Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genetic Diversity and Chemotypic Variation of Gentiana rhodantha: Insights into the Genetic Basis of Active Compound Accumulation

Qingqing Ye
Guizhou University of Traditional Chinese Medicine
Lingling Wei
Guizhou University of Traditional Chinese Medicine
Weixiang Xiao
Bijie Vocational and Technical College
Xing Tan
Guizhou University of Traditional Chinese Medicine
Li Yang
Guizhou University of Traditional Chinese Medicine
Deqiang Ren
Guizhou University of Traditional Chinese Medicine
Ning Ding
Guizhou University of Traditional Chinese Medicine
Zhikun Wu
47390933@qq.com
Guizhou University of Traditional Chinese Medicine

Research Article

Keywords: Gentiana rhodantha, Population genomics, Genotyping-by-sequencing (GBS), Genetic differentiation, Germplasm resources, Bioactive constituents

Posted Date: April 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9367916/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on June 15th, 2026. See the published version at https://doi.org/10.1186/s12870-026-09219-w.

Abstract

Gentiana rhodantha is a commonly used ethnic medicine in Southwest China, possesses high medicinal value. Although the quality of medicinal materials from different producing areas varies considerably, no studies to date have investigated the chemotypes or genetic diversity of *G. rhodantha* across its distribution range. In this study, we combined genotyping-by-sequencing (GBS) and high-performance liquid chromatography (HPLC) to analyze the genetic diversity and active component content of *G. rhodantha* from ten different producing areas. The relationships of these parameters with geographical and environmental factors were also examined and discussed. A total of 4,524,308 high-quality single nucleotide polymorphism (SNP) loci were obtained. Principal component analysis and phylogenetic analysis based on these SNPs divided the ten populations into three major genetic lineages. The cross-validation error rate reached its minimum at K = 10, suggesting the presence of ten distinct subgroups within the sampled populations. Genetic diversity within populations was low, whereas genetic differentiation among populations was significant. The contents of six active components varied widely among populations. The content of mangiferin varied by up to 3.25-fold across populations, while that of swertiamarin showed a 61-fold difference. Genetic clustering corresponded well with chemotypes. Moreover, both genetic diversity and the degree of genetic differentiation were significantly correlated with component content and the extent of variation. Active components exhibited specific associations with geographical factors (latitude, longitude, and altitude). Positive correlations predominated among active components, although negative correlations were also observed for certain component pairs.This study reveals the relationship between genetic variation and active component accumulation in *G. rhodantha*, and clarifies the role of geographical factors in shaping chemotypic differentiation. These findings provide an important scientific basis for the conservation, evaluation, elite germplasm screening, and standardized cultivation of this medicinal plant.

Introduction

Genetic diversity representing the total repertoire of heritable variations accumulated throughout a species' evolutionary history, constitutes the fundamental basis for its adaptation to environmental changes and its capacity for population survival and proliferation. Furthermore, it serves as a critical indicator for deciphering evolutionary trajectories and assessing future adaptive potential [1–2]. Genetic differentiation among geographical populations, arising from variations in genetic diversity and local adaptation processes, can further influence phenotypic traits, including the biosynthesis of secondary metabolites that are intrinsically linked to the quality of medicinal materials. For instance, in *Scutellaria baicalensis*, significant correlations have been established between population-level genetic diversity and the content of its major bioactive constituents, such as baicalin and baicalein. Population genetic structure analyses have revealed that specific populations harbor unique haplotypes of genes involved in flavonoid biosynthesis, thereby shaping the geographical chemotypic patterns of this medicinal plant [3]. Studies on *Panax ginseng* have demonstrated that disparities in genetic diversity between wild and cultivated populations directly impact the composition and accumulation of ginsenosides, with the higher genetic diversity observed in wild populations being associated with a more complex ginsenoside profile [4]. Similarly, in the analysis of cultivated *Panax quinquefolius* provenances, accessions exhibiting higher genetic diversity were frequently accompanied by an enriched content of flavonoid bioactive components, underscoring that genotypic diversity is a pivotal determinant of medicinal material quality [5]. Collectively, these investigations elucidate that genetic diversity not only underpins species evolution but also directly governs the formation and chemical stability of medicinal material quality by modulating genetic

variation in key genes within secondary metabolic pathways. This provides a crucial genetic foundation for the conservation and sustainable utilization of medicinal plant resources.

Genotyping-by-sequencing (GBS) technology, which integrates restriction enzyme digestion with high-throughput sequencing, enables the efficient acquisition of hundreds of thousands to millions of genome-wide single nucleotide polymorphisms (SNPs) [6–7]. Characterized by its high throughput, cost-effectiveness, and abundant information yield, GBS has become a cornerstone tool in population genetics and evolutionary research [8]. As a next-generation molecular marker, SNPs are distinguished by their high polymorphism, widespread genomic distribution, and co-dominant inheritance, offering unique value for dissecting genetic architecture and unraveling species evolutionary mechanisms at a genome-wide scale [9–10]. In recent years, GBS has been successfully applied in genetic studies across diverse species. For instance, in a population genomic analysis of *Anemone shikokiana*, 52,231 high-quality SNPs were generated via GBS, revealing two distinct genetic lineages with low intra-population diversity, where geographical isolation has driven significant genetic differentiation and limited gene flow [11]. Similarly, a GBS-based investigation of *Tsoongiodendron odorum* uncovered generally low genetic diversity within the species, with genetic variation primarily residing within populations, minimal differentiation among populations, and relatively high levels of gene flow [12]. Concurrently, GBS technology has demonstrated significant application potential in medicinal plant research. For example, a study employing GBS on 13 wild populations of *Ardisia violacea*, based on 246,000 SNP loci, systematically revealed a pattern of low overall genetic diversity coupled with significant inter-population differentiation [13]. In research on different chemotypes of *Artemisia annua*, GBS not only systematically elucidated the population genetic background but also, through genome-wide association analysis, identified genetic variants closely linked to the artemisinin biosynthetic pathway [14]. Collectively, these studies demonstrate that GBS, combined with SNP markers, provides an efficient and economical means of acquiring genome-wide genetic information. This approach offers robust support for elucidating species genetic structure, evolutionary history, and adaptive mechanisms, thereby laying a scientific foundation for the conservation of genetic resources in medicinal plants, the selection of elite germplasm, and the formulation of breeding strategies [15].

Gentiana rhodantha Franch. (Family Gentianaceae) is a perennial herbaceous species widely utilized as a traditional ethnic medicine among the Miao and Tujia communities in Southwest China. This species is officially documented in both the Chinese Pharmacopoeia (2025 edition) and the Quality Standards for Chinese Medicinal Materials and Ethnic Medicines of Guizhou Province (2003 edition), and is commonly known as “Xuelimei” or “Xuedanlong” [16]. The dried whole plant or root is used for medicinal purposes. In traditional Chinese medicine, it is characterized as having a cold property and a bitter taste, and is recognized for its effects in clearing heat, eliminating dampness, detoxifying, purging fire, alleviating coughing, and cooling the blood. Clinically, it is frequently employed in the treatment of conditions such as cough and dyspnea caused by lung heat, pneumonia, jaundice, dysentery, and sores or boils. Furthermore, it serves as a critical raw material in several Chinese patent medicines, including “Feilike Mixture” and “Xiao'er Xiaoji Zhike” Oral Liquid [17–18]. Phytochemical investigations have revealed that *G. rhodantha* primarily contains iridoids, flavonoids, and phenolic acids, which contribute to its diverse pharmacological properties, including anti-inflammatory, antioxidant, and antibacterial activities [19]. The species is predominantly distributed across the provinces of Guizhou, Yunnan, Sichuan, and Hunan in China, typically inhabiting slopes, shrublands, sparse forests, and rocky crevices at altitudes ranging from 300 to 2500 m [20]. However, due to multiple stressors, including habitat fragmentation, over-exploitation, and climate change, the distribution range of its wild populations is progressively contracting, and population sizes are continuously declining, leading to an increasingly precarious survival status. Consequently, a systematic

assessment of its genetic diversity and population structure is of substantial theoretical and practical significance for formulating science-based conservation strategies and ensuring the sustainable utilization of this medicinal resource.

Recent research on *G. rhodantha* has primarily focused on quality evaluation, analytical profiling, and the assessment of medicinal material characteristics across different production areas [21–23]. These investigations have yielded substantial knowledge regarding its growth characteristics and pharmacologically active constituents, thereby laying a foundation for its scientific cultivation, efficient extraction, and rational application. Relevant studies have demonstrated that the contents of bioactive compounds in *G. rhodantha* vary significantly among different habitats and geographical origins [24], suggesting differentiation in medicinal components across populations. However, the mechanisms underlying this differentiation remain unclear. To address this knowledge gap, the present study employed GBS and high-performance liquid chromatography (HPLC) to investigate 149 *G. rhodantha* samples collected from 10 populations across its main distribution areas in the four Chinese provinces of Hunan, Sichuan, Yunnan, and Guizhou. Through integrated analysis of population genetic diversity and bioactive compound content, we elucidated the association patterns between genetic variation and the accumulation of active constituents in *G. rhodantha*. These findings provide a critical scientific basis for germplasm resource evaluation, superior provenance selection, and the standardized cultivation and utilization of this medicinal species.

Materials and methods

Experimental samples

A total of 149 wild *G. rhodantha* individuals were sampled from its primary distribution areas, encompassing one population in Hunan Province and three populations each in Yunnan Province, Guizhou Province, and southwestern Sichuan Province. Fifteen individuals per population (14 individuals for the MB population). All samples were identified by Professor Zhikun Wu from Guizhou University of Traditional Chinese Medicine as *Gentiana rhodantha* (family Gentianaceae), and all voucher specimens are deposited in the Herbarium of Guizhou University of Traditional Chinese Medicine (GZTM, Cheng-gang Hu, 2274547063@qq.com). This sampling scheme was designed to systematically investigate the genetic diversity of the species across its representative geographical range. For the analysis of bioactive compound contents, eight populations were selected; the MB and QL populations were excluded from this component due to limited sample sizes available from field collections. All samples were preserved using silica gel desiccation at low temperatures to ensure the integrity of subsequent experimental analyses. Detailed information regarding the sampling locations and population characteristics is provided in Table 1 and illustrated in Fig. 1.

Table 1
Geographical distribution of the 10 populations of *Gentiana rhodantha*

Sample	Species	Location	Longitude (E)	Latitude (N)	Altitude (m)	Habitat	No. of voucher specimens
HY	*G. rhodantha*	Huayuan County, Hunan Province	109.26	28.47	391	Moist cliff	WLL-2024-005
XD	*G. rhodantha*	Xundian County, Yunnan Province	103.39	25.52	2085	Hillside shrubland	WLL-2024-008
PG	*G. rhodantha*	Puge County, Sichuan Province	102.6	27.37	2000	Sparse pine forest, grassy slope	WLL-2024-011
RH	*G. rhodantha*	Renhuai City, Guizhou Province	106.25	28.09	584	Roadside grassland	WLL-2024-012
XY	*G. rhodantha*	Xiangyun County, Yunnan Province	100.82	25.4	2187	Grassy slope, cliff	WLL-2024-013
HS	*G. rhodantha*	Huishui County, Guizhou Province	106.89	26	1246	Sparse pine forest	WLL-2024-015
HZ	*G. rhodantha*	Hezhang County, Guizhou Province	104.59	27.25	1868	Hillside shrubland	WLL-2024-016
ZY	*G. rhodantha*	Zhenyuan County, Yunnan Province	100.79	24.19	1338	Grassland on soil embankment	WLL-2024-017
MB	*G. rhodantha*	Mabian County, Sichuan Province	103.49	28.99	703	Edge of sparse bamboo forest, moist clif	WLL-2024-018
QL	*G. rhodantha*	Qionglai City, Sichuan Province	103.33	30.48	558	Edge of sparse forest	WLL-2024-019

GBS library construction and sequencing

Library construction, sequencing, and data analysis for all samples were performed on the BGI platform at Shanghai Majorbio Bio-pharm Technology Co., Ltd., utilizing the DNBSEQ-T7 sequencing system with the PE150

sequencing mode. The GBS library construction protocol commenced with an in silico evaluation and validation of the *G. rhodantha* reference genome to identify an optimal restriction enzyme combination, primarily employing MseI and TaqaI. This selection aimed to achieve appropriate fragment size ranges, an optimal number of tags, and sufficient reference genome coverage. Subsequently, genomic DNA (0.1–1 μg) was digested with the restriction enzyme MseI. Following digestion, Solexa P1 and P2 adapters, each incorporating a 6-bp barcode sequence, were ligated to the fragment ends. A secondary digestion using TaqaI was then performed to modulate the number of tags. The final GBS library was generated through PCR amplification, fragment size selection, and purification using AMPure XP beads, yielding high-quality libraries suitable for subsequent sequencing analysis.

Bioinformatic analysis methods

Quality control of raw sequencing data

To ensure high data quality for downstream analyses, rigorous quality control of the raw sequencing reads was performed using Fastp [25]. The quality control pipeline primarily involved the following steps: adapter sequences were first removed from the reads, and sequences containing non-AGCT bases at the 5' end were discarded. Subsequently, read ends with sequencing quality values below Q20 were trimmed, and reads containing up to 10 ambiguous 'N' bases were filtered out. Finally, after the removal of adapters and low-quality bases, short reads with a length of less than 25 bp were discarded, yielding high-quality sequencing data for subsequent analyses.

Variant discovery and molecular marker development

Following quality control of the raw sequencing data, the cleaned reads were assigned to individual samples using the axe R package. Sequencing data from each individual were subsequently pooled at the population level, and consensus sequences were generated using Stacks [26]. Variants were concurrently detected across all samples utilizing the populations module of Stacks. Single nucleotide polymorphisms (SNPs) with a call rate exceeding 70% and a minimum depth of coverage greater than> 2X were retained for subsequent population genetic analyses.

Phylogenetic analysis

Following the identification of high-quality SNP markers, a maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE2 under the GTR + I+G4 substitution model with 1,000 bootstrap replicates [27]. For comparative purposes, a neighbor-joining (NJ) tree was also constructed using FastTree with the -gtr -gamma parameters and 1,000 bootstrap replicates [28].

Population structure analysis

Population genetic structure was assessed using ADMIXTURE [29]. Clustering analysis was performed on all samples assuming a range of genetic clusters from 1 to 20. The optimal number of clusters was determined by identifying the K value corresponding to the lowest cross-validation (CV) error.

Principal component analysis (PCA)

Principal component analysis (PCA) was conducted using Plink to perform mathematical statistical analysis on the sequencing data [30]. This linear transformation reduced the dimensionality of the multivariate dataset, extracting the primary characteristic components to elucidate genetic relationships and distances among samples, thereby providing a foundation for subsequent evolutionary analyses.

Genetic diversity analysis

Genetic diversity within each population was quantified using the populations module of Stacks [26]. A comprehensive suite of key population genetic parameters was calculated to thoroughly characterize the level of genetic diversity within each group, offering robust data support for subsequent investigations into evolutionary relationships and population differentiation.

Determination of bioactive compound contents

To investigate the variation in accumulation patterns of pharmacologically active constituents among different geographical populations of *G. rhodantha*, the contents of six major bioactive compounds-loganic acid, neomangiferin, swertiamarin, sweroside, mangiferin, and isoorientin-were quantified using HPLC in eight populations (HY, XD, PG, RH, XY, HS, HZ, and ZY). The MB and QL populations were excluded from this analysis due to limited sample availability. Sample preparation involved accurately weighing 0.5 g of powdered material, followed by ultrasonic extraction with 15 mL of 70% methanol for 30 min. The resulting extract was filtered, and the subsequent filtrate was collected for analysis. A mixed standard solution containing the six reference compounds was prepared for calibration. Chromatographic separation was achieved using an Agilent TC-C18 column (250 mm × 4.6 mm, 5 um) maintained at 28°C, with detection at 242 nm. The mobile phase consisted of acetonitrile and 0.03% phosphoric acid applied in a gradient elution at 0.8 mL/min. The injection volume was 8 uL, with a total run time of 52 min. Under these conditions, all target compounds exhibited well-resolved peak shapes with retention times consistent with those of the corresponding reference standards, confirming the suitability of this method for quantitative analysis.

Integrative analysis of genetic and chemical data

To elucidate the intrinsic relationship between the population genetic structure and the accumulation of bioactive constituents in *G. rhodantha*, an integrative analysis was performed combining genetic data with quantitative chemical profiles. Initially, genetic clustering characteristics of the different populations were delineated based on principal component analysis (PCA), phylogenetic reconstruction, and population structure analysis. This facilitated an exploration of the correspondence between genetic relatedness and chemotypic similarity among populations. Subsequently, correlation analyses between genetic diversity parameters and bioactive compound contents were conducted to assess the potential influence of genetic variation on the accumulation of secondary metabolites. Finally, utilizing the pairwise genetic differentiation index (FST) matrix, the relationship between the degree of genetic differentiation and disparities in bioactive compound contents was quantitatively evaluated. This analysis, integrated with the observed patterns of chemotypic similarity among population pairs exhibiting varying levels of genetic differentiation, revealed the regulatory role of genetic isolation in shaping chemotypic divergence.

Correlation analysis between geographical factors and active ingredient contents

To investigate the influence of geographical factors on the accumulation of bioactive components in *G. rhodantha*, correlation analysis was performed between geographical variables (longitude, latitude, and altitude) and the contents of six major active ingredients (loganic acid, neomangiferin, swertiamarin, sweroside, mangiferin, and isoorientin). Pearson's correlation coefficient was employed to assess linear relationships among variables, with coefficient values ranging from − 1 to 1, where positive and negative values indicate positive and negative correlations, respectively, and values closer to absolute 1 represent stronger correlations. Statistical analyses were conducted using SPSS software (version 26.0). Based on the analysis results, correlation heatmaps were generated using Origin software (version 2021) to visually illustrate the correlation patterns between geographical factors and active components, providing a reference for further elucidating the mechanisms by which environmental factors affect the quality formation of *G. rhodantha*.

Results

Sequencing data summary

Utilizing GBS technology, sequencing libraries were successfully constructed for 149 *G. rhodantha* individuals collected from southwestern China. Sequencing generated a total of 2,866,380,310 raw reads, with an average of 19,109,202 reads per sample. The total number of bases obtained was 401,293,243,400 bp, corresponding to an average of 267,528,828,9 bp per sample. The GC content across samples ranged from 40.33% to 42.28%, with a mean value of 40.97%. All samples exhibited Q20 scores above 98.96% and Q30 scores above 96.74%, indicating high sequencing accuracy, low base error rates, and high-quality effective data suitable for subsequent analyses (Supplementary Table 1).

SNP detection and marker development

In this study, a total of 10,627,975 raw SNP loci were identified across the 149 *G. rhodantha* samples. Following stringent quality control, 4,524,308 high-quality SNPs were retained for subsequent analyses. The counts of transitions and transversions ranged from 12,665 to 21,934 and 7,868 to 14,059, respectively, with transition/transversion ratios varying between 1.52 and 1.67. All ratios exceeded the empirical threshold of 1.5, confirming high sequencing data quality and accurate SNP detection, thereby providing a reliable foundation for downstream population genetic analyses. Analysis of SNP-based genetic diversity revealed that the number of heterozygous and homozygous mutations per sample ranged from 3,720 to 17,902 and 7,352 to 30,216, respectively (Supplementary Table 2). The extensive variation observed in both mutation types indicates a relatively high level of genetic diversity within *G. rhodantha* populations. This characteristic likely reflects the combined influence of various evolutionary factors, including mutation accumulation, recombination events, and natural selection, and holds significant implications for the environmental adaptability and evolutionary potential of this species.

Phylogenetic and population structure analysis

Principal component analysis

To investigate the genetic structure and clustering patterns among different geographical populations of *G. rhodantha*, PCA was performed on the 10 populations. The results revealed that the first principal component (PC1) and the second principal component (PC2) explained 20.81% and 14.92% of the total genetic variation, respectively, with a cumulative contribution of 35.73%. This indicates substantial genetic differentiation among populations from different geographical origins, exhibiting clear population stratification at the genomic level. The two-dimensional scatter plot based on principal component scores further elucidated the clustering relationships among populations. Joint analysis of PC1 and PC2 demonstrated that the 10 populations could be partitioned into three primary genetic clusters: PG, HZ, HS, XD, and HY clustered together; QL, MB, and RH formed a second cluster; and ZY and XY constituted a third cluster. Analysis of the PC1 versus PC3 scatter plot revealed that the HY population clustered independently along the PC1 axis, ZY and XY formed a distinct branch, while the remaining populations grouped into another major cluster. The PC2 versus PC3 scatter plot similarly supported the distinctiveness of the HY population, with PG, HZ, HS, and XD clustering together and the remaining populations forming another cluster (Fig. 2). Synthesizing the multi-dimensional PCA results, the HY population exhibited pronounced genetic distinctiveness and a higher degree of genetic differentiation. Conversely, close clustering relationships were observed between the ZY and XY populations, between MB and RH, and among HS and XD, suggesting closer genetic affinities among these population pairs or potential influences from similar habitat selection pressures, resulting in greater consistency in their genetic architecture (Fig. 3, Supplementary Table 3).

Population structure analysis

To elucidate the genetic structure and determine the optimal number of genetic groups within the *G. rhodantha* germplasm, population genetic structure was analyzed for the 149 individuals using ADMIXTURE. The optimal number of genetic clusters was identified by examining the trend of CV error rates across varying numbers of putative subpopulations (K values). The results demonstrated a continuous decrease in CV error with increasing K, reaching its minimum value of 0.28341 at K = 10. This indicates that partitioning the 149 *G. rhodantha* accessions into 10 genetic subpopulations represents the optimal grouping, suggesting that all individuals can be traced to 10 distinct ancestral gene pools (Fig. 4).

The population structure plot based on K = 10 (Fig. 5) further elucidated the ancestry components characterizing each subpopulation. Across the 10 genetic subpopulations, the vast majority of individuals exhibited a highly homogeneous ancestral genetic background, with the proportion of their corresponding ancestral component exceeding 0.999. This suggests that these accessions possess relatively independent genetic origins without significant introgression or admixture. A minority of individuals, however, displayed pronounced mixed genetic backgrounds, harboring genetic information from two to five different ancestral populations, with their maximum ancestral component proportions all falling below 0.999. Individuals with mixed ancestral backgrounds were primarily distributed across three populations: RH, HZ, and MB, specifically including RH3, RH6, RH15, HZ1, HZ3, HZ7, HZ8, HZ14, MB5, MB6, and MB11. Notably, the HZ population contained the highest number and proportion of individuals with mixed genetic backgrounds, suggesting that this population may have experienced more frequent or complex historical inter-population gene flow events or admixture processes.

Further analysis of the patterns of genetic introgression among different populations revealed significant shared genetic components between the RH and QL populations. The HZ population exhibited weak gene flow signals with the RH, HS, XD, and PG populations. Additionally, a subtle signal of genetic admixture was detected between the MB and QL populations. These patterns of introgression indicate close genetic affinities or historical gene

flow events among the relevant populations, a finding that is highly consistent with the population clustering relationships revealed by the principal component analysis.

Phylogenetic analysis

To further elucidate the genetic evolutionary relationships among *G. rhodantha* germplasm resources, a phylogenetic tree was constructed for the 149 accessions using the NJ method based on genome-wide SNP data. The phylogenetic analysis revealed that the 149 *G. rhodantha* samples could be clustered into three primary genetic lineages (Fig. 6): Lineage I comprised populations HS, XD, HZ, and HY; Lineage II consisted of populations XY, ZY, and PG; and Lineage III encompassed populations QL, MB, and RH. This clustering pattern is highly congruent with the genetic differentiation patterns revealed by principal component analysis and population structure analysis. The consistency among these three analytical approaches collectively elucidates a clear geographical genetic structure within *G. rhodantha* populations. These findings further substantiate the existence of significant genetic differentiation among different geographical populations, providing a robust phylogenetic foundation for subsequent germplasm resource evaluation and the formulation of conservation strategies.

Linkage disequilibrium analysis

To evaluate the linkage disequilibrium (LD) characteristics and infer the genetic diversity levels among different geographical populations of *G. rhodantha*, the relationship between the average linkage disequilibrium coefficient (r^2) and physical distance (kb) among markers was calculated using 168,057 high-quality SNPs. This analysis generated LD decay curves for each population, effectively reflecting the trend of LD attenuation with increasing genetic distance. The rate of this decay serves as an indirect indicator of a population's effective size, historical recombination events, and potential selective pressures. The results revealed substantial variation in LD levels among the 10 *G. rhodantha* populations. Ranked by decreasing LD magnitude (i.e., from highest to lowest r^2 values), the populations were ordered as follows: QL > XY > ZY > HZ > HY > MB > RH > HS > PG > XD. The difference in LD values between the populations exhibiting the highest and lowest levels exceeded 18-fold. A negative correlation was observed between LD coefficients and population genetic diversity; that is, higher degrees of linkage corresponded to lower genetic diversity. Based on this relationship, the genetic diversity levels of the 10 populations were inferred to be inversely ordered relative to LD magnitude: QL < XY < ZY < HZ < HY < MB < RH < HS < PG < XD.

Further analysis of the LD decay rate, quantified as the physical distance at which r^2 decays to half its maximum value, revealed that populations PG, HS, and XD exhibited similar decay rates, with half-decay distances ranging from 12.640 to 13.183 kb. This rapid decay suggests relatively high genetic diversity within these populations. Populations MB and RH also displayed comparable decay rates, with half-decay distances between 20.768 and 22.921 kb, indicating intermediate levels of genetic diversity. In contrast, populations QL and XY exhibited the slowest LD decay, with half-decay distances substantially exceeding those of the other populations. This pattern suggests relatively low genetic diversity, restricted intra-population recombination, smaller effective population sizes, or potential exposure to strong natural selection pressures in these two populations (Fig. 7, Table 2).

Table 2
LD-decay values for each population

popID	half.decay(kb)	decay01(kb)
Hs	13.183	47.649
Hy	31.342	113.292
Hz	42.319	152.975
Mb	22.921	82.853
Pg	12.829	46.368
Ql	228.726	826.826
Rh	20.768	75.067
Xd	12.640	45.685
Xy	106.079	383.462
ZY	57.675	208.486

Genetic diversity and population differentiation analysis

Population genetic diversity analysis

To comprehensively characterize the level of genetic diversity within *G. rhodantha* germplasm resources, population genetic parameters were systematically evaluated for the 10 geographical populations based on genome-wide SNP markers. The results revealed generally low genetic diversity at the population level for *G. rhodantha* (Table 3). Specifically, observed heterozygosity (Ho) ranged from 0.036 to 0.109, with a mean value of 0.075, while expected heterozygosity (He) ranged from 0.025 to 0.153, with a mean of 0.097. With the exception of the QL and ZY populations, where Ho slightly exceeded He suggesting a minor excess of heterozygotes, the remaining eight populations exhibited Ho values lower than He, indicating varying degrees of heterozygote deficiency. This trend was further corroborated by the inbreeding coefficient (FIS), which was positive across all 10 populations, ranging from 0.020 to 0.162 with a mean of 0.075. This positive FIS values suggest a trend toward relative excess of homozygotes in most populations, implying potential inbreeding or a Wahlund effect (i.e., the presence of substructure within populations).

Nucleotide diversity (π), a core indicator reflecting the richness of genetic variation within a population, is closely associated with effective population size and selective pressures. In this study, π values varied significantly among populations, ranging from 0.026 to 0.160, with an overall mean of 0.102 (Table 3). Notably, populations HZ, PG, XD, and MB, located in the southwestern region, exhibited relatively higher π values (0.129–0.160), suggesting that these populations may have retained richer genetic variation and could represent a genetic diversity center for *G. rhodantha*. In stark contrast, populations QL, ZY, HY, and XY displayed markedly lower π values (0.026–0.070). Combined with field habitat surveys, these low-diversity populations are generally distributed across small, fragmented shrub-grassland areas. Limited habitat area may lead to smaller effective population sizes, thereby exacerbating genetic drift and inbreeding risks, ultimately constraining the long-term preservation and accumulation of genetic variation.

Additional auxiliary genetic parameters further support the conclusion of limited intra-population genetic variation in *G. rhodantha*. Polymorphism information content (PIC) analysis revealed that PIC values across all populations ranged from 0.020 to 0.123, with a mean of 0.078. All values fell below the 0.25 threshold, classifying them as low polymorphism levels according to common evaluation standards [31]. Furthermore, minor allele frequency (MAF) and Shannon's information index (I) were also at low levels, with mean values of 0.072 and 0.147, respectively. These multiple genetic parameters mutually corroborate each other, collectively revealing the current state of impoverished genetic variation within *G. rhodantha* populations and providing an important theoretical basis for formulating germplasm conservation strategies for this species.

Population differentiation analysis

The FST serves as a critical metric for quantifying genetic divergence among populations, with its value directly reflecting the intensity of geographical isolation and the extent of restricted gene flow. According to established criteria, $FST < 0.05$ indicates negligible differentiation, $0.05 \leq FST \leq 0.15$ represents moderate differentiation, $0.15 < FST \leq 0.25$ denotes high differentiation, and $FST > 0.25$ signifies extremely high genetic differentiation, implying pronounced reproductive isolation or geographical barriers [32]. Analysis of genetic differentiation among the 10 geographical populations of *G. rhodantha* in this study (Table 4, Fig. 8) revealed pairwise FST values ranging from 0.1392 to 0.7864, with a mean of 0.3789. The vast majority of pairwise FST values substantially exceeded 0.25, indicating that high levels of genetic differentiation have developed among populations, characterized by significant genetic structure divergence and geographical isolation patterns, with severely restricted inter-population gene flow.

Further analysis revealed considerable heterogeneity in the degree of genetic differentiation among different population pairs. The lowest FST value was observed between the XD and HS populations (0.1392), followed by the PG and HZ pair (0.1436). Although these values fall within the moderate differentiation range, they nonetheless suggest that these populations may share relatively recent common ancestors or have experienced more frequent historical gene flow compared to other population pairs, thereby slowing the progression of genetic differentiation. In stark contrast, the FST value between the ZY and QL populations reached 0.7863, the highest among all population pairs. These two populations are geographically distant, and this pronounced spatial isolation severely restricts pollen and seed dispersal, leading to near-complete cessation of gene flow. Under such circumstances, genetic drift becomes the predominant evolutionary force, accumulating extremely high levels of genetic differentiation [33]. These results further corroborate that geographical isolation is a key driver of population genetic differentiation in *G. rhodantha*.

Table 3
Genetic diversity statistics of *Gentiana rhodantha* across 10 populations

Pop ID	Number	Ho	He	π	Fis	PIC	MAF	Shannon
HS	15.000	0.081	0.121	0.127	0.120	0.097	0.089	0.183
HY	15.000	0.053	0.067	0.070	0.047	0.054	0.051	0.102
HZ	15.000	0.109	0.153	0.160	0.137	0.123	0.112	0.231
MB	14.000	0.090	0.123	0.129	0.104	0.098	0.090	0.184
PG	15.000	0.095	0.147	0.155	0.162	0.119	0.106	0.225
QL	15.000	0.036	0.025	0.026	0.020	0.020	0.020	0.038
RH	15.000	0.094	0.111	0.116	0.064	0.090	0.079	0.170
XD	15.000	0.087	0.124	0.130	0.113	0.100	0.091	0.189
XY	15.000	0.063	0.064	0.067	0.016	0.051	0.048	0.096
ZY	15.000	0.039	0.035	0.036	0.002	0.027	0.027	0.052

Table 4
Pairwise FST values between different populations of *Gentiana rhodantha*

POP2	HS	HY	HZ	MB	PG	QL	RH	XD	XY	ZY
HS	0	0.3485	0.1822	0.3204	0.1983	0.5092	0.3430	0.1392	0.4404	0.5009
HY	0.3485	0	0.3140	0.4300	0.2936	0.6730	0.4503	0.3549	0.5764	0.6542
HZ	0.1822	0.3140	0	0.2533	0.1436	0.4160	0.2734	0.1808	0.3729	0.4239
MB	0.3204	0.4300	0.2533	0	0.2291	0.1950	0.1672	0.3231	0.4737	0.5327
PG	0.1983	0.2936	0.1436	0.2291	0	0.3930	0.2541	0.1968	0.3174	0.3672
QL	0.5092	0.6730	0.4160	0.1950	0.3930	0	0.2330	0.5036	0.7009	0.7863
RH	0.3430	0.4503	0.2734	0.1672	0.2541	0.2330	0	0.3425	0.4841	0.5381
XD	0.1392	0.3549	0.1808	0.3231	0.1968	0.5036	0.3425	0	0.4377	0.5008
XY	0.4404	0.5764	0.3729	0.4737	0.3174	0.7009	0.4841	0.4377	0	0.2807
ZY	0.5009	0.6542	0.4239	0.5327	0.3672	0.7863	0.5381	0.5008	0.2807	0

Analysis of bioactive compound contents in different G. rhodantha populations

To further investigate the variation in the accumulation of medicinal bioactive constituents among different geographical populations of *G. rhodantha*, the contents of six major active compounds—loganic acid, neomangiferin, swertiamarin, sweroside, mangiferin, and isoorientin—were quantified in eight populations using HPLC. The results are presented in Table 5 and Fig. 9.

Mangiferin, the predominant bioactive constituent of *G. rhodantha* and the indicator component specified by the 2025 edition of the Chinese Pharmacopoeia, exhibited the most pronounced variation among populations. The XD population displayed the highest mangiferin content at 122.8439 mg/g, significantly exceeding all other

populations. This was followed by the XY population (78.3677 mg/g), PG (69.0507 mg/g), HS (59.4093 mg/g), HY (55.4126 mg/g), ZY (47.2276 mg/g), and RH (46.8388 mg/g), with the HZ population exhibiting the lowest content at 37.8322 mg/g. Across the eight populations, mangiferin content ranged from 37.8322 to 122.8439 mg/g, with a mean value of 64.6229 mg/g and a 3.25-fold variation, indicating substantial chemotypic differentiation among different geographical populations of *G. rhodantha*.

The remaining five compounds also demonstrated considerable inter-population variation. Swertiamarin content was highest in the ZY population (1.7057 mg/g) and lowest in the PG population (0.0278 mg/g), representing a 61-fold difference, with a mean value of 0.5262 mg/g. Neomangiferin content peaked in the XD population (1.1604 mg/g) and reached its minimum in the HY population (0.2522 mg/g), averaging 0.7717 mg/g. Isoorientin content was highest in the XY population (1.4998 mg/g) and lowest in the RH population (0.4008 mg/g), with a mean of 0.8704 mg/g. Loganic acid content ranged from a high of 1.0464 mg/g in the HY population to a low of 0.1251 mg/g in the HS population, averaging 0.3583 mg/g. Sweroside content varied from 0.4645 mg/g in the ZY population to 0.1068 mg/g in the HZ population, with a mean of 0.2259 mg/g. This pronounced chemotypic differentiation indicates that the accumulation patterns of secondary metabolites in *G. rhodantha* are shaped by the combined influence of genetic background and habitat conditions. The substantial differences in secondary metabolite accumulation patterns among different geographical populations establish a critical foundation for subsequent investigations into the mechanisms associating genetic background with bioactive constituent accumulation.

Table 5
Determination results of six compound

Sample	loganic acid	neomangiferin	swertiamarin	sweroside	mangiferin	isoorientin
HY	1.0464	0.2522	0.1534	0.4251	55.4126	0.6740
XD	0.1997	1.1604	0.4173	0.2458	122.8439	0.9994
PG	0.6275	0.9254	0.0278	0.1673	69.0507	0.9399
RH	0.2195	0.6038	0.5254	0.1192	46.8388	0.4008
XY	0.2920	0.9320	1.2016	0.1588	78.3677	1.4998
HS	0.1251	1.0912	0.1374	0.1199	59.4093	1.0735
HZ	0.1784	0.3942	0.0408	0.1068	37.8322	0.6766
ZY	0.1780	0.8146	1.7057	0.4645	47.2276	0.6995

Integrative analysis of genetic differentiation and bioactive compound accumulation

Correspondence between genetic clustering and chemotypes

Integrative comparison of bioactive compound contents with population genetic analysis results revealed a clear correspondence between genetic clustering and chemotypes. Based on principal component analysis and phylogenetic reconstruction, the 10 *G. rhodantha* populations were partitioned into three major genetic lineages. The first genetic lineage comprised populations HS, XD, HZ, and HY. Within this lineage, populations XD and HS

exhibited close genetic clustering, and both demonstrated relatively high mangiferin contents of 122.8439 mg/g and 59.4093 mg/g, respectively, significantly exceeding those of most other populations. Furthermore, both populations displayed neomangiferin contents exceeding 1.0 mg/g (XD: 1.1604 mg/g; HS: 1.0912 mg/g), reflecting similar chemotypic profiles. The second genetic lineage consisted of populations XY, ZY, and PG. Within this lineage, populations XY and ZY clustered closely in principal component analysis, and both exhibited relatively high swertiamarin contents (XY: 1.2016 mg/g; ZY: 1.7057 mg/g), significantly surpassing those of other populations. The third genetic lineage encompassed populations QL, MB, and RH. Within this lineage, population RH showed evidence of weak gene flow with MB and QL, and its mangiferin content (46.8388 mg/g) fell within the moderate-to-low range. These findings demonstrate a strong correspondence between genetic relatedness and chemotypic similarity, indicating that the accumulation of secondary metabolites is significantly influenced by genetic background. This further corroborates the regulatory role of genetic background in shaping secondary metabolite accumulation patterns.

Influence of genetic diversity levels on bioactive compound accumulation

Genetic diversity analysis revealed that populations HZ, PG, and XD exhibited the highest π values among all populations, with values of 0.160, 0.155, and 0.130, respectively. In contrast, population ZY displayed a substantially lower π value of 0.036. Correspondingly, bioactive compound analysis demonstrated that population XD possessed a remarkably high mangiferin content of 122.8439 mg/g, and population PG exhibited a relatively high mangiferin content of 69.0507 mg/g, both ranking among the higher levels. Conversely, population ZY showed a mangiferin content of only 47.2276 mg/g. These findings are mutually corroborated by the linkage disequilibrium (LD) analysis results: populations XD and PG exhibited rapid LD decay, with LD declining to half its maximum value within 12.640 kb and 12.829 kb, respectively, indicating larger effective population sizes and active recombination that facilitate the retention of genetic variation and adaptive evolution of genes involved in secondary metabolism. In contrast, population ZY displayed slow LD decay, suggesting a smaller effective population size and stronger genetic drift effects that constrain both genetic diversity and the accumulation of bioactive constituents. These results collectively indicate a positive correlation between genetic diversity levels and bioactive compound contents.

Regulatory role of genetic differentiation in chemotypic divergence

To investigate the quantitative relationship between the degree of genetic differentiation among *G. rhodantha* populations and the accumulation patterns of their medicinal bioactive constituents, the contents of six major active compounds across eight geographical populations were analyzed in conjunction with pairwise FST values. The results revealed a distinct and consistent association between the extent of inter-population genetic differentiation and disparities in bioactive compound profiles (Table 5). Highly differentiated population pairs generally exhibited pronounced chemotypic divergence. For instance, the ZY and HY populations, characterized by a high FST value of 0.6542, displayed significant differences in the contents of loganic acid, neomangiferin, and swertiamarin. Similarly, the XY and HY populations, with an FST of 0.5764, showed substantial variation in the levels of loganic acid, swertiamarin, sweroside, and isoorientin. In contrast, population pairs characterized by low genetic differentiation tended to exhibit high similarity across multiple bioactive constituents. The XD and HS populations, which showed the lowest FST value (0.1392), demonstrated highly consistent contents of

neomangiferin, isoorientin, swertiamarin, and sweroside. Likewise, the PG and HZ populations (FST = 0.1436) exhibited comparable levels of swertiamarin, sweroside, and isoorientin. In summary, the degree of genetic differentiation among *G. rhodantha* populations corresponds closely with the divergence in their medicinal bioactive constituent profiles: genetically proximate populations tend to share similar chemical fingerprints, whereas significantly differentiated populations tend to develop distinct chemotypes. This finding reveals the underlying regulatory influence of genetic variation on the accumulation of bioactive constituents and provides an important foundation for predicting chemotypic characteristics and facilitating targeted selection of elite germplasm sources.

Correlation analysis between geographical factors and active ingredient contents

To investigate the influence of geographical factors on the accumulation of bioactive components in *G. rhodantha*, Pearson's correlation coefficient matrix was employed to analyze the relationships between sampling site geographical factors (longitude, latitude, and altitude) and the contents of six major active ingredients (Fig. 10). The results revealed significant spatial associations among geographical factors, with a strong positive correlation observed between longitude and latitude, indicating a certain spatial regularity in the distribution of sampling sites. Altitude exhibited moderate negative correlations with both longitude and latitude, suggesting that high-altitude sampling sites within the study area tended to be located in regions with lower longitude and latitude.

Correlation analysis between geographical factors and active ingredients showed that most compound contents exhibited negative correlations with longitude, latitude, and altitude; however, the influence of geographical factors on the accumulation of active ingredients displayed clear component specificity. Specifically, neomangiferin and isoorientin showed moderate negative correlations with longitude and latitude but strong positive correlations with altitude. In contrast, loganic acid presented moderate positive correlations with longitude and latitude and a weak negative correlation with altitude. Swertiamarin exhibited a strong negative correlation with longitude, a significant negative correlation with latitude, and a weak correlation with altitude. Mangiferin showed weak negative correlations with longitude and latitude but a strong positive correlation with altitude. By comparison, sweroside displayed weak negative correlations with latitude and altitude, and no correlation with longitude.

Among the active ingredients, positive correlations predominated. Neomangiferin exhibited strong positive correlations with mangiferin and isoorientin, and a weak positive correlation with swertiamarin. Swertiamarin showed a moderate positive correlation with sweroside and a weak positive correlation with isoorientin. Mangiferin was moderately positively correlated with isoorientin, while sweroside demonstrated a moderate positive correlation with loganic acid. These patterns suggest that these compounds may share co-accumulation characteristics or be regulated by similar metabolic pathways within the plant. Meanwhile, certain negative correlations were also observed among components, with the strongest negative correlation occurring between loganic acid and neomangiferin, followed by loganic acid with swertiamarin, whereas loganic acid showed only a weak negative correlation with mangiferin and isoorientin. Sweroside also displayed weak negative correlations with neomangiferin and isoorientin. These positive and negative correlation patterns reflect potential synergistic and competitive relationships among different active ingredients in biosynthetic pathways,

providing important clues for elucidating the accumulation mechanisms of secondary metabolites in *G. rhodantha*.

In summary, geographical factors exert a significant negative influence on the accumulation of multiple active ingredients in *G. rhodantha*, while the positive and negative correlations among active ingredients reflect their co-accumulation characteristics and underlying metabolic regulatory relationships. These findings provide a reference for further elucidating the mechanisms by which environmental factors influence the quality formation of *G. rhodantha* and are consistent with the conclusions regarding chemotype differentiation among different populations drawn from the population genetic analysis in this study.

Discussion

Population genetic structure of G. rhodantha reveals pronounced differentiation driven by geographical isolation

This study reveals a distinctive low intra-population and high inter-population genetic structure pattern in the Chinese endemic species *G. rhodantha*, characterized by extremely low intra-population genetic diversity coupled with remarkably high inter-population genetic differentiation. Consistent analyses based on SNP markers demonstrated that the mean He, π, and PIC across all populations were only 0.097, 0.102, and 0.078, respectively. These values are substantially below the established threshold for low polymorphism (PIC < 0.25), confirming the paucity of genetic variation within populations [31]. Concurrently, the mean FST among populations reached 0.3789, with the vast majority of population pairs classified as exhibiting extremely high genetic differentiation according to established criteria [32]. Phylogenetic reconstruction and principal component analysis further partitioned the 10 populations into three primary genetic lineages, with these complementary approaches collectively revealing a pronounced and well-defined geographical genetic structure for this species.

The formation of this pattern arises from the combined effects of the species' biological characteristics and historical evolutionary processes. First, as a typical selfing or inbreeding species, the reproductive system of *G. rhodantha* inherently tends to maintain low intra-population genetic diversity, with the predominantly positive inbreeding coefficients (FIS ranging from 0.020 to 0.162) providing direct evidence for this. Second, and more critically, climatic fluctuations since the Quaternary period, particularly in the topographically complex southwestern region, have driven population contractions and habitat fragmentation for this species [34]. Field surveys revealed that populations exhibiting low genetic diversity (e.g., QL, ZY) are predominantly distributed in small, isolated shrub-grassland areas, where limited habitat extent severely constrains effective population sizes (Ne). Under these conditions, genetic drift rapidly becomes the predominant evolutionary force in small, isolated populations, not only accelerating the stochastic loss of genetic variation within populations but also driving profound differentiation among populations through the random fixation and loss of alleles [33]. The exceptionally high FST value of 0.7863 between the ZY and QL populations exemplifies the extreme outcome of prolonged spatial isolation coupled with genetic drift. This finding carries paramount implications for conservation strategies: conservation efforts should not focus solely on enhancing genetic diversity within individual populations but should prioritize the preservation of highly differentiated populations representing distinct evolutionary units and harboring unique allelic combinations, thereby maximizing the conservation of the species' evolutionary potential and overall genetic variation.

Population demographic history inferred from linkage disequilibrium decay

LD decay analysis provides critical temporal evidence for elucidating the evolutionary demographic history of *G. rhodantha* populations. In this study, substantial variation in LD decay rates was observed among the 10 populations: populations PG, HS, and XD exhibited the most rapid decay, with half-decay distances ranging from 12.640 to 13.183 kb, whereas populations QL and XY displayed markedly slower decay rates. LD decay rates are intimately associated with historical effective population size, recombination rates, and selective pressures: larger effective population sizes and more frequent historical recombination events promote rapid LD decay, while population bottlenecks, inbreeding, or strong selection result in delayed LD decay [35]. Based on these principles, populations such as PG, HS, and XD are inferred to have maintained relatively large historical effective population sizes and experienced frequent recombination events, thereby retaining richer genetic diversity. This inference aligns closely with the π analysis results, where populations HZ, PG, and XD exhibited the highest π values (0.130–0.160). Collectively, these findings suggest that the southwestern region, centered around populations HZ, PG, and XD, may represent a genetic diversity center or glacial refugium for *G. rhodantha*. During the Quaternary glaciations, the complex topography and diverse microhabitats of this region likely provided relatively stable environments, enabling the preservation of a comparatively intact reservoir of genetic variation [36]. Conversely, the slow LD decay observed in populations QL and XY, combined with their extremely low π values, strongly suggests that these populations have experienced severe population bottlenecks or have been maintained at small population sizes for extended periods, resulting in significant genetic diversity depletion. This finding carries direct implications for germplasm collection strategies: populations exhibiting high genetic diversity, such as PG and XD, should be prioritized for germplasm conservation and collection to maximize the capture of the species' genetic variation.

Positive association between genetic diversity and bioactive compound accumulation and their co-differentiation pattern

Integrating population genetic data with secondary metabolite content analysis represents a central breakthrough of this study. The results reveal, for the first time in *G. rhodantha*, a positive association trend between genetic diversity levels and the contents of major medicinal bioactive constituents. Populations exhibiting high genetic diversity, such as XD, demonstrated significantly higher mangiferin contents compared to low-diversity populations like ZY. Furthermore, the rapid LD decay observed in populations XD and PG facilitates the retention of genetic variation, potentially providing richer raw material for the adaptive evolution of genes involved in secondary metabolism [37]. High genetic diversity implies a greater abundance of allelic variation in genes comprising secondary metabolic pathways, increasing the likelihood of efficient bioactive compound accumulation through gene dosage effects or optimized regulatory networks. Conversely, genetically depauperate populations, such as ZY, characterized by small effective population sizes and pronounced genetic drift, may have lost advantageous alleles favorable for secondary metabolite biosynthesis. This observed positive trend suggests that future efforts to screen for high-content medicinal germplasm should prioritize populations exhibiting high genetic diversity and rapid LD decay.

Particularly noteworthy is the pronounced co-differentiation pattern observed between the degree of genetic differentiation among *G. rhodantha* populations and their chemotypic divergence. The three major genetic

lineages delineated by SNP analysis were significantly associated with distinct chemotypic profiles. Within the first lineage (comprising HS, XD, HZ, and HY), populations XD and HS not only exhibited close genetic clustering but also demonstrated highly consistent contents of mangiferin and neomangiferin. Within the second lineage (comprising XY, ZY, and PG), populations XY and ZY jointly exhibited relatively high levels of swertiamarin accumulation. This coupling of genetic relatedness with chemotypic similarity strongly suggests that the key genes governing the biosynthesis of these bioactive constituents, or their regulatory regions, harbor genetic variation paralleling that of neutral markers, a phenomenon indicative of co-evolution or genetic linkage [38]. Quantitative analysis further substantiated this observation: weakly differentiated population pairs, such as XD and HS (FST = 0.1392), exhibited comparable contents across multiple bioactive constituents, whereas highly differentiated pairs, such as ZY and HY (FST = 0.6542), displayed significant disparities in several compounds. This compellingly demonstrates that as gene flow among populations becomes interrupted and genetic differentiation accumulates, the genetic machinery governing secondary metabolism likewise diverges, ultimately giving rise to heritable, stable chemotypes. Mangiferin, serving as the indicator component, exhibited a 3.25-fold variation in content across the eight populations (ranging from 37.8322 to 122.8439 mg/g), underscoring substantial chemotypic differentiation within this species. This variation may originate from genetic polymorphisms in key genes of the xanthone biosynthetic pathway or from differential regulation by upstream transcription factors. Studies on the congeneric species *Gentiana macrophylla* provide corroborating evidence for this inference [39].

Conclusions

In summary, this study draws the following principal conclusions: (1) *G. rhodantha* populations in southwestern China exhibit a pronounced geographical genetic structure, characterized by extremely high levels of inter-population genetic differentiation (mean FST = 0.3789), with geographical isolation serving as the dominant factor driving this differentiation; (2) Intra-population genetic diversity is generally low (mean π = 0.102), with habitat fragmentation exacerbating genetic drift effects and inbreeding risks; (3) LD decay analysis reveals differential demographic histories among populations, suggesting that populations such as PG and XD may represent glacial refugia that have retained relatively high genetic diversity; (4) A positive trend exists between genetic diversity and the accumulation of bioactive constituents, and a pronounced co-differentiation pattern is observed between genetic differentiation and chemotypic divergence, providing direct evidence for the genetic basis underlying chemotypic differentiation. (5) Constituent-specific associations were observed between geographical factors and the accumulation of active ingredients. Neomangiferin, isoorientin, and mangiferin exhibited strong positive correlations with altitude, whereas loganic acid showed moderate positive correlations with longitude and latitude. Positive correlations predominated among the active ingredients, reflecting their co-accumulation characteristics and potential metabolic regulatory relationships.

Synthesizing these findings, highly differentiated populations (e.g., ZY and QL populations with FST = 0.7863) represent distinct evolutionary units and should be prioritized as separate conservation targets to prevent the loss of adaptive gene combinations through admixture between different genetic lineages. Populations exhibiting both high genetic diversity and exceptional bioactive compound contents (e.g., XD and PG populations) are suitable candidates for targeted selection as elite germplasm for propagation and utilization. For low-diversity populations experiencing severe habitat fragmentation (e.g., QL, XY, and ZY), priority should be given to habitat restoration and population reinforcement efforts, implementing assisted moderate gene flow to mitigate inbreeding depression risks and enhance the long-term survival potential of these populations.

Declarations

Acknowledgements

We are grateful to editors and anonymous reviewers for their helpful comments.

Authors' contributions

Qingqing Ye: Data curation, Software, Methodology, Formal analysis, Writing-review and editing. Lingling Wei: Investigation, Resources, Software, Data curation, Formal analysis, Writing-original draft. Weixiang Xiao: Investigation, Resources, Data curation. Xin Tan: Investigation, Resources, Data curation. Li Yang: Investigation, Resources, Supervision. Deqiang Ren: Software, Data curation, Supervision. Ning Ding: Software, Data curation,Supervision. Zhikun Wu: Conceptualization, Formal analysis, Investigation, Supervision, Writing-original draft, Writing-review and editing.

Funding

This research was supported by the Research on Green Prevention and Control Technology for Leaf Blight of Gentiana rhodantha under Forest Based on Pathogen-Biological Control-Plant-Environment Coordinated Regulation (Qianlin Kehe [2026] Zhicheng No. 004), the Chuan-Qian Collaboration Premium Seeds and Medicinals: Joint Research and Demonstration of Systematic Introduction, Domestication, and Key Industrialization Technologies of Gentiana rhodantha Germplasm Resources (2026YFHZ0116), the National Wild Plant Germplasm Resource Center for Guizhou University of Traditional Chinese Medicine (ZWGX—2405), and the Science and Technology Plan Project of Guizhou Province ([2022]—4016).

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Marchant 1JAS, Peñuelas R. Environmental change and the option value of genetic diversity. Trends Plant Sci. 2009;14:51–8. https://doi.org/10.1016/j.tplants.2008.10.002.
2. Frankham 2RDH. Correlation between fitness and genetic diversity. Conserv Biol. 2003;17:230–7. https://doi.org/10.1046/j.1523-1739.2003.01236.x.

3. Cui 3ZQ, Levsh MY, Yang O, Liu D, Li J. Two CYP82D enzymes function as flavone hydroxylases in the biosynthesis of root-specific 4'-deoxyflavones in scutellaria baicalensis. Mol Plant. 2018;11:135–48. https://doi.org/10.1016/j.molp.2017.08.009.
4. Jang 4JW, Kim Y, Waminal NH, Kim NE, Lee YC. Genetic diversity among cultivated and wild *Panax ginseng* populations revealed by high-resolution microsatellite markers. J Ginseng Res. 2020;44:637–43. https://doi.org/10.1016/j.jgr.2019.05.008.
5. Duan 5LR, Ran WY, Chen ZF, Yu XL, Fang HX. Diversity and correlation analysis of endophytes and metabolites of *Panax quinquefolius* L. in various tissues. BMC Plant Biol. 2023;23:275. https://doi.org/10.1186/s12870-023-04282-z.
6. Rife 6PJA. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome. 2012;5:92–102. https://doi.org/10.3835/plantgenome2012.05.0005.
7. Zhao 7HJ, Laroche X, Lu A, Liu ZX, Li H. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci. 2014;5:484. https://doi.org/10.3389/fpls.2014.00484.
8. Batley 8SA, Edwards J. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J. 2017;15:149–61. https://doi.org/10.1111/pbi.12645.
9. Glaubitz 9ERJ, Sun JC, Poland Q, Kawamoto JA, Buckler K. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6:e19379. https://doi.org/10.1371/journal.pone.0019379.
10. Hohenlohe 10DJW, Etter PA, Boone PD, Catchen JQ, Blaxter JM. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011;12:499–510. https://doi.org/10.1038/nrg3012.
11. Zhou 11AK, Tong CX, Liu BQ, Shan D, Zhang XH, Bian X. Population genetic differentiation and structure of rare plant *Anemone shikokiana* based on genotyping-by-sequencing (GBS). BMC Plant Biol. 2024;24:995. https://doi.org/10.1186/s12870-024-05705-1.
12. Li 12WYP, Yang HH, Liu ZJ, Liu BY, Zhong YJ. Genotyping-by-sequencing study of the genetic diversity and population structure of the endangered plant *tsoongiodendron odorum* Chun in China. Forests. 2024;15:910. https://doi.org/10.3390/f15060910.
13. Rao 13ZXY, Xia Y, Ma GH, Chen DD, Yao TM. Genetic structure analysis of ardisia violacea based on genotyping by sequencing (GBS) technology. J Plant Resour Environ. 2023;32:11–21. https://doi.org/10.3969/j.issn.1674-7895.2023.06.02.
14. Zhang 14SQ, Liao LD, Wang ZH, Yan SY, Shi TX. The genome of artemisia annua provides insight into the evolution of asteraceae family and artemisinin biosynthesis. Mol Plant. 2018;11:776–88. https://doi.org/10.1016/j.molp.2018.03.015.
15. Good 15AKR, Miller JM, Luikart MR, Hohenlohe G. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 2016;17:81–92. https://doi.org/10.1038/nrg.2015.28.
16. Tang 16CBT, Chen SJ, Xiao R, Zhu N, Tian JS et al. MX,. The Traditional Uses, Phytochemistry, Pharmacology, Toxicology, and Clinical Uses of *Metagentiana Rhodantha* (Franch.) T.N.Ho and S.W.Liu, an Ethnomedicine in Southwest China. Frontiers in Pharmacology. 2021; 12: 658628. https://doi.org/10.3389/fphar.2021.658628
17. Cao 17DXJ. Preliminary study on miao medicine *Gentiana rhodantha* in treating ulcerative colitis. J Med Pharm Chin Minorities. 2025;31:60–5. https://doi.org/10.16041/j.cnki.cn15–1175.2025.09.019.

18. Liu 18LHF, Liu X, Zhao S. The chemical structure, pharmacological activity, and clinical progress of Gentianae Radix et Rhizoma. Front Pharmacol. 2025;16:1656493. https://doi.org/10.3389/fphar.2025.1656493.
19. Pu 19LHF, Zhao WJ, Shu T, Lan XD, Chen JJ. Iridoid glycosides and phenolic glycosides from *Gentiana rhodantha* and their anti-inflammatory activities. Phytochemistry. 2026;241:114660. https://doi.org/10.1016/j.phytochem.2025.114660.
20. Xie 20WS, Qin GY. Advance in pharmaceutical research of *Gentiana rhodantha*. Chin Wild Plant Resour. 2017;36:53–9. https://doi.org/10.3969/j.issn.1006-9690.2017.04.012.
21. Zhang 21PY, Shen J, Zhao T, Wang YL, Li YZ. Comparative metabolic fingerprinting of *Gentiana rhodantha* from different geographical origins using LC-UV-MS/MS and multivariate statistical analysis. BMC Biochem. 2015;16:9. https://doi.org/10.1186/s12858-015-0038-5.
22. Wei 22GJL, Yang SH, Xiang L, Li WX, Zhou XM. Comparative study on the quality of wild and cultivated *Gentiana Rhodantha*. J Chin Med Mater. 2025;48:46–51. https://doi.org/10.13863/j.issn1001-4454.2025.01.008.
23. Xu 23YCK, Xu SJ, Sun WF, Wang QW, Guo B. Study on UPLC fingerprint and content determination of mangiferin of *Gentiana rhodantha*. China Pharm. 2023;34:34–9. https://doi.org/10.6039/j.issn.1001-0408.2023.01.07.
24. Wu 24WLL, Wu YD. Comprehensive evaluation on the quality of *Gentiana rhodantha* from different producing areas. Seed. 2025;44:82–92. https://doi.org/10.16590/j.cnki.1001-4705.2025.03.082.
25. Zhou 25CSF, Chen YQ, Gu YR. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinf (Oxford England). 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.
26. Hohenlohe 26CJ, Bassham PA, Amores S, Cresko A. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22:3124–40. https://doi.org/10.1111/mec.12354.
27. Schmidt 27MBQ, Chernomor HA, Schrempf O, Woodhams D, von Haeseler MD. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
28. Dehal 28PMN, Arkin PS. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5:e9490. https://doi.org/10.1371/journal.pone.0009490.
29. Novembre 29ADH, Lange J. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64. https://doi.org/10.1101/gr.094052.109.
30. Neale 30PS, Todd-Brown B, Thomas K, Ferreira L, Bender MA. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.
31. Poczai 31NS, Cernák P, Gorji I, Hegedűs AM, Taller G. PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. Biochem Genet. 2012;50:670–2. https://doi.org/10.1007/s10528-012-9509-1.
32. Weir 32HKE. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet. 2009;10:639–50. https://doi.org/10.1038/nrg2611.
33. Slatkin M. Gene flow and the geographic structure of natural populations. Science. 1987; 236: 787–92. https://doi.org/10.1126/science.3576198

34. Fu 34QYX, Comes CX. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol Phylogenet Evol. 2011;59:225–44. https://doi.org/10.1016/j.ympev.2011.01.012.
35. Slatkin M. Linkage disequilibrium–understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9:477–85. https://doi.org/10.1038/nrg2361.
36. Hewitt G. The genetic legacy of the Quaternary ice ages. Nature. 2000;405:907–13. https://doi.org/10.1038/35016000.
37. Wink M. Plant secondary metabolism: diversity, function and its evolution. Nat Prod Commun. 2008; 3: 1934578X0800300801. https://doi.org/10.1177/1934578X0800300801
38. Andrew 38MBD, Külheim RL, Foley C. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. New Phytol. 2014;201:733–50. https://doi.org/10.1111/nph.12526.
39. Fu 39YY, Mi H, Yang F, Wang Y, Li Y. Genomic characterization of WRKY transcription factors related to secoiridoid biosynthesis in *Gentiana macrophylla*. BMC Plant Biol. 2024;24:66. https://doi.org/10.1186/s12870-024-04727-z.

Figures

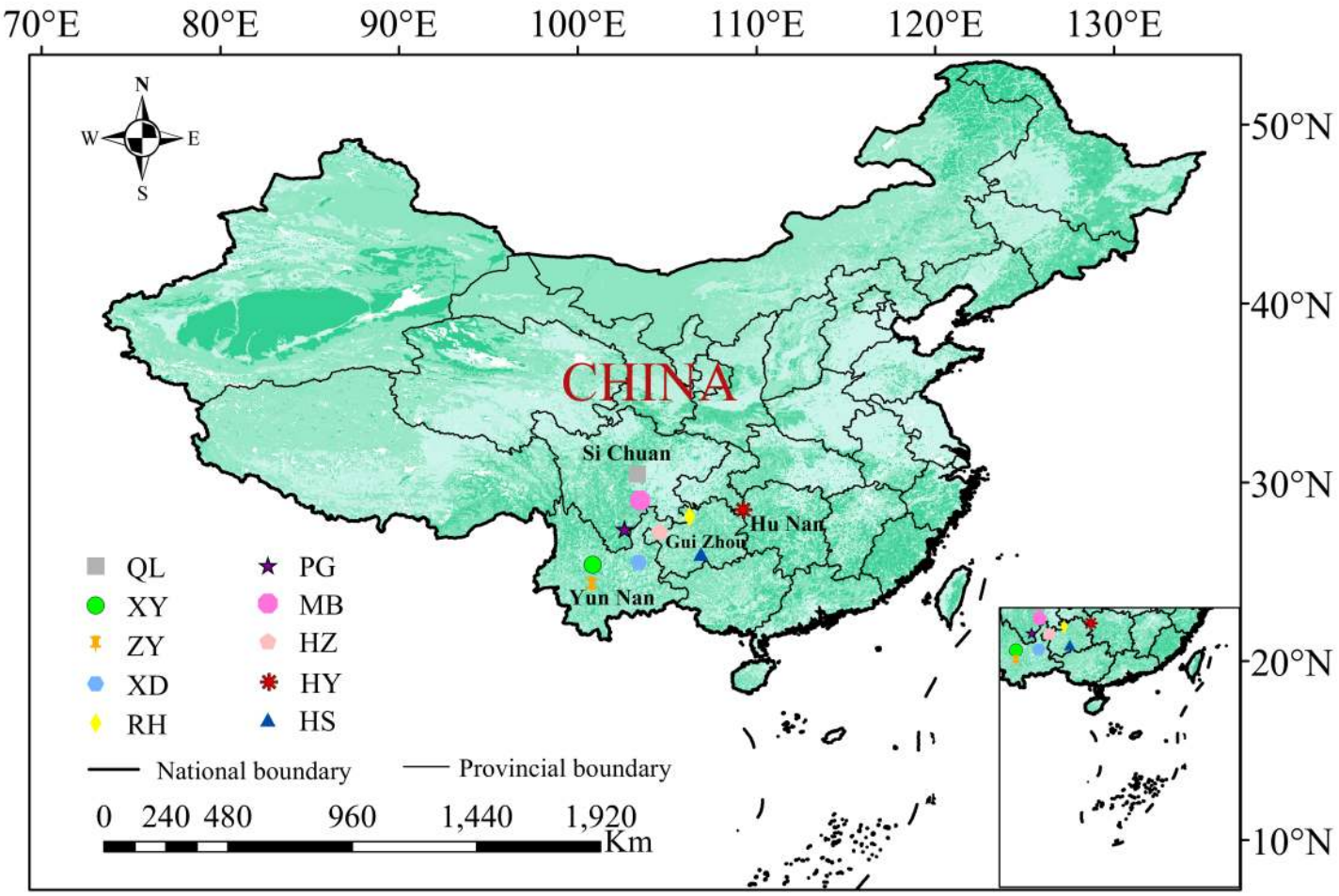


Figure 1

Geographic distribution of the 10 samples of *Gentiana rhodantha*

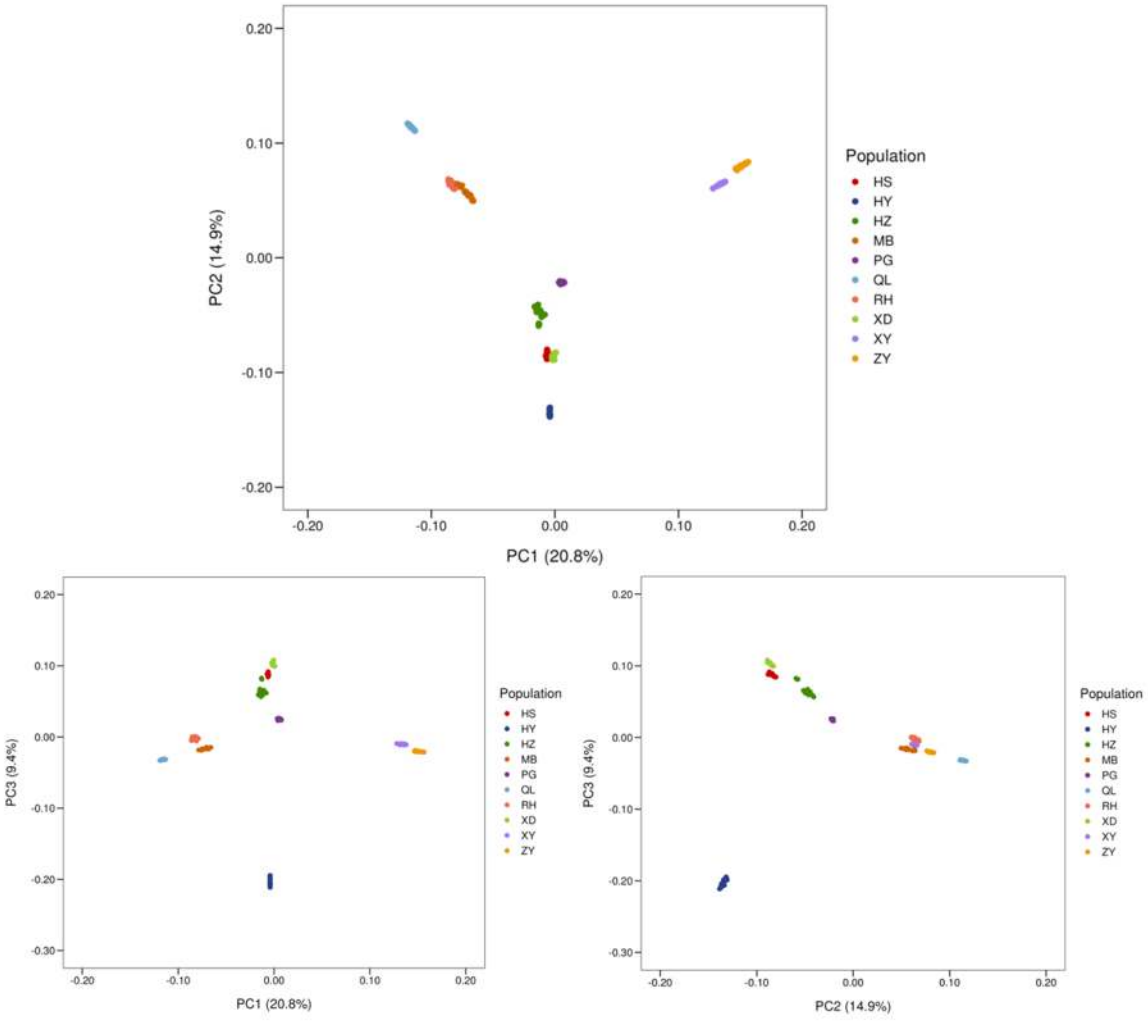


Figure 2

PCA scatter plot of *Gentiana rhodantha*

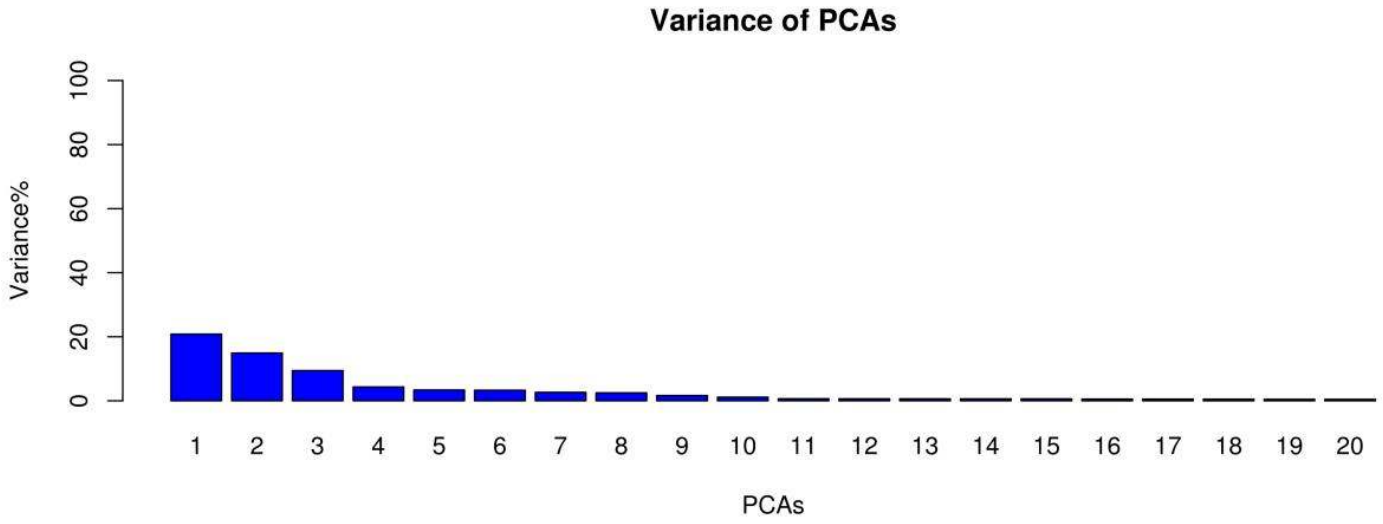


Figure 3

Distribution of explained variance by principal components in PCA

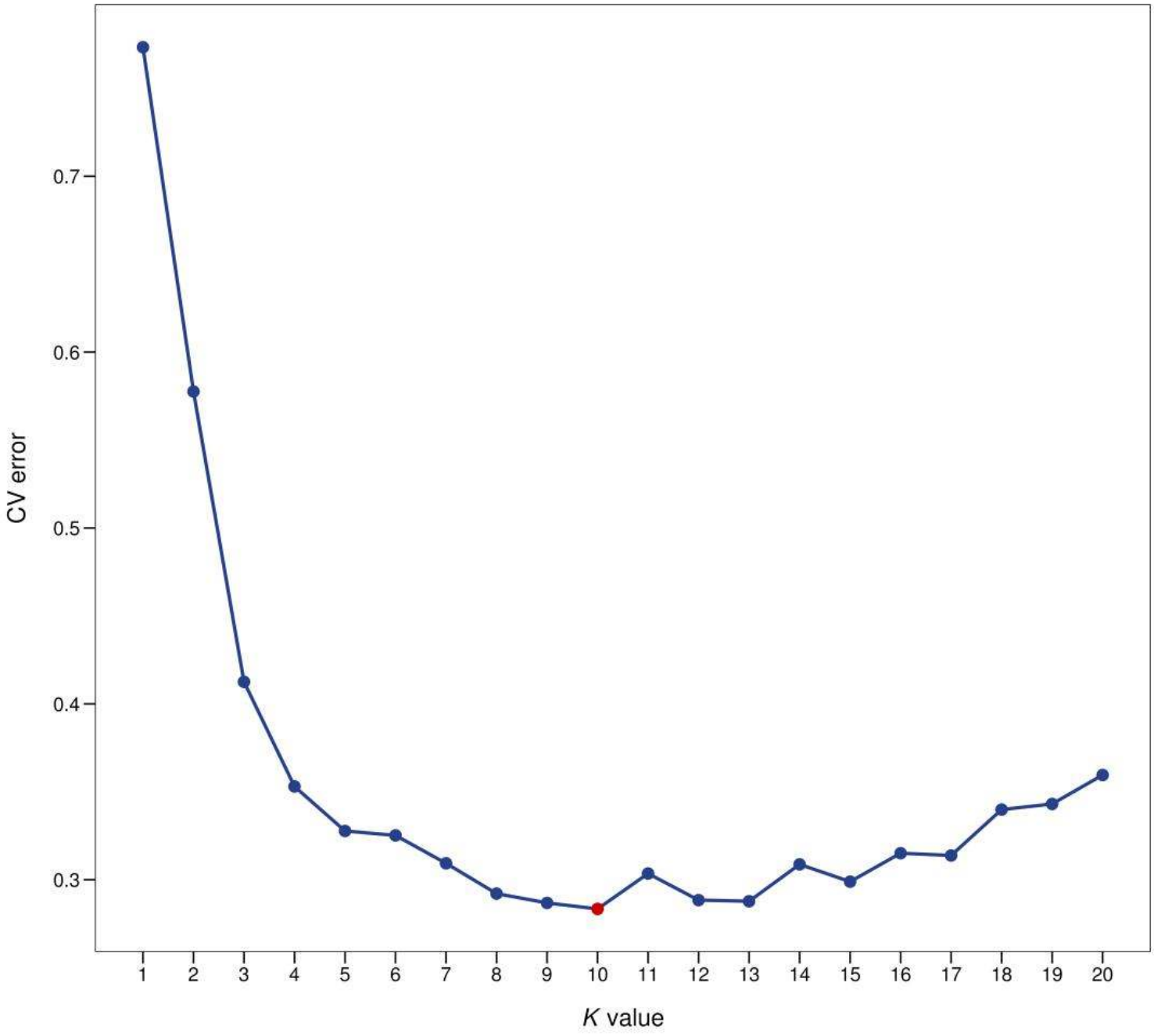


Figure 4

Cross-validation error across different K values in population structure analysis

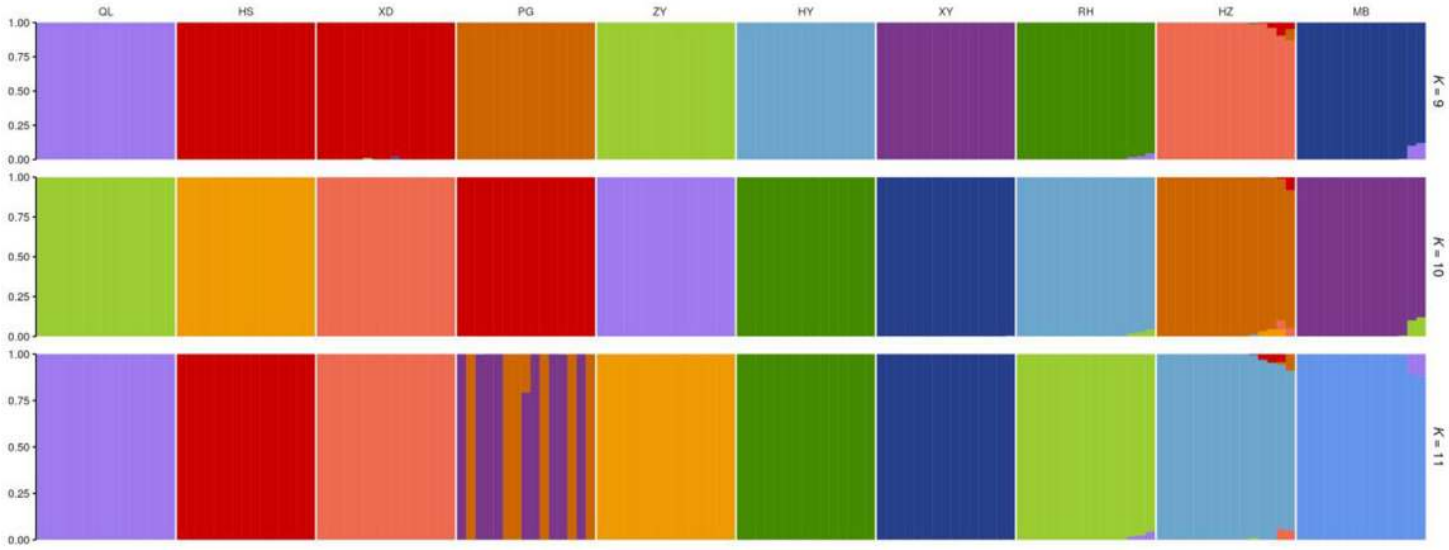


Figure 5

ADMIXTURE analysis of population genetic structure under assumed ancestral numbers K = 9 - 11

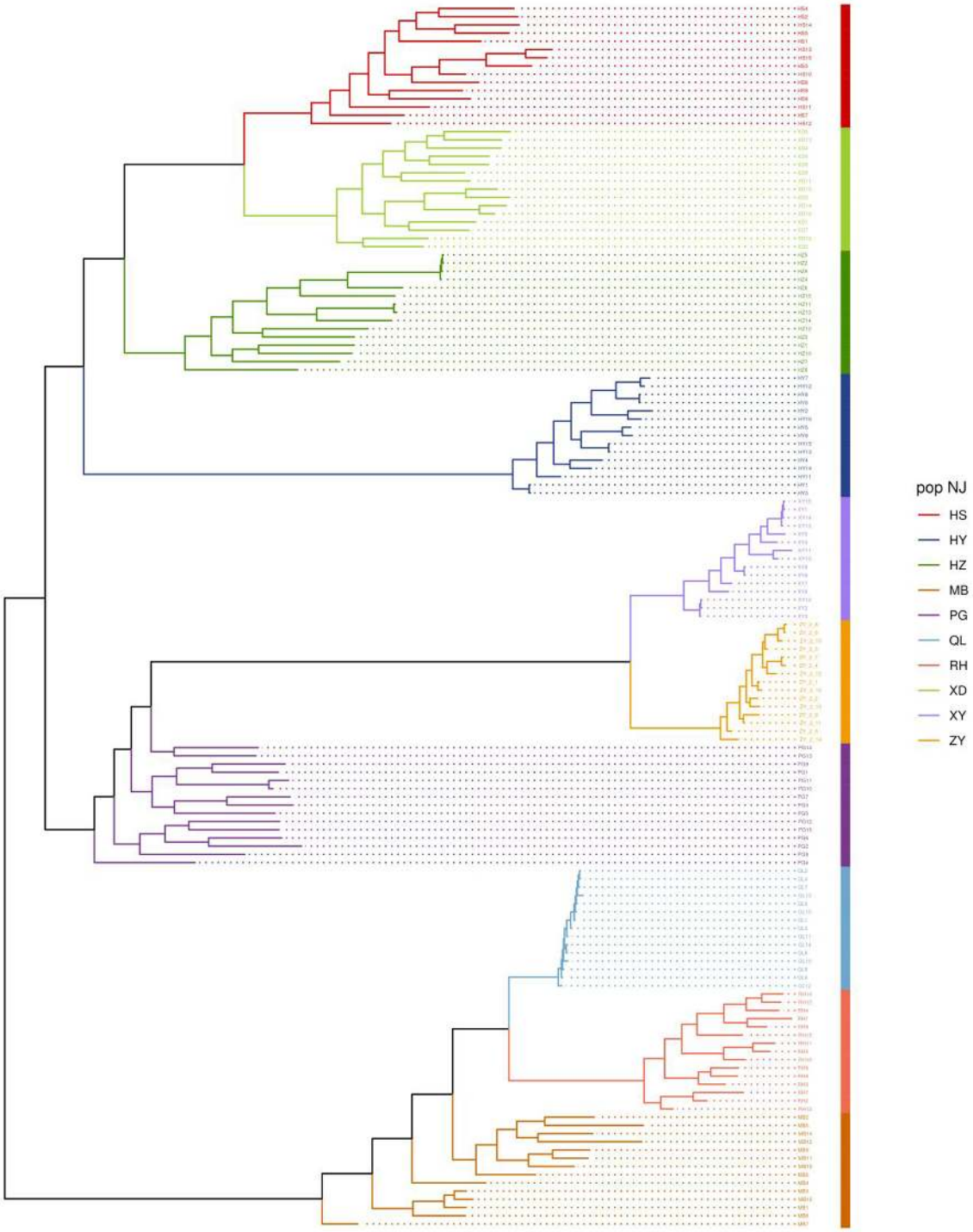


Figure 6

NJ phylogenetic tree of the 149 *Gentiana rhodantha* accessions

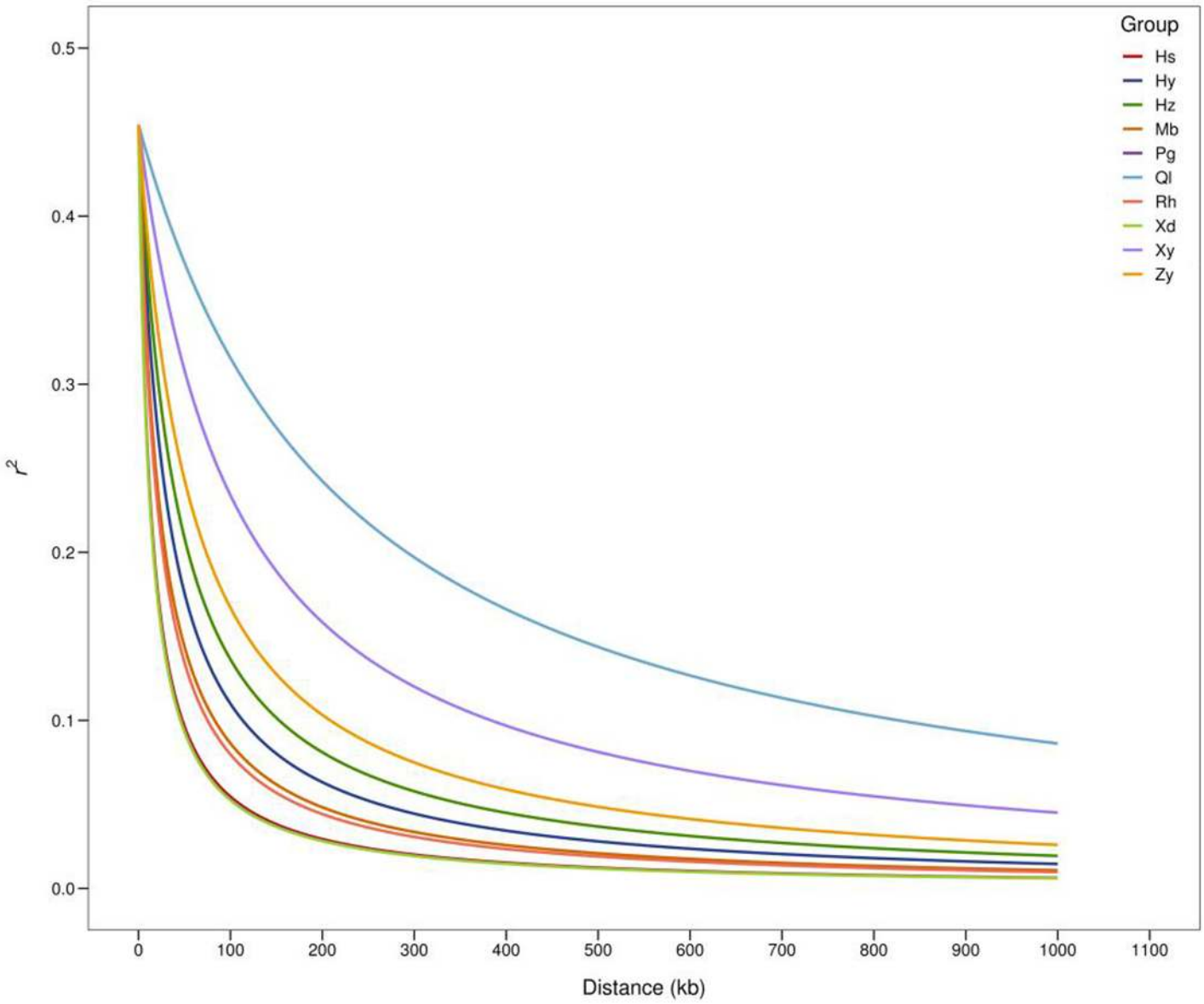


Figure 7

LD analysis of *Gentiana rhodantha*

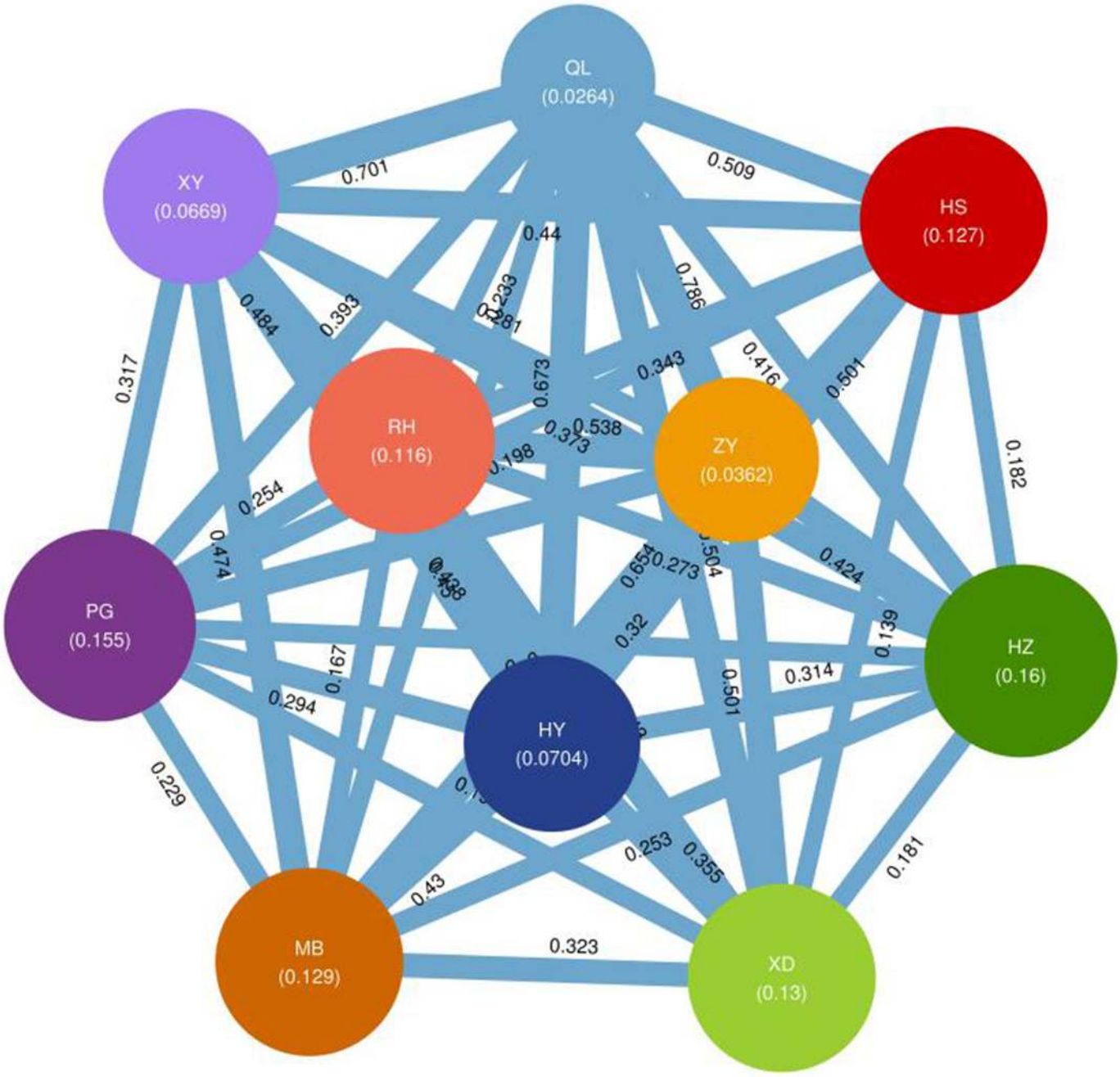


Figure 8

Schematic illustration of nucleotide diversity and population differentiation

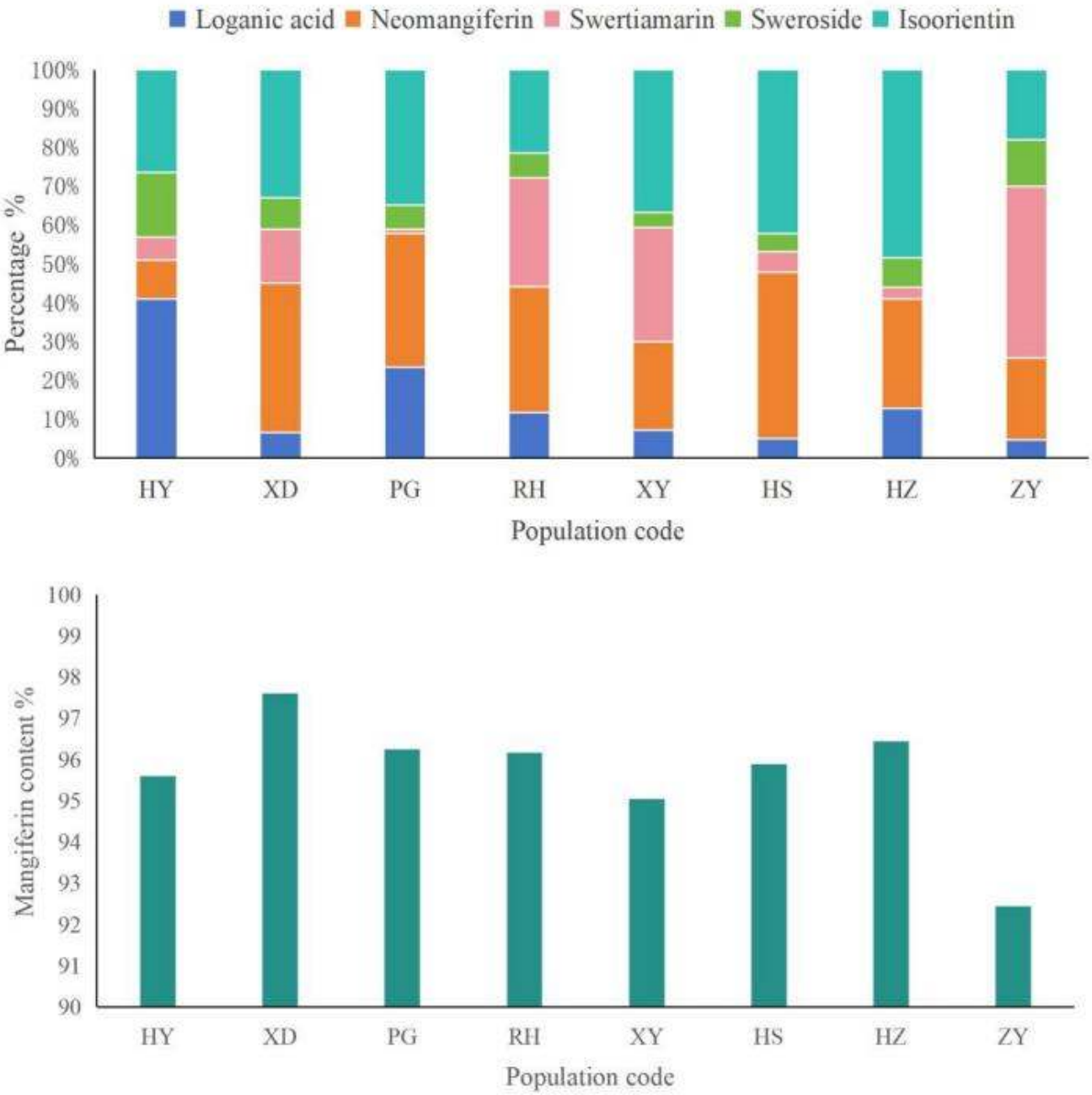


Figure 9

Contents of six compounds in eight medicinal materials

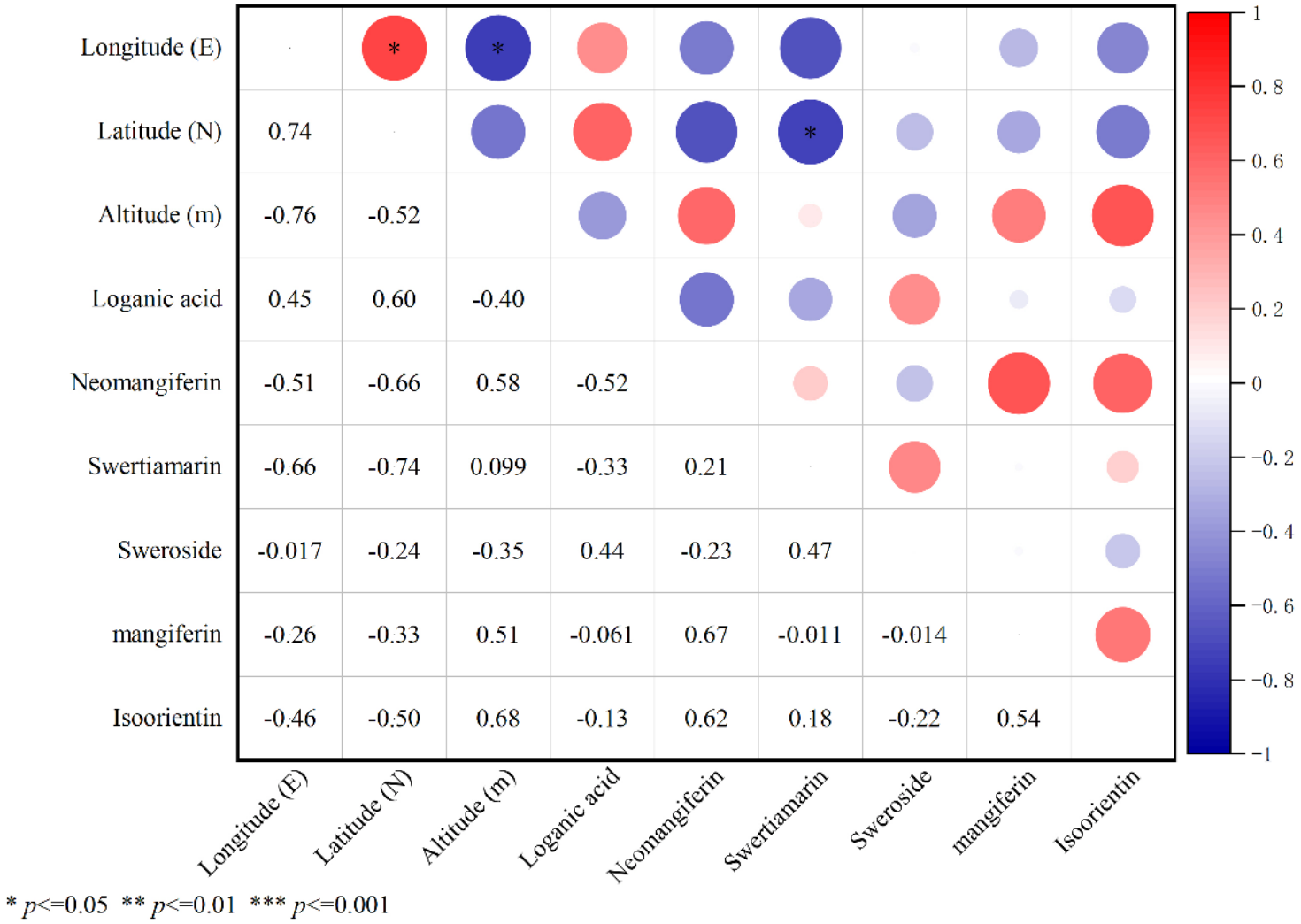


Figure 10

Correlation heatmap of geographical factors and seven chemical constituents.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx
- SupplementaryTable2.docx
- SupplementaryTable3.docx