

## Research paper

# Allopolyploid origin and niche expansion of *Rhodiola integrifolia* (Crassulaceae)



Da-Lv Zhong, Yuan-Cong Li, Jian-Qiang Zhang\*

National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, College of Life Sciences, Shaanxi Normal University, No. 620, West Chang'an Avenue, Chang'an District, Xi'an, 710119, China

## ARTICLE INFO

## Article history:

Received 26 December 2021

Received in revised form

18 August 2022

Accepted 19 August 2022

Available online 24 August 2022

## Keywords:

Allopolyploid

Hybridization

*ncpGS*

Niche shift

*rpb2*Schoener's *D*

## ABSTRACT

Polyploidy after hybridization between species can lead to immediate post-zygotic isolation, causing saltatory origin of new species. Although the incidence of polyploidization in plants is high, it is thought that a new polyploid lineage can succeed only if it establishes a new ecological niche divergent from its progenitor lineages. We tested the hypothesis that *Rhodiola integrifolia* from North America is an allopolyploid produced by *R. rhodantha* and *R. rosea* and determined whether its survival can be explained by the niche divergence hypothesis. To this end, we sequenced two low-copy nuclear genes (*ncpGS* and *rpb2*) in a phylogenetic analysis of 42 *Rhodiola* species and tested for niche equivalency and similarity using Schoener's *D* as the index of niche overlap. Our phylogeny-based approach showed that *R. integrifolia* possesses alleles from both *R. rhodantha* and *R. rosea*. Dating analysis showed that the hybridization event that led to *R. integrifolia* occurred ca. 1.67 Mya and niche modeling analysis showed that at this time, both *R. rosea* and *R. rhodantha* may have been present in Beringia, providing the opportunity for the hybridization event. We also found that the niche of *R. integrifolia* differs from that of its progenitors in both niche breadth and optimum. Taken together, these results confirm the hybrid origin of *R. integrifolia* and support the niche divergence hypothesis for this tetraploid species. Our results underscore the fact that lineages with no current overlapping distribution could produce hybrid descendants in the past, when climate oscillations made their distributions overlap.

Copyright © 2022 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Polyploidy has long been regarded as an important force in plant evolution (Stebbins, 1950; Grant, 1981), and can occur either within species (autopolyploidy) or after hybridization between species (allopolyploidy). Polyploidy after hybridization between species can lead to immediate post-zygotic isolation, causing the saltatory origin of new species (Mallet, 2007; Soltis and Soltis, 2009; Abbott et al., 2013). It has been estimated that 15% of speciation events in angiosperms and 31% ferns involved polyploidy events (Wood et al., 2009). Previous studies have identified progenitor lineages of allopolyploids (e.g., Robertson et al., 2010; Kelly et al., 2013) in various plant groups, providing important insights into the role hybridization plays in plant evolution.

Although the incidence of polyploidization in plants is high, it is thought that a new polyploid lineage, especially an autopolyploid lineage succeeds only if it establishes a new ecological niche divergent from its progenitor lineages (Levin, 1975, 2002). Otherwise, the lineage will likely be eliminated by 'minority cytotype disadvantage' (Levin, 1975). Recent studies based on global climate data and new analytical methods (Warren et al., 2008; Wang et al., 2022) have shown that the niches of allopolyploids differ from at least one of their progenitors (López-Alvarez et al., 2015; Ficetola and Stöck, 2016; Marchant et al., 2016; Akiyama et al., 2021; Han et al., 2022). Nevertheless, other studies have shown that allopolyploids occupy intermediate or non-divergent ecological niches compared to their progenitors (e.g., Casazza et al., 2017). Thus, additional empirical studies are needed to elucidate the role niche shift plays in the establishment of polyploid lineages.

The genus *Rhodiola* L. comprises 55 species in China (Fu and Ohba, 2001) and ca. 70 species worldwide, most of which are distributed in the Qinghai-Tibet Plateau and its adjacent areas. Phylogenetic and biogeographic studies have shown that *Pseudo-sedum* (Boiss.) Berger should be merged with *Rhodiola*, and that the

\* Corresponding author.

E-mail address: [jqzhang@snnu.edu.cn](mailto:jqzhang@snnu.edu.cn) (J.-Q. Zhang).

Peer review under responsibility of Editorial Office of Plant Diversity.

genus originated in the Qinghai-Tibet Plateau and its adjacent areas (Zhang et al., 2014a, b). These studies also demonstrated several cases of incongruences between phylogenies based on plastid and nuclear data sets, indicating potential reticulate evolution of the genus (Zhang et al., 2014a). Three species, i.e., *R. rosea* L., *R. integrifolia* Raf., and *R. rhodantha* (A. Gray) H. Jacobsen, expanded their distribution to North America via two dispersal events, which have been dated to be in the Pliocene and the Pleistocene, respectively (Guest and Allen, 2014; Zhang et al., 2014a). *R. rhodantha* is distinguished from the other two species by its elongated raceme, bisexual flowers, and greenish or white to rose petals. *R. integrifolia* is morphologically similar to *R. rosea* in bearing corymbose cymes and unisexual flowers. The two species differ in petal color and the color of their leaf blades. *R. integrifolia* has dark red petals and bright green leaf blades, whereas *R. rosea* has pale yellow petals and pale green and usually glaucous leaf blades (Clausen, 1975).

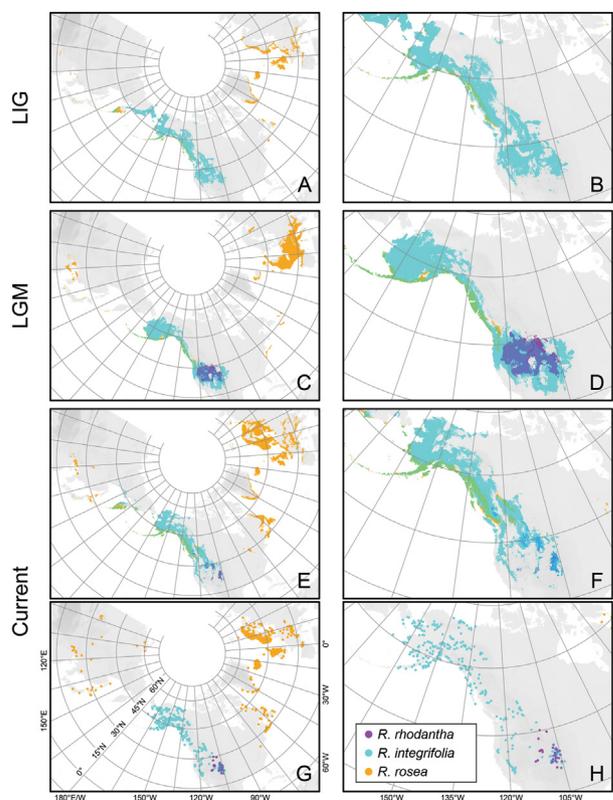
Previous cytogenetic study (Uhl, 1952) revealed the haploid chromosome numbers for *R. rhodantha* (7), *R. rosea* (11), and *R. integrifolia* (18). Uhl (1952) then hypothesized that the 18-chromosome strain originated either by hexaploidization of an ancestral *Rhodiola* species with a haploid chromosome number of  $x = 6$  or by an allopolyploidization of the 11-chromosome strain and 7-chromosome strain, which was likely *R. rhodantha*. This hypothesis was tested and confirmed by Hermsmeier et al. (2012) using sequences of nuclear-encoded chloroplast-expressed glutamine synthetase (*npcGS*) gene. However, this study included a limited number of samples with few Asian species, and no phylogenetic methods were employed. In addition, the hypothesis is contradicted by the following facts. First, the current distribution areas of *R. rosea* and *R. rhodantha* do not overlap. *R. rosea* grows in the arctic circumpolar area and alpine habitats of Asia, Europe, and eastern North America, from Nunavut to North Carolina. In contrast, *R. rhodantha* is an endemic of the Rocky Mountains of the United States (Fig. 1G). Second, the phylogeny based on the internal transcribed spacers (ITS) data, a marker from the nuclear genome, does not support hybrid origin of *R. integrifolia*'s (Zhang et al., 2014a, b), as sequences of *R. integrifolia* were clustered with *R. rhodantha*, consistent with the plastid data (Zhang et al., 2014b).

To reconcile these discordant facts, we hypothesize that: (1) *R. integrifolia* indeed originated via allopolyploidization; (2) *R. rosea* and *R. rhodantha* are its parental species; (3) the hybridization event occurred when the distributions of *R. rosea* and *R. rhodantha* overlapped, probably due to range expansions in the Quaternary climate oscillations; (4) rapid concerted evolution only retained ITS alleles of *R. rhodantha* in *R. integrifolia*. To test these hypotheses, we sequenced two low-copy nuclear markers that have been widely used in other phylogenetic studies (Zimmer and Wen, 2013), as well as in studies of *Rhodiola* (Hermsmeier et al., 2012; Li et al., 2019) to reconstruct a phylogeny for 42 species of *Rhodiola*. In addition, we used ecological niche modeling (ENM) to reconstruct the past distribution of the two parental species and to infer the time period when the distribution of the two species probably overlapped. Finally, we dated the divergence time between plastid haplotypes of *R. rhodantha* and *R. integrifolia* to infer the time of the hybridization event. Through these studies, we can not only confirm the hybrid origin of *R. integrifolia*, but also infer where and when this hybridization event happened.

## 2. Materials and methods

### 2.1. Sample collection

A total of 77 accessions of *Rhodiola* species (including 13 accessions contributed by colleagues) were used for low-copy nuclear



**Fig. 1.** Predicted distributions of *Rhodiola integrifolia*, *R. rosea* and *R. rhodantha*. The suitability threshold shown on the map is 0.4. (A) The predicted distributions of the three species during the Last Interglacial (LIG, ca. 130 kya); (B) Enlargement of the northwestern coast of North America; (C) The predicted distributions of the three species during the Last Glacial Maximum (LGM, ca. 20 kya); (D) Enlargement of the northwestern coast of North America; (E) The predicted distributions of the three species in the current period; (F) Enlargement of the northwestern coast of North America; (G) Localities used in the present study for species distribution modeling; (H) Enlargement of the northwestern coast of North America. In A, B, C, D, E, F, the green color is potential overlapping distribution between *R. integrifolia* and *R. rosea*; the pale purple color is potential overlapping distribution between *R. rhodantha* and *R. integrifolia*.

gene cloning and sequencing (Table S1). These samples represent 42 *Rhodiola* species, including four accessions of *R. integrifolia*, six accessions of *R. rhodantha* and five accessions of *R. rosea*. Two accessions from *Phedimus aizoon* (L.) Hart and *Sedum drymarioides* Hance were used as outgroups. For each individual, we collected fresh leaves into silica gel in the field.

### 2.2. DNA extraction, amplification and sequencing

Total genomic DNA was extracted from 20 to 30 mg of dried leaves using a modified CTAB method (Doyle and Doyle, 1987). Two bi-parentally inherited nuclear markers were cloned and sequenced: chloroplast-expressed glutamine synthetase *npcGS* (glutamate-ammonia ligase) (Hermsmeier et al., 2012), and *rpb2*, which encodes the largest subunit of RNA polymerase II (Zimmer and Wen, 2013). The PCR primers for *npcGS* were *npcGSF* (5'-GGTGATTGGAATGGTGC-3') and *npcGSR* (5'-GCCTTGTTTCTCAGTATCG-3'); for *rpb2* were *rpb2F* (5'-AGGCAACCAACAAGTAT-3') and *rpb2R* (5'-CGGTGAAGTGTCCGTAT-3'). The PCR reaction mix totaled 20  $\mu$ l, containing 10  $\mu$ l of  $2 \times$  Taq master mix reaction buffer (including dNTPs), 0.8  $\mu$ l of each primer at 100 ng/ $\mu$ l, 1  $\mu$ l of genomic DNA at 50 ng/ $\mu$ l, and 7.8  $\mu$ l of Milli-Q H<sub>2</sub>O. The annealing temperature was set to 52 °C. PCR fragments were visualized on

agarose gels. Purification using polyethylene glycol (PEG) were conducted prior to sequencing. All purified PCR products were sequenced by Sangon Biotech (Shanghai, China) using an ABI 3739XL sequencer (Applied Biosystems, Foster City, California, U.S.A.). The same primers in amplification were used for sequencing. For most accessions, PCR products were directly sequenced, while others had multiple bands (indicating multiple alleles) on agarose gels. These sequences were ligated onto pGEM-T Easy Vector using a Promega Kit (Promega Corporation, Madison, WI, USA). Six to eight plasmids containing insertions were chosen for sequencing reactions.

### 2.3. Sequence alignment and phylogenetic analysis

Contigs obtained were edited and assembled in Geneious R11 (<http://www.geneious.com/>). Sequence alignment was conducted using MUSCLE v.3.8.31 (Edgar, 2004), followed by a visual check in Geneious R11. Optimal nucleotide substitution models were chosen by jModeltest v.2.1 (Darrriba et al., 2012) according to the Akaike Information Criterion (AIC) (the *ncpGS* data set: T92 + G; the *rpb2* data set: TN93 + G). RAxML v.8.2.10 (Stamatakis, 2014) was used to infer the maximum likelihood (ML) tree with the GTRIG model and 1000 bootstrap replicates. The Neighbor-Net algorithm based on uncorrected *P*-distance was performed with SplitsTree v.4.14.1 (Huson et al., 2008).

### 2.4. Divergence time estimation

We used the maternal inherited plastid markers to estimate the divergence time between *R. integrifolia* and *R. rhodantha*. This estimation can be regarded as the time at which the hybridization event occurred, because after the hybridization event, the plastome of *R. integrifolia* began to differentiate from that of *R. rhodantha*. We downloaded the plastid data set of Zhang et al. (2014a) (including *psbA-trnH*, *trnL-F*, *ndhA* intron, *rpS16* intron, *rpl16* intron, *rbcl*, *matK*, and *trnS-G*) and only kept accessions of *R. rhodantha* and *R. integrifolia*, with *R. prainii* (Hamet) H. Ohba, *R. alterna* S.H. Fu, *R. stapfii* (Hamet) S.H. Fu, *R. chrysanthemifolia* (Levl.) S.H. Fu and *R. sinuata* (Royle ex Edgew.) S.H. Fu as outgroups. The molecular clock hypothesis was tested using a likelihood ratio test in PAUP\* v.4.0b10. BEAST v.1.7.1 (Drummond et al., 2012) was used to conduct the dating analysis. The program BEAUti was used to set parameters and priors for the analysis. All parameters were sampled every 1000 generations from 10,000,000 Markov chain Monte Carlo (MCMC) generations with the first 25% as burn-in. Tracer v.1.7.1 (Rambaut et al., 2018) was used to examine the convergence of MCMC chains. As there are no reliable fossils for *Rhodiola*, we used substitution rates instead. The evolutionary rate of the plastid markers was assumed to be  $2 \times 10^{-9}$  substitutions/site/year (Yamane et al., 2006). The computer program TreeAnnotator v.1.7.5 was used to produce the maximum clade credibility tree (MCC).

### 2.5. Niche-modeling analyses

We compiled species occurrence data from different sources, including the Global Biodiversity Information Facility (<http://www.gbif.org/>) and the Chinese Virtual Herbarium (CVH; <http://www.cvh.ac.cn>). After removing duplicated (no more than one point in each grid) and possible fallacious record (e.g., in botanic gardens or water areas), 351 georeferenced occurrence records for *R. rosea*, 272 for *R. integrifolia* and 34 for *R. rhodantha* were used for subsequent analyses (Table S2).

We downloaded 19 climatic factors (Table S3) for the Last Interglacial (LIG, ca. 130 kya, thousand years ago), the Last Glacial

Maximum (LGM; the MIROC model (Hasumi and Emori, 2004; ca. 20 kya)) and the current time from the Worldclim database at the resolution of 30 arc-seconds (Hijmans et al., 2005). To remove highly correlated environmental variables, we performed pairwise Pearson correlations, excluding variables with  $r > 0.8$ . Eight variables were kept for subsequent analysis: BIO1 (Annual mean temperature), BIO2 (Mean diurnal range), BIO4 (Temperature seasonality), BIO8 (Mean temperature of wettest quarter), BIO9 (Mean temperature of driest quarter), BIO15 (Precipitation seasonality), BIO18 (Precipitation of warmest quarter), BIO19 (Precipitation of coldest quarter). MaxEnt 3.3.3 (Phillips et al., 2006) was used to run species distribution models (SDMs). We used 75% of the occurrence data for training the model and 25% for testing. The area under the curve (AUC) was used to evaluate the model performance (Elith et al., 2006; Phillips et al., 2006).

The R package ‘ecospat’ (Broennimann et al., 2012; Di Cola et al., 2017) was used to estimate the three species’ ENMs. To estimate the climatic factor densities along environmental axes (PCA-env), an ordination approach (Principal Component Analysis, PCA) was used. We then used these densities to calculate ENM overlap. Schoener’s *D* metric (Schoener, 1968) was used to evaluate ENM overlap. This metric varies from 0 (no overlap) to 1 (complete overlap).

Niche equivalency and similarity tests were conducted according to Molina-Henao and Hopkins (2019). In summary, the niche equivalency test’s null hypothesis is that the ENMs of compared species are statistically identical. Using a bootstrap resampling approach (Warren et al., 2008; Broennimann et al., 2012), rejecting the null hypothesis means that the niches are not equivalent. The ENM similarity test also uses bootstrap resampling to evaluate if one ENM predicts the other better than a randomly generated ENM from the background range.

If compared niches were not significantly equivalent, we then calculated the optimum and breadth of ENMs. The median of the 95% inter-percentile interval along the first three PCA-env axes were regarded as the ENM optimum, and the length as the ENM breadth (Broennimann et al., 2012). We used one-way ANOVA to test if the ENM optimum along the tree axes were significantly different, and also conducted pairwise comparison using Tukey’s HSD Test. All statistical analyses were performed in R v.3.5.1 (R Core Team, 2018).

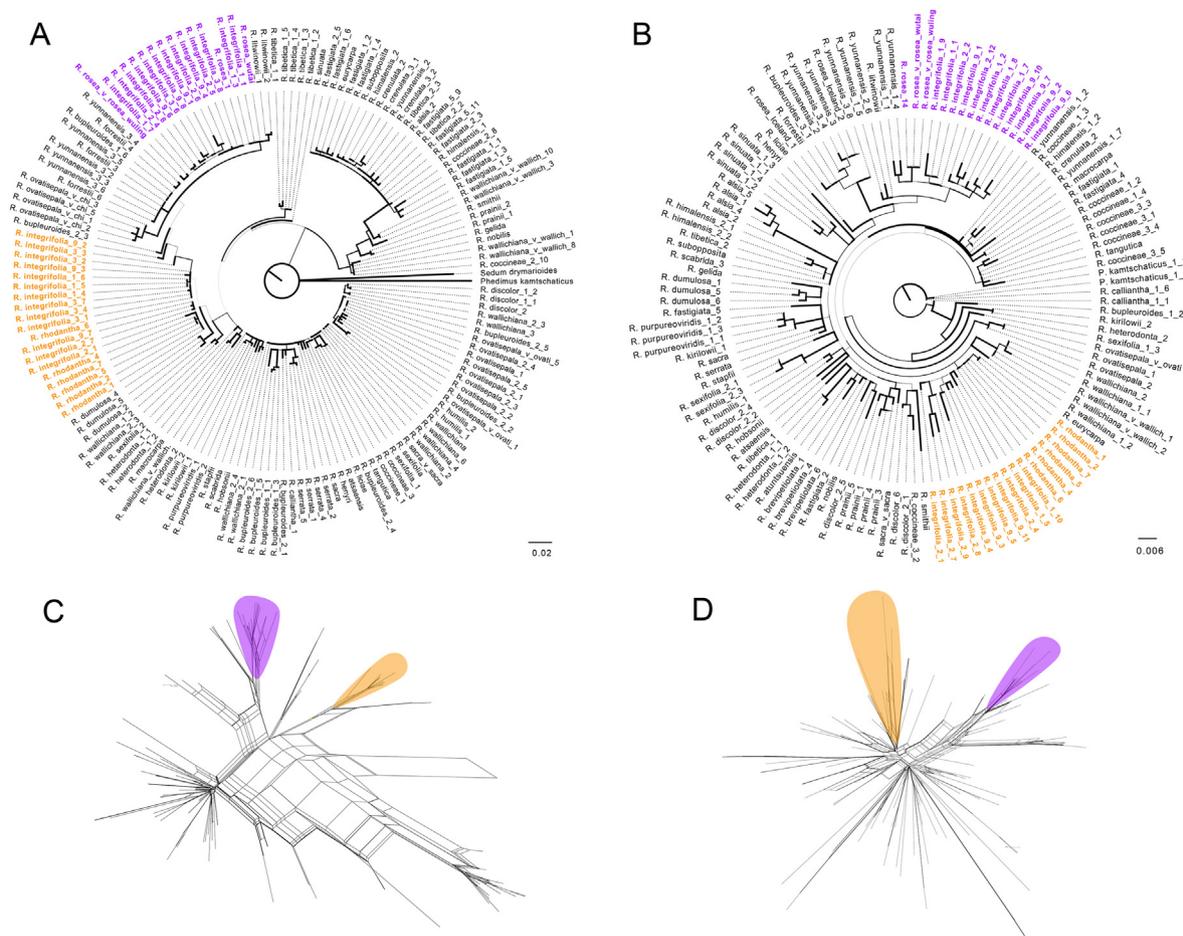
## 3. Results

### 3.1. Phylogenetic relationships

The aligned length of the *ncpGS* data set was 444 bp, with 174 variable sites and 100 parsimonious informative sites. The aligned length of the *rpb2* data set was 467 bp, and the number of variable and parsimonious sites were 163 and 95, respectively. All sequences obtained in this study have been deposited in the GenBank database (accession nos. OM976061–OM976334). The ML trees for both data sets are shown in Fig. 2A and B. In both the *ncpGS* and *rpb2* trees, alleles of *R. integrifolia* were split into two clusters, one with *R. rosea* and the other with *R. rhodantha* (Fig. 2A and B). The Neighbor-Net network results showed that there was substantial reticulate evolution or conflicting signals at the root of *Rhodiola*, typical in rapid radiations (Fig. 2C and D).

### 3.2. Divergence time

The likelihood ratio test showed that a constant molecular clock throughout the tree was rejected for the plastid data set ( $2\ln LR = 1440.2$ ,  $df = 14$ ,  $P < 0.01$ ). Thus, we used a Bayesian relaxed clock uncorrelated lognormal method in the BEAST analysis. The



**Fig. 2.** Phylogenetic trees and networks of *Rhodiola* species. The maximum likelihood (ML) phylogenetic trees based on the *ncpGS* data set (A) and the *rpb2* data set (B) reconstructed by RAxML v.8.2.10 are shown. Phylogenetic network of *Rhodiola* species reconstructed using Splitstree4 based on *ncpGS* (C) and *rpb2* (D) data sets are also shown. In A and B, branch width indicates bootstrap value, and bold lines have higher bootstrap value.

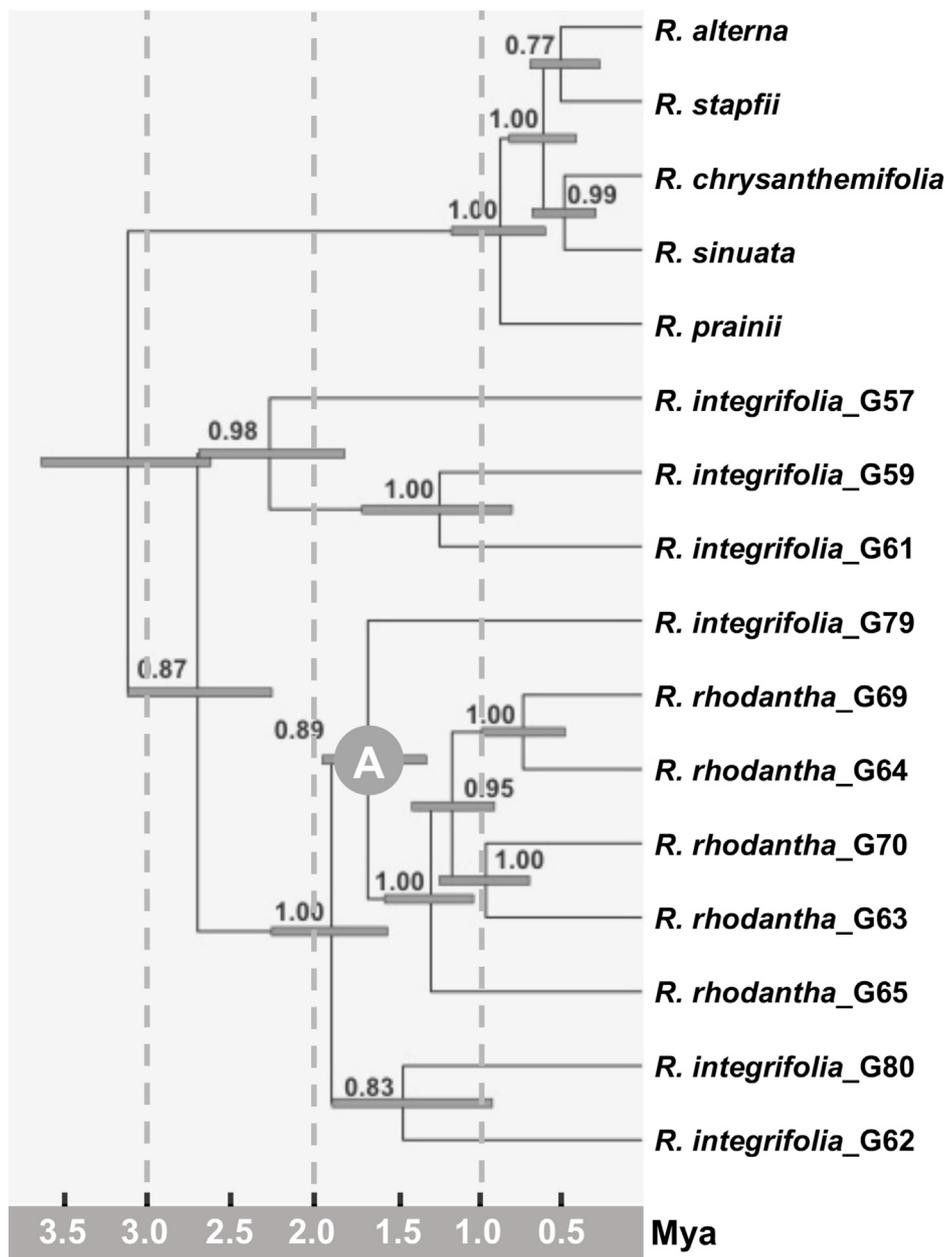
divergence of sequences between *R. integrifolia* and *R. rhodantha* was estimated at 1.67 Mya (Million years ago) (95% HPD: 1.35–1.94 Mya; Fig. 3, node A), in the Pleistocene.

### 3.3. Niche-modeling analyses

Our ENMs had high predictive ability, with the AUCs all above 0.95 and the standard deviation less than 0.01. The predicted potential distributions of *R. rosea*, *R. rhodantha* and *R. integrifolia* during the LIG, LGM and at the current time are shown in Fig. 1. In our model of the current time scenario, *R. rosea* occurs across the Bering Strait, overlapping with *R. integrifolia* (Fig. 1E and F, green color). However, *R. rosea* is currently not distributed in this area. Our model also indicated that the distribution range of *R. rosea* was lightly smaller during the LIG than at present, while that of *R. integrifolia* was comparable to its current distribution (Fig. 1A and 1B). Potential distribution areas of *R. rhodantha* during the LIG were not identified. In our model of the LGM, *R. rosea* migrated to the south both in North America and Europe as expected. In contrast, the distribution area of *R. integrifolia* expanded, especially in Alaska (Fig. 1C and D). The distribution area of *R. rhodantha* also expanded compared to its current distribution. Notably, in our model *R. rosea* had a possible distribution range along the west coast of North America in the LIG and the LGM.

The first three components of the principal component analysis explain 80.9% of the total variance in the climate data set. PC1 described 40.3% of the variance, and the greatest contributions were from BIO4 (Temperature seasonality, standard deviation × 100), BIO9 (Mean temperature of driest quarter), BIO19 (Precipitation of coldest quarter); PC2 explained 23.6% of the total variance and BIO1 (Annual mean temperature) and BIO2 (Mean diurnal range) were the most important contributors; PC3 described another 17% of the total variation, and the greatest contributions were from BIO18 (Precipitation of warmest quarter) (Table S4; Figs. 4 and 5). If we consider PC1 and PC2, the climate niche overlap between *R. integrifolia* and *R. rosea* was 54.0% (Schoener's *D* = 0.54; Table 1), that between *R. integrifolia* and *R. rhodantha* was 31.0%, and niche overlap between the two progenitor species was 30.0%. The corresponding number based on PC1 and PC3 is shown in Table 1. Schoener's *D* based tests of equivalency showed that all comparisons were not equivalent (Table 1). In the similarity tests, the niches of *R. integrifolia* and *R. rosea* were more similar than expected by chance, whereas in other comparisons, the niches compared were not more similar than expected by chance.

Analyses of niche optimum and breadth showed that along the PC1 axis, *R. integrifolia* niche was wider than those of *R. rosea* and *R. rhodantha*, but the breadth along the PC2 and PC3 axes were not. Optimums along PC1, PC2 and PC3 among the three species were all significantly different (one-way ANOVA, *P* < 0.01; Fig. 6).



**Fig. 3.** Divergence time of plastid sequences between *Rhodiola integrifolia* and *R. rhodantha* estimated with BEAST v.1.7.1 (Drummond et al., 2012). Grey bars on each node show 95% highest posterior density intervals, and numbers on each branch represent posterior probability of the clade.

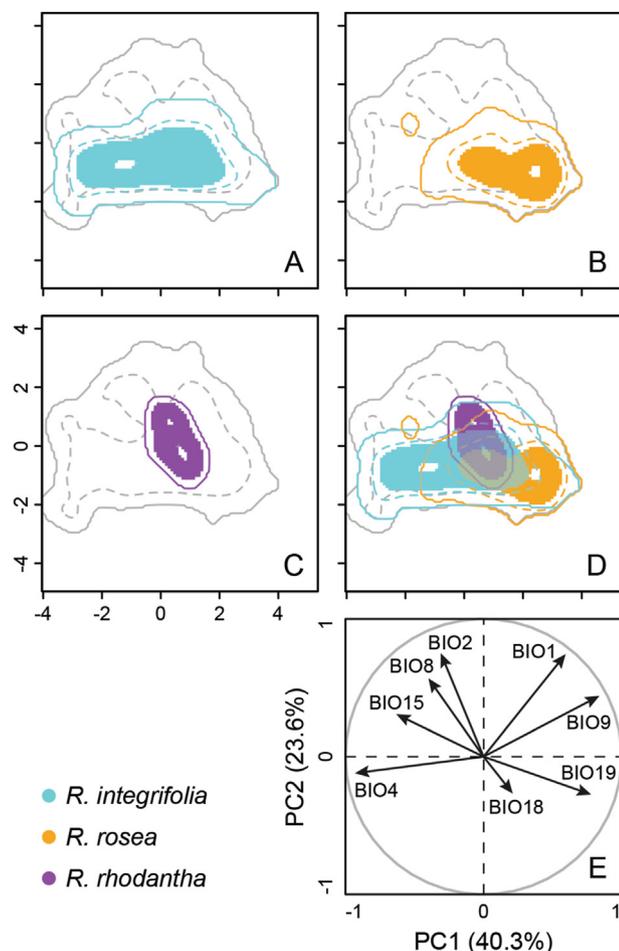
## 4. Discussion

### 4.1. Hybrid origin of *Rhodiola integrifolia*

Our data are in congruence with the hypothesis proposed by Uhl (1952) and partially verified by Hermsmeier et al. (2012) that *R. integrifolia* is an allopolyploid species between *R. rosea* and *R. rhodantha*. Using low-copy nuclear genes *ncpGS* and *rpb2*, we show clearly that *R. integrifolia* possesses alleles from both *R. rhodantha* and *R. rosea*, confirming the hybrid hypothesis. For both markers, we obtained multiple sequences (more than four) from individual plants of *R. integrifolia* as in Hermsmeier et al. (2012). These additional sequences may be recombinants caused by PCR recombination (Cronn et al., 2002). Another possibility is the template DNA was variable due to endopolyploidy, which is

common in Crassulaceae and other succulent plants (Barow, 2006). Nevertheless, two groups of *R. integrifolia* alleles were evident in both phylogenetic trees (Fig. 2A and B), one clustered with *R. rosea* and the other with *R. rhodantha*.

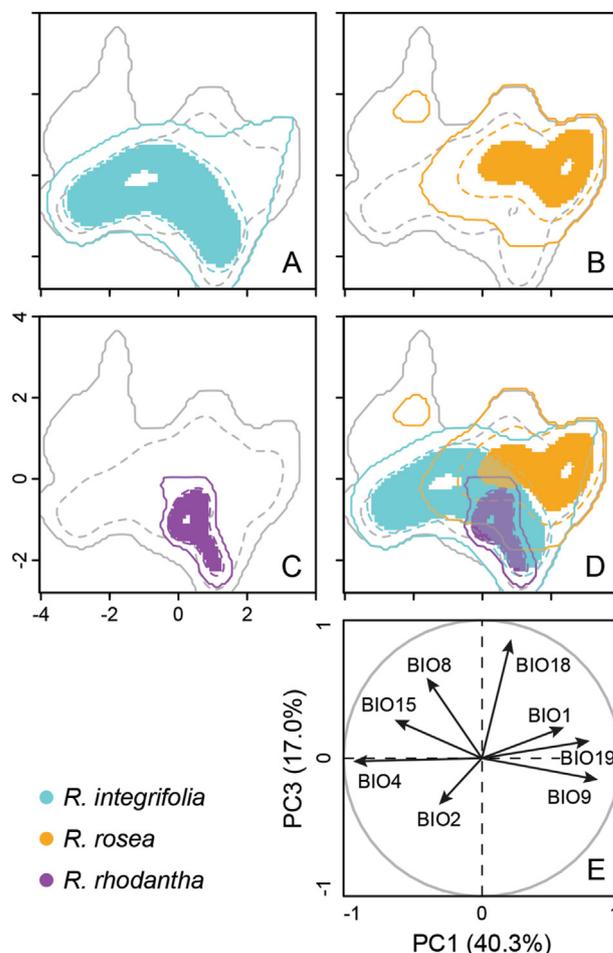
Although we tried to sample as many species as candidate parent species, our sampling was not exhaustive. For example, a previous phylogenetic study based on ITS data showed that *Rhodiola algida* (Ledeb.) Fisch. et Mey., a species distributed in the Altai Mountains, formed a clade with *R. rhodantha* + *R. integrifolia* (Zhang et al., 2014a). Therefore, it is possible that *R. algida* also contributed in the origin of *R. integrifolia*. Another caveat is that the two nuclear markers used in our study both contain limited phylogenetic information. The average pairwise genetic distance is relatively low (*ncpGS*, 0.04; *rpb2*, 0.02). Moreover, as *Rhodiola* may have undergone a rapid radiation, incomplete lineage sorting and



**Fig. 4.** Niche dynamics observed comparing *Rhodiola integrifolia* ( $n = 272$ ), *R. rosea* ( $n = 351$ ) and *R. rhodantha* ( $n = 34$ ) based on PC1 and PC2. (A) Niche model of *R. integrifolia*; (B) Niche model of *R. rosea*; (C) Niche model of *R. rhodantha*; (D) Overlap among the three species; (E) Correlation circle of WorldClim variables used in the PCA-env. In A, B, C, and D, continuous lines demarcate the full environmental space available within the background area or occupied by the species; dashed lines demarcate the 75th percentile of the environmental space available within the background area or occupied by the species. White dots represent niche centroids.

gene flow could also cause gene tree incongruities. For these reasons, accessions of other species sometimes formed a clade with *R. integrifolia* and *R. rosea* or *R. rhodantha*, e.g., *R. litwinowii* in Fig. 2A and *R. yunnanensis* in Fig. 2B. It is possible that these species also participated in the origin of *R. integrifolia*. However, considering the distribution pattern (both *R. litwinowii* and *R. yunnanensis* are confined to the Qinghai-Tibet Plateau and its adjacent area), this phylogenetic pattern might more likely have been caused by limited information and/or ancestor polymorphism. In addition, we only included limited individuals for species other than *R. rosea*, *R. rhodantha* and *R. integrifolia*; thus, the allelic variation of other species remains insufficiently known. Moreover, it is possible that other species may also have contributed to the origin of *R. integrifolia*. That said, to completely exclude contributions from other species, more markers with higher resolution are needed.

A previous phylogenetic study using both plastid markers and ITS data showed that in the ITS tree, *R. integrifolia* is most closely related to *R. rhodantha* (Zhang et al., 2014a). The reason for this pattern may be that concerted evolution has operated to maintain only one ITS parental lineage. For several decades, ITS sequences have been the most popular nuclear markers for phylogenetic



**Fig. 5.** Niche dynamics observed comparing *Rhodiola integrifolia* ( $n = 272$ ), *R. rosea* ( $n = 351$ ) and *R. rhodantha* ( $n = 34$ ) based on PC1 and PC3. (A) Niche model of *R. integrifolia*; (B) Niche model of *R. rosea*; (C) Niche model of *R. rhodantha*; (D) Overlap among the three species; (E) Correlation circle of WorldClim variables used in the PCA-env. In A, B, C, and D, continuous lines demarcate the full environmental space available within the background area or occupied by the species; dashed lines demarcate the 75th percentile of the environmental space available within the background area or occupied by the species. White dots represent niche centroids.

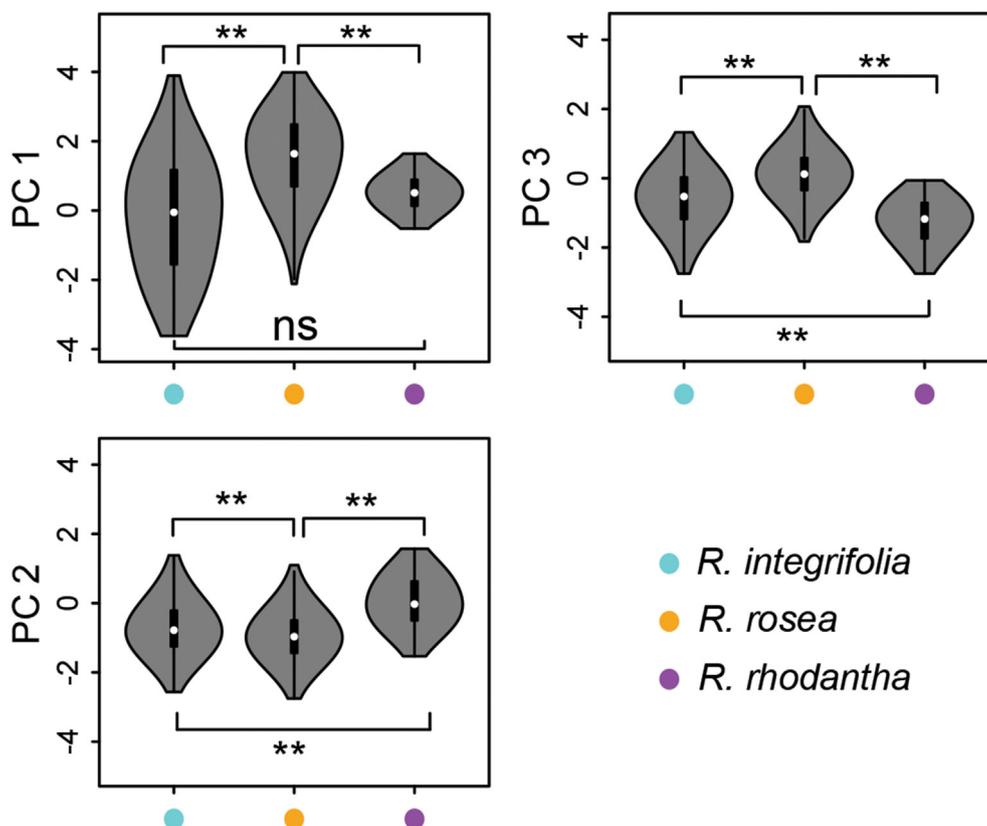
studies of a wide range of plant taxa (Álvarez and Wendel, 2003; Feliner and Rosselló, 2007). However, several potential problems have been raised with using ITS data in phylogenetic studies, including paralogs, pseudogenes and concerted evolution (may occur in other multiple-copy gene regions as well) (Arnheim, 1983). In hybrids and introgressants in which different copies from parent species have not been homogenized by concerted evolution, ITS data may be useful in identifying progenitors. Nonetheless, in cases where concerted evolution has occurred, the speed and direction of homogenization is hard to predict and is not consistent across different descendant lineages (Álvarez and Wendel, 2003). In the present study, concerted evolution of ITS copies in *R. integrifolia* were in the direction of *R. rhodantha*, the maternal parent. Interestingly, *R. integrifolia* is more similar to *R. rosea* in morphology. For a long time, *R. integrifolia* was treated as conspecific with *R. rosea* (Uhl, 1952; Clausen, 1975; Cody, 2000), as they are both dioecious, and the inflorescences are flat-topped cymes. Intriguingly, *R. integrifolia* inherited the sexual system of its paternal progenitor, *R. rosea* (dioecy). As the genetic mechanism controlling the sexual system in *Rhodiola* is unknown, it is not clear if this is a general pattern or a special case.

**Table 1**  
Results of Schoener’s *D*, equivalency test, and similarity test of *Rhodiola integrifolia*, *R. rosea* and *R. rhodantha*.

Compared species	PC 1&2 model			PC 1&3 model		
	Schoener’s <i>D</i>	Equivalency	Similarity	Schoener’s <i>D</i>	Equivalency	Similarity
<i>R. integrifolia</i> vs. <i>R. rosea</i>	0.54	<i>P</i> < 0.01	<i>P</i> < 0.01	0.39	<i>P</i> < 0.01	<i>P</i> = 0.37
<i>R. integrifolia</i> vs. <i>R. rhodantha</i>	0.31	<i>P</i> < 0.01	<i>P</i> = 0.24	0.26	<i>P</i> < 0.01	<i>P</i> = 0.15
<i>R. rosea</i> vs. <i>R. rhodantha</i>	0.30	<i>P</i> < 0.01	<i>P</i> = 0.39	0.11	<i>P</i> < 0.01	<i>P</i> = 0.40

Notably, the current distribution of *R. rhodantha* and *R. rosea* does not overlap. Thus, it is possible that one or both species were more widespread at some previous time point. Dating analysis in the present study showed that the hybridization event that led to *R. integrifolia* occurred at ca. 1.67 Mya (95%HPD: 1.35–1.94 Mya; Fig. 3, node A), in the Pleistocene. At this time, both *R. rosea* and *R. rhodantha* could have been present in Beringia and hybridized. For *R. rosea*, our niche modeling analyses clearly showed that during the LGM, there might be suitable habitats in Alaska and along the west coast of Canada that overlapped with *R. integrifolia*’s potential distribution area (Fig. 1C and D). Zhao et al. (2021) used plastome data to reconstruct the spatio-temporal dynamics of *Rhodiola* and showed that *R. rhodantha* formed a cluster with *R. dumulosa*, a species widespread in northern China, and that they diverged at around 3.5 Mya. This result echoed the conclusion of Zhang et al. (2014a) and Guest and Allen (2014) that *R. rhodantha* arrived in North America via the Bering Land Bridge, after the divergence from *R. dumulosa*. Based on the above evidence, we propose that as the climate oscillated in the Pleistocene (1.67 Mya),

*R. rosea* had expanded its distribution, and became sympatric with *R. rhodantha*, then hybridized either in eastern Siberia or the northwestern coast of North America. It is also possible that *R. rosea* hybridized with the ancestor of the clade comprising *R. rhodantha* and *R. dumulosa* (Zhao et al., 2021), probably in northeastern China and eastern Siberia. Following the hybridization event, climate oscillations caused a contraction of *R. rhodantha* to its current distribution area, and *R. integrifolia* expanded to its current distribution. High genetic diversity for *R. integrifolia* in Alaska and the Yukon (Guest and Allen, 2014) also support the idea that *R. integrifolia* may have originated in Beringia and dispersed multiple times to southern and eastern sites. However, more data, especially genomic data, are needed to reconstruct the detailed evolutionary history of these three species. We note that in our ENM analysis, there was no suitable distribution area for *R. rhodantha* during the LIG. Because *R. rhodantha* diverged from *R. dumulosa* much earlier than the LIG, it is less likely that *R. rhodantha* did not exist at the LIG. Instead, *R. rhodantha*, which is a cold-adapted species, may have been distributed in very small



**Fig. 6.** Principal components values along the three PCA-env axes. Niche optimum and breadth correspond to the median and the length of the 95% inter-percentile interval along the three PCA-env axes respectively.

refugia in mountainous areas during the LIG, and these regions went undetected in our ENM analysis.

#### 4.2. Niche expansion of *Rhodiola integrifolia* compared to its progenitors

The biological species concept (Mayr, 2000) acknowledges that polyploidy after hybridization between species may lead to immediate post-zygotic isolation and cause the saltatory origin of new species. However, polyploid lineages must overcome competition and costly hybridization with its diploid progenitors to persist. The niche shift hypothesis posits that successfully established polyploid lineages must have a divergent ecological niche (Levin, 1975; Fowler and Levin, 1984). A previous study on *Alyssum montanum* L. (Brassicaceae) supports the niche shift hypothesis and shows that allopolyploids expanded their niche compared to their diploid congeners (Arrigo et al., 2016). Our work also supports the hypothesis by showing that the niche of a tetraploid species differs from that of its progenitors in both niche breadth and optimum. Using Schoener's *D* as a summary statistic for niche overlap, the niche equivalency hypothesis was rejected for all comparisons, whereas in the similarity test, niches of *R. integrifolia* and *R. rosea* were more similar than expected by chance. These tests showed that the niche of *R. integrifolia* is not conserved, although it is statistically 'similar' with the niche of *R. rosea*. The significantly different niche between *R. integrifolia* and *R. rhodantha* is also evident in their sympatric areas: *R. rhodantha* grows on wet meadows and lake margins, whereas *R. integrifolia* grows on dry rock sites.

Kadereit (2015) showed that species formed by homoploidy and allopolyploidy are often displaced by the parental taxa. The author suggested that the reason for this phenomenon is that, as hybrids obtained evolutionary novelty by hybridization (Abbott et al., 2013), they would be more likely to adapt to changed environmental conditions after their parents retreated from areas where they were formerly in contact and hybridized. Due to the marked effects that genome duplication has on modifying genome structure and gene expression, considerable genetic novelty may be generated in the early stage of allopolyploid speciation. Our ENM analyses show that *R. integrifolia*'s niche has significantly expanded in PC1, in which BIO4 (Temperature seasonality) contributed the most, indicating that *R. integrifolia* may be more tolerant to temperature oscillations than *R. rosea*. In other words, *R. integrifolia*'s niche is statistically 'similar' to that of *R. rosea*, but with a wider breadth. This may be the reason why *R. integrifolia* does not co-occur with its paternal progenitor *R. rosea* in eastern Siberia, Alaska and western North America.

As our ENMs only incorporated 19 climatic factors, it is possible that we missed some important axes of ecological variation, thus underestimating the divergence of *R. integrifolia* niches. For example, soil type, phenology, or biotic interactions might also be important. Thus, *R. integrifolia* may have diverged in other aspects to compete or coexist with *R. rosea*. More detailed environmental and phenological data are needed to fully elucidate niche overlap and divergence of allopolyploids from their progenitors.

#### Author contributions

JQZ conceived and designed the study. DLZ conducted the ecological niche analysis, YCL performed molecular experiments and sequencing, and analyzed sequences with JQZ. JQZ wrote the

manuscript. All authors reviewed and approved the final submission.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors thank Dr. Jianwen Zhang and the five anonymous reviewers for their constructive suggestions, which greatly improved the original manuscript. We thank Profs. Geraldine Allen and Jun Wen for providing samples from North America. This work has been supported by the Fundamental Research Funds for the Central University of Shaanxi Normal University (GK202103077 to J.Q. Zhang) and the National Natural Science Foundation of China (Grant nos. 31870194, 32070236).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pld.2022.08.004>.

#### References

- Abbott, R., Albach, D., Ansell, S., et al., 2013. Hybridization and speciation. *J. Evol. Biol.* 26, 229–246.
- Akiyama, R., Sun, J., Hatakeyama, M., et al., 2021. Fine-scale empirical data on niche divergence and homeolog expression patterns in an allopolyploid and its diploid progenitor species. *New Phytol.* 229, 3587–3601.
- Arnheim, N., 1983. Concerted evolution of multigene families. In: Nei, M., Koehn, R. (Eds.), *Evolution of Genes and Proteins*. Sinauer, Sunderland, pp. 38–61.
- Arrigo, N., De La Harpe, M., Litsios, G., et al., 2016. Is hybridization driving the evolution of climatic niche in *Alyssum montanum*? *Am. J. Bot.* 103, 1348–1357.
- Álvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* 29, 417–434.
- Barow, M., 2006. Endopolyploidy in seed plants. *Bioessays* 28, 271–281.
- Broennimann, O., Fitzpatrick, M.C., Pearman, P.B., et al., 2012. Measuring ecological niche overlap from occurrence and spatial environmental data. *Global Ecol. Biogeogr.* 21, 481–497.
- Casazza, G., Boucher, F.C., Minuto, L., et al., 2017. Do floral and niche shifts favour the establishment and persistence of newly arisen polyploids? A case study in an Alpine primrose. *Ann. Bot.* 119, 81–93.
- Clausen, R.T., 1975. *Sedum* of North America North of the Mexican Plateau. Cornell University Press, Ithaca.
- Cody, W.J., 2000. *Flora of the Yukon Territory*, second ed. NRC Research Press, Ottawa.
- Cronn, R., Cedroni, M., Haselkorn, T., et al., 2002. PCR-mediated recombination in amplification products derived from polyploidy cotton. *Theor. Appl. Genet.* 104, 482–489.
- Darriba, D., Taboada, G.L., Doallo, R., et al., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772.
- Di Cola, V., Broennimann, O., Petitpierre, B., et al., 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 40, 774–787.
- Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19, 11–15.
- Drummond, A.J., Suchard, M.A., Xie, D., et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–1797.
- Elith, J., Graham, C.H., Anderson, R.P., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Feliner, G.N., Rosselló, J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol. Phylogenet. Evol.* 44, 911–919.
- Ficetola, G.F., Stöck, M., 2016. Do hybrid-origin polyploid amphibians occupy transgressive or intermediate ecological niches compared to their diploid ancestors? *J. Biogeogr.* 43, 703–715.
- Fowler, N.L., Levin, D.A., 1984. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *Am. Nat.* 124, 703–711.
- Fu, K.T., Ohba, H., 2001. *Crassulaceae*. In: Wu, C.Y., Raven, P.H. (Eds.), *Flora of China*, 8. Science Press, Beijing, pp. 202–268.
- Grant, V., 1981. *Plant Speciation*. Columbia University Press, New York.

- Guest, H.J., Allen, G.A., 2014. Geographical origins of North American *Rhodiola* (Crassulaceae) and phylogeography of the western roseroot, *Rhodiola integrifolia*. *J. Biogeogr.* 41, 1070–1080.
- Han, T.S., Hu, Z.Y., Du, Z.Q., et al., 2022. Adaptive responses drive the success of polyploid yellowcresses (*Rorippa*, Brassicaceae) in the Hengduan Mountains, a temperate biodiversity hotspot. *Plant Divers.* 44, 455–467.
- Hasumi, H., Emori, S., 2004. K-1 Coupled GCM (MIROC) Description. Center for Climate System Research, Univ. of Tokyo, Tokyo, Japan.
- Hermesmeier, U., Grann, J., Plescher, A., 2012. *Rhodiola integrifolia*: hybrid origin and Asian relatives. *Botany* 90, 1186–1190.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., et al., 2005. Very high-resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Huson, D.H., Klopper, T., Bryant, D., 2008. SplitsTree 4.0-Computation of phylogenetic trees and networks. *Bioinformatics* 14, 68–73.
- Kadereit, J.W., 2015. The geography of hybrid speciation in plants. *Taxon* 64, 673–687.
- Kelly, L.J., Leitch, A.R., Clarkson, J.J., et al., 2013. Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section *Suaevolentes*). *Evolution* 67, 80–94.
- Levin, D.A., 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24, 35–43.
- Levin, D.A., 2002. *The Role of Chromosomal Change in Plant Evolution*. Oxford University Press, New York.
- Li, Y.C., Wen, J., Ren, Y., et al., 2019. From seven to three: integrative species delimitation supports major reduction in species number in *Rhodiola* section *Trifida* (Crassulaceae) on the Qinghai-Tibetan Plateau. *Taxon* 68, 268–279.
- López-Alvarez, D., Manzaneda, A.J., Rey, P.J., et al., 2015. Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. *Am. J. Bot.* 102, 1073–1088.
- Mallet, J., 2007. Hybrid speciation. *Nature* 446, 279.
- Marchant, D.B., Soltis, D.E., Soltis, P.S., 2016. Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. *New Phytol.* 212, 708–718.
- Mayr, E., 2000. *The Biological Species Concept. Species Concepts and Phylogenetic Theory: a Debate*. Columbia University Press, New York, pp. 17–29.
- Molina-Henao, Y.F., Hopkins, R., 2019. Autopolyploid lineage shows climatic niche expansion but not divergence in *Arabidopsis arenosa*. *Am. J. Bot.* 106, 61–70.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Rambaut, A., Drummond, A.J., Xie, D., et al., 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67, 901–904.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Robertson, A., Rich, T.C., Allen, A.M., et al., 2010. Hybridization and polyploidy as drivers of continuing evolution and speciation in *Sorbus*. *Mol. Ecol.* 19, 1675–1690.
- Schoener, T.W., 1968. Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 49, 704–726.
- Soltis, P.S., Soltis, D.E., 2009. The role of hybridization in plant speciation. *Annu. Rev. Plant Biol.* 60, 561–588.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Stebbins, G.L., 1950. *Variation and Evolution in Plants*. Columbia University Press, New York.
- Uhl, C.H., 1952. Heteroploidy in *Sedum rosea* (L.) Scop. *Evolution* 6, 81–86.
- Wang, A., Melton, A.E., Soltis, D.E., et al., 2022. Potential distributional shifts in North America of allelopathic invasive plant species under climate change models. *Plant Divers.* 44, 11–19.
- Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868–2883.
- Wood, T.E., Takebayashi, N., Barker, et al., 2009. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13875–13879.
- Yamane, K., Yano, K., Kawahara, T., 2006. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Res.* 13, 197–204.
- Zhang, J.Q., Meng, S.Y., Allen, G.A., et al., 2014a. Rapid radiation and dispersal out of the Qinghai-Tibetan Plateau of an alpine plant lineage *Rhodiola* (Crassulaceae). *Mol. Phylogenet. Evol.* 77, 147–158.
- Zhang, J.Q., Meng, S.Y., Wen, J., et al., 2014b. Phylogenetic relationships and character evolution of *Rhodiola* (Crassulaceae) based on nuclear ribosomal ITS and plastid *trnL-F* and *psbA-trnH* sequences. *Syst. Bot.* 39, 441–451.
- Zhao, D.N., Ren, C.Q., Zhang, J.Q., 2021. Can plastome data resolve recent radiations? *Rhodiola* (Crassulaceae) as a case study. *Bot. J. Linn. Soc.* 197, 513–526.
- Zimmer, E.A., Wen, J., 2013. Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol. Phylogenet. Evol.* 66, 539–550.