

Phenotypic and genome-wide differentiation among three cultivated variants of *Gastrodia elata* and exploratory mapping of quality-related traits

Rui Yang

Guizhou University

FU-Lai Luo

luo512658@163.com

Guizhou University

Ping Wen

Guizhou Jiulong *Gastrodia elata* Co., Ltd

Article

Keywords: *Gastrodia elata*, cultivated variants, whole-genome resequencing, population genetic structure, exploratory GWAS, quality-related traits

Posted Date: June 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9200185/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Gastrodia elata Bl. is an important medicinal plant in China. Red, green and black cultivated variants differ in medicinal quality and processing-related traits. Here, 28 accessions were analysed using whole-genome resequencing and 30 phenotypic traits to compare phenotypic differentiation and population genetic background among the three variants and to screen quality-related candidate regions and genes. All 30 traits varied, and 16 remained significantly different after false discovery rate correction. We identified 2.486 million high-quality SNPs. Population genomic analyses indicated genome-wide differentiation among variants, with red *Gastrodia* more clearly separated from the other two variants and black *Gastrodia* showing a more complex internal genetic composition. Red *Gastrodia* had lower nucleotide diversity and slower LD decay. Comparison between red and black *Gastrodia* identified 1,679 relatively highly differentiated regions. Exploratory GWAS detected 225 significant association signals, mainly for polysaccharide, moisture and gastrodin contents. Integration of association intervals, genotype distributions and qPCR results highlighted Gel14G16652, Gel02G02702 and Gel06G08670 as candidate genes for further investigation. These results provide a population-level basis for comparing genetic differentiation among cultivated variants of *G. elata* and for screening quality-related candidate regions and genes.

1 Introduction

Gastrodia elata Bl. is a traditional medicinal herb widely used in China, and its dried tuber has a long history of medicinal application. Previous studies have shown that phenolic constituents in *G. elata*, including gastrodin, *p*-hydroxybenzyl alcohol and parishins, are closely associated with its major pharmacological activities and constitute an important chemical basis for quality evaluation [1, 2]. During long-term cultivation and utilization, different cultivated variants of *G. elata* have shown variation in tuber morphology, accumulation of bioactive constituents and processing-related traits [3], suggesting that these phenotypic differences may be associated with underlying genetic factors.

According to current classification and cultivation practice, red, green and black *Gastrodia* are among the most common cultivated variants in China. Existing studies have mainly focused on comparisons of morphological traits or phenotypic diversity, whereas systematic analyses of genome-wide differentiation, genetic diversity and the genetic basis of quality-related traits among these three major cultivated variants remain limited [4, 5]. In recent years, the publication of reference genomes for *G. elata* has provided a foundation for related research; however, most studies have focused on species-level genomic features, and population genomic studies targeting major cultivated variants are still relatively scarce [6, 7].

On the other hand, medicinal quality and processing-related traits of *G. elata* tubers are complex quantitative traits that are likely influenced by multiple genes and by the underlying population genetic background. Genome-wide association studies (GWAS) provide an effective framework for dissecting relationships between natural genetic variation and phenotypic traits. Previous studies have shown that clonally propagated crops often exhibit relatively long linkage disequilibrium blocks, which may facilitate the detection of association signals under certain conditions but may also reduce mapping resolution [8].

Therefore, under a framework in which population structure and genetic background are first clarified, combining whole-genome resequencing with multi-trait phenotypic data to conduct exploratory population genomic and association analyses of major cultivated variants of *G. elata* has clear practical relevance.

In this study, 28 accessions representing three cultivated variants of *G. elata* (red, green and black) were analysed using whole-genome resequencing and multi-trait phenotypic assessment. We compared phenotypic differentiation, population structure and genetic diversity among cultivated variants, identified relatively highly differentiated genomic regions, and performed exploratory GWAS for quality-related traits. On this basis, candidate genes were further screened by integrating association interval positions, genotype distributions and qPCR expression results. The results provide a population-level basis for evaluating cultivated germplasm of *G. elata*, analysing genetic differentiation among cultivated variants, and screening candidate genes related to quality traits.

2. Materials and methods

2.1 Plant materials

A total of 28 cultivated accessions of *Gastrodia elata* collected from major production areas in Guizhou Province, China, were used in this study, including 13 red *Gastrodia*, 9 black *Gastrodia*, and 6 green *Gastrodia*. All accessions were identified by Prof. Fulai Luo of Guizhou University as *Gastrodia elata* Bl. and the corresponding cultivated variants.

To minimize environmental effects on phenotypic comparisons, all materials were propagated and cultivated under uniform conditions. In February 2023, all source materials were planted indoors at 25°C with a relative humidity of 70–80%. Artificial self-pollination was conducted from April to May 2023 to obtain selfed seeds. In June 2023, the seeds were sown at a single experimental site (Dexi Xiangbosen Planting Base, Bijie, Guizhou Province, China; 27°31'39" N, 105°35'47" E; 1594 m a.s.l.) for cultivation. *Cyclobalanopsis glauca* logs were used as host wood, and the associated symbiotic fungi were *Mycena dendrobii* and *Armillaria* sp. strain A9. Three replicate plots were established for each accession, and sowing density, host wood dosage, inoculum amount, and field management practices were standardized across plots.

In December 2024, samples were harvested at the tuber dormancy stage. One tuber showing no obvious external abnormality and relatively similar morphology was selected from the three replicate plots of each accession as the analytical sample for agronomic trait measurement, quality-related trait measurement, whole-genome resequencing, and subsequent qPCR expression analysis. This sampling strategy was used to maintain one-to-one correspondence between resequencing samples and phenotypic measurements while minimizing the influence of conspicuous tuber morphological differences among accessions.

2.2 Major reagents and instruments

The major reagents included standards of gastrodin, *p*-hydroxybenzyl alcohol, parishins A, B, C and E, and glucose (all ≥ 98% purity; Shanghai Yuanye Bio-Technology Co., Ltd., China), single-element standard

solutions of P, K, Ca, Mg, Na, Fe, Mn, B, Zn and Cu (National Center of Analysis and Testing for Nonferrous Metals and Electronic Materials, China), and methanol and acetonitrile (HPLC grade), as well as petroleum ether, absolute ethanol, sulfuric acid and anthrone (analytical grade).

The major instruments included an electronic balance, a vernier caliper, an Agilent 1260 HPLC system (Agilent Technologies, USA), a Waters C18 column (150 mm × 3.9 mm, 4 μm), an Optima 8100 inductively coupled plasma optical emission spectrometer (ICP-OES; PerkinElmer, USA), a Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA), a BioTek Epoch microplate spectrophotometer (BioTek, USA), an Illumina NovaSeq 6000 sequencing platform (Illumina, USA), a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, USA), an Agilent Bioanalyzer 4150 (Agilent Technologies, CA, USA), and an ABI StepOnePlus real-time PCR system (Applied Biosystems, USA).

2.3 Measurement of agronomic traits

The agronomic traits measured included tuber fresh weight, tuber length, tuber width, tuber thickness, terminal bud length, basal scar diameter, and annulation number.

After fresh tubers were cleaned and surface moisture was removed, tuber fresh weight was measured using an electronic balance. Tuber length, width, thickness, terminal bud length, and basal scar diameter were measured using a vernier caliper, and annulation number was recorded manually.

2.4 Sample pretreatment and determination of quality-related traits

2.4.1 Sample pretreatment

After agronomic trait measurement, part of each tuber was excised under low-temperature conditions, rapidly frozen in liquid nitrogen, and stored at – 80°C for DNA and RNA extraction and subsequent qPCR expression analysis. The remaining tuber tissue was heat-inactivated at 105°C, dried at 55°C to constant weight, ground, and passed through a No. 3 sieve for the determination of quality-related traits.

2.4.2 Determination of bioactive constituents

Test solutions for gastrodin, *p*-hydroxybenzyl alcohol, and parishins A, B, C and E were prepared according to the General Chapters of the Chinese Pharmacopoeia (2020 edition) and the assay method for *Gastrodia elata* [9]. Chromatographic analysis was performed on an Agilent 1260 HPLC system (Agilent Technologies, USA) equipped with a Waters C18 column (150 mm × 3.9 mm, 4 μm). The mobile phase consisted of acetonitrile (A) and 0.1% phosphoric acid in water (B). The gradient elution program is shown in Table 1. The flow rate was 0.8 mL min⁻¹, the column temperature was 30°C, the detection wavelength was 220 nm, and the injection volume was 5 μL.

Table 1
Gradient elution program used for HPLC analysis.

Time (min)	Mobile phase A (%)	Mobile phase B (%)
0–10	3→10	97→90
10–15	10→12	90→88
15–25	12→18	88→82
25–40	18	82
40–42	18→95	82→5

2.4.3 Determination of physicochemical and edible quality traits

The physicochemical and edible quality traits included moisture content, dry matter ratio, ash content, ethanol-soluble extract content, polysaccharide content, protein content, and fat content.

Moisture content, ash content, and ethanol-soluble extract content were determined according to the corresponding General Chapters of the Chinese Pharmacopoeia (2020 edition). Dry matter ratio was calculated as the ratio of tuber dry weight to fresh weight. Polysaccharide content was determined using the anthrone-sulfuric acid colorimetric method according to General Chap. 0401 of the Chinese Pharmacopoeia. Protein and fat contents were determined according to GB 5009.5–2016 and GB 5009.6–2016, respectively.

Dry matter ratio (%) = tuber dry weight / tuber fresh weight × 100.

It should be noted that moisture content in this study refers to the residual moisture content of dried powder samples rather than the water content of fresh tubers.

2.4.4 Determination of mineral elements

The mineral elements determined included K, Ca, P, Mg, Na, Fe, Mn, B, Zn and Cu. After digestion, the samples were analysed according to GB 5009.268–2016 using an Optima 8100 ICP-OES system (PerkinElmer, USA).

2.5 Whole-genome resequencing and variant detection

Whole-genome resequencing was performed by Novogene Co., Ltd. Genomic DNA was independently extracted three times from the same tuber tissue using the CTAB method. After assessment of DNA integrity and concentration by 1% agarose gel electrophoresis and a Qubit 3.0 fluorometer, three sequencing libraries were constructed for each accession as technical replicates. DNA was fragmented to approximately 350 bp using Covaris ultrasonication, paired-end libraries were prepared following the Illumina standard protocol, and sequencing was performed on the Illumina NovaSeq 6000 platform in PE150 mode.

After quality control, the data from the three technical replicate libraries of each accession were merged and aligned to the *G. elata* reference genome (NGDC GWH accession: GWHBHOU000000000) using BWA with the parameters `mem -t 4 -k 32 -M` [10]. PCR duplicates were removed using Picard v2.27.5. SNPs and InDels were then called using the HaplotypeCaller module in GATK v4.3.0 [11], followed by hard filtering with VariantFiltration. The filtering criteria for SNPs were $QD < 2.0$, $FS > 60.0$, $MQ < 40.0$, $SOR > 3.0$, and $ReadPosRankSum < -8.0$; the filtering criteria for InDels were $QD < 2.0$, $FS > 200.0$, $SOR > 10.0$, and $ReadPosRankSum < -20.0$. Variants were finally annotated for genomic location and function using ANNOVAR [12] based on the reference genome annotation.

2.6 Population genetic analyses

2.6.1 Population structure analysis

The initial variant dataset was filtered using the SNPRelate package [13]. Only biallelic SNPs were retained, with a minor allele frequency (MAF) ≥ 0.03 and a missing rate ≤ 0.25 . Redundancy among strongly linked loci was reduced using a sliding-window approach (`ld.threshold = 0.5`, `slide.max.bp = 500000 bp`, `autosome.only = FALSE`). The resulting independent SNP dataset was used for principal component analysis (PCA), neighbour-joining (NJ) tree construction, and ancestry analysis. Genetic relationships among samples were estimated based on the IBS similarity matrix, and an NJ tree was constructed using 1 – IBS as the genetic distance. ADMIXTURE v1.3.0 [14] was used for ancestry inference with K values ranging from 1 to 6, and cross-validation errors were used to evaluate the most appropriate number of ancestral clusters.

2.6.2 Genetic differentiation and diversity analysis

For AMOVA and population differentiation analyses, SNPs were filtered using PLINK v2.0 [15] with $MAF > 0.05$ and missing rate < 0.2 . Molecular analysis of variance (AMOVA) was performed using Arlequin v3.5 [16] to estimate the proportions of genetic variation among and within cultivated variants. Nucleotide diversity (π) and the fixation index (F_{st}) were calculated using 20-kb sliding windows with a 5-kb step size, and pairwise comparisons were conducted among the three cultivated variants. Relatively highly differentiated regions were further defined using the threshold corresponding to the top 10% of windowed F_{st} values in the comparison between red and black *Gastrodia* and were used for downstream candidate interval analysis.

2.6.3 Linkage disequilibrium analysis

For linkage disequilibrium (LD) decay analysis, a relatively relaxed filtering criterion was applied ($MAF \geq 0.05$ and missing rate ≤ 0.5). The filtered SNP dataset was used to calculate pairwise SNP r^2 values with `snpGdsLDMat` in SNPRelate. LD decay curves were fitted using LOESS with 5-kb bins.

2.7 Exploratory genome-wide association analysis

The phenotypic dataset used for exploratory GWAS consisted of tuber measurements from 28 samples that corresponded one-to-one with the whole-genome resequencing data.

Exploratory GWAS for the 30 phenotypic traits was performed using the FarmCPU model implemented in the MVP package in R [17, 18]. Traits that clearly deviated from normality were analysed after inverse normal transformation. The model included the first three principal components, cultivated variant class, and a kinship matrix as covariates to correct for the effects of population structure and relatedness on association results.

P values were corrected using the Benjamini–Hochberg method, and loci with $Q < 0.05$ were considered significantly associated. Manhattan and QQ plots were generated using the CMplot package. For significant loci, the coefficient of determination (R^2) from single-marker linear regression was calculated to estimate the proportion of phenotypic variance explained (PVE).

2.8 Candidate gene screening and genotype clustering analysis of candidate intervals

Considering the LD decay range and the mapping resolution of the present population, a window extending 500 kb upstream and downstream of each significant association signal was used as the initial interval for candidate gene screening. Annotated genes within these intervals were extracted. BLAST searches against the NR, Swiss-Prot, GO and KEGG databases were used to identify genes related to processes such as secondary metabolism, water transport and carbohydrate metabolism and with potential biological relevance to the target traits. Priority candidate genes were further defined according to the positional relationship between significant loci and candidate genes.

For candidate loci, genotype matrices were extracted after removing missing genotypes. The frequencies of reference homozygotes, heterozygotes and alternative homozygotes were then calculated for each cultivated variant, and genotype frequency plots were generated. Based on the reference genome annotation, candidate gene structures and the distribution of key SNPs within the corresponding intervals were visualized. For variation patterns within candidate gene intervals, SNP genotype matrices were extracted and Manhattan distances among samples were calculated. Hierarchical clustering was then performed using the Ward.D2 method. Phenotypic differences among genotype groups were tested using the Kruskal–Wallis test, followed by pairwise Wilcoxon tests when necessary.

2.9 qPCR expression analysis

Candidate gene expression analysis was conducted in accordance with the MIQE guidelines [19]. Candidate reference genes were evaluated using the geNorm algorithm [20], and Gel02G03143 and Gel18G19311 were selected as internal reference genes. Primer information is listed in Table 2.

The samples used for qPCR were obtained from the same accessions and the same tuber tissues that corresponded one-to-one with the whole-genome resequencing and phenotypic measurements. Samples were ground in liquid nitrogen, and total RNA was extracted using TRIzol® reagent (Magen). RNA quality was assessed by measuring the A260/A280 ratio using a NanoDrop ND-2000 spectrophotometer (Thermo Scientific, USA), determining the RNA integrity number (RIN) using an Agilent Bioanalyzer 4150 (Agilent Technologies, CA, USA), and checking integrity by agarose gel electrophoresis. Samples with OD260/280 values between 1.8 and 2.1 were used for subsequent analyses.

Reverse transcription and qPCR assays were performed by Tsingke Biotechnology Co., Ltd. Reverse transcription was carried out using the SynScript® III RT SuperMix for qPCR kit, with genomic DNA removal prior to cDNA synthesis. The resulting cDNA was diluted three-fold and used as the qPCR template. qPCR amplification was performed using ArtiCanCEO SYBR qPCR Mix on an ABI StepOnePlus real-time PCR system. The total reaction volume was 20 µL, containing 10 µL SYBR qPCR Mix, 0.8 µL of each primer, 1 µL cDNA template, and 7.4 µL ddH₂O. The cycling program was as follows: 95°C for 5 min; 40 cycles of 95°C for 15 s, 60°C for 20 s, and 72°C for 20 s, with fluorescence acquisition during the extension step; a melting curve analysis was performed at the end of amplification.

Relative expression levels were calculated using the 2^{−ΔΔCt} method [21], with red *Gastrodia* sample H-1 used as the calibrator. Three technical replicates were included for each sample, and a coefficient of variation (CV) of technical replicate Cq values below 2% was used as the criterion for acceptable repeatability. Correlation analyses between gene expression and phenotypic traits were based on ΔCt values. Pearson correlation was used for normally distributed data, whereas Spearman rank correlation was used for non-normally distributed data.

Table 2
Primer sequences used for qPCR analysis.

Primer name	Sequence (5'–3')	Product size (bp)
1-Gel02G02702-F	ACGGGCTTTGGGATGACAAT	119
1-Gel02G02702-R	TCACTTGGTTGCTGCTGTGA	
2-Gel14G16652-F	GTCCTTCCCTCTCAGCTTGG	120
2-Gel14G16652-R	GGTAGATGAGCAGCCAACGA	
3-Gel12G14048-F	TGGAGCATCATGTGGACACC	110
3-Gel12G14048-R	ACTTTAGTCAGGGTCACGGC	
4-Gel06G08670-F	TTAGCACCACATCACC GCC	101
4-Gel06G08670-R	GGAAGTGGCATCATCCTCCA	
5-Gel10G12766-F	GTTCCATGCGACGGACTTTG	110
5-Gel10G12766-R	TGCTCTTCCTCTCCTTCCCA	
ref1-Gel02G03143-F	GGCCTCCTAATGCTGATGCT	134
ref1-Gel02G03143-R	AAAGGCAGTGGATAACGGCA	
ref2-Gel18G19311-F	GGGGAGCCTTTCTTCTTGGT	131
ref2-Gel18G19311-R	CGTCCGAGTGACAAAATCGC	

2.10 Statistical analysis

Group comparisons were conducted using tuber measurements corresponding to the 28 source accessions as the statistical units. Because the three cultivated variants had unequal sample sizes (13 red, 6 green and 9 black accessions), all traits were first tested for normality using the Shapiro–Wilk test and for homogeneity of variance using Levene’s test. For traits that met the assumptions of normality and homogeneity of variance, one-way ANOVA followed by Tukey’s HSD test was used. For traits that met the normality assumption but not the homogeneity assumption, Welch’s ANOVA followed by the Games–Howell test was used. For traits that did not meet the normality assumption, the Kruskal–Wallis test followed by Dunn’s test was applied. P values were corrected using the Benjamini–Hochberg method. All statistical analyses and visualizations were performed in R v4.5.1.

3. Results

3.1 Phenotypic variation and differences among cultivated variants in the study population

Among the 28 tuber samples corresponding one-to-one with the resequencing dataset, all 30 phenotypic and quality-related traits showed varying degrees of variation. Traits with relatively high coefficients of variation included parishin A content (52.11%), gastrodin content (48.86%), phosphorus content (40.49%), and potassium content (36.91%), whereas tuber width, fat content, and dry matter ratio showed relatively low coefficients of variation, at 8.88%, 10.20%, and 12.02%, respectively.

After false discovery rate (FDR) correction, 16 traits remained significantly different among the three cultivated variants, including annulation number, moisture content, total ash content, dry matter ratio, ethanol-soluble extract content, gastrodin content, parishin A, B, C, and E contents, potassium content, phosphorus content, sodium content, copper content, polysaccharide content, and fat content (Table S1).

Among the agronomic traits, only annulation number differed significantly among the three cultivated variants, with the highest value in green *Gastrodia* (14.167 ± 2.401), the lowest in red *Gastrodia* (10.308 ± 2.720), and an intermediate value in black *Gastrodia* (12.556 ± 1.878). Although the remaining agronomic traits showed some variation, none remained significant after FDR correction.

Among the physicochemical and processing-related indices, green *Gastrodia* showed relatively high moisture content ($6.053 \pm 0.291\%$) and total ash content ($3.942 \pm 0.325\%$), whereas black *Gastrodia* showed relatively high dry matter ratio ($21.531 \pm 1.269\%$) and ethanol-soluble extract content ($20.888 \pm 2.512\%$).

Among the medicinal quality-related traits, gastrodin content was highest in red *Gastrodia* ($1.483 \pm 0.471\%$), significantly higher than that in black *Gastrodia* ($0.819 \pm 0.364\%$) and green *Gastrodia* ($0.642 \pm 0.106\%$). Parishin A, B, and C contents were also highest in red *Gastrodia*, at $2.963 \pm 0.772\%$, $1.586 \pm 0.325\%$, and $0.636 \pm 0.110\%$, respectively. In contrast, parishin E content was highest in green *Gastrodia* ($1.817 \pm 0.271\%$), followed by black *Gastrodia* ($1.592 \pm 0.404\%$), and lowest in red *Gastrodia* ($1.117 \pm 0.214\%$). No significant difference in *p*-hydroxybenzyl alcohol content was detected among the three cultivated variants.

Among the edible quality-related traits, polysaccharide content was highest in black *Gastrodia* ($0.228 \pm 0.013\%$), followed by green *Gastrodia* ($0.189 \pm 0.006\%$), and lowest in red *Gastrodia* ($0.175 \pm 0.012\%$). Fat content also differed significantly among the three cultivated variants, with the highest value in black *Gastrodia* ($0.040 \pm 0.002\%$), the lowest in green *Gastrodia* ($0.035 \pm 0.001\%$), and an intermediate value in red *Gastrodia* ($0.039 \pm 0.005\%$). Protein content did not differ significantly among the three cultivated variants.

For mineral elements, green *Gastrodia* generally showed relatively high potassium, phosphorus, and copper contents. Potassium content reached 13.001 ± 0.458 g/kg in green *Gastrodia*, significantly higher than that in red *Gastrodia* (6.588 ± 0.997 g/kg) and black *Gastrodia* (6.266 ± 1.209 g/kg). Phosphorus content was also highest in green *Gastrodia* (2.317 ± 0.650 g/kg) and lowest in red *Gastrodia* (1.102 ± 0.237 g/kg). Sodium content was highest in black *Gastrodia* (1.811 ± 0.326 g/kg) and lowest in red *Gastrodia* (1.363 ± 0.187 g/kg). Although the remaining mineral elements showed some numerical differences, none remained significant after FDR correction.

Overall, among the 28 samples corresponding to the omics analyses, most agronomic traits did not differ significantly among the three cultivated variants, whereas several quality-related traits still showed clear inter-variant differences. Specifically, red *Gastrodia* generally showed higher levels of gastrodin and parishins A, B, and C; black *Gastrodia* showed relatively high dry matter ratio and polysaccharide content; and green *Gastrodia* showed relatively high parishin E as well as potassium, phosphorus, and copper contents.

1. (a) Geographic distribution of the 28 sampled *Gastrodia* accessions.
2. (b) Representative mature tuber morphology of the three cultivated variants, red, green, and black *Gastrodia*.
3. (c) Raincloud plot of moisture content in the three cultivated variants.
4. (d) Raincloud plot of gastrodin content in the three cultivated variants.
5. (e) Raincloud plot of polysaccharide content in the three cultivated variants.

Note

In (c–e), dots indicate individual sample observations, boxes represent the interquartile range, the central line indicates the median, and the cloud-shaped distribution indicates data density.

3.2 Whole-genome resequencing and variant detection results

A total of 364 Gb of clean data was generated from the 28 *Gastrodia* samples, with 11.19–18.39 Gb of clean data per sample. Sequencing quality was high across all samples, with Q20 values above 99.25% and Q30 values above 96.47%. For most samples, the mapping rate of clean reads to the *G. elata* reference genome ranged from 94% to 99%, whereas two samples showed mapping rates of approximately 86%. The average sequencing depth across the population was approximately 11×, with a range of 8.84×–14.94× per sample.

Across the whole genome, 2.486 million high-quality SNPs and 0.360 million InDels were identified. Functional annotation of SNPs showed that most were located in intergenic regions (44.8%) and intronic regions (20.4%), followed by upstream/downstream regulatory regions (3.8%) and exonic regions (0.3%). Among exonic SNPs, 3,327 nonsynonymous variants, 58 stop-gain variants, and 14 stop-loss variants were detected.

3.3 Population genetic structure, diversity, and differentiation

Population structure analyses indicated genome-wide differentiation among the three cultivated variants of *Gastrodia*. In the NJ tree, red *Gastrodia* samples clustered into a relatively distinct branch, whereas green and black *Gastrodia* still showed partial overlap. The PCA results were consistent with the NJ tree. The first principal component (PC1) explained 23.90% of the variation and separated red *Gastrodia* from green and black *Gastrodia*, whereas the second principal component (PC2) explained 8.84% of the variation and further separated green and black *Gastrodia* along this axis. ADMIXTURE analysis showed that at $K = 3$, red and green *Gastrodia* were each dominated by their own ancestral components, whereas black *Gastrodia* displayed a relatively more complex internal ancestry composition. AMOVA further showed that 24.37% of the total genetic variation occurred among cultivated variants ($F_{st} = 0.2437$, $P < 0.001$).

The three cultivated variants also differed in genetic diversity. The genome-wide average nucleotide diversity (π) of red *Gastrodia* was 3.07×10^{-6} , lower than that of green *Gastrodia* (5.07×10^{-6}) and black *Gastrodia* (8.21×10^{-6}). LD decay curves showed that r^2 decreased more rapidly with increasing physical distance in green *Gastrodia*, more slowly in red *Gastrodia*, and relatively slowly in black *Gastrodia*. Within the range evaluated in this study, the LD decay distance was approximately 120 kb in green *Gastrodia* and approximately 380 kb in red *Gastrodia*, whereas in black *Gastrodia*, r^2 remained 0.738 at 500 kb. Genome-wide heterozygosity across the three cultivated variant populations ranged from 3.7% to 9.2%.

Genome-wide windowed F_{st} distributions from pairwise comparisons among the three cultivated variants indicated different levels of differentiation among comparison groups. In the comparison between red and black *Gastrodia*, relatively highly differentiated regions were defined using the threshold corresponding to the top 10% of the windowed F_{st} distribution ($F_{st} > 0.0476$), resulting in a total of 1,679 relatively highly differentiated regions.

1. (a) NJ tree.
2. (b) Scatter plot of the first two principal components from PCA.
3. (c) Ancestry composition inferred by ADMIXTURE at $K = 3$.
4. (d) Cross-validation errors for $K = 1-6$.
5. (e) Genome-wide LD decay curves of red, green, and black *Gastrodia*.
6. (f) Genome-wide sliding-window distributions of nucleotide diversity (π) in the three cultivated variants.
7. (g) Genome-wide windowed F_{st} distributions from pairwise comparisons among the three cultivated variants.

3.4 Exploratory GWAS of agronomic and quality-related traits

Exploratory association analysis identified 225 SNPs that reached the significance threshold. Association signals were mainly concentrated in loci related to polysaccharide content, moisture content, gastrodin content, as well as some mineral elements and agronomic traits.

For polysaccharide content, a relatively concentrated association signal was detected in the 45.47–45.52 Mb region of scaffold GWHBHOU00000005. Additional association signals for moisture content were also detected in this region. The representative SNP GWHBHOU00000005-45493074 was located within the gene region of Gel14G16652. Single-marker regression indicated that this locus explained 52.59% of the phenotypic variance in polysaccharide content.

Significant association signals for gastrodin content were mainly located in the 72.59–72.63 Mb region of scaffold GWHBHOU00000011. The representative SNP GWHBHOU00000011-72631073 fell within the gene region of Gel02G02702.

Significant association signals for moisture content were mainly distributed in the 51.76–51.80 Mb region of scaffold GWHBHOU00000015. The representative SNP GWHBHOU00000015-51795705 was located downstream of Gel06G08670 and explained approximately 3.23% of the phenotypic variance.

In addition, a co-localized association signal for tuber fresh weight and phosphorus content was detected near 22.95 Mb on scaffold GWHBHOU00000002, with Gel02G13386 as the neighboring gene.

3.5 Candidate gene screening and expression analysis

By integrating association signal positions, functional annotation, genotype classification, and expression analysis, three candidate genes were identified: Gel14G16652, Gel02G02702, and Gel06G08670.

Genotype analysis showed that these candidate regions differed in genotype distribution among the cultivated variants. For the polysaccharide-associated SNP GWHBHOU00000005-45493074, all red *Gastrodia* samples were alternative homozygotes (Alt 1/1), whereas all green and black *Gastrodia* samples were heterozygotes (Het 0/1). For the gastrodin-associated SNP GWHBHOU00000011-72631073, polymorphism was observed in green and black *Gastrodia*, whereas in red *Gastrodia* the raw resequencing coverage at this locus was unstable, leading to some uncertainty in genotype assignment.

Genotype clustering within candidate gene intervals showed that different genotype groups of Gel14G16652 and Gel06G08670 differed significantly in polysaccharide content and moisture content ($P < 0.05$). For Gel06G08670, some variants within the candidate interval showed grouping differences among cultivated variants, and several loci were close to fixation in particular variants.

The Cq coefficients of variation of qPCR technical replicates were all below 2%. Gel14G16652 showed overall higher relative expression in black and green *Gastrodia* than in red *Gastrodia*, and its transcript level was significantly positively correlated with polysaccharide content ($r = 0.703$, $P < 0.001$). No significant

correlation was detected between Gel02G02702 expression and gastrodin content in dormant tubers. Gel06G08670 showed some variation in relative expression among cultivated variants, but its expression was not significantly correlated with moisture content.

Table 4 Association information and functional annotation of candidate genes.					
Gene	Associated trait(s)	Representative SNP	Best P value	Positional relationship to representative SNP	Functional annotation
Gel14G16652	Polysaccharide content, moisture content	GWHBHOU00000005-45493074	9.82898 × 10 ⁻⁶	Located within the gene region	Palmitoyl-protein thioesterase 1 (PPT1)
Gel02G02702	Gastrodin content	GWHBHOU00000011-72631073	1.12517 × 10 ⁻⁷	Located within the gene region	RING-type zinc finger E3 ubiquitin ligase
Gel06G08670	Moisture content	GWHBHOU00000015-51795705	9.82898 × 10 ⁻⁶	Located in the downstream flanking region	Transcription elongation factor 1 homolog (TEF-1/TFIIS)
Note: “Associated trait(s)” refers to the trait(s) associated with the genomic interval containing the candidate gene in the exploratory GWAS. “Representative SNP” refers to the SNP with the strongest association signal within the candidate gene-related interval. “Best P value” refers to the association P value of the representative SNP.					

- (a, b) Manhattan and QQ plots for gastrodin content.
- (c, d) Manhattan and QQ plots for polysaccharide content.
- (e, f) Manhattan and QQ plots for moisture content.
- (g–i) Gene structures of Gel02G02702, Gel06G08670, and Gel14G16652, together with the distribution of key SNPs within the corresponding intervals.
- (j–l) Relationships between genotype groups within candidate gene intervals and the corresponding phenotypic traits.
1. (m) Relative expression levels of the three candidate genes in the different cultivated variants.
 2. (n) Genotype distributions of the representative associated SNPs in red, green, and black *Gastrodia*.

Note

In (a–f), horizontal red lines indicate the significance threshold, and highlighted coloured points indicate significant association signals. In (g–i), dark boxes indicate exons and lines indicate introns. In (j–l),

different lowercase letters above the boxplots indicate significant differences ($P < 0.05$).

4. Discussion

Within the scope of the present materials, red, green and black *Gastrodia* showed differences at both the phenotypic and genome-wide levels. Population structure analyses showed that the three cultivated variants were not completely admixed. Red *Gastrodia* was relatively clearly separated from the other two variants, whereas green and black *Gastrodia* still showed a certain degree of overlap. These results suggest that the three cultivated variants have already developed recognizable genetic differentiation within the current material set, although this differentiation has not reached the level of complete independence. In particular, the genetic heterogeneity observed within black *Gastrodia* suggests that its population composition may be relatively complex. At the current stage, it may be more appropriate to regard black *Gastrodia* as a cultivated group with a relatively heterogeneous internal genetic background; whether it constitutes a single and relatively stable genetic unit remains to be tested in a broader set of materials.

The results for nucleotide diversity and linkage disequilibrium further suggest that the population backgrounds of the three cultivated variants are not identical. The lower π value and relatively slower LD decay observed in red *Gastrodia* are broadly consistent with the tendency toward reduced genetic diversity that is often observed during crop domestication and continued selection [8]. Considering the cultivation history and production use of *G. elata*, the lower genetic diversity of red *Gastrodia* may be related to the continued use and repeated selection of a limited germplasm base during long-term cultivation. By contrast, black *Gastrodia* showed a higher level of within-group variation and a relatively more complex internal genetic composition. This pattern may be associated with differences in material origin, germplasm maintenance, or selection intensity. However, because the materials analysed in this study were mainly collected from production areas in Guizhou Province, whether these differences are representative across a broader geographic range still requires verification in larger populations.

A total of 1,679 relatively highly differentiated regions were identified between red and black *Gastrodia*, suggesting that differentiation between the two variants is not evenly distributed across the genome but is concentrated in specific intervals. In the context of the present study, the main value of these regions lies in providing interval-level clues for subsequent screening of candidate loci and candidate genes. Their correspondence with specific trait formation still requires further evaluation in combination with association results and functional validation.

For materials with relatively small sample sizes and pronounced population structure, the interpretation of GWAS results needs to take both detection power and residual effects of population structure into account [22]. At the same time, effect-size estimates for significant loci may also be influenced by factors such as Winner's curse [23]. Under the current dataset, a total of 225 loci reached the significance threshold, and the main association signals were concentrated in quality-related traits such as polysaccharide content, moisture content and gastrodin content. These results were broadly consistent with the directions of inter-variant differences observed in the phenotypic analyses, suggesting that these traits have a certain genetic differentiation basis in the present population.

In terms of phenotypic variance explained, different traits appear to have different underlying genetic architectures. The relatively low PVE of the moisture-associated loci is more consistent with a scenario involving multi-locus effects, whereas the polysaccharide-associated locus GWHBHOU00000005-45493074 showed a PVE of 52.59%, suggesting that this interval may have a relatively large genetic effect in the present material set. However, this value still needs to be interpreted cautiously in light of the sample size. At the current stage, it is more appropriate to regard this interval as a priority candidate region rather than directly treat it as a stable major-effect locus. It should also be noted that, although principal components, cultivated variant class and a kinship matrix were included in the model to reduce the influence of population structure, some association signals may still reflect both phenotypic differences and background population differentiation, given that the three cultivated variants are themselves genetically differentiated. Therefore, the GWAS results reported here are better regarded as a basis for screening candidate regions and guiding follow-up validation than as broadly generalizable genetic rules for quality-related traits in *G. elata*. For this reason, retaining the qualifier “exploratory GWAS” throughout the manuscript is necessary.

By integrating association signal positions, genotype distributions and expression results, we focused on three candidate genes, namely Gel14G16652, Gel02G02702 and Gel06G08670. The strength of evidence supporting these genes is not identical and should therefore be considered separately.

Gel14G16652 corresponds to an association interval for polysaccharide content and moisture content. The representative SNP is located within the gene region, and genotype differences were observed among the cultivated variants. In addition, Gel14G16652 showed overall higher expression in black and green *Gastrodia* than in red *Gastrodia*, and its transcript level was significantly positively correlated with polysaccharide content. PPT1-like/de-S-acylation-related enzymes participate in reversible S-acylation processes, and S-acylation is closely associated with membrane protein localization, signal transduction and plant stress responses [24]. Taken together, the mapping result, genotype pattern and expression trend suggest that Gel14G16652 is one of the candidate genes currently supported by relatively more concentrated evidence. However, the available evidence still remains at the level of genetic and expression associations, and the specific relationship between this gene and polysaccharide accumulation in *G. elata* tubers requires further verification.

Gel02G02702 is associated with gastrodin content, and its representative SNP is located within the gene region. This gene encodes a RING-type zinc finger E3 ubiquitin ligase. Proteins of this class are widely involved in ubiquitination-mediated protein homeostasis and are associated with plant development, stress responses and some metabolic processes [25]. In the present study, the interval containing Gel02G02702 showed a relatively clear association signal, but its expression level did not show a significant correspondence with gastrodin content, and stable sequencing coverage at this locus was lacking in red *Gastrodia*. Therefore, the interval may be related to variation in gastrodin content, but its mode of action, and whether it involves post-transcriptional regulation or protein homeostasis, cannot yet be determined. Given that the expression data were obtained from a single time point during tuber dormancy, this gene is better regarded as a candidate requiring further follow-up.

Gel06G08670 is located near a moisture-associated signal, and the representative SNP lies in the downstream flanking region of the gene. Genotype clustering within the candidate interval showed differences among groups in both polysaccharide and moisture contents. This gene encodes a transcription elongation factor 1 homolog (TEF-1/TFIIS). Factors of the TFIIS class participate in RNA polymerase II-mediated transcription elongation and have been linked to transcriptional reprogramming under stress conditions [26]. In the context of the present results, Gel06G08670 is better regarded as a neighbouring candidate gene. That is, the interval in which it is located is associated with the target trait, but the representative SNP does not fall directly within the gene region. Therefore, the relationship between this gene and variation in moisture content still needs to be clarified through finer interval dissection and functional validation.

Overall, among the three candidate genes identified in this study, the currently available evidence is relatively more concentrated for Gel14G16652, whereas Gel02G02702 and Gel06G08670 remain at the candidate stage. Further confirmation of their roles will require validation of locus stability in larger populations, together with long-read sequencing, fine mapping and direct functional experiments.

This study has several limitations.

First, the sample size was relatively small, and most materials were collected from production areas in Guizhou Province, which limits population representativeness. The significant loci and candidate genes identified here are therefore better regarded as leads for further validation, and their broader applicability remains to be tested in larger populations with more diverse origins.

Second, the analysed materials were obtained through a uniform propagation procedure and are not completely identical to the original cultivated materials maintained through long-term clonal propagation in production systems. Therefore, caution is needed when extrapolating the present results to cultivated populations maintained under production conditions.

Third, the current evidence for candidate genes is mainly based on association mapping, genotype distribution and expression analysis, and direct functional evidence is still lacking. In particular, the expression data only reflect transcriptional states in a specific tissue and at a specific developmental stage, which limits inference regarding candidate gene function. Fourth, read alignment and variant detection in this study were performed against reference genome version GWHBHOU000000000. Although a chromosome-scale genome of *G. elata* has been published [7], the present analyses were not re-anchored to that assembly version. As a result, the identification of local structural variation and the fine dissection of associated intervals may still be constrained by the resolution of the reference version and by differences in annotation systems.

Further improvements in interval-level interpretation may be achieved through the use of higher-quality reference genomes or pan-genome resources.

5. Conclusion

In this study, 28 accessions representing three cultivated variants of *Gastrodia elata* (red, green, and black *Gastrodia*) were analysed using whole-genome resequencing and multi-trait phenotypic data to compare phenotypic differentiation, population genetic background, and quality-related candidate genomic regions and candidate genes.

The results showed that the three cultivated variants exhibited differentiation at both the phenotypic and genome-wide levels. Red *Gastrodia* showed lower nucleotide diversity and relatively slower LD decay, whereas black *Gastrodia* displayed a relatively more complex internal population composition. In the comparison between red and black *Gastrodia*, a total of 1,679 relatively highly differentiated regions were identified. Exploratory GWAS detected 225 loci that reached the significance threshold, with relatively concentrated association signals for polysaccharide content, moisture content, and gastrodin content.

By integrating association signal positions, genotype distributions, and expression results, Gel14G16652, Gel02G02702, and Gel06G08670 were identified as candidate genes for further validation. Among them, the currently available evidence is relatively more concentrated for Gel14G16652, whereas Gel02G02702 and Gel06G08670 still require further evaluation in larger populations and through more direct functional experiments.

Overall, this study provides a population-level data basis for comparing genetic differentiation among three cultivated variants of *G. elata* and for screening quality-related candidate regions, and it also provides a basis for subsequent candidate gene validation.

Declarations

Competing interests

The authors declare no competing interests.

Author Contribution

R.Y. and F.-L.L. conceived and designed the study. R.Y. performed the experiments, analysed the data, and drafted the manuscript. P.W. provided the *Gastrodia elata* germplasm resources used in this study. F.-L.L. critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors gratefully acknowledge the support of the College of Agriculture, Guizhou University, and the Guizhou Key Laboratory for Germplasm Innovation and Resource-Efficient Utilization of Dao-di Herbs.

Data Availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive [27] at the National Genomics Data Center [28], China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number CRA040483, and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. All other data supporting the findings of this study are included in this published article and its Supplementary Information files.

References

1. Zhan, H. D. et al. The rhizome of *Gastrodia elata* Blume—An ethnopharmacological review. *J. Ethnopharmacol.* **189**, 361–385 (2016).
2. Gong, M. Q. et al. Traditional uses, phytochemistry, pharmacology, applications, and quality control of *Gastrodia elata* Blume: a comprehensive review. *J. Ethnopharmacol.* **319**, 117128 (2024).
3. Ji, N. et al. Comparison on bioactivities and characteristics of polysaccharides from four varieties of *Gastrodia elata* Blume. *Front. Chem.* **10**, 956724 (2022).
4. Chen, Y. Y. et al. Genetic variation in cultivated populations of *Gastrodia elata*, a medicinal plant from central China analyzed by microsatellites. *Genet. Resour. Crop Evol.* **61**, 1523–1532 (2014).
5. Mei, Y. et al. Genetic diversity and construction of a core collection in Chinese *Gastrodia elata* germplasm resources. *Sci. Hortic.* **337**, 113595 (2024).
6. Yuan, Y. et al. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* **9**, 1615 (2018).
7. Xu, Y. et al. A chromosome-scale *Gastrodia elata* genome and large-scale comparative genomic analysis indicate convergent evolution by gene loss in mycoheterotrophic and parasitic plants. *Plant. J.* **108**, 1609–1623 (2021).
8. McKey, D. et al. The evolutionary ecology of clonally propagated domesticated plants. *New. Phytol.* **186**, 318–332 (2010).
9. Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China* 163–166 (China Medical Science, 2020).
10. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
11. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
12. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
13. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
14. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
15. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

16. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
17. Yin, L. et al. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteom. Bioinf.* **19**, 619–628 (2021).
18. Liu, X. et al. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* **12**, e1005767 (2016).
19. Bustin, S. A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
20. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034 (2002).
21. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
22. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Primers.* **1**, 59 (2021).
23. Forde, A., Hemani, G. & Ferguson, J. Review and further developments in statistical corrections for Winner's Curse in genetic association studies. *PLoS Genet.* **19**, e1010546 (2023).
24. Li, J., Zhang, M. & Zhou, L. Protein S-acyltransferases and acyl protein thioesterases, regulation executors of protein S-acylation in plants. *Front. Plant. Sci.* **13**, 956231 (2022).
25. Shu, K. & Yang, W. E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses. *Plant. Cell. Physiol.* **58**, 1461–1476 (2017).
26. Szádeczky-Kardoss, I. et al. Elongation factor TFII5 is essential for heat stress adaptation in plants. *Nucleic Acids Res.* **50**, 1927–1950 (2022).
27. Zhang, S. et al. The GSA family in 2025: a broadened sharing platform for multi-omics and multimodal data. *Genomics Proteom. Bioinf.* **23**, qzaf072 (2025).
28. CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center. China National Center for Bioinformation in 2025. *Nucleic Acids Res.* **53**, D30–D44 (2025).

Figures

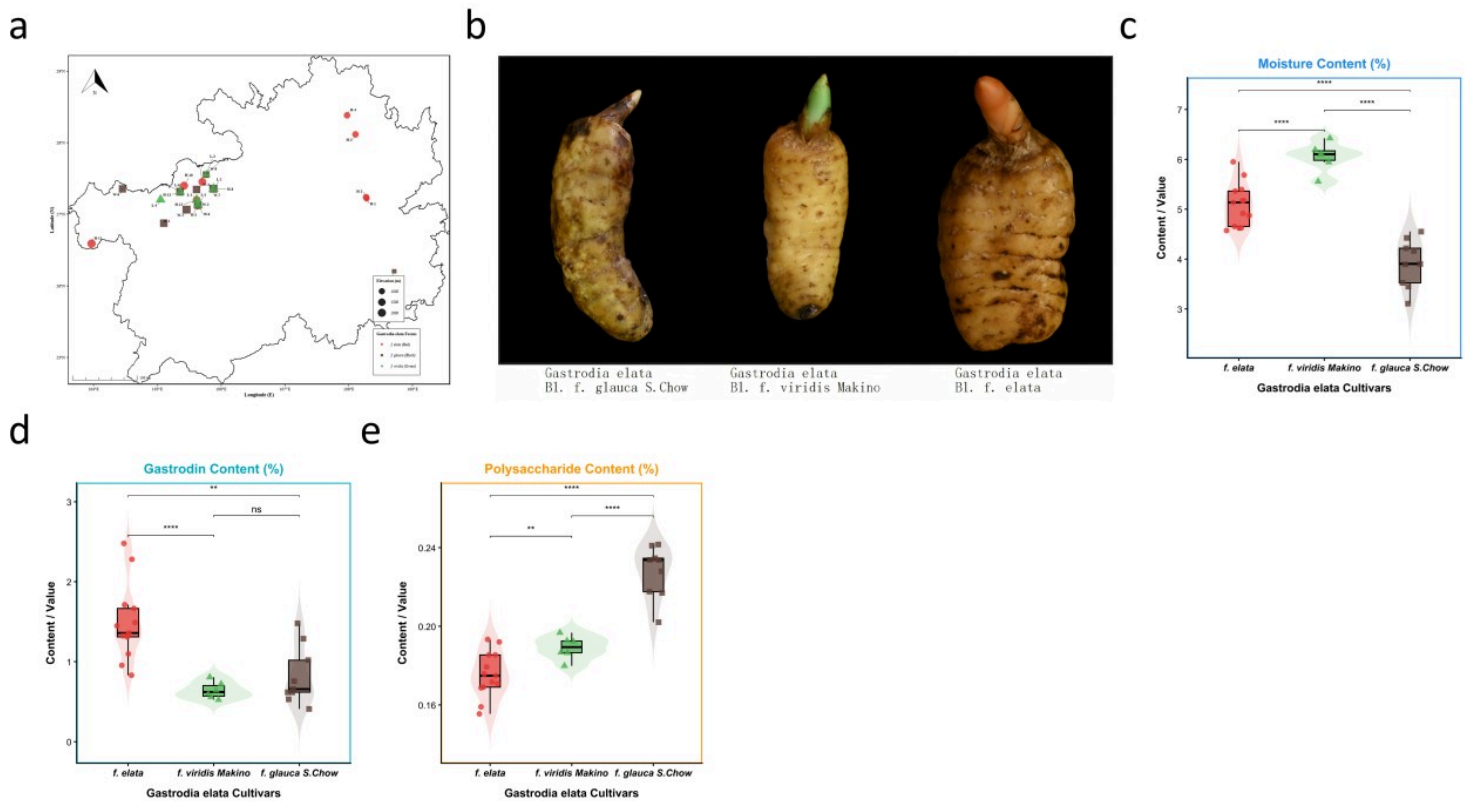


Figure 1

Sampling origins and phenotypic variation of the 28 *Gastrodia* accessions.

- (a) Geographic distribution of the 28 sampled *Gastrodia* accessions.
- (b) Representative mature tuber morphology of the three cultivated variants, red, green, and black *Gastrodia*.
- (c) Raincloud plot of moisture content in the three cultivated variants.
- (d) Raincloud plot of gastrodin content in the three cultivated variants.
- (e) Raincloud plot of polysaccharide content in the three cultivated variants.

Note: In (c–e), dots indicate individual sample observations, boxes represent the interquartile range, the central line indicates the median, and the cloud-shaped distribution indicates data density.

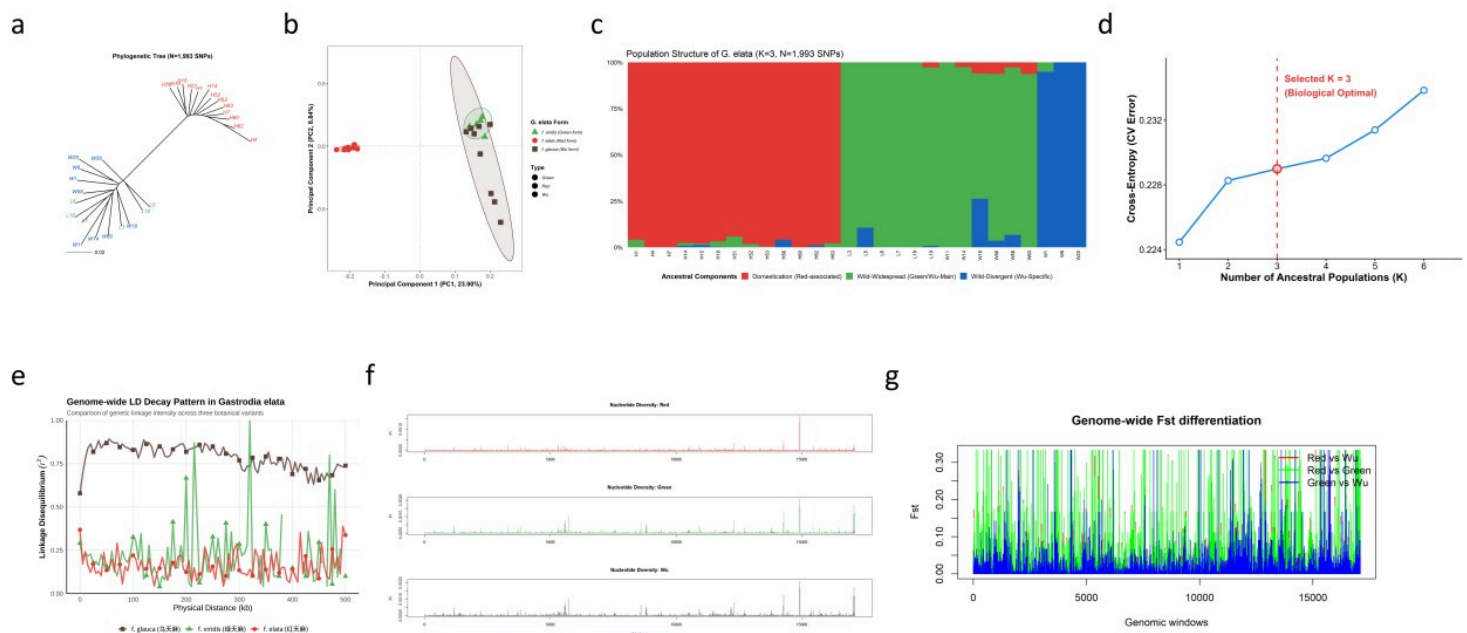


Figure 2

Genetic structure, diversity, and differentiation of the study population.

- NJ tree.
- Scatter plot of the first two principal components from PCA.
- Ancestry composition inferred by ADMIXTURE at K = 3.
- Cross-validation errors for K = 1–6.
- Genome-wide LD decay curves of red, green, and black *Gastrodia*.
- Genome-wide sliding-window distributions of nucleotide diversity (π) in the three cultivated variants.
- Genome-wide windowed F_{st} distributions from pairwise comparisons among the three cultivated variants.

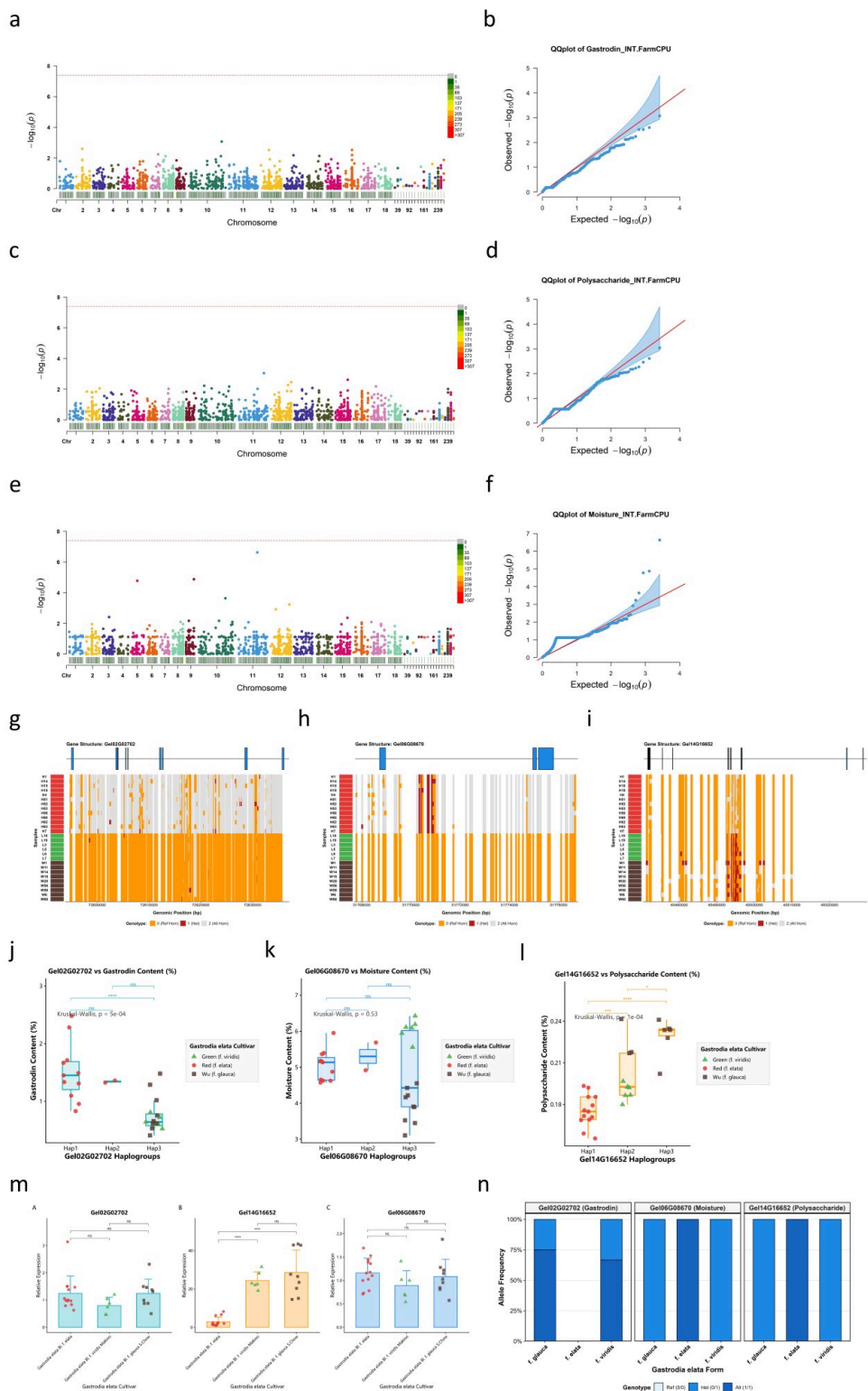


Figure 3

Exploratory GWAS results for quality-related traits, structures of candidate gene intervals, and expression analysis.

(a, b) Manhattan and QQ plots for gastrodin content.

(c, d) Manhattan and QQ plots for polysaccharide content.

(e, f) Manhattan and QQ plots for moisture content.

(g–i) Gene structures of Gel02G02702, Gel06G08670, and Gel14G16652, together with the distribution of key SNPs within the corresponding intervals.

(j–l) Relationships between genotype groups within candidate gene intervals and the corresponding phenotypic traits.

(m) Relative expression levels of the three candidate genes in the different cultivated variants.

(n) Genotype distributions of the representative associated SNPs in red, green, and black *Gastrodia*.

Note: In (a–f), horizontal red lines indicate the significance threshold, and highlighted coloured points indicate significant association signals. In (g–i), dark boxes indicate exons and lines indicate introns. In (j–l), different lowercase letters above the boxplots indicate significant differences ($P < 0.05$).

Supplementary Files

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

- [TableS1HPLCgradientelutionconditions.xlsx](#)
- [TableS2informationfor28accessions.xlsx](#)
- [TableS3raworsummarizeddatafor30phenotypictraits.xlsx](#)
- [TableS4sequencingqualitystatistics.xlsx](#)
- [TableS5SNPandinDelstatisticsbeforeandafterfiltering.xlsx](#)
- [StatisticalAnalysisResultsofExperimentalData.xlsx](#)
- [TableS8completelistof225significantassociationloci.xlsx](#)
- [TableS9annotatedgeneswithincandidateintervalsofthreecoreassociationsignals.xlsx](#)
- [TableS10qPCRprimerandreferenceinformation.xlsx](#)
- [SupplementaryMaterial.rar](#)