Postprint: Revealing the Endangerment Mechanism of Hydrocera Through Reduced-Representation Genome Sequencing

Authors: Wu Xinyi, Wang Meng, Zheng Xilong, Zhang Rui, He Song, YAN Yuehong

Date: 2023-07-26T00:00:00+00:00

Abstract

Genetic diversity constitutes a critical factor for species adaptability and viability. Habitat fragmentation represents a major driver of biodiversity loss and exerts significant impacts on the genetic diversity of plant populations. Hydrocera triflora is currently regionally extinct, with its genetic diversity status and endangerment mechanisms remaining unreported. This study collected 34 samples from seven populations of H. triflora and obtained single nucleotide polymorphism (SNP) markers via restriction-site associated DNA sequencing (RAD-seq). Through analyses of population genetic diversity and structure, combined with population historical dynamics and predictions of species potential distribution under various climate scenarios, we investigated the endangerment mechanisms of H. triflora. The results revealed: (1) H. triflora exhibited low genetic diversity (Ho=0.1569, He=0.1654, Pi=0.186 5) and high genetic differentiation coefficients; AMOVA analysis demonstrated that genetic variation occurred primarily within populations; (2) Mantel tests indicated significant positive correlations between environmental distance and both genetic distance (P=0.044 7) and geographic distance (P=0.008 2); (3) The effective population size of H. triflora has declined continuously since the mid-Holocene, with eruptions of the Qiongbei volcanic group and anthropogenic activities exerting non-negligible influences; (4) Ecological niche modeling indicated that compared with contemporary climate, the total area of potential distribution under future climate scenarios would remain relatively stable; however, under high CO2 emission scenarios, extensive highly suitable habitats would be lost and converted to low-suitability areas, particularly in the Malay Archipelago where suitable habitats would nearly vanish completely. These results demonstrate that habitat fragmentation has resulted in low genetic diversity and continuous decline in effective population size in H. triflora. Consequently, low natural regeneration capacity coupled with adverse environmental conditions such as anthropogenic disturbance and urbanization constitute the primary drivers of its endangerment. We recommend strengthening in-situ conservation efforts for H. triflora and implementing measures such as artificial pollination to enhance gene flow and increase population genetic diversity. Additionally, priority should be given to protecting wetlands from degradation.

Full Text

Preamble

The RAD-seq Revealed the Endangered Mechanism of *Hydrocera triflora*

WU Xinyi[1], WANG Meng[1], ZHENG Xilong[2], ZHANG Rui[3], HE Song[4], YAN Yuehong[1]*

[1] Orchid Conservation & Research Center of Shenzhen and the National Orchid Conservation Center of China, Shenzhen Key Laboratory for Orchid Conservation and Utilization, Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization, Shenzhen, 518114, Guangdong, China
[2] School of Traditional Chinese Medicine Resources, Guangdong Pharmaceutical University, Guangzhou 510006, China
[3] Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai 201602, China
[4] Haikou Wetland Protection and Management Center, Haikou 570135, China

Abstract

Genetic diversity is a critical factor determining species' adaptability and survival. Habitat fragmentation represents a major driver of biodiversity loss and significantly impacts plant population genetic diversity. *Hydrocera triflora* currently faces regional extinction, yet its genetic diversity status and endangerment mechanisms remain unreported. This study collected 34 samples from seven populations and obtained single nucleotide polymorphism (SNP) markers via restriction site-associated DNA sequencing (RAD-seq). We investigated the endangered mechanisms of *H. triflora* through population genetic diversity and structure analysis, combined with population history dynamics and species distribution modeling under different climate scenarios. The results revealed: (1) *H. triflora* exhibited low genetic diversity ($Ho = 0.1569$, $He = 0.1654$, $Pi = 0.1865$) with high genetic differentiation. AMOVA analysis indicated that genetic variation primarily occurred within populations. (2) Mantel tests showed significant positive correlations between environmental distance and genetic distance ($P = 0.0447$), as well as between environmental distance and geographic distance ($P = 0.0082$). (3) Effective population size has continuously declined since the mid-Holocene, with eruptions of the Qiongbei volcanic group and human activities exerting non-negligible impacts. (4) Ecological niche modeling revealed that while total potential distribution area changed little under future climate change, high CO_2 emission scenarios would cause substantial loss of highly suitable habitats, converting them to low-suitability areas. Particularly, suitable habitats in the Malay Archipelago would nearly disappear entirely. These results demonstrate that habitat fragmentation has led to low genetic diversity and continuous decline in effective population size in *H. triflora*. Consequently, low self-renewal capacity combined with adverse environmental conditions such as human disturbance and urbanization constitute the primary causes of its endangered status. We recommend strengthening in-situ conservation of *H. triflora* and implementing artificial pollination to enhance gene flow and increase population genetic diversity, while prioritizing wetland protection from destruction.

Keywords: habitat fragmentation, RAD-seq, genetic diversity, genetic structure, species distribution models, population history

Introduction

Conservation genomics broadly refers to the application of genomic technologies and methods to biodiversity conservation and species protection [**?**], focusing primarily on endangered mechanisms and conservation strategies. Assessing genetic diversity levels and their distribution patterns within individuals and populations constitutes essential content for conservation genetic research [**?**]. Molecular genetic markers serve as important tools for evaluating germplasm resource genetic diversity and have driven the development of conservation genetics. Although genetic information has long been applied in endangered species research, genome-wide information markers have only been utilized in the past two decades [**?**], initially restricted to a few model plants.

The rapid development of high-throughput sequencing technology has given rise to reduced-representation genome sequencing. RAD-seq (restriction site-associated DNA sequencing) is a common reduced-representation genome technology that sequences short fragments adjacent to restriction enzyme cleavage sites, offering advantages of high coverage and low cost [**?**, **?**]. It operates without reference genome constraints and can obtain large numbers of high-quality single-nucleotide polymorphism (SNP) markers. Currently, RAD-seq has been widely applied in population research and molecular breeding, including molecular marker development, population genetic analysis, genetic map construction, and genome-wide association studies. For instance, Sun et al. [**?**] employed RAD-seq to investigate the conservation genomics of the endangered fern *Adiantum nelumboides*, revealing its endangered mechanisms. Cai et al. [**?**] performed RAD-seq on the economically valuable tropical rainforest tree *Horsfieldia tetratepala* to develop more effective conservation strategies. Cao et al. [**?**] utilized RAD-seq for reference-free analysis of genetic characteristics in the critically endangered species *Rhododendron hemsleyanum*, demonstrating that reduced-representation genome sequencing can obtain sufficient SNP molecular markers for plant species with extremely small populations (PSESP) lacking reference genomes. Therefore, we believe that for endangered species without reference genomes, reduced-representation genome sequencing can acquire adequate genetic information, providing effective tools for conservation genomics research.

Estimation of effective population size is crucial for conserving and managing endangered species, understanding population dynamics, and predicting the direction and speed of genetic evolution. Population history dynamics represent important research content in conservation genomics [?]. During population evolution, climate change and geological events serve as critical external forces driving population dynamic changes.

Beyond species-specific characteristics and climate history, habitat fragmentation constitutes another major factor causing biodiversity loss [?], leading to reduced genetic diversity, inbreeding depression, decreased gene flow, and population size reduction [?, ?, ?]. Habitat fragmentation also reduces species habitats, affects pollinator abundance, decreases pollination rates and reproductive success, thereby influencing species diversity [?, ?]. Habitat fragmentation refers to the formation of new habitat patterns through the division of previously continuous habitats into fragments of varying sizes and shapes [?]. Currently, most studies indicate that habitat fragmentation exerts more negative than positive effects on genetic diversity [?]. Research on the obligate outcrossing species *Dianthus deltoides* in fragmented habitats revealed low seed set and potential self-fertilization, further reducing population genetic diversity [?]. The endangered plant *Banksia goodii* experienced reduced reproductive fitness in roadside populations due to habitat fragmentation [?]. Therefore, when endangered species are affected by habitat fragmentation, enhanced research is needed to better understand how habitat fragmentation influences plant genetic diversity, reveal endangered mechanisms, and propose effective conservation strategies.

Hydrocera triflora is a perennial aquatic herb in the family Balsaminaceae, distributed in Hainan (China), India, Sri Lanka, Thailand, Vietnam, Laos, Cambodia, Malaysia, and Indonesia. It prefers lakesides, swamp wetlands, or rice paddies. Investigations of *H. triflora* resources in Hainan Province revealed that with rural tourism development, water body modification, and extensive wetland exploitation, *H. triflora* habitats have suffered varying degrees of destruction, resulting in increasingly narrow distribution ranges and extreme rarity [?]. The IUCN assessment report classifies *H. triflora* as regionally extinct (RE). Related research remains scarce; the species was rediscovered in the Haikou area by the Kunming Institute of Botany, Chinese Academy of Sciences in 2014, marking its first domestic sighting in over 30 years. Floral development studies of *Hydrocera* and *Impatiens* have revealed evolutionary trends in the earliest diverging lineages of Balsaminaceae, showing that inflorescences gradually transformed from simple axial inflorescences to more complex axillary inflorescences, with perianth differentiation representing a gradual process [?]. As a monotypic genus in Balsaminaceae, studying the genetic diversity of *H. triflora* can not only assess the impact of habitat fragmentation on genetic diversity but also provide fundamental data for the evolution and systematic classification of Balsaminaceae.

This study utilized 34 *H. triflora* individuals from seven populations in the Haikou region as experimental materials, performing population genetic analysis through RAD-seq reduced-representation genome sequencing to investigate: (1) genome-level genetic diversity and structure of *H. triflora*; (2) historical changes in effective population size; and (3) potential distribution area changes under different climate scenarios. The findings will provide theoretical foundations for germplasm resource conservation and breeding strategy implementation.

Materials and Methods

1.1 Experimental Materials

Field collection of *H. triflora* distributed in Haikou City was conducted [Figure 1: see original paper], yielding 34 samples from seven populations. Each population had one voucher specimen deposited in the Herbarium of the Orchid Conservation & Research Center of Shenzhen (CNOCC). Detailed population information is provided in Table 1 .

1.2 DNA Extraction and RAD Library Construction

Genomic DNA from all samples was extracted using the Tiangen Plant Genomic DNA Extraction Kit (DP305) with appropriate modifications. Following agarose gel electrophoresis and Qubit concentration detection, samples were sent to Shanghai Yuanxin Biotechnology Co., Ltd. for library construction. The RAD library was constructed using EcoR I restriction enzyme digestion, with paired-end sequencing libraries ranging from 300–500 bp. Purified libraries were sequenced on the Illumina NovaSeq 6000 platform for reduced-representation genome sequencing.

1.3 SNP Discovery

After sequencing, FastQC software was used for quality control, including adapter removal, poly-A trimming, and low-quality data filtering to obtain high-quality clean data for subsequent analysis. We employed Stacks v2.54 [**?**] for de novo assembly and variant calling of *H. triflora* RAD-seq data. The process_{radtags} program in Stacks was run to filter samples by removing sequences lacking RAD tags. The ustacks program built loci for each sample with parameters M = 3 (controlling mismatches) and m = 3 (minimum fragments per allele). The cstacks program constructed a catalog containing all loci information with n = 3 (allowed mismatches between loci). The sstacks program aligned each sample to the created catalog, followed by tsv2bam, gstacks, and population programs to generate a VCF file containing 75,209 SNP loci. For reliable downstream analysis, VCFtools v0.1.13 [**?**] was used for further SNP filtering with parameters: –max-missing 0.9 (maximum missing rate 90%), –maf 0.05 (minimum allele frequency 0.05), and –minDP 5 (minimum genotype depth 5). After filtering, 38,667 high-quality SNPs remained for population genetic structure analysis.

1.4 Population Structure Analysis

Based on SNP loci from 34 *H. triflora* samples, the vcf2phylip script was used to construct an individual genetic distance matrix. IQ-TREE v1.6.9 [?] was then employed to build phylogenetic trees using maximum likelihood (ML) with 1,000 ultrafast bootstrap replicates [?] and SH-aLRT tests [?] to evaluate branch support. The resulting trees were visualized and edited using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

The high-quality SNP dataset was converted to ped and map files using VCFtools, then filtered to generate bed files. ADMIXTURE [?] was used to infer individual ancestry components based on genotypes, simulating clustering scenarios with K = 2 to 7 subpopulations. The optimal number of clusters was determined by the minimum cross-validation (CV) error across different K values.

GCTA v1.26.0 (https://yanglab.westlake.edu.cn/software/gcta/) was used to generate matrices for principal component analysis (PCA), with results plotted using the ggplot2 package in R.

1.5 Population Genetic Diversity Analysis

The Populations subprogram in Stacks calculated species-level inbreeding coefficient (FIS), observed heterozygosity (Ho), expected heterozygosity (He), and nucleotide diversity (Pi) based on all (variant and invariant) loci. The GroupGenome v2.6.1 package in R [?] was used to calculate interpopulation genetic differentiation coefficient (FST) to measure genetic differentiation levels among populations.

1.6 Mantel Correlation Tests and AMOVA

To evaluate correlations among genetic distance, geographic distance, and environmental distance for seven *H. triflora* populations, we extracted climate information for each population using the "raster" and "rgdal" packages in R to construct environmental distance matrices from 19 bioclimatic variables. Geographic distance matrices were built using "geosphere," genetic distance matrices were read using "ape," and pairwise matrix correlations were tested with 9,999 Mantel permutations using "ade4," with results plotted using "ggplot."

Additionally, Arlequin 3.5 software [?] performed analysis of molecular variance (AMOVA). AMOVA calculates variance (δ^2) among haplotypes (including alleles) or genotypes based on evolutionary distance to infer variation among different genetic components.

1.7 Effective Population Size Analysis

SMC-based methods for inferring population history heavily rely on recombination maps from complete genomes and are unsuitable for species lacking reference genomes [?]. In contrast, site frequency spectrum (SFS)-based methods do not depend on linkage or continuity between loci [?] and offer high resolution for recent population size changes [?, ?]. Therefore, we used Stairway Plot v2 [?], an SFS-based method, to infer Ne changes in *H. triflora* from 100 thousand years ago (kya) to 1 kya. A non-missing, linkage disequilibrium (LD)-free SNP dataset was generated. The VCF file was processed using the easySFS python script (https://github.com/isaacovercast/easySFS) to construct a one-dimensional (1D) SFS. The SFS was specified as folded to track minor allele frequencies. Projection values maximizing the number of retained loci were selected to output SFS information, which was then input into the blueprint file required by Stairway Plot v2. The mutation rate was set to 1.0×10^{-8} per site per generation, with a generation time of 2 years. Two hundred bootstrap replicates were used to calculate median Ne and 95% confidence intervals, with results visualized in R.

1.8 Potential Distribution Modeling

Distribution point coordinates were obtained through GPS field positioning during material collection and from databases including the Chinese Virtual Herbarium (http://www.cvh.org.cn/) and Global Biodiversity Information Facility (https://www.gbif.org/), yielding over 200 coordinate records for *H. triflora* across Southeast Asia. All data were standardized according to MaxEnt model requirements (http://biodiversityinformatics.amnh.org/open_{source}/maxent). To prevent overfitting during maximum entropy modeling, distribution points were filtered to retain only one point per 50 km^2, resulting in 29 distribution points with latitude and longitude information compiled into a CSV file.

Nineteen bioclimatic variables were downloaded from WorldClim 2.1 (http://www.worldclim.org) for the mid-Holocene (MID, ~6,000 years BP), contemporary period (1970–2000), and future period (2081–2100) at 2.5-minute spatial resolution (5 km). Future climate scenarios selected two representative CO_2 concentration pathways: RCP2.6 and RCP8.5, representing the lowest (1.0 °C) and highest (2.0 °C) projected global mean temperature increases relative to present.

MaxEnt software simulated potential distribution patterns under different climate scenarios using 75% training and 25% testing datasets. Species distribution and environmental data were imported into MaxEnt, with jackknife model iterations set to 500 to calculate contribution rates of environmental variables, plot response curves, and generate prediction maps. Results were loaded into ArcGIS 10.2 to extract the Southeast Asia region. Based on habitat suitability index (HSI), areas were classified into four levels: HSI ≤ 0.1 (unsuitable), $0.1 <$ HSI ≤ 0.3 (low suitability), $0.3 <$ HSI ≤ 0.5 (medium suitability), and HSI

> 0.5 (high suitability). Model performance was evaluated using ROC curves.

Results

2.1 Sequencing Data Overview

RAD-seq was performed on 34 *H. triflora* samples, with results stored in FASTQ format containing sequence and quality information. After quality assessment and filtering, 37.91 Gb of clean data were obtained, averaging 1.12 Gb per sample. Individual sample read counts ranged from 2,512,919 to 55,093,740, with mean GC content of 36.2% and mean Q30 of 92.9% across all samples, indicating high sequencing quality meeting downstream analysis requirements. De novo analysis using Stacks yielded an average of 68,226 loci per sample with mean fragment coverage depth of 8.61X. After VCFtools filtering, 38,667 high-density SNP loci were retained for analysis.

2.2 Population Genetic Structure Analysis

Structure analysis revealed the minimum CV error at $K = 3$ [Figure 2A: see original paper], indicating that 34 individuals could be divided into three genetic clusters [Figure 2B: see original paper]. The maximum likelihood phylogenetic tree constructed from 38,667 SNPs showed that individuals from each population clustered on the same genetic branch with 100% support [Figure 2C: see original paper]. Changwangxi (CWC), Bumaocun (BMC), and Dongxingcun (DXC) showed similar genetic structures as Group 1 (Cluster1). Bubicun (BBC), Bushicun (BSC), and Wenxuancun 2 (WXC2) displayed closer genetic relationships as Group 2 (Cluster2). Wenxuancun 1 (WXC1) exhibited a distinct genetic structure as Cluster3. Principal component analysis (PCA) based on principal components 1 and 2 [Figure 3: see original paper] showed each population clustering tightly, consistent with phylogenetic results. DXC, CWC, BMC, BBC, and WXC2 distributed relatively closely, while WXC1 was separated from other populations along PC2, consistent with Structure analysis results.

2.3 Population Genetic Diversity Analysis

Statistical analysis of genetic diversity across seven populations revealed mean observed heterozygosity (Ho), expected heterozygosity (He), and nucleotide diversity (Pi) of 0.1569, 0.1654, and 0.1865, respectively. All populations contained private alleles, averaging 828.29 per population, with BSC showing the highest number (1,445) and lowest observed heterozygosity (0.0791). Nucleotide diversity (Pi) ranged from 0.01502 to 0.2399, with Cluster1 populations in the eastern sampling area (BMC, CWC, DXC) exhibiting higher genetic diversity. Inbreeding coefficient (FIS) ranged from -0.0061 to 0.1683, with a mean of 0.06167; CWC and WXC2 showed slightly negative values, indicating heterozygote excess.

Pairwise genetic differentiation coefficients (FST) among seven populations based on 38,667 SNPs ranged from 0.2035 to 0.5846. The smallest FST (0.2035) between CWC and BMC indicated the most frequent genetic exchange. All other pairwise FST values exceeded 0.25, indicating high genetic differentiation among the seven populations. Wenxuancun 1 (WXC1) showed the greatest differentiation, with mean FST of 0.5281 relative to the other six populations.

2.4 AMOVA and Mantel Tests

AMOVA results showed that when seven populations were grouped into three clusters, 9.35% of genetic variation occurred among groups (FCT = 0.09353), indicating non-significant differentiation among groups. Variation within populations accounted for 52.65% of total variation, while variation among populations within groups represented 38%. Mantel tests revealed significant positive correlation between genetic distance and environmental distance (R = 0.43457, P = 0.0447) [Figure 4A: see original paper], but no significant relationship between genetic distance and geographic distance (P = 0.3393) [Figure 4B: see original paper]. Environmental distance showed significant positive correlation with geographic distance (R = 0.6432, P = 0.0082) [Figure 4C: see original paper].

2.5 Population Size History Analysis

To infer temporal changes in population size using Stairway Plot, 38,667 unlinked SNPs were first screened. Stairway Plot inferred effective population size changes from 100 kya to 1 kya [Figure 5: see original paper]. The population showed contraction trends after the Last Glacial Period (LGP) at 11.7 kya. During the mid-Holocene (MID), particularly between 4–2 kya, *H. triflora* experienced pronounced declines in effective population size.

2.6 Potential Distribution Patterns Under Different Scenarios

MaxEnt model AUC values exceeded 0.952, indicating reliable predictions of suitable distribution areas. Temperature seasonality, minimum temperature of the coldest month, mean annual temperature, and elevation contributed most to the model, identifying them as primary environmental factors influencing *H. triflora* distribution.

Results showed substantial reduction in total suitable area from the MID to contemporary period [FIGURE:6A, 6B; TABLE:5], primarily reflected in 29% and 34% decreases in highly and moderately suitable areas, respectively. Under future RCP2.6 scenarios [Figure 6C: see original paper], total suitable area increased by 4.74% compared to contemporary, mainly due to a 33.40% increase in low-suitability area. Under RCP8.5 scenarios [Figure 6D: see original paper], total suitable area changed little, but highly suitable area decreased substantially by 23.32%, moderately suitable area decreased by 4.91%, while low-suitability area increased by 72.85%. These findings indicate that future climate change, particularly rising CO_2 concentrations and global temperatures, will significantly impact highly suitable habitats, causing extensive loss or conversion to low-suitability areas, with suitable habitats in the Malay Archipelago nearly disappearing entirely.

Discussion

3.1 Genetic Diversity of *H. triflora*: Habitat Fragmentation Reduces Genetic Diversity

Since the 1990s, habitat fragmentation has gradually gained attention in China as a research hotspot in conservation biology and ecology, initially focusing on tropical rainforests [?, ?] and subsequently on subtropical forests [?]. Island studies account for only 2% of research across different habitat types [?]. Investigating species genetic diversity not only reflects potential for environmental adaptation but also enables effective assessment of current survival status, particularly for endangered species [?]. Studies have found that widespread species exhibit higher genetic diversity than geographically narrow, rare, and endangered species due to genetic drift and inbreeding [?], with low genetic diversity representing a common characteristic of endangered or narrowly distributed species [?, ?]. Nucleotide diversity serves as a comprehensive indicator of population genetic diversity [?]. Our RAD-seq-based population genetic analysis of seven *H. triflora* populations revealed low genetic diversity (Pi = 0.1865; Ho = 0.1569; He = 0.1654). Similar findings were reported in the endangered fern *Adiantum nelumboides*, with mean Pi, Ho, and He of 0.373, 0.106, and 0.135, respectively [?]. Herbaceous plants generally exhibit lower genetic diversity than woody plants [?]. The critically endangered tree *Rhododendron hemsleyanum* studied using RAD-seq showed genetic diversity of Pi = 0.2410 and He = 0.2267 [?]. Compared to congeneric species, *H. triflora*'s mean He (0.1654) was slightly lower than that of *Impatiens macrovexilla* (0.186) [?]. Additionally, mean observed heterozygosity was lower than expected heterozygosity, suggesting heterozygote deficiency and some degree of inbreeding—genetic consequences of habitat fragmentation. Dixo et al. [?] found positive correlation between genetic diversity and fragment area in toad populations in Brazil's Atlantic Coastal Forest, with habitat fragmentation negatively affecting gene flow among remaining habitats. Such restrictions can increase inbreeding, cause heterozygote deficiency, and reduce genetic diversity.

Genetic distance studies among populations facilitate investigation of genetic diversity differences and describe population genetic structure. Genetic differentiation coefficient (FST) serves as a key metric, with values above 0.25 indicating substantial differentiation [?]. Among 21 pairwise FST comparisons among seven *H. triflora* populations, 20 exceeded 0.25, indicating high differentiation. AMOVA showed that 38% of genetic variation occurred among populations within groups, while 52.65% occurred within populations. No previous studies have examined *H. triflora* genetic diversity. Balsaminaceae includes only *Hy-*

drocera and *Impatiens* genera, which are taxonomically closely related, allowing results from *Impatiens* studies to provide preliminary references. Zhong et al. [?] used ISSR markers to study *Impatiens hainanensis*, finding that 92% of variation occurred within populations and only 8% among populations—consistent with our *H. triflora* results showing greater within-population variation. Similarly, studies on *Impatiens macrovexilla* using ISSR [?] or inter-primer binding site (iPBS) markers [?] found greater within-population than among-population genetic diversity, indicating extensive genotypes at each collection site. Clustering, genetic structure, and PCA analyses revealed that our Hainan *H. triflora* populations could be divided into three groups with different genetic components. Despite these distinct components, differentiation among groups was low (FCT = 0.09353). Considering *H. triflora*'s habitat characteristics and anthropogenic barriers such as buildings and farmland, we hypothesize that subpopulation formation resulted from habitat isolation.

Mantel tests among environmental, genetic, and geographic distances showed significant positive correlation between environmental and geographic distances ($P = 0.0082$) and between environmental and genetic distances ($P = 0.0447$), indicating that genetic differentiation correlates with environmental differences, which are closely associated with geographic distance. However, no significant correlation existed between geographic and genetic distances ($P = 0.3393$), meaning genetic distance does not increase with isolation distance. This differs from *Impatiens hainanensis*, which showed significant high correlation between genetic and geographic distances ($r = 0.952$, $P < 0.01$) [?]. This may be because *H. triflora*, unlike typical aquatic plants, possesses fleshy pseudo-berries exposed on or submerged in water that can be consumed by birds or fish, making geographic distance a poor predictor of genetic distance. It may also relate to *H. triflora*'s narrow distribution and close geographic proximity among populations. Overall, high genetic differentiation, correlation between genetic and environmental distances but not geographic distances, suggests that genetic differentiation among *H. triflora* populations is primarily driven by environmental distance resulting from habitat fragmentation rather than geographic distance.

3.2 Population History of *H. triflora*: Drastic Population Decline Following Qiongbei Volcanic Eruptions ~10,000 Years Ago

Species population history dynamics result from combined effects of historical events and genetic characteristics [?], predicting current and future viability [?]. Using extensive SNP markers for population size history analysis, we found that *H. triflora* maintained relatively stable effective population size from 100 kya until before the Last Glacial Period, likely associated with global warming during the early Holocene and flourishing Hainan flora [?]. Effective population size began continuous decline after the Last Glacial Period (LGP: 11.2 kya) to present [Figure 5: see original paper]. Qiongbei volcanic group eruptions occurred approximately 10,000 years ago [?], coinciding with the onset of *H. triflora* population decline. We hypothesize that volcanic eruptions caused land destruction, climate change, or other environmental alterations that fragmented *H. triflora* habitats and restricted gene flow. During the mid-Holocene, effective population size declined markedly. The mid-Holocene represents an important period of global human activity expansion, during which human societies transitioned from hunter-gatherer lifestyles to agriculture and settlement, significantly impacting environmental evolution. In the mid-to-late Neolithic, over 10,000 people foraged in Hainan, competing with plants for land and causing substantial habitat changes [**?**]. Pollen studies from Hainan northern crater sediments revealed that agricultural activities began significantly affecting lowland plants as early as 3,000 years BP [**?**]. We therefore hypothesize that *H. triflora* effective population size was first impacted by Qiongbei volcanic eruptions and subsequently by human activities, with combined effects causing continuous habitat changes. This effective population size reduction likely led to genetic diversity loss and population differentiation, gradually forming the current distribution pattern.

3.3 Potential Distribution Patterns of *H. triflora*: High CO_2 Emission Scenarios Convert Highly Suitable Habitats to Low-Suitability Areas

Environmental factor contribution analysis revealed that temperature seasonality, minimum temperature of the coldest month, and mean annual temperature collectively contributed 67.6% to the model, consistent with *H. triflora*'s thermophilic nature and growth cessation below 15 °C [**?**]. Elevation contributed the fourth-highest rate (7.4%), identifying it as a primary environmental factor influencing distribution and supporting findings from *Impatiens hainanensis* studies showing higher genetic diversity in low-elevation than high-elevation populations, indicating elevation as an important factor affecting gene flow [**?**].

Potential distribution modeling for *H. triflora* across Southeast Asia under different climate scenarios revealed substantial total suitable area reduction from the mid-Holocene to contemporary period , with Hainan Island's suitable areas shifting from high- and medium-suitability distribution patterns to contemporary low-suitability dominance [FIGURE:6A, 6B]. This aligns with population history analysis, indicating threats to *H. triflora* survival since the mid-Holocene.

Compared to contemporary climate, future climate change scenarios showed little change in total potential distribution area but a 33.40% increase in low-suitability area. Particularly under RCP8.5, highly suitable area decreased by 23.32% while low-suitability area increased by 72.85%, with suitable habitats in the Malay Archipelago nearly disappearing entirely [Figure 6D: see original paper]. According to the IPCC Sixth Assessment Report, under SSP5-8.5 scenarios, CO_2 emissions will double by 2050, raising global mean temperature by 4.4 °C by 2100. We hypothesize this relates to northward shift of the subtropical zone under warming conditions, causing species distributions to move toward higher latitudes and elevations [**?**], resulting in near-complete disappearance of *H. triflora* suitable habitats in the Malay Archipelago. These results indicate that future climate change will further affect *H. triflora* suitable habitats, particularly threatening loss and conversion of highly suitable areas.

Human activities including land reclamation, logging, grazing, transportation, and reservoir construction constitute primary causes of global natural habitat loss and fragmentation threatening biodiversity [**?**, **?**]. Urbanization exacerbates habitat fragmentation, particularly for remnant small populations of native plants in urban areas.

For *H. triflora*, low genetic diversity may weaken adaptive potential during migration, leading to reduced effective population size. High within-population genetic differentiation indicates limited natural gene flow among populations. Significant positive correlations between environmental distance and both genetic and geographic distances suggest environmental factors play important roles in genetic differentiation, while geographic distance indirectly influences differentiation through environmental differences. Therefore, factors limiting *H. triflora* population development stem primarily from its inherently low genetic diversity and habitat conditions, specifically severe wetland habitat fragmentation compounded by human disturbance and urbanization that further destroy habitats and reduce population size, gradually leading to regional extinction. To address this endangered status, we recommend prioritizing wetland protection from anthropogenic destruction and implementing in-situ conservation to effectively improve population genetic characteristics. Additionally, artificial pollination should be employed to enhance gene flow and promote population development.

References

Aguilar R, Ashworth L, Galetto L, et al., 2006. Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis [J]. *Ecol Lett*, 9(8): 968-980.

Alexander DH, Novembre J, Lange K, 2009. Fast model-based estimation of ancestry in unrelated individuals[J]. *Genome Res*, 19(9): 1655-1664.

Allendorf FW, Lundquist LL, 2003. Introduction: population biology, evolution, and control of invasive species[J]. *Conserv Biol*, 24-30.

Cai C, Xiao J, Ci X, et al., 2021. Genetic diversity of *Horsfieldia tetratepala* (Myristicaceae), an endangered plant species with extremely small populations to China: implications for its conservation[J]. *Plant Syst Evol*, 307(4).

Cao YR, Ma YP, Zhang XJ, et al., 2022. Genetic characteristics of *Rhododendron hemsleyanum* based on SNP molecular markers[J]. *Journal of Yunnan University (Natural Sciences Edition)*, 44: 1-11.

Catchen J, Bassham S, Wilson T, et al., 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing[J]. *Mol Ecol*, 22(11): 2864-2883.

Catchen J, Hohenlohe PA, Bassham S, et al., 2013. Stacks: an analysis tool set for population genomics[J]. *Mol Ecol*, 22(11): 3124-3140.

Chen XY, 2000. Effects of habitat fragmentation on genetic structure of plant populations and implications for the biodiversity conservation[J]. *Acta Ecol Sin*, 20(5): 884-892.

Chung MY, Son SW, Herrando-Moraira S, et al., 2020. Incorporating differences between genetic diversity of trees and herbaceous plants in conservation strategies[J]. *Conserv Biol*, 34(5): 1142-1151.

Cole CT, 2003. Genetic variation in rare and common plants[J]. *Annu Rev Ecol Evol Syst*, 34(1): 213-237.

Danecek P, Auton A, Abecasis G, et al., 2011. The variant call format and VCFtools[J]. *Bioinform*, 27(15): 2156-2158.

Davey JW, Hohenlohe PA, Etter PD, et al., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing[J]. *Nat Rev Genet*, 12: 499-510.

Dixo M, Metzger JP, Morgante JS, et al., 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest[J]. *Biol Conserv*, 142(8): 1560-1569.

Excoffier L, Lischer HE, 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows[J]. *Mol Ecol Resour*, 10(3): 564-567.

Fan QC, Sun Q, Li N, et al., 2004. Periods of volcanic activity and magma evolution of Holocene in North Hainan Island[J]. *Acta Petr Sin*, 20(3): 533-544.

Girod C, Vitalis R, Leblois R, et al., 2011. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method[J]. *Genetics*, 188(1): 165-179.

Gitzendanner MA, Soltis PS, 2000. Patterns of genetic variation in rare and widespread plant congeners[J]. *Amer J Bot*, 87: 783-792.

González-Varo JP, Arroyo J, Aparicio A, 2009. Effects of fragmentation on pollinator assemblage, pollen limitation and seed production of Mediterranean myrtle (*Myrtus communis*) [J]. *Biol Conserv*, 142(5): 1058-1065.

Guindon S, Dufayard JF, Lefort V, et al., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0[J]. *Syst Biol*, 59(3): 307-321.

Gutenkunst RN, Hernandez RD, Williamson SH, et al., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data[J]. *PLoS Genet*, 5(10): e1000695.

Haddad NM, Brudvig LA, Clobert J, et al., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems[J]. *Sci Adv*, 1(2): e1500052.

Hamrick JL, Godt MJW, 1996. Effects of life history traits on genetic diversity in plants[J]. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1345): 1291-1298.

Hare MP, Nunney L, Schwartz MK, et al., 2011. Understanding and estimating effective population size for practical application in marine species management[J]. *Conserv Biol*, 25: 438-449.

Hewitt G, 2000. The genetic legacy of the Quaternary ice ages[J]. *Nature*, 405: 907-913.

Holsinger KE, Weir BS, 2009. Genetics in geographically structured populations: defining, estimating and interpreting[J]. *Nat Rev Genet*, 10(9): 639-650.

Honnay O, Jacquemyn H, 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation[J]. *Conserv Biol*, 21(3): 823-831.

Janssens SB, Smets EF, Alexander V, 2012. Floral development of *Hydrocera* and *Impatiens* (Balsaminaceae) reveals evolutionary trends in the most early-diverged lineages[J]. *Ann Bot*, (7): 1285-1296.

Jennersten O, 1988. Pollination in *Dianthus deltoides* (Caryophyllaceae): effect of habitat fragmentation on visitation and seed set[J]. *Conserv Biol*, 2: 359-366.

Jiang YX, Wang BS, Zang RG, et al., 2002. *Tropical Rainforest Biodiversity and Its Formation Mechanism in Hainan Island*[M]. Beijing: Science Press: 30-84.

Jing ZY, Cheng KG, Shu H, et al., 2023. Whole genome resequencing approach for conservation biology of endangered plants[J]. *Biodivers Sci*, (31): 22679.

Lamont BB, Klinkhamer PGL, Witkowski ETF, 1993. Population fragmentation may reduce fertility to zero in *Banksia goodii*—a demonstration of the Allee effect[J]. *Oecologia*, 94: 446-450.

Lam-Tung N, Schmidt HA, Arndt VH, et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Mol Biol Evol*, (1): 268-274.

Lapierre M, Lambert A, Achaz G, 2017. Accuracy of demographic inferences from the site frequency spectrum: the case of the Yoruba population[J]. *Genetics*, 206: 439-449.

Laurance WF, Clements GR, Sloan S, et al., 2014. A global strategy for road building[J]. *Nature*, 513: 229-232.

Li H, Durbin R, 2011. Inference of human population history from individual whole-genome sequences[J]. *Nature*, 475(7357): 493-496.

Li MH, Song RS, Jiang YF, et al., 2008. Plant diversity in six evergreen broad-leaved forest fragments in East China[J]. *Acta Ecol Sin*, 28(3): 1137-1146.

Li XJ, Tang WX, Huang SX, et al., 2013. ISSR analysis for genetic diversity of *Impatiens macrovexilla* in Guilin[J]. *Guangdong Agric Sci*, 40(13): 142-144.

Liu XM, Fu YX, 2015. Exploring population size changes using SNP frequency spectra[J]. *Nat Genet*, 47(5): 555-559.

Liu XM, Fu YX, 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra[J]. *Genome Biol*, 21: 280.

Mable BK, 2019. Conservation of adaptive potential and functional diversity: integrating old and new approaches[J]. *Conserv Genet*, 20(1): 89-100.

Minh BQ, Nguyen M, Haeseler AV, 2013. Ultrafast approximation for phylogenetic bootstrap[J]. *Mol Biol Evol*, 30(5): 1188-1195.

Morrell PL, Buckler ES, Ross-Ibarra J, 2012. Crop genomics: advances and applications[J]. *Nat Rev Genet*, 13: 85-96.

Narum SR, Buerkle CA, Davey JW, et al., 2013. Genotyping-by-sequencing in ecological and conservation genomics[J]. *Mol Ecol*, 22: 2841-2847.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, et al., 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R[J]. *Mol Biol Evol*, 31(7): 1929-1936.

Situ SJ, 1987. *Historical Land Development Research in Hainan Island*[M]. Haikou: Hainan People's Publishing House, 9-19.

Supple MA, Shapiro B, 2018. Conservation of biodiversity in the genomics era[J]. *Genome Biol*, 19: 131.

Sun WY, Shu JP, Gu YF, et al., 2022. Conservation genomics analysis revealed the endangered mechanism of *Adiantum nelumboides*[J]. *Biodivers Sci*, 30(7): 21508.

Tingley R, Vallinoto M, Sequeira F, et al., 2014. Realized niche shift during a global biological invasion[J]. *PNAS*, 111: 10233-10238.

Wang JF, Lv DR, Huang S, et al., 2017. The resource status and investigation analysis of endangered aquatic plants in Hainan Province[J]. *Chin Hort Abstr*, 33(12): 4.

Wang XL, Liu Q, Yu DL, et al., 2016. Diversity of epiphytic orchids in fragmental limestone forests in Xishuangbanna, China[J]. *Guihaia*, 36(2): 162-169.

Wilcove DS, McLellan CH, 1986. Habitat fragmentation in the temperate zone[J]. *Conserv Biol*, 6: 237-256.

Wilcove DS, Rothstein D, Dubow J, et al., 1998. Quantifying threats to imperiled species in the United States[J]. *BioScience*, 48(8): 607-615.

Wright S, 1978. *Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations*[M]. Chicago: University of Chicago Press.

Wu J, Liu ZM, 2014. Effect of habitat fragmentation on biodiversity: a review[J]. *Chin J Ecol*, 33(7): 1946-1952.

Wu QQ, Liang ZS, Liu JJ, et al., 2017. Effects of habitat fragmentation on biodiversity in China[J]. *Chin J Ecol*, 36(9): 2605-2614.

Young A, Boyle T, Brown T, 1996. The population genetic consequences of habitat fragmentation for plants[J]. *Trends Ecol Evol*, 11(10): 413-418.

Zhong YF, Hu XY, Song XQ, et al., 2014. ISSR analysis on population genetic diversity of *Impatiens hainanensis* (Balsaminaceae), an endemic species to Hainan Island[J]. *Chin J Trop Crops*, 35(6): 1041-1046.

Zhou J, Guan S, Song Q, et al., 2020. Retrotransposon-based genetic variation and population structure of *Impatiens macrovexilla* Y. L. Chen in natural habitats and the implications for breeding[J]. *Sci Hortic*, 276: 109753.

Zhu H, Xu ZF, Wang H, et al., 2001. Over 30-year changes of floristic composition and population structure from an isolated fragment of tropical rain forest in Xishuangbanna[J]. *Acta Bot Yunnan*, 23(4): 415-427.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.