JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12490

Research Article

Divergence and hybridization in the desert plant *Reaumuria soongarica*

Yong Shi[1,2], Xia Yan[1,3], Heng-Xia Yin[1,4], Chao-Ju Qian[1], Xing-Ke Fan[1,5], Xiao-Yue Yin[1,5], Yu-Xiao Chang[6], Cheng-Jun Zhang[2]*, and Xiao-Fei Ma[1]*

[1]Key Laboratory of Stress Physiology and Ecology in Cold and Arid Regions, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

[2]Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[3]Key Laboratory of Eco-hydrology of Inland River Basin, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

[4]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining 810016, China

[5]University of Chinese Academy of Sciences, Beijing 100049, China

[6]Agricultural Genomes Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

*Authors for correspondence. Xiao-Fei Ma. E-mail: maxiaofei@lzb.ac.cn; Cheng-Jun Zhang. E-mail: zhangchengjun@mail.kib.ac.cn

Received 29 September 2018; Accepted 25 February 2019; Article first published online 7 March 2019

Abstract Speciation is widely accepted to be a complex and continuous process. Due to complicated evolutionary histories, desert plants are ideal model systems to understand the process of speciation along a continuum. Here, we elucidate the evolutionary history of *Reaumuria soongarica* (Pall.) Maxim., a typical desert plant that is wildly distributed across arid central Asia. Based on variation patterns present at nine nuclear loci in 325 individuals (representing 41 populations), we examined the demographic history, patterns of gene flow, and degree of ecological differentiation among wild *R. soongarica*. Our findings indicate that genetic divergence between the ancient western and eastern lineages of *R. soongarica* occurred approximately 0.714 Mya, probably due to the Kunlun–Yellow River tectonic movement and the Naynayxungla glaciation. Later, multiple hybridization events between the western and eastern lineages that took place between 0.287 and 0.543 Mya, and which might have been triggered by the asynchronous historical expansion of the western and eastern deserts, contributed to the formation of a hybrid northern lineage. Moreover, despite continuing gene flow into this population from its progenitors, the northern lineage maintained its genetic boundary by ecological differentiation. The northern lineage could be an incipient species, and provides an opportunity to study the continuous process of speciation. This study suggests that two opposite evolutionary forces, divergence and hybridization, coexisting in the continuous speciation of the desert plant *R. soongarica* in a short time. Moreover, we provide evidence that this continuous speciation process is affected by geological events, climatic change, and ecological differentiation.

Key words: arid central Asia, ecological speciation, glaciation, hybrid speciation, Quaternary, speciation continuum.

1 Introduction

Speciation is a continuous and complex process rather than an instantaneous event (Darwin, 1859; Mallet, 2008; Via, 2009), and reproductive isolation is a critical component of the speciation process, which allows one species to split into two or more derived species. However, before reproductive isolation is complete, species can occupy many positions along the speciation continuum (Coyne & Orr, 2004; Nosil et al., 2009; Baack et al., 2015). Geographic isolation and/or ecological differentiation can initiate genetic divergence, which can facilitate reproductive isolation (Mayr, 1963; Barraclough & Vogler, 2000; Coyne & Orr, 2004; Hendry, 2009). The level of reproductive isolation, from minor to complete isolation, can be enhanced by natural selection (Hendry et al., 2009). However, hybridization, a ubiquitous force opposing divergence, plays an important role in regulating the speciation process (Mallet, 2007). Moreover, more evidence of hybridization in plant speciation has been found than previously thought (Grant, 1981; Rieseberg et al., 1993; Abbott et al., 2013). Hybridization occurs when, due to incomplete reproductive isolation, differentiated lineages come into contact and produce crossbreed offspring (Lexer et al., 2003; Arnold, 2004; Rieseberg & Willis, 2007; Abbott et al., 2013). Divergence and hybridization act together to ensure that speciation is a continuous process that cannot be divided into distinct stages. Up to the present, many studies of speciation have supported a continuous model of speciation, and this interpretation is now widely accepted (Coyne & Orr, 2004). However, much still remains to be learned regarding the nature of speciation (Schluter, 2001; Via & West, 2008; Schluter,

2009), especially in species with complicated speciation histories. Rather than studying species with pure divergence or hybridization, studying species with a more complicated speciation background could deepen our understanding of the timing of genetic changes that contribute to the speciation continuum.

Arid central Asia (ACA) is one of the most arid temperate regions in the world, and is likely to be strongly impacted by global warming (Wang et al., 2016). The disconnected deserts found in the ACA region could trigger speciation events in desert plants, which provide excellent opportunities for studying speciation. The desertification of ACA is accompanied by many geological events (i.e., multiple uplifts of the Qinghai–Tibet Plateau [QTP]) and climatic changes (i.e., glacial–interglacial cycles), which fragment the habitats of desert plants and increase variation among different populations, thereby influencing the genetic structure within species (Templeton et al., 1990; Bermingham et al., 1992). Moreover, the desertification of ACA creates novel niches for desert plants. Novel or extreme niches located far from ancestral landscapes could facilitate hybrid speciation as adaptive divergence in response to differing ecological factors can establish reproductive isolation (Buerkle et al., 2000; Rieseberg et al., 2003; Rieseberg & Willis, 2007). Thus, the ACA region might have strongly affected the evolutionary history of desert plants, many of which are valuable model organisms that can be used to understand continuous speciation.

Reaumuria soongarica Maxim. (Tamaricaceae) is a typical relic Tertiary desert plant that is dominant in many ACA deserts (Shi et al., 2013) and is a useful model for studying continuous speciation. A previous study showed that the phylogeography of *R. soongarica* was significantly influenced by the uplift of the QTP and by the East Asian monsoon system (Yin et al., 2015). This evidence suggested that the speciation history of *R. soongarica* has been influenced by historical geological events, including climatic oscillation (Yin et al., 2015). According to our field observations, although no apparent reproductive isolation has been detected in populations of *R. soongarica*, its flowering time varies among different populations (Fan et al., unpublished data). Thus, despite considerable genetic divergence, *R. soongarica* does not appear to be far along the speciation continuum. Furthermore, all haplotypes found in the northern region are derived haplotypes, and the coexistence of eastern and western ribotypes/chlorotypes in the northern region suggests the simultaneous colonization of eastern and western populations of *R. soongarica* in northern areas, which might provide a clue that the northern lineage originated as a hybrid. In such a case, *R. soongarica* would have diverged and hybridized within a short time, and therefore provides an ideal model to investigate a speciation continuum. However, the nature of its complex demographic history remains unclear.

To decode the complicated speciation history of *R. soongarica*, we used nuclear multiloci methods, Approximate Bayesian Computation (ABC) simulation, and ecological niche modelling (ENM) to investigate its detailed evolutionary history. Specifically, we asked the following questions:

1. Do nuclear genes of the eastern and western lineages show significant genetic isolation?
2. Did hybridization contribute to the origination of the northern lineage? If so, how?
3. As the desertification of ACA creates different niches, did different lineages occupy different ecological niches? If so, how does gene flow influence the ecological differentiation between distinct geographical lineages?

To answer these questions, we undertook a detailed investigation of the speciation history of *R. soongarica*. We believe this study provides an important opportunity to advance our understanding of the nature of the speciation continuum.

2 Material and Methods

2.1 Plant material and DNA extraction

All samples used in this study were collected with the permission of the local government and approval of the appropriate ethics committee. We collected *Reaumuria soongarica* samples from different locations within its range in ACA, and obtained samples from populations in the Taklimakan Desert, Gurbantünggüt Desert, Qaidam Basin, Kumtag Desert, Gashun Gobi Desert, Tengger Desert, and Badain Jaran Desert. In each population, we sampled young leaves of *R. soongarica* plants located at least 30 m apart from each other to minimize sampling bias. Sampled leaves were immediately stored in silica gel. The latitude and longitude of each population were recorded using a Global Positioning System device (Garmin, Taiwan, China). We extracted the total genomic DNA of all samples using the Qiagen DNeasy plant DNA extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. Seven to eight samples from each population were used for sequence analysis; this sampling protocol yielded a total of 325 individuals from 41 populations. A geographical distribution map of the 41 sampled populations is shown in Fig. 1. This map was produced using the R package ggmap (Kahle & Wickham, 2013). Further information on these samples is presented in Table S1.

2.2 DNA amplification and sequencing

To screen for loci suitable for demographic analysis, we searched for homologs to 211 single-copy genes identified in *Arabidopsis thaliana* (Wang et al., 2014a). Homologous copies of these genes were identified by local BLAST of a database of RNA-seq unigene sequences of *R. soongarica* (Shi et al., 2013). The BLAST results were used to design primers to amplify the corresponding genes in *R. soongarica*. Purified Polymerase Chain Reaction products were processed using the ABI-PRISM BIGDYE Terminator Cycle Sequencing Ready Reaction Kit, version 3.1 (Applied Biosystems, Foster, CA, USA). Both forward and reverse directions were sequenced. The resulting primers were evaluated by the quality and sequence-specificity of their polymerase chain reaction products as well as the level of DNA polymorphism at the corresponding locus. In total, we identified 13 loci (Table S2) for further amplification and sequencing.

Nuclear haplotypes were inferred using the PHASE algorithm (Stephens & Donnelly, 2003) implemented in DnaSP version 5.10.1 (Librado & Rozas, 2009). In addition, DNA from one to five heterozygous individuals at each locus was cloned into a

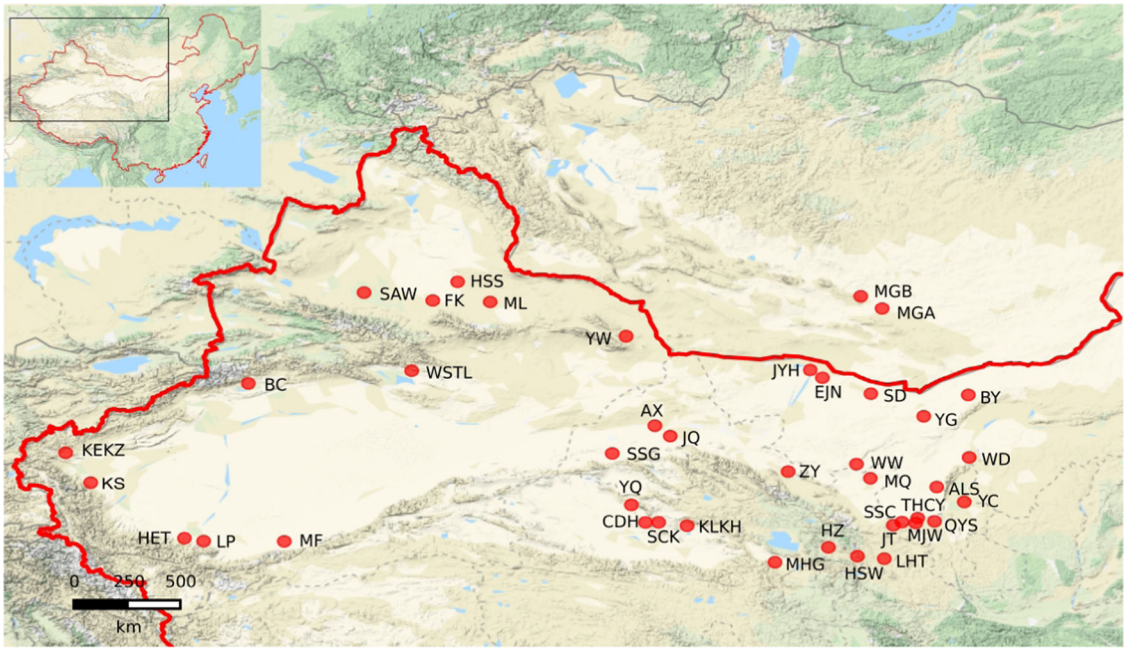

Fig. 1. Sampling locations of *Reaumuria soongarica*. Most populations were collected from China except MGA and MGB, which were collected from Mongolia. Map was retrieved from Google Earth (https://www.google.com/earth/). ALS, Alxa left Banner; AX, Anxi; BC, Baicheng; BY, Bayannaoer; CDH, Chaidam Lake; EJN, Ejin Banner; FK, Fukang; HET, HeTian; HSS, Huoshaoshan; HSW, Haishiwan; HZ, Huzhu; JQ, Jiuquan; JT, Jingtai; JYH, Juyanhai; KEKZ, Keerkezi; KLKH, Keluke Lake; KS, Kashi; LHT, Laohutai; LP, Luopu; MF, Mingfeng; MGA, Bulgan Mongolia; MGB, Bogd Mongolia; MHG, Mahuanggou; MJW, Mengjiawan; ML, Mori; MQ, Mingqin; QYS, Quanyanshan; SAW, Shawan; SCK, Sanchakou; SD, Provincial Road No. 312; SSC, Shashichang; SSG, Shashangou; THCY, Tonghucaoyuan; WD, Wuda; WSTL, Wusitala; WW, Wuwei; YC, Yinchuan; YG, Yingen Sumu; YQ, Yuqia; YW, Yiwu; ZY, Zhangye.

pGEM-T Easy Vector (Promega, Madison, WI, USA). Six to eight clones per individual were randomly selected and sequenced to reconstruct haplotypes for heterozygous individuals. Singletons were verified by resequencing to exclude sequencing errors.

All sequences were aligned using ClustalW implemented in BioEdit version 7.2.2 (Hall, 1999), and final alignments were carefully refined by eye twice. All haplotype sequences were deposited in GenBank under the accession numbers MH688942–MH697391. Outgroup sequences from *R. trigyna* Maxim. (Dang et al., 2013) were downloaded from the NCBI database.

2.3 Nucleotide diversity and tests of neutrality

We calculated the classic genetic parameters for each genetic cluster. The coding and noncoding regions of all sequences were aligned with their corresponding mRNA sequences, and DnaSP was used to calculate the number of segregating sites (S), the numbers of synonymous and non-synonymous sites, Watterson's theta (θ_w; Watterson, 1975), the degree of polymorphism (π; Tajima, 1983), the minimum number of recombination events (R_m; Hudson & Kaplan, 1985), the number of haplotypes (N_h), and haplotype diversity (H_e; Nei, 1987).

As the assumption of neutrality is crucial for estimating demographic parameters, we applied the maximum frequency of derived mutations (MFDM) test (Li, 2011); this test is a recently developed method to detect variants of neutrality, and is not affected by changing population size. The MFDM method has higher power and a lower false-positive rate compared to other comparable methods (Li, 2011). Alignment gaps were excluded, and the MFDM test was undertaken for each locus, and for each test *R. trigyna* was used as an outgroup. Loci under selection were excluded from subsequent analyses.

2.4 Population structure

The population substructure of *R. soongarica* was analyzed using a Bayesian clustering method implemented by structure version 2.3.4 (Hubisz et al., 2009). First, Single Nucleotide Polymorphisms under divergent selection were identified by BayeScan version 2.1 (Foll & Gaggiotti, 2008). We ran the program using 20 pilot runs of 5000 iterations and an additional burn-in of 50 000 iterations. Afterwards we used 100 000 iterations for posterior estimation (i.e., with a sample size of 5000 and a thinning interval of 20). *P*-values less than 0.05 were identified as outliers. Second, after removing outliers, we obtained a total of 503 single nucleotide polymorphisms to be used as the input data set for structure analyses. We examined the number of clusters (K) from $K = 1$ to $K = 10$ using the admixture model. For each cluster number (K), we ran the simulation 20 times with different seed numbers. For each run, a chain length of 500 000 steps followed a burn-in of 200 000 steps. The delta K method was implemented in structure harvester version 0.6.94 (Earl & Vonholdt, 2012) to determine the most likely number of clusters for *R. soongarica*. distruct version 1.1 (Rosenberg, 2004) was used to visualize our results from $K = 2$ to $K = 7$, showing the highest value of LnPD for each K.

In addition to STRUCTURE analyses, we undertook two additional analyses to infer the genetic clusters present in our *R. soongarica* data set. First, we used Populations version 1.2.32 (Langella, 1999) to construct a phylogenetic tree. This tree was constructed using the neighbor-joining (NJ) method based on Nei's DA distances (Nei et al., 1983) with 1000 bootstrap runs. Second, we applied principal coordinate analysis (PCoA), a genetic distance-based clustering method, to determine the genetic relationships between different populations without group identity. The distance matrix was calculated using the Distance function of GenAlex version 6.5 (Peakall & Smouse, 2012) and was visualized using the PCoA tools implemented in the same program. Finally, we divided the 41 populations into four groups: the Taklimakan Desert (TaD), the Gurbantünggüt Desert (GuD), the Qaidam Basin–Kumtag Desert–Gashun Gobi Desert (QBKG), and the Tengger Desert–Badain Jaran Desert–Mongolia–Gobi Desert (BJTD). These four groups clustered into three lineages (see section 3.1 for details): a western lineage (TaD), an eastern lineage (QBKG and BJTD), and a northern lineage (GuD).

2.5 Phylogenetic relationships among different groups

Next, *BEAST version 2.4.1 (Bouckaert et al., 2014) was used to construct a multilocus species tree to determine the phylogenetic relationships reflecting the speciation history of *R. soongarica*. The BEAUti function implemented in *BEAST was used to generate an input file with haplotype sequences of nine neutral loci and the *R. trigyna* sequence as an outgroup. A strict clock was used as the molecular clock, and the Yule model was used as speciation prior. The best substitution model for each locus (Table S3) was determined using jModeltest version 2.1.7 (Posada, 2008). For each run, the Markov chain Monte Carlo length was set to 80 million generations, and trees were stored every 8000 generations. Several runs were carried out, and Tracer version 1.7.1 (Drummond & Rambaut, 2007) was used to ensure that each independent run converged to a similar stationary distribution. LogCombiner version 1.7.4 (Drummond & Rambaut, 2007) was used to combine log and tree files with 10% burn-in. TreeAnnotator version 1.7.4 (Drummond & Rambaut, 2007) was used to recover the highest clade probability tree from consensus trees.

Due to the ambiguous results obtained from BEAST (see section 3.2 for details), we tested the possibility of a hybrid origin of the northern lineage. PhyloNet version 3.5.4 (Than et al., 2008) was used to infer a network that accounted for both incomplete lineage sorting and gene flow using a maximum likelihood method. We used the best gene trees generated using BEAST for each locus, and datasets that included genotypic data for nine loci were compiled. We ran PhyloNet 10 times from different starting seeds, searching a maximum of 50 network topologies; the top five optimal maximum likelihood networks were inferred.

2.6 *Reaumuria soongarica* demographic history scenarios

To reconstruct the demographic history of *R. soongarica*, we used DIYABC version 2.0.4 (Cornuet et al., 2014) to undertake a two-step simulation analysis. First, rather than examining an exhaustive list of scenarios, we identified the eight most likely scenarios (Fig. S1) based on previous knowledge of *R. soongarica* (Li et al., 2012a; Yin et al., 2015). These scenarios differ in their treatment of hybridization/admixture after splitting events and whether GuD acts as an ancestral population (notably, sequences from WSTL and YW were excluded because of their ambiguous classification; see Figs. 2, S2 for further detail). Second, based on the results of the first step, four scenarios differing in their treatment of admixture events in the GuD population (Fig. S3) were analyzed in a subsequent step. The specific prior settings for all scenarios are shown in Table S4. The summary statistics include number of haplotypes, number of segregating sites, mean of pairwise differences, variance of pairwise differences, Tajima's *D*, private segregating sites, and mean of numbers of the rarest nucleotide at segregating sites for one sample summary statistics, and number of segregating sites, mean of pairwise differences (*W*) and (*B*), and F_{ST} for the two sample summary statistics. For each scenario, we simulated at least 1 000 000 datasets to ensure statistically robust results in downstream analyses.

To determine the best scenario fitting the data, we compared the difference between simulated data and observed data to calculate the posterior probabilities of each scenario. Logistic regression and direct approaches were undertaken to estimate the relative likelihoods of different scenarios, using 10 000 closest simulated data and 500 data from each scenario, respectively.

To validate the accuracy of our model selection, we estimated the confidence in scenario choice for each of the top four models in the first-round simulation, and estimated the false positive and false negative error rates of each scenario. Finally, we undertook model checking and assessed measures of bias and error to evaluate the goodness of fit between the observed and simulated data.

2.7 Gene flow estimation

IMa2 version 2.0 (Hey, 2010) and Migrate-N version 3.6.4 (Beerli, 2006) were used to estimate the gene flow between different groups. As recombination is not modelled by either the "Isolation with Migration" model or Migrate-N, we ran IMgc (Woerner et al., 2007) to identify the largest single block without recombination at each locus. Nine trimmed neutral loci datasets were used for the IMa2 and Migrate-N analyses.

Due to the complex demographic history of the four *R. soongarica* groups, we carried out pairwise analyses in IMa2 among the four groups. An infinite allele mutation model was implemented for all loci, and several preliminary runs were undertaken to constrain the upper boundary of priors and to determine the geometric scheme for good mixing of Markov chain Monte Carlo chains. After obtaining an appropriate prior distribution, final simulations were carried out using 60–120 chains, with $h_1 = 0.98–0.99$, $h_2 = 0.55–0.75$ to yield high swap rates and sufficient mixing. After a burn-in period of 10–20 million steps (depending on the population used), a stationary phase was reached (i.e., low parameter autocorrelation, no trends in the plots, and a high effective sample size). Trees were then sampled using Markov chain state files. Two well-mixed runs were carried out using different seed numbers to confirm accurate parameter estimation. After sampling 300 000 trees, all results were combined and log-likelihood ratio (LLR) tests (Nielsen & Wakeley, 2001) were applied to determine whether or not gene flow between different groups was greater than zero.

Next, we ran Migrate-N to determine the effective population size for each group and to estimate migration

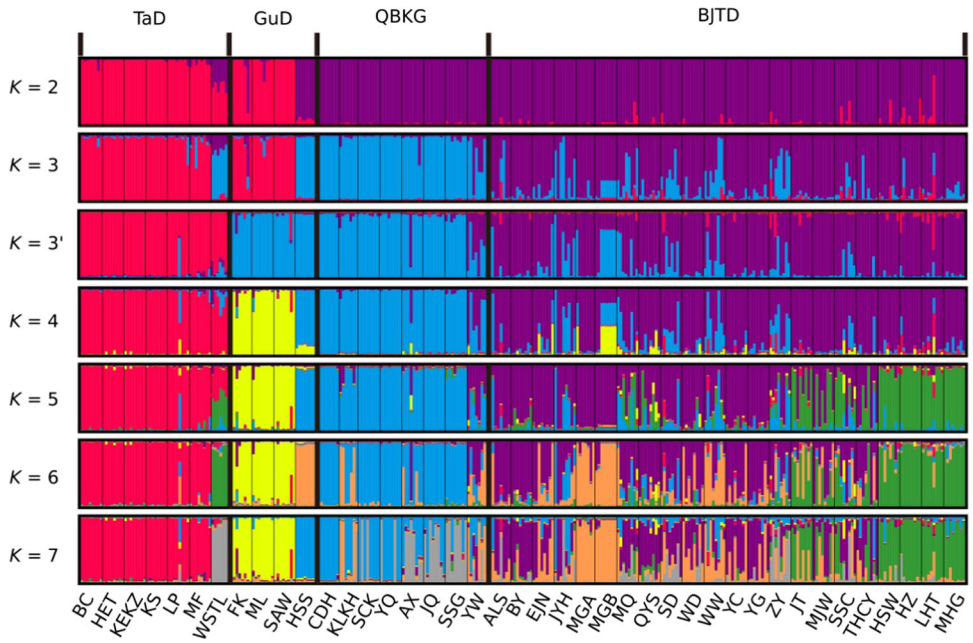

Fig. 2. Structure plots depicting the population assignment of 41 *Reaumuria soongarica* populations estimated under $K = 2$ to $K = 7$. $K = 3$' shows the alternative assignment of Gurbantünggüt Desert (GuD) when populations are clustered into three putative groups. BJTD, Badain Jaran–Tengger Desert; QBKG, Qaidam Basin–Kumtag Desert–Gashun Gobi; TaD, Taklimakan Desert. Population collection sites: ALS, Alxa left Banner; AX, Anxi; BC, Baicheng; BY, Bayannaoer; CDH, Chaidam Lake; EJN, Ejin Banner; FK, Fukang; HET, HeTian; HSS, Huoshaoshan; HSW, Haishiwan; HZ, Huzhu; JQ, Jiuquan; JT, Jingtai; JYH, Juyanhai; KEKZ, Keerkezi; KLKH, Keluke Lake; KS, Kashi; LHT, Laohutai; LP, Luopu; MF, Mingfeng; MGA, Bulgan Mongolia; MGB, Bogd Mongolia; MHG, Mahuanggou; MJW, Mengjiawan; ML, Mori; MQ, Mingqin; QYS, Quanyanshan; SAW, Shawan; SCK, Sanchakou; SD, Provincial Road No. 312; SSC, Shashichang; SSG, Shashangou; THCY, Tonghucaoyuan; WD, Wuda; WSTL, Wusitala; WW, Wuwei; YC, Yinchuan; YG, Yingen Sumu; YQ, Yuqia; YW, Yiwu; ZY, Zhangye.

rates between them. These tests were undertaken using the same input file as for IMa2. A Bayesian inference module was used for each run, and the theta and migration values were initiated from the F_{ST} calculation. Uniform priors and slice sampling for parameter distributions were used. For static heating, we ran four chains with temperatures of 1.0, 1.5, 3.0, and 1×10^6 from cold to hottest. The increment was set to 100, and a total of 1.5×10^8 genealogies were sampled after discarding 1.5×10^6 trees per chain as burn-in. Convergence was verified by investigating the smoothness of posterior distribution histograms and whether similar results were generated from different independent runs.

2.8 Tests of demographic expansion

The historical demography of *R. soongarica* was determined using several different methods. An Extended Bayesian Skyline Plot (EBSP) implemented by BEAST version 1.75 (Drummond et al., 2012) was used to test historical changes in the population size of different groups. Moreover, we used Tajima's *D* and Fu's F_s tests implemented by DnaSP to test for deviations from neutrality at each locus. These results were visualized using the ggplot2 package in R (Wickham, 2016).

Independent EBSP analyses were executed for each group. For each locus, the best substitution model was set to HKY (Hasegawa et al., 1985), which was determined by jModeltest. A strict clock for each analysis was selected, and each run started with a random tree. To estimate the mutation rate for all nuclear loci, all chloroplast DNA (cpDNA) sequences from our previous study were added and a substitution rate of 4.87×10^{-4} substitutions per cpDNA site per million years (Yin et al., 2015) was specified. The prior distribution was set to log normal with respect to the clock rate. Three independent runs were performed for each group. For each run, trees were sampled every 5000 generations, yielding a total of one billion generations; TRACER version 1.7.1 (Drummond & Rambaut, 2007) was used to ensure that an effective sample size greater than 200 and a stationary state for each parameter was reached.

2.9 Ecological niche modeling and identity tests

A maximum entropy model implemented by MAXENT version 3.4.1 (Phillips et al., 2006) was used to determine the distribution ranges of the western lineage, eastern lineage, and northern lineage, respectively. A total of 126 non-duplicate occurrence data points were collected from field-work and published reports; these included 14, 96, and 16 points for the western, eastern, and northern lineages, respectively. Climatic variables were downloaded from the WorldClim database (Hijmans et al., 2005). After removing highly correlated bioclimatic variables (i.e., those with Pearson correlation coefficients > 0.9) using ENMtools version 1.4.3 (Warren et al., 2010), 12 environmental variables were retained for subsequent analyses (Table S5). The present distribution models of the three lineages were

generated by MAXENT using the default settings. The random tests used 25% of the data set for testing and 75% for training. Model accuracy was determined by the area under the receiver operating characteristic curve, which ranges from 0 (no discrimination) to 1 (perfect discrimination), and a score of 0.5 suggests that the discriminatory power of the model is no better than random prediction (Elith et al., 2006).

In addition, identity tests were also applied to compare the similarity of the distribution models of each pair of ecotypic clusters. The null hypothesis supposes that each pair of groups is distributed in an identical environmental area. The Schoener's *D* similarity index (Schoener, 1970) and Warren's *I* (Warren et al., 2008) measure of niche overlap were calculated suing ENMTools (Warren et al., 2010). Finally, to test the significance of divergence between distribution models, 100 simulations were carried out for each pairwise comparison.

3 Results

We sequenced a total of 13 unlinked single-copy nuclear loci from 325 individuals found in 41 populations of *Reaumuria soongarica*. After excluding gaps and missing data, the sequenced fragments ranged from 394 to 1756 bp with a total length of 9740 bp. The basic genetic information is summarized in Table S6. The MFDM test detected signals of adaptive evolution at four loci: A7, A105, H13, and H22 ($P = 0.0061$, $P = 0.0062$, $P = 0.0063$, and $P = 0.0031$, respectively, at a significance level of $P < 0.01$). Thus, we excluded these loci from subsequent demographic analyses.

3.1 Genetic differentiation and population structure

To delimit the genetic boundaries of *R. soongarica*, we used STRUCTURE to infer the most likely genetic clusters without prior geographic information. The delta *K* method (Fig. S4) supported $K = 2$ as the best-fit number of genetic clusters, which are consistent with our previous study (Yin et al., 2015). However, the log-likelihood values peaked at $K = 4$ and then plateaued, remaining similar up to $K = 6$, after which they dropped. This pattern suggested that it would be reasonable to cluster *R. soongarica* into four groups ($K = 4$, Fig. S4). Moreover, PCoA results (Fig. S5) as well as an NJ tree (Fig. S2), both supported the four-group classification. Thus, to obtain a detailed understanding of the population history of *R. soongarica*, we clustered individuals in four groups from four distinct geographic areas: TaD, GuD, QBKG, and BJTD. Due to the recent divergence and substantial gene flow between the QBKG and BJTD populations (see section 3.4 for details), these four groups could be further clustered into three lineages, that is, a western lineage (TaD), an eastern lineage (QBKG and BJTD), and a northern lineage (GuD). Interestingly, we obtained alternative classification results when we set $K = 3$ (Fig. 2). In this case, GuD could be clustered either with TaD or QBKG, both of which had similar Ln-likelihood values, supporting the hypothesis that GuD originated by hybridization.

3.2 Phylogenetic relationships among the four groups

Next, we used BEAST and PhyloNet to obtain a clear understanding of the phylogenetic relationships among the four groups. These analyses yielded consistent phylogenetic relationships among the TaD, BJTD, and QBKG populations (posterior probability > 0.90). However, we did not obtain consistent results for the GuD group (the northern lineage) from BEAST, as the position of the GuD group (the northern lineage) varied among the three trees produced by BEAST (Fig. S6). Moreover, each of the top five trees obtained from PhyloNet supported the hypothesis that the northern lineage originated by hybridization (Fig. S7).

3.3 Demographic history of scenario

To obtain a detailed picture of the demographic history of *R. soongarica*, we ran a two-Step DIYABC simulation analysis. In the end, our simulation results supported the divergence and hybrid scenario. After the first round of simulations, we obtained a set of four most likely scenarios (Fig. S8), each of which had a high posterior probability among the eight putative models. A second round of simulations that considered these top four scenarios generated the best results, and identified the best scenario overall, by incorporating two admixture events (Fig. S9). We then estimated the demographic parameters of the best scenario; the posterior probability distributions for each parameter are summarized in Table 1. The best scenario suggested that *R. soongarica* emerged approximately 0.714 Mya, after which *R. soongarica* soon diverged into the western and eastern lineages. The northern lineage then originated from two admixture events between the western and eastern lineages that occurred sometime between 0.287 and 0.543 Mya. Finally, the eastern lineage recently diverged into the BJTD and QBKG groups around 0.159 Mya (Fig. 3).

To evaluate the accuracy of the DIYABC models, we calculated the likelihood of type I and type II errors in each evolutionary model using a logistic method based on 1% of the simulated datasets. The results of this analysis (shown in Table 2) suggested that the likelihood of type I and type II errors occurring in scenarios with admixture events (i.e., scenarios 1 and 4) were lower than those of the other scenarios. We also evaluated demographic parameter estimation for the best scenario (see Table 3). The root of the relative mean square error (RRMSE) was 0.38 or less for all parameters except r_b and N_5. We noticed that the RRMSE of r_a (the first admixture proportion from the Western lineage, TaD, RRMSE = 0.377) was more accurate than r_b (the second admixture proportion from the western lineage, TaD, RRMSE = 9.187). This result could be caused by the uncertainty of N_5 (i.e., the population size of the unsampled ghost population, RRMSE = 26.796). However, a 95% credible interval of both r_a and r_b did not include zero or one, implying that the northern lineage did likely originate from admixture between the western and eastern lineages.

3.4 Gene flow estimation

To estimate gene flow among the four groups, we undertook pairwise comparisons using IMa2 and Migrate-N. To better interpret the migration rate, we explained the migration (2NeM) results of IMa2 in forward (natural) time instead of coalescent time. Our IMa2 (Fig. S10) and Migrate-N (Fig. S11) results revealed similar migration patterns, both suggesting that small migrations occurred between TaD and GuD, and between BJTD and TaD. However, an exceptionally high level of migration was detected between QBKG and BJTD; this finding is consistent with their recent divergence in the best DIYABC scenario. Due to their weak genetic isolation (NeM > 4, Wright, 1931; Waples &

Table 1 Estimation of the posterior distribution of parameters of the DIYABC under the best scenario for the demographic history of *Reaumuria soongarica*

Parameter	N_1	N_2	N_3	N_4	N_5	N_a	r_a	r_b	T_1	T_{m1}	T_{m2}	T_3
Mean	1.60E + 06	1.10E + 06	9.90E + 05	8.27E + 06	2.67E + 05	3.76E + 05	0.47	0.49	7.14E + 05	5.43E + 05	2.87E + 05	1.59E + 05
Median	1.64E + 06	1.12E + 06	1.00E + 06	8.32E + 06	2.72E + 05	3.95E + 05	0.46	0.5	6.99E + 05	5.16E + 05	2.67E + 05	1.61E + 05
Mode	1.65E + 06	1.15E + 06	1.03E + 06	8.03E + 06	4.47E + 05	4.38E + 05	0.18	0.48	6.21E + 05	4.77E + 05	2.34E + 05	1.67E + 05
95% CI	1.03E + 06	5.84E + 05	6.29E + 05	4.07E + 06	2.23E + 04	1.56E + 05	0.03	0.04	2.53E + 05	2.00E + 05	8.49E + 04	8.43E + 04
	1.96E + 06	1.54E + 06	1.30E + 06	1.17E + 07	4.88E + 05	4.87E + 05	0.96	0.94	1.21E + 06	1.04E + 06	6.21E + 05	2.24E + 05

Estimation was based on 1% of the closest simulated datasets. Parameters were logit-transformed. N_1, effective population size of Taklimakan Desert (TaD); N_2, effective population size of Gurbantünggüt Desert (GuD); N_3, effective population size of Qaidam Basin–Kumtag Desert–Gashun Gobi Desert (QBKG); N_4, effective population size of Tengger Desert–Badain Jaran Desert–Mongolia–Gobi Desert (BJTD); N_5, effective population size of ghost population; r_a, the admixture proportion from the ancestral western lineage; r_b, the admixture proportion from the ancestral eastern lineage; T_1, time since divergence between the ancestral western and eastern lineages; T_{m1}, time since the first admixture between the ancestors of the western and eastern lineages; T_{m2}, time since the second admixture between the ghost population and the ancestral eastern lineage; T_3, time since the divergence between QBKG and BJTD.

Gaggiotti, 2006; Reilly et al., 2012), we considered QBKG and BJTD to be a panmictic lineage (i.e., the eastern lineage) for subsequent analyses. We used LLR tests (Nielsen & Wakeley, 2001) to examine 25 models listed in Table S7. The results (Table S8) showed significant bidirectional migration among most pairwise comparisons, including significant gene flow from TaD to QBKG and from BJTD to GuD. Finally, we used ggmap (Kahle & Wickham, 2013) to represent the final gene flow patterns in Fig. 4.

3.5 Population expansion in BJTD and TaD groups

We carried out EBSP analyses using BEAST to infer historical changes in population size for each of the four groups. No demographic changes were detected in the GuD and QBKG groups. However, the EBSP (Fig. 5) for the TaD group revealed a population size expansion initiated approximately 0.20–0.25 Mya, and the EBSP for the BJTD group showed a population size expansion that took place approximately 0.55–0.60 Mya; both of these findings were consistent with their negative Tajima's *D* and Fu's F_s statistics (Fig. S12).

3.6 Identity test and ENM

Ecological niche modeling revealed that the eastern, western, and northern *R. soongarica* lineages occupied distinct niches. Fig. 6A shows the reliable distribution predicted by ENM for each lineage; this suggests that the northern lineage might have colonized an entirely new niche. We also applied identity tests using ENMTools to undertake pairwise comparison among the eastern, western, and northern lineages. The Schoener's *D* and *I* values for each comparison (Fig. 6B) suggested that each of the three areas significantly differed from a random distribution ($P < 0.01$), suggesting that the niches occupied by the different lineages were not identical. Moreover, the *D* and *I* values also supported the hypothesis that these three lineages are somewhat ecogeographically isolated, and especially that the northern lineage is relatively isolated from both of the other two lineages.

4 Discussion

The evolutionary history of *Reaumuria soongarica* can be generally divided into three stages: (a) the divergence of ancestral *R. soongarica* into the western and eastern lineages; (b) the hybrid formation of the northern lineage; and (c) the stabilization of the northern lineage. Our results suggest that geographic isolation and ecological differentiation caused *R. soongarica*'s western–eastern genetic differentiation. In addition, the asynchronous desertification of desert regions, which likely caused the expansion of the western and eastern lineages, facilitated the hybrid origin of the northern lineage. Moreover, although gene flow from other populations is ongoing, the northern lineage is genetically stable, and this stability could be caused by its significant ecological differentiation.

4.1 Genetic divergence between western and eastern lineages

Multiple lines of evidence suggest that the ancestors of *R. soongarica* first divided into the western and eastern lineages. First, significant genetic differentiation was detected between the two lineages. Although we clustered *R. soongarica* into four groups to obtain a detailed demographic history, the delta

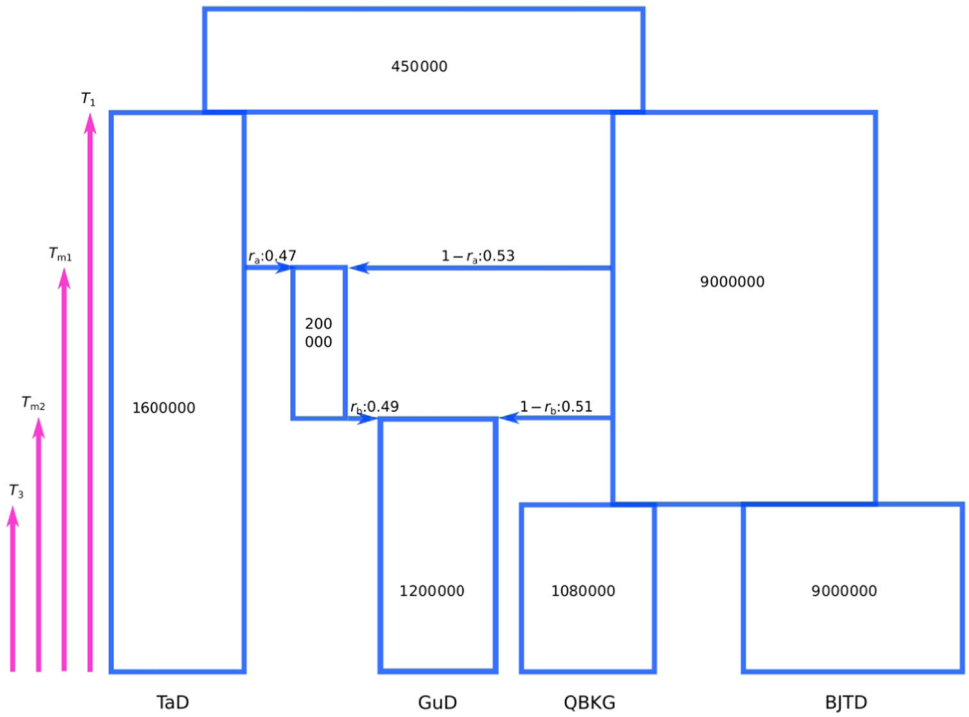

Fig. 3. DIYABC scenario shows the demographic history of *Reaumuria soongarica*. Numbers inside the box indicate the effective population size. r_a, the admixture proportion from the ancestral western lineage; r_b, the admixture proportion from the ancestral eastern lineage; T_1, time since divergence of the ancestral western and eastern lineages (0.714 Mya); T_{m1}, time since the first admixture of the ancestral western and eastern lineages (0.543 Mya); T_{m2}, time since the second admixture between the ghost population and the ancestral eastern lineage (0.287 Mya); T_3, time since the divergence of the Qaidam Basin–Kumtag Desert–Gashun Gobi Desert (QBKG) and Badain Jaran–Tengger Desert (BJTD) groups (0.159 Mya).

K method implemented in STRUCTURE HARVEST showed a peak at $K = 2$ (Fig. S4), suggesting that *R. soongarica* is chiefly clustered into western and eastern lineages. This finding is consistent with the divergence revealed by cpDNA and internal transcribed spacer (ITS) analyses (Li et al., 2012a; Yin et al., 2015). Second, the divergence between the western and eastern lineages was also supported by the best DIYABC scenario (Fig. 3), which showed that *R. soongarica* first divided into these two lineages around 0.714 Mya. Third, other analyses, such as our PCoA (Fig. S5) and NJ tree (Fig. S2) results, were also found to support a divergence between the western and eastern lineages. Taken together, it seems incontrovertible that during the evolutionary history of *R. soongarica*, a clear genetic divergence between western and eastern lineages occurred.

Table 2 Type I and type II error rates for the second round DIYABC simulated data

	Logistic regression	
Scenario	False negatives	False positives
1	0.216	0.176
2	0.328	0.456
3	0.302	0.298
4	0.268	0.184

Type I and type II error rates of each evolutionary model were calculated using a logistic method based on 1% of the simulated datasets. Scenarios with admixture events (scenarios 1 and 4) had lower type I and type II error rates than other scenarios.

The divergence between western and eastern lineages could be attributable to geographic isolation and/or ecological differentiation. The divergence time (0.714 Mya [0.253–1.210 Mya]) is consistent with the Kunlun–Yellow River tectonic movement (1.2–0.7 Mya, Cui et al., 1998) and the Naynayxungla glaciation (0.72–0.50 Mya, Zheng et al., 2002). The Kunlun–Yellow River tectonic movement caused the rise of the Kunlun Mountains and the uplift of the QTP to 3000 m. It also induced the most extensive glaciations of the Quaternary in China, the Naynayxungla Glaciation, which began approximately 1.20 Mya and reached its maximum between 0.60 and 0.80 Mya (Shi & Ren, 1990; Zhou & Li, 1998; Zheng et al., 2002). During this period, the atmospheric circulation in these areas changed dramatically, resulting in a change in the climate of China from hot and humid to cold and dry, and causing an especially strong drying-out of the northwestern QTP (Cui et al., 1998). In response to these changes, many other species, such as schizothoracine fishes (He et al., 2016) and *Hippophae tibetana* Schltdl. (Jia et al., 2011), also diverged. Thus, it is possible that the genetic divergence and the reduction of gene flow between the western and eastern lineages of *R. soongarica* might have been caused by geographic isolation.

In addition to geographic isolation, the genetic divergence between western and eastern lineages would also be established by ecological differentiation. Before restricted geographic

Table 3 Precision of parameter estimation based on 500 pseudo-observed datasets

Parameter	RRMISE	RMeanAD	RRMSE	50% Coverage	95% Coverage	Factor 2	RMB
N_1	0.159	0.120	0.117	0.456	0.954	1.000	−0.008
N_2	0.213	0.158	0.158	0.538	0.948	1.000	−0.008
N_3	0.163	0.121	0.114	0.528	0.956	1.000	−0.007
N_4	0.249	0.179	0.177	0.554	0.968	0.996	−0.011
N_5	26.796	2.580	6.319	0.466	0.968	0.608	0.719
N_a	0.294	0.216	0.200	0.532	0.954	0.998	−0.012
T_1	0.230	0.169	0.154	0.520	0.970	1.000	0.002
T_{m1}	0.369	0.273	0.247	0.566	0.968	0.980	−0.069
T_{m2}	0.385	0.271	0.248	0.536	0.950	0.994	0.003
T_3	0.208	0.154	0.150	0.498	0.942	1.000	−0.016
r_a	0.377	0.282	0.309	0.516	0.956	0.972	0.038
r_b	9.187	1.877	9.367	0.504	0.960	0.524	0.990

Estimates were made under the best DIYABC scenario, with the square root of the relative mean square error (RRMSE) and the relative median absolute deviation (RMeanAD). Factor 2, proportion of the test datasets for which the point estimate is at least half and at most twice the true value; RMB, relative median bias; RRMISE, square root of the relative mean integrated square error.

isolation builds up, the genetic divergence between western and eastern lineages might also have been established by ecological differentiation, even in the presence of substantial intralineage gene flow (Turelli et al., 2014). Bidirectional gene flow between the TaD and BJTD groups was a model that could not be rejected by our LLR results (Table S8), suggesting that western and eastern lineages were not strictly geographically isolated. After the emergence of geographical boundaries between the lineages, reproductive isolation could be facilitated by existing patterns of ecological differentiation (Mayr, 1942, 1947; Ellstrand, 1992; Mallet et al., 2009). The dramatic climate changes in the Quaternary period might have enhanced the patterns of ecological differentiation between the western and eastern lineages. For example, the formation of deserts in the west, such as the TaD, were less affected by the East Asian monsoon system than their eastern desert counterparts (Yin et al., 2015). Moreover, the western and eastern deserts had opposite air current patterns during the Quaternary glacial–interglacial cycle

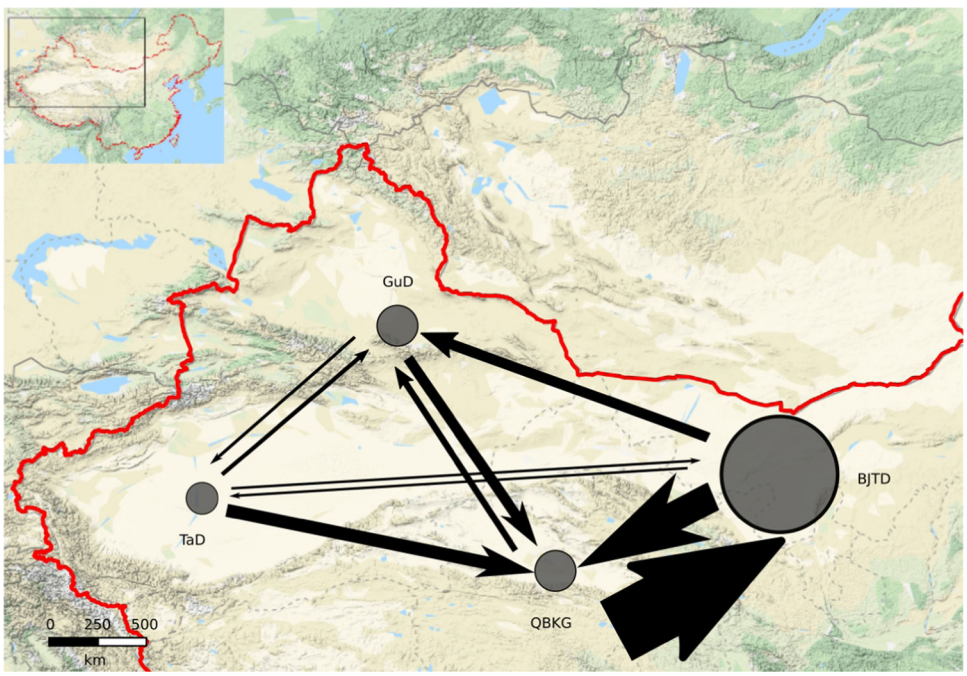

Fig. 4. Summary of estimated historical migration patterns among four groups of *Reaumuria soongarica*. Migration rates are scaled against mutation rates, and the width of the arrows reflects relative migration rates, with migration from Badain Jaran–Tengger Desert (BJTD) to Taklimakan Desert (TaD) being the lowest. Due to the large arrow scale, the migration rates between the Qaidam Basin–Kumtag Desert–Gashun Gobi Desert (QBKG) and BJTD were divided by 10. Map was retrieved from Google Earth (https://www.google.com/earth/). GuD, Gurbantünggüt Desert.

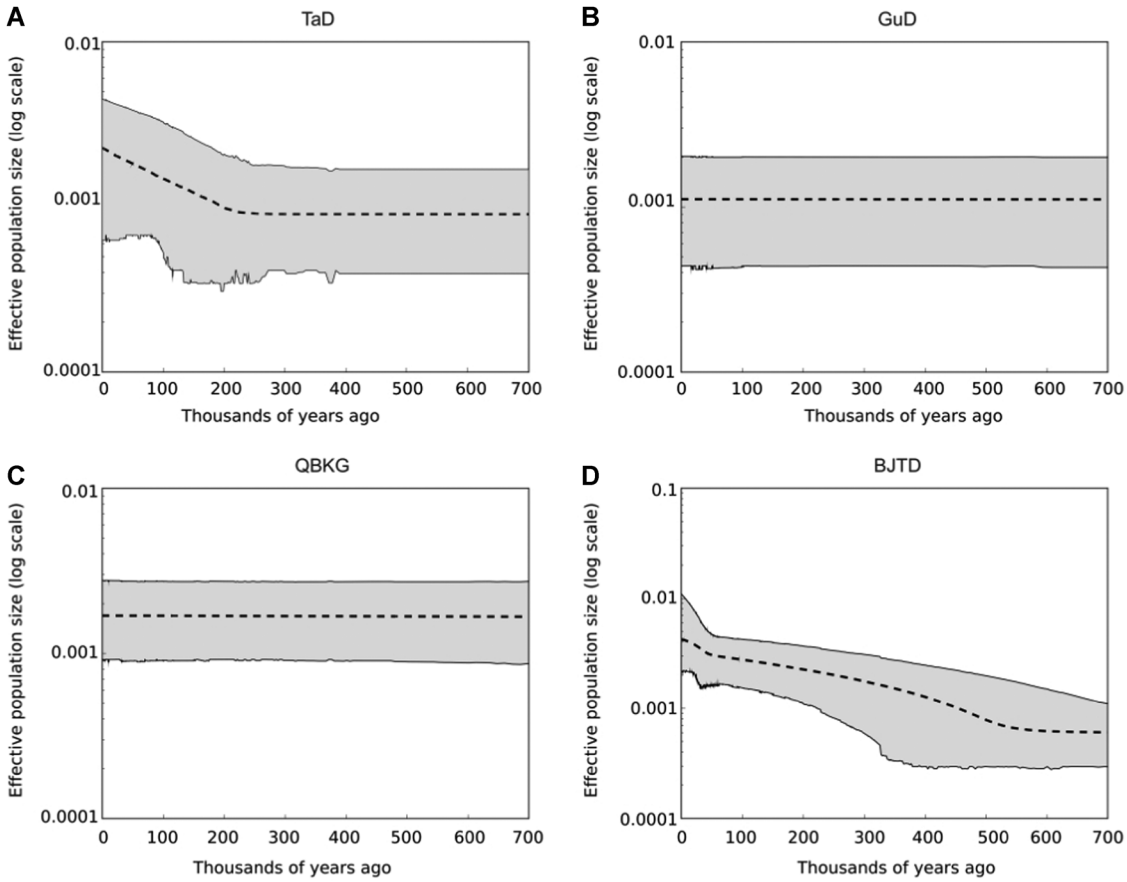

Fig. 5. Variation in effective population size through time based on Extended Bayesian Skyline Plot (EBSP) analyses. **A,** Taklimakan Desert (TaD) showed population expansion that started approximately 0.20–0.25 Mya. **B,** Gurbantünggüt Desert (GuD) showed no historical changes in population size. **C,** TaD showed no historical changes in population size. **D,** Badain Jaran–Tengger Desert (BJTD) showed a population expansion that started approximately 0.55–0.60 Mya. The dotted horizontal line shows the median estimate of the EBSP and the gray line shows the upper and lower 95% highest posterior density limits. QBKG, Qaidam Basin–Kumtag Desert–Gashun Gobi Desert.

(Fang et al., 2002). In addition, identity test results (Fig. 6) suggested that the western and eastern lineages showed significant ecological niche differentiation, and this pattern is consistent with differentiation patterns found in other species (e.g., the ecological differentiation between the western desert population of *Populus pruinosa* Schrenk and the eastern desert population of *P. euphratica* Olivier due to differing underground water levels) (Wang et al., 2014b). Taken together, our results suggest that the genetic divergence between the western and eastern lineages of *R. soongarica* can be attributed to both geographic isolation and ecological differentiation.

4.2 Hybrid origin of Northern lineage of *R. soongarica*

Our data also suggested that contact between the western and eastern lineages facilitated the hybrid origin of the northern lineage. Rather than evolving directly from ancestors or other lineages, the northern lineage might have originated through hybridization. In our previous study, we found that the cpDNA and ITS haplotypes present in the northern lineage were all non-ancestral (i.e., derived) (Yin et al., 2015). This finding was consistent with the recent formation of the flora of the Junggar Basin, many of which were not relic species (Liu, 1995). Thus, the GuD was unlikely to be a refuge for ancestral *R. soongarica*. In addition, the hypothesis that the northern lineage had differentiated directly from the ancestral population was rejected by the DIYABC results (Fig. S1, first-round simulation model 5). Moreover, unstable phylogenetic trees produced by BEAST (Fig. S6), and the alternative assignment of the northern lineage by STRUCTURE at $K = 3$ (Fig. 2) together suggested that the northern lineage probably did not originate either from only the eastern or western lineages. In addition, the five trees generated by PhyloNet (Fig. S7) implied a hybrid origin of the northern lineage. Taken together, these results indicate that, rather than having a single origin, the northern lineage originated by hybridization.

The hybrid origin of the northern lineage could be attributed to multiple (i.e., two or more) contacts between the western and eastern lineages. Our data best fits the DIYABC scenario (Fig. 3) that specifies two hybridization events between them. This scenario is reliable as most estimated parameters were robust. Although estimates of r_b and N_5 showed uncertainty, the 95% credible interval of the parameters crucial for the hybrid hypothesis (i.e., r_a, 0.03–0.96 and r_b, 0.04–0.94) did not include one or zero, indicating that the genome of the northern lineage

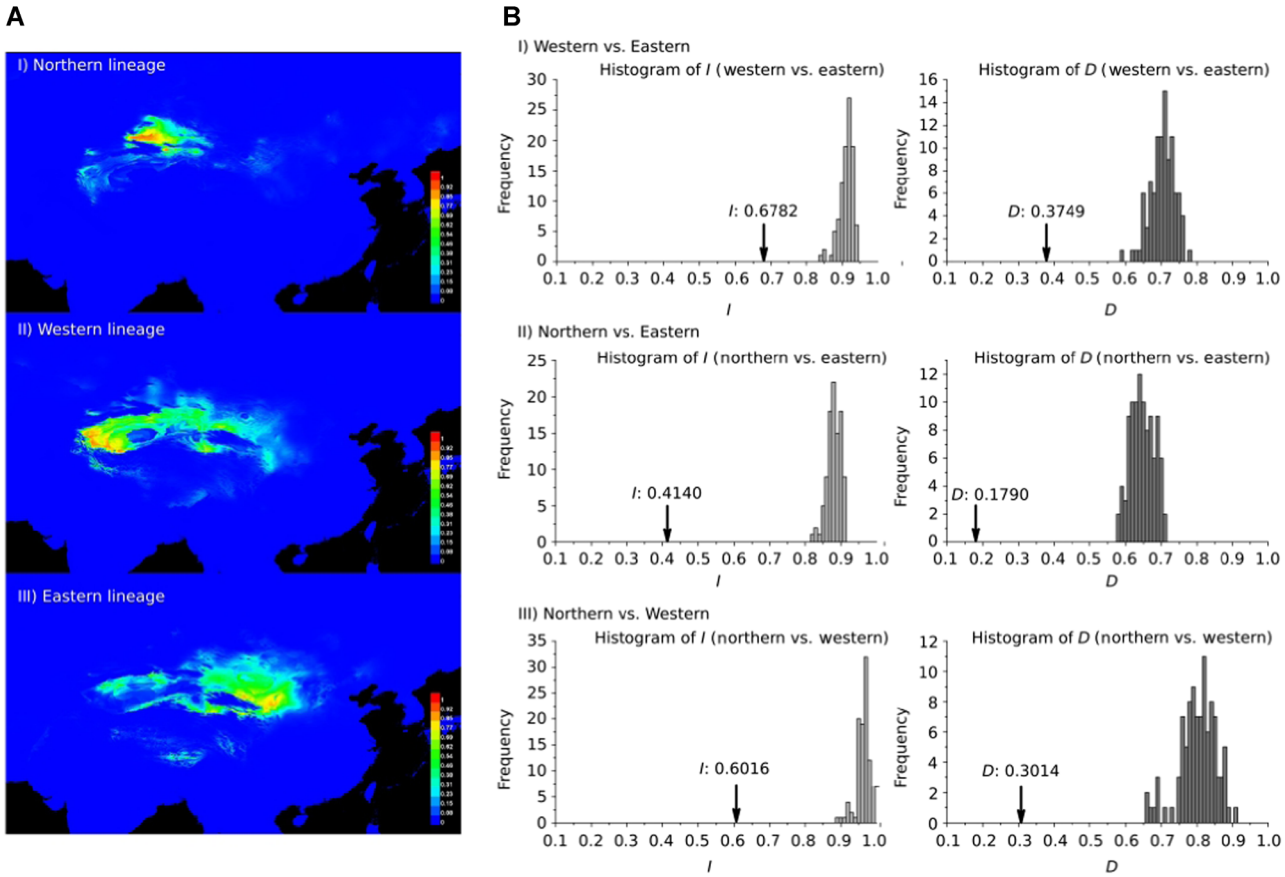

Fig. 6. Ecological niche modelling (ENM) results suggest three significant divergent niches of *Reaumuria soongarica*. **A,** Geographic projections of the northern, western, and eastern lineages of *R. soongarica* niche derived from the MaxEnt model. **B,** Results of identity tests. Observed niche overlap values of Schoener's *D* and *I* for each pairwise combination of lineages compared with null distributions (bars indicate the null distributions of *D* or *I*). All pairwise comparisons show significant niche divergence ($P < 0.05$).

was composed of both the western and eastern lineages to some degree. Moreover, the type I error (scenarios 1, 0.216; and scenarios 4, 0.268 in round 1 simulation) and type II error (scenarios 1, 0.176; and scenarios 4, 0.184 in round 1 simulation) rates were the lowest in the hybridization/admixture models, suggesting that models without hybridization could be rejected. The two-hybridization model (model 4) displayed a much higher posterior probability than the one-hybridization model (model 1), and was supported by both the direct and logistical methods (Fig. S9). Owing to the uncertainty of ghost populations (N_5), we did not test scenarios with more than two contacts. However, we could not rule out the possibility of a higher number (i.e., ≥ 2) contacts between the western and eastern lineages. Multiple hybridization events were also discovered in the homoploid hybrid speciation patterns of other plants (note that in this case the northern lineage of *R. soongarica* can be considered a homoploid hybrid species due to the fact that it shows no changes in ploidy level [data not shown]). Such events have been found in sunflowers (Rieseberg, 1991; Schwarzbach & Rieseberg, 2002) and *Pinus densata* Masters (Wang et al., 2001; Ma et al., 2006; Gao et al., 2012). Taken together, our data suggest that the formation of the northern lineage of *R. soongarica* can be attributed to at least two hybridization events.

The hybridization events were likely caused by asynchronous historical expansions of the western and eastern desert ecosystems, which expanded the historical habitat of *R. soongarica* in the TaD and BJTD populations. The expansions of the TaD and BJTD populations are supported by multiple lines of evidence. First, the Tajima's *D* and Fu's F_s values of the TaD and BJTD populations were both negative, with BJTD showing a more significant result (Fig. S12). Negative values of Tajima's *D* indicate an excess of rare alleles (Tajima, 1989); negative values of Fu's F_s indicate an excess number of alleles relative to the null expectation (Fu, 1997). In addition, MFDM tests suggested no significant deviation from neutrality. Thus, rather than showing a selective sweep, our data is consistent with a recent population expansion (Haddrill et al., 2005; Pyhajarvi et al., 2007; Li et al., 2010, 2012b). Second, the EBSP test results showed significant historical expansions of the BJTD and TaD populations (Fig. 5), and these were consistent with the two periods of glaciation in the Pleistocene (i.e., the Naynayxungla and Guxiang glaciations). The EBSP results suggested an expansion of the BJTD group initiated approximately 0.55–0.60 Mya, as well as an expansion of the TaD group initiated approximately 0.20–0.25 Mya. These expansion periods were all found in the 95% credible interval of the two hybridization events simulated

by DIYABC (Table 1; 95% credible interval of T_{m1} and T_{m2}, 0.138–0.870 Mya). During these periods, the Naynayxungla glaciation (the largest glaciation in the Middle Pleistocene, 0.72–0.50 Mya) and the Guxiang glaciation (0.30–0.13 Mya) appeared sequentially (Zheng et al., 2002). In both cases, the climate became colder and dryer, which aggravated the aridity of ACA and expanded the habitat available to *R. soongarica*. Finally, the GuD (formed before 0.80 Mya) oscillated dramatically during the period from 0.4 to 0.13 Mya (Shi et al., 2006), increasing the possibility of multiple contact points between the western and eastern lineages. For these three reasons, we propose that recent changes in climate and continuing desertification resulted in a massive desert expansion, thereby expanding the habitat available to the existing TaD and BJTD groups, and facilitating multiple hybridization events between the western and eastern lineages of *R. soongarica*.

4.3 Northern lineage might be an incipient species located at a rare point along the speciation continuum

We further propose that the northern lineage could be an incipient species, located at a special point along the speciation continuum. This could be the case because the northern lineage might be genetically stable, with gene flow balanced with ecological differentiation, and because it has arisen from hybridization recently. Further discussion of each of these two points follows.

First, the genetic boundary of the northern lineage, although facing ongoing gene flow, has been maintained by significant ecological differentiation. Our analyses suggested that incoming gene flow is consistent (Figs. S10, S11). Signatures of smaller bidirectional migration were detected between the TaD and GuD populations (Fig. 4), but larger bidirectional migrations were detected between the QBKG and GuD groups, and strong unidirectional gene flow was detected from the BJTD to GuD group. These gene flow patterns confirm our previous conclusion that East Asian monsoons play an essential role in shaping the pattern of migration between the northern and eastern lineages (Yin et al., 2015). Moreover, they are similar to those of *P. euphratica*, another desert plant (Zeng et al., 2018). Taken together, our data suggest that the northern lineage is still facing incoming gene flow, but it is also strongly genetically differentiated from both the western and eastern lineages. In addition, the population size and genetic composition of the northern lineage are not undergoing dramatic changes in response to this gene flow. Furthermore, distinct ecological niches have been detected for each of the three lineages.

The possibility that the northern lineage could be an incipient species (i.e., an emerging hybrid species) is intriguing because it is at a rare point along the speciation continuum. For a long time, species were defined as reproductively isolated populations (Mayr, 1969). However, speciation is now generally considered to be a continuous process (Harrison, 1998; Via, 2009); for this reason, we can regard a cluster of multilocus genotypes as a species, so long as it remains genetically stable and distinct if it comes into contact with either parent (Mallet, 1995; Coyne & Orr, 2004). This definition describes the northern lineage of *R. soongarica* fairly well: it is both distinctly genetically differentiated from western and eastern lineages and remains genetically stable. In addition, there is increasing genetic evidence that hybridization is more common in plant evolutionary history than has been generally accepted in the past (Grant, 1981; Mallet, 2007). However, given the ubiquity of hybridization, hybrid speciation is less well studied, even in plants. Few situations where hybrid speciation, and especially homoploid hybrid speciation, have yet been documented. The limited number of examples includes sunflowers (Rieseberg, 1991; Schwarzbach & Rieseberg, 2002) and *P. densata* (Wang et al., 2001; Ma et al., 2006; Gao et al., 2012). The fact that the data show that hybrid speciation is rare reflects the difficulties in establishing an independent hybrid species: hybrid speciation requires a new niche that is not occupied by parents (Gross & Rieseberg, 2005), but it also requires a reduction in gene flow from parents to prevent hybrids from disappearing by genetic fusion (Mallet, 2007; Ward et al., 2012). In this study, we detected significant ecological differentiation between the northern lineage and the others, supporting the conclusion that the northern lineage colonized new habitats far from the original parental habitat. Moreover, we found that the flowering time of the northern lineage (i.e., from end of June to end of August) is different from the western lineage (i.e., from end of May to end of June) and the eastern lineage (i.e., from end of June to early September). This might be an early indication of the initiation of reproductive isolation, and if true, in the future we should expect reduced gene flow between the northern lineage and the other lineages. For these reasons, it is reasonable for us to consider the northern lineage to be an emerging hybrid species. Clearly, more detailed evidence is needed to fully understand the complexity of the phenotype and evolutionary history of the northern lineage. For example, future studies should identify the phenotypic differentiation and to assess the depth of their reproductive isolation, and should attempt to characterize the genomic diversity of the northern lineage. Regardless of of whether the northern lineage will eventually evolve into an entirely new species, this emerging lineage offers a unique opportunity to investigate the evolutionary dynamics of a hybrid population shortly after secondary contact (Via, 2009; Nolte & Tautz, 2010). The northern lineage could thereby provide us with an additional vantage point to better understand the mechanisms associated with hybrid species formation.

5 Conclusion

Our data suggested that the desert plant *Reaumuria soongarica* underwent a multiphase speciation process caused by various geological events and climatic changes in the Quaternary. First, due to the geographic isolation and ecological differentiation caused by the Kunlun–Yellow River tectonic movement and the Naynayxungla glaciation, *R. soongarica* was genetically divided into western and eastern lineages approximately 0.714 Mya. Second, the origination of the northern lineage was attributed to multiple hybridization events between the western and eastern lineages, both of which were dramatically influenced by asynchronous desertification caused by the Naynayxungla and Guxiang glaciation events. Third, although facing ongoing gene flow from other populations, the genetic background of the northern lineage has been maintained by significant ecological differentiation, suggesting that it could be an incipient species at a critical point along the speciation continuum. The divergence and

hybridization history of *R. soongarica* provides compelling evidence that plant speciation can be a multilevel process that is strongly determined by geological events and climatic changes (Abbott et al., 2013). Our study confirmed that, despite its complicated evolutionary history, desert plants such as *R. soongarica* are useful model systems for understanding the nature of the speciation continuum.

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (NSFC, Grant Nos. 31370395, 31500266, 31770416, and 31571311), the "One Hundred Talents" projects of the Chinese Academy of Sciences (Grant No. 29Y127E71), the Shenzhen Science and Technology Research Fund (JCYJ20150630165133393), and by the Postdoctoral Targeted Fund of the Ministry of Human and Social Resources of Yunnan Province, China. We have no conflicts of interest to declare with this manuscript. We sincerely appreciate the valuable comments and suggestions on this manuscript from Dr. Daoura Goudia Bachir. We are indebted to the heartfelt help of the sampling work of Yuan Liu, Haiting Hao, and Zhengyin Xu, as well as laboratory assistance from Lirong Wang and Wen Zhang. We acknowledge the computing resources and time on the Linux cluster of the Supercomputing Center of the Cold and Arid Region Environment and Engineering Research Institute, Chinese Academy of Sciences.

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J, Keller B, Marczewski T, Mallet J, Martinez-Rodriguez P, Möst M, Mullen S, Nichols R, Nolte AW, Parisod C, Pfennig K, Rice AM, Ritchie MG, Seifert B, Smadja CM, Stelkens R, Szymura JM, Väinölä R, Wolf JBW, Zinner D. 2013. Hybridization and speciation. *Journal of Evolutionary Biology* 26: 229–246.

Arnold ML. 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology* 13: 997–1007.

Baack E, Melo MC, Rieseberg LH, Ortiz-Barrientos D. 2015. The origins of reproductive isolation in plants. *New Phytologist* 207: 968–984.

Barraclough TG, Vogler AP. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155: 419–434.

Beerli P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22: 341–345.

Bermingham E, Rohwer S, Freeman S, Wood C. 1992. Vicariance biogeography in the pleistocene and speciation in North American wood warblers: A test of Mengel's model. *Proceedings of the National Academy of Sciences USA* 89: 6624–6628.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Buerkle CA, Morris RJ, Asmussen MA, Rieseberg LH. 2000. The likelihood of homoploid hybrid speciation. *Heredity* 84: 441–451.

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.

Coyne JA, Orr HA. 2004. *Speciation*. Sunderland: Sinauer Associates.

Cui Z, Wu Y, Liu G, Ge D, Pang Q, Xu Q. 1998. On Kunlun–Yellow River tectonic movement. *Science in China Series D: Earth Sciences* 41: 592–600.

Dang ZH, Zheng LL, Wang J, Gao Z, Wu SB, Qi Z, Wang YC. 2013. Transcriptomic profiling of the salt-stress response in the wild recretohalophyte *Reaumuria trigyna*. *BMC Genomics* 14: 29.

Darwin C. 1859. *On the origins of species by means of natural selection*. London: John Murray.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214–221.

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.

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.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.

Ellstrand NC. 1992. Gene flow among seed plant populations. *New Forests* 6: 241–256.

Fang XM, Lu LQ, Yang SL, Li JJ, An ZS, Jiang P, Chen XL. 2002. Loess in Kunlun Mountains and its implications on desert development and Tibetan Plateau uplift in west China. *Science in China Series D: Earth Sciences* 45: 289–299.

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.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.

Gao J, Wang B, Mao JF, Ingvarsson P, Zeng QY, Wang XR. 2012. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. *Molecular Ecology* 21: 4811–4827.

Grant V. 1981. *Plant speciation*. New York: Columbia University Press.

Gross BL, Rieseberg LH. 2005. The ecological genetics of homoploid hybrid speciation. *Journal of Heredity* 96: 241–252.

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. 2005. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research* 15: 790–799.

Hall TA. 1999. Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95–98.

Harrison RG. 1998. Linking evolutionary pattern and process. In: Howard DJ, Berlocher SH eds. *Endless forms: Species and speciation*. New York: Oxford University Press. 19–31.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160–174.

He D, Chen Y, Liu C, Tao J, Ding C, Chen Y. 2016. Comparative phylogeography and evolutionary history of schizothoracine fishes in the Changtang Plateau and their implications for the lake level and Pleistocene climate fluctuations. *Ecology and Evolution* 6: 656–674.

Hendry AP. 2009. Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences* 66: 1383–1398.

Hendry AP, Bolnick DI, Berner D, Peichel CL. 2009. Along the speciation continuum in sticklebacks. *Journal of Fish Biology* 75: 2000–2036.

Hey J. 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution* 27: 905–920.

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.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164.

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.

Kahle D, Wickham H. 2013. ggmap: Spatial visualization with ggplot2. *The R Journal* 5: 144–161.

Langella O. 1999. Populations version 1.2.30. Available from http://bioinformatics.org/populations/ [accessed September 2018].

Lexer C, Welch ME, Raymond O, Rieseberg LH. 2003. The origin of ecological divergence in *Helianthus paradoxus* (asteraceae): Selection on transgressive characters in a novel hybrid habitat. *Evolution* 57: 1989–2000.

Li HP. 2011. A new test for detecting recent positive selection that is free from the confounding impacts of demography. *Molecular Biology and Evolution* 28: 365–375.

Li Y, Stocks M, Hemmila S, Kallman T, Zhu HT, Zhou YF, Chen J, Liu JQ, Lascoux M. 2010. Demographic histories of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple nuclear loci. *Molecular Biology and Evolution* 27: 1001–1014.

Li ZH, Chen J, Zhao GF, Guo YP, Kou YX, Ma YZ, Wang G, Ma X-F. 2012a. Response of a desert shrub to past geological and climatic change: A phylogeographic study of *Reaumuria soongarica* (Tamaricaceae) in western China. *Journal of Systematics and Evolution* 50: 351–361.

Li ZH, Zou JB, Mao KS, Lin K, Li HP, Liu JQ, Kallman T, Lascoux M. 2012b. Population genetic evidence for complex evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan Plateau. *Evolution* 66: 831–845.

Librado P, Rozas J. 2009. Dnasp v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Liu Y. 1995. A study on origin and formation of the Chinese desert floras. *Acta Phytotaxonomica Sinica* 33: 131–143.

Ma X-F, 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.

Mallet J. 1995. A species definition for the modern synthesis. *Trends in Ecology and Evolution* 10: 294–299.

Mallet J. 2007. Hybrid speciation. *Nature* 446: 279–283.

Mallet J. 2008. Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363: 2971–2986.

Mallet J, Meyer A, Nosil P, Feder JL. 2009. Space, sympatry and speciation. *Journal of Evolutionary Biology* 22: 2332–2341.

Mayr E. 1942. *Systematics and the origin of species, from the viewpoint of a zoologist*. Cambridge: Harvard University Press.

Mayr E. 1947. Ecological factors in speciation. *Evolution* 1: 263–288.

Mayr E. 1963. *Animal species and evolution*. Cambridge: Harvard University Press.

Mayr E. 1969. The biological meaning of species. *Biological Journal of the Linnean Society* 1: 311–320.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* 19: 153–170.

Nielsen R, Wakeley J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* 158: 885–896.

Nolte AW, Tautz D. 2010. Understanding the onset of hybrid speciation. *Trends in Genetics* 26: 54–58.

Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution* 24: 145–156.

Peakall R, Smouse PE. 2012. Genalex 6.5: Genetic analysis in Excel. Population genetic software for teaching and research: An update. *Bioinformatics* 28: 2537–2539.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Posada D. 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

Pyhajarvi T, Garcia-Gil MR, Knurr T, Mikkonen M, Wachowiak W, Savolainen O. 2007. Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics* 177: 1713–1724.

Reilly SB, Marks SB, Jennings WB. 2012. Defining evolutionary boundaries across parapatric ecomorphs of black salamanders (*Aneides flavipunctatus*) with conservation implications. *Molecular Ecology* 21: 5745–5761.

Rieseberg LH. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): Evidence from ribosomal genes. *American Journal of Botany* 78: 1218–1237.

Rieseberg LH, Ellstrand N, Arnold M. 1993. What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences* 12: 213–241.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.

Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* 317: 910–914.

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

Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology and Evolution* 16: 372–380.

Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737–741.

Schoener TW. 1970. Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology* 51: 408–418.

Schwarzbach AE, Rieseberg LH. 2002. Likely multiple origins of a diploid hybrid sunflower species. *Molecular Ecology* 11: 1703–1715.

Shi Y, Yan X, Zhao PS, Yin HX, Zhao X, Xiao HL, Li XR, Chen GX, Ma X-F. 2013. Transcriptomic analysis of a tertiary relict plant, extreme xerophyte *Reaumuria soongorica* to identify genes related to drought adaptation. *PLoS ONE* 8: e63993.

Shi YF, Ren JW. 1990. Glacier recession and lake shrinkage indicating a climatic warming and drying trend in central Asia. *Annals of Glaciology* 14: 261–265.

Shi ZT, Song YG, An ZS. 2006. Evolution of Gurbantunggut Desert recorded by Tianshan Loess. *Journal of Desert Research* 26: 675–679.

Stephens M, Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 73: 1162–1169.

Tajima F. 1983. Evolutionary relationship of DNA-sequences in finite populations. *Genetics* 105: 437–460.

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

Templeton AR, Shaw K, Routman E, Davis SK. 1990. The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden* 77: 13–27.

Than C, Ruths D, Nakhleh L. 2008. Phylonet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9: 322.

Turelli M, Lipkowitz JR, Brandvain Y. 2014. On the Coyne and Orr-Igin of species: Effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on *Drosophila* speciation. *Evolution* 68: 1176–1187.

Via S. 2009. Natural selection in action during speciation. *Proceedings of the National Academy of Sciences USA* 106: 9939–9946.

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

Wang BH, Zhang Y, Wei PP, Sun M, Ma X-F, Zhu XY. 2014a. Identification of nuclear low-copy genes and their phylogenetic utility in rosids. *Genome* 57: 547–554.

Wang J, Kallman T, Liu JQ, Guo Q, Wu Y, Lin K, Lascoux M. 2014b. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity* 112: 156–164.

Wang X, Wei HT, Taheri M, Khormali F, Danukalova G, Chen FH. 2016. Early Pleistocene climate in western arid central Asia inferred from loess-palaeosol sequences. *Scientific Reports* 6: 20560.

Wang XR, Szmidt AE, Savolainen O. 2001. Genetic composition and diploid hybrid speciation of a high mountain pine, *Pinus densata*, native to the Tibetan Plateau. *Genetics* 159: 337–346.

Waples RS, Gaggiotti O. 2006. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* 15: 1419–1439.

Ward JL, Blum MJ, Walters DM, Porter BA, Burkhead N, Freeman B. 2012. Discordant introgression in a rapidly expanding hybrid swarm. *Evolutionary Applications* 5: 380–392.

Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution* 62: 2868–2883.

Warren DL, Glor RE, Turelli M. 2010. ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography* 33: 607–611.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256–276.

Wickham H. 2016. *ggPlot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Woerner AE, Cox MP, Hammer MF. 2007. Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23: 1851–1853.

Wright S. 1931. Evolution in mendelian populations. *Genetics* 16: 97–159.

Yin H, Yan X, Shi Y, Qian C, Li Z, Zhang W, Wang L, Li Y, Li X, Chen G, Li X, Nevo E, Ma X-F. 2015. The role of East Asian monsoon system in shaping population divergence and dynamics of a constructive desert shrub *Reaumuria soongarica*. *Scientific Reports* 5: 15823.

Zeng YF, Zhang JG, Abuduhamiti B, Wang WT, Jia ZQ. 2018. Phylogeographic patterns of the desert poplar in northwest China shaped by both geology and climatic oscillations. *BMC Evolutionary Biology* 18: 75.

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.

Zhou S, Li J. 1998. The sequence of Quaternary glaciation in the Bayan Har Mountains. *Quaternary International* 45: 135–142.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12490/suppinfo:

Fig. S1. Eight scenarios tested in the first-round DIYABC simulation.

Fig. S2. Neighbour-joining tree of 41 populations with the support of 1000 bootstraps.

Fig. S3. Four scenarios tested in the second-round DIYABC simulation.

Fig. S4. Results indicating the best fit for the number of K groups of *R. soongarica*.

Fig. S5. Principal coordinate analyses (PCoA) results illustrating the genetic relationships of 41 *R. soongarica* populations.

Fig. S6. Three alternative trees produced by *BEAST.

Fig. S7. Top five best topology-based phylogeny trees inferred by PhyloNet.

Fig. S8. The best four scenarios from the first-round simulation.

Fig. S9. The best scenario from the second-round simulation.

Fig. S10. Migration histograms estimated by IMa2.

Fig. S11. Migration rates with confidence intervals produced by MIGRATE-N.

Fig. S12. Boxplot of Tajima's *D* and Fu's F_s values among four groups.

Table S1. Detailed geographical location and grouping of 41 sampled populations of *R. soongarica* from China.

Table S2. The primer sequences, annealing temperature, gene length and putative function of thirteen nuclear loci used in the current study.

Table S3. The best-fit substitution model for each locus in each group as determined using jModeltest to compare Akaike Information Criterion values.

Table S4. The prior distribution of parameters used for the description of the scenarios analyzed in DIYABC.

Table S5. List of bioclimatic variables defining the ecological niche of *R. soongarica*.

Table S6. Nucleotide variation, haplotype diversity and MFDM tests for thirteen nuclear loci of *R. soongarica*.

Table S7. All 25 possible nested models within the full two-population IM model.

Table S8. Results of the LLR tests of pairwise comparisons among the four groups.