

Genome-wide assessment of genetic variation and genetic differentiation for *Gastrodia elata* germplasm based on SLAF sequencing

Zhe XU (✉ xuzheT_T@163.com)

Yunnan University of Traditional Chinese Medicine <https://orcid.org/0000-0002-9876-6519>

Yana Shi

Yunnan Academy of Agricultural Sciences

Lei Zhang

Yunnan University of Traditional Chinese Medicine

Huali Qian

Yunnan University of Traditional Chinese Medicine

Xiaolei Chen

Yunnan University of Traditional Chinese Medicine

Jiayun Su

Yunnan University of Traditional Chinese Medicine

Hao Chen

Yunnan University of Traditional Chinese Medicine

Jiahong Dong

Yunnan University of Traditional Chinese Medicine

kun Cong

Yunnan Academy of Agricultural Sciences

Pengzhang Ji

Yunnan University of Traditional Chinese Medicine

Research Article

Keywords: *Gastrodia elata* Bl., Genetic diversity, SLAF-seq SNP

Posted Date: August 12th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1875947/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on March 2nd, 2023. See the published version at <https://doi.org/10.1007/s10722-023-01548-1>.

Abstract

Gastrodia elata Bl. is an important cultivated medicinal plant in China. To analyze the genetic diversity and evolutionary relationship of the germplasm resources of *G. elata*, specific Single nucleotide polymorphism (SNP) markers were developed. SLAF analysis was used to compare 28 samples of the same *G. elata* cultivar. Plants from 4 different varieties or different habitats were collected to explore intraspecific variation and to lay a foundation for resource protection. This will facilitate improved variety breeding in future. In this study, Single nucleotide polymorphism (SNP) genetic variation and differentiation of *G. elata* f. *glauca*, *G. elata* f. *viridis*, and *G. elata* f. *elata* were analyzed using Specific-Locus Amplified Fragment Sequencing (SLAF-seq). A total of 75.95M reads with an average sequencing depth of $19.32 \times$ and a mean Q30 of 91.71% were obtained. Based on the 19,675 polymorphic SLAF tags, 60,238 SNPs were identified and a subset of 22,737 SNPs with minor allele frequency > 0.05 and integrity > 0.5 were selected. A model-based analysis divided the accessions into two groups, wild type *G. elata* f. *glauca* and *G. elata* f. *viridis* groups. Phylogenetic analysis also clustered the samples into the two major groups. *G. elata* has high genetic diversity. Population diversity was highest in *G. elata* f. *elata* and lowest in *G. elata* f. *viridis*. Analysis of molecular variance (AMOVA) revealed significant variations within individuals (92.23%). This study provides new insights into the genetic variation and differentiation of *G. elata*, which can be exploited to improve existing commercial cultivars.

Introduction

G. elata is a perennial Chinese medicinal plant belonging to the *Orchidaceae* family. Dried tubers of the plant species are precious traditional Chinese medicinal materials. These have been used as food and medicine for at least two thousand years. The product is believed to have several therapeutic effects, including alleviating epilepsy convulsion, headache, dizziness, rheumatic arthralgia, and tetanus. Its other beneficial effects include dispelling wind and dredging collaterals, translating liver yang, calming wind, and stopping spasm [Pharmacopoeia of the People's Republic of China. Part I. 2020]. *G. elata* is mainly distributed in Yunnan, Sichuan, Guizhou, northwest and northeast China, Japan, and South Korea [Chen et al. 1999].

In recent years, excessive mining of wild *G. elata* has made it has become increasingly scarce. Therefore, in order to ensure sufficient abundance of *G. elata*, artificial cultivation has become the main source of *G. elata* medicinal materials. In China, the national *G. elata* planting area is stable at about 40000 hectares, with an annual output of about 100,000 tons of fresh *G. elata*. Its output value is about 5 billion yuan, and the export turnover is 5 million dollars and is gradually increasing. About 200 tons of *G. elata* are exported overseas (Zhaotong Tianma specialty bureau). In addition, some *G. elata* products such as *G. elata* wine, *G. elata* powder, *G. elata* tea, etc., have been developed for industrial use [Wang 2016].

There are thirteen known species of *Gastrodia* in China [Xu 1993], namely *Gastrodia elata*, *G. javanica*, *G. gracilis*, *G. angusta*, *G. tuberculata*, *G. flabilabella*, *G. hiemalis*, *G. menghaiensis*, *G. autumnalis*, *G. peichatieniana*, *G. confusa*, *G. appendiculata* and *G. fontinalis*. Only *G. elata* is listed in Pharmacopoeia

of the People's Republic of China. Various *G. elata* varieties are widely cultivated and distributed in China with minor phenotypic differences, such as flower color, tuber shape, and stem color. Currently, there are five *G. elata* varieties that include *G. elata f. elata*, *G. elata f. glauca* S. Chow, *G. elata f. viridis* Mak, *G. elata f. alba* S. Chow, and *G. elata f. flavida* S. Chow [Zhou 2013]. *G. elata* is rich in bioactive components such as gastrodin, p-hydroxybenzaldehyde, p-hydroxybenzyl alcohol, vanillin, parishin, and plant polysaccharides [Yu et al. 2016]. *G. elata f. glauca* contains the highest levels of gastrodin, followed by *G. elata f. viridis*, *G. elata f. elata* and *G. elata f. flavida* [Li et al. 2007]. Traditional breeding mainly depends on the selection of plant phenotype, which is influenced by many factors such as the environment, genetics, and the interactions between these two. *G. elata* is mainly cultivated for its active ingredients, which vary between its different varieties. Therefore, it is necessary to analyze and characterize the germplasm of *G. elata*.

The wild germplasms of *G. elata* are currently on the verge of extinction due to over-exploitation. Moreover, mass cultivation of the few *G. elata* varieties has led to low genetic diversity. Biodiversity conservation is one of the keys to protect species. More specifically, it protects the genetic diversity or evolutionary potential of species. The richer the genetic diversity or variability within a species, the greater the adaptability of the species to environmental changes, and the greater its evolutionary potential. This will help protect the diversity of the species and the entire ecosystem or slow down the extinction process caused by adaptation and evolution [Shi 1990]. Only by mastering the knowledge about the species diversity and the genetic structure of the population can we formulate effective protection strategies and measures. Analysis of DNA fragments among organisms can reveal the molecular variations underlying their phenotypic traits. Molecular markers have been extensively applied to analyze the genetic diversity of germplasm resources [Li et al. 2019; Huang et al. 2018; Yousef et al. 2018; Yang et al. 2019; Arif et al. 2011], construct genetic maps [Chang et al. 2018; Su et al. 2017; Mishra 2014], and map quantitative trait loci (QTL) [Wen et al. 2020; Lv et al. 2021]. Commonly used DNA molecular markers include Restriction Fragment Length Polymorphism (RFLP) [Liang et al. 2019], Single Nucleotide Polymorphism (SNP) [Yang et al. 2020], Random Amplified Polymorphic DNA (RAPD) [Kazemeini et al. 2020], Simple Sequence Repeat (SSR) [He et al. 2020], Amplified Fragment Length Polymorphism (AFLP) [Sorkheh et al. 2019], Sequence-related Amplified Polymorphism (SRAP) [Ma et al. 2021], Inter-simple Sequence Repeat (ISSR) [Luc et al. 2010] and DNA barcode [Lou et al. 2010]. Genetic diversity of *G. elata* has been analyzed using AFLP, RAPD, and ISSR [Wang et al. 2019; Wang et al. 2019; Zou et al. 2006; Zhao et al. 2008] markers. SNP has many advantages, such the ability to locate many loci, wide distribution, genetic stability, representativeness, and exists in the form of biallelic polymorphism [Brookes 1999]. In this study, third-generation molecular marker SNP was used to analyze the genetic diversity of *G. elata f. glauca*, *G. elata f. viridis* and *G. elata f. elata* for the first time. We aimed to examine levels of genetic variation and differentiation in *G. elata* for protection and utilization of *G. elata* germplasm resources. These can be exploited to improve the existing commercial cultivars in Yunnan province, China and beyond.

Materials And Methods

Plant materials

Samples from four *G. elata* populations were collected from Yiliang, Weixi, Yunnan, and Yichang regions in Hubei, China. The Weixi population (Ww) was *G. elata* f. *glauca*; the Yiliang population (W) was *G. elata* f. *glauca*; the other Yiliang population (L) was *G. elata* f. *viridis*, and the Yichang population (H) was *G. elata* f. *elata* (Figs. 1 and 2). The Weixi population (Ww) is a wild-type *G. elata*, whereas the other populations are cultivated varieties (Table 1).

Table 1
Details of the *G. elata* samples used in this study

Code	Variety	Origin	Sample size	Longitude (E)	Latitude (N)	Altitude(m)	Type
Ww	<i>Gastrodia elata</i> f. <i>glauca</i>	Weixi, Diqing, Yunnan	7	27.2713	99.2774	2700	wild
W	<i>G. elata</i> f. <i>glauca</i>	Yiliang, Zhao tong, Yunnan	7	27.7926	104.2656	1716	cultivated
L	<i>G. elata</i> f. <i>viridis</i>	Yiliang, Zhao tong, Yunnan	7	27.7926	104.2656	1716	cultivated
H	<i>G. elata</i> f. <i>elata</i> .	Yiling, Yichang, Hubei	7	30.9599	111.1670	88	cultivated

Genomic DNA extraction and SLAF sequencing

Genomic DNA was extracted from rhizomes of *G. elata* using the CTAB method with minor modification [Murray et al. 1980]. The quality and concentration of the extracted DNA samples were determined using a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific, US). The concentrations of the DNA samples were then standardized to 50 ng/μL.

Subsequently, genomic DNA was digested using HaeIII restriction enzyme (new England laboratory, Nebraska). A poly-adenine tail was added to the digested DNA fragments using Klenow (3'~5' exonuclease) and dATP at 37°C. A double-labeled sequencing adapter (PAGE separation and purification, American Life Technology Company) was then connected to the DNA fragment at the poly A tail by T4 DNA ligase. The digested samples were then diluted and ligated. Subsequently, PCR was performed using dNTP, Q5 high-fidelity DNA polymerase, and PCR primers: AATGATACGGCGACCACCGA and CAAGCAGAAGACGGCATACG (PAGE separation and purification, American life technology company). The PCR products were purified and enriched by agt AM Pure XP magnetic beads and separated by 2% agarose gel electrophoresis. Fragments ranging from 400–450 bp were isolated and purified using rapid

gel extraction kit (QIA quick Gel Extraction, QIAGEN). The gel purified products were sequenced on Illumina HiSeq2500 platform (Illumina company, San Diego, California, USA).

SNP calling and data analysis

SNPs were called using BWA software package [Li et al. 2009]. The data were filtered according to the workflow of "the most effective method" proposed by McKenna et al [McKenna et al. 2010]. The most labeled samples were used as references. Variable sites were detected using GATK [McKenna et al. 2010] and SAMTOOLS [Li et al. 2009]. SNP markers identified by both methods were shortlisted, and the results were considered as the final and reliable SNP marker data set.

Data processing and analysis

The genetic diversity of the population was measured based on the conventional percentage of polymorphic loci (p), Nei's gene diversity index, and Shannon diversity index using POPGENE [Yeh et al. 1999; Yeh et al. 1997; Nei 1987]. The population structure of *G. elata* was constructed using Admixture software. Phylogenetic analysis was performed using MEGA 7 software [Kumar et al. 2016]. AMOVA analysis was performed using AMOVA software [Excoffier et al. 2007] and the genetic variation among and within populations was calculated. PCA analysis was performed using smart software [Patterson et al. 2006].

Results

Identification of SLAF Tag and SNP Marker

SLAF tags were screened in 28 samples of *G. elata*. A total of 471,001 SLAF tags were obtained, of which 19,675 were polymorphic with a sequencing depth of 63,772,545 reads. The sequencing depth was 13.28 ~ 26.5 X for a single sample, with an average sequencing depth of 19.32x (Table 1). Further analysis showed that each sample had 21848 ~ 45240 SNP markers with integrity of 18.34 ~ 54.43% (25.69% average) and heterozygosity rate of 36.26% ~ 75.1% (47.58% average). After considering MAF and integrity, 60,238 SNPs were finally obtained and further utilized to analyze the genetic background of the population (Table 2).

Table 1
The SLAF labels of *G. elata*

Number	Sample number	SLAF tag number	Total depth of sequencing	Average sequencing depth
1	Ww1	123,112	2,166,004	17.59
2	Ww2	258,203	3,429,227	13.28
3	Ww3	125,828	2,432,004	19.33
4	Ww4	121,818	1,683,874	13.82
5	Ww5	119,236	2,780,116	23.32
6	Ww6	120,653	2,576,764	21.36
7	Ww7	124,873	2,607,156	20.88
8	W1	138,733	2,690,240	19.39
9	W2	123,457	2,126,627	17.23
10	W3	113,944	2,545,067	22.34
11	W4	120,016	2,265,255	18.87
12	W5	127,430	2,689,731	21.11
13	W6	130,228	2,702,234	20.75
14	W7	112,327	2,144,808	19.09
15	L1	101,658	1,936,550	19.05
16	L2	109,983	2,728,377	24.81
17	L3	109,128	2,892,365	26.50
18	L4	112,211	2,899,183	25.84
19	L5	108,604	2,554,627	23.52
20	L6	103,507	2,013,389	19.45
21	L7	95,970	1,436,508	14.97
22	H1	104,180	1,753,314	16.83
23	H2	103,075	2,038,579	19.78
24	H3	94,673	1,471,044	15.54
25	H4	129,323	1,748,843	13.52
26	H5	102,510	1,644,827	16.05

Number	Sample number	SLAF tag number	Total depth of sequencing	Average sequencing depth
27	H6	107,770	2,035,052	18.88
28	H7	98,702	1,780,780	18.04

Ww, Weixi population of *G. elata f. elata*; W, Yiliang population of *G. elata f. elata*; L, Yiliang population of *G. elata f. viridis*; H, Yiling population of *G. elata f. elata*.

Table 2
Details of SNPs in *G. elata*

Number	Sample number	Total number of SNPs	Corresponding SNP number	SNP integrity	SNP heterozygosity
1	Ww1	60,238	33,146	26.29	55.02
2	Ww2	60,238	45,240	54.43	75.1
3	Ww3	60,238	34,722	23.97	57.64
4	Ww4	60,238	34,799	26.01	57.76
5	Ww5	60,238	30,332	26	50.35
6	Ww6	60,238	30,492	24.05	50.61
7	Ww7	60,238	23,766	18.34	39.45
8	W1	60,238	35,189	28.78	58.41
9	W2	60,238	30,950	25.87	51.37
10	W3	60,238	28,021	23.48	46.51
11	W4	60,238	32,037	25.74	53.18
12	W5	60,238	31,067	24.72	51.57
13	W6	60,238	32,205	25.54	53.46
14	W7	60,238	29,590	25.33	49.12
15	L1	60,238	24,774	22.21	41.12
16	L2	60,238	25,817	24.22	42.85
17	L3	60,238	25,825	25.75	42.87
18	L4	60,238	28,603	24.55	47.48
19	L5	60,238	26,251	25.16	43.57
20	L6	60,238	24,444	22.16	40.57
21	L7	60,238	22,567	21.49	37.46
22	H1	60,238	24,678	25.46	40.96
23	H2	60,238	24,884	24.47	41.3
24	H3	60,238	21,848	22.55	36.26
25	H4	60,238	27,951	22.11	46.4
26	H5	60,238	23,091	27.81	38.33

Number	Sample number	Total number of SNPs	Corresponding SNP number	SNP integrity	SNP heterozygosity
27	H6	60,238	27,742	25.06	46.05
28	H7	60,238	22,596	27.7	37.51

Ww, Weixi population of *G. elata* f. *elata*; W, Yiliang population of *G. elata* f. *elata*; L, Yiliang population of *G. elata* f. *viridis*; H, Yiling population of *G. elata* f. *elata*.

Genetic structure analysis

The SNP markers from 60,238 samples were developed by SLAF-seq. The population structure of *G. elata* was constructed based on the different K values (1–10) and the minimum CV error related to K values was computed using admission software. The results showed that four *G. elata* germplasms mainly originated from two different primitive ancestors (K = 2) (Fig. 3). Specifically, wild *G. elata* f. *glauca* and *G. elata* f. *viridis* were from two different original ancestors (Q1 and Q2). On the other hand, some samples of cultivated *G. elata* f. *glauca* and *G. elata* f. *elata* were in Q1, and the others were in Q2 (Table 3).

Table 3
The relationship between sample and population of 28 *G. elata* samples.

Number	Sample number	Q1	Q2	Group
1	Ww1	0.999990	0.000010	Q1
2	Ww2	0.999990	0.000010	Q1
3	Ww3	0.999990	0.000010	Q1
4	Ww4	0.999990	0.000010	Q1
5	Ww5	0.999990	0.000010	Q1
6	Ww6	0.999990	0.000010	Q1
7	Ww7	0.999990	0.000010	Q1
8	W1	0.999990	0.000010	Q1
9	W2	0.589848	0.410152	Q1
10	W3	0.999990	0.000010	Q1
11	W4	0.221396	0.778604	Q2
12	W5	0.209159	0.790841	Q2
13	W6	0.545204	0.454796	Q1
14	W7	0.633856	0.366144	Q1
15	L1	0.000010	0.999990	Q2
16	L2	0.000010	0.999990	Q2
17	L3	0.000010	0.999990	Q2
18	L4	0.000010	0.999990	Q2
19	L5	0.000010	0.999990	Q2
20	L6	0.000010	0.999990	Q2
21	L7	0.000010	0.999990	Q2
22	H1	0.312125	0.687875	Q2
23	H2	0.313867	0.686133	Q2
24	H3	0.999990	0.000010	Q1
25	H4	0.964960	0.035040	Q1
26	H5	0.999990	0.000010	Q1
27	H6	0.558055	0.441945	Q1

Number	Sample number	Q1	Q2	Group
28	H7	0.436665	0.563335	Q2

Notes
Ww, Weixi population of *G. elata f. elata*; W, Yiliang population of *G. elata f. elata*; L, Yiliang population of *G. elata f. viridis*; H, Yiling population of *G. elata f. elata*. Q1, Possibility from the first primitive ancestor; Q2, Possibility from the second primitive ancestor.

Phylogenetic analysis

MEGA7 software was used to screen for MAF > 0.05 with integrity > 0.5. A total of 22,737 SNPs were obtained for further analysis. A phylogenetic tree comprising 28 samples was constructed to analyze the evolutionary genetic relationships (Fig. 4). Seven Ww (wild *G. elata f. glauca*) samples, 3 W (cultivated *G. elata f. glauca*) samples, and 3 H (*G. elata f. elata*) samples were clustered into one group. Meanwhile, 7 L (*G. elata f. viridis*) samples, 4 W samples, and 4 H samples were clustered into another group. These results further confirm that wild type *G. elata f. glauca* (Ww) and *G. elata f. viridis* (L) came from two different original ancestors (Q1, Q2), consistent with the results of genetic structure analysis.

Population genetic diversity

To evaluate the genetic diversity of the four inferred groups, genetic parameters including Nei's genetic diversity index (N), observed heterozygosity, PIC, Shannon Wiener index (I), and MAF were estimated separately. (Table 4). The average effective allele (Ae) of the four populations was 1.5852. The percentage of polymorphic loci varied from 14.2 to 30.51%, with an average of 20.94%. The average Nei's genetic diversity (H) was 0.3059, while Shannon's polymorphic information index (I) was 0.3885. Genetic diversity of *G. elata f. elata*. is the highest PIC = 0.3051, N = 0.4547, I = 0.5666, and *G. elata. f. viridis* is the lowest with P = 0.1420, N = 0.2067, I = 0.2632.

Table 4
Summary statistics of genetic variation in *G. elata* population-based on SNP markers

Code	Ho	Ae	N	MAF	I	PIC
W	0.2681	1.6933	0.4547	0.3214	0.5666	0.3051
H	0.2291	1.7123	0.3275	0.3330	0.4234	0.2283
L	0.2160	1.4315	0.2067	0.3623	0.2632	0.1420
Ww	0.2139	1.5037	0.2347	0.3420	0.3007	0.1621

Genetic differentiation analysis Notes

Ho: observed heterozygosity; Ae, mean number of alleles per locus; N, Nei's genetic diversity index; MAF, minor allele frequency; I, Shannon Wiener index; PIC, polymorphic information content.

**indicates significant difference in genetic diversity at 0.01 level between the entire and core germplasm set collection as revealed by t-test.

Analysis of molecular variance (AMOVA) indicated that 14.31% of the total variation was distributed between populations, 6.54% among individuals within populations, and 92.23% within individuals. Variation among individuals (92.23%) was significantly higher than among populations (14.31%). The fixed coefficient $F_{ST}=0.14310$ ($P<0.05$) indicated that the genetic differentiation among different populations of *G. elata* was not significant ($F_{ST} < 0.25$) (Table 5).

Table 5
Results of AMOVA between the core collection and the remaining collection

Source of variation	DF	SS	VC	Percentage of variation
Among populations	3	28898.250	493.20493	14.31
Among individuals within populations	24	65469.143	-225.39881	-6.54
Within individuals	28	89003.000	3178.67857	92.23
Total	55	94367.393	267.80612	

Notes

DF, degree of freedom; SS, sum of square; VC, Variance components.

These variances were statistically significant at $P < 0.001$.

Genetic distance analysis

Nei's unbiased genetic distance was used to calculate the average genetic distance and genetic identity among populations (Table 6). The results showed that the maximum genetic distance was 0.007 between L and H, while the minimum genetic distance was 0.003 between H and Ww. Meanwhile, the maximum genetic identity was 0.548 between H and W, while the minimum genetic distance was 0.406 between L and Ww.

Table 6
Matrix of pairwise Nei's genetic distance and genetic distance of the four inferred groups.

	W	H	L	Ww
W		0.006	0.005	0.005
H	0.548		0.007	0.003
L	0.507	0.461		0.006
Ww	0.479	0.430	0.406	

Notes

Below diagonal Nei's genetic identity; above diagonal genetic distance.

Principal component analysis

The first three principal coordinates (the first, second, and third principal coordinates) explained 44.52% of the total variation (Fig. 5). Specifically, the first, second, and third principal coordinates explained 11.79, 11.88, and 20.85% of the total variation, respectively. PCA was used to evaluate the genetic evolution of 28 *G. elata* samples. WW and L populations were divided into two groups, while W and H populations were scattered over the WW and L populations. The results of PCA analysis were consistent with those of genetic structure and phylogenetic tree analyses.

Discussion

Molecular markers have been widely used to characterize traditional Chinese medicinal plants [Sun et al. 2017; Mir et al. 2021; Ma et al.2021; Agyenim-Boateng et al. 2019]. High-throughput DNA sequencing technology has promoted gene discovery and genome exploration, providing a new strategy for SNP genotyping. Genome-wide resequencing strategy can be used to identify genotypic variations among samples. SLAF sequencing technology has been applied to screen specific SNP loci in many plant species [He et al. 2017; Chen et al. 2022; Chang et al. 2019; Du et al. 2019; Zhao et al. 2019]. Molecular marker density directly affects population heredity. In this study, high-density SNP markers were used to analyze the population and genetic diversity of *G. elata*. SLAF-seq can be used to develop SNP markers with high density, efficiency, and consistency and has been widely applied to analyze genetic variation among populations. Chen and others [Chen et al. 2022] constructed a phylogenetic tree involving 50 germplasms based on SNPs and grouped three *Miscanthus* species into two clades: one clade comprised *M. sinensis* alone, while the other included the offspring *M. lutarioriparius*, and *M. sacchariflorus*. Wang and others [Wang et al. 2020] analyzed the genetic diversity of 234 taro accessions from 16 provinces of China using 132,869 SNP markers identified by SLAF-seq and categorized the accessions into eight groups. AMOVA analysis showed that the variation among the eight inferred groups was higher than within groups, while a relatively small variance was found among the two morphological types and the 16 collection regions. SLAF-seq was applied to create a high-density genetic map for asparagus bean.

Subsequently, 55,437 polymorphic SLAF markers were developed and utilized to mine SNP markers. Comparative genome analysis was also performed, which revealed that asparagus bean is more related to adzuki bean than soybean, common bean, and mung bean [Huang et al. 2018]. SLAF-seq is a fast, accurate, efficient, and cost-effective strategy for de novo SNP discovery and genotyping of large populations [Sun et al. 2013].

There has been limited genetics and genomics research on *G. elata*. However, research on the genetic diversity of *G. elata* has been conducted using various markers, including AFLP [Tsai et al. 2014; Guan et al. 2013; Xie et al. 2007], RAPD [Zou et al. 2006], ISSR [Chen et al. 2007], SRAP [Chai et al. 2014], and ITS [Wang et al. 2010]. Guan and others [Guan et al. 2013] analyzed the polymorphism among *G. elata* accessions using ISSR and AFLP molecular markers. Chen et al. [Chen et al. 2007] revealed significant genetic differences between wild and cultivated *G. elata*; notably, the degree of genetic variation between *G. elata f. elata* and *G. elata f. flavida* S. Chow was less than that between *G. elata f. glauca* and *G. elata f. viridis*. Chai and others used SRAP molecular marker to assess the genetic diversity of three *G. elata* germplasm resources and found that the genetic traits of *G. elata f. elata* were relatively stable. They also found that there was a lack of gene communication between *G. elata f. viridis*, *G. elata f. glauca*, and *G. elata f. glauca* variants with very little genetic diversity. However, the variants of *G. elata f. glauca* and *G. elata f. viridis* had high genetic diversity. Existing reports have diverse views on the causes of genetic variations among *G. elata* varieties. Guan et al. [Guan et al. 2013] believe that the environment exerts significant influence on the genetic diversity of *G. elata*, arguing that the genetic of *G. elata* in the same area is obviously different. Other factors have also been shown to affect the genetic variation in *G. elata*, including genetics [Xie et al. 2007], genetic background, and heredity [Chai et al. 2014; Wang et al. 2010].

In this study, AMOVA analysis showed that genetic variation among *G. elata* populations was mainly among individuals of the same population. The variation between populations was small, and genetic differentiation between populations was not significant ($F_{ST} = 0.14310$, $P < 0.05$). Chai et al. [Chai et al. 2014] analyzed genetic diversity of *G. elata* using SRAP molecular markers and revealed that the variation within *G. elata* populations was greater than between *G. elata* populations, but there was great genetic differentiation among *G. elata* populations. Zhou et al. [Zhou et al. 2018] analyzed molecular variance among *G. elata* populations and showed that 38.18% of the variation came from the variants, 17.68% from the populations within the variants, and the variation within the populations accounted for 44.14%. Notably, the genetic differentiation among *G. elata* populations was strong ($F_{ST} = 0.5586$), which may be related to their geographical origin, sample size, and the technical methods adopted.

Genetic evolution analysis performed in this study showed that four groups of *G. elata* probably came from two different primitive ancestors. The most representative of these two primitive ancestors were wild *G. elata f. glauca* and *G. elata f. viridis*. This may reveal the evolutionary route of the three varieties of *G. elata* and is consistent with the planting conditions of the various *G. elata* varieties. *G. elata f. glauca* and *G. elata f. elata* were scattered between two primitive ancestors based on the genetic structure and phylogenetic analysis, suggesting that W population has mixed genetic origins. These may be related to many "Wuhong Tianma" hybrids between *G. elata f. glauca* and *G. elata f. elata* in Zhao

tong, Yunnan, China (a major trading market of *G. elata* in China for the last 20 years). Wang and others [Wang et al. 2010] identified genetic variations among *G. elata* f. *flavida*, *G. elata* f. *viridis* and *G. elata* f. *glauca* using ITS and showed that *G. elata* f. *glauca* was an older type than *G. elata* f. *flavida* and *G. elata* f. *viridis*. This suggested that contrary to our findings, *G. elata* f. *flavida* and *G. elata* f. *viridis* evolved from *G. elata* f. *glauca*. *G. elata* variety may have originated from two different original ancestors: wild *G. elata* f. *glauca* and *G. elata* f. *viridis*. Samples of *G. elata* from the two provinces are distributed, between these two ancestors but the environmental conditions of the two provinces are different. Although there are different variants of black, red, and green, the effect of the environment in their generation cannot be ruled out. Transcriptomics and genomics studies can verify the findings of this study.

Declarations

Acknowledgments We thank Dr. Jian Fu (Biotechnology and Germplasm Resources Institute, Yunnan Academy OF Agricultural Sciences) for assisting us in data analysis.

Author contributions

ZX: performed data analysis and wrote the first draft of the manuscript; LZ: manuscript review; KC and PZJ: Collect *Gastrodia elata* resources; PZJ: Project designer and person in charge; HLQ, HC, XLC, JYS, and JHD: Guide experimental design, data analysis, paper writing, and revision. All authors read and approved the final manuscript.

Funding This research was funded by the Innovative Talents Foundation of Yunnan Province (2018HB085), the Yunnan Provincial Environmental Protection Special Fund Project (2014BI006), the high-level talents Foundation of Yunnan University of Chinese Medicine, and the Innovation Team of Yunnan Colleges and Universities.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Agyenim-Boateng KG, Lu J, Shi Y, Zhang D, Yin X 2019 SRAP analysis of the genetic diversity of wild castor (*Ricinus communis* L.) in South China. PLoS One 14(7): e0219667. <https://doi.org/10.1371/journal.pone.0219667>.
2. Arif IA, Khan HA, Bahkali AH, Al Homaidan AA, Al Farhan AH, Al Sadoon M, Shobrak M (2011) DNA marker technology for wildlife conservation. Saudi J Biol Sci 18(3):219-25. <https://doi.org/10.1016/j.sjbs.2011.03.002>.
3. Brookes AJ (1999) The essence of SNPs. Gene 234(2):177-86. [https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/10.1016/S0378-1119(99)00219-X).
4. Chai K, Liu HC, Li JL, Luo FL, Wang HL, Huang MJ, Luo CL, Zhao Z (2014) Genetic diversity of *Gastrodia elata* based on SRAP molecular markers. Chinese Herbal Medicine 45(20):2974-2981.

5. Chang E, Tian Y, Wang C, Deng N, Jiang Z, Liu C. Exploring the Phylogeography of Ancient *Platyclusus orientalis* in China by Specific-Locus Amplified Fragment Sequencing. *Int J Mol Sci*. 2019 Aug 8;20(16):3871. <https://doi.org/10.3390/ijms20163871>.
6. Chang Y, Ding J, Xu Y, Li D, Zhang W, Li L, Song J (2018) SLAF-based high-density genetic map construction and QTL mapping for major economic traits in sea urchin *Strongylocentrotus intermedius*. *Sci Rep* 8(1):820. <https://doi.org/10.1038/s41598-017-18768-y>.
7. Chen SC, Tsi ZH, Lang KY, Zhu GH (1999) *Flora Reipublicae Popularis Sinicae*. Science Press, Beijing (in Chinese with English abstract).
8. Chen Z, He Y, Iqbal Y, Shi Y, Huang H, Yi Z (2022) Investigation of genetic relationships within three *Miscanthus* species using SNP markers identified with SLAF-seq. *BMC Genomics* 23(1):43. <https://doi.org/10.1186/s12864-021-08277-8>.
9. Chen ZY, Wang XL, Song JX (2007) ISSR Preliminary Analysis of Genetic Polymorphism of *Gastrodia elata* in Guizhou. *China journal of traditional Chinese medicine and pharmacy* (07):436-439.
10. Du X, Zhang X, Bu H, Zhang T, Lao Y, Dong W (2019) Molecular Analysis of Evolution and Origins of Cultivated Hawthorn (*Crataegus* spp.) and Related Species in China. *Front Plant Sci* 10:443. <https://doi.org/10.3389/fpls.2019.00443>.
11. Excoffier L, Laval G, Schneider S (2007) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47-50.
12. Guan P, Shi JM, Chen F (2013) AFLP analysis of *Gastrodia elata* in different distribution areas. *journal of plant genetic resources* 14(01):70-77.
13. He D, Zhang J, Zhang X, He S, Xie D, Liu Y, Li C, Wang Z, Liu Y (2022) Development of SSR markers in *Paeonia* based on *De Novo* transcriptomic assemblies. *PLoS One* 15(1): e0227794. <https://doi.org/10.1371/journal.pone.0227794>
14. He Y, Yuan W, Dong M, Han Y, Shang F (2017) The First Genetic Map in Sweet Osmanthus (*Osmanthus fragrans* Lour.) Using Specific Locus Amplified Fragment Sequencing. *Front Plant Sci* 8:1621. <https://doi.org/10.3389/fpls.2017.01621>.
15. Huang H, Tan H, Xu D, Tang Y, Niu Y, Lai Y, Tie M, Li H (2018) High-density genetic map construction and comparative genome analysis in asparagus bean. *Sci Rep* 8(1):4836.
16. Huang HT (2018) Screening of Cold Tolerance Germplasm Resources, Construction of Genetic Map and QTL Mapping of Cold Tolerance in Pod Setting Stage. Dissertation, Sichuan Agricultural University.
17. Kazemeini F, Asri Y, Mostafavi G, Kalvandi R, Mehregan I (2020) RAPD-based evaluation of genetic diversity among populations of the Iranian endemic species *Rhabdosciadium aucheri* Boiss. (Apiaceae). *Mol Biol Rep* 47(12):9345-9352. <https://doi.org/10.1007/s11033-020-06028-8>.
18. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-4. <https://doi.org/10.1093/molbev/msw054>.
19. Liang YT, Jiang C, Zhou JH, Hu QT, Yuan Y. (2019) Molecular identification of *Armillaria gallica* by PCR-RFLP analysis. *Zhongguo Zhong Yao Za Zhi* 44(17):3622-3626.

<https://doi.org/10.19540/j.cnki.cjcmm.20190701.102>.

20. Li DX, Chen G, Li FB, Ma YX (2007) Comparison of gastrodin content in different variants of *Gastrodia elata*. Research and Practice of Modern Traditional Chinese Medicine, (03):23-24.
21. Li HF, Huang YM, Li YQ, Hua JF, Wu CR, Fan JZ, Chen TY (2019) Genetic evolution analysis of sweet potato germplasm resources based on SLAF-seq technology. Acta Tropical Crops Sinica, 40(12):2390-2396.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-9. <https://doi.org/10.1093/bioinformatics/btp352>.
23. Lou SK, Wong KL, Li M, But PP, Tsui SK, Shaw PC (2010) An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). BMC Genomics 11:402. <https://doi.org/10.1186/1471-2164-11-402>.
24. Luz GC, Strioto DK, Mangolin CA, Machado MFPS (2020) ISSR markers to assess genetic diversity of cultivated populations from artificial selection of *Stevia rebaudiana* (Bert.) Bertoni. Breed Sci. 70(4):508-514. <https://doi.org/10.1270/jsbbs.20014>.
25. Lv D, Zhang C, Yv R, Yao J, Wu J, Song X, Jian J, Song P, Zhang Z, Han D, Sun D (2021) Utilization of a Wheat50K SNP Microarray-Derived High-Density Genetic Map for QTL Mapping of Plant Height and Grain Traits in Wheat. Plants (Basel) 10(6):1167. <https://doi.org/10.3390/plants10061167>.
26. Ma M, Wang T, Lu B (2021) Assessment of genetic diversity in *Amomum tsao-ko* Crevost & Lemarié, an important medicine food homologous crop from Southwest China using SRAP and ISSR markers. Genet Resour Crop Evo 68(6):2655-2667. <https://doi.org/10.1007/s10722-021-01204-6>.
27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297-303. <https://doi.org/10.1101/gr.107524.110>.
28. Mir MA, Mansoor S, Sugapriya M, Alyemeni MN, Wijaya L, Ahmad P (2021) Deciphering genetic diversity analysis of saffron (*Crocus sativus* L.) using RAPD and ISSR markers. Saudi J Biol Sci 28(2):1308-1317. <https://doi.org/10.1016/j.sjbs.2020.11.063>.
29. Mishra RK, Tomar RS (2014) Molecular markers and their application in genetic mapping. Eur Acad Res 2(3):4012–4040
30. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8(19):4321-5. <https://doi.org/10.1093/nar/8.19.4321>.
31. National Pharmacopoeia Commission (2020) Pharmacopoeia of the People's Republic of China. Part I. China Medical Science and Technology Press, Beijing, pp 59-60.
32. Nei M (1987) Molecular evolutionary genetics. Columbia university press, New York.
33. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12):e190. <https://doi.org/10.1371/journal.pgen.0020190>.

34. Shi LM (1990) Genetic diversity and its preservation. *Information of Biological Sciences*, (04):158-164.
35. Sorkheh K, Masaeli M, Chaleshtori MH, Adugna A, Ercisli S (2019) Expression of Concern: AFLP-Based Analysis of Genetic Diversity, Population Structure, and Relationships with Agronomic Traits in Rice Germplasm from North Region of Iran and World Core Germplasm Set. *Biochem Genet* 57(5):745. <https://doi.org/10.1007/s10528-019-09933-1>.
36. Sun M, Zhang C, Zhang X, Fan Y, Fu K, Wu W, Bai S, Zhang J, Peng Y, Huang L, Yan Y, Ma X (2017) AFLP assessment of genetic variability and relationships in an Asian wild germplasm collection of *Dactylis glomerata* L. *C R Biol* 340(3):145-155. <https://doi.org/10.1016/j.crvi.2016.12.003>.
37. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H (2013) SLAF-seq: an efficient method of large-scale *De novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS One* 8(3): e58700. <https://doi.org/10.1371/journal.pone.0058700>.
38. Su W, Wang L, Lei J, Chai S, Liu Y, Yang Y, Yang X, Jiao C (2017) Genome-wide assessment of population structure and genetic diversity and development of a core germplasm set for sweet potato based on specific length amplified fragment (SLAF) sequencing. *PLoS One* 12(2): e0172066. <https://doi.org/10.1371/journal.pone.0172066>.
39. Tsai CC, Wu PY, Kuo CC, Huang MC, Yu SK, Hsu TW, Chiang TY, Chiang YC (2014) Analysis of microsatellites in the vulnerable orchid *Gastrodia flavilabella*: the development of microsatellite markers, and cross-species amplification in *Gastrodia*. *Bot Stud*. Dec;55(1):72.
40. Wang DX (2010) ITS sequence analysis and identification of variation types of *Gastrodia elata*. *Biotechnology* 20(06):33-35.
41. Wang Y, Shahid MQ, Ghouri F, Baloch FS (2020) De Novo Assembly and Annotation of the Juvenile Tuber Transcriptome of a *Gastrodia elata* Hybrid by RNA Sequencing: Detection of SSR Markers. *Biochem Genet* 58(6):914-934. <https://doi.org/10.1007/s10528-020-09983-w>.
42. Wang Y, Shahid MQ, Ghouri F, Ercisli S, Baloch FS (2019) Development of EST-based SSR and SNP markers in *Gastrodia elata* (herbal medicine) by sequencing, de novo assembly and annotation of the transcriptome. *3 Biotech* 9(8):292. <https://doi.org/10.1007/s13205-019-1823-4>.
43. Wang YS (2016) Development Status, Existing Problems and Countermeasures of *Gastrodia elata* in Qiandongnan Prefecture. *Guizhou Agricultural Science* 44(02):125-128.
44. Wang Z, Sun Y, Huang X, Li F, Liu Y, Zhu H, Liu Z, Ke W (2020) Genetic diversity and population structure of eddoe taro in China using genome-wide SNP markers. *PeerJ* 8: e10485. <http://doi.org/10.7717/peerj.10485>.
45. Wen Y, Fang Y, Hu P, Tan Y, Wang Y, Hou L, Deng X, Wu H, Zhu L, Zhu L, Chen G, Zeng D, Guo L, Zhang G, Gao Z, Dong G, Ren D, Shen L, Zhang Q, Xue D, Qian Q, Hu J (2020) Construction of a High-Density Genetic Map Based on SLAF Markers and QTL Analysis of Leaf Size in Rice. *Front Plant Sci* 11:1143. <https://doi.org/10.3389/fpls.2020.01143>

46. Xie Y, Li Y, Zhang XL, Jiang CH, Shan KR, Wu XL (2007) Preliminary application of AFLP technology in genetic variation of *Gastrodia elata*. Plant physiology communications, (01):141-144.
47. Xu JT (1993) Cultivation of *Gastrodia elata* in China. Beijing Medical University and Peking Union Medical College Joint Press, Beijing.
48. Yang S, Xue S, Kang W, Qian Z, Yi Z (2019) Genetic diversity and population structure of *Miscanthus lutarioriparius*, an endemic plant of China. PLoS One 14(2): e0211471. <https://doi.org/10.1371/journal.pone.0211471>.
49. Yang S, Yu W, Wei X, Wang Z, Zhao Y, Zhao X, Tian B, Yuan Y, Zhang X (2020) An extended KASP-SNP resource for molecular breeding in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). PLoS One 15(10): e0240042. <https://doi.org/10.1371/journal.pone.0240042>.
50. Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belgian Journal of Botany,; 129:157.
51. Yeh FC, Yang RC, Boyle T (1999) Popgene Version 1.31. Microsoft Window-based Freeware for Population Genetic Analysis. Department of Renewable Resources, University of Alberta, Edmonton, AB Canada.
52. Yousef EAA, Müller T, Börner A, Schmid KJ (2018) Comparative analysis of genetic diversity and differentiation of cauliflower (*Brassica oleracea* var. *botrytis*) accessions from two ex situ genebanks. PLoS One 13(2): e0192062. <https://doi.org/10.1371/journal.pone.0192062>.
53. Yu X, Hu WZ, Jiang AL, Li H, Yang L, Song CL (2016) Research progress on active components and functions of *Gastrodia elata*. Food Industry Science and Technology,37(08):392-395+399.
54. Zhao J, Xu Y, Li H, Yin Y, An W, Li Y, Wang Y, Fan Y, Wan R, Guo X, Cao Y (2019) A SNP-Based High-Density Genetic Map of Leaf and Fruit Related Quantitative Trait Loci in Wolfberry (*Lycium* Linn.). Front Plant Sci. Aug 10:977. <https://doi.org/10.3389/fpls.2019.00977>.
55. Zhou TH, Ding JX, Xu H, Chen C (2018) SSR fingerprint of *Gastrodia elata* germplasm resources. Journal of Northwest Botany 38(05):830-838.
56. Zhou X (2013) A variety of *Gastrodia elata* in today's drug market. Institute of Medicinal Plants, Chinese Academy of Sciences, Beijing.
57. Zhou YL, Fu TH, Fan QJ, Ren ZL (2008) Genetic polymorphism analysis of wild and cultivated *Gastrodia elata* Blume by RAPD and AFLP markers. Anhui Agricultural Sciences (17):7119-7123.
58. Zou JN, Song JX, Chang CR, Wang XL (2006) RAPD analysis on the germplasm resources of *Gastrodia elata* in Guizhou. Zhong Yao Cai 29(9):881-3.

Figures



Figure 1

Geographical location of *G. elata* accessions used in present study. (Maps sources: NGCC). Ww, Weixi population of *G. elata* f. *elata*; W, Yiliang population of *G. elata* f. *elata*; L, Yiliang population of *G. elata* f. *viridis* L, Yiling population of *G. elata* f. *elata*.



Figure 2

Morphology of samples collected from different *G. elata* populations.

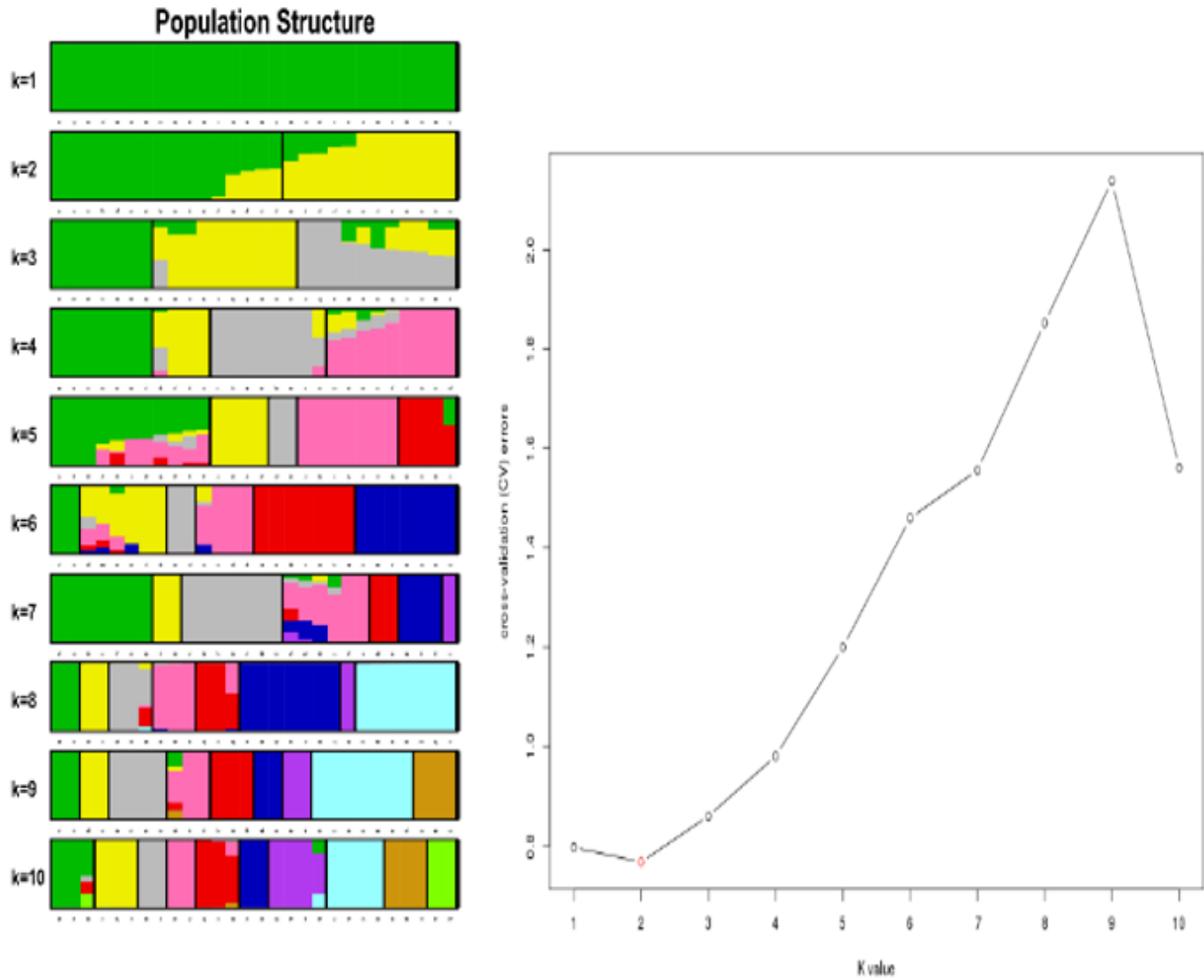


Figure 3

Cluster graphs of K=1–10 with 28 samples. Each color in the left figure represents a cluster, with each row representing a cluster value. In the right figure, ΔK is the value that corresponds to each K value. ΔK was minimized when K=2

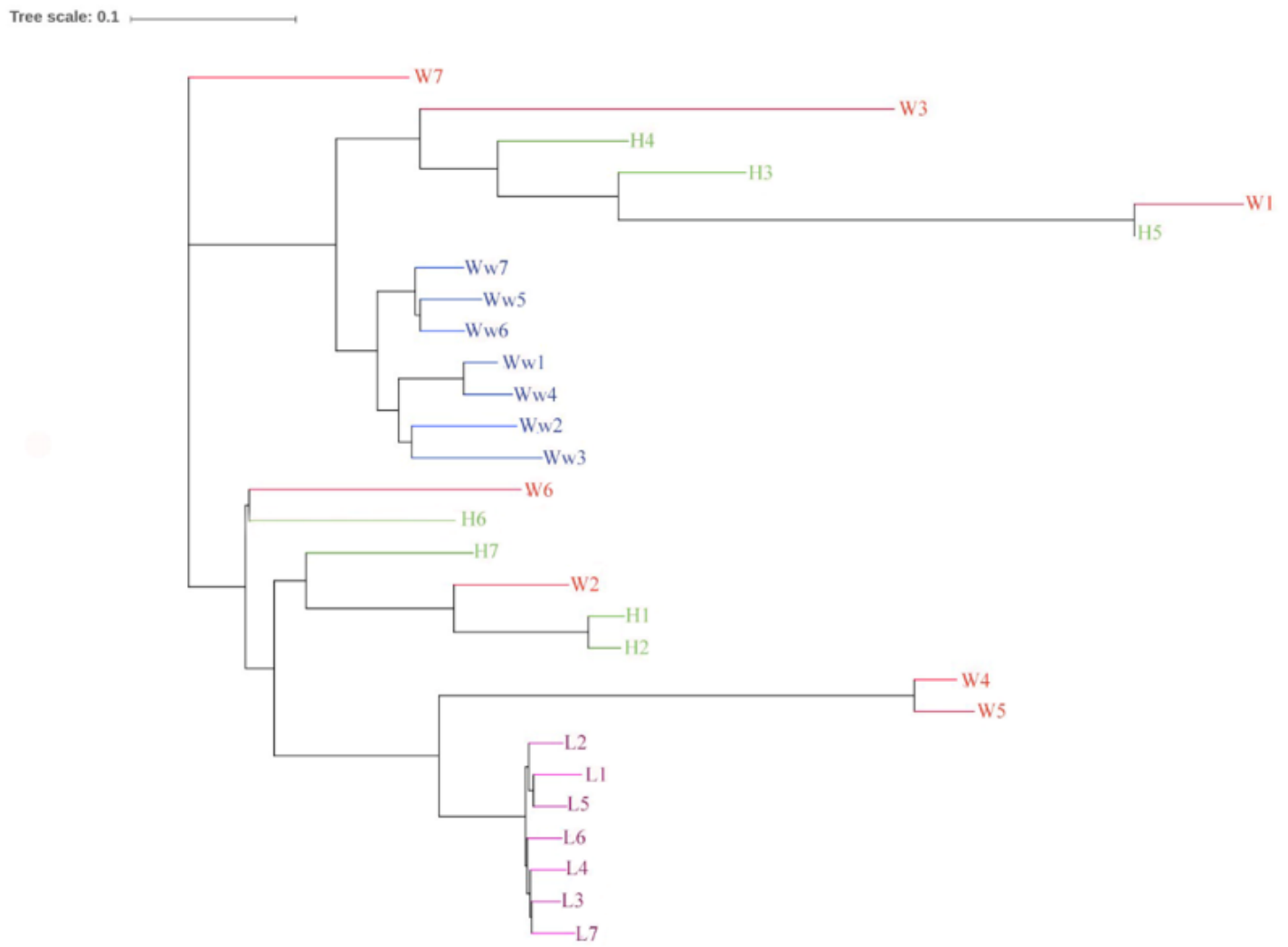


Figure 4

A maximum likelihood phylogenetic tree of the 28 *G. elata* accessions based on 60238 SNPs.

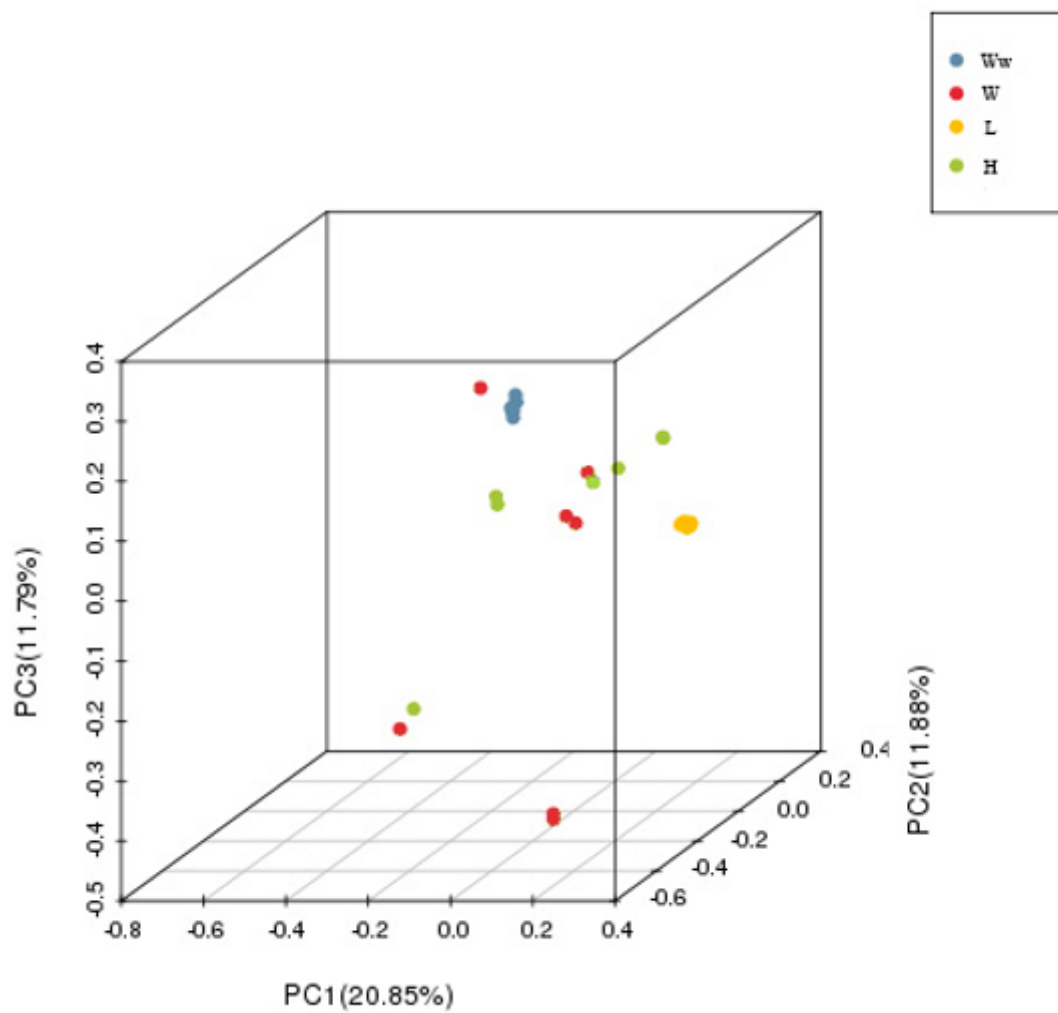


Figure 5

Principal component analysis was performed on 28 *G. elata* individuals based on highly consistent population SNPs.