates the maintenance of reproductive isolation and trait differences in hybrid zones between closely related species is a challenging topic of molecular ecology and evolutionary biology (Coyne & Orr, 2004; Arnold, 2006). A first question that remains is whether the formation of hybrid zones results from primary intergradation or secondary contact. Our results in chapters 2, 3 and 4 indicated that the three species might have retreated to different and isolated glacial refugia during glaciations, suggesting that the current contact zones were formed by secondary contact. Second, although our results showed little introgression in general among the three species, the sampling did not include any sympatric populations. Therefore, whether there is no hybridization/introgression in contact zones need further investigation by sampling more sympatric populations (i.e. mosaic contact zones). Whole genomic sequencing analyses will provide more detailed information on the effects of hybridization in these contact zones (e.g. Meier *et al.* 2016). Third, patterns of heterogeneous genomic

divergence may reflect differential introgression following secondary contact or could result from variation in selection and recombination across the genome in the absence of any gene flow (Payseur 2010; Nachman & Payseur 2012; Cruickshank & Hahn 2014). It is essential to distinguish between genome regions that are divergent between pure parental populations and regions that show restricted introgression where these populations interact in hybrid zones. The latter, more so than the former, reveal the likely genetic architecture of reproductive isolation and local adaptation (Harrison & Larson 2016). Finally, as a complement to genomic analyses, experimental and field studies should be explored to evaluate the evolutionary and ecological consequences of hybridization and introgression in contact zones (Christe *et al.* 2016; von Rönn *et al.* 2016). Combining all these knowledge will help to obtain a more complete understanding of the factors that may allow species to originate and be maintained in the face of gene flow (Abbott *et al.* 2016).

Finally, **in chapters 3 and 4**, I have reconstructed detailed demographic histories of *P. tibetica* and *P. fasciculata*, respectively, and combined with the SDMs to evaluate the effects of the Quaternary climatic oscillations in the QTP on the intraspecific divergence and range dynamics of the two species. Our results highlight the power of population genomic data in revealing fine-scale population genetic structure in mountains and the significance of combining genomic approaches with environmental data when evaluating the effects of past climatic changes. More specifically, the largest glaciation that occurred in the QTP has markedly affected the evolution and demography of these two species, which has probably caused extensive extinction of ancestral populations, while the LGM had little effects on the recent diverged lineages. Moreover, the response to climatic changes of populations of a species depends on its specific ecological preferences. Our results provide a new response pattern of alpine plant species to the past climatic changes in the QTP, and this pattern may be also applicable for other species that share a similar preference of ecological niches.

In contrast to the two species endemic to the QTP, *P. nutans* is widely distributed in NW China, Central Asia, N Mongolia, N Europe, W&E Siberia and NW North America (Richards, 2003). I have shown in chapter 2 that this species might originate from the QTP, but the routes and timing of migration events during the spread of this species into its current distribution remain unclear. A similar study involving a finer sampling across the entirely 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).

When populations inhabit heterogeneous environments, how do populations locally adapt to their ecological niches? Detecting the footprints of local selection at sequence level is difficult, because demographic histories such as expansions and bottlenecks may generate similar patterns of genetic variation (Tajima, 1989; Zeng *et al.* 2007). However, recent improved genomic tools now allow

genomic-wide studies of local adaptation, as only restricted genomic islands are under local selection underlying specific selective pressures while demographic histories affect genome-wide variation and differentiation (Wright & Gaut 2005; Excoffier *et al.* 2009). The results in chapter 2, 3 and 4 have shown that historical, spatial and environmental factors act as drivers of genomic variation, especially in the divergent selected genomic regions, which may primarily suggest local adaptation at different environments. However, more data (e.g. whole genomic data) are needed to provide more insights into the following key questions in local adaptation: what traits are involved (e.g. flowering time along altitude gradient)? What environmental variables are the most important (e.g. altitude)? Does local adaptation target the same genes in related species? Do loci responsible for local adaptation exhibit trade-offs across environments (Tiffin & Ross-Ibarra 2014)? Identifying which traits or genes are locally selected could contribute to a better understanding of the mechanisms involved in the processes during climate change adaptation, convergent adaptation and speciation, and improve predictions of long-term climatic change responses (Savolainen *et al.* 2013).

References

Abbott RJ, Barton NH, Good JM (2016) Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, **25**, 2325–2332.

Arnold ML (2006) Evolution Through Genetic Exchange. Oxford University Press, Oxford, UK.

Boucher FC, Casazza G, Szövényi P, Conti E (2016) Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.

Charlesworth B (2010) Molecular population genomics: a short history. *Genetics research*, **92**(5-6), 397–411.

Christe C, Stölting KN, Bresadola L *et al.* (2016) Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology*, **25**, 2482–2498.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, Massachusetts.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Favre A, Päckert M, Pauls SU *et al.* (2014) The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological reviews*, **90**, 236–253.

Guggisberg A, Mansion G, Conti E (2009) Disentangling Reticulate Evolution in an Arctic-Alpine Polyploid Complex. *Systematic Biology*, **58**, 35–73.

Harrison RG, Larson EL (2016) Heterogeneous genome divergence, differential introgression, and the origin and

structure of hybrid zones. *Molecular Ecology*, **25**, 2454–2466.

Hipp AL, Eaton DAR, Cavender-Bares J *et al.* (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE*, **9**, e93975.

Jia DR, Abbott RJ, Liu TL *et al.* (2012) Out of the Qinghai-Tibet Plateau: Evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of *Hippophaë rhamnoides* (Elaeagnaceae). *New Phytologist*, **194**, 1123–1133.

Li L, Abbott RJ, Liu B *et al.* (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.

Liu JQ, Duan YW, Hao G, Ge XJ, Sun H (2014) Evolutionary history and underlying adaptation of alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution*, **52**, 241–249.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Meier JI, Sousa VC, Marques DA *et al.* (2016) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology*, **26**, 123–141.

Nachman MW, Payseur BA (2012) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 409–421.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

von Rönn JAC, Shafer A, Wolf JBW (2016) Disruptive selection without evolution across a migratory divide. *Molecular ecology*, **25**, 2529–2541

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

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3), 585–595.

Tiffin P, Ross-ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.

Wagner CE, Keller I, Wittwer S *et al.* (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.

Wang LY, Ikeda H, Liu TL, Wang YJ, Liu JQ (2009) Repeated range expansion and glacial endurance of *Potentilla glabra* (Rosaceae) in the Qinghai-Tibetan Plateau. *Journal of Integrative Plant Biology*, **51**, 698–706.

Wang Q, Liu J, Allen GA *et al.* (2016) Arctic plant origins and early formation of circumarctic distributions: A case study of the mountain sorrel, Oxyria digyna. *New Phytologist*, **209**, 343–353.

Weigel D, Nordborg M (2015) Population Genomics for Understanding Adaptation in Wild Plant Species. *Annual Review of Genetics*, **49**, 315–338.

Wen J, Zhang JQ, Nie ZL, Zhong Y, Sun H (2014) Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics*, **5**, 1–16.

Wright SI, Gaut BS (2005) Molecular population genetics and the search for adaptive evolution in plants. *Molecular Biology and Evolution*, **22**, 506–519.

Zeng K, Mano S, Shi S, Wu C-I (2007) Comparisons of Site- and Haplotype-Frequency Methods for Detecting Positive Selection. *Molecular Biology and Evolution*, **24**, 1562–1574.

Zhang Q, Chiang TY, George M, Liu JQ, Abbott RJ (2005) Phylogeography of the Qinghai-Tibetan Plateau endemic *Juniperus przewalskii* (Cupressaceae) inferred from chloroplast DNA sequence variation. *Molecular Ecology*, **14**, 3513–3524.

Zhang XL, Wang YJ, Ge XJ, *et al* (2009) Molecular phylogeny and biogeography of *Gentiana* sect. *Cruciata* (Gentianaceae) based on four chloroplast DNA datasets. *Taxon*, **58**(3), 862–870.

Annex: Ren *et al.* in prep.

Cannabis whole-genome resequencing reveals domestication history and signatures of selection for drug and hemp

Guangpeng Ren, Kate E. Ridout, Martha Liliana Serrano-Serrano, Nicolas Salamin, Luca Fumagalli

Summary

Cannabis is a predominantly diecious phenotypically diverse domesticated genus with few if any extant natural populations. Despite millennia of cultivation and current widespread use across the globe, its genetic identities, origins of most varieties and selection signatures on different types (i.e. hemp or drug) are unknown. Here we analyzed whole-genome resequencing data for 69 samples, including cultivar populations from China and Europe, feral populations from central Asia and drug samples that are currently used in Europe, Asia and South America, and identified more than one million SNPs that present in more than 90% of the samples. Using a *Hop* whole-genome data as outgroup, I identified five distinct genetic lineages (Fig. 1). The two Chinese populations formed two genetic lineages were in the basal position of the phylogenetic tree, which is consistent with the early domestication in China. The cultivars from Europe, feral populations and drugs were grouped into another three well-supported lineages. The third Chinese population was nested in the European cultivar clade and clustered with monoecious samples (Fig. 1, 2). Coalescent modeling was used to further simulate the changes in population size, divergence times and gene flow among these lineages. 930 and 173 candidate genomic regions that may involve in selection in cultivars and drugs, respectively, were identified (Fig. 3). Next step will be annotation of these regions to test which functions were selected during its domestication.

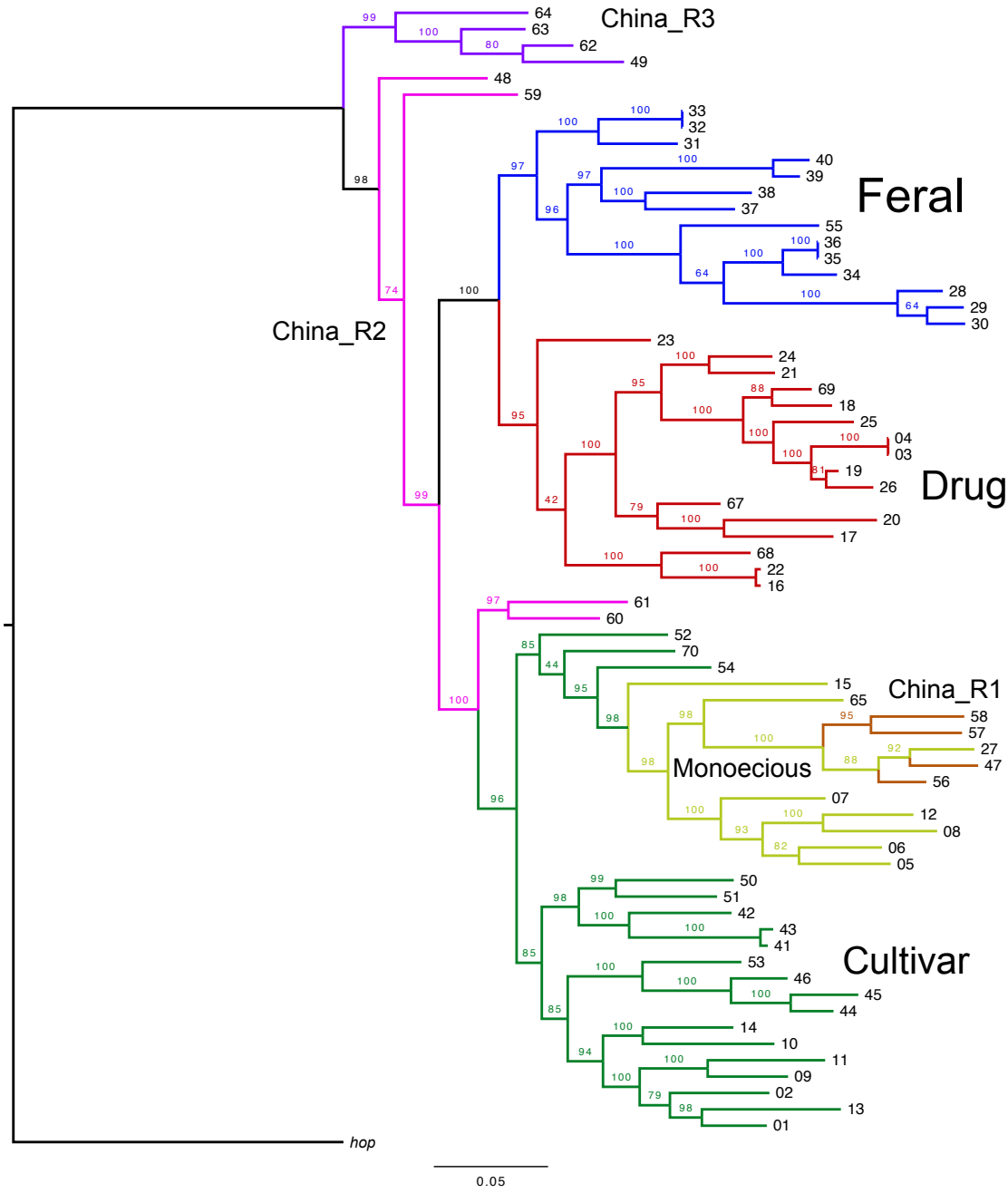


Fig. 1 Phylogenetic tree reconstructed based on 1.09 million SNPs for 69 samples of *Cannabis* with Hop as outgroup. Bootstrap values are shown on the top of braches.

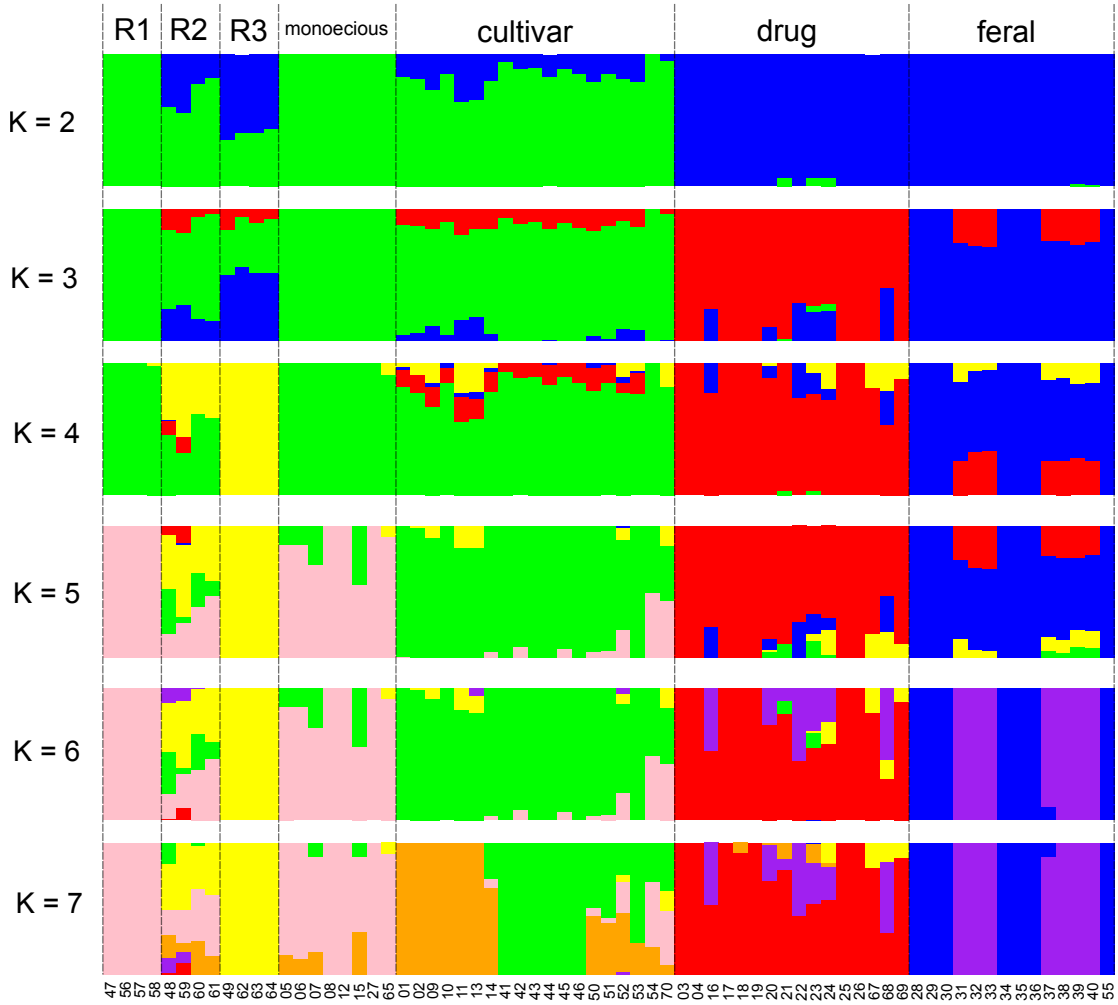


Fig. 2 Genetic structure of *Cannabis* inferred using NGSadmix. The y-axis quantifies subgroup membership, and the x-axis shows the sample ID for each individual.

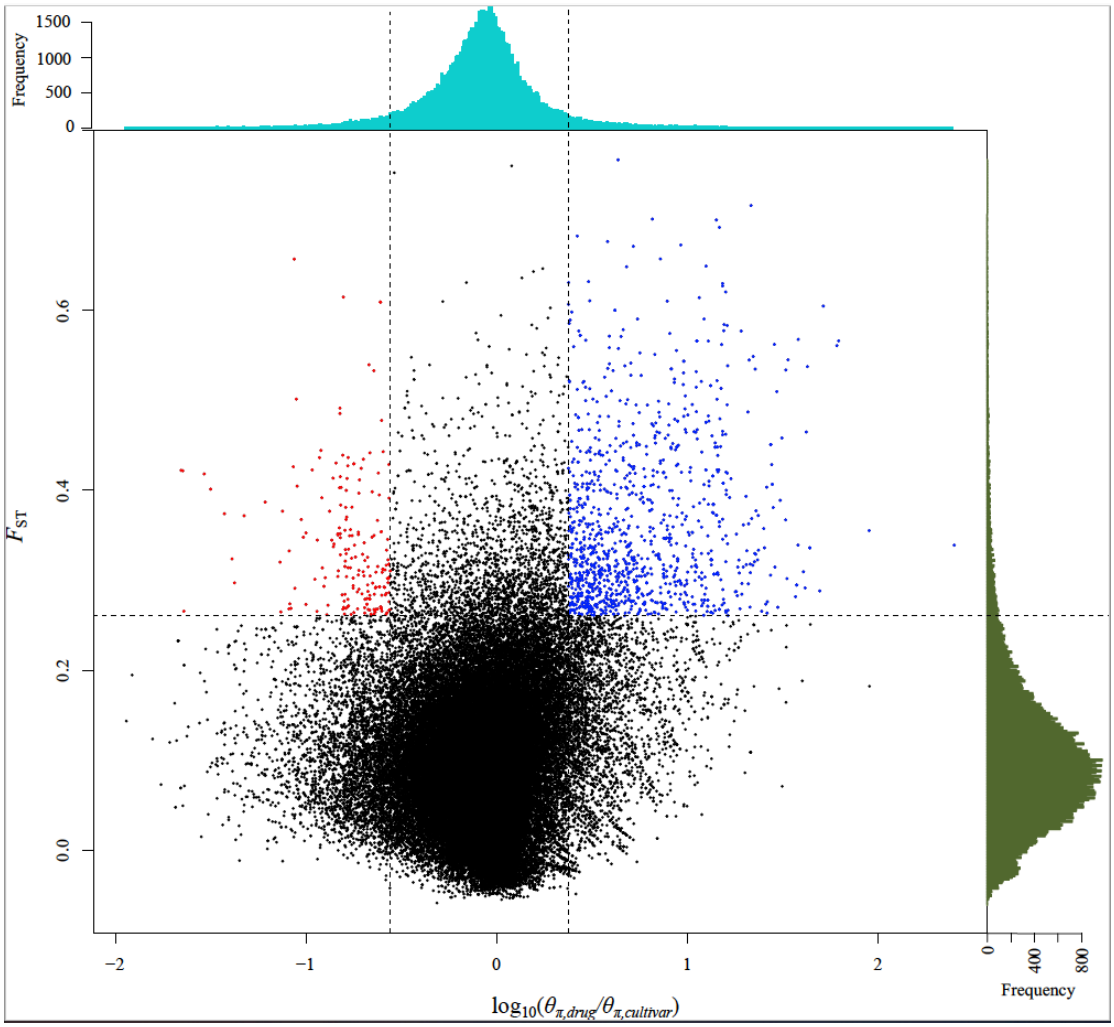


Fig. 3 Genomic regions with selection sweep signals in both drugs (red dots) and cultivars (blue dots) shown by plotting the distribution of log10 ratio ($\theta_{\pi,\text{drug}}/\theta_{\pi,\text{cultivar}}$) and F_{ST} of 50 kb windows with 10 kb steps. The windows with the top 5% values for the log10 ratio ($\theta_{\pi,\text{drug}}/\theta_{\pi,\text{cultivar}}$) and F_{ST} simultaneously as the candidate outliers under strong selective sweeps.