Genome wide sequencing provides evidence of adaptation to heterogeneous environments for the ancient relictual *Circaeaster agrestis* (Circaeasteraceae, Ranunculales)

Running title: Adaptive evolution of *Circaeaster agrestis*

Xu Zhang[1, 2, 3, #], Yanxia Sun[1, 2, #, *], Jacob B. Landis[4], Jianwen Zhang[5], Linsen Yang[6], Nan Lin[1, 2, 3], Huajie Zhang[1, 2], Rui Guo[1, 2, 3], Lijuan Li[1, 2, 3], Yonghong Zhang[5], Tao Deng[5], Hang Sun[5, *], Hengchang Wang[1,2, *]

[1]CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, China.

[2]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan 430074, Hubei, China.

[3]University of Chinese Academy of Sciences, Beijing, 100049 China.

[4]Department of Botany and Plant Sciences, University of California Riverside, Riverside, CA, USA.

[5]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, 650201 China.

[6]Hubei Key Laboratory of Shennongjia Golden Monkey Conservation Biology, Administration of Shennongjia National Park, Shennongjia, Hubei, China.

[#]These authors contributed equally to this work.

*Correspondence: sunyanxia@wbgcas.cn; sunhang@mail.kib.ac.cn; hcwang@wbgcas.cn

Summary

- Investigating the interaction between environmental heterogeneity and local adaptation is critical to understand the evolutionary history of a species, providing the premise for studying the response of organisms to rapid climate change. However, for most species how exactly the spatial heterogeneity promotes population divergence and how genomic variations contribute to adaptive evolution remain poorly understood.
- We examine the contributions of geographical and environmental variables to population divergence of the relictual, alpine herb *Circaeaster agrestis* as well as genetic basis of local adaptation using RAD-seq and plastome data.
- We detected significant genetic structure with an extraordinary disequilibrium of genetic diversity among regions, and signals of isolation-by-distance along with isolation-by-resistance. The populations were estimated to begin diverging in the late Miocene, along with a possible ancestral distribution of the Hengduan Mountains and adjacent regions. Both environmental gradient and redundancy analyses revealed significant association between genetic variation and temperature variables. Genome-environment association analyses identified 16 putatively adaptive loci related to biotic and abiotic stress resistance.
- Our genome wide data provide new insights into the important role of environmental heterogeneity in shaping genetic structure, and access the footprints of local adaptation in an ancient relictual species.

Keywords: local adaptation, *Circaeaster agrestis*, RAD sequencing, genetic structure, environmental heterogeneity, temperature variables

Introduction

A primary goal of phylogeography is understanding how geographical and ecological factors shape the genetic structure and evolutionary history of species. Environmental factors can act in concert and vary in space, resulting in divergent natural selection leading to genetic divergence during adaptation to heterogeneous environments (Savolainen *et al.*, 2007). Such genetic divergence along environmental gradients, or across varied ecological habitats, can be indicative of local adaptation, a

mechanism that is beneficial to the long persistence of a population and intensify differentiation among populations (Kawecki & Ebert, 2004; Conover *et al.*, 2009; Colautti & Barrett, 2013; Savolainen *et al.*, 2013; Lowry *et al.*, 2019). Understanding the genetic basis of local adaptation will provide deeper insights into the response of organisms to ongoing climate change, aiding efforts to the conservation of genetic resources (Aitken *et al.*, 2008; Ahrens *et al.*, 2019).

On the one hand, divergent natural selection is the driving force of genetic divergence (Darwin, 1859; Coyne & Orr, 2004), while on the other hand, gene flow can buffer genetic differentiation (Spieth, 1974; Goicoechea *et al.*, 2019). Evolution resulting from local adaptation depends on both the strength of local selective pressures and the homogenizing effect of connectivity among different localities. Additionally, genetic drift may reinforce the establishment and maintenance of genetic divergence, particularly in populations with small effective size (Wright, 1946; Kimura & Crow, 1964). Therefore, quantifying the relative contributions among natural selection, genetic drift and population connectivity to genetic divergence is crucial for understanding the role of local adaptation in adaptive evolution of a species (Savolainen *et al.*, 2007; Friis *et al.*, 2018).

Phenotypic patterns that are associated with divergent selection have been observed in model plants (Mitchell-Olds *et al.*, 2007; Anderson *et al.*, 2011; Leinonen *et al.*, 2013; Zou *et al.*, 2017). However, identification of genetic loci that govern phenotypic variations in response to environmental pressures is a challenging task, especially for non-model species with limited genomic resources. Facilitated by advancements in sequencing technology and population genomics, it is now possible to study hundreds of thousands of loci improving our understanding of genome-wide effects of accumulating genetic divergence and genomic properties that influence the process of local adaptation (McCormack *et al.*, 2013; Ellegren, 2014; Seehausen *et al.*, 2014; Weigel & Nordborg, 2015). In particular, highly divergent loci identified by Genome-Environment Association (GEA) analyses can be interpreted as potential targets of divergent selection associated with population-specific environmental covariables (Coop *et al.*, 2010; Rellstab *et al.*, 2015; Hoban *et al.*, 2016; Forester *et al.*, 2018). In this context, plant studies are poised to reveal the genomic signatures underlying the evolution of local adaptation (e.g. Seehausen *et al.*, 2014; Ahrens *et al.*,

2019). However, for most species, such as relictual alpine herbs, how exactly the spatial heterogeneity promotes population diversification and how genomic variations contribute to adaptive evolution, in particular their interaction, remains poorly understood.

In the present study, we focus on investigating the evolutionary history and local adaptation of *Circaeaster agrestis* Maxim, an annual alpine herb with a relatively narrow habitat (Fu & Bartholomew, 2001; Sun, Y *et al.*, 2017). *Circaeaster* Maxim., with its only alliance *Kingdonia* Balf.f. & W.W. Smith, constitute the early-diverging eudicot family Circaeasteraceae (Ranunculales) (The Angiosperm Phylogeny et al., 2016). Due to special morphological characters such as open dichotomous leaf venation, this genus is of great interest to many botanists (Foster, 1971; Ren *et al.*, 2003). *Circaeaster* exhibits low morphological diversity with only one species, garnering a status of critical endangerment (Wild Plants Under State Protection in China). The distributions of *C. agrestis* is confined to the Qinghai-Tibet Plateau (QTP) and adjacent areas where exhibit a wide range of elevations from 2,100-5,000 m (Fu & Bartholomew, 2001). The uplift of QTP has created extraordinarily geomorphological and climatic diversity in this area. Due to the high mountain barrier formed by the Himalayas in the south, the central and western QTP (i.e. the Himalayas) are characterized by a cold and dry climate (Favre *et al*., 2015). In contrast, the eastern QTP (i.e. the Hengduan Mountains) and adjacent arears are associated with deep valleys and characterized mainly by a warm and wet climate (Ren *et al.*, 2017). Given the high degree of geographical and ecological heterogeneity, QTP comprises a promising model system for studying genetic signatures of local adaptation.

Specifically, we generated and analyzed two types of genomic datasets: single nucleotide polymorphisms (SNPs) derived from the restriction site associated DNA sequencing (RAD-Seq) of 18 *C. agrestis* populations and 20 plastome sequences of *C. agrestis* populations with other Ranunculales representatives. We aim to illustrate the evolutionary patterns of population divergence interacting with environmental variables and access the potential genetic basis of local adaptation. Here, we raise the following questions: 1) how alpine environments affect genetic structure within species and drive population divergence, and 2) how genetic variation is associated

with local adaptation? Answering these questions will help us understand species adaptation to rapidly changing climate, and aid in conservation of small patchy diversity in biodiversity hotspots.

Materials and Methods

Sampling, RAD library preparation and sequencing

A total of 139 individuals of *C. agrestis* from 18 localities were sampled (Table S1). Sampling locations were chosen based on existing occurrence records covering the geographic and climatic distribution of the species (Fig. 1). All materials were dried and stored in silica gel, and all the voucher specimens were deposited in Herbarium of Wuhan Botanical Garden (HIB) (Table S1). Total DNA was extracted from silica gel-dried leaves with a modified CTAB (Cetyl trimethylammonium bromide) method (Yang *et al.*, 2014). For RAD library construction and sequencing, genomic DNA was digested with the restriction enzyme EcoRI in a 30 ul reaction. T4 ligase was used to ligate a P1 adapter to the fragments. The adapter-ligated fragments were subsequently pooled, randomly sheared, and size-selected with a length of 350–550 bp. DNA was then ligated to a second adapter (P2), a Y adapter hat has divergent ends to the chosen DNA fragments. The ligation products were purified and PCR amplified, followed by gel purification, and finally size selected to 350-550 bp. Agilent 2100 Bioanaylzer and qPCR were used to qualify and quantify the sample library. Paired-end sequencing was performed on two lanes of Illumina HiSeq 2000 platform at BGI-Shenzhen (Shenzhen, Guangdong, China).

For plastome sequencing, 20 samples covering all 18 localities were chosen. For each sample, a 500-bp DNA TruSeq Illumina (Illumina Inc., San Diego, CA, USA) sequencing library was constructed using 2.5-5.0 ng sonicated DNA as input. Then the libraries were quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and real-time quantitative PCR. Libraries were multiplexed and sequenced using a 2×125 bp run on one lane of Illumina HiSeq 2000 platform at BGI-Shenzhen (Shenzhen, Guangdong, China).

Processing of Illumina data

For RAD-sequencing, pair-end Illumina reads were processed into RAD-tags using the STACKS

v.2.30 software pipeline (Catchen *et al.*, 2013). Samples were initially demultiplexed with PROCESS_RADTAGS. Reads with an average Phred score of at least 30, an unambiguous barcode, and restriction cut site were retained. Then the perl wrapper DENOVO_MAP.PL was used to identify optimal parameters by executing USTACKS, CSTACKS, and SSTACKS (Catchen *et al.*, 2013). Following the suggested protocol, we investigated a range of parameter values with M and n values ranging from 1 to 9 (fixing $M = n$) and $m = 3$ (Rochette & Catchen, 2017). We plotted two statistics that were useful in making the appropriate parameter choice: the number of polymorphic loci shared across most samples (the r80 loci) and the distribution of the number of SNPs per locus.

We filtered the catalog of reads using the POPULATIONS module in the STACKS pipeline to produce datasets for downstream population genetic analyses. Polymorphic RAD loci that presented in all 139 individuals were retained for subsequent analyses. To validate the influence of missing data, we also employed a loose parameter of - $r = 0.8$, in which at least 80% of individuals in a population were required to process a locus for that population. Potential homologs were excluded by removing loci showing heterozygosity > 0.5 within samples. We further filtered our dataset with a minor allele frequency (MAF) > 0.01 and kept only biallelic SNPs. Finally, per locus, we filtered our dataset by selecting the most informative SNP based on the number of minor alleles.

For genome-skimming sequencing, raw sequence reads were subsequently filtered using Trimmomatic v.0.36 (Bolger *et al.*, 2014) by removing duplicate reads and adapter-contaminated reads. Remaining high-quality reads were directly mapped to the reported plastome of *C. agrestis* (NCBI accession number: KY908400) using Geneious v9.0.2 (Kearse *et al.*, 2012). Initial annotations were implemented in the Plastid Genome Annotator (PGA) (Qu *et al.*, 2019), and then were refined by manual corrections in Geneious. The annotated plastomes were deposited in GenBank under accession numbers xxx-xxx.

Genetic diversity and population structure

We estimated genetic diversity indices including nucleotide diversity (π), expected heterozygosity (H_e), observed heterozygosity (H_o) and inbreeding coefficients (F_{IS}) using POPULATIONS based

on the RAD-seq data. The polymorphic sites in *Phylip* format were exported from POPULATIONS and were used for calculating haplotypes in DnaSP v6.12 (Rozas *et al.*, 2017). The *hierfstat* package (Goudet, 2005) in R v3.6.1 (R Team, 2014) was used to estimate population and group level F_{ST} using Weir and Cockerham's method (Weir & Cockerham, 1984).

Population genetic structure was estimated using two methods, Bayesian clustering and principal coordinate analysis (PCoA). Bayesian clustering was performed in STRUCTURE v2.3.4 (Pritchard *et al.*, 2000) with the admixture model. As STRUCTURE analysis assumes that loci are unlinked, we thus filtered out linked SNPs using the *–write_single_snp* option in the POPULATIONS module (Catchen *et al.*, 2013). To determine the optimal number of groups (*K*), we ran STRUCTURE 10 times for each *K* value from 1 through 10. Each run was performed for 200,000 Markov Chain Monte Carlo (MCMC) generations with a burn-in period of 100,000 generations. The optimal *K* was chosen using both delta-*K* method implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012) and cross-entropy criterion implemented in R package *LEA* v2.8.0 (*snmf* function) (https://bioconductor.org/packages/LEA/). The coefficient for cluster membership of each individual was averaged across the 10 independent runs using CLUMPP (Jakobsson & Rosenberg, 2007) and plotted using DISTRUCT (Rosenberg, 2004). We performed PCoA analyses using the *dartR* v1.1.11 package (*gl.pcoa* function) in R. We tested the overall population structure using the G-statistic test (Goudet *et al.*, 1996) by *gstat.randtest* function in *adegenet* v2.1.1 package (Jombart, 2008). Additionally, an analysis of molecular variance (AMOVA) was performed to quantify genetic differentiation among and within populations and genetic groups (as identified by STRUCTURE and PCoA) using Arlequin v3.5 (Excoffier *et al.*, 2007) with significance tests of variance components based on 10,000 permutations.

Population divergence history

We adopted a two-step strategy for investigating the divergent time of *C. agrestis.* First, we used plastome sequences assembled from Ranunculales and newly sampled *C. agrestis* populations to estimate the diversification of *C. agrestis*. A total of 47 plastomes were included (Table S1). The 79 protein coding regions (CDS) were extracted and aligned using MAFFT v7.313 (Katoh & Standley,

2013) and were then concatenated using PhyloSuite v1.1.15 (Zhang *et al.*, 2019). The program BEAST2 v2.5.2 (Bouckaert *et al.*, 2014) was employed for molecular dating with concatenated matrix using a GTR+G+I substitution model selected by jModelTest v2.0.1 (Darriba *et al.*, 2012) under the Bayesian information criterion (BIC). A log-normal relaxed-clock mode and birth-death process were applied. According to fossil records, the time to most recent common ancestors (TMRCA) of Ranunculales and Berberidaceae were constrained to the minimum age of 112 Ma and 72 Ma, respectively. We set both calibrations as an exponential distribution (Ho & Phillips, 2009) with a mean of 1 and an offset (hard bound constraint) that equaled the minimum age of the calibrations. The Markov Chain Monte Carlo (MCMC) run was set as 100 million generations with sampling every 1,000 generations. Tracer v1.7.1 (Rambaut *et al.*, 2018) was used to assess the effective sample size (ESS > 200) of each parameter. A maximum clade credibility tree was built by TreeAnnotator v1.8.0 (Rambaut & Drummond, 2010) with the initial 20% of trees discarded as burn-in.

Second, we estimated divergent time between haplotypes of *C. agrestis* using the SNP sequences generated from RAD-seq. All the parameters of BEAST2 were kept the same as above, with the exception of the TMRCA of *C. agrestis* populations was constrained based on the previous results. To infer the biogeographic history, ancestral distributions of C. agrestis were estimated by statistical dispersal-extinction-cladogenesis (S-DEC) (Ree & Smith, 2008), as implemented in Reconstruct Ancestral State in Phylogenies (RASP) v4.0 (Yu *et al.*, 2015), using 50,000 randomly sampled trees from posterior distribution generated by BEAST2. We defined the following five areas according to geographic distributions of *C. agrestis*: A, Qinling Mountain and adjacent regions; B, Northwestern Sichuan-South of Gansu; C, North of Qinghai; D, Hengduan Mountains; E, Himalayas. To detect potential demographic expansions, temporal changes in the effective population size (Ne) were inferred with Extended Bayesian skyline plots (EBSP) as implemented in BEAST2 (Heled & Drummond, 2008). We used the SNP sequences of 139 individuals and the default substitution model. The MCMC was set to 50 million generations with sampling every 5000 generations, and the first 20% was discarded as burn-in. Final graphs were produced with the custom R script provided by Heled (2010), with a burn-in cutoff of 20%.

Effects of heterogeneous landscapes on genetic structure

To illustrate the effects of the heterogeneous landscapes on shaping genetic structure, we conducted Mantel tests for the presence of isolation-by-distance (IBD) and isolation-by-resistance (IBR). To avoid the potential correlation between geographical distance and resistance distance, a partial Mantel test was also conducted to test IBD and IBR by controlling the resistance and geographical distance, respectively. GenAlEx v6.5 (Peakall & Smouse, 2012) was employed to compute the pairwise geographic distance among 18 populations. Resistance distance was generated in CIRCUITSCAPE v4.0.5 based on circuit theory (McRae & Beier, 2007). Circuit resistance provide models of connectivity or resistance to predict patterns of dispersal among sites in a heterogeneous landscape (Dickson *et al.*, 2019). The current ecological niche model (ENM) of environmental niche was calculated in MAXENT (Phillips & Dudik, 2008). We used 19 bioclimatic variables available from the WorldClim2 database (Fick & Hijmans, 2017) at a 30 arc□seconds resolution (Table S1), and actual evapotranspiration (AET; http://www.physicalgeography.net/fundamentals/8j.html) 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 0.9 paired. A total 12 climatic layers was retained for analyses (Table S1).

We used environmental niches representing connectivity between populations as conductance grids to produce pairwise resistance distance, as high habitat suitability was assumed to have low resistance (Nowakowski *et al.*, 2015).Tests of the significance for the relationship between geographical/resistance distances and genetic distance among populations were implemented in the R package *ade4* v1.7 (https://CRAN.R-project.org/package=ade4) using *mantel.rtest* function with 9,999 permutations.

Environmental variables associated with genetic structure

To estimate the contributions of environmental variables to genetic differentiation and to understand the turnover of allele frequencies along an environmental gradient, a gradient forest (GF) analysis

was performed using the R package *gradientForest* v0.1(http://gradientforest.r-forge.r-project.org/). GF is a nonparametric, machine learning regression tree approach that allows for exploration of nonlinear associations of spatial, environmental, and allelic variables (Gugger *et al.*, 2017; Bay *et al.*, 2018). The analysis partitions the allele frequency data at split values along the environmental gradients and defines the amount of variation explained as 'split importance' values (Jiang *et al.*, 2019). The overall predictor importance plot of GF shows the mean importance weighted by allele R^2, and the cumulative plot for each predictor shows cumulative change along the environmental gradient. The split importance values are added cumulatively to produce a step-like curve, thus, areas with large steps in a row indicate significant influence on allelic change. The 12 climatic layers used in above analysis was employed as environmental variables in the GF analysis.

A redundancy analysis (RDA) was performed to understand the associations between genetic structure and environmental variables. We estimated the proportion of genetic variance in the populations that is explained by environmental variables using six important variables identified by GF (bio03: isothermality; AET: actual evapotranspiration; bio04: temperature seasonality; bio07: temperature annual range; bio12: annual precipitation and bio15: seasonality precipitation, see results). In our analysis, we constrained the dependent variables (individuals) by the explanatory variables (climate). The RDA analysis was performed using the *rda* function in the *vegan* v2.5 package (Oksanen *et al.*, 2018) in R (http://CRAN.R-project.org/package=vegan). The *anova.cca* function was used to calculate overall significance of the RDA plot, and if this was found to be significant, then significance of each climate variable was calculated, using a significance tests with 9,999 permutations.

Genome Environment Association (GEA) analysis

To access the evidence of divergent selection acting on the genome and the genetic signatures of local adaptation, tests of associated outlier loci with environmental variation were carried out using BAYESCENV v1.1 (de Villemereuil & Gaggiotti, 2015), which extends the capabilities of the BAYESCAN algorithm by including a model that incorporates environmental data. BAYESCENV has been shown to be fairly robust to isolation by distance and a hierarchically structured scenario

(de Villemereuil & Gaggiotti, 2015). For a comprehensive consideration of the environmental effect, the environmental covariable layer was calculated with principal component analysis (PCA) of the six important bioclimatic variables, and the first PC axe (PC1), which explained most of the variability (48.62%), was extracted as the input for BAYESCENV. We ran 20 pilot runs of 5,000 iterations and a burn□in of 50,000 iterations, ultimately obtaining 5,000 MCMC iterations for the analysis. Diagnostics of the log likelihoods and F_{ST} values for the 5,000 sampled iterations were checked using the R package *coda* v0.19 (Plummer *et al.*, 2006) to confirm convergence and sample sizes of at least 2,500.

For functional annotation, the locus consensus sequences of all 6,120 SNPs that were initially generated by POPULATIONS were exported in *fasta* format. Candidate loci identified in GEA were annotated with Blast2Go v2 (Gotz *et al.*, 2008), which performs automated annotation augmentation by searching the non-redundant *Arabidopsis* protein database with records from the NCBI ESTs databases (Blastx). All the parameters were set as default with e-value < 1.0E-5.

Results

Sequence data quality and processing

We sequenced 139 individuals of *C. agrestis* using Illumina that produced a total of 133 gigabyte (Gb) data containing 845,396,236 raw reads (Table S2). After filtering out low quality reads, the average number of used reads per sample was 4,542,944 (median: 4,883,567; Table S2). The depths of coverage for processed samples were ranged from 5.77x (WLG-2) to 22.44x (HZ2-3), with a mean coverage of 11.99x (Table S2). We optimized the parameters for STACKS analysis using five representative samples and identified optimal parameters for our dataset to be $M = n = 3$ (Fig. S1 and S2). There were 3,640,060 loci containing 2,519,268 variant sites after all samples *de novo* assembling. After strict filtering (- $r = 1$), we obtained 6,959 RAD loci containing 6,120 SNPs that were used for population genetic analyses. The mean genotyped sites per locus was 146.59bp (stderr 0.03). After a looser filtering (- $r = 0.8$), we obtained 24,247 RAD loci containing 22,915 variant sites. The sizes of 20 sequenced plastomes from 18 populations of *C. agrestis* were ranged from 150,979 bp (ZF) to 151,056 bp (SNJ) possessing the same 79 protein coding genes arranged in the

same order (Table S3).

Genetic diversity and structure

We detected an extraordinary difference of genetic diversity at population level, especially for the nucleotide diversity π, which ranged from 0.0008 (HZ2) to 0.1646 (DDL). The expected heterozygosity (H_e) and observed heterozygosity (H_o) were 0.0007 (HZ2) - 0.1555 (DDL) and 0.0014 (HZ2) - 0.2650 (DDL; Table 1), respectively. When a loose filtering strategy was employed, we obtained a similar genetic diversity pattern (Table S4). In total, 66 haplotypes from 139 individuals (Fig. 1; Table S5) were detected, and the haplotype diversity (H_d) was 0.9798. Genetic differentiation between YL and HZ3 is the highest (F_{ST}=0.9913), and that between ZG and DDL is the lowest (F_{ST}=0.0640; Table S6; Fig. S3a).

We conducted a Bayesian clustering analysis to explore the genetic structure of *C. agrestis*. The delta-*K* method identified best-fit number for clustering (highest ΔK value) was two (Fig. S4a), representing the East and West clades. While the method based on the cross-entropy criterion indicated a better number of clusters was six (Fig. S4b). Considering high F_{ST} among populations, we further subdivided two clades into two and four groups (E1, E2, W1, W2, W3 and W4), respectively, which was also supported by the PoCA results (Fig. 2). The AMOVA revealed that 77.59% of the overall variation was distributed among six groups, with 11.27% explained by variation among populations within groups (Table S7). Within the genetic groups, W2 exhibited the highest nucleotide diversity (π; 0.1646), while W3 had the lowest (0.0206). All genetic groups had lower observed heterozygosity than expected, with the exception of W2. In general, the East clade has lower levels of observed heterozygosity (H_o, E1=0.0029, E2=0.0034, W1=0.0548, W2=0.2650, W3=0.0134, W4=0.0088) and higher levels of inbreeding coefficient (F_{IS}, E1=0.1952, E2=0.1011, W1=0.0384, W2=-0.1806, W3=0.0227, W4=0.0510) than the West clade (Table 1). Among genetic groups, genetic differentiation between W3 and W4 is the highest (F_{ST}=0.8908), followed by E2 and W4 (F_{ST}=0.8451), and that between E1 and E2 is lowest (F_{ST}=0.1284; Table S8; Fig. S3b). The G-statistic test suggested a significant population structure (p = 0.01**; Fig. S3c).

Population divergence history

The 47-taxa 79-CDS matrix of Ranunculales used for estimating the original divergence of *C. agrestis* was 66,117 bp in length containing 12,998 parsimony-informative sites. The divergence of *C. agrestis* from its closely related *Kingdonia* was estimated to occur in the early Eocene (stem age, ca. 52.19 Ma; 95% highest posterior density [HPD] intervals, 26.12-83.13 Ma), whereas the early divergence within *C. agrestis* started in the late Miocene (ca. 6.33 Ma; 95% HPD, 1.96-23.35 Ma) (Fig. S5). The tree topology inferred by BEAST2 using the SNP sequences generated from RAD-seq identified two main clades (East and West clades) and six sub-groups (E1, E2, W1, W2, W3 and W4 groups), which was consistent with STRUCTURE and PoCA results. The divergence of two main clades was estimated at appropriately 6.71 Ma (95% HPD, 5.18-7.89 Ma). The subsequent divergences between six groups occurred mainly during the Pliocene to late Pleistocene (ca. 5.05Ma-2.90Ma) (Fig. 3a). Biogeographic analysis supported Hengduan Mountains and adjacent regions as the most likely ancestral region for C. agrestise (Fig. 3a and 3b). For the demographic inference, EBSP analysis showed no sign of population expansion or contraction in demographic history of *C. agrestis* (Fig. 3c).

Effects of topography and ecology

Given that the distribution range of *C. agrestis* is topographically complex, we evaluated the influence of the heterogeneous landscape in shaping genetic structure via Mantel test. Our analyses suggested both significant patterns of IBD (R^2= 0.443, P = 0.001**) and IBR (R^2= 0.315, P = 0.017*) (Fig. 4a and b), which were also supported by a partial Mantel test (R_1^2= 0.426, P = 0.001**; R_2^2= 0.192, P = 0.011**). In our case, geographical distance explained the genetic structure better than resistance distances based on current distribution of habitats. We further plotted the conductance grid derived from the IBR model to illustrate the connectivity among populations and thus represent the empirical costs of gene flow through a complex landscape. In general, connectivity gaps among populations were observed in the western clade, reflecting increased habitat isolation in this area resulted from complex topography. In contrast, locations of eastern clade had low resistance to dispersal thus high connectivity among populations (Fig. 4c).

Environmental variables associated with genetic structure

Among the 12 variables used for GF analysis, isothermality (bio03) was indicated as the most important predictor. Actual evapotranspiration (AET), temperature annual range (bio07) and temperature seasonality (bio04) were also of high importance. Seasonality precipitation (bio15), and annual precipitation (bio12) showed moderate importance to allele frequencies. The other six environmental variables had relatively lower contributions (Fig. 5a). Fig. 5b showed the cumulative change in overall allele frequencies, where changes evolved with the environmental gradient.

The RDA analysis revealed that a significant amount of genetic variation among populations is associated with the six important environmental variables (46.58%, $p = 0.001$). Each of the two axes explained a significant amount of variation (Axis 1: 58.55%, $P = 0.001$**; Axis 2: 27.31%, $P = 0.001$**; Fig. 6). We also conducted RDA analysis using each six environmental variables separately and found that all six variables were significant. The contribution on genetic variation of each variable generally consistent with the GF analysis. Isothermality (bio03; 18.93%, $P = 0.001$**) was the most important predictor, whereas the temperature annual range (bio07) was less important in RDA analysis (Table 2).

Genome□Environment Association (GEA) analysis

The BAYESCENV result revealed a total of 61 loci have a significant association with environmental variables. We then aligned consensus sequences of these candidate loci for annotation. Sixteen of the 61 loci were successfully annotated by Blast2Go with cut-off of e-value <1E-5 and a mean similarity more than 80 (Table 3; Table S9). Among these were genes related to abiotic stress responses and stress-induced morphological adaptations, such as *UGT74E2* (Loci ID: 2763) regulating auxin homeostasis (Tognetti *et al*., 2010), *SRK2E* (Loci ID: 8038) involving in the resistance to drought by avoiding water loss (Mustilli *et al.*, 2002) and *AHK5* (Loci ID: 203056) regulating stomatal state and transmitting the stress signal (Pham *et al.*, 2012), as well as genes required for the formation and development of leaves (*CYP71*, loci ID: 293398; *CSI1*, loci ID: 218068) (Li & Luan, 2011). These loci also included genes related to oxidation-reduction

process involved in cellular respiration, as for example, *NDUFS7* (Loci ID: 83859) and *NDUFB3* (Loci ID: 219291); and genes implicated in the defense of plants against pathogens, such as Glucan endo-1,3-beta-glucosidase B (Loci ID: 296907), *RPPL1* (Loci ID:146932) and *NUDT8* (Loci ID: 243120) (Stroud *et al.*, 2016).

Discussion

Effects of topography on genetic structure

Our study presents the genetic structure of the palaeoendemic alpine *C. agrestis.* High haplotype diversity, uneven distribution of genetic diversity, as well as significant IBD and IBR patterns suggest restricted gene flow is likely a key factor in the high genetic differentiation among *C. agrestis* populations. The distribution area of *C. agrestis* is remarkable for its extraordinarily topographic diversity with some of the world's most rugged mountain ranges and altitudinal gradients spanning over 5,000 m (Favre *et al.*, 2015). As a direct consequence of the uplift of the QTP, habitat fragmentation induced by complex topography and frequented climatic fluctuation was invoked as an important mechanism for the high diversity of alpine plants in this area (Wen *et al.*, 2014; Sun, H *et al.*, 2017). Hence, dramatically topographic and climatic changes triggered by mountain uplift probably have influenced genetic connectivity and shaped current genetic structure of *C. agrestis.*

The high IBR pattern of western clade, reflecting the habitat isolation and perhaps caused by complex topography, may be the cause for limited gene flow as well as observed high differentiation. The Himalayan orogeny likely triggered the vicariance among different *C. agrestis* groups. The north-south mountain chains and valleys may provide corridors for flora exchange between the north and south, but represent barriers to migration between the east and west (Chen *et al.*, 2017); this likely resulted in the observed meridional differentiation of *C. agrestis* in the western clade. Previous studies (e.g., Gao *et al.*, 2007; Li *et al.*, 2011) have also suggested phylogeographic isolation associated with mountains and valleys with north-south direction. Notably, the populations located in the Hengduan Mountains (HM), exhibit high levels of genetic diversity (Table 1). As the QTP blocks the cold and dry air toward HM (Liu *et al.*, 2013), the HM

region is characterized by a much warmer and wetter climate, providing a comparatively stable environment for species. Although the western individuals are likely to possess higher genetic diversity, they are also likely to be more vulnerable to shifting disturbance regimes and environmental changes due to lack of genetic connectivity resulted from habitats isolation.

Compared to the western clade, north-south mountain barriers among different latitudes may be the reason for the divergence among the two groups in the eastern clade. A conductance grid derived from the IBR model showed low resistance to dispersal, thus high connectivity, among populations in the eastern clade. In addition, E1 and E2 exhibit a relatively low level of genetic differentiation. This phenomenon is likely due to the similar local environments among different populations resulting from the quite moderate topographical and climatic characteristics in this region (Chen *et al.*, 2017). We detected a significantly high F_{IS} values combined with low level of observed heterozygosity in the eastern clade (Table 1), indicating high levels of inbreeding, which may attribute to disturbance caused by human activities or the possible bottleneck events (Rands et al., 2010). There is consensus that breeding between related individuals can lead to a decrease in heterozygosity and fixation of deleterious alleles that may reduce reproductive fitness and increase the population susceptibility to environmental pressure (Kramer & Havens, 2009; Poudel *et al.*, 2014), making populations vulnerable to cope with rapid climate changes. Notably, the highest F_{IS} value was detected in WLG (F_{IS}: 0.0621), indicating a key protected population for *C. agrestis* populations in this region.

Impact of environmental heterogeneity in population divergence

Our molecular dating based on plastome sequences suggested an early Eocene origin of *C. agrestis*, which is accordant with the previous estimation (Ruiz-Sanchez *et al.*, 2012) and our results from nuclear data (Sun *et al.*, 2020). The diversification of *C. agrestis* has increased since the Pliocene (Fig. 3), which is likely related with this period of intense uplift of the HM during the late Miocene to late Pliocene (Sun *et al.*, 2011; Favre *et al.*, 2015). The environmental and climatic fluctuation triggered by mountain uplifts can always led to divergent selection and species adaptation associated with dramatic ecological niches changes (Ren *et al.*, 2017). Orogenic activities not only

provided numerous new habitats favoring for diversification, but also promoted population divergence due to the hindered gene flow across a continuous altitudinal difference. Our biogeographic analysis revealed that the demographic evolution of *C. agrestis* has not simply been influenced by the orogeny at a particular time in history, but is more complex, involving repeated episodes of dispersal and subsequent vicariance events over multiple timescales. Given the EBSP analysis showing a relatively stable population size in demographic history of *C. agrestis*, geographical events induced environmental and climatic fluctuation that promotes population divergence, while likely had a small effect on demographic changes. Coupled with the evidence of significant IBD and IBR patterns, we speculate natural selection driven by extraordinary environmental heterogeneity and geomorphological dynamics may have profound impact in the diversification of *C. agrestis*. In addition, during our field work we found population sizes of *C. agrestis* in the Himalaya-Hengduan Mountains to be much smaller than other sites. Previous studies have indicated a significant negative effect of small population size on adaptive evolution (Nevado *et al.*, 2019). Hence, we speculate that natural selection may play double-edged roles in the evolutionary history of any given species. On one hand, heterogeneous environments could provide more new ecological niches conducive for speciation, on the other hand, it may put the populations with weak adaptive ability to new environment in danger of extinction.

Reproductive isolation induced by geographical distance is likely to be another important factor driving the population divergence of *C. agrestis*. The significant population structure from AMOVA analysis, along with high level of F_{ST} values, and small population numbers discovered in the field commonly reflect there is a potential for reproductive isolation among different *C. agrestis* populations, especially in the western clade. For populations with small size, deleterious mutations and compensatory adaptive mutations can be quickly and independently fixed, causing reproductive isolation coupled with high level of genetic differentiation (Nielsen, 2005). However, *C. agrestis* exhibits low morphological diversity and no sign of expanding population demographic history. We speculate that strong environmental selective pressures are constraining morphological diversity while mutations are leading to observable population differentiation. Nonetheless, cryptic speciation cannot be ruled out, which needs further crossing experiments or/and testing for

ecological niche divergence/overlap. Furthermore, significant IBR pattern suggest that genetic structure would be better explained by habitat isolation between populations than just by geographical distance alone. Therefore, habitat changes may have large effect on genetic structure for *C. agrestis*, indicating its sensitivity to environmental niche.

Studying local adaptation contributes to understanding the ability of populations to sustain or adapt to rapid climate change (Jia *et al.*, 2019). We associated environmental variables with genetic structure to demonstrate the impact of environmental heterogeneity on population divergence. GF and RDA analyses indicated that covariation between genetic structure and environmental variables have played a role in promoting the diversification of *C. agrestis*. Specifically, both analyses found a significant association between genetic variation and temperature variables (Fig. 4; bio03, AET, bio04, bio07), suggesting temperature is potentially an important driver of genetic variation within *C. agrestis*, as detected in previous studies of angiosperm trees (Ahrens *et al.*, 2019; Jia *et al.*, 2019; Jiang *et al.*, 2019). In fact, there is growing evidence that high-mountain environments experience more rapid changes in temperature than environments at lower elevations (Pepin et al., 2015; Palazzi, Mortarini, Terzago, & von Hardenberg, 2019). Despite that, little attention has been given to understand the mechanisms of adaptive response to climate changes in high altitude areas. *C. agrestis* is habitually confined to grow under trees or shrubs, indicating it is sensitive to temperature. Moreover, the vertical distribution of *C. agrestis* is extensively wide ranging from 2,100 to 5,000 meters above sea level (Fu & Bartholomew, 2001), exhibiting a considerable differentiation in temperature. Therefore, it is reasonable that temperature variables play a vital role in promoting genetic divergence within *C. agrestis.* Furthermore, given the ongoing climate warming, temperature is likely still to be a key driver for adaptation of *C. agrestis* in future.

Genomic signatures associated with local adaptation

Simulation studies showed that the GEA method, which incorporates environmental information in the form of “environmental differentiation”, has a much lower false positive rate than genome-scan methods, and is robust to IBD pattern as well as a hierarchically structured scenario

(de Villemereuil & Gaggiotti, 2015). In the present study, we first employed PCA to calculate the overall environmental differentiation and provide the first axis (PC1) to BAYESCENV as a description of the environmental distance between populations. Considering that all environmental pressures drive species divergence together, this approach could provide a realistically and comprehensively environmental gradient for studying local adaptation.

In our analysis, BAYESCENV identified a set of genes that may be involved in a continuously adaptive process. Some genes as pivotal kinases involved in signaling pathway regulation, may induce abiotic stress responses for defending harsh climate in QTP. For example, genes associated with abscisic acid (ABA) signaling pathway (*SRK2EA*) can regulate numerous ABA responses, such as stomata closure, to response to high altitudes with extreme difference in temperature (Sierla *et al.*, 2018). Considering the tiny leaves of *C. agrestis* and their sensitivity to temperature, we speculate that stomatal regulation would be a vital process for maintaining hydration and regulating temperature. As altitude increases, hypobaric hypoxia becomes the main factor interfering with life activity. Genes related to oxidation-reduction processes involved in cellular respiration, such as *NDUFS7* and *NDUFB3*, are of importance in facilitating adaptation of *C. agrestis* to high altitude areas of hypoxia. In addition, some putatively adaptive genes (*CYP71* and *CSI1*) are associated with physiological trait regulation of vegetative and reproductive organs (Gu *et al.*, 2010). Thus, morphological adaptations may be an essential alternative for *C. agrestis* to respond to environmental stress.

Despite limited available gene information in this study, it should be noted that genetic patterns of adaptation to environmental heterogeneity are likely driven by a wide range of alleles of small-effect for complex traits (Savolainen *et al.*, 2013). The physiological processes involved in the adaptation of *C. agrestis* to alpine environments are most likely complex traits determined by multiple genes with potentially adaptive mechanisms including regulation of expression and epigenetic alteration (Gomulkiewicz *et al.*, 2010; Ahrens *et al.*, 2019). The genes identified as putatively adaptive signals in *C. agrestis* are most likely a part of the broader process of adaptation to climate and should constitute the first step in studying genetic mechanism of local adaptation in alpine environments. Further work with characterizing whole genome could further illuminate

patterns of functional genes controlling the response to climate changes. Generally, these patterns can serve as a proxy for understanding and monitoring the adaptive process in rapidly changing environment (Ahrens *et al.*, 2019) and inform future management strategies.

Conclusions

In this study, we demonstrate two major advances. First, we provide new insights into the genetic structure and evolutionary history of a relictual alpine herb surviving in heterogeneous environment, which can inform further conservation. Our results shed light on vital roles of mountain uplift in promoting population diversification and the dual effect of environmental heterogeneity on species evolutionary history. Second, our analyses provide a roadmap to study evolution of local adaptation of non-model species for which common gardens and genomic resources are limited. By conducting multiple analyses, we associated environmental variables with genetic variation and explicitly measured the relative influence of bioclimate on population divergence. In turn, the important environmental predictors were implemented in identifying potential loci that may under divergent selection to gain in-depth knowledge of genetic basis involved in local adaptation. More importantly, our results improve our understanding of drivers and mechanisms of adaptive evolution, aiding efforts on developing sound conservation programs in facing of rapid changing climate.

Acknowledgments

We are grateful to Jie Cai for help with sample collection in Tibet. This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the Programme Foundation for the Backbone of Scientific Research by Wuhan Botanical Garden, Chinese Academy of Sciences (Y855241G01), the Major Program of National Natural Science Foundation of China (31590823), and the National Key R and D Program of China (2017YFC0505200).

Author contributions

HCW, HS, and YXS developed the idea and designed the experiment. JWZ, LSY, NL, HJZ, RG, LJL, YHZ and TD collected the leaf materials. XZ and YXS performed the statistical analyses; XZ, YXS, JBL, HS and HCW interpreted the results and wrote the manuscript. All authors read, edited and approved the final manuscript. XZ and YXS contributed equally to this work.

Data accessibility

All newly sequenced plastomes were deposited in National Center for Biotechnology Information (NCBI) with accession numbers xxx-xxx. RAD raw read data are stored at the NCBI Sequence Read Archive in the Bioproject xxx. Genomic data sets and R scripts used in this study are available in the DRYAD archives under accession xxx.

REFERENCE:

Ahrens CW, Byrne M, Rymer PD. 2019. Standing genomic variation within coding and regulatory regions contributes to the adaptive capacity to climate in a foundation tree species. *Mol Ecol* **28**(10): 2502-2516.

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol Appl* **1**(1): 95-111.

Anderson JT, Willis JH, Mitchell-Olds T. 2011. Evolutionary genetics of plant adaptation. *Trends Genet* **27**(7): 258-266.

Bay RA, Harrigan RJ, Underwood VL, Gibbs HL, Smith TB, Ruegg K. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**(6371): 83.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15): 2114-2120.

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 Comput Biol* **10**(4): e1003537.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular ecology* **22**(11): 3124-3140.

Chen Y-S, Deng T, Zhou Z, Sun H. 2017. Is the East Asian flora ancient or not? *National Science Review* **5**(6): 920-932.

Colautti RI, Barrett SCH. 2013. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science* **342**(6156): 364.

Conover DO, Duffy TA, Hice LA. 2009. The Covariance between Genetic and

Environmental Influences across Ecological Gradients. *Annals of the New York Academy of Sciences* **1168**(1): 100-129.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics* **185**(4): 1411.

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

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**(8): 772.

Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, UK: John Murray.

de Villemereuil P, Gaggiotti OE. 2015. A new FST-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution* **6**(11): 1248-1258.

Dickson BG, Albano CM, Anantharaman R, Beier P, Fargione J, Graves TA, Gray ME, Hall KR, Lawler JJ, Leonard PB, et al. 2019. Circuit-theory applications to connectivity science and conservation. *Conservation Biology* **33**(2): 239-249.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**(2): 359-361.

Ellegren H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution* **29**(1): 51-63.

Excoffier L, Laval G, Schneider S. 2007. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online* **1**: 47-50.

Favre A, Packert M, Pauls SU, Jahnig SC, Uhl D, Michalak I, Muellner-Riehl AN. 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev Camb Philos Soc* **90**(1): 236-253.

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**(12): 4302-4315.

Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology* **27**(9): 2215-2233.

Foster AS. 1971. Additional studies on the morphology of blind vein-endings in the leaf of circaeaster agrestis. *American Journal of Botany* **58**(3): 263-272.

Friis G, Fandos G, Zellmer AJ, McCormack JE, Faircloth BC, Milá B. 2018. Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology* **27**(24): 5137-5153.

Fu D, Bartholomew B 2001. CIRCAEASTERACEAE. In: Wu ZY, Raven, P. H. & Hong, D. Y. ed. *Flora of China*. Beijing & St. Louis: Science Press & Missouri

Botanical Garden Press, 439.

Gao LM, MÖLler M, Zhang XM, Hollingsworth ML, Liu J, Mill RR, Gibby M, Li DZ. 2007. High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Molecular Ecology* **16**(22): 4684-4698.

Goicoechea PG, Guillardín L, Fernández-Ibarrodo L, Valbuena-Carabaña M, González-Martínez SC, Alía R, Kremer A. 2019. Adaptive Introgression Promotes Fast Adaptation In Oaks Marginal Populations. *bioRxiv*: 731919.

Gomulkiewicz R, Holt RD, Barfield M, Nuismer SL. 2010. Genetics, adaptation, and invasion in harsh environments. *Evolutionary Applications* **3**(2): 97-108.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* **36**(10): 3420-3435.

Goudet J, Raymond M, de Meeüs T, Rousset F. 1996. Testing Differentiation in Diploid Populations. *Genetics* **144**(4): 1933.

Gu Y, Kaplinsky N, Bringmann M, Cobb A, Carroll A, Sampathkumar A, Baskin TI, Persson S, Somerville CR. 2010. Identification of a cellulose synthase-associated protein required for cellulose biosynthesis. *Proceedings of the National Academy of Sciences* **107**(29): 12866.

Gugger PF, Liang CT, Sork VL, Hodgskiss P, Wright JW. 2017. Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications* **11**(2): 231-242.

Heled J. 2010. Extended Bayesian Skyline Plots Tutorial.

Heled J, Drummond AJ. 2008. Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology* **8**(1): 289.

Ho SYW, Phillips MJ. 2009. Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology* **58**(3): 367-380.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. 2016. Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist* **188**(4): 379-397.

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**(14): 1801-1806.

Jia K-H, Zhao W, Maier PA, Hu X-G, Jin Y, Zhou S-S, Jiao S-Q, El-Kassaby YA, Wang T, Wang X-R, et al. 2019. Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evolutionary Applications* https://doi.org/10.1111/eva.12891.

Jiang X-L, Gardner EM, Meng H-H, Deng M, Xu G-B. 2019. Land bridges in the

Pleistocene contributed to flora assembly on the continental islands of South China: Insights from the evolutionary history of *Quercus championii*. *Molecular Phylogenetics and Evolution* **132**: 36-45.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**(11): 1403-1405.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**(4): 772-780.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* **7**(12): 1225-1241.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12): 1647-1649.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. *Genetics* **49**(4): 725-738.

Kramer AT, Havens K. 2009. Plant conservation genetics in a changing world. *Trends in Plant Science* **14**(11): 599-607.

Leinonen PH, Remington DL, LeppÄLÄ J, Savolainen O. 2013. Genetic basis of local adaptation and flowering time variation in *Arabidopsis lyrata*. *Molecular Ecology* **22**(3): 709-723.

Li H, Luan S. 2011. The cyclophilin AtCYP71 interacts with CAF-1 and LHP1 and functions in multiple chromatin remodeling processes. *Mol Plant* **4**(4): 748-758.

Li Y, Zhai S-N, Qiu Y-X, Guo Y-P, Ge X-J, Comes HP. 2011. Glacial survival east and west of the 'Mekong–Salween Divide'in the Himalaya–Hengduan Mountains region as revealed by AFLPs and cpDNA sequence variation in *Sinopodophyllum hexandrum* (Berberidaceae). *Molecular Phylogenetics and Evolution* **59**(2): 412-424.

Liu J, Möller M, Provan J, Gao L-M, Poudel RC, Li D-Z. 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist* **199**(4): 1093-1108.

Lowry DB, Lovell JT, Zhang L, Bonnette J, Fay PA, Mitchell RB, Lloyd-Reilley J, Boe AR, Wu Y, Rouquette FM, et al. 2019. QTL × environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proceedings of the National Academy of Sciences* **116**(26): 12933.

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**(2): 526-538.

McRae BH, Beier P. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* **104**(50): 19885.

Mitchell-Olds T, Willis JH, Goldstein DB. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews*

Genetics **8**: 845.

Mustilli AC, Merlot S, Vavasseur A, Fenzi F, Giraudat J. 2002. Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *The Plant cell* **14**(12): 3089-3099.

Nevado B, Wong ELY, Osborne OG, Filatov DA. 2019. Adaptive Evolution Is Common in Rapid Evolutionary Radiations. *Current Biology*.

Nielsen R. 2005. Molecular Signatures of Natural Selection. *Annual Review of Genetics* **39**(1): 197-218.

Nowakowski AJ, DeWoody JA, Fagan ME, Willoughby JR, Donnelly MA. 2015. Mechanistic insights into landscape genetic structure of two tropical amphibians using field-derived resistance surfaces. *Mol Ecol* **24**(3): 580-595.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2018. Vegan: Community Ecology Package. https://CRAN.R-project.org/package=vegan.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**(19): 2537-2539.

Pham J, Liu J, Bennett MH, Mansfield JW, Desikan R. 2012. Arabidopsis histidine kinase 5 regulates salt sensitivity and resistance against bacterial and fungal infection. *New Phytol* **194**(1): 168-180.

Phillips SJ, Dudik M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**(2): 161-175.

Plummer M, Best N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**(1): 7-11.

Poudel RC, Möller M, Liu J, Gao L-M, Baral SR, Li D-Z. 2014. Low genetic diversity and high inbreeding of the endangered yews in Central Himalaya: implications for conservation of their highly fragmented populations. *Diversity and Distributions* **20**(11): 1270-1284.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**(2): 945-959.

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**(1): 50.

R Team C 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rambaut A, Drummond A. 2010. TreeAnnotator version 1.6. 1. *University of Edinburgh, Edinburgh, UK.* Available at: http://beast.bio.ed.ac.uk.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* **67**(5): 901-904.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst Biol* **57**(1): 4-14.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**(17): 4348-4370.

Ren G, Mateo RG, Liu J, Suchan T, Alvarez N, Guisan A, Conti E, Salamin N. 2017. Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing. *New Phytologist* **213**(3): 1500-1512.

Ren Y, Li Z, Hu Z. 2003. Approaches to the systematic position of *Circaeaster* based on the morphological data. *Acta Botanica Boreali-occidentalia Sinica* **23**(7): 1091-1097.

Rochette NC, Catchen JM. 2017. Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols* **12**: 2640.

Rosenberg NA. 2004. Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**(1): 137-138.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* **34**(12): 3299-3302.

Ruiz-Sanchez E, Rodriguez-Gomez F, Sosa V. 2012. Refugia and geographic barriers of populations of the desert poppy, *Hunnemannia fumariifolia* (Papaveraceae). *Organisms Diversity & Evolution* **12**(2): 133-143.

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

Savolainen O, Pyhäjärvi T, Knürr T. 2007. Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics* **38**(1): 595-619.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, et al. 2014. Genomics and the origin of species. *Nature Reviews Genetics* **15**: 176.

Sierla M, Horak H, Overmyer K, Waszczak C, Yarmolinsky D, Maierhofer T, Vainonen JP, Salojarvi J, Denessiouk K, Laanemets K, et al. 2018. The Receptor-like Pseudokinase GHR1 Is Required for Stomatal Closure. *The Plant cell* **30**(11): 2813-2837.

Spieth PT. 1974. Gene flow and genetic differentiation. *Genetics* **78**(3): 961.

Stroud DA, Surgenor EE, Formosa LE, Reljic B, Frazier AE, Dibley MG, Osellame LD, Stait T, Beilharz TH, Thorburn DR, et al. 2016. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature* **538**(7623): 123-126.

Sun B-N, Wu J-Y, Liu Y-S, Ding S-T, Li X-C, Xie S-P, Yan D-F, Lin Z-C. 2011. Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography, Palaeoclimatology, Palaeoecology* **304**(3): 328-336.

Sun H, Zhang J, Deng T, Boufford DE. 2017. Origins and evolution of plant diversity in the Hengduan Mountains, China. *Plant Diversity* **39**(4): 161-166.

Sun Y, Deng T, Zhang A, Moore MJ, Landis JB, Lin N, Zhang H, Zhang X, Huang J, Zhang X, et al. 2020. The draft genome of the endangered, relictual plant *Kingdonia uniflora* (Circaeasteraceae, Ranunculales) reveals potential mechanisms and perils of evolutionary specialization. *bioRxiv*: 2020.2001.2008.898460.

Sun Y, Moore MJ, Lin N, Adelalu KF, Meng A, Jian S, Yang L, Li J, Wang H. 2017. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family. *BMC Genomics* **18**(1): 592.

Tognetti, V. B., Van Aken, O., Morreel, K., Vandenbroucke, K., van de Cotte, B., De Clercq, I., et al. 2010. Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates Arabidopsis architecture and water stress tolerance. *The Plant cell*, **22**(8), 2660-2679.

Weigel D, Nordborg M. 2015. Population Genomics for Understanding Adaptation in Wild Plant Species. *Annual Review of Genetics* **49**(1): 315-338.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**(6): 1358-1370.

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

Wright S. 1946. Isolation by distance under diverse systems of mating. *Genetics* **31**(1): 39-59.

Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour* **14**(5): 1024-1031.

Yu Y, Harris AJ, Blair C, He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol Phylogenet Evol* **87**: 46-49.

Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. 2019. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour* doi:10.1111/1755-0998.13096.

Zou Y-P, Hou X-H, Wu Q, Chen J-F, Li Z-W, Han T-S, Niu X-M, Yang L, Xu Y-C, Zhang J, et al. 2017. Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. *Genome Biology* **18**(1): 239.

Table 1 Summary of statistics calculated for the 6120 variant positions. n, number of genotype; π, average nucleotide diversity; H_e, average expected heterozygosity per locus; H_o, average observed heterozygosity per locus; the inbreeding coefficients (F_{IS}). The statistics of six defined groups are bolded. E1, eastern group1; E2, eastern group2; W1, western group1; W2, western group2; W3, western group3; W4, western group4

Group IDs	Population IDs	n	π	H_e	H_o	F_{IS}
E1	**-**	**75**	**0.0464**	**0.0461**	**0.0029**	**0.1952**
	SNJ	13	0.0207	0.0199	0.0038	0.0383
	TBS	17	0.0218	0.0212	0.0029	0.0532
	HZ2	9	0.0008	0.0007	0.0014	-0.0010
	HZ3	10	0.0013	0.0012	0.0022	-0.0017
	KMG	8	0.0015	0.0014	0.0027	-0.0023
	TSM	9	0.0214	0.0203	0.0027	0.0425
	WLG	4	0.0306	0.0287	0.0034	0.0621
	XLS	5	0.0223	0.0210	0.0026	0.0375
E2	**-**	**19**	**0.0384**	**0.0374**	**0.0034**	**0.1011**
	BLS	9	0.0216	0.0204	0.0035	0.0529
	SGN	10	0.0254	0.0241	0.0033	0.0573
W1	**-**	**7**	**0.0688**	**0.0639**	**0.0548**	**0.0384**
	MXG	4	0.0215	0.0189	0.0220	-0.0005
	YL	1	0.0016	0.0008	0.0016	-
	ZG	2	0.1633	0.1225	0.1470	0.0244
W2	**DDL**	**9**	**0.1646**	**0.1555**	**0.2650**	**-0.1806**
W3	**SJL**	**10**	**0.0206**	**0.0196**	**0.0134**	**0.0227**
W4	**-**	**19**	**0.0249**	**0.0242**	**0.0088**	**0.0510**
	YH	8	0.0150	0.0140	0.0057	0.0238
	GS	1	0.0255	0.0128	0.0255	-
	ZF	10	0.0136	0.0130	0.0095	0.0100

Table 2 RDA results based on six important environmental variables identified by GF analysis.

Environmental variables	Constrained proportion	*F*	*p*
Bio03	0.1893	31.98	0.001**
AET	0.1712	28.31	0.001**
Bio04	0.1408	22.45	0.001**
Bio12	0.0902	13.59	0.001**
Bio07	0.0627	9.17	0.001**
Bio15	0.0203	2.84	0.001**

Table 3 Annotation information of 16 candidate loci under selection that the E-value of blastx is less than 1E-5. Detailed functional annotation is provided in Table S9.

Loci ID	Gene names	Description	Length	E-Value	similarity
2763	*UGT74E2*	UDP-glycosyltransferase 74E2-like	146	1.16E-19	82.08
8038	*SRK2E*	Serine/threonine-protein kinase TIO-like isoform X1	146	4.97E-07	81.56
8150	*CSC1*	Calcium permeable stress-gated cation channel 1-like	146	5.12E-20	92.71
19591	*At1g18270*	Ketose-bisphosphate aldolase class-II family protein	146	6.34E-06	87.07
83859	*NDUFS7*	NADH dehydrogenase subunit 7 (mitochondrion)	146	7.55E-26	93.96
126986	*RTL*	Retrotransposon gag protein	147	4.78E-14	84.5
146932	*RPPL1*	Putative disease resistance RPP13-like protein 1 isoform X1	146	2.78E-06	87.52
203056	*AHK5*	Histidine kinase 5	146	7.55E-14	84.61
206146	*ABCC2*	ABC transporter C family member 2-like	146	2.77E-15	89.16
218068	*CSI1*	Protein cellulose synthase interactive 1	146	3.74E-14	86.88
219291	*NDUFB3*	NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 3-B-like	146	8.26E-20	82.13
243120	*NUDT8*	Nudix hydrolase 8	146	8.49E-14	81.72
275249	*N/A*	Glycosyl transferase (glycosyl transferase family 2)	146	3.38E-24	85.59
293398	*CYP71*	Peptidyl-prolyl cis-trans isomerase CYP71	146	7.81E-17	92.11
296907	*N/A*	Glucan endo-1,3-beta-glucosidase 13-like	146	1.45E-20	83.33
305334	*CSLC6*	Probable xyloglucan glycosyltransferase 6	146	7.31E-26	98.54

Figure legends

Fig. 1 Geographical distribution of haplotypes detected in polymorphic sites of Circaeaster agrestis populations. Colors indicate genetic groups based on STRUCTURE and PoCA results (see also Fig. 2), with the Eastern clade being made up of the orange and dark green circles, with the Western clade consisting of the light green, light pink, purple, and blue circles. Circle sizes reflect the number of individuals sampled.

Fig. 2 Population genetic structure in Circaeaster agrestis based on 6,120 SNPs from (a) PCoA

for all populations, east-clade and west-clade; and (b) STRUCTURE with K=2 (showing the highest ΔK; Fig. S4a) and 6 (representing the divergence within east and west clades).

Fig. 3 Evolutionary history of Circaeaster agrestis. (a) Divergent time estimation and ancestral area reconstructions of 66 haplotypes based on SNPs sequences. Mean divergence dates and 95% HPDs for major nodes (1–5) are summarized in top left. Pie charts on each node indicate marginal probabilities for each alternative ancestral area derived from S-DEC. Results are based on a maximum area number of two. Inferred dispersal and vicariance events are indicated by green and red circle respectively. (b) Five defined areas according to geographic distributions of C. agrestis: A, Qinling Mountains; B, Sichuan-South of Gansu; C, North of Qinghai; D, Hengduan Mountains; E, Himalayas. (c) Temporal changes in the effective population size (Ne) changes inferred with Extended Bayesian skyline plots (EBSP) using BEAST2. Re, reproductive number, δ, death rate.

Fig. 4 Topographic and ecological effects on genetic structure. Relationship of genetic distance and (a) geographical distance, (b) resistance distance based on climatic niche suitability, as tested by a Mantel test. (c) Conductance grid derived from the Isolation-by-Resistance (IBR) model. Grid with warm colour indicates high conductance, cold colour indicates high resistance.

Fig. 5 (a) R2-weighted importance of environmental variables that explain genetic gradients from gradient forest (GF) analysis. (b) Cumulative importance of allelic change along the first six environmental gradients.

Fig. 6 Redundancy analysis showing the relationship between the independent climate parameters and population structure. Individuals are colored points and colors represent six groups (E1, E2, W1, W2, W3, W4). Small black points are SNPs. The percentage of explained variance for each axis is provided.

Supporting information

Fig. S1 The number of 80% polymorphic loci shared across most samples (the r80 loci) as M=n increases.

Fig. S2 The distribution of the number of SNPs per locus for a range of M=n values.

Fig. S3 (a) F_{ST} values within populations estimated by *hierfstat*, Group names of populations were

labeled in the top of boxes. (b) F_{ST} values within genetic groups estimated by *hierfstat.* (c) The results of G-statistic test.

Fig. S4 The optimal *K* value identified using (a) delta-*K* method in STRUCTURE HARVESTER, and (b) cross-entropy criterion in R package *LEA.*

Fig. S5 Divergence times of Ranunculales estimated by BEAST2 with a relaxed molecular clock based on the combined protein-coding region sequences. A and B indicate fossil calibration points. C and D indicate the origination and diversification of Circaeaster agrestis, respectively. Median ages of nodes are shown with bars indicating the 95% highest posterior density intervals for each node.

Table S1 Summary of population information of Circaeaster agrestis, including the number of individuals genotyped at each localities (n), geographic information,voucher specimens information, and 20 bioclimate values (bio01-bio19 and AET) of each location. AET, the actual evapotranspiration.

Table S2 Numbers of reads of RAD-seq for each individual of Circaeaster agrestis.

Table S3 Assemble and annotation information of newly sequenced plastomes of Circaeaster agrestis populations.

Table S4 Summary of statistics calculated for the 22915 variant positions. n, number of genotype; π, average nucleotide diversity; He, average expected heterozygosity per locus; Ho, average observed heterozygosity per locus; the inbreeding coefficients (FIS). The statistics of six defined groups are bolded. E1, eastern group1; E2, eastern group2; W1, western group1; W2, western group2; W3, western group3; W4, western group4.

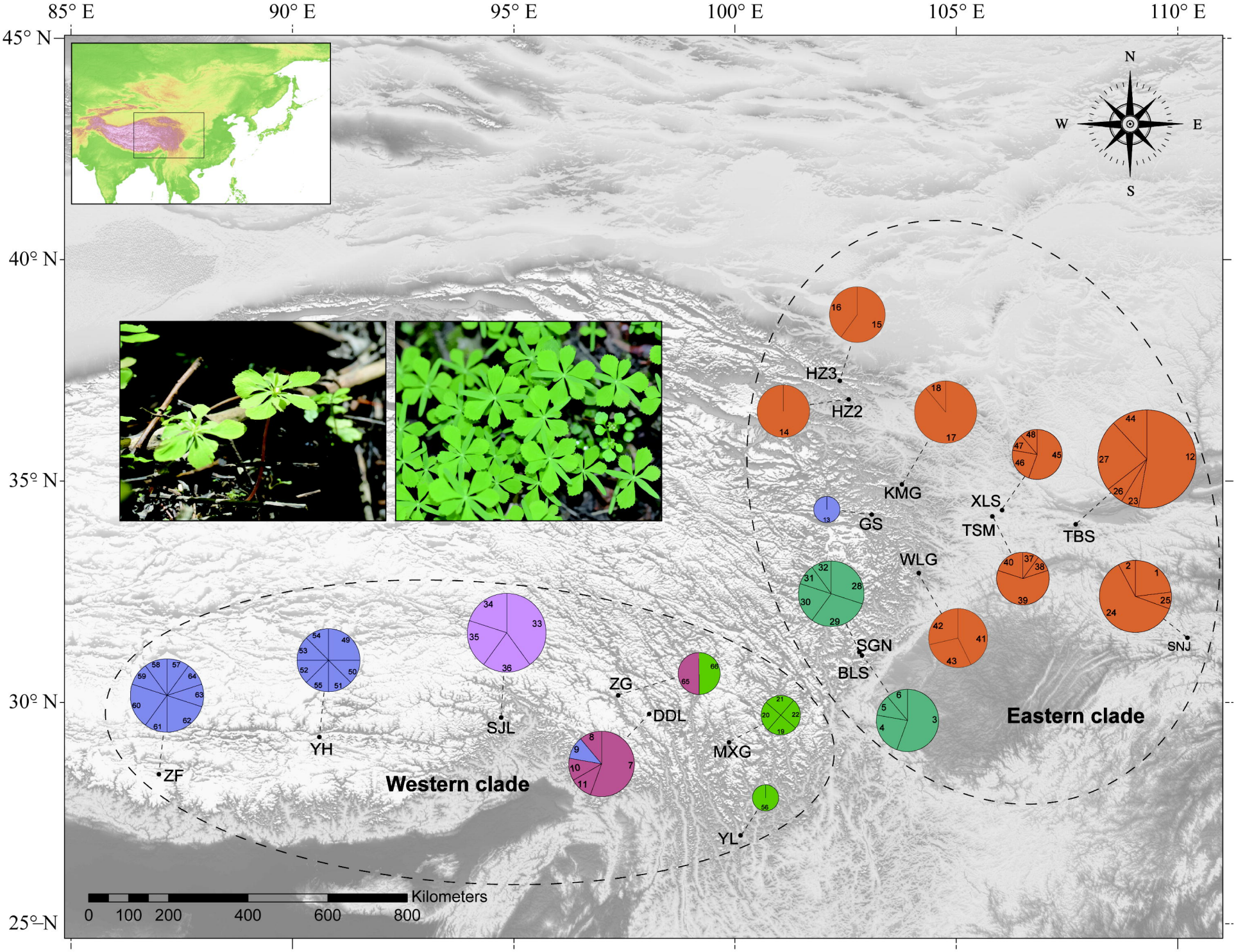
Table S5 Haplotypes information detected in 18 populations based on 6120 SNPs.

Table S6 Paired population-level FST estimated by hierfstat using Weir and Cockerham's method.

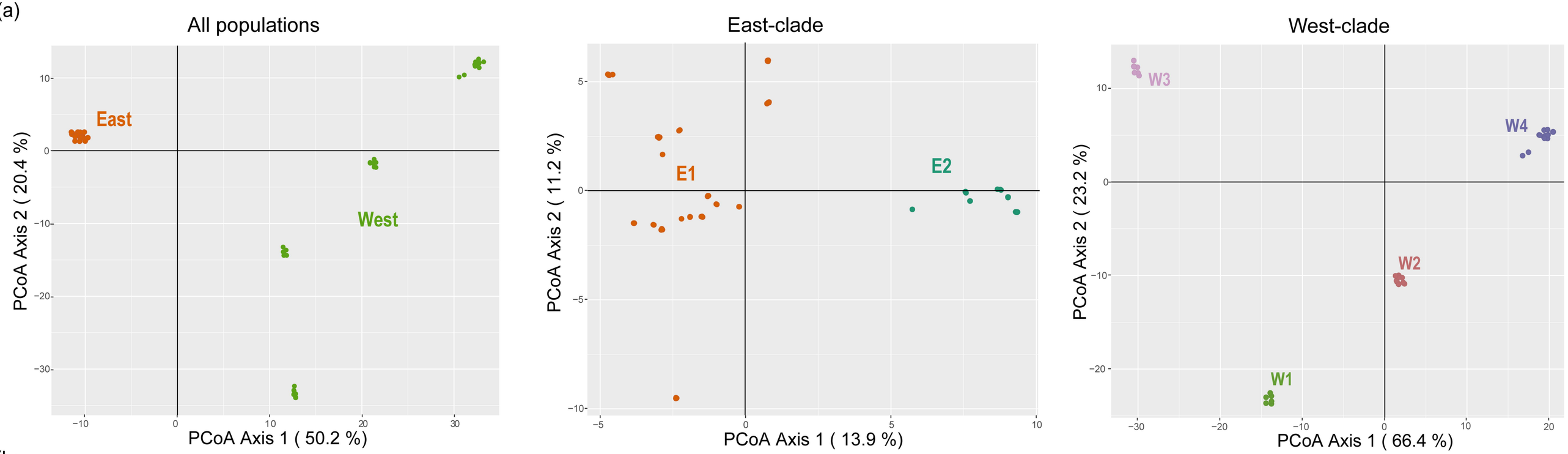
Table S7 The analysis of molecular variance (AMOVA) for SNP data among six genetic groups (E1, E2, W1, W2, W3 and W4)

Table S8 Paired group-level FST estimated by hierfstat using Weir and Cockerham's method.

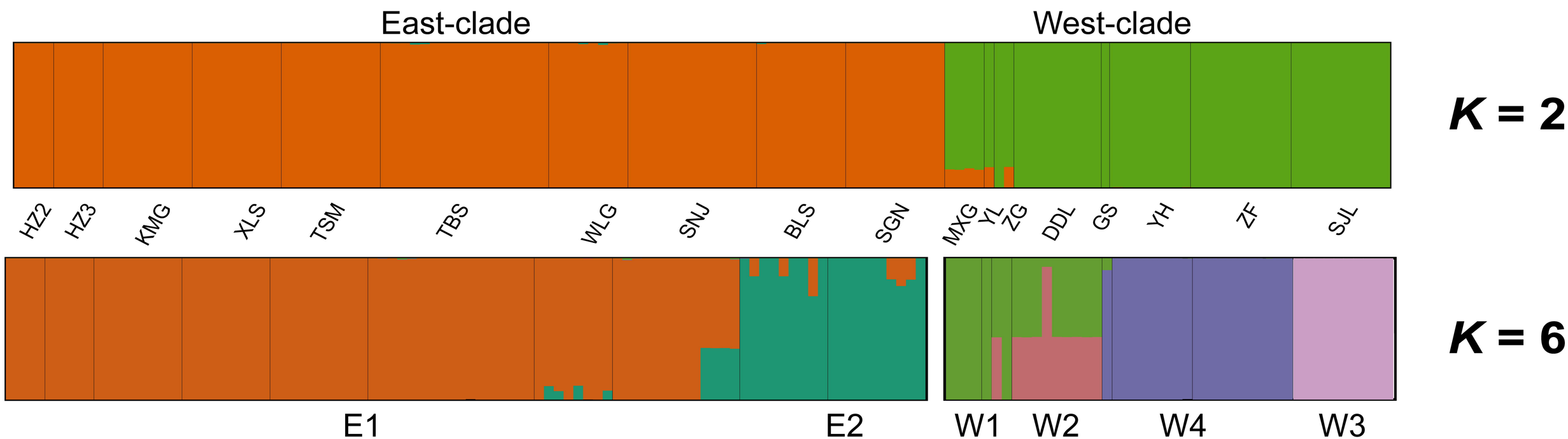
Table S9 Detailed annotation information of sixteen genes under potential divergent selection.

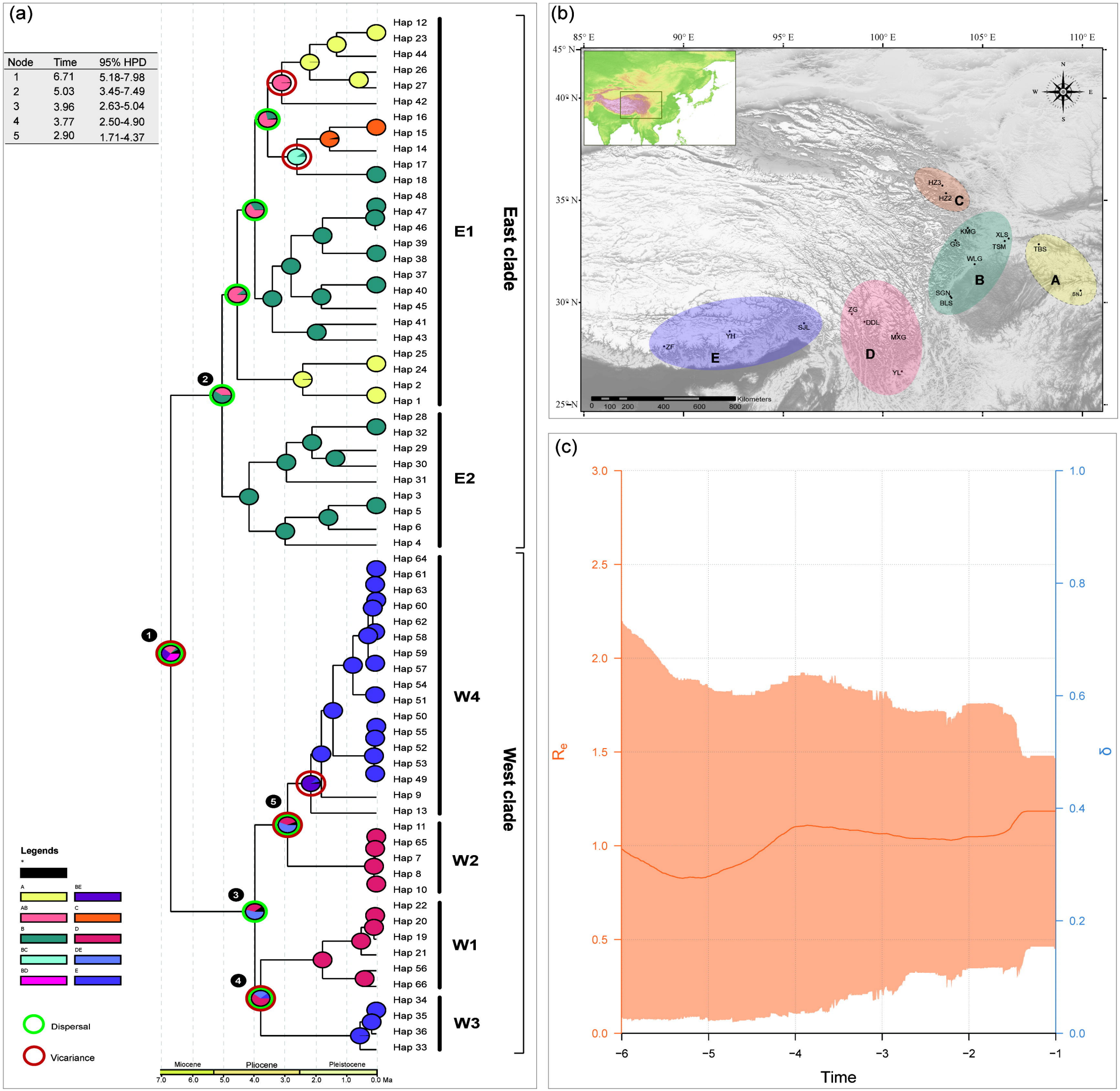

85° E
90° E
95° E
100° E
105° E
110° E
45° N
40° N
35° N
30° N
25° N
N
S
W
E
HZ3
HZ2
KMG
GS
XLS
TSM
TBS
WLG
SGN
BLS
SNJ
ZG
DDL
SJL
YH
ZF
MXG
YL
Eastern clade
Western clade
Kilometers
0 100 200 400 600 800

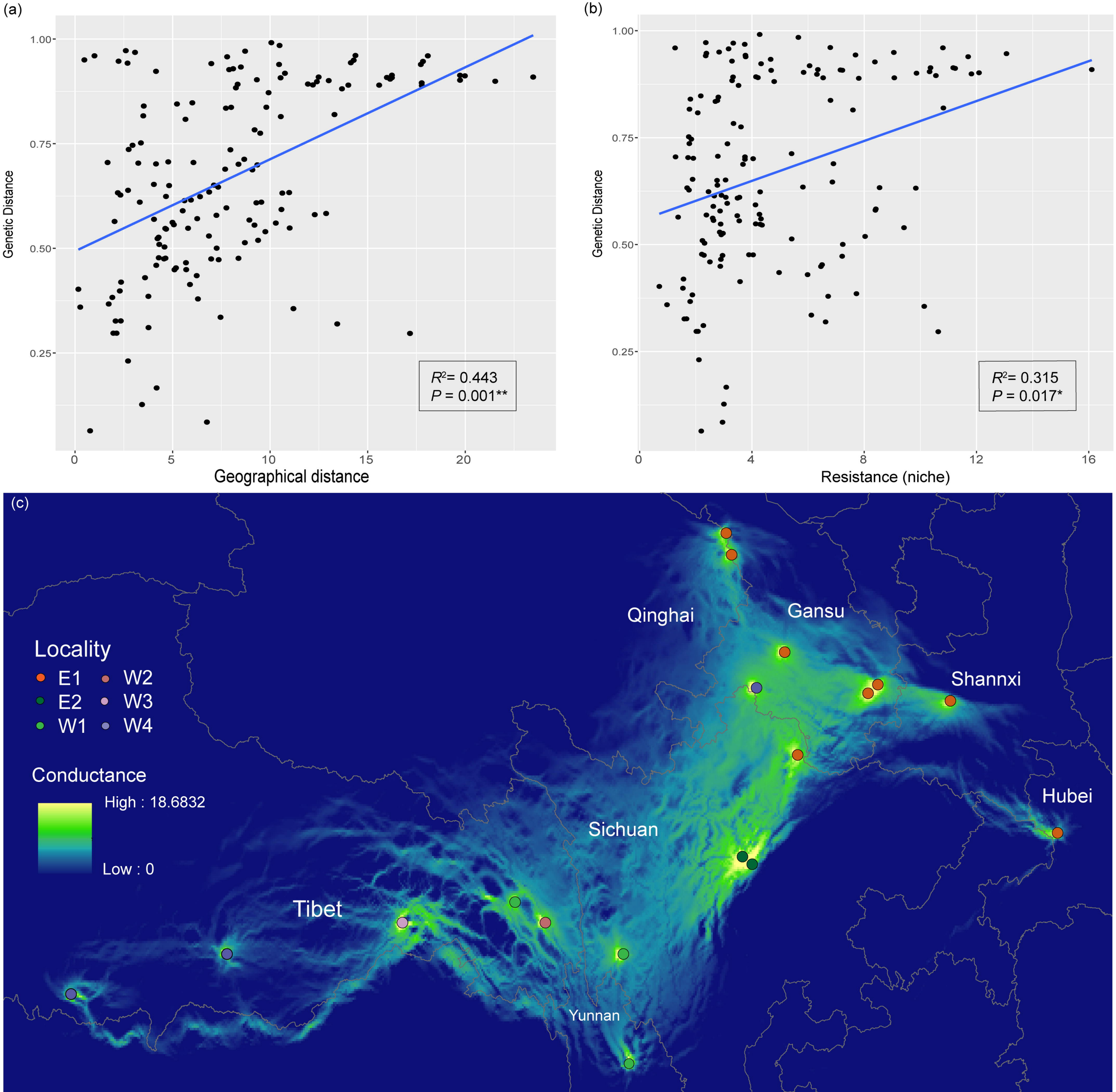
(a)


All populations
East
West
PCoA Axis 2 (20.4 %)
PCoA Axis 1 (50.2 %)
East-clade
E1
E2
PCoA Axis 2 (11.2 %)
PCoA Axis 1 (13.9 %)
West-clade
W3
W4
W2
W1
PCoA Axis 2 (23.2 %)
PCoA Axis 1 (66.4 %)

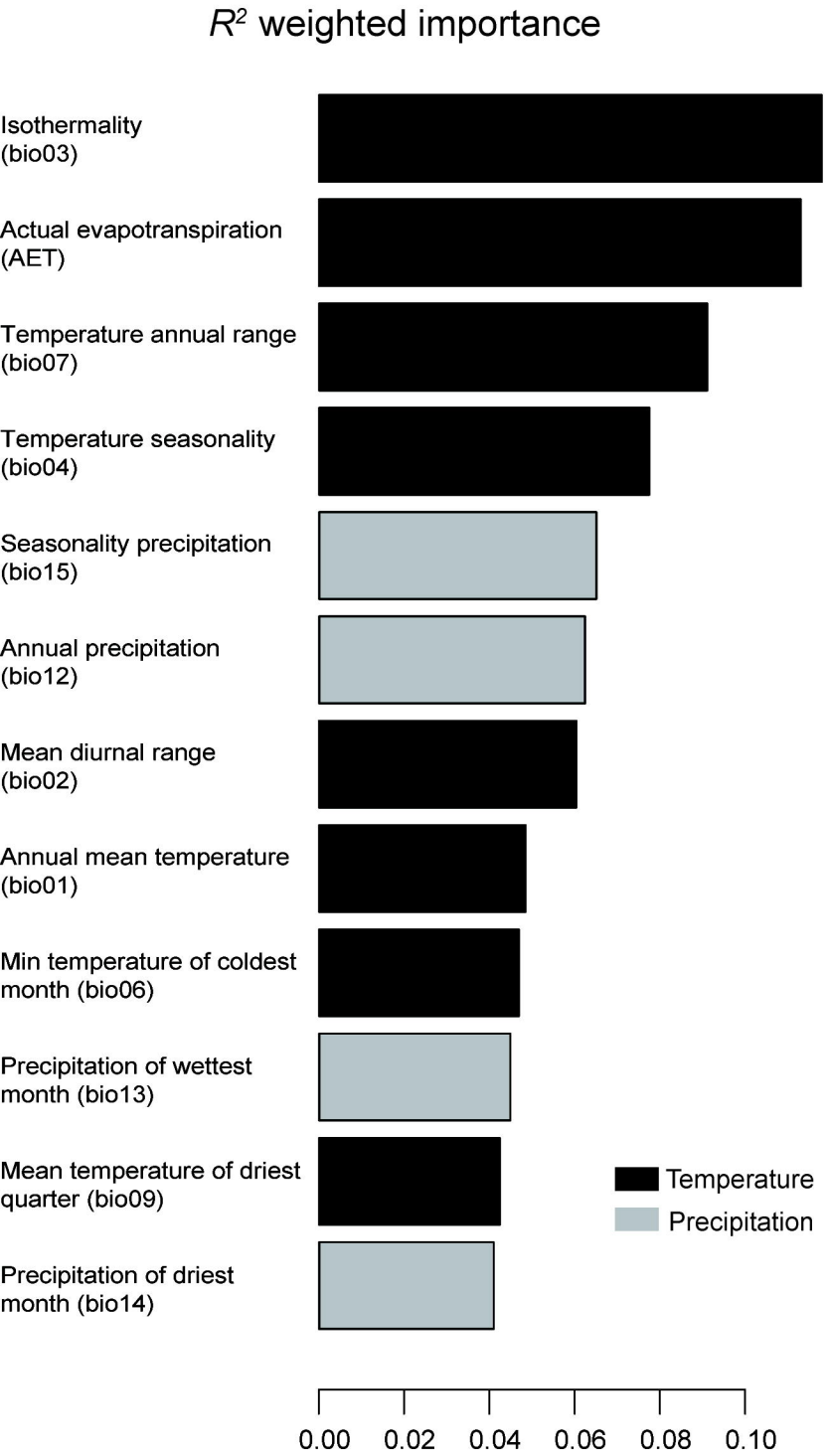
(b)


East-clade
West-clade
K = 2
HZ2
HZ3
KMG
XLS
TSM
TBS
WLG
SNJ
BLS
SGN
MXG
YL
ZG
DDL
GS
YH
ZF
SJL
K = 6
E1
E2
W1
W2
W4
W3

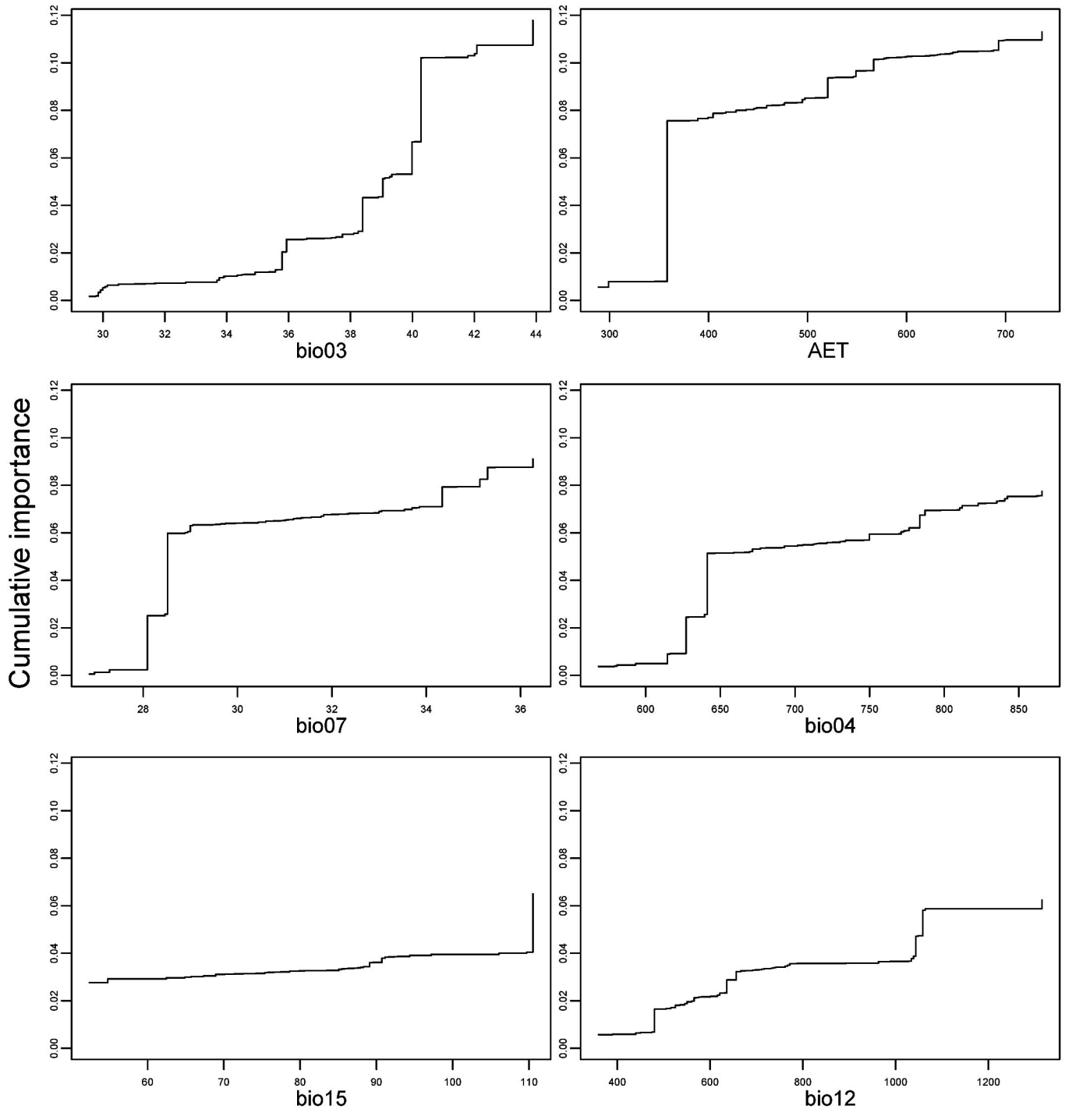

(a)
Node Time 95% HPD
1 6.71 5.18-7.98
2 5.03 3.45-7.49
3 3.96 2.63-5.04
4 3.77 2.50-4.90
5 2.90 1.71-4.37
East clade
West clade
E1
E2
W4
W2
W1
W3
Legends
Dispersal
Vicariance
Miocene
Pliocene
Pleistocene
7.0 6.0 5.0 4.0 3.0 2.0 1.0 0.0 Ma
(b)
85° E 90° E 95° E 100° E 105° E 110° E
45° N 40° N 35° N 30° N 25° N
HZ3 HZ2 C
KMG GS XLS TSM WLG B
TBS A SNJ
SGN BLS
ZG DDL MXG D
YH SJL ZF E
YL
Kilometers
0 100 200 400 600 800
(c)
R_e
3.0 2.5 2.0 1.5 1.0 0.5 0.0
1.0 0.8 0.6 0.4 0.2 0.0
c
−6 −5 −4 −3 −2 −1
Time

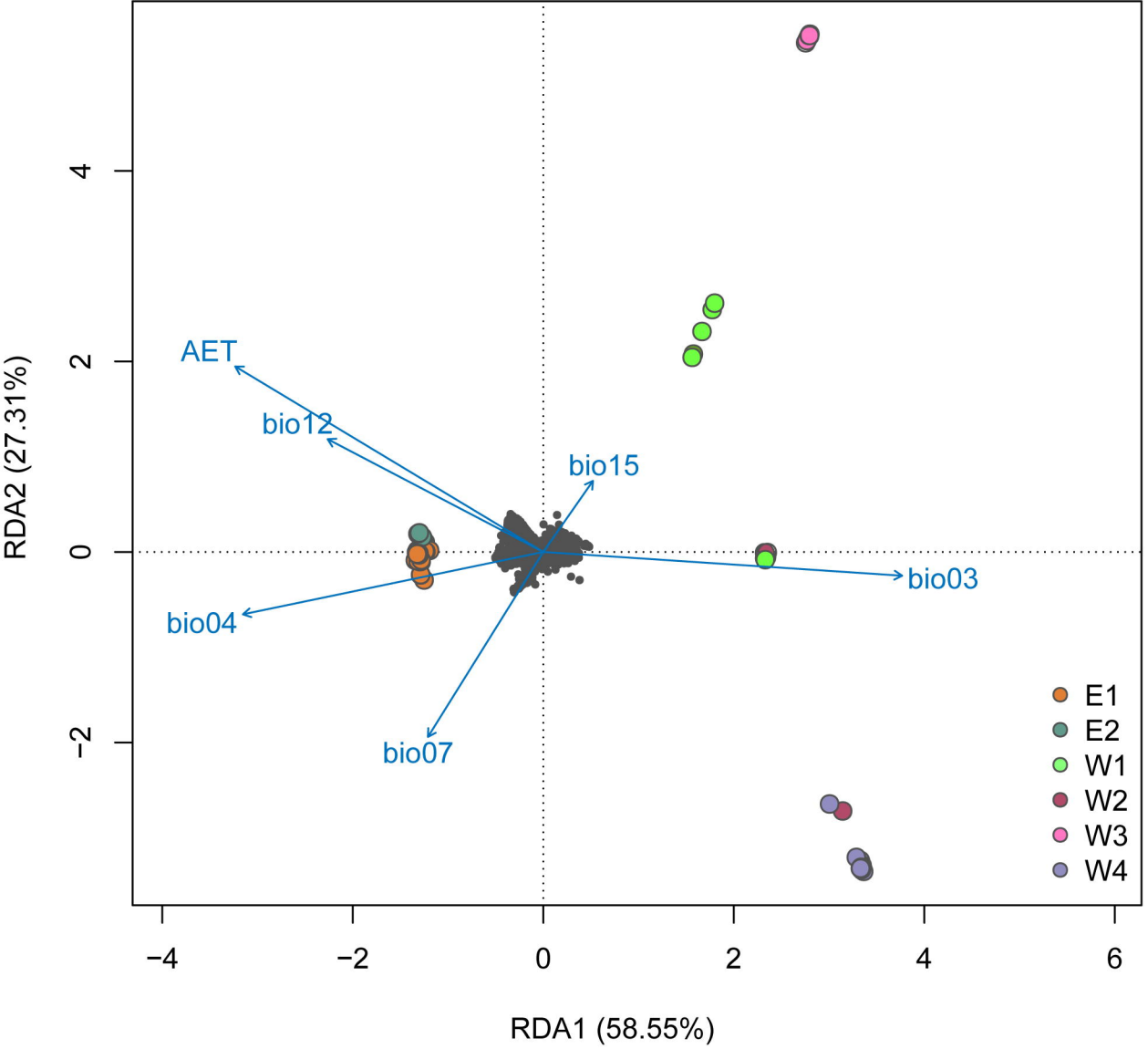

(a)
Genetic Distance
Geographical distance
R^2= 0.443
P = 0.001**
(b)
Genetic Distance
Resistance (niche)
R^2= 0.315
P = 0.017*
(c)
Locality
E1
E2
W1
W2
W3
W4
Conductance
High : 18.6832
Low : 0
Qinghai
Gansu
Shannxi
Hubei
Sichuan
Tibet
Yunnan

(a)


R^2 weighted importance
Isothermality (bio03)
Actual evapotranspiration (AET)
Temperature annual range (bio07)
Temperature seasonality (bio04)
Seasonality precipitation (bio15)
Annual precipitation (bio12)
Mean diurnal range (bio02)
Annual mean temperature (bio01)
Min temperature of coldest month (bio06)
Precipitation of wettest month (bio13)
Mean temperature of driest quarter (bio09)
Precipitation of driest month (bio14)
Temperature
Precipitation
0.00
0.02
0.04
0.06
0.08
0.10

(b)


Cumulative importance
bio03
AET
bio07
bio04
bio15
bio12


AET
bio12
bio15
bio03
bio04
bio07
E1
E2
W1
W2
W3
W4
RDA2 (27.31%)
RDA1 (58.55%)
4
2
0
−2
−4
−2
0
2
4
6