

RESEARCH ARTICLE

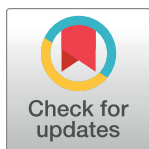
Genome-wide association study of the loci and candidate genes associated with agronomic traits in *Amomum villosum* Lour

Wenxiu Li¹✉, Ping Luo¹✉, Yunfeng Shi², Hualin Zhang¹, Qing Yan¹, Yana Ye¹, Yanli Yao³, Junjun He³*

1 Chinese Academy of Tropical Agricultural Sciences/Zhanjiang Rubber Forest Economic Engineering Technology Research Center, Zhanjiang Experimental Station, Zhanjiang, Guangdong, China, **2** Huazhong Agricultural University, Wuhan, Hubei, China, **3** Chinese Academy of Tropical Agricultural Science/Key Laboratory of Tropical Fruit Biology, Ministry of Agriculture & Rural Affairs, Key Laboratory of Hainan Province for Postharvest Physiology and Technology of Tropical Horticultural Products, South Subtropical Crop Research Institute, Zhanjiang, Guangdong, China

✉ These authors contributed equally to this work.

* hbj46@163.com



OPEN ACCESS

Citation: Li W, Luo P, Shi Y, Zhang H, Yan Q, Ye Y, et al. (2024) Genome-wide association study of the loci and candidate genes associated with agronomic traits in *Amomum villosum* Lour. PLoS ONE 19(8): e0306806. <https://doi.org/10.1371/journal.pone.0306806>

Editor: Shailender Kumar Verma, University of Delhi, INDIA

Received: January 11, 2024

Accepted: June 24, 2024

Published: August 5, 2024

Copyright: © 2024 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: Our study received funding for the operation and maintenance of the germplasm resource garden for southern traditional Chinese medicine *Amomum villosum* Lour (grant number 2023-MBH-00-023, awarded to JH), through the Open Competition Program of Top Ten Critical Priorities of Agricultural Science and Technology Innovation for the 14th Five-Year Plan of

Abstract

Amomum villosum Lour. (*A. villosum*) is a valuable herbaceous plant that produces the famous traditional Chinese medicine Amori Fructus. Identifying molecular markers associated with the growth of *A. villosum* can facilitate molecular marker-assisted breeding of the plant. This study employed 75 *A. villosum* accessions as the test material and utilized 71 pairs of polymorphic simple sequence repeat (SSR) molecular markers to genotype the population. The study analyzed the association between SSR markers and phenotypic traits through the linkage imbalance and population structure analysis. Candidate genes associated with the molecular markers were also identified. The results showed that the phenotypic diversity index range of the 12 agronomic traits was 4.081–4.312 and conformed to a normal distribution. Moreover, 293 allelic variations were detected in the 75 accessions, with an average of 5.32 amplified alleles per loci, ranging from 3 to 8. The maximum number of amplified alleles for AVL12 was 8. The population structure and cluster analysis indicated that the accessions could be divided into two subgroups. Using the mixed linear model (MLM) model of population structure (Q)+kinship matrix (K) for association analysis, three SSR molecular markers significantly associated with the agronomic traits were detected. Fluorescence quantification was used to analyze the expression levels of six candidate genes, and it was found that three of the genes were differentially expressed in phenotypically different accessions. This study is the first to use SSR markers for genome-wide association study (GWAS) mapping and identification of the associated agronomic traits in *A. villosum*. The results of this study provide a basis for identifying genetic markers for growth traits for marker-assisted breeding in *A. villosum*.

Guangdong Province (grant number 2022SDZG07, awarded to JH), and from the earmarked fund for CARS 33 (grant number CARS-33-GD2, awarded to PL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Wurfbainia villosa ($2n = 48$), also known as *Amomum villosum* Lour., is a perennial plant belonging to the monophyletic genus *Wurfbainia* of the family Zingiberaceae [1]. *Amomi Fructus*, the dry and mature fruit of *A. villosum*, is one of the famous Chinese traditional medicines [2], with the potential to prevent miscarriage, warm the spleen and eliminate dampness [3]. *A. villosum* has been cultivated for over 1300 years in China [4] and currently has many accessions from different regions with significant phenotypic differences [5]. Evaluating these accessions and their genetic background is beneficial for improving the utilization rate, selection efficiency, and selective breeding of *A. villosum* [6]. Additionally, studying the correspondence between phenotypes and genotypes is crucial for marker-assisted breeding (MAS) [7].

Genetic diversity research is important for accession evaluation, core germplasm bank construction, and hybrid breeding of new varieties [8]. The most direct method for evaluating genetic diversity in plants involves morphological observation, but morphological observation is easily influenced by the environment. Therefore, genetic diversity evaluation often combines morphological observation with molecular marker technology to evaluate germplasm resources. Simple sequence repeats (SSRs) are common and important molecular markers with high polymorphism and repeatability [9]. SSRs have been widely used in genetic diversity [10] and association analysis-related research [11]. Genetic diversity analysis makes it possible to understand the genetic distance and variation between different accessions [12]. Genetic diversity analysis also helps eliminate accessions with high similarity based on cluster analysis results to construct a core germplasm bank while preserving as much genetic information as possible within the population [13]. In addition, it is necessary to select distant accessions or germplasm for hybridization to fully utilize heterosis in breeding. The genetic relationships between different accessions can be understood [14], and suitable parental materials can be easily selected for breeding [15] through genetic diversity analysis, thereby reducing the selection of parental combinations with similar genetic information and the breeding workload. Thus, genetic diversity analysis plays an important role in the selection of parental material for combinations [12]. Several studies have shown that the accessions of *A. villosum* are abundant and have significant differences in their main chemical components, content, and quality in different growing areas [5, 16, 17]. However, the number of collected and studied *A. villosum* accessions is relatively limited (currently up to 29) [17], and there is no detailed evaluation and identification of their growth traits.

Association analysis is an effective method for identifying genetic markers associated with quantitative traits [18]. Genome-wide association study (GWAS) is suitable for detecting the relationship between molecular markers or candidate genes and target traits in natural populations based on linkage disequilibrium (LD) [19]. GWAS can directly detect the relationship between phenotypes and genotypes in large populations without the need to construct hypotheses between phenotypes and genes, thus reducing the analysis cost at the genomic level for large samples [20]. The technique was first applied in plants in 2001 when Hansen et al. [19] conducted research on Sea beet. Several researchers subsequently utilized SSRs in plants, such as maize [21, 22] and wheat [11], to identify the molecular markers closely linked to the target traits. Although GWAS has been widely used in plant research, there are no reports on its usage on *A. villosum*. Studies have shown that growth traits (such as leaf width, ligule length, etc.) of *A. villosum* are closely related to the number of capsules [23]. Therefore, identifying genetic markers associated with these growth traits would provide a reference basis for breeding high-yielding *A. villosum* varieties.

This study employed 75 accessions of *A. villosum* Lour. (which are currently the most reported) as the test material and utilized 71 pairs of polymorphic SSR molecular markers to

genotype the population. The study analyzed the association between SSR markers and phenotypic traits through the linkage imbalance and population structure analysis. The candidate genes were selected based on linkage markers, and fluorescence quantitative expression pattern analysis was conducted, laying the foundation for genetic research and molecular marker-assisted breeding of complex quantitative traits in *A. villosum*.

Materials and methods

Plant material

A natural population of 75 *A. villosum* accessions were collected from Guangdong (31 samples), Guangxi (20 samples), Yunnan (22 samples), Fujian (1 samples) and Hainan (1 samples) provinces in China (S1 File). After collection, the accessions were intercropped in shaded areas under the rubber forest. Due to the asexual reproduction ability of *A. villosum*, many new shoots were produced in each material after planting for a period of time. Ten seedlings with relatively consistent sizes were randomly selected from each material for planting. The experiment was conducted in a greenhouse, each plot measuring 2 × 5 meters and containing 10 plants. The experimental site was located in Zhanjiang City, with a geographical location of 110° a geographical location of and an elevation of about 21 meters. The region had a tropical northern monsoon climate, with an annual average temperature of 23°C, precipitation of 1200 mm–1800 mm, sunshine hours of 1700 h–2100 h, annual average total radiation of 175 W/m², and average light intensity of about 897Lx. The soil type was laterite.

The germplasm utilized in this study was grown in the germplasm resource garden situated at No. 1 Huxiu Road, Guangdong Province, China. The garden is affiliated with the Zhanjiang Experimental Station of the Chinese Academy of Tropical Agricultural Sciences. We obtained the permissions (verbal consent) and can confirm no conflicting interests.

Phenotyping

After one year of planting, five plants were randomly selected from each test plot for investigation, according to the "*Amomum villosum* Lour." section of the "Specification for the Description of Germplasm Resources of Southern Medicines" [24] and the Chinese agricultural industry standard "Guidelines for Plant Variety Specificity (Distinguishability), Consistency, and Stability Testing *Amomum villosum* Lour." [25]. Twelve agronomic traits were investigated, including the plant height (PH), stem diameter (SD), number of blades (NB), number of ramets (NR), leaf length (LL), leaf width (LW), ligule length (LEL), total length of stolons (TLS), stolon thickness (ST), stolon internode length (SIL), number of new shoots (NNS), and total number of ramets (TNR). PH, LL, LW, LEL, TLS, and SIL were measured using a ruler, while LL, LW, and LEL were measured using the third leaf blade from the top. SD and ST were measured at the breast height using vernier calipers. The average value of each trait was used for subsequent analysis.

DNA extraction

The DNA was extracted using a DP305 plant DNA kit (TIANGEN Biotech Co., Ltd., Beijing, China) [16]. The quality and concentration of the DNA were determined by agarose gel electrophoresis and NanoDrop2000 ultra-micro ultraviolet spectrophotometer (Thermo Scientific, MIT, USA) [26], and the DNA was stored at -20°C for subsequent analysis.

Primer screening, PCR amplification and sequencing

A total of 71 SSR primer pairs were used, of which 63 pairs were obtained from the transcriptome data analysis of *A. villosum* developed in our laboratory (at Zhanjiang Experimental

Station, Chinese Academy of Tropical Agricultural Sciences), and 8 were obtained from a previous study [27] (S1 File). The SSR primers were synthesized by Shanghai Shengsong Biotechnology Co., Ltd. SSR markers were screened using five morphologically distinct germplasm, including S8, S10, S40, S52, and S92, to identify polymorphic markers and those with good amplification effects were selected. Subsequently, 55 pairs of SSR markers were selected for genotyping the 75 accessions of *A. villosum* using PCR amplification.

The PCR reaction volume was 10 μ l, consisting of 5 μ l of 2x Taq plus Master Mix (Bishsarp Life Sciences Co., Ltd, China), 0.5 μ l of the forward and reverse primers (each 0.5 μ M), 1 μ l of genomic DNA (10ng), and 3 μ l of double distilled water (ddH₂O). The PCR program setting and PCR product analysis were conducted according to the method by Yeboah et al. [26].

Data analysis

Phenotypic data analyses. Descriptive statistics, viz. minimum, maximum, mean, range, and standard deviation values, were calculated using SPSS version 21.0, while the coefficient of variation (CV) and diversity index (H_e') were calculated using EXCEL 2021 [28]. Pearson's correlation analysis of the different traits was conducted using the *corrplot* package in R 3.4.5 [29].

Genotypic data analysis. Each amplification product was considered a unit character, with the SSR-PCR bands being scored as either present (1) or absent (0) and missing bands recorded as 9. Various parameters related to marker polymorphism, such as major allele frequency (MAF), the number of alleles, the number of genotypes, gene diversity, and the polymorphism information content (PIC), were calculated using Powermaker V3.25 [30].

Population structure, relative kinship, LD, marker-trait association, and mining of beneficial alleles. Population structure (Q matrix) and a kinship coefficient estimation matrix (K) of the tested *A. villosum* accessions were analyzed using the STRUCTURE 2.3.4 software [13] and SPAGeDi software [31], respectively, as previously described [26]. LD, genome-wide association mapping, manhattan plot, and quantile-quantile (QQ) plots were generated using TASSEL 3.0.174 software [32] according to the Bonferroni test threshold [33, 34], which was set as 0.05/total SSR makers ($P < 0.05/293 = 1.6 \times 10^{-4}$). To calculate the coefficient of genetic differentiation (F_{ST}) and gene flow (N_m), use Excel and GenAlEx 6.5 (Due to the small sample size in Hainan and Fujian, no statistics were conducted).

Candidate gene prediction and expression analyses. Based on the GWAS analysis results, functionally annotated genes were searched within the 200 Kb interval upstream and downstream of the molecular markers in the *A. villosum* genome [1, 9]. The candidate genes related to traits were selected based on the annotation results, and the identified regions were subjected to BLAST analysis using BlastP on the National Center for Biotechnology Information (NCBI) website to predict potential candidate genes. The fresh roots stems and leaves of the plant materials with significantly different LEL and NNS were selected for quantitative fluorescence expression pattern analysis. Total RNA was extracted using FastPure® Tissue Total RNA Isolation Kit (Vazyme #RC101, China). The oligo-(dT)18 primer and 1 μ g of total RNA were used to synthesize cDNA using the HiScript® III RT SuperMix for qPCR (+gDNA wiper)(Vazyme #R323, China). The sequences of candidate genes were retrieved from the NCBI database (<https://doi.org/10.6084/m9.figshare.19200005.v1>), and the primers targeting these genes were designed using the qPCR 2 Primers Intercalating Dyes tool (<https://sg.idtdna.com/pages>) and were synthesized by Sangon Biotechnology Co., Ltd. Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) was performed on LightCycler480 II using ChamQ™ Universal SYBR® qPCR Master Mix (Vazyme #Q711, China). The reaction conditions were 95°C for 2 min, followed by 45 cycles of 95°C for 20 s, 55°C for 15 s, and 72°C for 30 s. The *TUA* gene was selected as the endogenous control, and the relative gene

expression was measured using the delta-delta Ct ($2^{-\Delta\Delta Ct}$) method. The data were analyzed in Microsoft Excel 2013 and compiled in GraphPad Prism.

Results

Phenotypic characterization

The Shannon index (H_e') reflects species richness and evenness. The more uniform the number of individuals, the greater the H_e' and the smaller the degree of variation. Conversely, the coefficient of variation reflects the opposite result, i.e., the larger the coefficient of variation, the greater the degree of dispersion. Statistical analysis showed that the H_e' values of the 12 phenotypic traits of *A. villosum* accessions had a range of 4.146–4.312 (Table 1), and those of SD, NB, NR, LL, LW, and ST were higher than 4.3. The H_e' of NR was the lowest at 4.146, while that of LW was the highest at 4.312. The maximum coefficient of variation for NR was 0.691, and the minimum coefficient of variation for LW was 0.105. Overall, these results indicate that NR had the highest variation, while LW had the lowest. As shown in Fig 1, the frequency distribution of the 9 phenotypic traits conformed to normal distribution.

Correlation analyses of the phenotypic traits

The correlation analysis of the 12 traits was conducted, and the correlation coefficients are shown in Fig 3. The results indicate that PH had a highly significant positive correlation with SD, NB, LL, LW, LEL, TLS, ST, SIL, and TNR at the $P < 0.001$ significance level and with NNS at $P < 0.01$ significance level (Fig 2). Additionally, SD had a highly significant positive correlation with NB, LL, LW, LEL, TLS, ST, and SIL ($P < 0.001$) and a significant positive correlation with TNR ($P < 0.01$). NB was not correlated with NR, LW, ST, NNS, and TNR but had a highly significant positive correlation with LL, LEL, and TLS ($P < 0.001$) and a significant positive correlation with SIL ($P < 0.05$). NR only had a significant positive correlation with NNS at the $P < 0.05$ significance level. Furthermore, LL exhibited a highly significant positive correlation with LW, LEL, SIL, and TNR at the $P < 0.001$ significance level and a highly significant

Table 1. Diversity analysis of the growth traits of *Amomum villosum* Lour.

Trait ^(a)	Minimum	Maximum	Mean	Range	SD ^(b)	CV ^(c)	H_e' ^(d)	K-S ^(e)
PH (cm)	52.58	177.94	116.72	125.36	24.80	0.212	4.295	0.941
SD/(mm)	5.60	12.78	9.33	7.18	1.21	0.130	4.309	0.000
NB	15.00	30.60	21.55	15.60	3.33	0.154	4.306	0.164
NR	1.60	4.40	3.08	2.80	0.45	0.147	4.307	0.000
LL/(cm)	23.78	46.04	32.55	22.26	3.88	0.119	4.311	0.111
LW/(cm)	3.48	5.74	4.67	2.26	0.49	0.105	4.312	0.986
LEL/(cm)	0.20	1.02	0.42	0.82	0.14	0.346	4.265	0.001
TLS/(cm)	7.48	42.80	24.94	35.32	6.46	0.259	4.283	0.878
ST/(mm)	4.58	9.15	6.36	4.58	0.91	0.143	4.308	0.857
SIL/(cm)	0.20	1.88	0.43	1.68	0.22	0.510	4.227	0.720
NNS	21.00	491.00	97.35	470.00	67.28	0.691	4.146	0.070
TNR	48.00	1148.00	281.21	1100.00	173.60	0.617	4.163	0.060

^(a) PH = plant height; SD = stem diameter; NB = number of blades; NR = number of ramet; LL = leaf Length; LW = leaf width; LEL = ligule length; TLS = total length of stolons; ST = stolon thickness; SIL = stolon internode length; NNS = number of new shoots; TNR = total number of ramets.

^(b) Standard deviation.

^(c) Coefficient of variation.

^(d) Diversity index.

^(e) Kolmogorov-Smirnov test, '>0.05' indicates normal distribution; Extreme maximum values were removed during NNS and TNR statistics.

<https://doi.org/10.1371/journal.pone.0306806.t001>

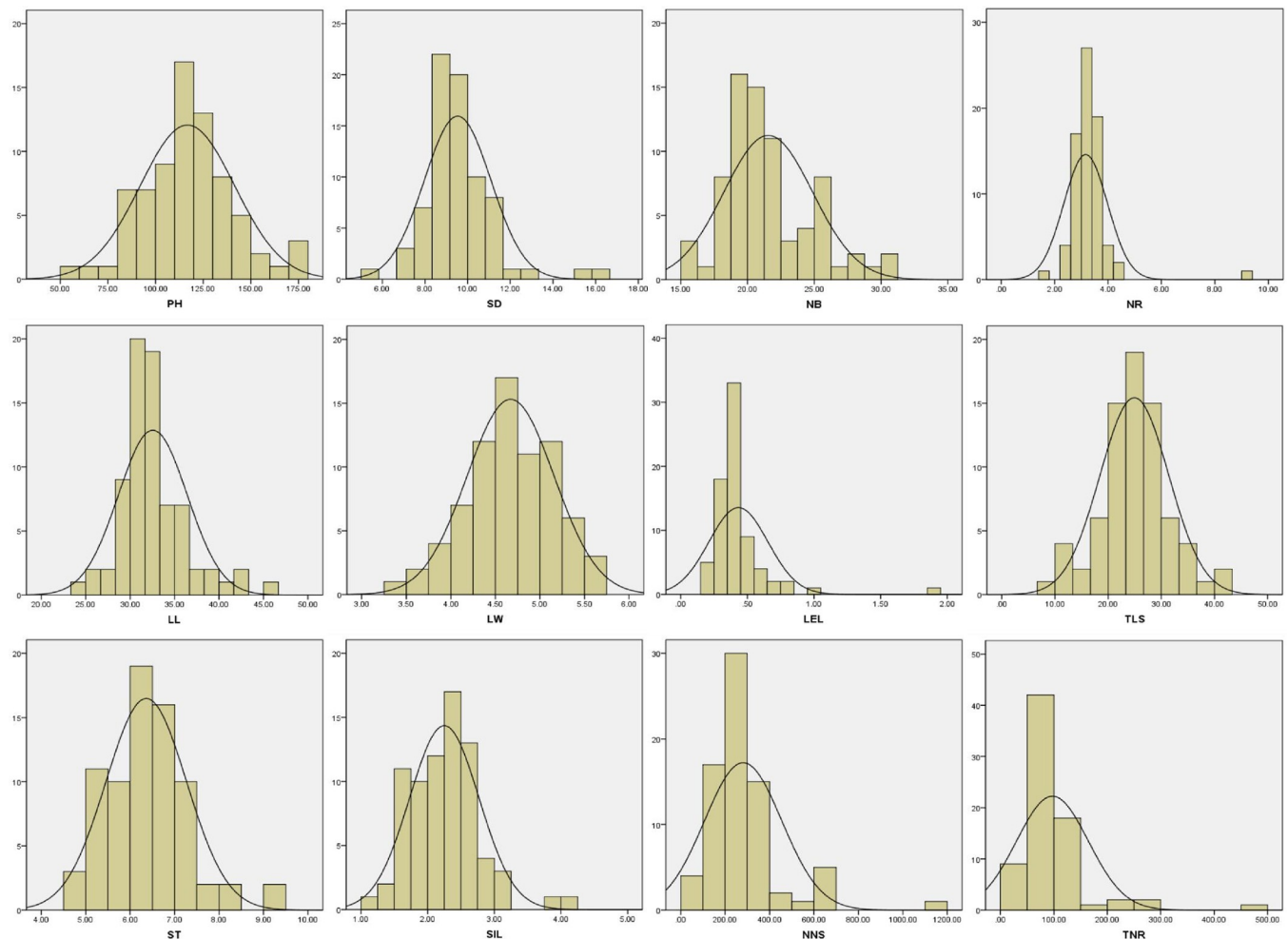


Fig 1. Frequency distribution of the 12 agronomic traits of *Amomum villosum* Lour. PH = height; SD = stem diameter; NB = number of blades; NR = number of ramets; LL = leaf Length; LW = leaf width; LEL = ligule length; TLS = total length of stolons; ST = stolon thickness; SIL = stolon internode length; NNS = number of new shoots; TNR = total number of ramets.

<https://doi.org/10.1371/journal.pone.0306806.g001>

positive correlation with ST and NNS at the $P < 0.01$ significance level. Similarly, LW demonstrated a highly significant positive correlation with LEL, TLS, and SIL at the $P < 0.001$ significance level and a highly significant positive correlation with ST and TNR at the $P < 0.01$ significance level. LEL showed a highly significant positive correlation with TLS and SIL at the $P < 0.001$ significance level and a highly significant positive correlation with ST at the $P < 0.01$ significance level. Furthermore, a highly significant positive correlation existed between TLS and the ST, SIL, NNS, and TNR ($P < 0.001$). ST exhibited a highly significant positive correlation with SIL ($P < 0.001$) and a significant positive correlation with TNR ($P < 0.05$). There was also a highly significant positive correlation between SIL and TNR ($P < 0.01$) and a significant positive correlation between SIL and NNS ($P < 0.05$).

Population structure and cluster analysis

The population structure and cluster analysis indicated that the *A. villosum* accessions could be categorized into two major subpopulations. The group population structure analyzed through the “Structure Harvester” program showed a sharp peak of the *ad hoc* quantity (ΔK)

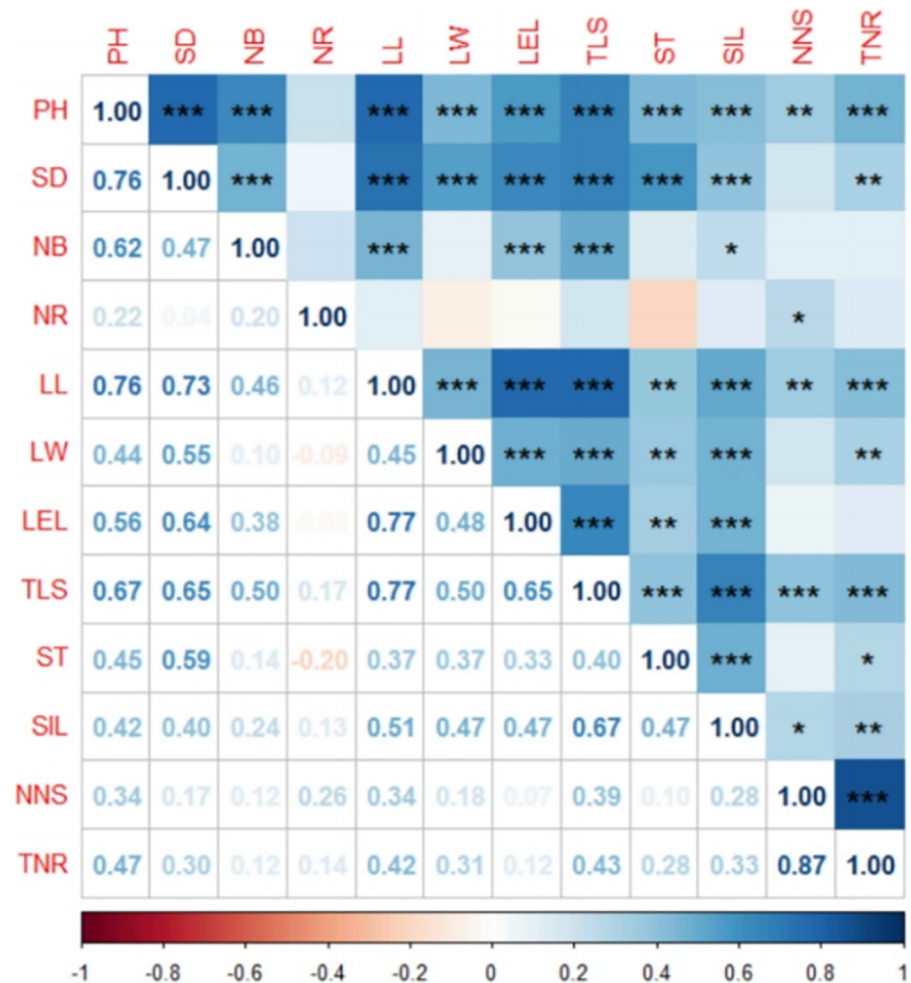


Fig 2. Correlation heat map of the 12 agronomic traits of *Amomum villosum* Lour. “*” indicates correlation significance at the 0.05 level, while “**” represents significance at the 0.01 level and “***” shows significance at the 0.001 level. PH = height; SD = stem diameter; NB = number of blades; NR = number of ramets; LL = leaf Length; LW = leaf width; LEL = ligule length; TLS = total length of stolons; ST = stolon thickness; SIL = stolon internode length; NNS = number of new shoots; TNR = total number of ramets.

<https://doi.org/10.1371/journal.pone.0306806.g002>

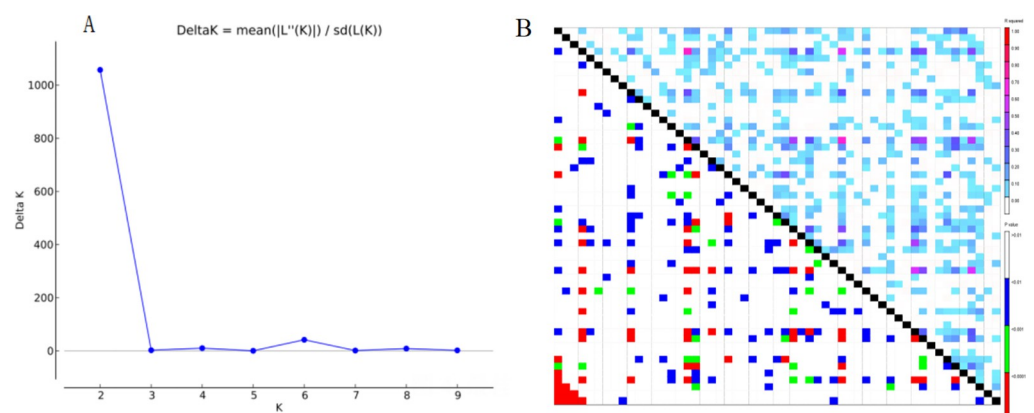


Fig 3. Population structure (A) and cluster analysis (B) of the 75 *Amomum villosum* accessions.

<https://doi.org/10.1371/journal.pone.0306806.g003>

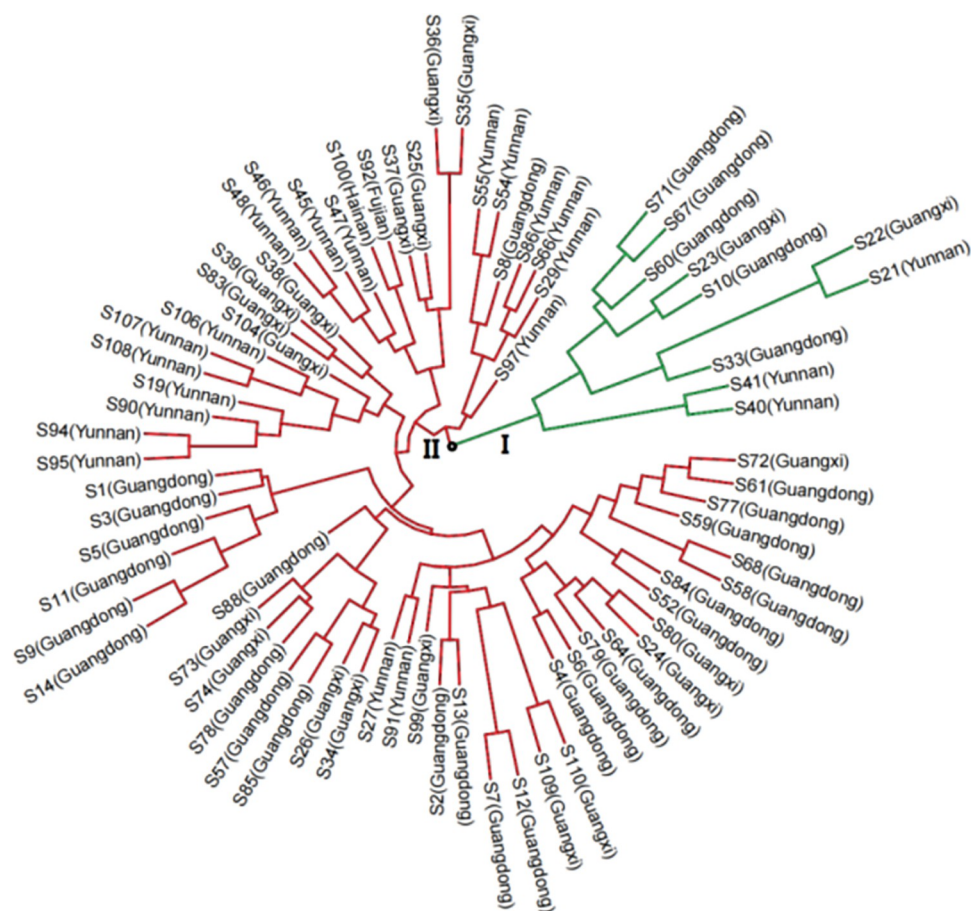


Fig 4. The linkage disequilibrium (LD) distribution among 55 simple sequence repeat (SSR) loci in 75 *Amomum villosum* accessions.

<https://doi.org/10.1371/journal.pone.0306806.g004>

at $K = 2$ (Fig 3A). Cluster analysis was performed using the RStudio software, and the clustering tree showed the most likely genetic relationship between the *A. villosum* accessions (Fig 4). The cluster analysis categorized all materials into two major clusters, with the first cluster containing 10 materials and the second 65. Cluster I comprised two branches, one containing two materials from Yunnan and the other having eight materials from Guangdong, Guangxi, and Yunnan. This branch indicated a closer genetic relationship among Guangdong, Guangxi, and Yunnan materials. Cluster two also had two major branches, with one consisting of 7 materials from Yunnan and the other encompassing 58 materials. Overall, the cluster analysis results showed that the *A. villosum* accessions from different places in China were dependent but were not significantly correlated with their origin, indicating that though the accessions were from different sources, they were not significantly different.

The analysis of the F_{ST} and N_m yielded the following results in the Table 2 below. The coefficient of F_{ST} between populations ranges from 0.016 to 0.027, with the lowest F_{ST} between populations A and B (0.016) and the highest F_{ST} between populations B and C (0.027). However, all F_{ST} values are less than 0.05, indicating a low degree of genetic differentiation between populations. In terms of N_m , the highest N_m is observed between populations A and B, at 15.335, suggesting high genetic similarity. Conversely, the lowest N_m is found between populations B and C, at 8.6, indicative of low genetic similarity. Selecting these two populations (B and C) for hybrid breeding is more likely to yield new variations.

Table 2. Values of F_{ST} (below diagonal) and N_m (above diagonal) among 3 populations of *Amomum villosum* Lour.

Population	A	B	C
A		15.335	9.398
B	0.016		8.896
C	0.026	0.027	

Note: A = population of Guangdong; B = population of Guangxi; C = population of Yunnan

<https://doi.org/10.1371/journal.pone.0306806.t002>

The degree of LD was evaluated using the average coefficient of determination (R^2) at 0.01 (Fig 3B). The distribution analysis of LD showed that 13.06% and 6.32% of R^2 were statistically significant at $P < 0.01$ and $P < 0.001$, respectively.

Marker-trait association

The association between the SSR markers and the traits was analyzed using a mixed linear model (MLM). Three significant SSR markers were found to be associated with growth traits in *A. villosum* (Fig 5A and 5C). NNS was linked to two makers, AVL14 and AVL22, which had the phenotypic variance explained (PVE) proportions of 41.31% and 38.07%, respectively. The other marker (AVL11) was associated with LEL and had a PVE value of 60.33%. The QQ and Manhattan plots for the maker-trait associations are shown in Fig 5D. There was evident statistic inflation for LEL and NNS. Based on the physical location of the molecular marker sites associated with the GWAS analysis, six candidate genes with functional annotation were identified within the 200 Kb interval upstream and downstream of the molecular markers in the *A. villosum* genome (Table 3).

Candidate genes and their relative expression in four different accessions

The relative expression levels of six candidate genes in different tissues of *A. villosum* accessions (GX16 with the most NNS, YN9 with the least NNS, YN17 with the longest LEL, and

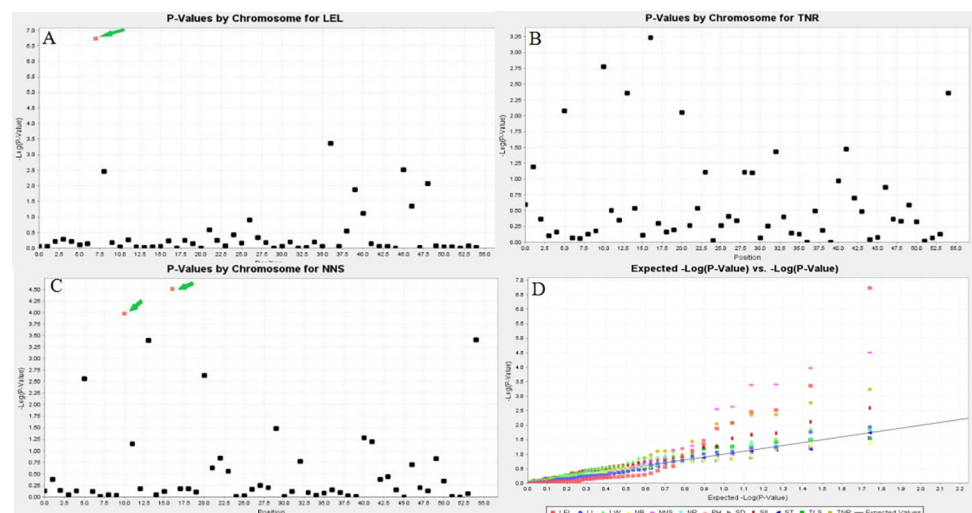


Fig 5. Genome-wide association analysis of the growth traits of *Amomum villosum* Lour. using a mixed linear model. Manhattan plot of (A) ligule length, (B) total number of ramets and (C) number of new shoots. (D) Quantile-quantile (QQ) plot of all the 12 traits.

<https://doi.org/10.1371/journal.pone.0306806.g005>

Table 3. Simple sequence repeat (SSR) markers associated with target traits (MLM) of *Amomum villosum* Lour.

Trait	Marker	F-value	P-value	Phenotypic variance explained (%)	Chromosome	Position (bp)	Candidate gene ID	Description
LEL	AVL11	14.64	1.86E-07	60.33	LG23	9730351>>—85—>> 9730475	Wv_072909	LRR receptor kinase SERK2-like (LOC122018488)
							Wv_072917	oleosin
NNS	AVL22	9.30	3.10E-05	38.07	LG06	17632519>>—85—>> 17632644	Wv_033832	BEL1-like homeodomain protein 4 (LOC121981163)
							Wv_033833	Histone-lysine N-methyltransferase SUV5-like
NNS	AVL14	6.07	1.05E-04	41.31	LG06	53462795>>—111—>> 53462946	Wv_038222	Alanine-tRNA ligase, chloroplastic/mitochondrial-like
							Wv_038224	uncharacterized

<https://doi.org/10.1371/journal.pone.0306806.t003>

YN9 with the shortest LEL) were determined using qRT-PCR. As shown in Fig 6, three candidate genes showed significant differences in different tissues of the four accessions. The relative expression level of *Wv_038222* significantly differed in the roots and leaves of GX16 and YN9, while that of *Wv_038224* significantly differed in the roots of GX16 and YN9. Moreover, *Wv_072909* showed highly significant differences in the roots, stems, and leaves of YN17 and YN9.

Discussion

Phenotypic characterization, population structure and LD analysis of the 75 *A. villosum* accessions

Descriptive statistics and correlation analysis reflect the differences and similarities in population phenotypic traits [35]. However, the magnitude of variation amplitude and coefficient of variation indicate the phenotypic richness in the population. A larger variation amplitude and coefficient of variation mean that the population has a high degree of trait variation and phenotypic diversity [36]. The diversity index of *A. villosum* accessions ranged from 4.081 to 4.312, with a coefficient of variation ranging from 0.105 to 0.691. This indicated that the *A. villosum* accessions collected from Yunnan, Guangxi, Guangdong, Fujian, and Hainan provinces in China were different, probably due to geographical isolation restrictions, which resulted in complex genetic components [37], leading to differences in the phenotypic traits of germplasm in different regions. Through correlation analysis, we found that most growth traits were correlated, indicating that they have a promoting effect on each other.

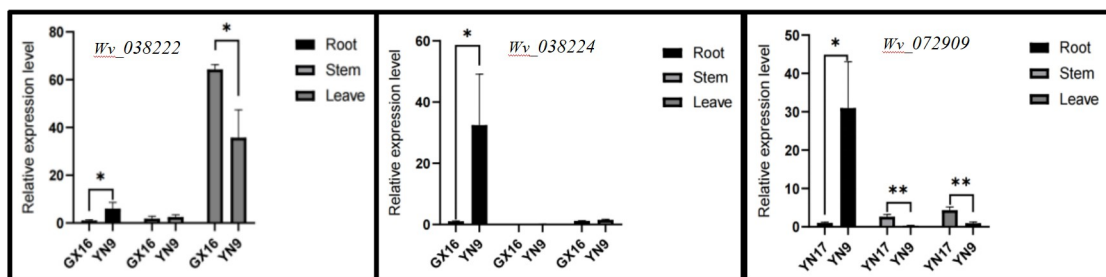


Fig 6. The expression levels of three candidate genes showing significant differences in different tissues of the phenotypically different accessions (GX16 with the most NNS, YN9 with the least NNS, YN17 with the longest LEL, and YN9 with the shortest LEL). Data represent the mean of three replicates. Significant differences are indicated as * $P < 0.05$ and ** $P < 0.01$. NNS = number of new shoots; LEL = ligule length.

<https://doi.org/10.1371/journal.pone.0306806.g006>

Genetic diversity and LD analysis of the 75 *A. villosum* accessions

The population structure is an important factor affecting the association analysis results [36]. False-positive associations should be eliminated in cases where the population structure of the association is simple and conforms with the cluster analysis [26]. In this study, the population structure analysis showed that the *A. villosum* accessions were categorized into two subgroups. Similarly, the cluster analysis also grouped the accessions into two groups, indicating that *A. villosum* generally has a simple population structure, similar to the results of Li et al. [37] and Xu et al. [5]. Notably, *A. villosum* accessions from Guangdong province were categorized into different subgroups and had close genetic relationships with germplasm from different regions. This may be because Guangdong is the main production area of *A. villosum* and distributes germplasm resources to other regions, resulting in a high degree of similarity.

LD is the foundation of association analysis, and the degree of LD in the population determines the resolution of association analysis. Evaluating the whole genome LD before conducting GWAS simplifies the process of association analysis [38]. R^2 is an indicator for measuring the degree of LD, with a range of 0–1. The closer the R^2 is to 1, the closer the linkage. The average R^2 obtained in this study was 0.099, which was higher than that reported in castor bean [26] and cotton [39] but lower than reported in soybean [34] and rice [40]. However, overall, the degree of LD in the *A. villosum* population was still low, probably because *A. villosum* is a cross-pollinated species with a higher recombination rate and LD break, resulting in a lower level of LD [39]. This lower level of LD results in weaker association signals and fewer associated marker loci. Therefore, if conditions permit, it is necessary to appropriately increase the number of markers to increase the strength of the associated signals.

Marker-trait associations and mining of beneficial alleles in *A. villosum*

This study is the first to perform GWAS in *A. villosum*. Three SSR molecular markers (AVL11, AVL22 and AVL14) significantly associated with growth traits of *A. villosum* were detected using the MLM (Q+K) method. Two loci, AVL22 and AVL14, with a PVE of 38.73% and 42.27%, respectively, were linked to NNS, suggesting that a trait could be affected by multiple genes [41, 42]. Meanwhile, LEL was associated with the AVL11 locus with a PVE of 60.33%. The association results showed that 55 marker loci were related to growth traits, explaining 38.73% to 60.33%, which was higher than those reported in other crops, such as soybean (5.42–27.6%) [34], castor (7.32–20.47%) [26], and peanut (11.22–32.30%) [43]. This indicated that these loci are promising genetic markers that should be evaluated for use in MAS [44].

Predicted putative candidate genes and their relative expression levels

Based on the expression analysis, the three genes exhibited significant differences in their expression levels across various tissues of different germplasms. Among them, *Wv-038222* and *Wv-038224* are potential candidates for the same marker AVL14 and are both associated with the number of new shoots in the plant. The primary factor influencing the number of new shoots is the meristematic ability of the root and stem. *Wv-038222* and *Wv-038224* had low expression in GX16 and high expression in YN9, indicating that high expression of these genes inhibits the growth of the root and stem, thereby decreasing the number of new shoots. Through CD-Search comparison in NCBI, it was discovered that *Wv-038222* shares similarities with the Alanine-tRNA ligase (*AlaRS*) gene. Previous studies indicated that the *AlaRS* gene is linked to embryonic development and functions as an enzyme in aminoacylation [45]. While the specific role of the *AlaRS* gene in root and stem differentiation remains unknown, mutations in *AlaRS* gene expression can result in cell or embryo death, thus impacting plant

growth and development. Hence, it can be inferred that the *Wv-038222* gene may play a certain role in the meristematic ability of the root and stem tissues in *A. villosum*.

Furthermore, *Wv-038222* was also characterized by significantly higher expression levels in YN9 leaves than GX16. However, there are no comparison data on the *Wv-038224* gene in the NCBI database, and the action mechanism of the gene requires further research.

The *Wv.072909* gene is associated with ligule length. The expression level of the gene in the root of YN17 was significantly lower than that of YN9, while the expression level in the stem and leaves of YN17 was notably higher than that of YN9. This suggests that the high expression in the leaves and stems promotes leaf ligule elongation, whereas high expression in the root system has no effect. Upon comparison, it was discovered that the *Wv-072909* gene is similar to the FRQ1 superfamily member COG5126, identified as a Ca²⁺-binding protein gene. Research indicates that this gene may be closely linked to plant stress resistance [46]. The leaf ligule of *A. villosum* resembles that of cereal crops in the Poaceae family, containing numerous stomata that effectively reduce water evaporation. For example, rice maintains water in intense ultraviolet environments through the stomata, thus enhancing its drought resistance. Additionally, these stomata can create a confined space, maintaining a relatively stable temperature and resisting the intrusion of cold weather. Therefore, the presence of leaf ligule in *A. villosum* may hold a significance akin to that of cereal crops. It can be inferred that the expression of *Wv-072909* influenced the leaf length of *A. villosum*, thereby impacting the stress resistance of *A. villosum*.

In summary, the three differentially expressed genes may be candidate genes associated with the number of new shoots and ligule length of *A. villosum*. However, it is necessary to verify these candidate genes through transgenic validation to accurately obtain their functions.

Conclusion

In summary, the 12 agronomic traits of *A. villosum* germplasm resources exhibited a normal distribution, indicating that these traits are mainly controlled by multiple genes. We detected 293 allelic variations using SSR primers, with an average amplification of 5.32 alleles and a variation range of 3–8, indicating that the SSR markers were feasible and highly polymorphic. Using the MLM model of Q+K for association analysis, three SSR molecular markers (AVL11, AVL22 and AVL14) significantly associated with the agronomic traits were detected. The genes on the left and right borders of the markers were selected as candidate genes, and their differential expression in different tissues was analyzed in four phenotypically different accessions. Finally, three related candidate genes were identified. The NNS was found to be associated with the AVL22 and AVL14 loci, while the LEL was linked to the AVL11 locus. This indicates the potential of these loci as genetic markers, and their application in MAS should be further validated.

Supporting information

S1 File. Supplementary table includes accessions, phenotypic, genotype, primer information, and coding sequences of the candidate genes.
(XLSX)

Acknowledgments

We are thankful to Gang Zhang, and all anonymous reviewers for making this manuscript a better one.

Author Contributions

Conceptualization: Yunfeng Shi, Junjun He.

Data curation: Wenxiu Li, Junjun He.

Formal analysis: Wenxiu Li, Hualin Zhang, Yanli Yao.

Funding acquisition: Ping Luo, Junjun He.

Investigation: Qing Yan, Yana Ye.

Methodology: Wenxiu Li, Ping Luo.

Project administration: Junjun He.

Resources: Junjun He.

Validation: Yanli Yao.

Writing – original draft: Wenxiu Li.

Writing – review & editing: Wenxiu Li, Yunfeng Shi, Junjun He.

References

1. Yang P, Zhao H-Y, Wei J-S, Zhao Y-Y, Lin X-J, Su J, et al. Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of *Wurfbainia villosa*. *Plant J*. 2022; 112(3): 630–645. <https://doi.org/10.1111/tpj.15968> PMID: 36071028
2. Yue J, Zhang S, Zheng B, Raza F, Luo Z, Li X, et al. Efficacy and Mechanism of Active Fractions in Fruit of *Amomum villosum* Lour. for Gastric Cancer. *J Cancer*. 2021; 12(20): 5991–5998. <https://doi.org/10.7150/jca.61310> PMID: 34539873
3. Xuefeng X, Xueying LI, Feilong C, Wenxue XU. Study on Chemical Constituents of Essential Oil from Fruits and Leaves of *Amomum villosum* Lour. by GC-MS. *Traditional Chinese Drug Research and Clinical Pharmacology*. 2012; 5(5): 760–765. <https://doi.org/10.4236/ajps.2014.55090>
4. De Boer H, Newman M, Poulsen AD, Droop AJ, Fér T, Hiên LTT, et al. Convergent morphology in *Alpinieae* (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. *Taxon*. 2018; 67(1): 6–36. <https://doi.org/10.12705/671.2>
5. Xu J, Yang B, Li M, Li Z, Tu Y, Tang L, et al. Research on germplasm diversity of *Amomum villosum* Lour in genuine producing area. *PLoS One*. 2022; 17(8): e0268246. <https://doi.org/10.1371/journal.pone.0268246> PMID: 36044470
6. Yang T, Jiannong LU, Yeboah A, Shuailei GU, Dongna LI, Shi Y, et al. Construction of castor functional markers fingerprint and analysis of genetic diversity. *Biocell*. 2020; 44(3): 381–388 <https://doi.org/10.32604/BIOCELL.2020.09112>
7. Banerjee N., Siraree A., Yadav S., Kumar S., Singh J., Kumar S., et al. Marker-trait association study for sucrose and yield contributing traits in sugarcane (*Saccharum* spp. hybrid). *Euphytica*. 2015; 205(1): 185–201. <https://doi.org/10.1007/s10681-015-1422-3>
8. Kaur J, Akhatar J, Goyal A, Kaur N, Banga SS. Genome wide association mapping and candidate gene analysis for pod shatter resistance in *Brassica juncea* and its progenitor species. *Molecular Biology Reports*. 2020; 47(4): 2963–2974. <https://doi.org/10.1007/s11033-020-05384-9> PMID: 32219770
9. Kalyana Babu B, Mathur RK, Venu. MVB, Shil Sandip, Ravichandran G, Anita P, et al. Genome-wide association study (GWAS) of major QTLs for bunch and oil yield related traits in *Elaeis guineensis* L. *Plant science: an international journal of experimental plant biology*. 2021; 305: 110810. <https://doi.org/10.1016/j.plantsci.2020.110810> PMID: 33691957
10. Jian Zhang, Jingjing Yang, Yanling Lv, Xiaofei Zhang, Changxuan Xia, Hong Zhao, et al. Genetic diversity analysis and variety identification using SSR and SNP markers in melon. *BMC plant biology*. 2023; 23(1): 39. <https://doi.org/10.1186/s12870-023-04056-7> PMID: 36650465
11. Alsaleh A. SSR-based genome-wide association study in turkish durum wheat germplasms revealed novel QTL of accumulated platinum. *Mol Biol Rep*. 2022; 49(12): 11289–11300. <https://doi.org/10.1007/s11033-022-07720-7> PMID: 35819556
12. Singh N, Choudhury DR, Tiwari G, Singh AK, Kumar S, Srinivasan K, et al. Genetic diversity trend in Indian rice varieties: an analysis using SSR markers. *BMC Genetics*. 2016; 17(1): 127. <https://doi.org/10.1186/s12863-016-0437-7> PMID: 27597653

13. Tajner N. Genetic Structure and Core Collection of Olive Germplasm from Albania Revealed by Microsatellite Markers. *Genes*. 2021; 12(2): 256. <https://doi.org/10.3390/genes12020256> PMID: 33578843
14. Ji R, Yu X, Ren T, Chang Y, Li Z, Xia X, et al. Genetic diversity and population structure of *Caryopteris mongholica* revealed by reduced representation sequencing. *BMC Plant Biology*. 2022; 22(1): 1–13. <https://doi.org/10.1186/s12870-022-03681-y> PMID: 35710341
15. Chen J, Liu Q, Lu C, Liu Q, Pan J, Zhang J, et al. Genetic diversity of *Prunus armeniaca* L. var. *ansu* Maxim. germplasm revealed by simple sequence repeat (SSR) markers. *PLoS One*. 2022; 17(6): e0269424. <https://doi.org/10.1371/journal.pone.0269424> PMID: 35657925
16. Gong L, Zhang D, Ding X, Huang J, Guan W, Qiu X, et al. DNA barcode reference library construction and genetic diversity and structure analysis of *Amomum villosum* Lour. (Zingiberaceae) populations in Guangdong Province. *PeerJ Inc*. 2021; 9: e12325. <https://doi.org/10.7717/peerj.12325> PMID: 34721994
17. Huang Q, Duan Z, Yang J, Ma X, Zhan R, Xu H, et al. SNP Typing for Germplasm Identification of *Amomum villosum* Lour. Based on DNA Barcoding Markers. *PLoS ONE*. 2014; 9(12): e114940. <https://doi.org/10.1371/journal.pone.0114940> PMID: 25531885
18. Elango D, Wang W, Thudi M, Sebastiar S, Ramadoss BR, Varshney RK. Genome-wide association mapping of seed oligosaccharides in chickpea. *Front Plant Sci*. 2022; 13: 1024543. <https://doi.org/10.3389/fpls.2022.1024543> PMID: 36352859
19. Hansen M., Kraft T., Ganestam S., TORBJÖRN SÄLL, & Nilsson N. O. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genetics Research*. 2001; 77(1): 61–66. <https://doi.org/10.1017/s0016672300004857> PMID: 11279831
20. Elango D, Xue W, Chopra S. Genome wide association mapping of epi-cuticular wax genes in *Sorghum bicolor*. *Physiol Mol Biol Plants*. 2020; 26(8): 1727–1737. <https://doi.org/10.1007/s12298-020-00848-5> PMID: 32801499
21. Parisseaux B., Bernardo R. In silico mapping of quantitative trait loci in maize. *Theoretical & Applied Genetics*. 2004; 109(3): 508–514. <https://doi.org/10.1007/s00122-004-1666-0> PMID: 15150690
22. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet*. 2005; 111(4): 723–730. <https://doi.org/10.1007/s00122-005-2057-x> PMID: 15997389
23. Ge LI, Yan-Fang W, Ling T, Chun-Yong Y, Rong-Ying LI, Li-Xia Z, et al. Phenotypic Trait Variation, Correlation and Path Analysis of *Amomum villosum* Lour in Yunnan. *Journal of Yunnan University of Traditional Chinese Medicine*. 2015; 38(1): 38–41+47. <https://doi.org/10.19288/j.cnki.issn.1000-2723.2015.01.010>
24. Wang Z. Specification for the Description of Germplasm Resources of Southern Medicines. Beijing, Agricultural Press of China. 2006. Available: <https://book.duxiu.com/bookDetail.jsp?dxNumber=000004932152&d=EEE07B23E676343AA0BF61228423B476&fenlei=17090704&sw=%E5%8D%97%E8%8D%AF%E7%A7%8D%E8%B4%A8%E8%B5%84%E6%BA%90%E6%8F%8F%E8%BF%B0%E8%A7%84%E8%8C%83>
25. NY/T 3725–2020. Guidelines for testing plant variety specificity (distinguishability), consistency, and stability of *Amomum villosum*. Beijing, China Agricultural Publishing House. 2020. Available: <https://book.duxiu.com/StdDetail.jsp?dxid=320151509555&d=2887E3A27F6FB8F202ACDDBDC7A6F33C&fenlei=0>
26. Yeboah A, Lu J, Yang T, Karikari B, Yin X. Genome-wide association study identifies loci, beneficial alleles, and candidate genes for cadmium tolerance in castor (*Ricinus communis* L.). *Industrial Crops and Products*. 2021; 171: 113842. <https://doi.org/10.1016/j.indcrop.2021.113842>
27. Li M., Screening of Excellent Germplasm of *Amomum villosum* in the Real Estate District (Master's degree) Guangzhou University of Traditional Chinese Medicine. 2019. <https://doi.org/10.27044/d.cnki.ggzsu.2019.000443>
28. Shannon C. E. A Mathematical Theory of Communication. *Bell Systems Technical Journal*. 1948; 27(3): 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
29. Agyenim-Boateng KG, Zhang S, Islam MS, Gu Y, Li B, Azam M, et al. Profiling of naturally occurring folates in a diverse soybean germplasm by HPLC-MS/MS. *Food Chemistry*. 2022; 384: 132520–. <https://doi.org/10.1016/j.foodchem.2022.132520> PMID: 35217465
30. Kejun L, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005; 21(9): 2128–2129. <https://doi.org/10.1093/bioinformatics/bti282> PMID: 15705655
31. Hardy OJ, Vekemans X. SPAGeDI: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*. 2002; 2(4): 618–620. <https://doi.org/10.1046/j.1471-8286.2002.00305.x>

32. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES, et al. Genetics and population analysis TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics*. 2008; 23(19): 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: 17586829
33. Li Y, Cao K, Zhu G, Fang W, Chen C, Wang X, et al. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome biology*. 2019; 20(1): 36. <https://doi.org/10.1186/s13059-019-1648-9> PMID: 30791928
34. Karikari B, Bhat JA, Denwar NN, Zhao TJ. Exploring the genetic base of the soybean germplasm from Africa, America and Asia as well as mining of beneficial allele for flowering and seed weight. *3Biotech*. 2020; 10(5): 195. <https://doi.org/10.1007/s13205-020-02186-5> PMID: 32296618
35. Brownstein NC, Adolfsson A, Ackerman M. Descriptive statistics and visualization of data from the R datasets package with implications for clusterability. *Data Brief*. 2019; 25: 104004. <https://doi.org/10.1016/j.dib.2019.104004> PMID: 31317060
36. Yang T., Lu J., Zhang D., Shi Y., Li D., Gu S., et al. Classification and evaluation of wild castor germplasm resources in South China. *Chinese Journal of Oil Crops*. 2020; 42(3): 420–431. <https://doi.org/10.19802/j.issn.1007-9084.2020026>
37. Li W-X, Li J-L, He J-J, Zhang H-L, Luo P, Wei Y, et al. Genetic diversity and population structure of germplasm resources of *Amomum villosum* based on SSR markers. *Zhongguo Zhong Yao Za Zhi*. 2022; 47(17): 4618–4626. <https://doi.org/10.19540/j.cnki.cjcm.20220514.101> PMID: 36164867
38. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015; 47(3): 291–295. <https://doi.org/10.1038/ng.3211> PMID: 25642630
39. Nie X, Huang C, You C, Li W, Zhao W, Shen C, et al. Genome-wide SSR-based association mapping for fiber quality in nation-wide upland cotton inbred cultivars in China. *BMC Genomics*. 2016; 17(1): 352. <https://doi.org/10.1186/s12864-016-2662-x> PMID: 27177443
40. Liu S, Zhong H, Meng X, Sun T, Li Y, Pinson SRM, et al. Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. *BMC Plant Biol*. 2020; 20(1): 441. <https://doi.org/10.1186/s12870-020-02603-0> PMID: 32972357
41. Cober ER, Morrison MJ. Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. *Theor Appl Genet*. 2010; 120(5): 1005–1012. <https://doi.org/10.1007/s00122-009-1228-6> PMID: 20012856
42. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, et al. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics*. 2011; 188(2): 395–407. <https://doi.org/10.1534/genetics.110.125062> PMID: 21406680
43. Zhao J, Huang L, Ren X, Pandey MK, Wu B, Chen Y, et al. Genetic Variation and Association Mapping of Seed-Related Traits in Cultivated Peanut (*Arachis hypogaea* L.) Using Single-Locus Simple Sequence Repeat Markers. *Front Plant Sci*. 2017; 8: 2105. <https://doi.org/10.3389/fpls.2017.02105> PMID: 29321787
44. Cai D, Xiao Y, Yang W, Ye W, Wang B, Younas M, et al. Association mapping of six yield-related traits in rapeseed (*Brassica napus* L.). *Theor Appl Genet*. 2014; 127(1): 85–96. <https://doi.org/10.1007/s00122-013-2203-9> PMID: 24121524
45. Mireau H, Cosset A, Marechal-Drouard L, Fox TD, Small ID, Dietrich A. Expression of Arabidopsis thaliana mitochondrial alanyl-tRNA synthetase is not sufficient to trigger mitochondrial import of tRNA^{Ala} in yeast. *J Biol Chem*. 2000; 275(18): 13291–13296. <https://doi.org/10.1074/jbc.275.18.13291> PMID: 10788435
46. Wagner S, Behera S, De Bortoli S, Logan DC, Fuchs P, Carraretto L, et al. The EF-Hand Ca²⁺ Binding Protein MICU Choreographs Mitochondrial Ca²⁺ Dynamics in Arabidopsis. *Plant Cell*. 2015; 27(11): 3190–3212. <https://doi.org/10.1105/tpc.15.00509> PMID: 26530087