



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Allopatric divergence and hybridization within *Cupressus chengiana* (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China

Citation for published version:

Li, J, Milne, R, Ru, D, Miao, J, Tao, W, Zhang, L, Xu, J, Liu, J & Mao, K 2020, 'Allopatric divergence and hybridization within *Cupressus chengiana* (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China', *Molecular Ecology*. <https://doi.org/10.1111/mec.15407>

Digital Object Identifier (DOI):

[10.1111/mec.15407](https://doi.org/10.1111/mec.15407)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Molecular Ecology

Publisher Rights Statement:

"This is the peer reviewed version of the following article: Li, J, Milne, R, Ru, D, et al. Allopatric divergence and hybridization within *Cupressus chengiana* (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China. *Mol Ecol*. 2020; 00: 1– 17. <https://doi.org/10.1111/mec.15407>, which has been published in final form at <https://doi.org/10.1111/mec.15407>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions."

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Title:

Allopatric divergence and hybridization within *Cupressus chengiana* (Cupressaceae), a threatened conifer in the northern Hengduan Mountains of western China

Authors:

Jialiang Li¹, Richard Ian Milne², Dafu Ru³, Jibin Miao¹, Wenjing Tao¹, Lei Zhang¹, Jingjing Xu¹, Jianquan Liu¹, Kangshan Mao¹

Affiliation:

¹ Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, Sichuan, P.R. China;

² Institute of Molecular Plant Sciences, The University of Edinburgh, Edinburgh EH9 3JH, UK;

³ State Key Laboratory of Grassland Agro-Ecosystem, Institute of Innovation Ecology, Lanzhou University, Lanzhou 730000, China

Author for correspondence:

Kangshan Mao

Tel: +8613608091356

Email: maokangshan@163.com, maokangshan@scu.edu.cn

Abstract

Having a comprehensive understanding of population structure, genetic differentiation and demographic history is important for conservation and management of threatened species. High-throughput sequencing (HTS) provides exciting opportunities to address a wide range of factors for conservation genetics. Here, we generated HTS data and identified 266,884 high-quality SNPs from 82 individuals, to assess population genomics of *Cupressus chengiana* across its full range, comprising the Daduhe River (DDH), Minjiang River (MJR) and Bailongjiang River (BLJ) catchments in western China. [Each of ADMIXTURE](#), PCA and phylogenetic analyses indicated that each region contains a distinct lineage, with high levels of differentiation between them (DDH, MJR and BLJ lineages). MJR was newly distinguished compared to previous surveys, and evidence including coalescent simulations supported a hybrid origin of MJR during the Quaternary. Each of these three lineages should be recognized as an evolutionarily significant unit (ESU), due to isolation, differing genetic adaptations and different demographic history. Currently, each ESU faces distinct threats, and will require different conservation strategies. Our work shows that population genomic approaches using HTS can reconstruct the complex evolutionary history of threatened species in mountainous regions, and hence inform conservation efforts, and contribute to the understanding of high biodiversity in mountains.

Keywords: Population genomics, Mountainous regions, Threatened species, ESUs, Hybridization

1 **Introduction**

2 Resolving taxonomic uncertainties and identifying conservation units (CUs) are crucial for the
3 conservation of biological diversity, providing managers and policy makers with a clear
4 understanding of the population unit boundaries of endangered species (Funk, McKay,
5 Hohenlohe, & Allendorf, 2012). Accurate determination of taxonomic status can avoid both
6 underestimation of the necessary protection status for endangered species, and wasted effort on
7 abundant species (Frankham, Ballou, & Briscoe, 2010). Moreover, a species may comprise
8 several genetically distinct evolutionary units, each of which warrants conservation in its own
9 right (Palsbøll, Berube, & Allendorf, 2007). In recent decades, genetic markers that have been
10 used to define CUs have included microsatellite loci (SSR) (Wang, Liang, Hao, Chen, & Liu,
11 2018a), nuclear DNA (nrDNA) (Shang et al., 2015), chloroplast DNA (cpDNA) (Feng, Xu, &
12 Wang, 2018; Petit, El Mousadik, & Pons, 1998) and mitochondrial DNA (mtDNA) (Moritz,
13 1994; Torres-Cambas, Ferreira, Cordero-Rivera, & Lorenzo-Carballa, 2017). However, these
14 markers only yield a few variable loci, and so are generally inadequate for characterizing the
15 population genetic structure of species with complex demographic history and adaptive patterns
16 (Funk et al., 2012).

17 Hybridization, including subsequent introgression, either between species or across
18 intraspecific lineages, can complicate the identification of taxonomic and conservation units,
19 and hence the assignment of priorities when allocating conservation efforts (Allendorf, Leary,
20 Spruell, & Wenburg, 2001; Naciri & Linder, 2015). Nevertheless, hybridization among
21 diverging lineages is prevalent in nature, and about 25% of plant and 10% of animal species are
22 known to have undergone hybridization (Mallet, 2007). Hybridization becomes a conservation
23 issue when gene flow erodes population distinctions, especially when the distinctness of a rare
24 species or race is threatened by introgression from a commoner, sometimes alien, species or
25 race (Allendorf et al., 2001). Equally, however, hybridization is increasingly recognized as a
26 generator of adaptation and biodiversity (Lamichhaney et al., 2018; Rieseberg, 2019). For
27 example, adaptive traits transferred between species by introgression can promote adaptive
28 radiations (Edelman et al., 2019; Rieseberg, 2019). Combining distinct genomes in novel ways,
29 coupled with stabilization and isolation from parents, may form new lineages or species very
30 quickly (Barrera-Guzman, Aleixo, Shawkey, & Weir, 2018; Lamichhaney et al., 2018), which
31 can result in rapid speciation. Thus, hybridization has played an important role in the evolution
32 of many species/lineages (Goulet, Roda, & Hopkins, 2017). Therefore, policy making requires
33 an understanding of the roles hybridization has played in any threatened species. However,
34 detection of hybridization is difficult using traditional molecular markers (Allendorf,
35 Hohenlohe, & Luikart, 2010), with hundreds of markers usually required for accurate
36 determination of the dynamics of hybridization (Allendorf, Hohenlohe, & Luikart, 2010).
37 Recently, the application of high-throughput sequencing (HTS) technologies has made rapid
38 collection of genomic data much easier (Funk et al., 2012), providing exciting opportunities to
39 quantify adaptive variation (Hämälä & Savolainen, 2019; Ma et al., 2019), accurately delimit
40 taxa within critical species complexes (Fennessy et al., 2016; Liu et al., 2018) and assess
41 complex genetic structure, including the effects of hybridization (Ru et al., 2018; Sun et al.,
42 2018; vonHoldt, Brzeski, Wilcove, & Rutledge, 2018). This expanded genomic data will permit
43 many new questions to be addressed regarding conservation (Allendorf, Hohenlohe, & Luikart,

44 2010), which will make the conservation and management of threatened species more effective.

45 The Hengduan Mountains (HDMs) region, at the eastern edge of the Qinghai-Tibetan Plateau
46 (QTP), possesses exceptional richness in plant diversity, with about 12,000 species in 1500
47 genera of vascular plants (Li & Li, 1993; Liu, Duan, Hao, Ge, & Sun, 2014b; Wu, 1988), of
48 which >3300 species (>27.5%) and 90 genera (>6%) are endemic (Sun, Zhang, Deng, &
49 Boufford, 2017). Many of these occur in specific habitats that are also rare and threatened, e.g.
50 *Larix mastersiana*, *Cephalotaxus lanceolata* and *Parakmeria omeiensis* (Fu, 1992; Yong, Bing,
51 & Njenga, 2017). The genetic structure and demographic history of species in the HDMs have
52 been shaped by local orogenetic events and climate oscillations (Favre et al., 2015; Liu et al.,
53 2014b). Mountain uplifts generated geographic barriers that limited gene flow among
54 populations, affecting divergence of lineages, genetic structure, and the evolution of alpine
55 plants (Liu, Sun, Ge, Gao, & Qiu, 2012; Shahzad, Jia, Chen, Zeb, & Li, 2017; Wen, Zhang, Nie,
56 Zhong, & Sun, 2014). This region was also affected by a series of Quaternary glaciations
57 (Zheng, Xu, & Shen, 2002; Zhou & Li, 1998), among which the two largest on the QTP were
58 the Xixiabangma Glaciation and the Naynayxungla Glaciation, which occurred around 1.2-0.8
59 million years ago (Mya) and 0.72-0.5 Mya, respectively (Zheng et al., 2002; Zhou & Li, 1998).
60 Many tree species on the QTP moved south and/or to lower altitudes during the ice ages (Liu
61 et al., 2014b; Qiu, Fu, & Comes, 2011), which could drive intraspecific divergence, or
62 hybridization if diverged lineages share a refugium (Du, Hou, Wang, Mao, & Hampe, 2017;
63 Liu, Abbott, Lu, Tian, & Liu, 2014a; Liu et al., 2013; Ren et al., 2017; Sun et al., 2014). Hence
64 species distributed in the HDMs may have complex evolutionary histories, necessitating large
65 numbers of markers to accurately delimit both closely related species and intraspecific lineages.

66 The Minjiang Cypress, *Cupressus chengiana* S.Y. Hu, is a threatened conifer that occurs
67 around the northern HDMs, where it is a vital ecological component of arid valley ecosystems,
68 and is regularly used for house construction and furniture production. It has suffered a sharp
69 decline in range and population size because of logging (Hao et al., 2006; Zeng & Yang, 1992),
70 and is now classified as “Vulnerable” by the IUCN (Zhang & Christian, 2013), and as a
71 “Second-Class Endangered Plant” of China (Fu, 1992). Early studies using three regions of
72 chloroplast genome (Xu et al., 2010), six pairs of nuclear microsatellite markers (Lu et al., 2014)
73 and ten nuclear DNA sequence loci (Xu et al., 2017) demonstrated clear genetic differentiation
74 between Bailongjiang river material in Gansu province (hereafter labelled BLJ) and material
75 from the Daduhe and Minjiang rivers in Sichuan Province. This suggested that *C. chengiana*,
76 comprises two evolutionary significant units (ESUs): one in BLJ, and the other Daduhe (DDH)
77 plus Minjiang (MJR). The BLJ material would currently satisfy the IUCN (2012) criterion of
78 “Endangered” if treated alone. However, currently available data cannot provide a
79 comprehensive understanding of its genetic status, and therefore population genomic data are
80 needed to address broader factors of conservation for this threatened species. Here, we collected
81 HTS data to characterise genetic variation across *C. chengiana* populations to address the
82 following questions. (i) How many ESUs can be identified within *C. chengiana* based on HTS
83 data? (ii) What roles have past environmental changes and hybridization played in its
84 evolutionary and population history? (iii) Do adaptive differences exist among the ESUs? (iv)
85 What conservation implications and recommendations can be inferred from our data, for this
86 rare conifer? A robust inference for the genetic status and lineage evolutionary history of *C.*

87 *chengiana* would facilitate conservation and management of this threatened species, as well as
88 shedding light on the evolution of species and [populations](#) within the HDM biodiversity hotspot.
89

90 **Materials and Methods**

91 **Sampling and RNA sequencing**

92 *Cupressus chengiana* is now restricted to three isolated arid valleys between 800 and 2900 m
93 a.s.l. in the upper reaches of the Daduhe (DDH), Minjiang (MJR) and Bailong (BLJ) rivers (Fu,
94 Yu, & Farjon, 1999; Hao et al., 2006; Xu et al., 2017). The Minjiang river material lies roughly
95 between the other two regions in both location and altitude, and is separated from Bailongjiang
96 and Daduhe rivers by the Minshan and Qionglai Mountains, respectively. We collected across
97 the full range of *C. chengiana* from 2016 to 2018, and sampled fresh leaves in thirteen locations
98 across the three river catchments for RNA-seq: 35 individuals for BLJ, 17 for MJR, and 30 for
99 DDH (Table 1, Figure 1). In each location, the distance from every sampled individual to any
100 other was more than 50m, to avoid the impact of potential clonal reproduction. Five samples
101 each of *C. duclouxiana* and *C. gigantea* that were collected in our previous work (Ma et al.,
102 2019) were included as outgroups.

103 Fresh leaves were put in liquid nitrogen immediately and kept below -80°C before extraction.
104 RNeasy Pure Plant Plus Kits (TIANGEN® Biotech, Beijing, China), which provide an
105 efficient method for purification of total RNA from plant tissues rich in polysaccharides and
106 polyphenolics, was used to isolate total RNAs. A NanoPhotometer® spectrophotometer
107 (IMPLEN, CA, USA) was used to check RNA purity, and a Qubit® RNA Assay Kit in Qubit®
108 2.0 Fluorometer (Life Technologies, CA, USA) was used to measure RNA concentration. RNA
109 integrity was assessed via the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100
110 system (Agilent Technologies, CA, USA).

111 Sequencing libraries were generated using a NEB Next® Ultra™ RNA Library Prep Kit for
112 Illumina® (NEB, USA) following manufacturer's recommendations. Briefly, index codes were
113 first added to attribute sequences to each sample, and then mRNAs were fragmented into short
114 sequences. cDNA was synthesized for each RNA [fragment](#), and NEBNext Adaptors with
115 hairpin loop structure were ligated to prepare for each cDNA fragments. PCR was then
116 performed with Phusion High Fidelity DNA polymerase. Finally, PCR products were purified
117 and library quality was assessed on the Agilent Bioanalyzer 2100 system. After the above steps,
118 the library preparations were sequenced on Illumina HiSeq X Ten platforms to generate 150bp
119 paired-end raw reads.

120

121 **SMRT sequencing of full-length transcriptome**

122 To get a high-quality reference, single-molecule real-time (SMRT) sequencing was used to
123 obtain the full-length transcriptome of *C. chengiana*. Fresh leaves, stems and female seed cones
124 of one individual were sampled and then snap-frozen in liquid nitrogen. After RNA extraction
125 and quality checking (see above), pooled RNA comprising equal amounts of high-quality RNA
126 from the three tissues was used for cDNA synthesis and library construction using a SMART
127 PCR cDNA kit (Clontech, Mountain View, CA, USA) and the BluePippin Size Selection
128 System. This library was subsequently sequenced on a Pacific Biosciences (PacBio) RS
129 sequencing instrument.

130 Raw reads of PacBio full-length isoform sequencing were processed using the SMRT Link
131 ver. 4.0 software (<https://www.pacb.com/support/softwaredownloads>). From this, Circular
132 Consensus Sequences were generated; these were then classified into full length and non-full
133 length reads by examining for poly(A) signal and 5' and 3' adaptors. Consensus isoforms were
134 identified from full-length non chimera sequences (FLNC), and polished with non-full length
135 reads to obtain high-quality isoforms (HQ, above 99% accuracy) using the Quiver algorithm
136 from SMRT Link. The Illumina RNA-seq data from the same individual was used to correct the
137 PacBio sequences performed in LoRDEC (Rivals & Salmela, 2014).

138 To eliminate confounding effects from microbial and plastid DNA, we removed sequences
139 showing high similarity with either microbial DNA sequences (MBGD, downloaded from
140 http://mbgd.genome.ad.jp/htbin/view_arch.cgi (Uchiyama, Higuchi, & Kawai, 2010) or any
141 part of the complete chloroplast genome of *C. jiangeensis* (GenBank accession: NC_036939.1)
142 (Li et al., 2019). The HQ full-length polished consensus transcripts had their redundancy
143 removed by CD-HIT-EST ver. 4.6.1 (Li & Godzik, 2006), and were then processed with Cogent
144 ver. 3.1 (<https://github.com/Magdoll/Cogent>) to obtain a final set of unique transcript isoforms
145 (referred to as UniIsoforms).

146

147

148 **Read mapping and SNP calling**

149 Illumina raw reads were filtered via Trimmomatic ver. 0.36 (Bolger, Lohse, & Usadel, 2014).
150 This involved first removing adapters or bases from either the start or the end of reads with base
151 Phred quality score (Q) < 3, and then discarding poly-N reads (those with >10% unidentified
152 nucleotides) and low-quality reads (those with over 50% of bases with Q < 3). Finally, reads
153 with more than 36 bases after trimming were retained as quality-filtered reads.

154 We used BWA-MEM ver. 0.7.12 (Li & Durbin, 2009) with default parameters to align the
155 quality-filtered reads of each individual to the nuclear transcriptome sequences (UniIsoforms).
156 SAMTOOLS ver. 1.2 (Li et al., 2009a) was run to convert Sequence Alignment/Map (SAM)
157 files to Binary Alignment/Map (BAM) files, and sort BAM files. We used PICARDTOOLS ver.
158 2.8.1 (<http://broadinstitute.github.io/picard/>; Broad Institute, GitHub Repository) to mark and
159 remove duplicate reads. The regions around indels were realigned using the
160 RealignerTargetCreator and IndelRealigner tools in GATK ver. 3.7 (DePristo et al., 2011). We
161 used the “mpileup” command in SAMTOOLS (Li et al. 2009) to identify SNPs with parameters
162 “-q 20 -Q 20 -t AD,ADF,ADR,DP,SP”. Data were filtered with the following processes: SNPs
163 with a mapping quality <30, a mapping depth <10, genotyping rate <50% per group, minor
164 allele frequency (MAF) <5%, or in 5bps windows around any indel. The program SnpEff
165 (Cingolani et al., 2012) was used to annotate SNPs.

166

167 **Genetic structure and phylogenetic inference**

168 We used VCFtools (Danecek et al., 2011) and a perl script (Ru et al., 2018) to estimate the value
169 of Tajima’s *D*, population genetic differentiation (F_{ST}), absolute differentiation (D_{XY}) and
170 nucleotide diversity (π , for all callable sites). To keep rare variants, the MAF control was not
171 performed for the data set that was used to calculate Tajima’s *D* and π .

172 A model-based evolutionary clustering analysis via ADMIXTURE ver. 1.23 (Alexander &

173 Lange, 2011) was used to identify evolutionary clusters. We used VCFtools and PLINK ver.
174 1.90 (Purcell et al., 2007) to convert input data and remove linkage disequilibrium sites with
175 the parameter set as “--indep-pairwise 50 5 0.4”. The most likely number of genetic clusters (K)
176 was estimated in ADMIXTURE ver. 1.23, by computing parameters’ maximum-likelihood
177 estimates. Ten independent simulations were run for each value of K from one to ten with cross
178 validation to investigate the convergence of samples. The minimization of cross-validation error
179 among all runs was used to determine the most likely number of clusters. In order to compare
180 with results from ADMIXTURE, principal component analysis (PCA) on *C. chengiana*
181 individuals was conducted to explore the species’ genetic structure, using the SMARTPCA
182 program in the software EIGENSOFT ver. 6.1.3 (Price et al., 2006).

183 A perl script (Ru et al., 2018) was then used to generate concatenated sequences of each
184 individual. Here, only neutral sites (4DTv, four-fold degenerate sites) were retained to construct
185 phylogenetic inference. The software jModelTest (Darriba, Taboada, Doallo, & Posada, 2012)
186 was used to select the best-fit model of nucleotide substitution using Akaike Information
187 Criterion. Maximum-likelihood (ML) trees were reconstructed in RAxML ver. 8.2.9
188 (Stamatakis, 2014) using *C. duclouxiana* and *C. gigantea* as outgroups. We performed 200
189 bootstrap replicates to calculate the node support values.

190 **Phylogenetic-network analysis**

191 To obtain single-copy genes, one individual from each of the three groups (BLJ, MJR and DDH)
192 within *C. chengiana*, plus one *C. gigantea* accession, were selected for orthologous sequences
193 searching. Quality-filtered reads were assembled into contigs in Trinity ver. 2.8.4 (Grabherr et
194 al., 2011) with default parameters. We used the BUSCO (Simão, Waterhouse, Ioannidis,
195 Kriventseva, & Zdobnov, 2015) database to assess the transcriptome assembly. The longest
196 transcript for each gene was selected by a custom python script (see Supplemental Information),
197 and then we used CD-HIT-EST ver. 4.6.1 (Li & Godzik, 2006) to eliminate redundancies.
198 Coding and peptide sequences were predicted by TransDecoder ver. 5.5.0 (Haas et al., 2013).
199 The 1:1:1:1 orthologous gene data set was generated for BLJ, MJR, DDH and *C. gigantea*
200 (outgroup) in Orthofinder ver. 2.3.3 (Emms & Kelly, 2015). The corresponding coding
201 sequences of each orthogroup were aligned via MAFFT (Katoh & Standley, 2013), and trimAL
202 ver. 1.4.1 (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009) was used to remove positions
203 with more than 50% missing data. Those aligned sequences that were longer than 300
204 nucleotides were retained to generate the best rooted ML trees in RAxML under the
205 GTR+GAMMA substitution model using rapid-bootstrapping approach. Meanwhile, we used
206 ASTRAL ver. 5.6.3 (Mirarab et al., 2014) to estimate the species tree with 100 bootstrap
207 replicates. Finally, a set of 10,227 orthologous gene trees was examined using PhyloNet ver.
208 3.6.10 (Than, Ruths, & Nakhleh, 2008) to infer reticulate evolutionary relationships for *C.*
209 *chengiana*. A custom python script (see Supplemental Information) was used to convert the
210 format for input files. We used the command InferNetwork_ML_Bootstrap with the parameter
211 “InferNetwork_ML_Bootstrap 2 -pl 6 -di” to infer a species network, where the maximum
212 number of reticulations was set as 2 and the sampling process was repeated 100 times in
213 parametric bootstrap by default.

214

215 **Demographic modelling and gene flow**

216 Although some synonymous sites are expected to evolve under purifying selection (Lawrie,
217 Messer, Hershberg, & Petrov, 2013), they are generally assumed to be under weak selection and
218 nearly neutral (Akashi, 1995; Yang & Nielsen, 2008). Therefore, many studies used 4DTv sites
219 to minimize the bias in demographic inferences when more neutral sites were unavailable
220 (Marburger et al., 2019; Zhang et al., 2017). Here, we also used SNPs at 4DTv sites for
221 demographic inference to reduce the impact of natural selection. SNPs without MAF filtering
222 were further filtered to remove all missing data across all individuals sampled. We used a
223 perl script (Ru et al., 2018) to generate folded two-dimensional joint site frequency spectra (2D-
224 SFS). The 2D-SFS for all *C. chengiana* individuals was estimated by *fastsimcoal2* (FSC2)
225 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). We used the Akaike information
226 criterion (AIC) to rank all models, and chose the model that was the best fit to the data, to reduce
227 subjective bias. Based on early results confirming that MJR was roughly intermediate between
228 the other two populations, we selected four possible scenarios, and compared 11 possible
229 models based on these. The scenarios were: (i) a hybrid origin of the MJR lineage from the
230 other two lineages (models 1-3, Figure S1); (ii) divergence of the MJR and BLJ lineages from
231 a recent common ancestor ((MJR, BLJ), DDH), models 4-6, Figure S1); (iii) divergence of the
232 MJR and DDH lineages from a recent common ancestor ((MJR, DDH), BLJ), models 7-9,
233 Figure S1); and (iv) radiative evolution among all three lineages ((MJR, DDH, BLJ), models
234 10-11, Figure S1). The mutation rate was set as 9.7×10^{-9} per site per generation following Li
235 et al. (2012)'s estimate for Cupressaceae species. Because of long generation times in
236 gymnosperms (De La Torre, Li, Van de Peer, & Ingvarsson, 2017), we assumed an average
237 generation time of 50 years for *C. chengiana*, which is about three to five times the age at first
238 reproduction but less than the maximum expected lifespan of conifers (Bouillé & Bousquet,
239 2005). This generation time was also commonly adopted in many studies for conifers (Bouillé
240 & Bousquet, 2005; Li et al., 2009b; Ma et al., 2019). A parameter bootstrapping approach was
241 used to construct 95% confidence intervals (CI) with 50 independent runs. We used the Stairway
242 plot method (Liu & Fu, 2015) to investigate the detailed population demographic history for
243 each lineage using the folded one-dimensional SFS (Figure S2) from 4DTv sequences.

244 We further used the ABBA-BABA test (*D*-statistics) (Durand, Patterson, Reich, & Slatkin,
245 2011) to test for the possibility of gene flow among the three groups. Based on the patterns of
246 ancestral and derived alleles in the ingroups and outgroups, this analysis can distinguish
247 between incomplete lineage sorting and hybridization (Elgvin et al., 2017; Zhang et al., 2019).
248 Two topologies ((BLJ, MJR), DDH) and ((DDH, MJR), BLJ) were selected to calculate a *D*
249 value using ANGSD's ABBA-BABA multipopulation tool (Korneliussen, Albrechtsen, &
250 Nielsen, 2014) with five *C. duclouxiana* individuals as the outgroup. We used a *Z*-test to
251 determine if the *D* value was significantly deviated from zero. We considered *Z* scores >3 to be
252 significant.

253 **Detection of candidate genes and GO annotation**

254 In endangered species, adaptive differences between populations can create differing ecological
255 requirements, necessitating distinct management strategies for each of them (Crandall, Bininda-
256 Emonds, Mace, & Wayne, 2000). At first, we assumed no *a priori* information [in order to](#)
257 [examine](#) adaptive differentiation patterns for *C. chengiana*, and to test if there were genes under
258 strong directional selection in any set of populations (Funk, McKay, Hohenlohe, & Allendorf,

259 2012). We performed a global F_{ST} outliers test in BAYESCAN ver. 2.1 (Foll & Gaggiotti, 2008)
260 with default parameters. A subpopulation specific fixation index (F_{ST}) was used to estimate the
261 difference in allele frequency between the total population (all individuals of *C. chengiana*) and
262 each subpopulation (BLJ, MJR and DDH). To reduce the false discovery rate when making
263 decisions, q-values were calculated in BAYESCAN. Among outliers, those having a q-value
264 lower than 0.001 were treated as false positives, and hence the remaining 99.9% of
265 corresponding outliers were expected to not be false positives.

266 DDH group occupies the highest habitats (>2000m) among the three groups of *C. chengiana*.
267 Therefore, we further used the population branch statistic (PBS) (Yi et al., 2010) to identify
268 candidate genes related to high-altitude adaptation in DDH, comparing to the lowland BLJ
269 group (<1500m). Because individuals of *C. duclouxiana* and *C. gigantea* were clustered
270 together as a clade in an outgroup position to *C. chengiana* (Figure 2c, 2d), all ten individuals
271 of *C. duclouxiana*+*C. gigantea* were treated as a single outgroup population. First, pairwise F_{ST}
272 values were calculated in VCFtools, and the population divergence time T , in units scaled by
273 the population size, was obtained as $T = -\log(1-F_{ST})$ (Cavalli-Sforza 1969). Next, the length of
274 the branch leading to the DDH population since the divergence from the BLJ was estimated as:

$$275 \quad PBS_{DDH} = \frac{T_{DDH-BLJ} + T_{DDH-outgroup} - T_{BLJ-outgroup}}{2}.$$

276 Genes with the highest 1% of PBS were recognized as highly divergent genes, which could
277 result from positive selection (Wang et al., 2018b).

278 The software ANGEL ver. 2.4 (<https://github.com/PacificBiosciences/ANGEL>) was used to
279 predict open reading frames, and translate protein codes. We used BLASTP ver. 2.2.23
280 (Altschul et al., 1997) to compare the protein sequences, using the Swiss-Prot protein sequence
281 database for homology search analysis. Gene ontology (GO) terms for each gene were searched
282 using the Blast2GO program (Conesa et al., 2005). For the functions of genes and gene set
283 enrichment analysis, the analysis tool Singular Enrichment Analysis (SEA) in agriGO ver. 2.0
284 (Tian et al., 2017) was used. The Chi-squared test was used to calculate the statistical
285 significance of enrichment, with P -values below than 0.05 treated as significant following
286 adjustment via the Benjamini–Yekutieli procedure to control for false discovery rate.

287

288 Chloroplast phylogenetic analysis

289 Previous research has shown that chloroplast (cp) sequences represent a large fraction of the
290 plant transcriptome (Osuna-Mascaró, Rubio de Casas, & Perfectti, 2018; Ru et al., 2018). We
291 used the BWA-MEM algorithm of BWA ver. 0.7.12 (Li & Durbin, 2009) to map quality-filtered
292 reads of each individual against the published complete cp genome sequence of *C. jiangeensis*
293 (Li et al., 2019) to examine cpDNA variation. After removing duplicate reads, and realigning
294 regions around indels (see above), SAMTOOLS (Li et al. 2009) was used to identify SNPs.
295 SNPs were filtered with the following processes: SNPs were removed if they had a mapping
296 quality <30, a mapping depth <3, genotyping rate <50% per group, minor allele frequency
297 (MAF) <5%, or in 5bps windows around any indel. Concatenated sequences of each individual
298 were used to reconstruct ML trees in RAXML using *C. duclouxiana* and *C. gigantea* as
299 outgroups.

300

301 **Ecological niche modelling**

302 Current potential distributions for each group were predicted using ecological niche modelling
303 (ENM) in MAXENT ver. 3.3.4 (Phillips & Dudík, 2008) with the parameters set as “replicates:
304 20 replicates; type: subsample; maximum iterations: 5000; random test points: 25”. Climate
305 layers comprising 19 bioclimatic variables of a 2.5 arc minute resolution were downloaded from
306 WorldClim database (version 1.4, <http://www.worldclim.org>), and in addition, one altitude layer
307 was downloaded from the SRTM elevation database (<https://www2.jpl.nasa.gov/srtm/>). We
308 calculated pairwise Pearson’s correlation coefficients (r) (Dormann et al., 2013) for current
309 climate and altitude data across distributions of all trees performed in ENMTools ver. 1.4.3
310 (Warren, Glor, & Turelli, 2008; 2010). Any factor that had a correlation coefficient greater than
311 0.7 with two or more other factors was excluded. The geographic coordinates of 18 locations
312 from DDH, 13 from MJR and 14 from BLJ (Table S1) were collected from field investigations
313 or previous publications (Xu et al., 2010, 2017), which covered most of the known area of
314 occupancy of this species; these were inputted into MAXENT. We conducted a hierarchical
315 partitioning approach (Chevan & Sutherland, 1991) to confirm which variable independently
316 contributed most, using the R package hier.part (Walsh, Mac Nally, & Walsh, 2003). The
317 performance of models was predicted by comparing their AUC values (the area under the
318 receiver operating characteristic curve). AUC values range from 0 to 1, where a score of 1
319 indicates perfect discrimination (Fielding & Bell, 1997). Niche overlap and identity tests were
320 performed in ENMTools to measure niche differences between groups by calculating
321 Schoener’s D (Schoener, 1968) and standardized Hellinger distance (I). The values of D and I
322 both ranged from 0 to 1, which indicated no niche overlap or identical niches respectively.

323 To further examine the patterns of distribution shifts within *C. chengiana*, we also used ENM
324 to predict potential distributions during the Last interglacial (LIG, ~120,000 - 140,000 years
325 ago), the last glacial maximum (LGM, about 22,000 years ago) and the future (2050, average
326 for 2041-2060) for each group. For the period of LGM, layers of four models available at the
327 WorldClim database were downloaded to generate average-over-pixel bioclimatic variables
328 following Zheng et al. (2017). Future climate data was available from the Fifth Phase of the
329 Coupled Model Intercomparison Project (CMIP5), while the climate data during LIG was
330 downloaded from WorldClim database (source: Otto-Bliesner et al. 2006).

331

332 **Results**

333 **Full-length transcriptome analysis using PacBio Iso-Seq**

334 Using mixed RNA samples of leaf, stem and female cone, we obtained 19.32G of nucleotide
335 (nt) reads of inserts (ROIs) from three SMRT cells. The number of ROIs was 13,637,084, and
336 the mean length was 1,417nt. The Iso-seq classification and clustering protocol yielded 47,546
337 polished high-quality (HQ) transcripts, while the N50 was 3,117nt (Table S2). UniIsoforms
338 were excluded from the final set if they had similarity to either microorganisms or the plastid
339 genome that was used as a reference. The total size of the reference UniIsoforms data set was
340 50.506M nt, and the N50 was 3,133nt (Table S2).

341

342 **SNP calling**

343 After removing low quality sequences, 3.4 billion filtered-quality reads were obtained for the

344 82 individuals from an Illumina platform. By mapping these filtered-quality reads to the
345 reference UniIsoforms, we identified 5.82 million nuclear SNPs. After quality control, a total
346 of 266,884 high-quality nuclear SNPs was retained. A total of 5,202 cpDNA SNPs was
347 successfully identified using the complete chloroplast genome of *C. jiangeensis* as a reference.
348 After all filtering steps, we finally retained 1,251 SNPs from which to reconstruct cp
349 phylogenetic trees.

350

351 **Population genetic structure and genetic diversity**

352 Three distinct genetic clusters were detected by both PCA and ADMIXTURE analyses. From
353 the PCA plot, the first principal component (PC1), which explained 12.26% of all genetic
354 variance, differentiated the three geographically distinct *C. chengiana* groups: MJR, BLJ and
355 DDH, with MJR occupying an intermediate space between BLJ and DDH according to PC1
356 (Figure 2b). Results of ADMIXTURE also indicated that three genetic groups ($K=3$) were
357 optimal (Figure S3). For $K=3$, a clear genetic differentiation among the same three groups was
358 detected, with the clearest differentiation between DDH and MJR (Figure 2a). In the scenario
359 of $K=2$, individuals of the BLJ and DDH clades clustered into two distinct groups, while the
360 MJR group contained a mixture of genetic components of the other two groups (Figure 2a),
361 which is consistent with MJR being of hybrid origin.

362 Of the entire set of 266,884 nuclear SNPs, 11,913 were specific to DDH, 11,489 to BLJ and
363 3,213 to MJR (Table 2). DDH and BLJ shared the fewest SNPs (14,748), whereas MJR shared
364 more with each of DDH (20,245) and BLJ (27,524, Table 2), consistent with a hypothesis of a
365 hybrid origin for MJR.

366 Regarding population differentiation, genetic distance was highest between BLJ and DDH
367 ($F_{ST}=0.1752$), whereas MJR had F_{ST} values of 0.1066 and 0.1397 with BLJ and DDH,
368 respectively. The average value of absolute divergence (D_{XY}) between BLJ and DDH
369 ($D_{XY}=0.3440$) was also greater than that between MJR and either BLJ ($D_{XY}=0.3008$) or DDH
370 ($D_{XY}=0.3104$, Table 2).

371 SNPs without MAF filtering were used to calculate the π and Tajima's D values. The average
372 π value for BLJ (0.0069, Table. 2) is less than that for MJR (0.0072, Table 2), while DDH has
373 the lowest π value (0.0064, Table 2). The average Tajima's D values are -0.1790, 0.0470, and
374 -0.1449 for BLJ, MJR and DDH respectively (Table 2).

375

376 **Phylogenetic inference for nuclear and chloroplast SNPs**

377 Based on the results of jModeltest (Table S3), we used the GTR+GAMMA model for ML tree
378 reconstruction. From the phylogeny for nuclear SNPs, three distinct lineages were detected,
379 corresponding exactly to BLJ, DDH and MJR (Figure 2c), with MJR closer to BLJ than to DDH.
380 A coalescent-based species tree generated by ASTRAL produced a very similar result (Figure
381 S4d). In contrast, an ML tree constructed from cp SNPs identified two distinct clades within *C.*
382 *chengiana*, with one comprising BLJ and the other DDH+MJR (Figure 2d).

383

384 **Reticulate evolutionary relationships within *C. chengiana***

385 The contig N50 of the assembled transcriptome for BLJ, MJR, DDH and *C. gigantea* is 1,592,
386 1,682, 1,595 and 1,143, respectively (Table S4). More than 80% of the genes in the BUSCO

387 plant set were covered by all four of the assembled transcriptomes (Table S4). A total of 10,233
388 single copy orthogroups was identified in Orthofinder, and 10,227 of them were retained to
389 reconstruct [gene trees using ML](#). Hence a total of 10,227 gene trees were generated in RAxML,
390 and of these, 3,975 (38.87%) showed the closest relationship between MJR and BLJ (Figure
391 S4a), whereas 3,426 (33.50%) clustered MJR with DDH (Figure S4b), and 2,826 (27.63%)
392 clustered DDH with BLJ (Figure S4c). Results of PhyloNet showed reticulate evolutionary
393 relationships among BLJ, MJR and DDH, indicating a hybrid origin for MJR. The inheritance
394 probability between MJR and BLJ was 58.87%, which was higher than that between MJR and
395 DDH (41.13%) (Figure 3c), indicating a greater genomic contribution of BLJ to MJR.

396

397 **Demographic history and gene flow**

398 By comparing the AIC values for all 11 models, the hybrid speciation model with continuous
399 migration among the three groups was the best-fitting model (model3, Figure 3b, Table S5).
400 Divergence between BLJ and DDH was dated to (4.23-) 4.56 (-4.87) Mya (incorporating 95%
401 CI; Table 3). The estimated hybrid parameter (α) indicated that ~62% of the nuclear genome of
402 the initial MJR population came from BLJ, and ~38% from DDH, which was consistent with
403 genetic admixture (Table 3; Figure 2a; Table S6). The population sizes for BLJ, MJR and DDH
404 were estimated to be 238,794, 114,433 and 166,952 respectively (Table 3; Table S6). The
405 ancestral effective population size ($N_A=323,866$, Table 3) was estimated to be larger than any
406 of these (Table 3). A stairway plot analysis showed that a decline of population size for DDH
407 occurred from 0.9 to 0.6 Mya, followed by an expansion 0.6-0.4 Mya, coinciding respectively
408 with the Naynayxungla glaciation, and its end (Figure 3a). In contrast, BLJ maintained a stable
409 and high effective population size (~160,000) over the past seven million years (Figure 3a). The
410 population size of MJR expanded rapidly until approximately 10-7 Mya, and declined to
411 ~110,000 around 1.5 Mya, and then maintained that size with little fluctuations.

412 Asymmetric gene flow between MJR and the other two groups was detected, with the rates
413 of migration from MJR to each of BLJ ($M_{1\leftarrow 2} = 9.22E-6$) and DDH ($M_{3\leftarrow 2} = 1.00E-5$) being
414 higher than those in the opposite direction ($M_{2\leftarrow 1} = 3.65E-6$; $M_{2\leftarrow 3} = 4.94E-6$, respectively)
415 (Figure 3b, Table 3). Results of the ABBA-BABA test suggested that significant gene flow had
416 occurred between MJR and both of BLJ and DDH on the genomic level (Table 4), which was
417 not consistent with the genetic pattern of MJR being the result of incomplete lineage sorting.

418

419 **Identification and characterization of outlier loci**

420 With a 0.1% threshold for the q-value, we identified 575 outlier SNPs (Figure 3d), which
421 suggested a divergent differentiation, and that these markers could have [been subject](#) to
422 divergent selection among MJR, BLJ and DDH, based on the Bayesian method performed in
423 BAYESCAN. The average F_{ST} estimated in BAYESCAN was (0.1718-) 0.1934 (-0.6522).
424 Nearly 90% of the SNPs (237,792 of 266,844; 89.10%) showed $F_{ST} < 0.2$, while the F_{ST} value
425 for outlier SNPs was high, i.e. (0.4628-) 0.5303 (-0.6522), suggesting that the three groups were
426 indeed greatly differentiated at outlier SNPs. These outliers might putatively be under divergent
427 selection, representing evidence of adaptive [differentiation](#) between the three groups. These
428 outliers were located in 226 genes, of which 157 were annotated in the Swiss-Prot protein
429 sequence database. Gene ontology enrichment analyses of all outliers detected 12 significantly

430 over-represented GO terms ($P < 0.05$, $FDR < 0.05$), including: “stilbene biosynthetic process”,
431 “coumarin biosynthetic process”, “lignin metabolic process”, “L-phenylalanine metabolic
432 process” and “double-stranded DNA binding” (Table S7).

433 We further used the PBS approach to identify genes potentially under positive selection in
434 the DDH group. A total of 127 genes (top 1%) were identified in DDH with $PBS_{DDH} \geq 0.7610$,
435 and 74 of them were annotated. In total, 18 significantly over-represented GO terms with
436 corrected P -value < 0.05 were identified (Table S8). Among the 18 GO terms in DDH, six had
437 also been identified by BAYESCAN, including the “stilbene biosynthetic process” and
438 “stilbene metabolic process” (Table S8). Furthermore, although no significant GO category
439 with corrected P -value < 0.05 was found to be involved in the functions of response to
440 abiotic/biotic stresses in highland environments, some genes with extreme PBS_{DDH} exhibited
441 the signature of high-altitude adaptation in DDH. These included six genes involved in “cellular
442 response to DNA damage stimulus”, 11 related to “positive regulation of response to stimulus”,
443 34 related to “response to abiotic stimulus”, and 14 related to “response to abscisic acid” (Table
444 S9). These genes were also extremely differentiated between DDH and either BLJ or MJR, and
445 different alleles for all of these genes were fixed between DDH and BLJ (Figure S5).

446

447 **Ecological niche differences among *C. chengiana* populations**

448 ENMs were constructed for the three *C. chengiana* groups to predict their current potential
449 distributions and then the model was projected to past and future scenarios. Seven bioclimatic
450 variables (Alt: altitude, Bio2: mean diurnal range, Bio3: isothermality, Bio4: temperature
451 seasonality, Bio15: precipitation seasonality, Bio16: precipitation of wettest quarter, and Bio19:
452 precipitation of coldest quarter) were retained with $r < 0.7$ in each pair. Values of AUC of all
453 models were 0.989 ± 0.007 for BLJ, 0.984 ± 0.006 for MJR, and 0.998 ± 0.001 for DDH, indicating
454 that all models performed better than random expectation. The environmental variables that
455 showed the highest independent contributions were Bio19 (37.32%), Bio19 (27%), and Alt
456 (22.04%) for BLJ, MJR, and DDH respectively (Figure S6). Observed measures of niche
457 similarity (D and I) were lower than null distributions for DDH vs either BLJ or MJR,
458 suggesting high niche differentiation between DDH and both of BLJ and MJR (Figure 4b).
459 However, D and I fell within the range of null distributions for BLJ vs MJR, suggesting that
460 few niche differences exist between these two (Figure 4b).

461 For the LIG model, all groupings were predicted to have undergone clear southward range
462 shifts (Figure S7). For both BLJ and MJR individually, northward distribution shifts were
463 predicted for LGM model, while the predicted present-day and LGM distributions were nearly
464 identical for the DDH group (Figure S7). The future model showed clear range expansions for
465 both MJR and DDH relative to the present day, while a clear distribution contraction was
466 predicted for BLJ (Figure S7).

467

468 **Discussion**

469 ***Cupressus chengiana* comprises three ESUs, and the Minjiang river ESU is of hybrid 470 origin**

471 Here, we employed population genomic data to explore the genetic diversity, genetic structure
472 and demographic history for the threatened conifer *C. chengiana*, to aid in its conservation.

473 Multiple lines of evidence presented here suggested that material from each of the three river
474 catchments (BLJ, DDH and MJR) forms a distinct genetic lineage, with a high level of genetic
475 differentiation between the three. ADMIXTURE and PCA demonstrated that no overlaps
476 existed between lineages, whereas the ML tree constructed from the nuclear SNPs demonstrated
477 that each lineage was reciprocally monophyletic. These three lineages might represent the early
478 stages of speciation by isolation, and each forms an important component of the conifer
479 diversity of the world. Hence, each of BLJ, DDH and MJR represents an evolutionarily
480 significant unit (ESU).

481 The major difference from previous results is the clear differentiation between DDH and
482 MJR, which had not been detected in past studies based on limited data (Lu et al., 2014; Xu et
483 al., 2017). Here, our analyses based on population genomic data presented evidence that the
484 newly recognized MJR ESU had a hybrid origin from the other two ESUs. Population genetic
485 analyses indicate that this ESU is genetically admixed between BLJ and DDH (Table 2, Figure
486 2a), and demonstrate a reticular evolutionary relationship (Figure 3c). Results of coalescent
487 analysis strongly favored a hybrid origin over non-hybrid scenarios (Figures 3b, S1, Table S5).
488 Thus, multiple analyses indicate that the MJR ESU might be a lineage of hybrid origin, with
489 ~62% of its nuclear composition derived from BLJ, and ~38% from DDH (Figures 3b).
490 Detection of admixture signals had been difficult in earlier studies using <10 loci (Allendorf et
491 al., 2010), but HTS data as used here provides abundant markers that can contribute to the
492 accurate description of dynamics of hybridization and introgression (Allendorf et al., 2010;
493 Witherspoon et al., 2007). Our work confirmed the advantages of population genomic
494 approaches using HTS for research concerning conservation genetics.

495

496 **Demographic history and gene flow among *C. chengiana* lineages**

497 The strong geographic structure here detected for the three *C. chengiana* lineages implied that
498 the Qionglai and Minshan Mountains may have been able to limit gene flow between
499 populations occupying separate valleys, at least in *C. chengiana*. According to the optimal
500 model from FSC2 analysis (Figure 3b, Table 3), divergence between DDH and BLJ occurred
501 (4.23-) 4.46 (-4.87) Mya. This falls within the timescale of the intense uplifts of the HDMs that
502 occurred from the Late Miocene onwards, approaching their highest elevation before the Late
503 Pliocene (Favre et al., 2015; Sun et al., 2011; Xing & Ree, 2017). Other lineage divergence
504 events in the HDMs or QTP during this period, include intraspecific differentiation in *Taxus*
505 *wallichiana* (~4.2 Mya) (Liu et al., 2013) and *Quercus aquifolioides* (~4.4 Mya) (Du et al.,
506 2017), plus interspecific differentiation between *C. gigantea* and *C. duclouxiana* (~3.35 Mya)
507 (Ma et al., 2019), all of which might be the results of uplift events in this area. If the Qionglai
508 and Minshan Mountains were uplifted at this time, these might have caused the divergence of
509 DDH from BLJ.

510 The hybridization event that formed the MJR group was estimated to have occurred at (1.14-)
511 1.34 (-1.41) Mya (Figure 3b), a little before the start of the Xixiabangma Glaciation (around
512 1.2 Mya). If hybridization occurred during this or the previous glaciation, then it might have
513 been triggered by southward and downhill migration of *C. chengiana* in response to climate
514 cooling. The Daduhe and Minjiang rivers meet at 400 a.s.l., whereas BLJ material might have
515 spread across suitable ground at lower altitudes to meet them. Previous studies showed that the

516 largest Quaternary glaciation had begun causing alterations to plant distributions in this period
517 (Li & Fang, 1999; Lisiecki & Raymo, 2007; Sun et al., 2014), and the homoploid hybrid species
518 *Picea purpurea* might also have originated during this time (~1.3 Mya) (Sun et al., 2014).

519 Alternatively, upward migration during a warmer interglacial might have reduced the barrier
520 presented by the Qionglai and Minshan Mountains, and led to contact at higher altitudes. Either
521 way, it is likely that the Minjiang river basin was first occupied by individuals of one lineage
522 (more likely BLJ, based on genetic similarity) and then invaded by genetic material of the other
523 ~1.34 Mya. This is also supported by phylogenomic analyses based on cp genome-wide SNPs,
524 wherein MJR and DDH were clustered together and barely distinguishable, whereas BLJ
525 formed a distinct and separate clade (Figure 2d), in common with Xu et al., (2010). In conifers,
526 cpDNA is transmitted via pollen, and nrDNA biparentally by seeds (Mogensen, 1996; Sun et
527 al., 2014), hence DDH appears to have been the pollen parent of MJR. This fits an idea that
528 BLJ material around the Minjiang river was invaded via DDH pollen to form the MJR lineage.
529 Taken overall, we can infer that both past orogeny and climatic events in HDMs region may
530 have acted as major factors in shaping the evolutionary history of *C. chengiana*.

531 FSC2 analysis indicates that gene flow between all three populations might have occurred
532 since they diverged (models 2 or 3, Fig S1). Admixture of cpDNA haplotypes between DDH
533 and MJR does not provide much evidence for gene flow due to poor resolution, except for one
534 sub-clade that also contains material of both groups. This could be a signature of a later contact
535 and/or gene flow event. Hence gene flow between populations does not necessarily occur at
536 present; perhaps more likely is that there were further periods of contact at high altitudes during
537 warmer interglacials, and/or low altitudes during glacial maxima.

538

539 **Adaptive distinctiveness of the three ESUs**

540 Outlier loci identified by BAYESCAN included many that were involved in biosynthetic and
541 metabolic processes of such secondary metabolites as stilbene, coumarin, lignin and L-
542 phenylalanine (Table S7). This might be evidence of adaptive differences among the three
543 lineages, because these compounds, especially stilbene, play key roles in defense mechanisms
544 in plants (Chong, Poutaraud, & Hugueney, 2009). Because of long lifespans, conifers are
545 vulnerable to attack by insects and pathogens, especially the combination of bark beetles and
546 their symbiotic pathogenic fungi (Fettig et al., 2007; Krokene, 2015). Both beetles and fungi
547 can be inhibited by stilbenes, which may be constantly present in bark and/or synthesized
548 following initial attack (Chong et al., 2009; Fettig et al., 2007; Kolosova & Bohlmann, 2012;
549 Krokene, 2015). Although there is little data available for the distribution of insects or fungal
550 pathogens, results from our ongoing analysis of soil microorganisms from *C. chengiana*
551 populations showed an un-even distribution of plant pathogens among the three ESUs (Wang
552 et al., unpublished data). In addition, Stilbenes may also be involved in responses to abiotic
553 stresses like wounding or ozone generated by ultraviolet radiation (Chiron et al., 2000; He, Wu,
554 Pan, & Jiang, 2008; Rosemann, Heller, & Sandermann, 1991). Therefore, evidence detected
555 here of strong selective pressure upon stilbene production mechanisms, implies that each ESU
556 might differ genetically from the others in how it responds to pathogens or abiotic stresses,
557 emphasizing the need to conserve all three to best protect the species.

558 Interestingly, many of the genes with high PBS_{DDH} were likewise involved in metabolic

559 and biosynthetic processes of the same compounds, indicating that genes related to these
560 compounds might have played an important role in the adaption of *C. chengiana* to different
561 environments. Comparing to BLJ, DDH is distributed at higher altitudes and might have
562 experienced stronger selection pressures. Those genes identified by PBS as under positive
563 selection in DDH did not include genes with significantly enriched functions directly related to
564 to high-altitude adaptation, however many genes that are functionally related to local adaptation
565 did have extreme PBS_{DDH} values and hence might underlie ecological divergence between DDH
566 and BLJ ESU. Of these genes, 34 and 11 were involved in the response to abiotic and biotic
567 stimuli, respectively, and 14 of them were involved in the defense response, all of which might
568 be the signature of DDH ESU to adapt to harsh environment (Table S9). In particular, these
569 genes included six involved in cellular response to DNA damage stimuli, and 13 related to
570 response to radiation (Table S9), which might be of have been important in adapting DDH
571 material to high altitudes, because such habitats are exposed to increased UV radiation which
572 can result in cell and DNA damage (Zeng et al., 2020).

573

574 **Conservation implications**

575 The three ESUs separated here each face very different conservation issues, so we recommend
576 that each ESU should be assessed independently with regard to its threatened status. Currently,
577 the BLJ ESU has a relatively large extent of occurrence, encompassing Jiuzhaigou County in
578 Sichuan, Wen and Zhouqu County and Wudu district in Gansu. However, its area of occupancy
579 is very narrow, only between 800m to 1500m in the dry valley along Bailongjiang River where
580 there are few steep rocky slopes. These populations are much more accessible to humans than
581 DDH or MJR, and while all areas of *C. chengiana* have suffered from long-term logging (Hao
582 et al., 2006; Zeng & Yang, 1992), the resulting sharp decline in population size may be worst
583 for BLJ, with populations fragmented, and only a few trees remaining. The estimated effective
584 population size by FSC2 is 238,794 (the greatest of the three ESUs), and the Stairway Plot
585 indicates that its effective population size has been stable at approximately 160,000 for the past
586 seven million years (Figure 3a). However, given the long generation times and that most
587 logging is recent, these statistics probably estimate population size before logging. Each habitat
588 of the BLJ ESU should be conserved carefully, and artificial transplanting among fragmented
589 habitats should be undertaken to reduce inbreeding and minimize the bottleneck effect.

590 Populations in DDH tend to occupy deep slopes near the river, restricting accessibility to
591 loggers. Hence many mature trees of *C. chengiana* remain as part of a wide extent of natural
592 pure forest along the Daduhe River. Noteworthy, our results suggest that a distinct series of
593 locally adapted genetic variations are harbored in DDH populations (Table S8; Figure S5).
594 However, this river is very suitable for hydropower development, with many stations built and
595 others on the way; therefore, large potential habitats of the DDH ESU could be flooded and
596 wiped out (Peng, Li, Wang, Xie, & Cao, 2011). Hence the DDH ESU faces a potentially high
597 risk of extinction, and seed collection for *ex situ* conservation is necessary from populations
598 threatened by development.

599 Our results suggest that the MJR ESU might be a lineage that experienced an independent
600 evolutionary history following a hybrid origin, making it eligible for protection as a “type 1”
601 taxon (Allendorf et al., 2001). This lineage is now endangered because the extent of suitable

602 arid habitats is very limited: its range comprises only a small natural forest around Li County
603 plus several fragmented patches in Mao and Wenchuan County. Consistent with this, its current
604 effective population size is the least of the three ESUs as estimated by both FSC2 ($N_2 = 114,433$;
605 Figure 3b) and Stairway Plot (Figure 3a). Our results indicate that this ESU merits conservation
606 in its own right, for which we recommend protection of existing populations and *in situ*
607 augmentation by planting more material. Because each ESU is genetically distinct, any
608 reintroductions or plantings should involve material from the same ESU, to avoid outbreeding
609 depression and preserve genetic distinctness.

610 From a wider perspective, our study emphasizes the utility of HTS data in conservation
611 genetics for threatened species that have complex genetic structure and evolutionary history. At
612 the same time, our findings also shed light on the formation of lineage diversity in biodiversity
613 hotspots like the HDMs, highlighting the likely roles of hybridization, local adaptation, orogeny
614 and climatic changes.

615

616 **Acknowledgements**

617 This work was supported by the National Natural Science Foundation of China (grant numbers
618 31622015, 31590821, 31370261), the National Basic Research Program of China (grant
619 number 2014CB954100) and Sichuan University (Fundamental Research Funds for the Central
620 Universities, SCU2019D013, SCU2018D006).

621

622

623 **References**

624

625 Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in
626 *Drosophila* DNA. *Genetics*, *139*(2), 1067.

627 Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry
628 estimation. *BMC Bioinformatics*, *12*(1), 246. doi:10.1186/1471-2105-12-246

629 Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation
630 guidelines. *Trends in Ecology & Evolution*, *16*(11), 613-622. doi: 10.1016/S0169-5347(01)02290-X

631 Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature*
632 *Reviews Genetics*, *11*(10), 697-709. doi:10.1038/nrg2844

633 Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation
634 guidelines. *Trends in Ecology & Evolution*, *16*(11), 613-622. doi:10.1016/s0169-5347(01)02290-x

635 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped
636 BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*,
637 *25*(17), 3389-3402. doi:10.1093/nar/25.17.3389

638 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data.
639 *Bioinformatics*, *30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

640 Bouillé, M., & Bousquet, J. (2005). Trans-species shared polymorphisms at orthologous nuclear gene loci among
641 distant species in the conifer *Picea* (Pinaceae): implications for the long-term maintenance of genetic
642 diversity in trees. *American Journal of Botany*, *92*(1), 63-73. doi:10.3732/ajb.92.1.63

643 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment
644 trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972-1973.

645 doi:10.1093/bioinformatics/btp348

646 Cavalli-Sforza, L. L. (1969). Human diversity. *Proc. 12th Int. Congr. Genet.*, 2, 405-416.

647 Chevan, A., & Sutherland, M. (1991). Hierarchical Partitioning. *The American Statistician*, 45(2), 90-96.

648 doi:10.1080/00031305.1991.10475776

649 Chiron, H., Drouet, A., Lieutier, F., Payer, H.-D., Ernst, D., & Sandermann, H. (2000). Gene Induction of Stilbene

650 Biosynthesis in Scots Pine in Response to Ozone Treatment, Wounding, and Fungal Infection. *Plant*

651 *Physiology*, 124(2), 865. doi:10.1104/pp.124.2.865

652 Chong, J., Poutaraud, A., & Huguene, P. (2009). Metabolism and roles of stilbenes in plants. *Plant Science*, 177(3),

653 143-155. doi:10.1016/j.plantsci.2009.05.012

654 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for

655 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80-92.

656 doi:10.4161/fly.19695

657 Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool

658 for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676.

659 doi:10.1093/bioinformatics/bti610

660 Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary

661 processes in conservation biology. *Trends in Ecology & Evolution*, 15(7), 290-295. doi:10.1016/S0169-

662 5347(00)01876-0

663 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant

664 call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. doi:10.1093/bioinformatics/btr330

665 Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and

666 parallel computing. *Nature Methods*, 9(8), 772-772. doi:10.1038/nmeth.2109

667 De La Torre, A. R., Li, Z., Van de Peer, Y., & Ingvarsson, P. K. (2017). Contrasting Rates of Molecular Evolution

668 and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution*,

669 34(6), 1363-1377. doi:10.1093/molbev/msx069

670 DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., . . . Daly, M. J. (2011). A

671 framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature*

672 *Genetics*, 43(5), 491-498. doi:10.1038/ng.806

673 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., . . . Lautenbach, S. (2013). Collinearity:

674 a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1),

675 27-46. doi:10.1111/j.1600-0587.2012.07348.x

676 Du, F. K., Hou, M., Wang, W., Mao, K., & Hampe, A. (2017). Phylogeography of *Quercus aquifolioides* provides

677 novel insights into the Neogene history of a major global hotspot of plant diversity in south-west China.

678 *Journal of Biogeography*, 44(2), 294-307. doi:10.1111/jbi.12836

679 Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., . . . Mallet, J. (2019). Genomic

680 architecture and introgression shape a butterfly radiation. *Science*, 366(6465), 594-599.

681 doi:10.1126/science.aaw2090

682 Elgvin, T. O., Trier, C. N., Torresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., . . . Saetre, G. P. (2017). The

683 genomic mosaicism of hybrid speciation. *Science Advances*, 3(6), e1602996. doi:10.1126/sciadv.1602996

684 Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons

685 dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1), 157. doi:10.1186/s13059-015-

686 0721-2

687 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference

688 from genomic and SNP data. *PLoS Genetics*, 9(10), e1003905. doi:10.1371/journal.pgen.1003905

689 Favre, A., Packert, M., Pauls, S. U., Jahnig, S. C., Uhl, D., Michalak, I., & Muellner-Riehl, A. N. (2015). The role
690 of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol Rev Camb Philos Soc*,
691 90(1), 236-253. doi:10.1111/brv.12107

692 Feng, L., Xu, Z.-Y., & Wang, L. (2018). Genetic diversity and demographic analysis of an endangered tree species
693 *Diplopanax stachyanthus* in subtropical China: implications for conservation and management. *Conservation*
694 *Genetics*. doi:10.1007/s10592-018-1133-0

695 Fennessy, J., Bidon, T., Reuss, F., Kumar, V., Elkan, P., Nilsson, M. A., . . . Janke, A. (2016). Multi-locus Analyses
696 Reveal Four Giraffe Species Instead of One. *Curr Biol*, 26(18), 2543-2549. doi:10.1016/j.cub.2016.07.036

697 Fettig, C. J., Klepzig, K. D., Billings, R. F., Munson, A. S., Nebeker, T. E., Negrón, J. F., & Nowak, J. T. (2007).
698 The effectiveness of vegetation management practices for prevention and control of bark beetle infestations
699 in coniferous forests of the western and southern United States. *Forest Ecology and Management*, 238(1),
700 24-53. doi:10.1016/j.foreco.2006.10.011

701 Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation
702 presence/absence models. *Environmental Conservation*, 24(1), 38-49. doi:10.1017/S0376892997000088

703 Foll, M., & Gaggiotti, O. (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both
704 Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, 180(2), 977-993.
705 doi:10.1534/genetics.108.092221

706 Frankham, R., Ballou, J. D., & Briscoe, D. A. (2010). *Introduction to Conservation Genetics*, 2nd edition.
707 Cambridge, UK: Cambridge University Press.

708 Fu, L., Yu, Y., & Farjon, A. (1999). Cupressaceae. In Z. Wu & P. Raven (Eds.), *Flora of China* (Vol. 4, pp. 62-77).
709 Beijing: Science Press.

710 Fu, L. K. (1992). *Red book of China: rare and threatened plants*. Beijing: Science Press.

711 Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating
712 conservation units. *Trends in Ecology & Evolution*, 27(9), 489-496. doi:10.1016/j.tree.2012.05.012

713 Goulet, B. E., Roda, F., & Hopkins, R. (2017). Hybridization in plants: old ideas, new techniques. *Plant Physiology*,
714 173(1), 65-78. doi:10.1104/pp.16.01340

715 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-
716 length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7),
717 644-652. doi:10.1038/nbt.1883

718 Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Regev, A. (2013). De novo
719 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and
720 analysis. *Nature Protocols*, 8, 1494. doi:10.1038/nprot.2013.084

721 Hämmälä, T., & Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in *Arabidopsis lyrata*.
722 *Molecular Biology and Evolution*, 36(11), 2557-2571. doi:10.1093/molbev/msz149

723 Hao, B., Li, W., Linchun, M., Li, Y., Rui, Z., Mingxia, T., & Weikai, B. (2006). A Study of Conservation Genetics
724 in *Cupressus chengiana*, an Endangered Endemic of China, Using ISSR Markers. *Biochemical Genetics*,
725 44(1), 29-43. doi:10.1007/s10528-006-9011-8

726 He, S., Wu, B., Pan, Y., & Jiang, L. (2008). Stilbene Oligomers from *Parthenocissus laetevirens*: Isolation,
727 Biomimetic Synthesis, Absolute Configuration, and Implication of Antioxidative Defense System in the Plant.
728 *The Journal of Organic Chemistry*, 73(14), 5233-5241. doi:10.1021/jo8001112

729 IUCN. (2012). *IUCN red list categories and criteria: version 3.1*. Switzerland: IUCN Gland.

730 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in

731 performance and usability. *Molecular Biology and Evolution*, 30(4), 772-780. doi:10.1093/molbev/mst010

732 Kolosova, N., & Bohlmann, J. (2012). Conifer Defense Against Insects and Fungal Pathogens. In R. Matyssek, H.

733 Schnyder, W. Oßwald, D. Ernst, J. C. Munch, & H. Pretzsch (Eds.), *Growth and Defence in Plants: Resource*

734 *Allocation at Multiple Scales* (pp. 85-109). Berlin, Heidelberg: Springer Berlin Heidelberg.

735 Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing

736 Data. *BMC Bioinformatics*, 15(1), 356. doi:10.1186/s12859-014-0356-4

737 Krokene, P. (2015). Chapter 5 - Conifer Defense and Resistance to Bark Beetles. In F. E. Vega & R. W. Hofstetter

738 (Eds.), *Bark Beetles* (pp. 177-207). San Diego: Academic Press.

739 Lamichhane, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid

740 speciation in Darwin's finches. *Science*, 359(6372), 224-227. doi:10.1126/science.aao4593

741 Lawrie, D. S., Messer, P. W., Hershberg, R., & Petrov, D. A. (2013). Strong Purifying Selection at Synonymous

742 Sites in *D. melanogaster*. *PLoS Genetics*, 9(5), e1003527. doi:10.1371/journal.pgen.1003527

743 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform.

744 *Bioinformatics*, 25(14), 1754-1760. doi:doi.org/10.1093/bioinformatics/btp324

745 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009a). The sequence

746 alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

747 doi:10.1093/bioinformatics/btp352

748 Li, J., & Fang, X. (1999). Uplift of Tibetan Plateau and environmental changes. *Chinese Sci Bull*, 44, 2117-2124.

749 doi:10.1007/BF03182692

750 Li, J., Wu, J., Zhang, L., Zhang, L., Wang, L., & Mao, K. (2019). The complete chloroplast genome of *Cupressus*

751 *jiangeensis* (cupressaceae), a critically endangered conifer species in China. *Conservation Genetics*

752 *Resources*, 11(1), 67-69. doi:10.1007/s12686-017-0970-3

753 Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide

754 sequences. *Bioinformatics*, 22(13), 1658-1659. doi:10.1093/bioinformatics/btl158

755 Li, X.-W., & Li, J. (1993). A preliminary floristic study on the seed plants from the region of Hengduan Mountain.

756 *Acta Botanica Yunnanica*, 15(3), 217-231.

757 Li, Y., Stocks, M., Hemmilä, S., Källman, T., Zhu, H., Zhou, Y., . . . Lascoux, M. (2009b). Demographic histories

758 of four spruce (*Picea*) species of the Qinghai-Tibetan Plateau and neighboring areas inferred from multiple

759 nuclear loci. *Molecular Biology and Evolution*, 27(5), 1001-1014. doi:10.1093/molbev/msp301

760 Li, Z., Zou, J., Mao, K., Lin, K., Li, H., Liu, J., . . . Lascoux, M. (2012). Population genetic evidence for complex

761 evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan Plateau. *Evolution*, 66(3),

762 831-845. doi:10.1111/j.1558-5646.2011.01466.x

763 Lisiecki, L., & Raymo, M. (2007). Plio-Pleistocene climate evolution: Trends and transitions in glacial cycle

764 dynamics. *Quaternary Science Reviews*, 26, 56-69. doi:10.1016/j.quascirev.2006.09.005

765 Liu, B., Abbott, R. J., Lu, Z., Tian, B., & Liu, J. (2014a). Diploid hybrid origin of *Ostryopsis intermedia*

766 (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change. *Molecular Ecology*,

767 23(12), 3013-3027. doi:10.1111/mec.12783

768 Liu, J.-Q., Duan, Y.-W., Hao, G., Ge, X.-J., & Sun, H. (2014b). Evolutionary history and underlying adaptation of

769 alpine plants on the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution*, 52(3), 241-249.

770 doi:10.1111/jse.12094

771 Liu, J.-Q., Sun, Y.-S., Ge, X.-J., Gao, L.-M., & Qiu, Y.-X. (2012). Phylogeographic studies of plants in China:

772 Advances in the past and directions in the future. *Journal of Systematics and Evolution*, 50(4), 267-275.

773 doi:10.1111/j.1759-6831.2012.00214.x

774 Liu, J., Möller, M., Provan, J., Gao, L.-M., Poudel, R. C., & Li, D.-Z. (2013). Geological and ecological factors
775 drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist*, *199*(4), 1093-1108.
776 doi:10.1111/nph.12336

777 Liu, X., & Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*,
778 *47*(5), 555-559. doi:10.1038/ng.3254

779 Liu, Y. C., Sun, X., Driscoll, C., Miquelle, D. G., Xu, X., Martelli, P., . . . Luo, S. J. (2018). Genome-Wide
780 Evolutionary Analysis of Natural History and Adaptation in the World's Tigers. *Current Biology*, *28*(23),
781 3840-3849.e3846. doi:10.1016/j.cub.2018.09.019

782 Lu, X., Xu, H., Li, Z., Shang, H., Adams, R. P., & Mao, K. (2014). Genetic diversity and conservation implications
783 of four *Cupressus* species in China as revealed by microsatellite markers. *Biochemical Genetics*, *52*(3-4),
784 181-202. doi:10.1007/s10528-013-9638-1

785 Ma, Y., Wang, J., Hu, Q., Li, J., Sun, Y., Zhang, L., . . . Mao, K. (2019). Ancient introgression drives adaptation to
786 cooler and drier mountain habitats in a cypress species complex. *Commun Biol*, *2*, 213. doi:10.1038/s42003-
787 019-0445-z

788 Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279-283. doi:10.1038/nature05706

789 Marburger, S., Monahan, P., Seear, P. J., Martin, S. H., Koch, J., Paajanen, P., . . . Yant, L. (2019). Interspecific
790 introgression mediates adaptation to whole genome duplication. *Nature Communications*, *10*(1), 5218.
791 doi:10.1038/s41467-019-13159-5

792 Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: genome-
793 scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), i541-i548.
794 doi:10.1093/bioinformatics/btu462

795 Mogensen, H. L. (1996). The Hows and Whys of Cytoplasmic Inheritance in Seed Plants. *American Journal of*
796 *Botany*, *83*(3), 383-404. doi:10.2307/2446172

797 Moritz, C. (1994). Defining 'Evolutionarily Significant Units' for conservation. *Trends in Ecology & Evolution*,
798 *9*(10), 373-375. doi:10.1016/0169-5347(94)90057-4

799 Naciri, Y., & Linder, H. P. (2015). Species delimitation and relationships: the dance of the seven veils. *Taxon*, *64*(1),
800 3-16. doi:10.12705/641.24

801 Osuna-Mascaró, C., Rubio de Casas, R., & Perfectti, F. (2018). Comparative assessment shows the reliability of
802 chloroplast genome assembly using RNA-seq. *Scientific Reports*, *8*(1), 17404. doi:10.1038/s41598-018-
803 35654-3

804 Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H., & Hu, A. (2006). Simulating Arctic climate
805 warmth and icefield retreat in the last interglaciation. *Science*, *311*(5768), 1751-1753.
806 doi:10.1126/science.1120808

807 Palsbøll, P. J., Berube, M., & Allendorf, F. W. (2007). Identification of management units using population genetic
808 data. *Trends in Ecology & Evolution*, *22*(1), 11-16. doi:10.1016/j.tree.2006.09.003

809 Peng, C., Li, X.-j., Wang, H., Xie, Y., & Cao, H. (2011). The influence upon *Cupressus chengiand* space
810 Distribution by Hydroelectric Development of dod River. *Sichuan Forestry Exploration and Design*, *2011*(1),
811 11.

812 Petit, R. J., El Mousadik, A., & Pons, O. (1998). Identifying populations for conservation on the basis of genetic
813 markers. *Conservation Biology*, *12*(4), 844-855. doi:10.1046/j.1523-1739.1998.96489.x

814 Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a
815 comprehensive evaluation. *Ecography*, *31*(2), 161-175. doi:10.1111/j.0906-7590.2008.5203.x

816 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal

817 components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8),
818 904-909. doi:10.1038/ng1847

819 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Daly, M. J. (2007). PLINK:
820 a tool set for whole-genome association and population-based linkage analyses. *The American Journal of*
821 *Human Genetics*, 81(3), 559-575. doi:10.1086/519795

822 Qiu, Y.-X., Fu, C.-X., & Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions:
823 tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse
824 temperate flora. *Molecular Phylogenetics and Evolution*, 59(1), 225-244. doi:10.1016/j.ympev.2011.01.012

825 Ren, G., Mateo, R. G., Liu, J., Suchan, T., Alvarez, N., Guisan, A., . . . Salamin, N. (2017). Genetic consequences
826 of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction
827 site-associated DNA sequencing. *New Phytologist*, 213(3), 1500-1512. doi:10.1111/nph.14221

828 Rieseberg, L. H. (2019). Mapping footprints of past genetic exchange. *Science*, 366(6465), 570-571.
829 doi:10.1126/science.aaz1576

830 Rivals, E., & Salmela, L. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics*,
831 30(24), 3506-3514. doi:10.1093/bioinformatics/btu538

832 Rosemann, D., Heller, W., & Sandermann, H. (1991). Biochemical Plant Responses to Ozone. *Plant Physiology*,
833 97(4), 1280. doi:10.1104/pp.97.4.1280

834 Ru, D., Sun, Y., Wang, D., Chen, Y., Wang, T., Hu, Q., . . . Liu, J. (2018). Population genomic analysis reveals that
835 homoploid hybrid speciation can be a lengthy process. *Molecular Ecology*, 27(23), 4875-4887.
836 doi:doi:10.1111/mec.14909

837 Schoener, T. W. (1968). The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna. *Ecology*, 49(4),
838 704-726. doi:10.2307/1935534

839 Shahzad, K., Jia, Y., Chen, F.-L., Zeb, U., & Li, Z.-H. (2017). Effects of Mountain Uplift and Climatic Oscillations
840 on Phylogeography and Species Divergence in Four Endangered *Notopterygium* Herbs. *Frontiers in Plant*
841 *Science*, 8, 1929. doi:10.3389/fpls.2017.01929

842 Shang, H.-Y., Li, Z.-H., Dong, M., Adams, R. P., Miehle, G., Opgenoorth, L., & Mao, K.-S. (2015). Evolutionary
843 origin and demographic history of an ancient conifer (*Juniperus microsperma*) in the Qinghai-Tibetan Plateau.
844 *Scientific Reports*, 5, 10216. doi:10.1038/srep10216

845 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing
846 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-
847 3212. doi:10.1093/bioinformatics/btv351

848 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.
849 *Bioinformatics*, 30(9), 1312-1313. doi:10.1093/bioinformatics/btu033

850 Sun, B.-N., Wu, J.-Y., Liu, Y.-S., Ding, S.-T., Li, X.-C., Xie, S.-P., . . . Lin, Z.-C. (2011). Reconstructing Neogene
851 vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeography,*
852 *Palaeoclimatology, Palaeoecology*, 304(3-4), 328-336. doi:10.1016/j.palaeo.2010.09.023

853 Sun, H., Zhang, J., Deng, T., & Boufford, D. E. (2017). Origins and evolution of plant diversity in the Hengduan
854 Mountains, China. *Plant Diversity*, 39(4), 161-166. doi:10.1016/j.pld.2017.09.004

855 Sun, Y., Abbott, R. J., Li, L., Li, L., Zou, J., & Liu, J. (2014). Evolutionary history of Purple cone spruce (*Picea*
856 *purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Molecular*
857 *Ecology*, 23(2), 343-359. doi:10.1111/mec.12599

858 Sun, Y. S., Abbott, R. J., Lu, Z. Q., Mao, K. S., Zhang, L., Wang, X. J., . . . Liu, J. Q. (2018). Reticulate evolution
859 within a spruce (*Picea*) species complex revealed by population genomic analysis. *Evolution*, 72(12), 2669-

860 2681. doi:10.1111/evo.13624

861 Than, C., Ruths, D., & Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing
862 reticulate evolutionary relationships. *BMC Bioinformatics*, 9(1), 322. doi:10.1186/1471-2105-9-322

863 Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., . . . Su, Z. (2017). agriGO v2. 0: a GO analysis toolkit for the
864 agricultural community, 2017 update. *Nucleic Acids Research*, 45(W1), W122-W129.
865 doi:10.1093/nar/gkx382

866 Torres-Cambas, Y., Ferreira, S., Cordero-Rivera, A., & Lorenzo-Carballa, M. O. (2017). Identification of
867 evolutionarily significant units in the Cuban endemic damselfly *Hypolestes trinitatis* (Odonata:
868 Hypolestidae). *Conservation Genetics*, 18(5), 1229-1234. doi:10.1007/s10592-017-0959-1

869 Uchiyama, I., Higuchi, T., & Kawai, M. (2010). MBGD update 2010: toward a comprehensive resource for
870 exploring microbial genome diversity. *Nucleic Acids Research*, 38(suppl_1), D361-D365.
871 doi:10.1093/nar/gkp948

872 vonHoldt, B. M., Brzeski, K. E., Wilcove, D. S., & Rutledge, L. Y. (2018). Redefining the Role of Admixture and
873 Genomics in Species Conservation. *Conservation Letters*, 11(2), e12371. doi:10.1111/conl.12371

874 Walsh, C., Mac Nally, R., & Walsh, M. C. (2003). The hier. part package version 1.4-0. Retrived from [https://cran.r-](https://cran.r-project.org)
875 [project.org](https://cran.r-project.org)

876 Wang, G.-D., Zhang, B.-L., Zhou, W.-W., Li, Y.-X., Jin, J.-Q., Shao, Y., . . . Che, J. (2018b). Selection and
877 environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*. *Proceedings of*
878 *the National Academy of Sciences*, 115(22), E5056. doi:10.1073/pnas.1716257115

879 Wang, Y., Liang, Q., Hao, G., Chen, C., & Liu, J. (2018a). Population genetic analyses of the endangered alpine
880 *Sinadoxa corydalifolia* (Adoxaceae) provide insights into future conservation. *Biodiversity and Conservation*,
881 27(9), 2275-2291. doi:10.1007/s10531-018-1537-7

882 Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism:
883 quantitative approaches to niche evolution. *Evolution*, 62(11), 2868. doi:10.1111/j.1558-5646.2008.00482.x

884 Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: a toolbox for comparative studies of environmental
885 niche models. *Ecography*, 33(3), 607-611. doi:10.1111/j.1600-0587.2009.06142.x

886 Wen, J., Zhang, J.-Q., Nie, Z.-L., Zhong, Y., & Sun, H. (2014). Evolutionary diversifications of plants on the
887 Qinghai-Tibetan Plateau. *Frontiers in Genetics*, 5, 4. doi:10.3389/fgene.2014.00004

888 Witherspoon, D. J., Wooding, S., Rogers, A. R., Marchani, E. E., Watkins, W. S., Batzer, M. A., & Jorde, L. B.
889 (2007). Genetic Similarities Within and Between Human Populations. *Genetics*, 176(1), 351.
890 doi:10.1534/genetics.106.067355

891 Wu, Z. Y. (1988). Hengduan Mountain flora and her significance. *Journal of Japanese Botany*, 63(9), 297-311.

892 Xing, Y., & Ree, R. H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity
893 hotspot. *Proceedings of the National Academy of Sciences*, 114(17), E3444-E3451.
894 doi:10.1073/pnas.1616063114

895 Xu, T., Abbott, R. J., Milne, R. I., Mao, K., Du, F. K., Wu, G., . . . Liu, J. (2010). Phylogeography and allopatric
896 divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *Bmc*
897 *Evolutionary Biology*, 10(1), 194. doi:10.1186/1471-2148-10-194

898 Xu, T., Wang, Q., Olson, M. S., Li, Z., Miao, N., & Mao, K. (2017). Allopatric divergence, demographic history,
899 and conservation implications of an endangered conifer *Cupressus chengiana* in the eastern Qinghai-Tibet
900 Plateau. *Tree Genetics & Genomes*, 13(5). doi:10.1007/s11295-017-1183-3

901 Yang, Z., & Nielsen, R. (2008). Mutation-Selection Models of Codon Substitution and Their Use to Estimate
902 Selective Strengths on Codon Usage. *Molecular Biology and Evolution*, 25(3), 568-579.

903 doi:10.1093/molbev/msm284
904 Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., . . . Wang, J. (2010). Sequencing of 50
905 Human Exomes Reveals Adaptation to High Altitude. *Science*, 329(5987), 75. doi:10.1126/science.1190371
906 Yong, Y., Bing, L., & Njenga, D. M. (2017). Red list assessment and conservation status of gymnosperms from
907 China. *Biodiversity Science*, 25(7), 758-764. doi:10.17520/biods.2017145
908 Zeng, P. A., & Yang, Q., Z. (1992). *Minjiang cypress forest*. Beijing: China Forestry Press.
909 Zeng, X., Yuan, H., Dong, X., Peng, M., Jing, X., Xu, Q., . . . Gao, M. (2020). Genome-wide Dissection of Co-
910 selected UV-B Responsive Pathways in the UV-B Adaptation of Qingke. *Molecular Plant*, 13(1), 112-127.
911 doi:10.1016/j.molp.2019.10.009
912 Zhang, D., & Christian, T. (2013). *Cupressus chengiana*. The IUCN Red List of Threatened Species 2013.
913 Retrieved from <http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T42217A2962398.en>
914 Zheng, B., Xu, Q., & Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on
915 the Qinghai-Tibetan Plateau: review and speculation. *Quaternary International*, 97-98(1), 93-101.
916 Zheng, H., Fan, L., Milne, R. I., Zhang, L., Wang, Y., & Mao, K. (2017). Species Delimitation and Lineage
917 Separation History of a Species Complex of Aspens in China. *Frontiers in Plant Science*, 8, 375.
918 doi:10.3389/fpls.2017.00375
919 Zhang, L., Su, W., Tao, R., Zhang, W., Chen, J., Wu, P., . . . Kuang, H. (2017). RNA sequencing provides insights
920 into the evolution of lettuce and the regulation of flavonoid biosynthesis. *Nature Communications*, 8(1), 2264.
921 doi:10.1038/s41467-017-02445-9
922 Zhou, S., & Li, J. (1998). The sequence of Quaternary glaciation in the Bayan Har Mountains. *Quaternary*
923 *International*, 45-46(1), 135-142. doi:10.1016/S1040-6182(97)00011-6
924

925 **Data Accessibility Statement**

926 The transcriptome sequencing data have been deposited in NCBI with the BioProject ID:
927 PRJNA556937 that is publicly accessible at
928 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA556937>. The filtered SNP matrices file is
929 available in Dryad at <https://doi.org/10.5061/dryad.70rxwdbtf>. The custom scripts have been
930 made available in Supplemental Information (Supplementary Text: Python scripts).
931

932 **Author Contributions**

933 K.M. and Jianquan Liu designed and supervised this study. Jialiang Li, J.M., W.T., L.Z., D.R.
934 and J.X. managed fieldwork and collected the materials. Jialiang Li and D.R. analyzed the data.
935 Jialiang Li, R.M. and K.M. wrote the manuscript. Jianquan Liu revised the manuscript.
936
937

938 **Tables and Figures**939 **Tables**

940

941 **Table 1** Location information of the sampled *Cupressus chengiana* populations (GS =
 942 Gansu province, SC = Sichuan province, N = number of individuals analyzed).

Group	Code	N	Location	Latitude (N)	Longitude (E)	Altitude (m)
BLJ	ZR-17	8	Zhouqu, GS	33°52.908'	104°08.156'	1521
	ZR-18	7	Wudu, GS	33°12.030'	105°02.130'	1025
	ZR-20	8	Longnan, GS	33°15.197'	104°59.024'	1634
	ZR-21	7	Wenxian, GS	32°49.364'	104°45.576'	1535
	LXT-18	5	Jiuzhaigou, SC	33°06.946'	104°19.491'	1351
MJR	ZR-01	5	Lixian, SC	31°40.420'	103°49.448'	1500-1938
	ZR-05	7	Lixian, SC	31°23.480'	103°03.680'	2106
	LXT-16	5	Maoxian, CS	31°38.392'	103°48.350'	1742
DDH	ZR-27	7	Jinchuan, SC	31°54.630'	102°01.797'	2410
	ZR-28	6	Maerkang, SC	31°47.460'	101°56.480'	2470
	ZR-30	7	Xiaojin, SC	31°09.790'	102°26.614'	2571
	LXT-05	5	Xiaojin, SC	31°01.749'	102°15.016'	2252
	LXT-08	5	Danba, SC	30°42.954'	101°59.692'	2211

943

944

945 **Table 2** Summary of genomic polymorphisms and variants in different *C. chengiana* groups

Parameters		Groups		
		BLJ	MJR	DDH
SNPs	266,884	231,513	228,734	224,658
Private SNPs	-	11,489	3,213	11,913
π	0.0077	0.0069	0.0072	0.0064
Tajima's D	-0.4310	-0.1790	0.0470	-0.1449
		B vs. D	M vs. B	M vs. D
Shared SNPs	-	14,748	27,524	20,245
F_{ST}	-	0.1752	0.1066	0.1397
D_{XY}	-	0.3440	0.3008	0.3104

946 The π and Tajima's D were calculated using the data set without MAF filtering, and the other parameters
 947 were calculated based on the data set after MAF control.

948

949

950

951 **Table 3** Inferred demographic parameters for the best-fitting FSC2 model shown in Figure 3b,
 952 including 95% confidence intervals.

Parameters	Point estimation	95% confidence intervals	
		Lower bound	Upper bound
N_A	323,866	310,722	330,971
N_1	238,794	227,162	248,078
N_2	114,433	97,828	131,628
N_3	166,952	155,815	172,291
T1	1,344,300	1,140,450	1,408,600
T2	4,559,100	4,225,050	4,869,500
α	0.62	0.54	0.65
$M_{1\leftarrow 2}$	9.22E-06	4.95E-06	1.22E-05
$M_{2\leftarrow 1}$	3.65E-06	1.37E-06	5.55E-06
$M_{2\leftarrow 3}$	4.94E-06	3.17E-06	6.44E-06
$M_{3\leftarrow 2}$	1.00E-05	6.36E-06	1.19E-05
$M_{1\leftarrow 3}$	9.83E-07	3.29E-07	1.71E-06
$M_{3\leftarrow 1}$	1.70E-06	6.83E-07	2.49E-06
$MA_{3\leftarrow 1}$	2.95E-07	1.58E-07	1.43E-06
$MA_{1\leftarrow 3}$	5.72E-07	1.35E-07	1.35E-06

953 Parameters included here comprise population size measures (N_A , N_1 , N_2 and N_3 , indicating ancestral
 954 population, BLJ, MJR and DDH, respectively), population divergence time (T2, years) and hybrid origin
 955 time (T1, years), hybrid parameter (α), migration per generation after hybridization (M) between each pair
 956 of ESUs in each direction, and migration per generation before hybridization (MA) between BLJ and DDH.

957
 958

959
960

Table 4 Results of the ABBA-BABA test. Patterson's *D* value for introgression between lineages with *Z* score and significance values were shown.

P ₁	P ₂	P ₃	Patterson' <i>D</i>	<i>Z</i> score	<i>p</i> ⁹⁶¹
BLJ	MJR	DDH	0.0780	20.2321	0
DDH	MJR	BLJ	0.1222	28.5418	0

962 **Figure Legends**

963

964 **Figure 1** Geographic distributions of sampled *Cupressus chengiana* populations. Those
965 individuals in the BLJ, MJR and DDH group are distributed in the upper reaches of the
966 Bailongjiang River, the Minjiang River and the Daduhe River, respectively.

967 **Figure 2** Genetic structure and Phylogenetic relationships of the three *C. chengiana* groups
968 (BLJ, MJR, DDH). (a) Admixture proportions of genetic clusters for each individual of the three
969 groups. The scenarios of $K=2$ and $K=3$ are shown, and $K=3$ is the best value according to cross-
970 validation analysis. (b) Principal component analysis (PCA) plot for the 82 *C. chengiana*
971 individuals based on the first two principal components. (c) An ML tree based on 31,527 SNPs
972 in 4DTV of nuclear genome, with three distinct lineages (BLJ, MJR, DDH) detected, among
973 which the relationship between the MJR and BLJ groups is the closest. (d) An ML tree based
974 on 1,251 SNPs of the chloroplast genome, in which the MJR group was not distinguished from
975 the DDH group, while all individuals in the BLJ group form a separate lineage. The supporting
976 values from bootstrap analyses are labeled beside the nodes. Group information is shown in
977 Table 1 and Figure 1.

978 **Figure 3** (a) The detailed population demographic history of BLJ, MJR and DDH over the last
979 10 million years inferred by Stairway Plot method. Thick lines represent the median, and thin
980 light lines represent the 95% pseudo-CI defined by the 2.5% and 97.5% estimations from the
981 SFS analysis. The periods of the Xixiabangma Glaciation and the Naynayxungla Glaciation are
982 highlighted in gray vertical bars. (b) Maximum likelihood parameter estimates of the best fit
983 models (model3) in FSC2. (c) An ML-bootstrap network for 10,227 orthologous gene trees
984 yielded in PhyloNet with a maximum of two reticulations allowed. The light blue curves
985 represent reticulations with inheritance probabilities behind them. (d) Results of Bayesian
986 outlier analysis for 266,884 SNPs. SNPs with $q\text{-value} < 0.001$ were recognized as outliers. A
987 total of 575 positive outlier SNPs were identified in this analysis.

988 **Figure 4** ENMs for three *C. chengiana* lineages, and identity tests results between paired groups.
989 (a) Current potential distributions of BLJ, MJR and DDH groups, predicted by Maxent. (b)
990 Results of identity tests of three comparisons (BLJ vs. DDH, MJR vs. DDH, MJR vs. BLJ).
991 The grey bars indicate the null distributions of D, while the black bars indicate I. Arrows
992 indicate values of D (gray) and I (black) in actual Maxent runs.

993

994

995

996

997

998