



## OPEN ACCESS

## EDITED BY

Francesco Sunseri,  
Mediterranea University of Reggio Calabria,  
Italy

## REVIEWED BY

Margret Veltman,  
Institut de Recherche Pour le Développement  
(IRD), France  
Antonio Mauceri,  
Mediterranea University of Reggio Calabria,  
Italy

## \*CORRESPONDENCE

Weibang Sun  
✉ wbsun@mail.kib.ac.cn

†These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 28 September 2023

ACCEPTED 08 January 2024

PUBLISHED 31 January 2024

## CITATION

Lin L, Cai L, Huang H, Ming S and Sun W  
(2024) Transcriptome data reveals the  
conservation genetics of *Cypripedium  
forrestii*, a plant species with extremely small  
populations endemic to Yunnan, China.  
*Front. Plant Sci.* 15:1303625.  
doi: 10.3389/fpls.2024.1303625

## COPYRIGHT

© 2024 Lin, Cai, Huang, Ming and Sun. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Transcriptome data reveals the conservation genetics of *Cypripedium forrestii*, a plant species with extremely small populations endemic to Yunnan, China

Liewen Lin<sup>1,2,3†</sup>, Lei Cai<sup>1,2†</sup>, Hua Huang<sup>4</sup>,  
Shengping Ming<sup>4</sup> and Weibang Sun<sup>1,2,4\*</sup>

<sup>1</sup>Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Populations, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China, <sup>2</sup>Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China, <sup>3</sup>School of Life Science, University of Chinese Academy of Sciences, Beijing, China, <sup>4</sup>Lijiang Alpine Botanic Garden/ Kunming Botanical Garden, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

The *Cypripedium forrestii* is an orchid species with extremely small populations (PSESP) in Yunnan, China. *C. forrestii* is range-restricted and less-studied than many orchid species, and it is exposed to various threats to its survival. We investigated its potential habitats and collected 52 samples from eight locations, as well as two outgroup species for reference. We developed genetic markers (SNPs) for *C. forrestii* based on transcriptome sequencing (RNA-seq) data, and analyzed the genetic diversity, population structure, gene flow and demographic history of *C. forrestii* in detail. *C. forrestii* is a taxonomically independent species to protect. We found that the genetic diversity of *C. forrestii* was very low ( $1.7e^{-4}$ ) compared with other endangered species. We identified three genetic clusters, and several populations with distinct genetic backgrounds. Most genetic diversity was found within sampling sites (87.87%) and genetic clusters (91.39%). Gene flow has been greatly limited over the most recent generations, probably due to geographical distance, historical climate change and habitat fragmentation. We also detected a severe bottleneck event brought about by the recent population constraints. These factors, together with its reproductive characteristics, contribute to the population fragmentation and low genetic diversity of *C. forrestii*. Based on our findings, we suggest an integrative conservation strategy to protect and recover the genetic diversity of *C. forrestii* and a further comprehensive study of its ecological traits in the future.

## KEYWORDS

*Cypripedium forrestii*, conservation genetics, RNA, plant species with extremely small populations (PSESP), orchid

## 1 Introduction

Species with extremely small populations may have difficulties in reproduction and regeneration (Kéry et al., 2000), as well as higher chances of extinction in a single incidental event (Ralls et al., 2018). Small and fragmented populations shaped by bottleneck events may also suffer aggravated inbreeding and the accumulation of deleterious mutations (Ma et al., 2022b). Moreover, species with small populations and/or restricted distributions may also have limited genetic diversity. The loss of genetic diversity due to genetic drift may jeopardize species' evolutionary potential and lead to outbreeding depression between populations (Spielman et al., 2004; Bijlsma and Loeschcke, 2012). Therefore, the protection of plant species with extremely small populations (PSESP) is urgent, and vital for the conservation of biodiversity (Ma et al., 2013; Sun et al., 2019a). The PSESP concept, which was initiated in Yunnan, China, proposes that emergency conservation of a species requires multidisciplinary study of the species in question, as well as concrete actions by different stakeholders (Sun et al., 2019a; Sun et al., 2019b). The PSESP project promotes initial field investigations and consultation with stakeholders, in order to provide guidelines to prioritize the conservation needs of various plant species, concentrating resources on the most desperate cases (Volis, 2016; Yang et al., 2020). By utilizing and coordinating the strength of government agents, researchers, botanic gardens, local residents and NGOs, the PSESP rescue plan (2010–2020) has achieved a number of conservation objectives (Crane, 2020; Cogoni et al., 2021).

*Cypripedium* L. (Orchidaceae), known as the lady's slipper orchids, is a terrestrial orchid genus with more than 50 species, primarily distributed in the temperate regions of the Northern Hemisphere (Cribb, 1997). More than half of the known species in this genus, 32 species, are native to China (Cribb, 1997; Chen, 2013), and all of the Chinese species of *Cypripedium* (except the relatively common *C. plectrochilum*) are now one of the National Key Protected Wild Plants of China. China has several endemic species of lady's slipper orchids, and these species tend to have limited distributions and small population sizes (Cribb, 1997; Chen, 2013). Currently, four particularly rare *Cypripedium* species are considered as PSESP for priority protection (Sun, 2021). The distribution patterns and population genetics of *Cypripedium* orchids are, as is the case for many orchids, limited by their relatively short seed dispersal distance (Brzosko et al., 2017; Kotlínek et al., 2020), low fruiting rate (Suetsugu and Fukushima, 2014; Gargiulo et al., 2021) and the heterogeneity of available microhabitats (Diez, 2007; Hollick et al., 2007; Jacquemyn et al., 2012; McCormick et al., 2012). Human activities, including over-collection for folk medicine and horticultural use, habitat destruction and disturbance also contribute to the decline of wild orchid populations (Brundrett, 2007; Brundrett, 2019). Moreover, genetic studies into various *Cypripedium* species suggest that habitat fragmentation may hamper gene flow between populations (Brzosko et al., 2002; Minasiewicz et al., 2018). This may impair potential resilience to environmental change or habitat loss, and may lead to a higher risk of extinction for these species (Brzosko et al., 2002; Izawa et al., 2007; Qian et al., 2014).

Some of the rarest *Cypripedium* species with extremely small populations, including our target species, *C. forrestii* P.J.Cribb (Figures 1A, B), have barely received any research attention. Previous studies have focused on its morphology (Cribb, 1997; Chen, 2013) and its phylogenetic position in the genus *Cypripedium* (Li et al., 2011). These studies suggested that its most closely related species are species from Sect. *Trigonopedia*, including the sympatric but relatively broadly distributed *C. bardolphianum* (Figure 1C) and *C. lichiangense* (Figure 1D). *C. forrestii* can be distinguished by its unique floral traits (Cribb, 1997; Chen, 2013) and leaf polymorphism. Two morphological forms of *C. forrestii* coexist in all populations, one having plain, unspotted green leaves (Figure 1A) and one with blackish spotted leaves (Figure 1B). As suggested by its Chinese common name, the 'Yulong slipper orchid' was first collected on Yulong Snow Mountain (YLSX) in Lijiang, northwest Yunnan, which, for a long time, was the only known population of this species. Our recent field surveys confirmed further populations in the nearby Shangri-la Tibetan Autonomous Prefecture (ZZ, NR, HBA and HBB, see Figures 1E–G). *C. forrestii* grows on banks in scrub and at the edge of conifer woodland, at c. 3000–3800 m. It produces small dark-colored flowers in July, which are deceptive to pollinators as no reward (e.g., nectar) is provided. It is a typical clonal plant, and its populations are composed of a series of similar elements, i.e., ramets. Threatened by anthropogenic disturbance, landslides and a low fruiting rate, there are only about 1000 ramets in total. Because of the tillering nature of this plant, there are likely to be fewer than 500 genetically distinct individuals in the wild. *C. forrestii* has been assessed as Critically Endangered by the International Union for Conservation of Nature due to its extremely narrow distribution and small population sizes.

The conservation of *C. forrestii* has been promoted since 2022. Preliminary conservation work has included field investigations, artificial pollination at Lijiang Alpine Botanic Garden, followed up by the germination of seeds and the aseptic culture of seedlings in Kunming Botanical Garden. However, its genetic diversity and population structure are still unknown, and a better understanding of the genetic background is essential for the further study and conservation of this PSESP (Sun et al., 2019a; Yang et al., 2020).

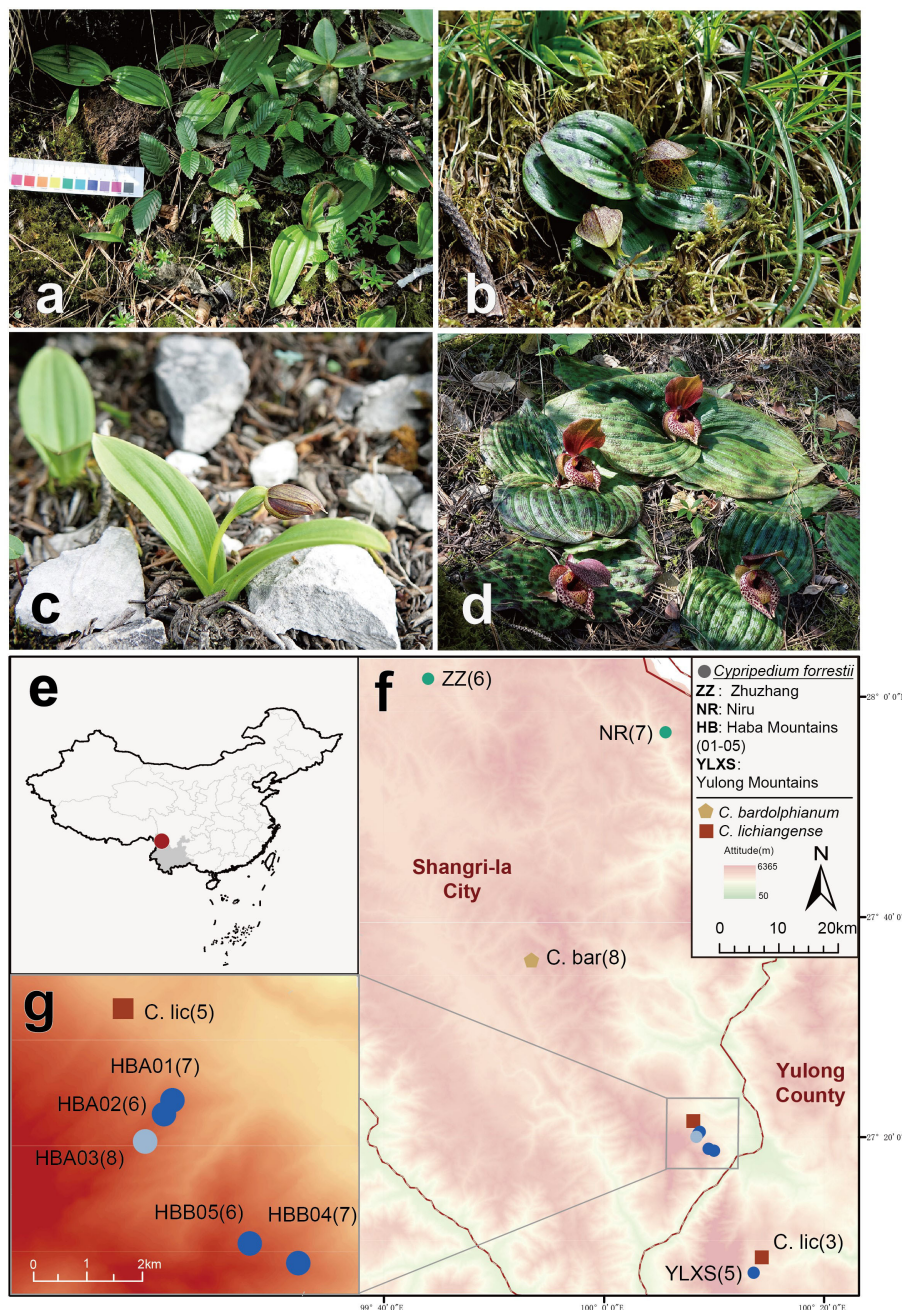
RNA-seq is now a powerful and cost-effective tool to tackle questions in evolutionary history, molecular ecology and conservation genetics (Alvarez et al., 2015; Tyagi et al., 2022), and we utilized transcriptome sequencing data as part of our investigation into the conservation genetics of *C. forrestii*. As reliable reference genome data is lacking for many species in need of conservation, including *Cypripedium* species, a transcriptome generated from RNA-seq data can be an economical and reliable alternative (Alvarez et al., 2015; Ma et al., 2019b). In this study, a full-length transcriptome and genetic marker (SNP) data were developed from RNA-seq data of *C. forrestii* to address the following objectives: (1) to clarify its population genetic structure; (2) to estimate the genetic diversity and genetic differentiation in different populations of this species; (3) to detect any bottlenecks arising from recent population declines; and (4) to evaluate gene flow among isolated habitats and small populations.

## 2 Methods

### 2.1 Sample collection and mRNA preparing

Samples of *C. forrestii* were collected from eight locations (Figures 1E, F), covering every location in China where this species is known to occur. Geographical coordinates and altitude information were recorded at all sample locations. 5-8 Leaf samples

from each *C. forrestii* population were collected in July, the peak growing season for this species. The distance between each sample was at least 10 m to ensure these ramets are genetically independent individuals, as its tillering tolns are mostly 30-50cm long. We avoided unnecessary damage and collected only the minimum amount of leaves from mature individuals to minimize our impact on these rare plants. Only fresh and undamaged leaves were sampled. All leaves were washed and then frozen immediately



**FIGURE 1** (A-D) Photographs of three related *Cypripedium* species. (A) *C. forrestii* individuals without spots on the leaves; (B) *C. forrestii* individual with spotted leaves; (C) *C. bardolphianum*; (D) *C. lichiangense*. (E, F) Map of sampling sites. (E) Position of sampling area in NW Yunnan, China; (F) Ten sampling sites of three *Cypripedium* species (*C. forrestii*, *C. bardolphianum* and *C. lichiangense* in Shangri-la and Yulong; (G) details of sampling site in the Haba Snow mountains. Numbers in parentheses are the sample sizes at each site. Different colored circles represent three genetic clusters of *C. forrestii* inferred in this study. C.bar and C. lic are the abbreviation of *C. bardolphianum* and *C. lichiangense*, respectively.

in liquid nitrogen. In total, 63 samples of *C. forrestii* were sequenced; 34 samples were from individuals with spotted leaves and 29 samples were from non-spotted plants. 11 samples (6 spotted and 5 non-spotted, 1-2 from each location) were used for long-read sequencing and a hybrid transcriptome assembly. 52 samples (28 spotted and 24 non-spotted, 5-8 from each location) were used for NGS sequencing (Table 1). Both morphotypes were collected in all locations. We also collected 8 samples from both outgroup species (*C. bardolphianum* and *C. lichiangense*) for further study (Figure 2). All outgroup samples were used for NGS sequencing.

The total RNA of each sample was extracted using RNAprep Pure Plant Kits in accordance with the manufacturer's instructions (Tiangen, Beijing, China). The mRNA was then purified using mRNA Capture Beads (oligo dT magnetic beads). A NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) was used to detect RNA purity (A260/280 and A260/230). A Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA) was used to measure RNA concentration. The integrity of RNA and the presence of DNA contamination were assessed using 1% agarose gel electrophoresis (200V, 10min). Samples were loaded with 400ng according to Nanodrop concentration measurement. Samples for long-read sequencing were pooled together in equimolar ratios, and were sequenced on the Pacific Bioscience RSII platform (Pacific Biosciences, Menlo Parks, CA, USA). The NGS sequencing was performed using a paired-end library with 150bp read length on an Illumina HiSeq 2000 instrument.

## 2.2 Processing of sequencing data

PacBio sequencing data was used for the hybrid assembly of full-length transcriptome. Raw reads were processed in accordance with the Iso-seq workflow recommended by Pacific Biosciences [https://iseq.how/clustering/high-level-workflow.html, (17 Feb 2023, date last accessed)]. In brief, raw reads were first input into

Pbccts v6.4.0 to generate consensus sequences (CCS). Next, primers were removed and the CCSs were demultiplexed using lima v2.6.0. Isoseq3 v3.8.0 was then used for the removal of poly(a) tails and for the clustering and polishing of the outputted isoforms. We obtained a raw reference transcriptome after completing the above steps.

The NGS reads from one sample (YL03, having the highest alignment rate with the Pacbio data) were used for the transcriptome refinement. Trimmomatic v0.39 (Bolger et al., 2014) was used to remove low-quality reads and adapters of NGS raw reads. FastQC v0.11.9 (Andrews, 2010) was then used to evaluate the quality of the trimmed reads. To obtain a high-quality reference transcriptome, the isoform outputs of PacBio data were further corrected using processed NGS reads in Pilon v1.24 (Walker et al., 2014) with the default parameters. After that, CD-HIT v4.8.1 was used to remove redundant isoforms (parameters: -c 0.99 -aS 0.99 -g 1 -d 0 -G 0). To minimize noise from contamination, sequences highly similar to rRNA sequences were removed (inclusion threshold of e-value < 0.01). rRNA sequences were download from the SILVA database (https://www.arb-silva.de). The viral, bacterial and fungal sequences were removed using the standard pipeline of DecontaMiner v1.4 (Sangiovanni et al., 2019) and its pre-built database (based on NCBI database, see https://github.com/amarinderthind/decontaminer). Finally, we obtained the first high-quality full-length transcriptome of *C. forrestii*, which we could then use as the reference sequence for this study.

## 2.3 SNP calling and filtering

Short read NGS data were used for SNP calling. We followed the GATK Best Practices Workflow to identify the SNPs in our RNAseq data [https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-, (17 Feb 2023, date last accessed)]. First, trimmed NGS reads from each sample were mapped to reference sequences using the two-

TABLE 1 Sampling information and genetic statistics of *C. forrestii* from eight sampling sites and three genetic groups.

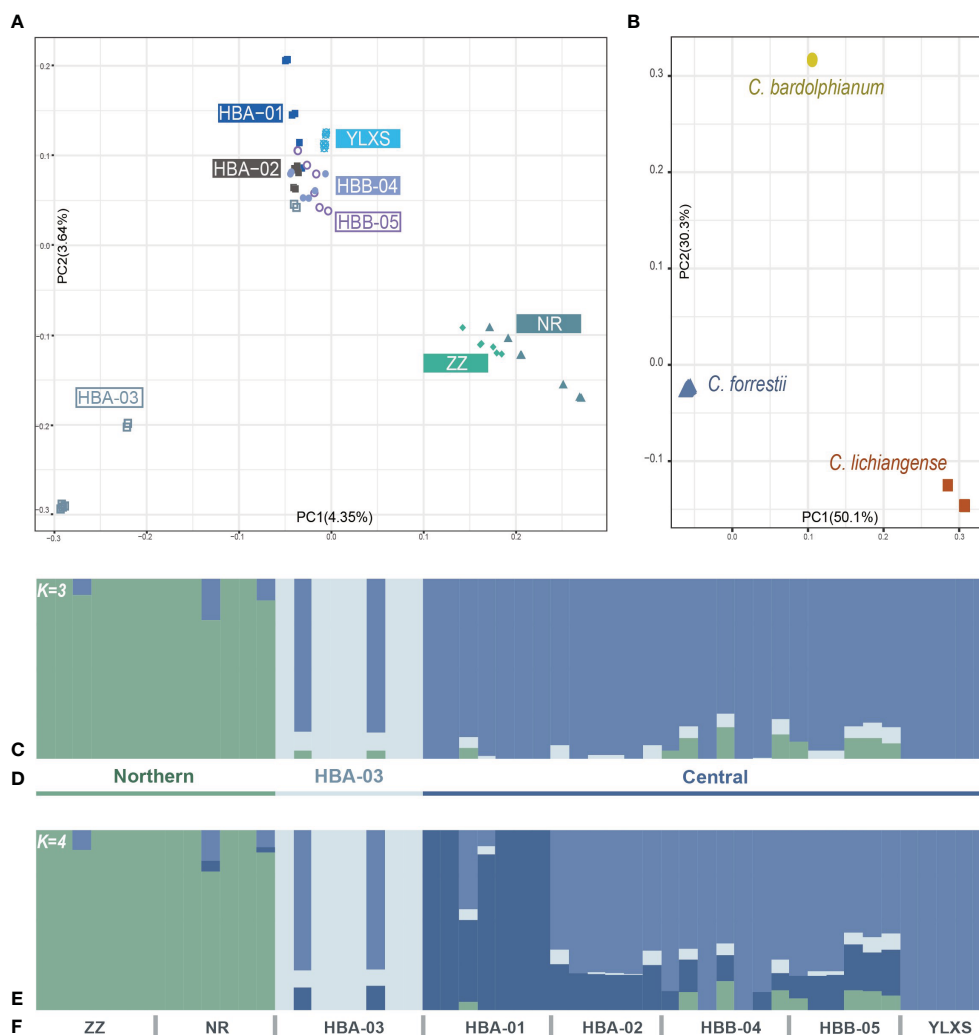
Genetic groups	Sampling locations	Number of ramets	Altitude (m)	Number of samples (NGS)	$\pi \times 10^4$	TajimaD	Ho	He	Fis	Ne (LD)	Ne (HE)	No (CO)
Northern				13	1.80	0.77	0.41	0.37	-0.11	4.3	12.2	3.8
	ZZ	50	3526	6	1.91	0.58	0.48	0.41	-0.17	27.8	4.9	5.3
	NR	50	3465	7	1.91	0.67	0.48	0.40	-0.18	3.1	4.6	4.5
Central				31	1.72	0.86	0.37	0.33	-0.12	15.5	10.8	5.7
	YLSX	50	2997	5	2.03	0.69	0.56	0.44	-0.28	Infinite	2.8	4.4
	HBA-01	200	3401	7	1.89	0.66	0.50	0.40	-0.26	2.2	3.1	4.0
	HBA-02	50	3576	6	2.02	0.73	0.50	0.42	-0.20	1.8	3.6	3.4
	HBB-04	200	3664	7	1.90	0.48	0.45	0.38	-0.17	Infinite	4.7	4.5
	HBB-05	200	3792	6	1.93	0.40	0.44	0.39	-0.12	Infinite	7.2	4.2
HBA-03	HBA-03	200	3835	8	1.89	0.79	0.53	0.40	-0.32	1.1	2.5	2.4
Overall		1000	-	52	1.71	1.05	0.35	0.32	-0.08 (Fit)	16.5	78	2.5

pass mode in STAR v 2.7.10b (Dobin et al., 2013). SAMBAMBA v0.6.6 (Tarasov et al., 2015) was used to filter out reads with mapping quality < 30, and then to sort and index the BAM files. A series of tools in GATK v4.1.9.0 were then applied to the raw mapped reads. GATK MarkDuplicates (Picard) was used to exclude duplicates and GATK SplitNCigarReads was employed to reformat alignments, after which we called SNPs using the default ploidy setting of GATK HaplotypeCaller, since *C. forrestii* and its sister species are diploid (unpublished data). The intermediate GVCFs generated by the ERC mode of HaplotypeCaller were then input into the GATK GenotypeGVCFs for joint calling. GATK SelectVariants was used to select raw SNPs and to remove indels from the joint VCF.

We then filtered the raw SNPs. Since it is not a model organism, we adopted the hard-filtering setting of the VariantFiltration tool as recommended by GATK [<https://gatk.broadinstitute.org/hc/en-us/articles/360035890471>, (17 Feb 2023, date last accessed)]: -filter-

expression “QD < 2.0 || MQ < 40.0 || FS > 30.0 || SOR > 3.0”. SNPs with >20% missing rate and minor allele frequency >0.05 were also excluded. We called the remaining data SNP Dataset1. In addition, we repeated the above-mentioned mapping, SNP calling and hard filtering procedures with the 16 outgroup samples added to the 52 *C. forrestii* samples, and obtained Dataset O.

As unlinked SNPs are required in some of the downstream analysis (e.g., ADMIXTURE and PCA), cluster SNPs were filtered out using GATK VariantFiltration (parameter: -cluster 3 -window 10). We also filtered out potentially linked SNPs with PLINK v1.90b6.21 (Purcell et al., 2007) (parameter: -indep-pairwise 50 5 0.5). Moreover, VCFtools v0.1.16 (Danecek et al., 2011) was used to retain only bi-allelic sites. SNP Dataset2 was thus derived from Dataset1. We also extracted the synonymous SNPs of four-fold degenerate third-codon transversion sites (4Dtv sites) from Dataset1 with the getCdsPep function in ReSeqTools (He et al., 2013). The output, Dataset 3, was used for the analysis of demography history.



**FIGURE 2**  
 (A, B) Population structure of *C. forrestii* inferred by PCA results. (A) 52 sampled *C. forrestii* individuals from eight sampling sites; (B) *C. forrestii* and its two sister species (*C. bardolphianum* and *C. lichiangense*); (C-F) Population structure inferred by Admixture results with (C)  $K = 3$  and (D) the putative three genetic groups; (E) second-best inference when  $K = 4$ ; (F) and the designation of each sampling sites.

## 2.4 Population structure and genetic diversity

Population structure was inferred using Principal Components Analysis (PCA) and ancestral population estimation based on Dataset 2. PCA was performed in PLINK v1.90b6.21 (Purcell et al., 2007) with the default parameters. The same method was employed to examine potential differentiation between spotted and non-spotted individuals. Ancestral population estimation was performed using ADMIXTURE v1.3.0 (Alexander et al., 2009). The optimal number of ancestral populations (K) was determined by the lowest cross-validation error of K values between 1 and 8. The results from the ADMIXTURE analysis were visualized in pophelper [<http://pophelper.com/>, (17 Feb 2023, date last accessed)].

The average nucleotide diversity ( $\pi$ ) and Tajima's D value per site were calculated for Dataset 1 and Dataset O, using VCFtools v0.1.16 (Danecek et al., 2011) with non-overlapping 1000bp windows. Heterozygosity indexes ( $H_O$  and  $H_E$ ), pairwise genetic differentiation ( $F_{ST}$ ) and inbreeding coefficient ( $F_{IS}$ ) of each sampling location and genetic group were calculated in Arlequin v3.5.2.2 (Excoffier et al., 2005) using Dataset 2. The same methods were employed to investigate differences between spotted and non-spotted individuals. An AMOVA analysis was also conducted in Arlequin to assess the genetic variance within and among sampling sites.

NeESTIMATOR v2.1 (Do et al., 2014) was used to estimate the contemporary effective population size ( $N_e$ ) using Dataset 2. The heterozygote excess method (HE), the molecular coancestry method (CO) and the linkage disequilibrium method (LD) were all employed. The cutoff was set based on the minor allele frequency within the following interval:  $1/(2n) \leq \text{PCRIT} \leq 1/n$ , where n is the number of samples (Waples and Do, 2010). Demographic trajectories of *C. forrestii* were inferred from Dataset 3 using Stairway Plot2 v2.1.1 (Liu and Fu, 2020). We estimated the mutation rate to be about  $3.9 \times 10^{-9}$  per site per year. The generation time was set as 3 years. The site frequency spectrum (SFS) was inferred from SNP data using the realsfs program implemented in ANGSD v0.938 (Korneliussen et al., 2014). The program returned no valid results when *C. bardolphianum* or *C. lichiangense* was used to infer the ancestral state of SFS, maybe because of the loss of low frequent polymorphism sites. We therefore calculated the folded SFS of all *C. forrestii* samples, since the ancestral state may not be correctly designated without an outgroup reference.

## 2.5 Evaluation of gene flow

The R package "ade4" was employed to test for the presence of isolation-by-distance (IBD) among sampling sites. Recent gene flow among sampling sites was evaluated with a Bayesian inference approach, using BayesAss (Wilson and Rannala, 2003) in BA3SNP v3.0.4 (Mussmann et al., 2019) based on Dataset 2, with  $1 \times 10^7$  iterations, a burn-in of  $1 \times 10^6$  steps and a sampling frequency of 1,000. We fine-tuned the two parameters (-a and -f) to

adjust acceptance rates for allele frequencies and inbreeding coefficients according to the official user manual (Mussmann et al., 2019).

Since results from BayesAss only represent gene flow among the most recent generations (the last 1–3 generations), Dataset 2 was also analyzed using TreeMix v. 1.12 (Pickrell and Pritchard, 2012). TreeMix requires an outgroup population setting. First, we rooted the graph with the outgroup species using dataset O. We then inferred the gene flow of *C. forrestii*, using its basal lineage suggested by the previous graphs. We constructed graphs of 0–10 migration edges (m), the Optimal m values and graphs were inferred using the *ad hoc* statistic ( $\Delta m$ ) and residuals were computed with OptM (Fitak, 2021).

## 3 Results

### 3.1 Sequencing and SNP calling

PacBio sequencing generated 29,383,943 subreads with a mean length of 1,438 bp and a mean quality of 40.1. The raw data of PacBio sequencing was used for a hybrid transcriptome assembly. The CCS process produced 402,442 reads with a mean length of 1,502 bp. After removing primers, demultiplexing, and poly(A) tail removal, 348,844 high-quality reads were used as input for Isoseq clustering, resulting in 32,297 reads. Two passes of error correction were conducted using the NGS sample with the highest alignment rate. The corrected reads had a mean length of 1,576 bp. After subsequently removing redundant isoforms, 15,553 transcripts remained. Any potentially contaminating sequences from rRNA and microbial genomes were then removed. The final reference transcriptome comprised 15,363 transcripts with a mean length of 1,662 bp.

The NGS sequencing of 52 *C. forrestii* samples generated an average of 62.8 million raw reads per sample. The length of the raw reads was 150 bp, and the average Q20, Q30 and GC ratios were 98.5%, 95.0% and 47.9%, respectively. The mean Phred score of nucleotides was 37. After the trimming process, an average of 29.2 million clean reads with a mean length of 135 bp were retained for each sample.

The clean NGS reads were mapped to this reference transcriptome. The mapping rates ranged from 83.87% to 89.49%, with an average of 86.35%. GATK identified 176,657 raw SNPs. After the hard-filtering using GATK, 44,970 SNPs were retained in Dataset 1. The mean depth per site averaged across all individuals was 118. The second round of filtering generated 20,973 bi-allelic and unlinked SNPs, which formed Dataset 2. After extraction of the 4Dtv sites from Dataset 1, a total of 7,876 SNPs were obtained for Dataset 3. Dataset O comprised 245,346 filtered SNPs.

### 3.2 Population structure and genetic diversity

The results of the PCA (Figure 2A) indicate that samples from the northern sites, ZZ and NR, form a distinct group (the Northern

genetic cluster). Most samples from HBA-03 are also strongly separated from samples from other sites (the HBA-03 genetic cluster). The remaining samples from the Haba and Yulong Mountains (HBA-01, 02, B04, 05, YLXS), which form the central part of the *C. forrestii* distribution area, are more closely related to each other and clustered together in one group (the Central genetic cluster). ADMIXTURE (Figures 2C, E) results also support that the optimal number of ancestral populations (K) is 3 (Supplementary Figure S1A), and demonstrate that most of the samples from HBA-03 and the two northern sites represented rather “unmixed” genetic backgrounds. The second-best estimation of K is 4 (Supplementary Figure S1A). When K=4 (Figure 2E), HBA-01 is differentiated from the other central populations, and samples from HBA-01 also stood apart from the other central samples in the PCA graph.

The fixation index values ( $F_{ST}$ ) also suggest that samples from the northern sites and from HBA-03 are genetically distinctive from other populations (Table 2), as the highest  $F_{ST}$  is found between ZZ and YLXS (0.183), follow by that between HBA-03 and YLXS (0.182). However, the  $F_{ST}$  values between YLXS and other populations are relatively high, even within the central genetic cluster (0.103 to 0.157 when compared with other central sites). Moreover, YLXS shows no genetic mixture with other genetic clusters in the ADMIXTURE clustering analysis, which suggests that YLXS has a unique genetic background. Meanwhile, other samples from the central group are relatively similar genetically, with pairwise  $F_{ST}$  values ranging from 0.029 to 0.096. The lowest pairwise  $F_{ST}$  is found between HBB-04 and HBB-05 (0.029), and the second lowest is between HBA-02 and HBB-05.

The individuals with spotted leaves are not significantly separated from those with non-spotted leaves in the PCA (Supplementary Figure S2). Pairwise  $F_{ST}$  between spotted and non-spotted individuals is 0.00379, much lower than the pairwise  $F_{ST}$  values between any of the sampling sites (Supplementary Table S2), demonstrating that the two leaf polymorphisms have many fewer genetic differences between them than any pair of sampling sites. AMOVA analysis reveal that the vast majority of genetic variation in *C. forrestii* is distributed within 8 sampling sites (87.87%) and within 3 genetic clusters (91.39%) (Supplementary Table S3).

PCA, ADMIXTURE and  $F_{ST}$  are also calculated based on Dataset O. The  $F_{ST}$  values between *C. forrestii* and *C. bardolphianum*, and *C. forrestii* and *C. lichiangense* are 0.91 and 0.86, respectively (Supplementary Table S1). Each species is distinct and well-separated in the PCA (Figure 2B) and ADMIXTURE

analyses. ADMIXTURE results do not reveal notable genetic mixture among the three species (Supplementary Figures S1C, D).

The nucleotide diversity ( $\pi$ ) of *C. forrestii* is  $1.71e^{-4}$  (Table 1). At the sampling site level, the diversity values are roughly similar. YLXS has the highest value ( $2.03e^{-4}$ ), follow by HBA-02 ( $2.02e^{-4}$ ), whereas the HBA-03 and HBA-01 have the lowest  $\pi$  values ( $1.89e^{-4}$ ). By comparison,  $\pi$  values for *C. bardolphianum* and *C. lichiangense* calculated from Dataset O are  $2.10e^{-4}$  and  $4.36e^{-4}$ , respectively (Supplementary Table S4).

The Tajima's D values at the site level range from 0.40 (HBB-05) to 0.79 (HBA-03), with an overall average of 1.05 (Table 1). The inbreeding coefficient among all sites,  $F_{IT}$ , is -0.08. The inbreeding coefficients within each site,  $F_{IS}$ , are all negative. The smallest value is found at HBA-03 (-0.32), while HBB-05 has the largest (-0.12). This is consistent with the estimation of contemporary effective population size ( $N_e$ , Table 1), i.e., HBA-03 has the smallest  $N_e$  of all the sampling sites. The  $N_e$  values vary greatly among the three methods used (HE, LD, CO, as mentioned in the methods section). The results of the LD method may be unreliable in this study, as the confidence intervals of several  $N_e$  values are outputted as “infinite”.

Stairway Plot2 reveals a significant drop in effective population size that happened about 1k years ago, and a second decline in the central group about 250 years ago (Supplementary Figure S5).

### 3.3 Limitation of gene flow

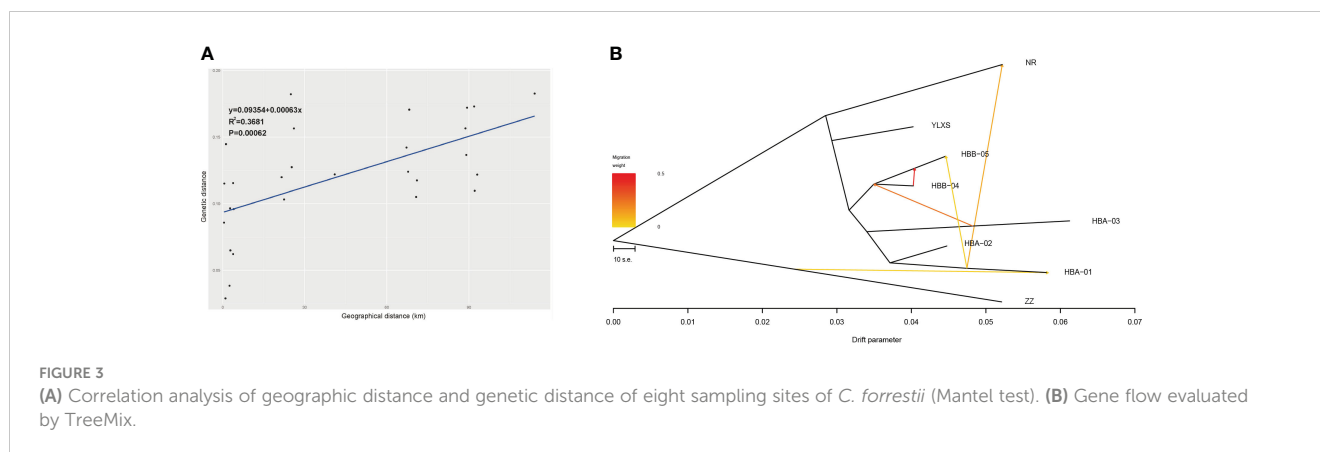
The Mantel test of IBD suggests that the divergence and gene flow of *C. forrestii* should be significantly limited by geographical distance ( $R^2 = 0.37$ ,  $P < 0.001$ , Figure 3A). However, geographical distance should not be the only constraint. As mentioned above,  $F_{ST}$  values between HBA-03 and other central sampling sites are relatively high, although some samples are collected only a few hundred meters away from each other. Some sites, such as HBA-03 and YLXS, should have lower gene flow to the others, because of the relatively high genetic distances between adjacent sites.

BayesAss analysis of recent gene flow (Supplementary Table S5) indicates that gene flow among most sampling sites should be scarce when standard errors are taken into account. Plausible gene flow is likely to occur within the central genetic clusters, especially between HBB-04 and HBB-05 (0.0471 and 0.0541). Moreover, HBB-05 seem to be a major sink for gene flow from nearby sites.

The NJ trees of Treemix suggest that ZZ (from the Northern cluster) is the basal lineage of *C. forrestii* when the root is set as

TABLE 2 Pairwise  $F_{ST}$  of *C. forrestii* between eight sampling sites.

	ZZ	NR	HBA-03	HBA-01	HBA-02	HBB-04	HBB-05	YLXS
ZZ	-							
NR	0.122	-						
HBA-03	0.172	0.171	-					
HBA-01	0.157	0.142	0.145	-				
HBA-02	0.137	0.124	0.115	0.086	-			
HBB-04	0.122	0.118	0.116	0.096	0.062	-		
HBB-05	0.110	0.105	0.097	0.065	0.039	0.029	-	
YLXS	0.183	0.173	0.182	0.157	0.127	0.120	0.103	-



*C. bardolphianum* or *C. lichiangense*. No gene flow is detected between these sister species (Supplementary Figure S3). When the outgroup species were included, Treemix fails to predict additional migration events within *C. forrestii* (except between NR and HBA-01). This could be due to the smaller intra-species differences, which are less detectable compared to the inter-species differences. Next, ZZ is set as the root to calculate the gene flow, resulting in a fairly stable topology. The number of predicted migration edges in TreeMix is 5 (Supplementary Table S6; Supplementary Figure S4), as it corresponds to the highest likelihood ( $\Delta m$  score) when  $m=5$ . In this model, Treemix suggests that previous gene flow or admixture events occurred among most sampling sites (Figure 3B). The strongest migration event occurred from HBB-05 into HBB-04, and migration between HBA-01 and NR was also detected. No gene flow was detected between YLXS and the other sampling sites. Besides the NR and ZZ populations, the HBA01 and 03 populations also had large drift parameters, indicating that they may be genetically well separated from the other populations.

## 4 Discussion

### 4.1 *C. forrestii* is a taxonomically independent species to protect

Recent studies reveal that the extremely small populations of endangered plant species may result from various factors, including over exploitation (Yu et al., 2021; Ma et al., 2022a), deforestation (Liu et al., 2019; Ma et al., 2022b; Yang et al., 2022) as well as from severe bottleneck events due to historical climate change (Ma et al., 2022b; Yang et al., 2022). The uplift of mountains and climate fluctuations may also facilitate inter-species hybridization (Ma et al., 2019b; Yang et al., 2023). Moreover, studies investigating *Primula* (Ma et al., 2019a) and *Cypripedium* (Hu et al., 2011) have suggested that asymmetric gene flow from sympatric aliens may put rare species at increased risk of extinction by decreasing reproductive fitness of parental species and inducing genetic swamp. Our genetic study tentatively rules out the possibility of a hybridization event between *C. forrestii* and *C. bardolphianum*, or between *C. forrestii* and *C. lichiangense*. *C. forrestii* should therefore be studied and protected as an independent species. In addition, the

individuals with and without spotted leaves are two morphotypes of one independent species. As suggested by the AMOVA analysis (Supplementary Table S3) and the pairwise  $F_{ST}$  values (Supplementary Table S4), most RNA-sequence variants occur within populations and genetic clusters, instead of different leaf morphs.

### 4.2 Relatively low genetic diversity compared to other threatened species

The nucleotide diversity of *C. forrestii* ( $0.17e^{-3}$  at the species level) is very low. Previous genetic studies of other orchid species data are based on chloroplast DNA fragments, and yielded much higher nucleotide diversity. For example, the nucleotide diversity of *C. japonicum* is  $0.88e^{-3}$  (Han et al., 2022); *C. tibeticum* has a much higher nucleotide diversity of  $1.52e^{-3}$  (Guo et al., 2019). However, nucleotide diversity estimated using transcriptome or genome data should be more accurate, as it was computed based on a much higher number of loci. In comparison to other threatened or endangered species studied with transcriptome data, the nucleotide diversity of *C. forrestii* is still very low. For example, *Pseudotaxus chienii* has a nucleotide diversity of  $0.7e^{-3}$  (Liu et al., 2021). *Cupressus gigantea* and *Cupressus duclouxiana* exhibit nucleotide diversities of  $2.9e^{-3}$  and  $3.1e^{-3}$ , respectively (Ma et al., 2019b). Additionally, the protein coding region of *Acer yangbiense* shows a nucleotide diversity of  $1.88e^{-3}$  [Supplementary Data of (Ma et al., 2022b)]. Moreover, the whole species level nucleotide diversity of *Cypripedium forrestii* was still lower than the very few samples of *Cypripedium lichiangense* ( $0.4e^{-3}$ ).

Nucleotide diversity values can be greatly altered by the type of genetic marker used in the analysis, the sequencing technique and the sampling size, and therefore, any direct comparison between species may be arbitrary. Furthermore, nucleotide diversity values and conservation status are not necessarily correlated. However, it is generally believed that a higher genetic diversity should enhance the potential of species to adapt to the ever-changing environment (Zhao et al., 2019; Ma et al., 2021). Therefore, a low nucleotide diversity could be a critical indicator of species that require the most urgent protection. The low nucleotide diversity may reflect highly conserved protein-coding genes and functional elements. However,



a better explanation may be that *C. forrestii* accumulates fewer mutations over time. This hypothesis is supported by our field observations and horticultural experience, i.e., the plant mainly reproduces asexually though tillering, and has low rates of flowering and fruiting. Another possible explanation for the low nucleotide diversity could be the loss of polymorphism sites, especially low frequency alleles, due to genetic drift. This will be discussed in the following section.

### 4.3 The bottleneck and declining population of *C. forrestii*

Negative  $F_{IS}$  values are observed at all sampled locations and in all genetic clusters. This phenomenon has also been reported in many other orchid species at the population and/or at the species level, e.g. *Pelatantheria scolopendrifolia* (Yun et al., 2020); *Serapias lingua* (Pellegrino et al., 2015); *Spiranthes spiralis* (Machon et al., 2003); *Cypripedium calceolus* (Minasiewicz et al., 2018; Gargiulo et al., 2021) and *C. reginae* (Kennedy and Walker, 2007); as well as plants from other families (Stoeckel et al., 2006; Cabrera-Toledo et al., 2008; Yang et al., 2022). The negative inbreeding values indicate the heterozygote excess within the population, which may result from certain kinds of heterozygote advantage during natural selection (Alvarez-Buylla et al., 1996). However, a small sample size may also cause negative inbreeding values (Balloux, 2004). Notably, the sampling site with the lowest  $F_{IS}$  values (HBA-03) has the highest Tajima's D value and the lowest  $N_E$ , and the site with the highest  $F_{IS}$  values (HBB-05) has the lowest Tajima's D and highest  $N_E$  values. Though originally designed to test the neutrality of DNA polymorphism data (Tajima, 1989), Tajima's D value can be an effective tool in the detection of selective forces and bottleneck events (Simonsen et al., 1995; Nielsen, 2001). The overall Tajima's D value of 1.05 may suggest the loss of low frequency alleles across the transcriptome, which is a typical consequence of a bottleneck event.

Stairway Plot2 suggests a continuous decline of the population size. *C. forrestii* is a typical alpine plant that favors a cold environment, and may be susceptible to global warming and contemporary climate fluctuations. During our field surveys, we found that most of its growing area was affected by livestock grazing. Increasing human activities in its suitable habitat may also contribute to its recent population decline.

The severe population decline brought by the bottle neck events is one of the initial causes of heterozygote excess, which can be enhanced by plant's pollination characteristics like dioecism, self-incompatibility and unbalanced sex ratios (Balloux, 2004; Stoeckel et al., 2006; Cabrera-Toledo et al., 2008). Self-pollination can result in the clearance of heterozygote excess, whereas several studies have shown that deceptive pollination promotes cross-pollination in many orchid species (Jersáková et al., 2006). As mentioned, *C. forrestii* produces deceptive flowers, therefore its heterozygote excess should be reinforced by the cross-pollination between different genetic individuals.

### 4.4 The limitation of gene flow

AMOVA analysis reveals that the vast majority of genetic variation is found within populations.  $F_{IS}$  values between some sites are very small, which suggest that individuals from these sites are very similar genetically. It is possible that all sites and genetic clusters used to be a whole entity with extensive genetic connection. The Treemix result suggests that most sites used to be connected, and that there may have been previous long-distance gene flow from the Haba mountains and the northern sites to other sites. This connection, if it ever existed, has been lost, possibly during the most recent population decline and habitat fragmentation. BayesAss analysis suggests that there has been barely any gene flow between most sites in recent generations. We do not find any populations between the Haba mountains and the two northern sites. Also, the hot and dry valley of the Jinsha river separates YLXS from the other populations, and villages, farmland and ranches have been replacing forest and shrubland in the lower altitude areas.

A certain degree of connectivity still remains between HBB-04 and HBB-05, probably due to the short geographical distance between the two sites. However, interestingly, the short geographical distance between HBA-03 and its neighbor does not seem to increase connectivity between these sites. Noticeably, as demonstrated in Figures 2A and C, two individuals from HBA-03 show a certain degree of genetic mixture with other genetic clusters, while other individuals have a "unmixed" genetic background. This is also observed in some individuals from the northern group and HBB-04 (Figures 2C, E). One possible explanation is that, although rather scarce, there is a certain amount of seed flow between the sites. However, this seed flow does not necessarily bring in actual gene flow. As most individuals do not flower annually and the fruiting rate is low, the chance of producing offspring with mixed genetic background is likely to be slim, even given occasional seed flow between populations. In contrast, genetically mixed offsprings seem to be more common in the other central sites. This specie may merely exist and reproduces asexually in the "second-rate" habitats (the northern sites and HBB-03) and is only inclined to flower in particular environments. It is also possible that presence or abundance of pollinators vary greatly in different sites. Further field observations and pollination ecology studies should help answer this question.

### 4.5 Conservative suggestion of *C. forrestii* based on genetic study

According to our field investigation and genetic study, *C. forrestii* is facing four major threats: (1) a very narrow distribution; (2) low genetic diversity; (3) a small population size and (4) the loss of gene flow between most sites. In view of these factors, the upcoming conservation effort should cover these aspects: (a) *in situ* protection; (b) *in vitro* conservation; (c) cultivation and reintroduction; (d) further exploration of other biological and ecological characteristics.

First, the *in situ* protection should be enhanced. All known sites of *C. forrestii* except ZZ are located within protected areas such as national parks or nature reserves. Cooperation between local government and related authorities can play a vital role in the protection of biodiversity in such areas. Practicable measures include setting up infrared cameras, warning signs and forest ranger patrol. One population is located in the Yulong Snow Mountains, which is one of the most famous scenic spots in Yunnan. Public education and tourism regulation should be enhanced in this particular site. On the other hand, the nearby mountains in Sichuan, for example, may comprise habitat suitable for *Cypripedium*, but to date, these mountains have not been thoroughly investigated. Therefore, field investigations in close-lying areas should be conducted to search for new populations.

The site ZZ is not located in any protected area and is facing the strongest disturbance. HBA-01 and HBA-03 comprise healthy numbers of individuals with a unique genetic background. Collection of ramet, pollen and fruit can be considered for these sites. Artificial pollination could be an effective way to enhance the fruit production and genetic diversity of *C. forrestii*. Since the genetic distances between most sites are relatively small, cross-pollination between sites, at least within genetic clusters, should be appropriate. Our study shows that gene flow between the Yulong Snow Mountains and other sites has been cut-off for a long time. Therefore, cross pollination between individuals in the Yulong Snow Mountains and other central genetic clusters (e.g., HBB-04 and 05) should be a priority to restore gene flow and promote genetic diversity.

The Kunming Botanical Garden is currently engaged in the *in vitro* conservation of *C. forrestii*, which includes aseptic cultivation of seedlings. Additionally, *in situ* seed baiting experiments can be conducted to detect mycorrhizal fungi that enhance the germination. These efforts aim to provide a large number of seedlings and establish a fundamental understanding of the species, which will serve as a foundation for future reintroduction work. Site monitoring is also crucial in understanding the relationship between the environment and reproductive fitness. It is important to identify suitable reintroduction sites where the plant can regenerate and thrive successfully.

The pollination ecology and determinants of leaf vaiegation of *C. forrestii* are still not fully understood. Further observation is required to identify the pollinators of this species and determine if there are any reproductive difficulties. Furthermore, its leaf patterns may have implications for photosynthesis and interactions with herbivores. Investigating leaf polymorphism and its impact on the growth and fitness of the species would provide valuable insights for population management strategies.

*C. forrestii* is now included in the PSESP protection project. Preliminary conservation actions, including artificial pollination and the aseptic cultivation of seedlings, are already underway. These initial conservation efforts are crucial for the long-term survival and recovery of this specie. By actively cooperating with various stakeholders and implementing conservation measures, there is hope for the population to stabilize in the future.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: BioProject, PRJNA1029356.

## Author contributions

LL: Writing – original draft, Writing – review & editing. LC: Writing – review & editing. HH: Writing – review & editing. SM: Writing – review & editing. WS: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by: the Second Tibetan Plateau Scientific Expedition and Research Program (2019QZKK0502); Science & Technology Basic Resources Investigation Program of China for Survey and Germplasm Conservation of PSESP in Southwest China (2017FY100100); the PSESP project of Yunnan Forestry and Grassland Bureau (2021SJ14X-09); and the project “Collection and Conservation of Plant Species with Extremely Small Populations of *Polystichum glaciale* and *Cypripedium forrestii* in Lijiang” (2021SJ14X-11).

## Acknowledgments

We would like to express our gratitude to Liu Yang from the Kunming Institute of Botany, He Zhixun from the Lijiang Alpine Botanic Garden, and Fang Ye from the Shangri-la Alpine Botanical Garden for their valuable assistance during the field survey.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1303625/full#supplementary-material>

## References

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Alvarez, M., Schrey, A. W., and Richards, C. L. (2015). Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? *Mol. Ecol.* 24, 710–725. doi: 10.1111/mec.13055
- Alvarez-Buylla, E., Garcia-Barrios, R., and Lara-Moreno and M. Martínez-Ramos, C. (1996). Demographic and genetic models in conservation biology: applications and perspectives for tropical rain forest tree species. *Annu. Rev. Ecol. Sys.* 27, 387–421. doi: 10.1146/annurev.ecolsys.27.1.387
- Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*, pp (Cambridge, United Kingdom: Babraham Bioinformatics, Babraham Institute).
- Balloux, F. (2004). Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution* 58, 1891–1900. doi: 10.1111/j.0014-3820.2004.tb00477.x
- Bijlsma, R., and Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Appl.* 5, 117–129. doi: 10.1111/j.1752-4571.2011.00214.x
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brundrett, M. C. (2007). Scientific approaches to Australian temperate terrestrial orchid conservation. *Aust. J. Bot.* 55, 293–307. doi: 10.1071/BT06131
- Brundrett, M. C. (2019). A comprehensive study of orchid seed production relative to pollination traits, plant density and climate in an urban reserve in Western Australia. *Diversity* 11, 123. doi: 10.3390/d11080123
- Brzosko, E., Ostrowiecka, B., Kotowicz, J., Bolesta, M., Gromotowicz, A., Gromotowicz, M., et al. (2017). Seed dispersal in six species of terrestrial orchids in Biebrza National Park (NE Poland). *Acta Societatis Botanicorum Poloniae* 86, 3557. doi: 10.5586/asbp.3557
- Brzosko, E., Wróblewska, A., and Ratkiewicz, M. (2002). Spatial genetic structure and clonal diversity of island populations of lady's slipper (*Cypripedium calceolus*) from the Biebrza National Park (northeast Poland). *Mol. Ecol.* 11, 2499–2509. doi: 10.1046/j.1365-294X.2002.01630.x
- Cabrera-Toledo, D., Gonzalez-Astorga, J., and Vovides, A. P. (2008). Heterozygote excess in ancient populations of the critically endangered *Dioon caputoi* (Zamiaceae, Cycadales) from central Mexico. *Botanical J. Linn. Soc.* 158, 436–447. doi: 10.1111/j.1095-8339.2008.00868.x
- Chen, S. (2013). *The genus Cypripedium in China* (Beijing, China: Science Press).
- Cogoni, D., Fenu, G., Dessì, C., Deidda, A., Giotta, C., Piccitto, M., et al. (2021). Importance of plants with extremely small populations (PSESPs) in endemic-rich areas, elements often forgotten in conservation strategies. *Plants* 10, 1504. doi: 10.3390/plants10081504
- Crane, P. (2020). Conserving our global botanical heritage: The PSESP plant conservation program. *Plant Diversity* 42, 319. doi: 10.1016/j.pld.2020.06.007
- Cribb, P. (1997). *The genus Cypripedium* (Portland, Oregon: Timber Press).
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Diez, J. M. (2007). Hierarchical patterns of symbiotic orchid germination linked to adult proximity and environmental gradients. *J. Ecol.* 95, 159–170. doi: 10.1111/j.1365-2745.2006.01194.x
- Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Mol. Ecol. Resour.* 14, 209–214. doi: 10.1111/1755-0998.12157
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinf.* 1, 117693430500100003. doi: 10.1177/117693430500100003
- Fitak, R. R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* 6, 1–6. doi: 10.1093/biomethods/bpab017
- Gargiulo, R., Adamo, M., Cribb, P. J., Bartolucci, F., Sarasan, V., Alessandrelli, C., et al. (2021). Combining current knowledge of *Cypripedium calceolus* with a new analysis of genetic variation in Italian populations to provide guidelines for conservation actions. *Conserv. Sci. Pract.* 3, e513. doi: 10.1111/csp2.513
- Guo, J.-L., Cao, W.-J., Li, Z.-M., Zhang, S., and Volis, Y. H. (2019). Conservation implications of population genetic structure in a threatened orchid *Cypripedium tibeticum*. *Plant Diversity* 41, 13–18. doi: 10.1016/j.pld.2018.12.002
- Han, L. X., Jin, Y., Zhang, J. L., Li, X. L., Chung, M. Y., Herrando-Morairae, S., et al. (2022). Phylogeography of the endangered orchids *Cypripedium japonicum* and *Cypripedium formosanum* in East Asia: Deep divergence at infra- and interspecific levels. *Taxon* 71, 733–757. doi: 10.1002/tax.12710
- He, W., Zhao, S., Liu, X., Dong, S., Lv, J., Liu, D., et al. (2013). ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. *Genet. Mol. Res.* 12, 6275–6283. doi: 10.4238/2013.December.4.15
- Hollick, P. S., McComb, J. A., and Dixon, K. W. (2007). Introduction, growth and persistence *in situ* of orchid mycorrhizal fungi. *Aust. J. Bot.* 55. doi: 10.1071/BT06073
- Hu, S. J., Hu, H., Yan, N., Huang, J. L., and Li, S. Y. (2011). Hybridization and asymmetric introgression between *Cypripedium tibeticum* and *C. yunnanense* in Shangrila County, Yunnan Province, China. *Nordic J. Bot.* 29, 625–631. doi: 10.1111/j.1756-1051.2010.00918.x
- Izawa, T., Kawahara, T., and Takahashi, H. (2007). Genetic diversity of an endangered plant, *Cypripedium macranthos* var. *rebunense* (Orchidaceae): background genetic research for future conservation. *Conserv. Genet.* 8, 1369–1376. doi: 10.1007/s10592-007-9287-1
- Jacquemyn, H., Brys, R., Honnay, O., Roldán-Ruiz, I., Lievens, B., Wiegand, T., et al. (2012). Nonrandom spatial structuring of orchids in a hybrid zone of three Orchis species. *New Phytol.* 193, 454–464. doi: 10.1111/j.1469-8137.2011.03913.x
- Jersáková, J., Johnson, S. D., and Kindlmann, P. (2006). Mechanisms and evolution of deceptive pollination in orchids. *Biol. Rev. Cambridge Philos. Soc.* 81 (2), 219–235. doi: 10.1017/S1464793105006986
- Kennedy, A. H., and Walker, G. L. (2007). The population genetic structure of the showy lady's-slipper orchid (*Cypripedium reginae* Walter) in its glaciated and unglaciated ranges. *Castanea* 72, 248–261. doi: 10.2179/06-30.1
- Kéry, M., Matthies, D., and Spillmann, H. H. (2000). Reduced fecundity and offspring performance in small populations of the declining grassland plants *Primula veris* and *Gentiana lutea*. *J. Ecol.* 88, 17–30. doi: 10.1046/j.1365-2745.2000.00422.x
- Korneliusson, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* 15, 1–13. doi: 10.1186/s12859-014-0356-4
- Kotilinek, M., Těšitelová, T., Košnar, J., Fibich, P., Hemrová, L., Koutecký, P., et al. (2020). Seed dispersal and realized gene flow of two forest orchids in a fragmented landscape. *Plant Biol.* 22, 522–532. doi: 10.1111/plb.13099
- Li, J. H., Liu, Z. J., Salazar, G. A., Bernhardt, P., Perner, H., Tomohisa, Y., et al. (2011). Molecular phylogeny of *Cypripedium* (Orchidaceae: Cypripedioideae) inferred from multiple nuclear and chloroplast regions. *Mol. Phylogenet. Evol.* 61 (2), 308–320. doi: 10.1016/j.ympev.2011.06.006
- Liu, X., and Fu, Y.-X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* 21, 1–9. doi: 10.1186/s13059-020-02196-9
- Liu, D., Sun, W., and Ma and Z. Fang, Y. (2019). Rediscovery and conservation of the critically endangered *Rhododendron griersonianum* in Yunnan, China. *Oryx* 53, 14–14. doi: 10.1017/S0030605318001278
- Liu, L., Wang, Z., Su, Y., and Wang, T. (2021). Population transcriptomic sequencing reveals allopatric divergence and local adaptation in *Pseudotsaxus chienii* (Taxaceae). *BMC Genomics* 22 (1), 388. doi: 10.1186/s12864-021-07682-3
- Ma, Y., Chen, G., Edward Grumbine, R., Dao, Z. L., Sun, W. B., Guo, H. J., et al. (2013). Conserving plant species with extremely small populations (PSESP) in China. *Biodiversity Conserv.* 22, 803–809. doi: 10.1007/s10531-013-0434-3
- Ma, Y., Li, C., Jin, J., Liao, C., Yang, J., and Sun, W. B. (2022a). Conservation genetics of *Firmiana major*, a threatened tree species with potential for afforestation of hot, arid climates. *Global Ecol. Conserv.* 36, e02136. doi: 10.1016/j.gecco.2022.e02136
- Ma, H., Liu, Y. B., Liu, D. T., Sun, W. B., Liu, X. F., Wan, Y. M., et al. (2021). Chromosome-level genome assembly and population genetic analysis of a critically endangered rhododendron provide insights into its conservation. *Plant J.* 107, 1533–1545. doi: 10.1111/tpj.15399
- Ma, Y., Liu, D., Wariss, H. M., Zhang, R. G., Tao, L. D., Milne, R. I., et al. (2022b). Demographic history and identification of threats revealed by population genomic analysis provide insights into conservation for an endangered maple. *Mol. Ecol.* 31, 767–779. doi: 10.1111/mec.16289
- Ma, Y. P., Marczewski, T., Xue, D., Wu, Z. K., Liao, R. L., Sun, W. B., et al. (2019a). Conservation implications of asymmetric introgression and reproductive barriers in a rare primrose species. *BMC Plant Biol.* 19, 1–11. doi: 10.1186/s12870-019-1881-0
- Ma, Y. Z., Wang, J., Hu, Q. J., Li, J. L., Sun, Y. S., Zhang, L., et al. (2019b). Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Commun. Biol.* 2, 213. doi: 10.1038/s42003-019-0445-z
- Machon, N., Bardin, P., Mazer, S. J., Moret, J., Godelle, B., Austerlitz, S., et al. (2003). Relationship between genetic structure and seed and pollen dispersal in the endangered orchid *Spiranthes spiralis*. *New Phytol.* 157, 677–687. doi: 10.1046/j.1469-8137.2003.00694.x
- McCormick, M. K., Lee Taylor, D., Juhaszova, K., Burnett, R. K. Jr., Whigham, D. F., and Neill O', J. P. (2012). Limitations on orchid recruitment: not a simple picture. *Mol. Ecol.* 21, 1511–1523. doi: 10.1111/j.1365-294X.2012.05468.x

- Minasiewicz, J., Znaniecka, J. M., Górniak, M., and Kawiński, A. (2018). Spatial genetic structure of an endangered orchid *Cypripedium calceolus* (Orchidaceae) at a regional scale: limited gene flow in a fragmented landscape. *Conserv. Genet.* 19, 1449–1460. doi: 10.1007/s10592-018-1113-4
- Musmann, S. M., Douglas, M. R., Chafin, T. K., and Douglas, M. E. (2019). BA3-SNPs: Contemporary migration reconfigured in BayesAss for next-generation sequence data. *Methods Ecol. Evol.* 10, 1808–1813. doi: 10.1111/2041-210X.13252
- Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity* 86, 641–647. doi: 10.1046/j.1365-2540.2001.00895.x
- Pellegrino, G., Bellusci, F., and Palermo, A. M. (2015). Effects of population structure on pollen flow, clonality rates and reproductive success in fragmented *Serapias lingua* populations. *BMC Plant Biol.* 15, 1–10. doi: 10.1186/s12870-015-0600-8
- Pickrell, J., and Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nat. Precedings*, 1–1. doi: 10.1038/npre.2012.6956.1
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qian, X., Li, Q.-J., Liu, F., Gong, M.-J., Wang, C.-X., Tian, M., et al. (2014). Conservation genetics of an endangered Lady's slipper orchid: *Cypripedium japonicum* in China. *Int. J. Mol. Sci.* 15, 11578–11596. doi: 10.3390/ijms150711578
- Ralls, K., Ballou, J. D., Dudash, M. R., Eldridge, M. D., Fenster, C. B., Lacy, R. C., et al. (2018). Call for a paradigm shift in the genetic management of fragmented populations. *Conserv. Lett.* 11, e12412. doi: 10.1111/conl.12412
- Sangiovanni, M., Granata, I., Thind, A. S., and Guarracino, M. R. (2019). From trash to treasure: detecting unexpected contamination in unmapped NGS data. *BMC Bioinf.* 20, 1–12. doi: 10.1186/s12859-019-2684-x
- Simonsen, K. L., Churchill, G. A., and Aquadro, C. F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* 141, 413–429. doi: 10.1093/genetics/141.1.413
- Spielman, D., Brook, B. W., and Frankham, R. (2004). Most species are not driven to extinction before genetic factors impact them. *Proc. Natl. Acad. Sci.* 101, 15261–15264. doi: 10.1073/pnas.0403809101
- Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., Mariette, S., et al. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. *Mol. Ecol.* 15, 2109–2118. doi: 10.1111/j.1365-294X.2006.02926.x
- Suetsugu, K., and Fukushima, S. (2014). Pollination biology of the endangered orchid *Cypripedium japonicum* in a fragmented forest of Japan. *Plant Species Biol.* 29, 294–299. doi: 10.1111/1442-1984.12016
- Sun, W. B. (2021). *List of Yunnan Protected Plant Species with Extremely Small Population, (2021)* (Yunnan, China: Yunnan Science and Technology Press CO., LTD.).
- Sun, W. B., Ma, Y. P., and Blackmore, S. (2019a). How a new conservation action concept has accelerated plant conservation in China. *Trends Plant Sci.* 24, 4–6. doi: 10.1016/j.tplants.2018.10.009
- Sun, W. B., Yang, J., and Dao, Z. (2019b). *Study and Conservation of Plant Species with Extremely Small Populations (PSESP) in Yunnan Province, China* (Beijing: Science Press).
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595. doi: 10.1093/genetics/123.3.585
- Tarasov, A., Vilella, A. J., Cuppen, E., and Nijman and P. Prins, I. J. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. doi: 10.1093/bioinformatics/btv098
- Tyagi, P., Singh, D., Mathur, S., and Singh and R. Ranjan, A. (2022). Upcoming progress of transcriptomics studies on plants: An overview. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1030890
- Volis, S. (2016). How to conserve threatened Chinese plant species with extremely small populations? *Plant Diversity* 38, 45–52. doi: 10.1016/j.pld.2016.05.003
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. doi: 10.1371/journal.pone.0112963
- Waples, R. S., and Do, C. (2010). Linkage disequilibrium estimates of contemporary *N<sub>e</sub>* using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Appl.* 3, 244–262. doi: 10.1111/j.1752-4571.2009.00104.x
- Wilson, G. A., and Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163, 1177–1191. doi: 10.1093/genetics/163.3.1177
- Yang, F. M., Cai, L., Dao, Z. L., and Sun, W. B. (2022). Genomic data reveals population genetic and demographic history of *Magnolia fistulosa* (Magnoliaceae), a plant species with extremely small populations in Yunnan province, China. *Front. Plant Sci.* 13, 811312. doi: 10.3389/fpls.2022.811312
- Yang, J., Cai, L., Liu, D. T., Chen, G., Gratzfeld, J., and Sun, W. B. (2020). China's conservation program on plant species with extremely small populations (PSESP): progress and perspectives. *Biol. Conserv.* 244, 108535. doi: 10.1016/j.biocon.2020.108535
- Yang, F. M., Ge, J., Guo, Y., Olmstead, R., and Sun, W. B. (2023). Deciphering complex reticulate evolution of *Asian Buddleja* (Scrophulariaceae): insights into the taxonomy and speciation of polyploid taxa in the Sino-Himalayan region. *Ann. Bot.* 132, 15–28. doi: 10.1093/aob/mcad022
- Yu, Y.-L., Wang, H.-C., Yu, Z.-X., Schinnerl, J., Tang, R., Geng, Y. P., et al. (2021). Genetic diversity and structure of the endemic and endangered species *Aristolochia delavayi* growing along the Jinsha River. *Plant Diversity* 43, 225–233. doi: 10.1016/j.pld.2020.12.007
- Yun, S. A., Son, H.-D., Im, H. T., and Kim, S. C. (2020). Genetic diversity and population structure of the endangered orchid *Pelatantheria scolopendrifolia* (Orchidaceae) in Korea. *PLoS One* 15, e0237546. doi: 10.1371/journal.pone.0237546
- Zhao, Y. P., Fan, G., Yin, P.-P., Sun, S., Li, N., Hong, X. N., et al. (2019). Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. *Nat. Commun.* 10, 4201. doi: 10.1038/s41467-019-12133-5