




# COMMUNICATIONS BIOLOGY

## ARTICLE

<https://doi.org/10.1038/s42003-019-0445-z>

OPEN

## Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex

Yazhen Ma<sup>1,4</sup>, Ji Wang<sup>1,4</sup>, Quanjun Hu<sup>1,4</sup>, Jialiang Li<sup>1</sup>, Yongshuai Sun<sup>2</sup>, Lei Zhang<sup>1</sup>, Richard J. Abbott<sup>3</sup>, Jianquan Liu<sup>1</sup> & Kangshan Mao<sup>1</sup> 

Introgression may act as an important source of new genetic variation to facilitate the adaptation of organisms to new environments, yet how introgression might enable tree species to adapt to higher latitudes and elevations remains unclear. Applying whole-transcriptome sequencing and population genetic analyses, we present an example of ancient introgression from a cypress species (*Cupressus gigantea*) that occurs at higher latitude and elevation on the Qinghai-Tibet Plateau into a related species (*C. duclouxiana*), which has likely aided the latter species to extend its range by colonizing cooler and drier mountain habitats during postglacial periods. We show that 16 introgressed candidate adaptive loci could have played pivotal roles in response to diverse stresses experienced in a high-elevation environment. Our findings provide new insights into the evolutionary history of Qinghai-Tibet Plateau plants and the importance of introgression in the adaptation of species to climate change.

<sup>1</sup>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, 610065 Chengdu, Sichuan, P. R. China. <sup>2</sup>Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, 666303 Mengla, P. R. China. <sup>3</sup>School of Biology, Mitchell Building, University of St Andrews, St Andrews, Fife KY16 9TH, UK. <sup>4</sup>These authors contributed equally: Yazhen Ma, Ji Wang, Quanjun Hu. Correspondence and requests for materials should be addressed to J.L. (email: [liujq@nwipb.cas.cn](mailto:liujq@nwipb.cas.cn)) or to K.M. (email: [maokangshan@163.com](mailto:maokangshan@163.com))

Rapid climate change is poised this century to be one of the greatest global challenges to the stability and cohesion of ecosystems<sup>1,2</sup>. It is argued that to survive such change, terrestrial species might migrate across land by tracking favorable conditions or persist in situ either through adaptive phenotypic plasticity<sup>3</sup> and/or genetic modification caused by natural selection favoring advantageous standing genetic variation and mutations<sup>1,4</sup>. A further, but until recently, overlooked potential source of genetic variation that can aid adaptation to climate change is introgression from closely related species<sup>5,6</sup>.

Recent studies on humans, dogs, cattle, sunflowers, poplars, and oaks suggest that introgression from other species can be a very important source of new genetic variation in the adaptation of species to new environments<sup>7–12</sup>. Such adaptive introgression resulting from hybridization and backcrossing with another species<sup>13–15</sup> has the potential to allow adaptation in recipient populations at rates considerably higher than if mutation was the only source of variation<sup>16–18</sup>. Though adaptive introgression has long been considered an important evolutionary mechanism<sup>13</sup>, with notable examples reported in both plant and animal genera<sup>10,18–22</sup>, most studies of adaptive introgression in plants using a genomic approach have focused on short generation species (but see refs. <sup>11,23</sup>). A major challenge to detecting introgression between species with long generation times and large effective population sizes, such as trees, is to distinguish between the causes of shared polymorphisms due to incomplete lineage sorting or introgression<sup>24,25</sup>. Notwithstanding this difficulty, a growing number of studies have shown that introgression occurs in at least some tree genera (reviewed in refs. <sup>5,26,27</sup>), particularly from local tree species into widespread or invasive ones<sup>28</sup>, involving genomic fragments exhibiting low rather than high intraspecific gene flow<sup>29</sup>. How commonly such introgression is adaptive, however, remains a topic of active research.

Adaptive genetic variation is common within tree species<sup>30,31</sup> with clines reflecting local adaptation across environmental gradients<sup>1,31</sup> and ecotypic variation is often recorded<sup>4</sup>. Moreover, candidate loci associated with adaptive changes across environmental gradients are increasingly implicated in studies of adaptive divergence in trees<sup>31–34</sup>. Although evidence of adaptive introgression in trees is accumulating (e.g., refs. <sup>35,36</sup>), only recently has detailed evidence of it emerged from genomic analyses in model tree genera such as poplars<sup>11,23,37</sup> and oaks<sup>38</sup>. In contrast, little is known of it occurring in conifers, partly due to their huge genome size<sup>39</sup> which makes reference genome assembly difficult. Moreover, very few studies have focused on introgression across elevation gradients (but see ref. <sup>27</sup>), especially in topographically complex regions, such as the Qinghai-Tibet Plateau and adjacent regions in Asia (but see ref. <sup>40</sup>). Here, we examine the possible role of introgressive hybridization in the adaptation of a cypress (*Cupressus*) species to higher elevations and latitudes in the Qinghai-Tibet Plateau and neighboring areas.

*Cupressus gigantea*, which occurs along the Yarlung Tsangpo river valley in Tibet (Xizang), grows at higher altitudes (2950–3430 m.a.s.l.) relative to other species in the genus<sup>41</sup>, whereas *C. duclouxiana*, occupies a wider altitude range (1400–3300 m.a.s.l.), and is mainly distributed in central and northwest Yunnan, and southwest Sichuan, China<sup>42</sup>. Previous work has indicated a close phylogenetic relationship between these two species based on cpDNA variation<sup>43</sup>, and an absence of gene flow between them based on nuclear microsatellite markers<sup>44</sup>. Importantly, populations of *C. duclouxiana* in the northern and southern parts of its range are genetically divergent, forming two separate groups based on microsatellite variation<sup>44</sup>. The northern populations tend to occur at elevations and latitudes between those of southern *C. duclouxiana* and *C. gigantea* populations, although their habitat is similar to that of

*C. gigantea*. Thus, both *C. gigantea* and northern populations of *C. duclouxiana* occur in mountain valleys that are cooler and drier than the habitat of southern populations of *C. duclouxiana*. The distribution of *C. duclouxiana* is also notably more strongly influenced by human activities than *C. gigantea*, with relict populations found around temples and in remote areas. Here, we tested whether adaptive introgression from *C. gigantea* to northern *C. duclouxiana* enabled the latter to colonize mountain habitats.

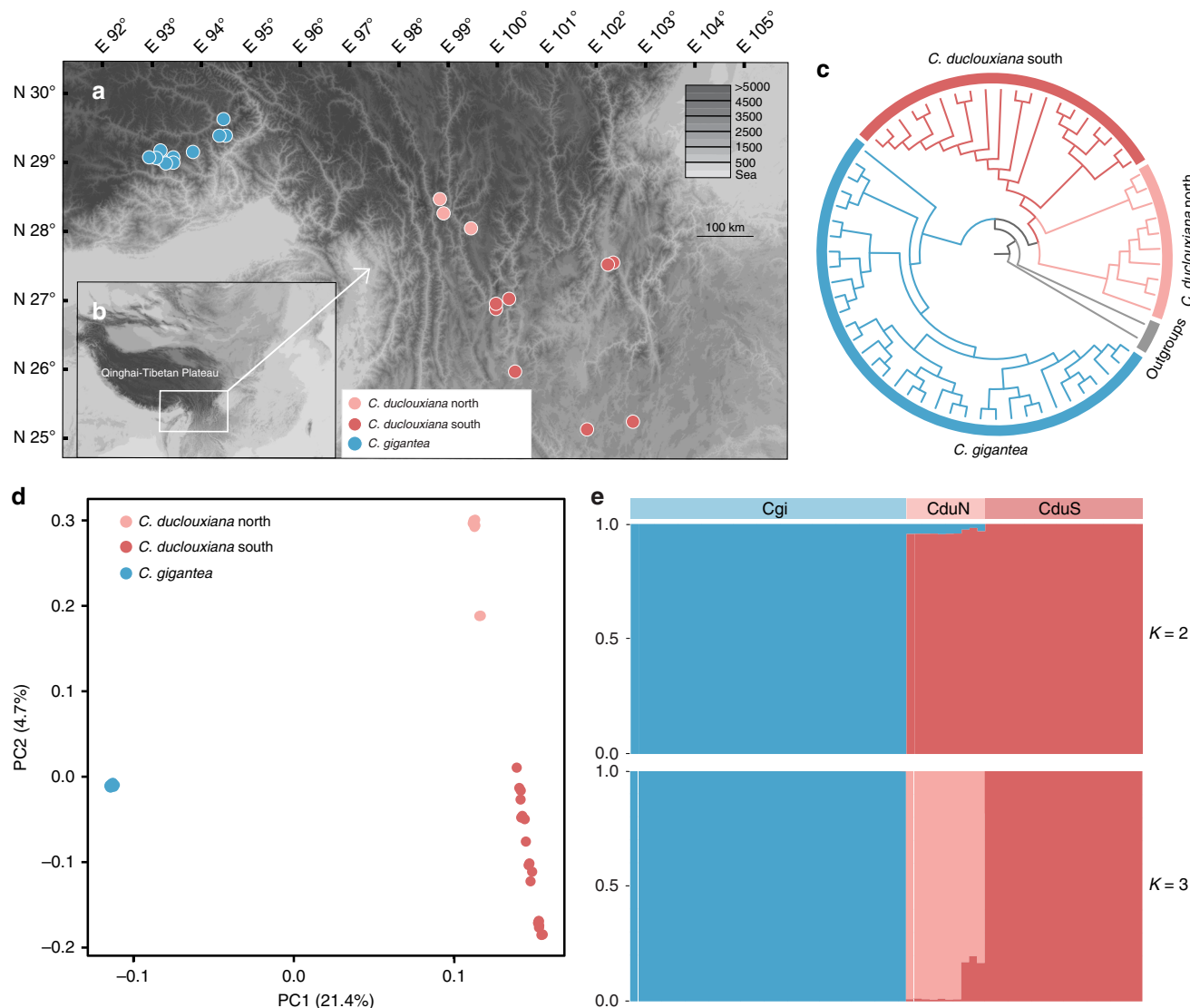
We analysed RNA-sequences from leaf samples representing both *Cupressus* species to address the following questions: What is the genetic affinity of northern to southern populations of *C. duclouxiana* and to populations of *C. gigantea*? What is the level and direction of gene flow between the different genetic groups detected? What role might adaptive introgression have played in shaping the distribution pattern of genetic variation in this complex, enabling *C. duclouxiana* to grow over a wide range of elevations? Our results suggest that both *C. duclouxiana* and *C. gigantea* are monophyletic; however, genetic introgression from *C. gigantea* to northern *C. duclouxiana* is supported by multiple lines of evidence. We detected 1285 loci likely to be introgressed from the former to the latter, with 16 of these identified as candidate adaptive loci that might play pivotal roles in response to diverse stresses in high-elevation environments. These loci were positively selected in *C. gigantea* and northern *C. duclouxiana*, but not in southern *C. duclouxiana*. This introgression most likely occurred during the last glacial maximum (ca. 0.022 million years ago) or during earlier glacial periods, and contributed to the adaptation of northern *C. duclouxiana* to cooler and drier conditions in mountain habitats.

## Results

**Habitat differentiation.** Leaf samples were collected from a total of 65 individuals comprising 30 *Cupressus duclouxiana* trees (from 11 populations) and 35 *C. gigantea* trees (from 12 populations) (Fig. 1a, b; Supplementary Table 1). A scatter plot of the latitudes and elevations of populations shows that northern populations of *C. duclouxiana* tend to occur at elevations and latitudes between those of southern *C. duclouxiana* and *C. gigantea* populations (Supplementary Fig. 1a). Principal component analysis of climate data (obtained from climate stations) further indicates that northern populations of *C. duclouxiana* occur in habitats similar to those of *C. gigantea* that are very different from those of southern populations (Supplementary Fig. 1b), with mean annual precipitation as the major contributor (Supplementary Fig. 1c). Climate records show that both *C. gigantea* and northern *C. duclouxiana* occur in mountain valleys with mean annual precipitation between 640 and 702 mm, and mean annual temperature between 5.85 and 9.1 °C, while southern *C. duclouxiana* occurs in areas with mean annual precipitation between 936 and 1007 mm (with one outlier as 838 mm) and mean annual temperature between 12.6 and 20.8 °C (Supplementary Tables 2, 3).

## Reference transcriptome assembly and DNA polymorphism.

The assembled reference transcriptome of *C. duclouxiana* contained 85,401 unigenes with an average length of 676 bp and a contig N50 equaling 1027 bp, and 28,414 of these unigenes were annotated with reference to the NCBI (National Center for Biotechnology Information) nr (non-redundant) database. After aligning transcriptome reads from all individuals to the reference transcriptome and undertaking stringent quality filtering, we identified a total of 878,614 single nucleotide polymorphisms (SNPs) across all 65 individuals of *C. duclouxiana* and *C. gigantea* examined. For 513,615 of these SNPs, covering 16,884 unigenes, there were no missing data across all samples.



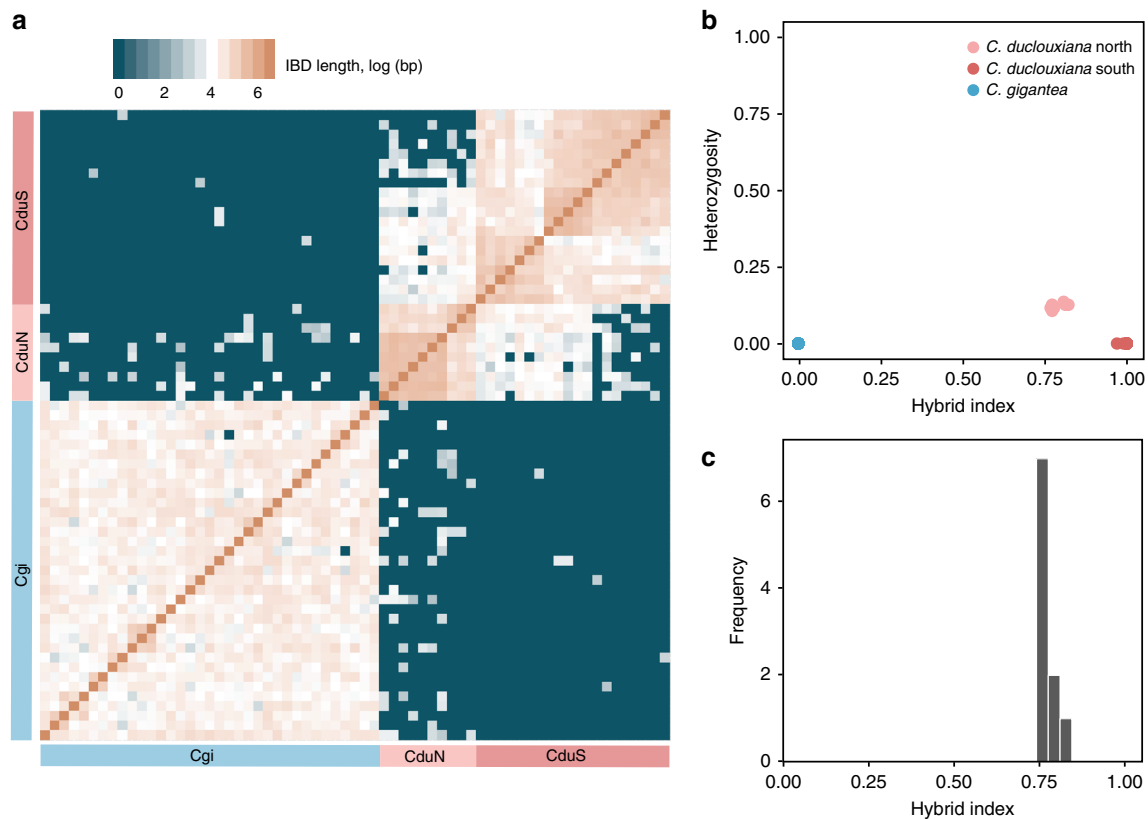
**Fig. 1** Phylogenetic and population genetic analyses of *C. gigantea* and *C. duclouxiana* based on SNP variation. **a** Map showing the geographic distribution of sampling locations for *C. gigantea* and *C. duclouxiana* populations, with **b** the location of the Qinghai-Tibet Plateau. **c** A maximum likelihood phylogenetic tree. **d** Principal component analysis (PCA) plot of the first two components. **e** Population structure plots with  $K = 2$  and  $3$ . The x-axis shows the different individuals of *C. gigantea* (Cgi) and of *C. duclouxiana* northern (CduN) and southern (CduS) populations; the y-axis quantifies the proportion of an individual's variation from inferred ancestral populations

The transcriptome-wide average values for both indicators of nucleotide diversity,  $\theta_{\pi}$  and  $\theta_w$ , were higher for *C. duclouxiana* (0.0031, 0.0151) than for *C. gigantea* (0.0029, 0.0144; Supplementary Table 4).

**Population structure and polymorphism sharing.** Phylogenetic analysis of SNP variation showed that samples of *C. gigantea* and *C. duclouxiana* were clustered into separate monophyletic clades (Fig. 1c). Principal component analysis (PCA) based on SNPs also distinguished the two species along PC1 (variance explained = 21.41%, Tracy-Widom  $P = 5.99 \times 10^{-10}$ ) and further separated northern from southern populations of *C. duclouxiana* along PC2 (variance explained = 4.72%, Tracy-Widom  $P = 8.14 \times 10^{-41}$ ) (Fig. 1d). Analysis of the genetic structure of *C. duclouxiana* and *C. gigantea* using Frappe<sup>45</sup> assigned most individuals to one or other of two species-specific groups with  $K = 2$ . However, all individuals representing northern *C. duclouxiana* were of mixed ancestry with a minority of their ancestry derived from *C. gigantea* (Fig. 1e). Some of these individuals were assigned to a

third genetic group when  $K = 3$  with the remainder indicated to be of mixed ancestry between this third group and the group comprising all southern *C. duclouxiana* individuals (Fig. 1e). When  $K = 4$ , both *C. gigantea* and *C. duclouxiana* were each divided into two subgroups with some southern *C. duclouxiana* individuals showing mixed ancestry between northern *C. duclouxiana* and the remainder of southern *C. duclouxiana* (Supplementary Fig. 2).

The relationships between *C. gigantea* and the northern and southern types of *C. duclouxiana* were further explored by a refined identical-by-descent (IBD) approach<sup>46</sup>. Pairwise comparisons between individuals showed that haplotype sharing was much more common between *C. gigantea* and northern *C. duclouxiana* (50 of 350 pairwise comparisons indicated sharing) than between *C. gigantea* and southern *C. duclouxiana* (only 6 of 700 pairwise comparisons indicated haplotype sharing) (Fig. 2a). As expected, haplotype sharing was most common between northern and southern *C. duclouxiana*, with 143 of 200 pairwise comparisons indicating sharing. In all cases, shared haplotypes



**Fig. 2** Inference of gene flow between *C. gigantea* and *C. duclouxiana* based on haplotype sharing and hybrid index estimation. **a** Estimated haplotype sharing between individuals of *C. gigantea* (Cgi) and *C. duclouxiana* (CduN: northern population; CduS: southern population). Heatmap colors represent the total length of IBD blocks for each pairwise comparison. **b** Scatter plot of hybrid index (x-axis) and heterozygosity (y-axis) for 65 individuals of *C. gigantea*, northern *C. duclouxiana* and southern *C. duclouxiana*. Note: there is considerable overlap of dots representing different individuals within each lineage. **c** A histogram showing frequency of hybrid indices for individuals of northern *C. duclouxiana*

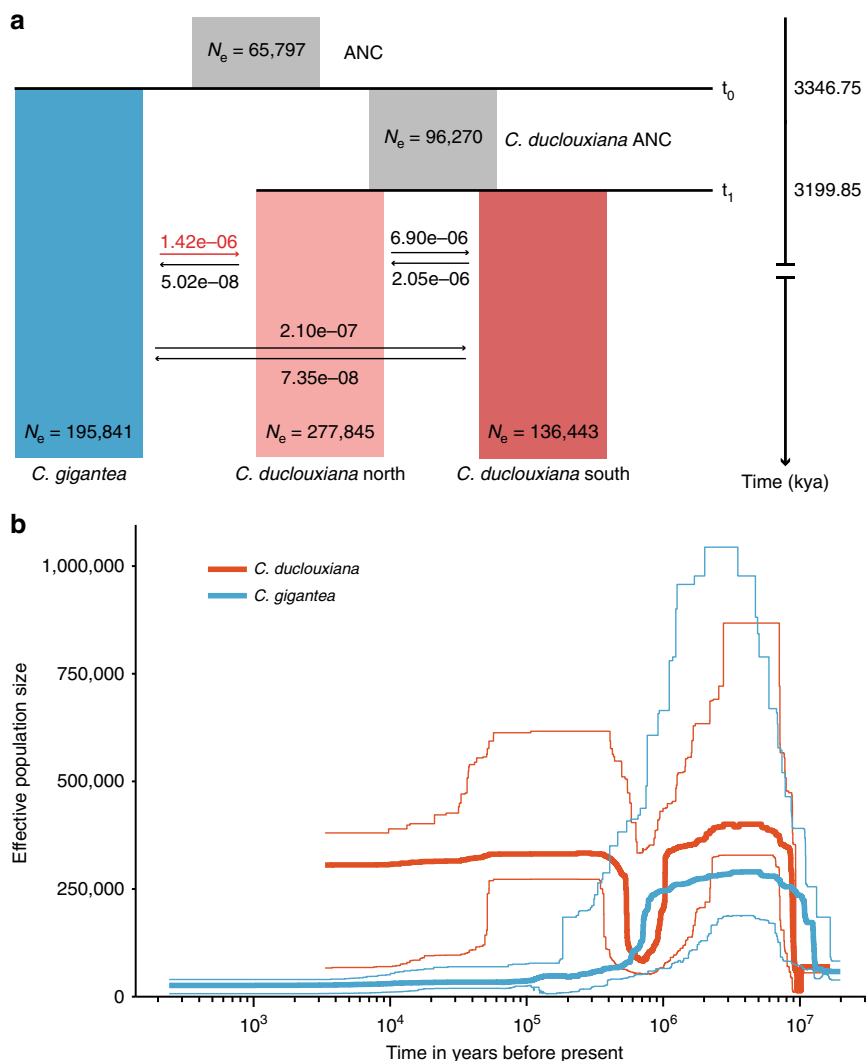
were short in length, due to the relatively short length of unigenes in the transcriptome dataset, yet were of sufficient length to reveal intra- and inter-specific gene flow. Finally, a Hlest analysis<sup>47</sup> estimated that northern *C. duclouxiana* individuals had heterozygosity indices between 0.11 and 0.14 (Fig. 2b) and hybrid indices between 0.76 and 0.815 (Fig. 2c), indicating they are advanced generation backcrosses between *C. gigantea* and *C. duclouxiana*. In summary, these results, together with the finding that both  $F_{ST}$  and  $d_{XY}$  were slightly lower between *C. gigantea* and northern *C. duclouxiana* than between *C. gigantea* and southern *C. duclouxiana* (Supplementary Table 4), strongly indicate that historical gene flow has occurred between *C. gigantea* and *C. duclouxiana* resulting in northern populations of *C. duclouxiana* being composed entirely of advanced generation backcrosses.

**Gene flow and population demography.** Levels of gene flow among the three genetic groups identified by phylogenetic and population structure analyses were estimated using coalescent-based simulations<sup>48</sup>. Of the 16 candidate migration models compared (Supplementary Fig. 3), the best-fitting one (with maximum Akaike's weight value, Supplementary Table 5) contained six of the eight possible migration parameters (Fig. 3a). This indicated that *C. gigantea* and *C. duclouxiana* diverged from their common ancestor approximately 3.35 million years ago (Mya), while the northern and southern types of *C. duclouxiana* diverged approximately 3.2 Mya (Fig. 3a; Table 1). Estimated gene flow from *C. gigantea* to the northern type of *C. duclouxiana* was much higher than in the reverse direction or between *C. gigantea*

and southern *C. duclouxiana* in either direction, but lower than migration between northern and southern *C. duclouxiana*. The best-fitting model further indicated an absence of historical gene flow between *C. gigantea* and the *C. duclouxiana* ancestral population. A Stairway Plot analysis<sup>49</sup> showed that population size ( $N_e$ ) declined in both species until approximately 0.6–1.0 Mya (Fig. 3b). It then expanded rapidly in *C. duclouxiana*, but declined further in *C. gigantea* (Fig. 3b).

Analysis of rare SNPs shared among populations indicated that the proportion of rare alleles shared between northern *C. duclouxiana* and *C. gigantea* was much higher than between southern *C. duclouxiana* and *C. gigantea* (Fig. 4). This suggests that though recent introgression has occurred between *C. gigantea* and both types of *C. duclouxiana* it has been much greater between *C. gigantea* and northern *C. duclouxiana*. The results also indicated that the highest level of recent introgression was between northern and southern *C. duclouxiana* populations (Fig. 4).

**Introgressed loci and positively selected genes.** To identify loci introgressed between *C. gigantea* and northern *C. duclouxiana*, we calculated the ABBA-BABA statistic<sup>50,51</sup> and the modified  $f$  statistic ( $f_{dM}$ )<sup>52,53</sup> for all unigenes detected across the whole transcriptome. First, we identified candidate signals of introgression only if the  $D$ -statistic was greater than or equal to 0.7. Next, we extracted those with averages (over 200 bp windows) of both  $D$  and  $f_{dM}$  significantly different from zero ( $P < 0.01$ ). Because introgressed regions generally show lower absolute genetic divergence, we further calculated sequence divergence ( $d_{XY}$ ) for each candidate introgressed locus and compared this with the whole

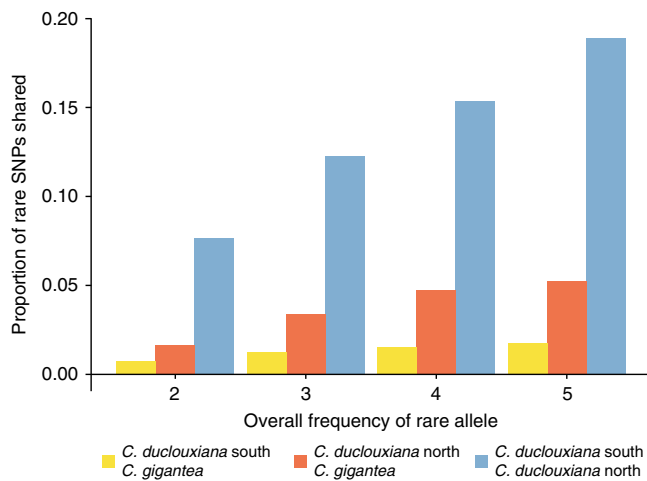


**Fig. 3** Demographic history of *C. gigantea* and *C. duclouxiana*. **a** Schematic of demographic scenario modeled in *fastsimcoal2*. Estimated effective population sizes ( $N_e$ ) and divergence times are indicated. The numbers next to arrows denote the per generation migration rate between populations. **b** Changes in effective population size ( $N_e$ ) over time in *C. gigantea* and *C. duclouxiana* inferred by Stairway Plot method. Thick lines represent the median and thin light lines the 95% pseudo-CI defined by the 2.5% and 97.5% estimations of the SFS analysis

**Table 1** Inferred demographic parameters for the best-fitting demographic model shown in Fig. 3a

Parameters	Point estimation	95% confidence intervals	
		Lower bound	Upper bound
$N_{e-ANC}$	65,797	55,016	98,062
$N_{e-C. duclouxiana ANC}$	96,270	81,860	151,345
$N_{e-C. gigantea}$	195,841	170,091	207,325
$N_{e-Northern C. duclouxiana}$	277,845	220,670	302,080
$N_{e-Southern C. duclouxiana}$	136,443	114,886	173,003
$m_{C. gigantea \rightarrow northern C. duclouxiana}$	$1.42 \times 10^{-06}$	$1.51 \times 10^{-07}$	$1.75 \times 10^{-06}$
$m_{Northern C. duclouxiana \rightarrow C. gigantea}$	$5.02 \times 10^{-08}$	$3.14 \times 10^{-10}$	$4.91 \times 10^{-07}$
$m_{C. gigantea \rightarrow southern C. duclouxiana}$	$2.10 \times 10^{-07}$	$1.24 \times 10^{-09}$	$2.15 \times 10^{-07}$
$m_{Southern C. duclouxiana \rightarrow C. gigantea}$	$7.35 \times 10^{-08}$	$5.25 \times 10^{-10}$	$2.57 \times 10^{-07}$
$m_{Northern C. duclouxiana \rightarrow southern C. duclouxiana}$	$6.90 \times 10^{-06}$	$3.44 \times 10^{-06}$	$9.60 \times 10^{-06}$
$m_{Southern C. duclouxiana \rightarrow northern C. duclouxiana}$	$2.05 \times 10^{-06}$	$1.11 \times 10^{-06}$	$6.57 \times 10^{-06}$
$t_0$	3,346,750	2,951,850	3,542,350
$t_1$	3,199,850	2,771,150	3,338,300

$N_e$ : the effective population sizes of each population;  $t$ : estimated divergence times;  $m$ : the per generation migration rate between populations



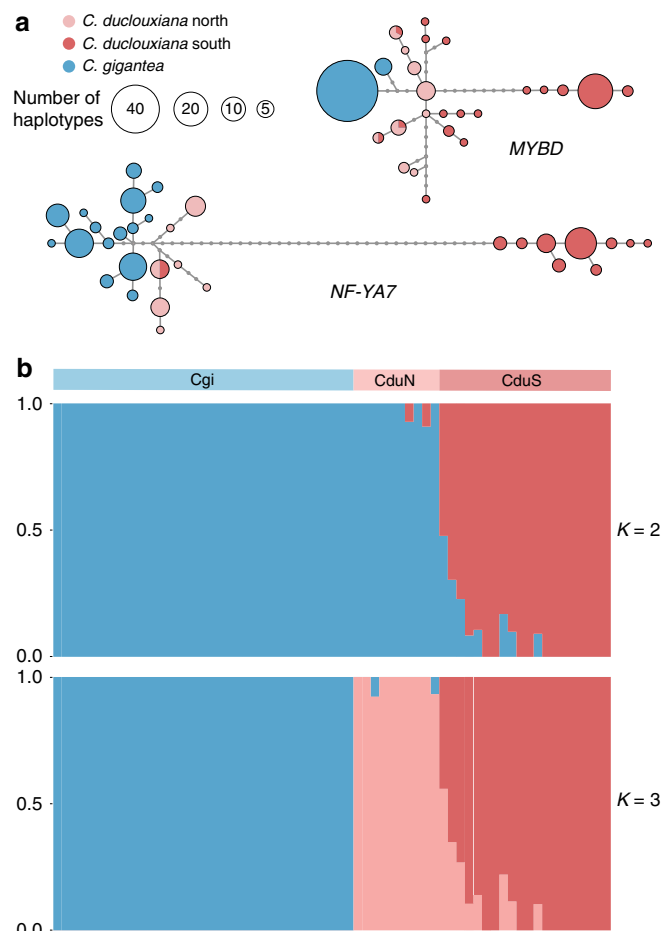
**Fig. 4** Sharing of rare SNPs among *C. gigantea*, northern *C. duclouxiana*, and southern *C. duclouxiana*. The proportion of all rare SNPs that are only shared between two of the three populations is plotted in each frequency category

transcriptome mean  $d_{XY}$ . In combination, these tests detected a subset of 1285 candidate loci introgressed from *C. gigantea* into northern *C. duclouxiana*. We performed Gene Ontology (GO) enrichment analysis for these introgressed loci and identified 65 enriched GO terms (corrected  $P < 0.05$ ), including response to salt stress, response to extracellular stimulus, detoxification, and response to metal ion (Supplementary Table 6).

To further explore the genetic basis of possible high elevation adaptation, we performed population branch statistic<sup>54</sup> and Hudson-Kreitman-Aguadé<sup>55</sup> tests to identify genes under positive selection within *C. gigantea* and northern *C. duclouxiana*. In total, 730 and 248 positively selected genes were identified within *C. gigantea* and northern *C. duclouxiana*, respectively. GO enrichment analyses identified 46 overrepresented GO terms for positively selected genes within *C. gigantea*, including photosynthesis, light harvesting, regulation of ion transport, and response to ethylene (Supplementary Table 7). For positively selected genes within northern *C. duclouxiana*, 26 GO categories were enriched with some associated with regulation of phospho-protein phosphatase activity, IMP biosynthetic process, and regulation of proteolysis (Supplementary Table 8).

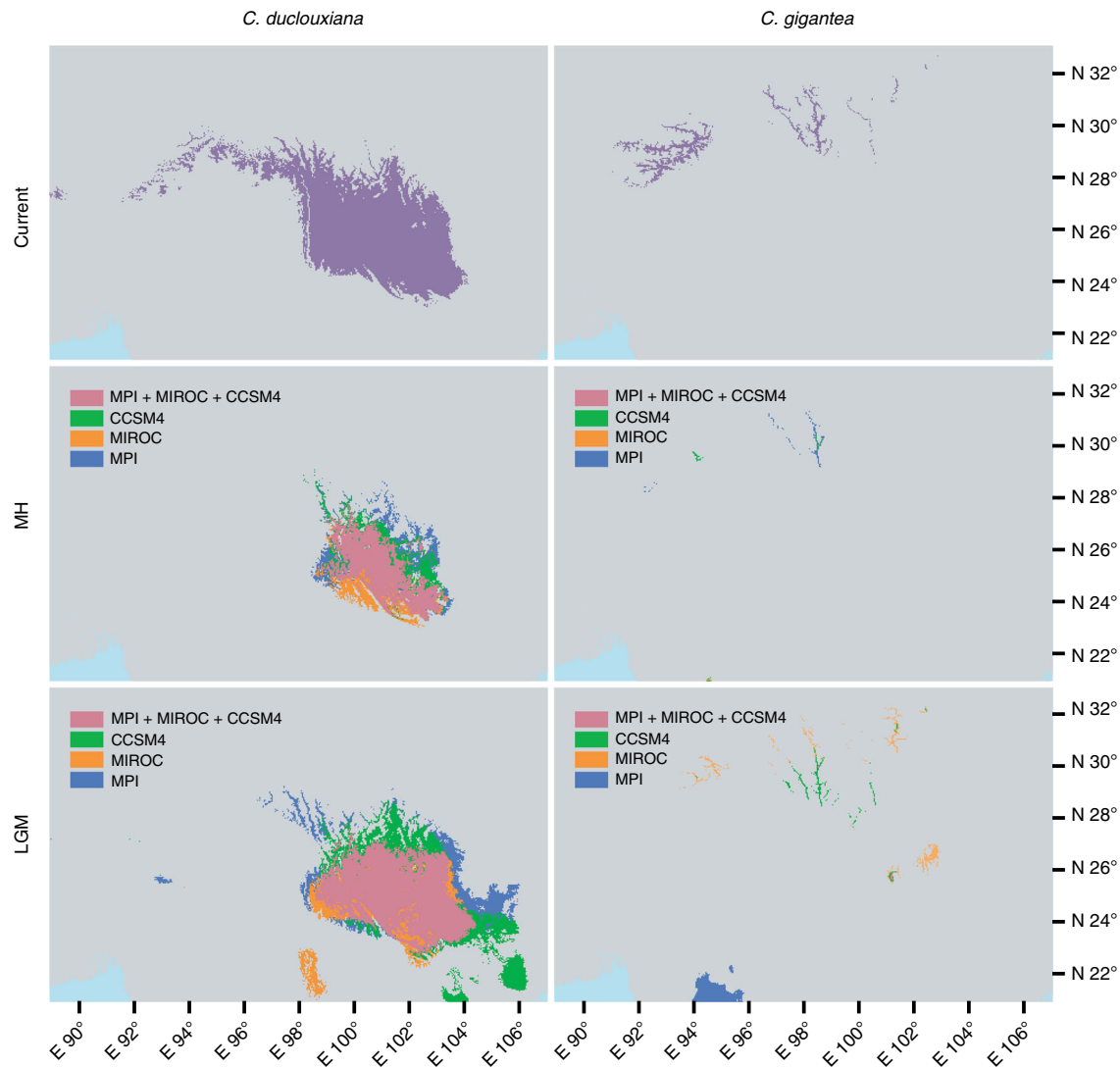
Sixteen introgressed loci shown to be positively selected, were shared between *C. gigantea* and northern *C. duclouxiana*. To test whether the number of these genes was higher than expected by chance, we randomly sampled 1285, 730, and 248 samples from an available pool of 16,884 samples (without replacement), respectively, and a maximum of 7 overlaps among three samplings were observed after 100,000 replications. These 16 loci were, therefore, considered as introgressed genes likely to be important to the adaptation of northern *C. duclouxiana* to its local environment at high latitudes and elevations on the Qinghai-Tibet Plateau (Fig. 5a; Supplementary Table 9). As a result of the sharing of these 16 positively selected genes between *C. gigantea* and northern *C. duclouxiana*, population structure analyses limited to the 881 SNPs located within these genes (Fig. 5b) revealed a very different picture of population structure to that produced from the analysis of the entire set of 513,615 SNPs (Fig. 1e). Thus, with  $K=2$ , northern *C. duclouxiana* individuals formed one genetic group with *C. gigantea*, while all southern *C. duclouxiana* individuals were placed in the second group (Fig. 5b).

**Distribution models for *C. duclouxiana* and *C. gigantea*.** Jackknife tests of ecological niche models showed that temperature



**Fig. 5** Genetic introgression from *C. gigantea* to *C. duclouxiana*. **a** Haplotype trees for two of 16 introgressed candidate adaptive loci, MYBD and NF-YA7, generated by Haplotype Viewer (<http://www.cibiv.at/~greg/haploviewer>) indicate their introgression from *C. gigantea* into *C. duclouxiana*. **b** Population structure plots with  $K=2$  and  $K=3$  based on SNPs within 16 introgressed candidate adaptive loci (Cgi, *C. gigantea*; CduN, northern *C. duclouxiana*; CduS, southern *C. duclouxiana*)

seasonality contributed most to model predictions for the distribution of *C. duclouxiana*, while annual mean temperature and precipitation of driest month contributed most to predicting the distribution of *C. gigantea*. All models exhibited high predictive ability with  $AUC > 0.9$ . Based on three general circulation models (GCMs: CCSM4, MIROC, and MPI), the ecological niche models for the middle Holocene and last glacial maximum predicted a wide potential distribution for *C. duclouxiana*, but a narrow one for *C. gigantea*, during both periods (Fig. 6). These different distribution patterns for the two taxa concur with results indicated by the Stairway Plot of effective population sizes (Fig. 3b). For both species, a narrower distribution was predicted for the middle Holocene relative to present-day distributions and that during the last glacial maximum period. Furthermore, the distribution of *C. gigantea* is predicted to have overlapped that of *C. duclouxiana* during the last glacial maximum (Fig. 6). Based on Welch's  $t$ -test, it is also predicted that the mean elevation for the distribution of *C. duclouxiana* increased gradually with time and is significantly greater now than it was during the middle Holocene and last glacial maximum periods (middle Holocene:  $t = 58.151$ , d.f. = 15,668,  $P < 2.2 \times 10^{-16}$ ; last glacial maximum:  $t = 69.725$ , d.f. = 28,776,  $P < 2.2 \times 10^{-16}$ ).



**Fig. 6** Potential distributions of *C. duclouxiana* and *C. gigantea* predicted by ecological niche modeling for the present, the mid-Holocene and last glacial maximum periods. Ecological niche models for the mid-Holocene and last glacial maximum are based on three general circulation models (GCMs): CCSM4 (green), MIROC (orange), and MPI (blue). The potential distributions supported by all three GCMs are indicated in pink

## Discussion

Our large-scale analysis of transcriptome-based SNP variation confirmed that the two cypress species, *Cupressus gigantea* and *C. duclouxiana*, are highly divergent genetically. This was evident from phylogenetic, principal component, and genetic clustering analyses conducted on a large set of transcriptomic SNPs. Our analyses also showed that within *C. duclouxiana* a northern group of populations is genetically divergent from a southern group, and further that individuals of the northern group are of mixed ancestry between *C. gigantea* and southern *C. duclouxiana* with most of their ancestry equating to that of southern *C. duclouxiana*. A similar pattern of differentiation is also evident for morphological characters. Thus, while *C. gigantea* has a wider terminal shoot diameter than both northern and southern types of *C. duclouxiana*, the terminal shoot diameter of northern *C. duclouxiana* is greater than that of southern *C. duclouxiana* (Supplementary Fig. 4). An analysis of demographic history using coalescent-based simulations indicated that *C. gigantea* and *C. duclouxiana* diverged from their common ancestor approximately 3.35 Mya, while northern and southern types of *C. duclouxiana* diverged approximately 3.2 Mya (Fig. 3a; Table 1). An earlier divergence of the two species indicated by a Stairway

Plot of changes in effective population sizes over time (Fig. 3b) should be treated with caution, given the reduced accuracy and resolution of this approach for inferring more ancient history<sup>49</sup>. Divergence within this *Cupressus* complex could have been triggered by uplift of southeastern Qinghai-Tibet Plateau, which according to Sun et al.<sup>56</sup> took place during the late Neogene, although this is disputed by Renner<sup>57</sup>.

Northern *C. duclouxiana* was inferred to be of mixed ancestry in an analysis using Frappe<sup>45</sup> with  $K = 2$  (Fig. 1e), but, rather surprisingly, was shown to be an independent cluster (i.e., not admixed) when  $K = 3$ . Such a change in genetic pattern has also been observed in some other studies of hybrid lineages. For example, an admixture analysis showed that the Italian sparrow was clearly a hybrid between the House sparrow and the Spanish sparrow when  $K = 2$ , but indicated it was an independent (non-admixed) genetic group when  $K = 3$ <sup>58</sup>. Confirmation that individuals of northern *C. duclouxiana* were hybrid emerged from a Hlest analysis (Fig. 2b, c) which showed their hybrid indices equated to those expected for advanced generation backcrosses.

Further evidence that northern *C. duclouxiana* (Figs. 1e, 2b) originated from introgression of *C. gigantea* genes into *C.*

*duclouxiana* comes from the greater haplotype sharing between *C. gigantea* and northern *C. duclouxiana*, as revealed by an identity-by-descent (IBD) blocks analysis (Fig. 2a), the slightly lower  $F_{ST}$  value between *C. gigantea* and northern *C. duclouxiana*, and the marginally higher nucleotide diversity values ( $\theta_{\pi}$  and  $\theta_w$ ) of northern relative to southern *C. duclouxiana* (Supplementary Table 4). Because it is possible that haplotype sharing might be caused by incomplete lineage sorting from the most recent common ancestor of the two species, rather than by introgression, we conducted an ABBA-BABA analysis, along with tests using a modified  $f$  statistic and estimates of sequence divergence ( $d_{XY}$ ) on each unigene, to distinguish between these two possibilities. These analyses clearly identified 1285 loci likely to be introgressed from *C. gigantea* into *C. duclouxiana*.

Additional evidence that introgression has occurred at a much higher rate from *C. gigantea* into northern than southern *C. duclouxiana* came from coalescent-based simulations (Fig. 3a). These indicated that gene flow from *C. gigantea* into northern *C. duclouxiana* ( $1.42 \times 10^{-6}$ ) was approximately 7 times greater than into southern *C. duclouxiana* ( $2.10 \times 10^{-7}$ ) and that gene flow in the opposite direction was much lower in each case (approximately 28-fold and 20-fold lower, respectively). Highest levels of gene flow were estimated to have occurred, as expected, between northern and southern *C. duclouxiana* ( $6.90 \times 10^{-6}$  in the forward direction vs.  $2.05 \times 10^{-6}$  in the reverse direction).

Of the 1285 loci indicated to have been introgressed from *C. gigantea* into northern *C. duclouxiana*, 16 were positively selected in both *C. gigantea* and northern *C. duclouxiana* but not in southern *C. duclouxiana*. These may be considered as candidate loci that have contributed to the adaptation of *C. gigantea* and northern *C. duclouxiana* to local environments, with some directly involved in the ability of plants to grow at higher latitudes and elevations relative to southern *C. duclouxiana* (Supplementary Table 9). Functional analysis of some of these genes (based on studies conducted on other plants) lends support to this hypothesis. For example, *MYBD* (Fig. 5a) encodes a MYB-like Domain transcription factor, MYBD, that plays a positive role in anthocyanin regulation by stress<sup>59</sup>. Foliar anthocyanins can protect plants facing abiotic or biotic stress, e.g., as sunscreens and antioxidants, thus helping plants to adapt to a wide range of environmental conditions<sup>60</sup>. In *Arabidopsis thaliana*, *MYBD* expression can be stimulated by high light and promote anthocyanin accumulation via inhibiting the expression of *MYBL2*, which encodes a repressor of this process<sup>61</sup>. In addition, *NF-YA7* (Fig. 5a) encodes a multifunctional transcription factor belonging to the NF-YA family which participates in several types of abiotic stress response, including heat, cold, flooding, drought, and nutrient stress<sup>62</sup>. Overexpression of *NF-YA7* in *A. thaliana* leads to the development of a dwarf late-senescent phenotype with enhanced abiotic stress tolerance<sup>62</sup>. Similarly, overexpression of *PtNF-YA9* (the *Populus* homolog of *NF-YA7*) produces *A. thaliana* plants that have a small leaf area, a dwarf phenotype and strong tolerance to salt and drought stress<sup>63</sup>. Thus, it is feasible that introgression of particular alleles of these and other genes (Supplementary Table 9) from *C. gigantea* into *C. duclouxiana* enabled *C. duclouxiana* to adapt to cooler and drier conditions at higher latitudes and elevations and extend its range as a consequence. Future studies of the 16 genes identified as candidates of adaptive introgression from *C. gigantea* into northern *C. duclouxiana* should aim to test their adaptive significance in more detail and involve both functional and fitness analyses under common garden and reciprocal transplant conditions.

Although, *C. gigantea* and *C. duclouxiana* are allopatrically distributed in the Qinghai-Tibet Plateau, it is feasible that they

have been in contact and exchanged genes at various times since they diverged from each other approximately 3.35 Mya. Our analysis of the demographic history of the two species using *fastsimcoal2* indicated that they did not exchange genes prior to divergence of the northern and southern forms of *C. duclouxiana* approximately 3.2 Mya. However, gene flow mainly from *C. gigantea* into *C. duclouxiana* may have triggered divergence of the two forms of *C. duclouxiana*, and occurred at various times since then.

The effective population sizes of both species have gone through massive changes during the Quaternary most likely as a result of the notable climatic oscillations that occurred during this period. Thus, both species experienced a major bottleneck in effective population size approximately 0.6–1.0 Mya, coinciding with the beginning of the largest glaciation that took place on the Qinghai-Tibet Plateau 0.6–0.8 Mya<sup>64</sup> (Fig. 3b). However, whereas *C. duclouxiana* largely recovered its former population size following this glaciation, the effective population size of *C. gigantea* has remained relatively unchanged since this event. During this and other glacial periods of the Quaternary, it is feasible that the distributions of the two species were brought together and exchanged genes. Thus, the results of ecological niche modeling indicated that at the time of the last glacial maximum (approximately 0.02 Mya) *C. gigantea* was distributed to the south of its current distribution and overlapped the distribution of *C. duclouxiana* (Fig. 6).

The occurrence of recent gene flow between *C. gigantea* and *C. duclouxiana*, possibly during the last glacial maximum, is supported by the results of an analysis of rare SNP sharing between these two taxa. Because rare alleles are expected, on average, to have originated only recently, they are likely to be restricted to a single population or species in the absence of recent gene flow<sup>65</sup>. In contrast, our analysis revealed that between approximately 2–5% of rare SNPs were shared between *C. gigantea* and northern *C. duclouxiana* while between 1 and 2% were shared between *C. gigantea* and southern *C. duclouxiana*. This indicates that recent gene flow between the species has occurred and was greater between *C. gigantea* and the northern relative to the southern form of *C. duclouxiana*. However, because the geographical distributions of *C. gigantea* and *C. duclouxiana* do not overlap at the current time, and according to our ecological niche model analysis have not done so after the last glacial maximum, introgression is unlikely to have occurred throughout the Holocene.

Our study has revealed a new example of introgression between two conifers in the mountainous region of the eastern Qinghai-Tibet Plateau and adjacent areas. We identified 1285 genes introgressed into northern *C. duclouxiana* from *C. gigantea*. Among these loci, 16 could be related to habitat adaptation, which were positively selected in *C. gigantea* and northern *C. duclouxiana*, but not in southern *C. duclouxiana*. These loci may have contributed, in turn, to the adaptation of northern *C. duclouxiana* to cooler and drier conditions experienced at higher latitudes and elevations. The two cypress species are currently geographically isolated from each other; however, it is likely that they have been in contact and exchanged genes at various times since they diverged from each other approximately 3.35 Mya. Evidence of recent gene flow was obtained from the sharing of rare SNPs between the two species, and may reflect an exchange of genes during the last glacial maximum when the distributions of the two species likely overlapped according to ecological niche models. The results of our study highlight the likely importance of introgression in the adaptation of species to climate change in the past, and emphasize the potential of introgressed populations for adaptation to future climate change.



## Methods

**Sample collection and mRNA sequencing.** Leaf samples were collected from a total of 65 individuals comprising 30 *Cupressus duclouxiana* and 35 *C. gigantea* trees (Fig. 1a, b; Supplementary Table 1). In addition, one *C. funebris*, one *Juniperus microsperma* and nine *C. chengiana* trees were sampled as outgroups. Young leaves were collected during day-time between 11 a.m. to 3 p.m., frozen instantly using liquid nitrogen and transported to the laboratory. Total RNAs were isolated from each sample using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and the RNeasy Kit (Qiagen, Hilden, Germany) approach. RNAs with poly (A) tails were purified from total RNA using oligo (dT) magnetic beads before fragmenting into short sequences and synthesizing and purifying cDNAs with a PCR extraction kit (QiaQuick). RNA sequencing was performed using paired-end libraries with 300 bp insert size on an Illumina HiSeq 2000 instrument with 125 bp read length. Raw reads were filtered for downstream analyses, during which reads were removed that contained adapters, poly(N) or were of low-quality.

**Transcriptome de novo assembly and annotation.** Filtered reads from one *C. duclouxiana* individual were assembled into contigs using Trinity v2.0.6<sup>66</sup> with default parameters. Transcripts of less than 200 bp were removed and longest transcripts (in case of alternative splice variants) were selected for the final assembly. CD-HIT-EST v4.6 (<https://github.com/weizhongli/cdhit>) was used in the final assembly to eliminate redundancies. To obtain high-quality contigs for further annotation and analysis, we removed sequences showing high similarity with known non-coding RNA sequences in the Rfam database (<http://rfam.xfam.org/>), and also contigs assigned to microbial (MBGD, <http://mbgd.genome.ad.jp/>), fungal, virus and bacterial sources (sequences downloaded from the NCBI database). In addition, sequences for which 50% of bases aligned with sequences in UTRdb (<http://utrdb.ba.itb.cnr.it/>) or contained <200 non-UTR bases were excluded. In this way, a final reference set of 85,401 contiguous expressed sequences (unigenes) was obtained. We compared all high-quality unigenes against the NCBI nr database using BLASTX<sup>67</sup>. Functional classification of GO categories for these sequences was performed using the Blast2GO program<sup>68</sup>.

**Reads mapping and SNPs calling.** For each *C. duclouxiana* and *C. gigantea* sample, high quality reads were aligned to the reference transcriptome using Bowtie2 v2.2.5 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) with default parameters. To identify the ancestral state of these two species and for the identification of positively selected genes (see below), we also mapped the reads of one *C. funebris* sample, one *J. microsperma* sample and nine *C. chengiana* samples to the reference assembly. Picard v1.128 (<https://github.com/broadinstitute/picard>) was used to remove PCR duplicates and to assign read group information containing library, lane and sample identity. RealignerTargetCreator and IndelRealigner in GATK v3.3 (<https://software.broadinstitute.org/gatk/>) were used to realign indels. With parameters set as -q 20 -Q 20 -t DP -m 2 -F 0.002, we executed the mpileup command in SAMtools v1.2 (<https://github.com/samtools/samtools>) to identify SNPs. Sites with coverage depth <2 and >250 and mapping quality <20 were filtered out. SNPs were retained only if they were present in 60 percent of data across all sampled individuals. For all filtered sites in both species, we defined alleles that were the same as those found in *C. funebris* and *J. microsperma* as the ancestral allelic state.

**Phylogenetic and population genetic analysis.** A maximum likelihood (ML) phylogenetic tree for *C. duclouxiana* and *C. gigantea* samples, based on SNP variation, was constructed using RAxML v8.1.24<sup>69</sup> and the GTRGAMMA model for heuristic tree search, with *C. funebris* and *J. microsperma* as outgroups. Estimates of nucleotide diversity,  $\theta_\pi$  (based on pairwise differences between sequences, see ref. 70) and  $\theta_w$  (based on number of segregating sites between sequences, see ref. 71), were calculated using VCFtools v0.1.14 (<http://vcftools.sourceforge.net/>) and the equations of Watterson<sup>71</sup>, respectively. The pairwise genetic differentiation ( $F_{ST}$ ) between species or groups was estimated according to Weir and Cockerham<sup>72</sup>. To further investigate the population structure of *C. duclouxiana* and *C. gigantea*, a principal component analysis (PCA) was conducted on SNP variation using EIGENSOFT v6.0 (<https://github.com/DReichLab/EIG>) after converting the SNP variant calling format to binary ped format using VCFtools and PLINK v1.07 (<http://zzz.bwh.harvard.edu/plink/>). Significance levels of principal components were determined using the Tracy-Widom test and the first two significant components were plotted. In addition, the software Frappe v1.1<sup>45</sup> was used to examine population structure with number of clusters ( $K$ ) ranging from 2 to 10 and each run comprising 10,000 iterations. Also, a refined identity-by-descent (IBD) blocks analysis<sup>46</sup> was performed using the algorithm from BEAGLE v4.1 (window = 600 overlap = 50 ibd = true ibdtrim = 20 ibdld = 3.0) to detect shared haplotypes between individuals of both species. Finally, the heterozygosity and hybrid index of each of the 65 individuals were calculated using the R package Hiest<sup>47</sup> on 144,027 SNPs that had no missing data for each parental population (*C. gigantea* and southern *C. duclouxiana*, respectively) and for which parental populations had allele frequency differences exceeding 0.4.

**Demographic history.** We used *fastsimcoal2*<sup>48</sup> to infer divergence times and gene flow based on the multidimensional site frequency spectrum for *C. gigantea* and the

northern and southern populations of *C. duclouxiana*. To minimize the effects of selection on demographic inference, only SNPs at fourfold degenerate sites (with no missing data across all individuals sampled) were analyzed. SNPs located within 5 bp of each other were also excluded, as were those that significantly deviated ( $P < 0.05$ ) from Hardy-Weinberg equilibrium when tested using VCFtools. The mutation rate per site per generation was estimated as:  $\mu = D \times g/2T$ , where  $D$  is the observed frequency of pairwise differences between the two species,  $T$  is the estimated divergence time and  $g$  is the estimated generation time. Average generation time ( $g$ ) was set to 50 years and estimated divergence time between *Juniperus* and *Cupressus* was set to 62.8 Mya according to previous field surveys and studies of other Cupressaceae species<sup>73</sup>, respectively. These values yielded an estimated mutation rate of  $7.0 \times 10^{-9}$  mutations per site per generation. Parameter estimates were obtained using the composite ML approach for 16 models that differed in presence or absence of migration (Supplementary Fig. 3). Global ML estimates were derived from 40 independent runs, with 50,000–100,000 coalescent simulations and 10–40 likelihood maximization algorithm cycles. Akaike's information criterion and Akaike's weight of evidence were used to assess the relative fit of each model<sup>48</sup>. We chose the model with the highest Akaike's weight as the one of best fit and constructed 90% nonparametric bootstrap confidence intervals by sampling the fourfold degenerate SNP matrix with replacement. In addition, a Stairway Plot analysis was performed to investigate changes in effective population size over time for both *C. gigantea* and *C. duclouxiana*, using fourfold degenerate SNP frequency spectra for each species, respectively<sup>49</sup>. Two hundred SFS subsamples were generated with each containing a random selection of 2/3 sites and the median of these was used as the final estimation.

To test the relative levels of gene flow between *C. gigantea* and northern and southern *C. duclouxiana*, the proportion of rare SNPs shared among the three types was calculated. We defined SNPs with frequencies from 2 (1.54%) to 5 (3.85%) in all 65 individuals, i.e., 130 alleles, as rare and calculated the proportion of them shared only between two of the three populations in each frequency category. Because rare alleles are expected on average to have originated only recently, they will be limited to a single population or species if recent gene flow is absent<sup>65</sup>. Hence, this approach helps to determine whether introgression has occurred recently between the species and populations examined.

**Screening for introgressed genes.** Based on the tree topology presented in Supplementary Fig. 5, we employed both Patterson's  $D$  statistic<sup>50,51</sup> and a modified  $f$  statistic ( $f_{AM}$ )<sup>52,53</sup> to identify potential loci introgressed from *C. gigantea* to the northern population of *C. duclouxiana*, using *C. funebris* and *J. microsperma* as outgroups. The  $D$  statistic was used to examine the phylogenetic distribution of derived alleles at loci that display either an ABBA or BABA allelic configuration using Eq. (1):

$$D(P_1, P_2, P_3, O) = \frac{\sum_{i=1}^n ((1 - \hat{p}_{i1})\hat{p}_{i2}\hat{p}_{i3}(1 - \hat{p}_{i4}) - \hat{p}_{i1}(1 - \hat{p}_{i2})\hat{p}_{i3}(1 - \hat{p}_{i4}))}{\sum_{i=1}^n ((1 - \hat{p}_{i1})\hat{p}_{i2}\hat{p}_{i3}(1 - \hat{p}_{i4}) + \hat{p}_{i1}(1 - \hat{p}_{i2})\hat{p}_{i3}(1 - \hat{p}_{i4}))}, \quad (1)$$

where  $P_1, P_2, P_3$  and  $P_4$  (O) are the four taxa compared and  $p_{ij}$  is the derived allele frequency of a site  $i$  in population  $j$ .

The modified  $f$  statistic,  $f_{AM}$ , was calculated based on Eq. (2):

$$S(P_1; P_2; P_3; O) = \sum_i ((1 - p_{i1})p_{i2}p_{i3}(1 - p_{i4})) - \sum_i (p_{i1}(1 - p_{i2})p_{i3}(1 - p_{i4})) \quad (2)$$

If  $p_{i2} > p_{i1}$ , then  $f_{AM} = S(P_1; P_2; P_3; O)/S(P_1; P_2; P_3; O)$  and  $P_D$  is the population ( $P_2$  or  $P_3$ ) that has the higher frequency of the derived allele; if  $p_{i2} < p_{i1}$ , then  $f_{AM} = S(P_1; P_2; P_3; O)/-S(P_1; P_2; P_3; O)$  and  $P_D$  is the population ( $P_1$  or  $P_3$ ) that has the higher frequency of the derived allele<sup>53</sup>.

For each of all the assembled unigenes, the standard error was calculated using a moving block bootstrap approach with optimal block size being 200 bp. All tests were followed by two tailed  $Z$ -tests to determine if each  $D$  or  $f_{AM}$  value was significantly different from zero, indicating potential gene flow.

To rule out false positive introgressed loci, resulting from incomplete lineage sorting, mean pairwise sequence divergence ( $d_{XY}$ ) of the whole transcriptome and  $d_{XY}$  for each unigene were calculated between *C. gigantea* and northern *C. duclouxiana* as a complementary analysis to the  $D$  statistic and modified  $f$  statistic<sup>74</sup>, using python scripts ([https://github.com/simonmartin/genomics\\_general](https://github.com/simonmartin/genomics_general); see ref. 52). For each candidate introgressed locus,  $d_{XY}$  was calculated using a smaller block size of 200 bp for moving block bootstrapping. If the mean  $d_{XY}$  of a putatively introgressed locus was lower than the mean value of the whole transcriptome, the two values were compared statistically using a Mann-Whitney  $U$ -test. The Benjamini-Hochberg false discovery rate method was performed to decrease false positive rate for all tests ( $D$  statistic,  $f$  statistic and  $d_{XY}$ ;  $P < 0.01$ ).

Having identified candidate introgressed loci in the above way, GO enrichment analysis was conducted using the TopGO package in Bioconductor (<http://www.bioconductor.org>). Fisher's exact tests with weight01 algorithms were employed to examine the significance of enrichment, with corrected  $P$ -values < 0.05 considered significant.

**Identification of positively selected genes.** Two methods, the population branch statistic<sup>54</sup> and the Hudson-Kreitman-Aguadé test<sup>55</sup>, were applied to identify genes under positive selection within two target populations, *C. gigantea* and northern *C. duclouxiana*, separately. Samples from southern *C. duclouxiana* were assumed as the control population. As two populations (MSZ-43 and MSZ-51) from southern *C. duclouxiana* may have been recipients of gene flow from northern *C. duclouxiana* (Supplementary Fig. 2), we discarded them within these analyses to avoid spurious signals of positive selection. We carried out the population branch statistics for two triples, *C. gigantea*-southern *C. duclouxiana*-outgroup and northern *C. duclouxiana*-southern *C. duclouxiana*-outgroup. Nine individuals of *C. chengiana* are assumed as an outgroup population. For each unigene, we calculated  $F_{ST}$  between population pairs including the target population and control population, the target population and outgroup and the control population and outgroup. The population branch statistic (PBS) value of the target population was calculated using Eq. (3):

$$PBS_{popT} = \frac{T^{popT-C} + T^{popT-O} - T^{popC-O}}{2}, \quad (3)$$

where  $T = -\log(1-F_{ST})$  is the population divergence time  $T$  in units scaled by the population size.

For the Hudson-Kreitman-Aguadé tests, the number of polymorphic sites in the target population (*C. gigantea* or northern *C. duclouxiana*) was denoted as  $A$  and the number of fixed differences between the target population and both control populations (southern *C. duclouxiana*) and outgroup (*C. funebris*) was denoted as  $B$ . We compared the ratio of  $A/B$  for each unigene to the transcriptome-wide average and tested the null hypothesis  $A(\text{unigene})/B(\text{unigene}) = A(\text{transcriptome-wide})/B(\text{transcriptome-wide})$  using Pearson's chi-square test for the  $2 \times 2$  contingency table.

Unigenes with the highest 10% of the population branch statistic and a significant  $P$ -value ( $<0.05$ ) for the Hudson-Kreitman-Aguadé test were recognized as positively selected genes in *C. gigantea*/northern *C. duclouxiana*. We carried out the GO enrichment analysis for positively selected genes using the same method applied to candidate introgressed loci.

**Morphological differentiation and habitat differentiation.** To quantify morphological differentiation between *C. gigantea*, northern and southern *C. duclouxiana*, we compared the diameter of the last ramification shoots with leaves. Nine individuals across three populations for each group were examined, and for each individual five of the last ramification shoots were randomly collected, preserved as herbarium specimens and measured. Mann-Whitney  $U$ -test was used to test the differences of diameters among groups.

To better illustrate the habitat differentiation among the three groups, a scatter plot for latitude against elevation of all populations was constructed. Subsequently, we conducted a principal component analysis on climate data (26 climate variables) obtained from ten climate stations (Supplementary Tables 2, 3) located within the sampling areas of the three groups, from the National Meteorological Information Center, China (<http://data.cma.cn>). Principal component analysis and graphical illustrations were performed with R v3.4.0.

**Ecological niche modeling.** To explore the distributional shifts of the two *Cupressus* species in response to recent Quaternary climate change, we conducted ecological niche modeling with MAXENT v3.3.4<sup>75</sup>. This predicted the potential distribution of *C. duclouxiana* and *C. gigantea* at present, during the mid-Holocene (0.006 Mya) and at the last glacial maximum (0.022 Mya) using the following parameters: random test points = 25; replicates = 20; type = subsample; maximum iterations = 5000. Information from 56 presence sites for *C. duclouxiana* and 17 for *C. gigantea* were collected from field investigations and herbarium records (<http://www.gbif.org>). Data for 19 bioclimatic variables for each period were retrieved from the WorldClim database (<http://www.worldclim.com>) with a resolution of 2.5 arc-min. For the mid-Holocene and last glacial maximum, paleoclimate data, based on three different GCMs (CCSM4, MIROCESM, MPI-ESM-P), were employed. To avoid multicollinearity, we examined pairwise Pearson correlations ( $r$ ) among the bioclimatic variables and eliminated one of the variables in each pair with  $r > 0.7$ . Seven variables were finally retained (Bio1 annual mean temperature, Bio2 mean diurnal range, Bio3 isothermality, Bio4 temperature seasonality, Bio13 precipitation of wettest month, Bio14 precipitation of driest month, Bio15 precipitation seasonality). The AUC (area under the receiver operating characteristic curve) values were used to compare predictive performance of one model with another and values above 0.9 indicated better model performance. Jackknife tests were also performed to identify which variables contributed the most individually. In addition, we downloaded altitude information from the SRTM elevation database (<https://www2.jpl.nasa.gov/srtm/>). The elevation for each predicted species distribution model (SDM) was extracted using QGIS v2.18.10 (<https://qgis.org/en/site/>). Welch's  $t$ -test was employed to examine divergence between the average altitude values of the SDMs for *C. duclouxiana*.

Changes in distributional shifts of the two species in response to recent Quaternary climate change as revealed by our ecological modeling analysis provide an environmental context to the effects of adaptive introgression on the recent demography of these two species.

**Statistics and reproducibility.** Pearson's chi-square test and Welch's  $t$ -test were performed using R v3.4.0.  $Z$ -test was performed using perl module Statistics::Zed. Mann-Whitney  $U$ -test was performed using perl module Statistics::Test::WilcoxonRankSum. Benjamini-Hochberg FDR correction was performed using perl module Statistics::Multtest. All the results within this study are reproducible.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The transcriptome sequencing data have been deposited in the Short Read Archive at NCBI under the accession codes SAMN08634857 to SAMN08634891, SAMN08634904 to SAMN08634933, SAMN08638057, and SAMN08638058.

## Code availability

The custom scripts have been made available at <https://github.com/mayz11/cypress>.

Received: 19 November 2018 Accepted: 29 April 2019

Published online: 18 June 2019

## References

- Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. & Curtis-McLane, S. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evol. Appl.* **1**, 95–111 (2008).
- Stocker, T. F. et al. Technical summary. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Stocker, T. F. et al.) (Cambridge University Press, Cambridge, 2014).
- Firmat, C., Delzon, S., Louvet, J. M., Parmentier, J. & Kremer, A. Evolutionary dynamics of the leaf phenological cycle in an oak metapopulation along an elevation gradient. *J. Evol. Biol.* **30**, 2116–2131 (2017).
- Alberto, F. J. et al. Potential for evolutionary responses to climate change—evidence from tree populations. *Glob. Change Biol.* **19**, 1645–1661 (2013).
- Hamilton, J. A. & Miller, J. M. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conserv. Biol.* **30**, 33–41 (2016).
- Martin, S. H. & Jiggins, C. D. Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* **47**, 69–74 (2017).
- Huerta-Sánchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
- Miao, B., Wang, Z. & Li, Y. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan Plateau. *Mol. Biol. Evol.* **34**, 734–743 (2017).
- Wu, D. D. et al. Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex. *Nat. Ecol. Evol.* **2**, 1139–1145 (2018).
- Owens, G. L., Baute, G. J. & Rieseberg, L. H. Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. *Mol. Ecol.* **25**, 2630–2643 (2016).
- Suarez-Gonzalez, A. et al. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Mol. Ecol.* **25**, 2427–2442 (2016).
- Khodwekar, S. & Gailing, O. Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations. *Am. J. Bot.* **104**, 1088–1098 (2017).
- Anderson, E. *Introgressive Hybridization* (Wiley, Hoboken, 1949).
- Rieseberg, L. & Wendel, J. F. in *Hybrid Zones and the Evolutionary Process* (ed. Harrison, R.) (Oxford University Press, Oxford, 1993).
- Arnold, M. L. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *Plant Cell* **16**, 562–570 (2004).
- Barton, N. H. The role of hybridization in evolution. *Mol. Ecol.* **10**, 551–568 (2001).
- Abbott, R. J. et al. Hybridization and speciation. *J. Evol. Biol.* **26**, 229–246 (2013).
- Suarez-Gonzalez, A., Lexer, C. & Cronk, Q. C. B. Adaptive introgression: a plant perspective. *Biol. Lett.* **14**, 20170688 (2018).
- Dasmahapatra, K. K. et al. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
- Norris, L. C. et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl Acad. Sci.* **112**, 815–820 (2015).
- Arnold, B. J. et al. Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl Acad. Sci.* **113**, 8320–8325 (2016).

22. Arnold, M. L. & Kunte, K. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. *Trends Ecol. Evol.* **32**, 601–611 (2017).
23. Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Douglas, C. J. & Cronk, Q. C. B. Introgression from *Populus balsamifera* underlies adaptively significant variation and range boundaries in *P. trichocarpa*. *New Phytol.* **217**, 416–427 (2018).
24. Willyard, A., Cronn, R. & Liston, A. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Mol. Phylogenet. Evol.* **52**, 498–511 (2009).
25. Abbott, R. J., Barton, N. H. & Good, J. M. Genomics of hybridization and its evolutionary consequences. *Mol. Ecol.* **25**, 2325–2332 (2016).
26. De La Torre, A. R. in *Evolutionary Biology: Biodiversification from Genotype to Phenotype* (ed. Pontarotti, P.) (Springer, Cham, 2015).
27. Abbott, R. J. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *J. Syst. Evol.* **55**, 238–258 (2017).
28. Currat, M., Ruedi, M., Petit, R. J. & Excoffier, L. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**, 1908–1920 (2008).
29. Petit, R. J. & Excoffier, L. Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393 (2009).
30. Sork, V. L. Genomic studies of local adaptation in natural plant populations. *J. Hered.* **109**, 3–15 (2017).
31. Savolainen, O., Pyhäjärvi, T. & Knürr, T. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Syst.* **38**, 595–619 (2007).
32. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat. Rev. Genet.* **14**, 807–820 (2013).
33. Holliday, J. A., Zhou, L., Bawa, R., Zhang, M. & Oubida, R. W. Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytol.* **209**, 1240–1251 (2016).
34. Lind, B. M., Menon, M., Bolte, C. E., Faske, T. M. & Eckert, A. J. The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genet. Genomes* **14**, 29 (2018).
35. Hamilton, J. A., Lexer, C. & Aitken, S. N. Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* × *P. glauca*) hybrid zone. *New Phytol.* **197**, 927–938 (2013).
36. de Lafontaine, G., Prunier, J., Gérard, S. & Bousquet, J. Tracking the progression of speciation: variable patterns of introgression across the genome provide insights on the species delimitation between progenitor-derivative spruces (*Picea mariana* × *P. rubens*). *Mol. Ecol.* **24**, 5229–5247 (2015).
37. Chhatre, V. E., Evans, L. M., DiFazio, S. P. & Keller, S. R. Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Mol. Ecol.* **27**, 4820–4838 (2018).
38. Oney-Birol, S., Fitz-Gibbon, S., Chen, J., Gugger, P. F. & Sork, V. L. Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (*Quercus* sect. *Quercus*). *BMC Genet.* **19**, 88 (2018).
39. Nystedt, B. et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).
40. Ru, D. F. et al. Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. *Mol. Ecol.* **27**, 4875–4887 (2018).
41. Farjon, A. A *Monograph of Cupressaceae and Sciadopitys* (Royal Botanic Gardens, Kew, 2005).
42. Fu, L. K., Yu, Y. F. & Farjon, A. in *Flora of China* (eds Wu Z. Y. & Raven P. H.) (Science Press & Missouri Botanical Garden Press, St. Louis, 1999).
43. Xu, T. et al. Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *BMC Evol. Biol.* **10**, 194 (2010).
44. Lu, X. et al. Genetic diversity and conservation implications of four *Cupressus* species in China as revealed by microsatellite markers. *Biochem. Genet.* **52**, 181–202 (2014).
45. Tang, H., Peng, J., Wang, P. & Risch, N. J. Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* **28**, 289–301 (2005).
46. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
47. Fitzpatrick, B. M. Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evol. Biol.* **12**, 131 (2012).
48. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
49. Liu, X. & Fu, Y. X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).
50. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
51. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
52. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
53. Malinsky, M. et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493–1498 (2015).
54. Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
55. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
56. Sun, B. N. et al. Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **304**, 328–336 (2011).
57. Renner, S. S. data point to a 4-km-high Tibetan Plateau by 40 Ma, but 100 molecular-clock papers have linked supposed recent uplift to young node ages. *J. Biogeogr.* **43**, 1479–1487 (2016).
58. Elgvin, T. O. et al. The genomic mosaicism of hybrid speciation. *Sci. Adv.* **3**, e1602996 (2017).
59. Ma, D. & Constabel, C. P. MYB repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci.* **24**, 275–289 (2019).
60. Landi, M., Tattini, M. & Gould, K. S. Multiple functional roles of anthocyanins in plant–environment interactions. *Environ. Exp. Bot.* **119**, 4–17 (2015).
61. Nguyen, N. H. et al. MYBD employed by HY5 increases anthocyanin accumulation via repression of MYBL2 in *Arabidopsis*. *Plant J.* **84**, 1192–1205 (2015).
62. Leyva-González, M. A., Ibarra-Laclette, E., Cruz-Ramírez, A. & Herrera-Estrella, L. Functional and transcriptome analysis reveals an acclimatization strategy for abiotic stress tolerance mediated by *Arabidopsis* NF-YA Family Members. *PLoS ONE* **7**, e48138 (2012).
63. Lian, C. et al. *Populus trichocarpa* PtNF-YA9, a multifunctional transcription factor, regulates seed germination, abiotic stress, plant growth and development in *Arabidopsis*. *Front. Plant Sci.* **9**, 954 (2018).
64. Zheng, B., Xu, Q. & Shen, Y. The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quat. Int.* **97**, 93–101 (2002).
65. Stryjewski, K. F. & Sorenson, M. D. Mosaic genome evolution in a recent and rapid avian radiation. *Nat. Ecol. Evol.* **1**, 1912–1922 (2017).
66. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
67. Altschul, S. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
68. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
69. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
70. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci.* **76**, 5269–5273 (1979).
71. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
72. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
73. Mao, K. S., Hao, G., Liu, J. Q., Adams, R. P. & Milne, R. I. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytol.* **188**, 254–272 (2010).
74. Smith, J. & Kronforst, M. R. Do *Heliconius* butterfly species exchange mimicry alleles? *Biol. Lett.* **9**, 20130503 (2013).
75. Phillips, S. J. & Dudík, M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**, 161–175 (2008).

## Acknowledgements

We thank Wenjing Tao for helps on the analysis of morphological and habitat differentiation. This work was supported by the National Natural Science Foundation of China (grant numbers 31590821, 31622015, 31370261), the National Basic Research Program of China (grant number 2014CB954100), Sichuan Provincial Department of Science and Technology (grant number 2015JQ0018) and Sichuan University (Fundamental Research Funds for the Central Universities, SCU2019D013, SCU 2018D006).

## Author contributions

K.M. and Jianquan Liu designed and supervised this study. K.M., L.Z., Y.S. and Jialiang Li managed the fieldwork and collected the materials. Y.M., J.W., Q.H., Jialiang Li and Y.S.

performed the data analysis. Y.M., K.M., R.J.A., and Jianquan Liu prepared the manuscript. All authors read and approved the manuscript.

### Additional information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42003-019-0445-z>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019