

# DNA barcode reference library construction and population genetic diversity and structure analysis of *Amomum villosum* Lour. in Daodi production area

Zhang Danchun (✉ [13570254463@139.com](mailto:13570254463@139.com))

China Medical College Hospital: China Medical University Hospital <https://orcid.org/0000-0003-4045-0682>

**Xiaoxia Ding**

2nd Clinical Medical College of Guangzhou University of Chinese Medicine: Guangdong Provincial Hospital of Traditional Chinese Medicine

**Wan Guan**

2nd Clinical Hospital of Guangzhou Chinese Traditional Medicine College: Guangdong Provincial Hospital of Traditional Chinese Medicine

**Juan Huang**

2nd Clinical Hospital of Guangzhou Chinese Traditional Medicine College: Guangdong Provincial Hospital of Traditional Chinese Medicine

**He Su**

2nd Clinical Hospital of Guangzhou Chinese Traditional Medicine College: Guangdong Provincial Hospital of Traditional Chinese Medicine

**Xiaohui Qiu**

Guangdong provincial key Laboratory of Clinical Research on Traditional Chinese Medicine Syndrome

**Zhi hai Huang**

2nd Clinical Hospital of Guangzhou Chinese Traditional Medicine College: Guangdong Provincial Hospital of Traditional Chinese Medicine

**Lu Gong**

2nd Clinical College of Guangzhou University of Chinese medicine

---

## Research

**Keywords:** A. villosum, molecular markers, DNA barcode, ISSR, genetic diversity

**Posted Date:** December 23rd, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-133546/v1>

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

---

# Abstract

**Background:** The *Amomum villosum* has the situation that it is inferior and other other varieties are used as *A. villosum* in the market. In order to develop and utilize the genuine medicinal materials *A. villosum*, this experiment aims to carry out the identification and research of variety of the *A. villosum* and analyze its genetic diversity, constructing the DNA barcode database of the genuine medicinal materials *A. villosum* in Guangdong Province and providing recommendations for populations planting, which will be critical to the further research of *A. villosum*. (2) **Methods:** A total of 141 samples of *A. villosum* were analyzed by DNA barcoding to construct DNA barcode database. The genetic diversity of *A. villosum* sampled from 7 populations in Guangdong Province was detected based on ISSR molecular marker technology. (3)

**Results:** The success rates of PCR amplification and sequencing of five barcodes of *A. villosum* was *rbcL* > ITS > ITS2 > *psbA-trnH* > *matK*. 141 samples of *A. villosum* from 7 populations in Guangdong Province were used to construct a reference DNA barcode database containing 531 sequences. The results of genetic diversity were as follow, the number of alleles  $N_a$  ranged from 1.2879 to 1.7121, the effective number of alleles  $N_e$  ranged from 1.1848 to 1.4240, the gene diversity index (H) ranged from 0.2536 to 0.1117, and the Shannon index (I) ranged from 0.3816 to 0.1658, which indicated the genetic diversity of *A. Villosum* was rich. The total genetic diversity among the 7 populations ( $H_t$ ) was 0.3299, the genetic diversity within the populations ( $H_s$ ) was 0.1819, and the gene differentiation coefficient ( $G_{st}$ ) was 0.4487. AMOVA showed that the genetic variation within the populations and the genetic variation between the populations accounted for 68.74% ( $P < 0.05$ ) and 31.26% ( $P < 0.05$ ) respectively, indicating that the genetic variation of *A. villosum* was mainly within the populations. The gene flow  $N_m$  was 0.6143. The genetic distance of the 7 populations was 0.0844 - 0.3347, and the genetic identity was 0.7156 - 0.9191, confirming that the genetic relationship of each population was relatively close. The 7 populations were significantly grouped in the cluster analysis and the genetic level of each population from high to low was as follow: ZY (National Highway Roadside) > ZJD (Zhongjiaodong) > GY (Geopark) > MM (Dianbai) > YC (Dadong Village) > XFC (Xingfu Village) > TK (Tankui Village). There was no correlation between the geographic distance and the degree of genetic differentiation among populations.

**Conclusion:** By amplifying and sequencing five barcodes of ITS2, *psbA-trnH*, ITS, *matK* and *rbcL*, a reference DNA barcode database of *A. villosum* with 531 sequences was constructed. The results of genetic diversity showed that it is necessary to take appropriate in situ protection measures for the populations of *A. villosum* in Yangchun City and increase the genetic exchange between populations to improve the genetic diversity of *A. villosum*.

## 1. Introduction

*Amomum villosum* Lour. is a medicinal plant of Zingiberaceae family mainly grown in southern China. It's ripe and dried fruit Fructus Amomi is the famous traditional Chinese medicine (TCM) with the effects of dampening appetite, warming the spleen to stop diarrhea, regulating qi and relieving the fetus and has

been used for thousands of years. Modern pharmacological studies show that, Fructus Amomi has great activities of anti-ulceration, anti-diarrhea, anti-inflammatory and antimicrobial[1]. In addition, Fructus Amomi is also used in food, liquors, and tea as the health product and condiment. Yangchun City located in Guangdong Province is considered the Daodi (genuine) production area of Fructus Amomi for its high quality. With the rapid development of the city as well as traditional Chinese medicine business, the habitat of *A. villosum* has been frequently destroyed that seriously threatened its germplasm resources. In 2016, Fructus Amomi was selected as one of the eight legally protected TCM variety in Guangdong Province.

The genetic diversity of a species is the basis for its survival and evolution, which is of great significance to the analysis of evolutionary polymorphism, genetic relationship, optimization of germplasm resources and protection of populations[2]. Polymerase Chain Reaction (PCR)-based molecular markers have been widely used in the analysis of plant genetic diversity[3]. Among them, Inter-Simple Sequence Repeat (ISSR) is a fast and efficient marker, with the characteristics of high polymorphism, high reliability, low cost, and does not require pre-determination of target sequence information[4, 5], which is widely used in germplasm identification and genetic diversity analysis[6]. Another marker, DNA barcodes, proposed in 2003 can not only be used in biological identification, but also in genetic diversity analysis[7].

In the current study, ISSR and DNA barcoding markers were used to investigate the genetic diversity of seven populations of *A. villosum* in its Daodi production area, Yangchun City of Guangdong Province. Five barcodes ITS2, *psbA-trnH*, ITS, *matK* and *rbcL* were amplified and DNA barcode reference libraries were constructed. This study will provide insights into the identification, conservation, domestication, and breeding of *A. villosum*.

## 2. Results

### 2.1. DNA barcode reference library construction

We extracted the genomic DNA from 141 samples of *A. villosum*. The OD<sub>260/280</sub> was 1.76–1.98 for all the DNA samples and the concentration was 73.70–1294.80 ng/μL. Five DNA barcodes of all the samples were amplified and sequenced. The PCR amplification and sequencing results were shown in Table 1. It showed that the success rate of sequencing for each barcode was *rbcL* (100.00%) > ITS (98.58%) > ITS2 (95.04%) > *psbA-trnH* (53.90%) > *matK* (29.08%). The ranking of the success rate of PCR amplification was consistent with that of sequencing. Thus, these sequences constructed the DNA barcode reference library of *A. villosum* in Guangdong Province.

Table 1  
DNA amplification and sequencing results of *A. villosum*

	sequence/ bar	PCR amplification success rate (%)	sequencing success rate (%)
ITS2	134	95.74	95.04
<i>psbA-trnH</i>	76	57.45	53.90
ITS	139	99.29	98.58
<i>matK</i>	41	29.08	29.08
<i>rbcL</i>	141	100.00	100.00

We analyzed the sequences of each barcode we obtained. Sequence characterization for each barcodes was shown in Table 2. All the sequences of the five barcodes had no variation sites, showing a strong conservation. ITS2 had the shortest sequence length and the highest GC content. The sequences of each barcode were shown in Fig. 1 to Fig. 5.

Table 2  
Analysis of the sequences of DNA barcode of *A. villosum*

	average sequence length /bp	GC content /%	conservative sites	variation site /Information site
ITS2	229	59.83	229	0
<i>psbA-trnH</i>	667	28.49	667	0
ITS	653	56.20	653	0
<i>matK</i>	787	29.22	787	0
<i>rbcL</i>	729	41.84	729	0

## 2.2.ISSR polymorphism and genetic diversity

In the present study, ISSR-PCR fragments of *A. villosum* populations ranged from 300 bp to 2000 bp. An example of amplified profiles was shown in FigueS1. The samples for ISSR analysis was shown in TableS1. A total of 66 ISSR bands were generated from 7 *A. villosum* populations by the selected 6 primers. Among them, 56 bands were polymorphic, thus the percentage of polymorphic bands was 84.85% for all the 7 *A. villosum* populations. For each primer, it amplified 6–14 bands with the polymorphic ratio was 64.29%-100.0%. The results were shown in Table 3.

Table 3  
ISSR banding patterns of 7 *A. villosum* populations.

primers	sequences	Tm(°C)	amplification bands	polymorphic bands	PPB(%)
UBC808	(AG) <sub>8</sub> C	52.00	11	11	100.00
UBC817	(CA) <sub>8</sub> A	49.24	11	10	90.91
UBC825	(AC) <sub>8</sub> T	52.00	12	12	100.00
UBC840	(GA) <sub>8</sub> YT	52.74	12	10	83.33
UBC866	(CTC) <sub>6</sub>	56.60	6	4	66.67
UBC889	DBD-(AC) <sub>7</sub>	46.83	14	9	64.29
sum			66	56	84.85
mean			11	9.33	84.82

Genetic diversity information of *A. villosum* populations analyzed by Popgene 32 was shown in Table 4. Among the populations of *A. villosum*, population ZY had the highest genetic diversity (PPB = 71.21%, Na = 1.7121, Ne = 1.4240, H = 0.2536, I = 0.3816) while population TK had the lowest genetic diversity (PPB = 28.79%, Na = 1.2879, Ne = 1.1848, H = 0.1117, I = 0.1658). At the species level, the value of Na, Ne, H, I and PPB were 1.4834, 1.3115, 0.1820, 0.2689 and 47.19, respectively.

Table 4  
Genetic diversity of *A. villosum*

	NO.	Na	Ne	H	I	PPB(%)
ZJD	5	1.5303	1.3492	0.1998	0.2954	53.03
TK	4	1.2879	1.1848	0.1117	0.1658	28.79
ZY	4	1.7121	1.4240	0.2536	0.3816	71.21
XFC	7	1.4394	1.2685	0.1624	0.2434	43.94
GY	5	1.4848	1.3516	0.1998	0.2897	48.48
MM	5	1.4901	1.3273	0.1818	0.2604	40.91
YC	3	1.4394	1.2753	0.1648	0.2460	43.94
mean	-	1.4834	1.3115	0.1820	0.2689	47.19

## 2.3. Genetic structure of *A. villosum* populations

Genetic structure information of *A. villosum* populations analyzed by Popgene 32 was shown in Table 5. Total genetic diversity (Ht) of the 7 populations was 0.3299, while the within population genetic diversity was 0.1819. Gene differentiation coefficient (Gst) was 0.4487, indicating that 55.13% of the genetic variation was existed within populations. The result was similar to that of molecular variance analysis

(AMOVA), which showing 68.74% ( $P = 0.001$ ) genetic variation was in within populations while 31.26% ( $P = 0.001$ ) was between populations (Table 6). Additionally, the gene flow ( $N_m$ ) among different populations was 0.4487.

Table 5  
Genetic structure of *A. villosum*

samples size	genetic diversity of the total populations ( $H_t$ )	genetic diversity within populations ( $H_s$ )	gene differentiation coefficient ( $G_{st}$ )	gene flow ( $N_m$ )
141	0.3299	0.1819	0.4487	0.6143
$\delta$	0.0272	0.0102		

Table 6  
AMOVA analysis of *A. villosum* based on ISSR

source of variation	degree of freedom /df	mean square deviation /SS	mean square value /MS	variance component	variance component percentage (%)	$P$
among populations	6	167.835	27.972	3.718	31.26	0.001
within populations	31	253.455	8.176	8.176	68.74	0.001

## 2.4. Genetic distance, genetic identity and cluster analysis

Genetic distance and genetic identity are the main indicators to examine the degree of genetic differentiation and the relationship between groups [8]. The genetic distances of 7 populations were between 0.0844 and 0.3347, and the genetic identity were between 0.7156 and 0.9191 (Table 7). Among them, the smallest genetic distance was between ZJD and TK population (0.0844), and the largest one was between XFC and YC population (0.3347). Mantel test carried out with NTSYS-pc 2.0 indicated that the genetic distance and geographical distance were not significantly correlated ( $r = 0.02698$ ,  $P = 0.5504$ ) (Fig. 6).

Table 7  
Genetic distance and genetic identity of *A. villosum* populations

	ZJD	TK	ZY	XFC	GY	MM	YC
ZJD	****	0.9191	0.9073	0.7908	0.7977	0.7569	0.7636
TK	0.0844	****	0.8827	0.7177	0.7636	0.7538	0.7620
ZY	0.0973	0.1247	****	0.8222	0.8231	0.8011	0.8045
XFC	0.2347	0.3317	0.1957	****	0.8639	0.7912	0.7156
GY	0.2260	0.2697	0.1946	0.1463	****	0.8930	0.8182
MM	0.2785	0.2827	0.2218	0.2342	0.1132	****	0.8709
YC	0.2697	0.2719	0.2176	0.3347	0.2003	0.1382	****
Note: Genetic identity above diagonal, genetic distance below diagonal							

The UPGMA clustering map of populations based on genetic similarity coefficient was constructed using the data of Nei's genetic identity and genetic distance of 7 populations (Fig. 7). 7 populations were divided into three groups at the similarity coefficient of 0.84. Three populations ZJD, TK and ZY formed one group. Three populations GY, MM and YC formed another group. One population XFC formed a single group. The results of PCoA based on the unbiased pair  $\Phi_{st}$  matrix of Nei were consistent with UPGMA cluster analysis (Fig. 8).

### 3. Discussion

In this study, five DNA barcodes ITS2, *psbA-trnH*, ITS, *matK* and *rbcl* were amplified and sequenced from 141 individuals of 7 *A. villosum* populations and finally 531 sequences were obtained. Thus a local DNA barcode reference library of *A. villosum* in its Daodi production area was constructed. Many DNA barcode reference libraries have been constructed for the purpose of a more rapid and accurate species identification [9]. And DNA barcoding has been used in *A. villosum* identification [10–12]. COI was an efficient species identification tool and frequently used in genetic diversity analysis of animals [13–17]. In plants, however, low substitution rates of mitochondrial DNA have made it unsuitable, and some other barcoding regions were searched as alternatives just like we used in the current study [18–20]. An ideal DNA barcode should be easily retrievable and bidirectionally sequenced, and provide maximal discrimination among species [21]. Among the five barcodes, the highest PCR amplification and sequencing success rate for *A. villosum* was *rbcl*. For a more comprehensive assessment of discrimination power of DNA barcodes, there needs more investigations of the DNA barcodes with adulterants of *A. villosum* included.

We aligned the sequences within the DNA barcodes, and no variation sites of the five barcodes were found. Therefore, genetic diversity could not be analyzed by these DNA barcodes. The main reason for Intra-specific divergence could occur at a very high rate within geographically isolated populations [22]. But genetic diversity at the sub-specific level is best explored with a multi-locus approach such as

fingerprinting techniques [23]. Consequently, we may say that ITS method is more suitable for genetic diversity analyses of population with wide geographic areas[24].

For species identification, geographic structure and plant diversity of the populations might be the problems for the barcoding approach and these problems have to be dealt at the library construction stage[25]. Here, no polymorphisms were examined in any DNA barcode sequence indicating that these DNA barcodes were suitable for *A. villosum* identification. How much variation is actually needed to separate species is not known with certainty because intra-specific sampling has generally been limited to narrow geographic locales[26].

Molecular markers have been used extensively to determine genetic diversity and genetic relationships in plant science[27, 28]. Moradkhani et al.[29] expressed that the ISSR marker was used among marker systems as a desirable marker in a wide range of genetic variations in various plants. In the ISSR marker, most amplified fragments were between 200 and 1500 bp.

The richer the genetic diversity of species, the stronger the ability of species to adapt to the natural environment. The level of genetic diversity of plants can be influenced by a variety of factors, including the breeding system, the mechanism of seed transmission, geographical distribution, and natural selection[30]. The species with high genetic variation can resist the survival pressure caused by various environmental changes. The loss of genetic diversity will reduce the ability of species to adapt to environmental changes and affect the survival ability of species. Therefore, it is helpful to evaluate the genetic diversity level of *A. villosum* populations to analyze its evolutionary potential and provide reference for the identification, preservation and utilization of the germplasm resources of *A. villosum*. This provides a reference for increasing the genetic diversity and planting of *A. villosum* in different populations.

According to the ISSR markers, there was a high level of inter-population genetic variability and a relatively low level of genetic diversity within populations. Then we analyzed the genetic diversity and genetic structure of *A. villosum* populations by ISSR marker through the whole genome. Genetic diversity parameters indicated that the genetic diversity of germplasm materials of *A. villosum* in Yangchun was relatively rich (PPB = 47.19%, H = 0.1820, I = 0.2689). The results of G<sub>st</sub> (55.13%) and AMOVA (68.74%) showed that more genetic variations existed in within populations.

Our results show that ISSR markers can effectively reveal the polymorphism among materials. Genetic diversity also differed somewhat within the 7 populations in this study. ISSR markers indicated that population ZY had the highest genetic diversity and TK had the lowest. This variation may be due to human activity, random genetic drift and/or inbreeding variation.

The higher the genetic differentiation index between populations, the more obvious the differentiation between populations and the more genetic difference between populations. Wright et al.[31] believed that the G<sub>st</sub> value of genetic differentiation coefficient is between 0 and 0.05, and the genetic differentiation of the populations is weak; between 0.05 and 0.15, the genetic differentiation of the populations is moderate;

between 0.15 and 0.25, meaning a large genetic differentiation of the populations; when the  $G_{st}$  value is higher than 0.25, the differentiation is extremely large. According to Nei's analysis of genetic diversity, the  $G_{st}$  value among *A. villosum* populations was 0.448, which was greater than 0.25, finding that the genetic differentiation between the populations was extremely large. AMOVA analysis showed that genetic variation within populations of *A. villosum* accounted for 68.74% ( $P < 0.05$ ) of the total genetic variation in the populations, and genetic variation among populations accounted for 31.26% ( $P < 0.05$ ), indicating that most of the genetic variation of *A. villosum* occurred within the populations. Gene flow is the movement of genes within and between populations, and its intensity has an important effect on populations differentiation. In this study, the gene flow  $N_m$  between different populations was 0.6143. According to Slatkin[32], the fraction of  $N_m > 1$  between any population reflects that it is resistant to the influence of genetic drift, with sufficient communication and no obvious differentiation, so it can prevent populations segmentation. Therefore, the  $N_m$  value of 0.6143 indicated that genetic drift was the main factor affecting the genetic variation between populations and the genetic communication between populations is difficult, and high genetic variation was maintained within the population, which can be considered as an independent population. It was also confirmed by the genetic differentiation index between populations. The similarity and genetic relationship between plant populations can be expressed through genetic distance. Some scholars believe that genetic distance and geographic distance have a positive correlation[33]. While some scholars also believe that geographic distance and genetic distance are not significantly correlated[34].

The genetic distance of 7 populations was between 0.0844 and 0.3347, and the genetic similarity coefficient was between 0.7156 and 0.9191, indicating that the kinship of the populations was relatively close. If gene flow and seed transmission through the mating system are the main causes of populations variation, the closer the geographic distance between populations, the smaller the genetic differentiation. However, the result of the Mantel test indicated that the distribution of genetic diversity among populations may not be explained by obvious geographic distance and this results can be explained by enhancing the geographical distribution of gene flow, therefore, we can analyze the results through the grouping situation generated by the PCoA diagram and UPGMA. The populations in this study were clustered according to the similarity of habitats, regardless of geographic location, and populations with similar habitats are clustered together first, and 7 populations were clearly divided into three major clusters: ZJD, TK and ZY were clustered together, GY, MM and YC were clustered together, while XFC was a separate cluster. It could be an introduction problem. Isolates in different groups had a similarity range of 78–92%, this high level of genetic diversity can be obtained through a series of evolutionary processes, including mutation, recombination and migration. The genetic level among populations was consistent with the results of PCoA and UPGMA cluster analysis revealed by ISSR markers.

In this study, ISSR marker technology is used to preliminarily analyze the genetic relationship of 7 populations by analysing band polymorphism, populations polymorphism, populations clustering based on genetic distance, populations genetic distance and genetic consistency, clustering result and PCoA, which can all verify that 7 populations have certain genetic diversity and ability to resist external invasion, which is undoubtedly a good news for the protection and cultivation of germplasm. It provides a

theoretical basis for further research on the classification of *A. villosum* populations and lays a theoretical foundation for the protection and sustainable utilization of germplasm resources of Southern Medicine *A. villosum*.

## 4. Materials And Methods

### Plant Material Sampling

A total of 141 samples of *A. villosum* were collected from 7 populations in Yangchun City, Guangdong Province from August to November 2018. The sampled plants were identified by Huang Zhihai, the chief Chinese pharmacist of the Second Clinical College of Guangzhou University of Chinese Medicine. Fresh and healthy leaves were moved from the plants, dried and preserved in silica gel right now in the field, then stored in an ultra-low temperature refrigerator (-80 °C) when came back to the laboratory. Detailed information and geographic location of samples see Table 8 and Fig. 9.

Table 8  
Sampling information of *A. villosum* from different populations in Guangdong Province

No.	Pop.	Serial No.	longitude /N	Latitude /E	samples size
1	ZhongjiaoDong	ZJD	112°04' 01"	22°24' 16"	33
2	Tankui Village	TK	112°04' 30"	22°24' 22"	16
3	G325 National Roadside,	ZY	112°03' 59"	22°24' 26"	25
4	XingfuVillage,	XFC	112°00' 41"	22°21' 54"	18
5	National Geopark,	GY	111°49' 51"	22°35' 19"	19
6	Dianbai District, Maoming	MM	111°12' 21"	21°45' 52"	10
7	Datong Village	YC	111°58' 48"	22°23' 25"	20

### DNA extraction, PCR amplification and sequencing

The total DNA was extracted using Tiangen DP305 plant DNA kit method. The NanoDrop2000 ultra-micro ultraviolet spectrophotometer was used to determine the DNA concentration and purity.

The PCR amplification reaction system of the experiment contained 2×Taq PCR Mix 12.5 μL, forward primer (2.5 μM) 1.0 μL, reverse primer (2.5 μM) 1.0 μL, genomic DNA 2.0 μL and added up to 25 μL with ddH<sub>2</sub>O. The primer sequences and amplification conditions of different DNA barcodes were shown in Table 9. All amplification reactions were completed on the ProFlex PCR instrument (Life Technologies, USA). PCR products were sent to Shanghai Meiji Biotechnology Company Guangzhou Branch to be sequenced.

Table 9  
Primers informations and PCR amplification conditions

primers		base(5'-3')	amplification conditions
ITS2	2F	ATGCGATACTTGGTGTGAAT	94°C,5 min
	3R	GACGCTTCTCCAGACTACAAT	94°C,30s;56°C×30s×72°C×45 s. 35 circles 72°C,10 min
<i>psbA-trnH</i>	fwd PA	GTTATGCATGAACGTAATGCTC	94°C,4 min
	rev TH	CGCGCATGGTGGATTACAATCC	94°C,30s;55°C×1min×72°C×1 min. 35 circles 72°C,10 min
<i>matK</i>	3F_KIM	CGTACAGTACTTTTGTGTTTACGAG	94°C,1 min
	1R_KIM	ACCCAGTCCATCTGGAAATCTTGGTTC	94°C,30s;52°C×20s×72°C×50 s. 35 circles 72°C,5 min
<i>rbcL</i>	1F	ATGTCACCACAAACAGAAAC	95°C,4 min
	724R	TCGCATGTACCTGCAGTAGC	94°C,1min;55°C×1min×72°C×2 min. 38 circles 72°C,7 min
ITS	5F	GGAAGTAAAAGTCGTAACAAGG	94°C,5 min
	4R	TCCTCCGCTTATTGATATGC	94°C,1min;50°C×1min×72°C×1 min30s. 30 circles 72°C,7 min
ISSR-PCR amplification system			

100 ISSR universal primer sequences published by Columbia University were screened[35]. And 6 primers that produced clear and reproducible banding patterns were selected (Table 5). ISSR-PCR amplifications were performed on 141 individuals from 7 *A. villosum* populations using the selected primers in a 20µL reaction volume including 10µL2X PCR Mix (containing dye, MgCl<sub>2</sub>, dNTPs), 2µL template DNA, 1µLISSR primer and 7µL ddH<sub>2</sub>O. The ISSRPCR amplification procedures were programmed in the ProFlexrmocycler as follows: predenaturation at 94 °C for 5 min, 35 cycles of denaturation at 94 °C for 45 s, annealing at the T<sub>m</sub> value of the primer for 45 s, extension at 72 °C for 2 min, with a final extension at 72 °C for 5minand preservation at 4 °C. ISSR-PCR products were separated on a 2% agarose gel stained with Goldview by electrophoresis in 1 × TAE buffer at 80 V. The gels were visualized under UV light and photographed with Bio-Rad ChemiDoc imagingsystem (USA). The molecular weights of ISSR-PCR products were estimated using a 100 bp Plus DNA Ladder (TIANGEN Biotech Co., Ltd., Beijing, China).

# Data analysis

The two-way sequenced peaks of DNA barcodes were evaluated and assembled by CondonCode Aligner v8.0.1 software. Low-quality areas at both ends of the assembled sequences were removed. ITS2 barcodes were annotated by cutting off the conserved 5.8S and 28S motifs based on HMM[23] and the ITS2 database[24]. Mega6.0 software was used to align DNA barcode sequences and calculate sequence statistics including the base composition ratio, GC content, heterotopic site information, conservative site and parsimony informative sites. Haplotype sequences for each barcodes were exhibited in the two-dimensional code picture. In the picture, each vertical line represented a base, and the two-dimensional code on the right could be scanned directly read the DNA sequence.

Reproducible ISSR-PCR bands were determined with the help of the GelPro32 software and manual correction. These clear bands were scored as either present (1) or absent (0), thus generating an ISSR phenotype data matrix. And the data matrix was imported in Popgene32 software to analyze genetic diversity and genetic structure. Genetic diversity parameters included percentage of polymorphic sites (PPB), number of alleles (Na), effective number of alleles (Ne), Nei's gene diversity index (H) and Shannon's polymorphism information index (I) were calculated. Genetic structure parameters included Nei's gene differentiation coefficient (Gst), total population genetic diversity (Ht), intra-group genetic diversity (Hs) and gene flow (Nm) were calculated. GenAlEx 6.502 software was used to estimate the components of genetic variance within and among populations by analysis of molecular variance (AMOVA) and to assess the correlation between population genetic distance and geographic distance by Mantel tests. Genetic distance and genetic similarity coefficient among populations were calculated and a UPGMA dendrogram was constructed by using NTSYS 2.10e.

## 5. Conclusion

A total of 141 samples of *A. villosum* from 7 populations in Guangdong Province were used to construct a reference DNA barcode reference library containing 531 sequences. On the other hand, the 7 populations were significantly grouped in the cluster analysis and the genetic level of each population from high to low was as follows: ZY > ZJD > GY > MM > YC > XFC > TK. Based on the above research results, the following planting recommendations for *A. villosum* are proposed: priority should be given to the populations ZY, ZJD and GY with rich genetic diversity, in order to preserve as much genetic diversity as possible. At the same time, considering the significant genetic differentiation between *A. villosum*, in-situ conservation should be stepped up for every existing population, and it is recommended to add a protected area in Yangchun City.

A high level of genetic diversity is very important for the long-term survival of the species. *A. villosum*, which is used for both medicine and food from Zingiberaceae, has a very narrow distribution range due to its special requirements for the growth environment. This study found that the genetic diversity of *A. villosum* is relatively rich. It also confirms from the molecular level that the Yangchun area is the origin of *A. villosum*. But currently, *A. villosum* is facing many problems such as high incidence of pests and

diseases, unstable yield, lack of cultivation management, difficulty in breeding seedlings, and so on. Based on the results of this research, establish a germplasm resource nursery for *A. villosum*, extensively collect germplasm resources, and carry out research on excellent germplasm selection, seedling breeding, and high-yield and high-quality cultivation techniques of *A. villosum*, which is crucial to the protection and utilization resources of *A. villosum*.

## Declarations

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:**The authors agree to publish this research.

**Availability of data and materials:**The extracted features datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. All the other data generated or analyzed during this study are included in this article.

**Competing of interests:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Funding:** This work was supported by project 20184013, Demonstration Research on precise powder decoction pieces in traditional Chinese Medicine from Traditional Chinese Medicine Bureau of Guangdong Province of China; 002009, Quality standard system construction for the whole industry chain of Chinese medicinal decoction pieces from Guangdong Provincial Drug Administration of China.

**Author's contributions:** Conceptualization, Zhihai Huang; methodology, Xiaohui Qiu.; software, He Su; validation, Wan Guan; formal analysis, Danchun Zhang; investigation, Lu Gong; resources, Zhihai Huang; data curation, Zhihai Huang; writing—original draft preparation, Danchun Zhang; writing—review and editing, Xiaoxia Ding; visualization, He Su; supervision, Zhihai Huang and Juan Huang; project administration, Zhihai Huang; funding acquisition, Zhihai Huang. All authors have read and agreed to the published version of the manuscript.

**Acknowledgements:**Thanks to the experimental platform of Guangdong Provincial Hospital of Chinese Medicine and Guangdong Provincial Key Laboratory of Clinical Research on Traditional Chinese Medicine Syndrome and funders.

## References

1. Therapeutic Effect of Amomum villosum on Inflammatory Bowel Disease in Rats. *Frontiers in Pharmacology* **2018**, *9*.
2. Zhanhuan, S. RESEARCH MEANS OF GENETIC DIVERSITY AND ITS PROTECTIVE MEASURES. *Journal of Ningxia Agricultural College* **2002**.

3. Jun, P.; Fuyao, Z.; Xin, L.; Zhihong, D.; Huiming, L.; Zhaoxiang, T.; Tingting, Y. A Review of DNA Molecular Markers and Its Application in Sorghum Breeding. *chinese agricultural science bulletin* **2011**.
4. Souframanien, J.; Gopalakrishna, T. A comparative analysis of genetic diversity in blackgram genotypes using RAPD and ISSR markers. *Theoretical and Applied Genetics* **2004**, *109*.
5. A., L.; E., T.; T., J.; X., Z.; A., D. A genetic linkage map for watermelon derived from a testcross population: ( *Citrullus lanatus* var. *citroides* x *C. lanatus* var. *lanatus*) x *Citrullus colocynthis*. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* **2002**, *105*.
6. S, A.A.L.; P, C.P.; A, M.M.; P, d.P.C.M.; S, S.F. High degree of genetic diversity among genotypes of the forage grass *Brachiaria ruziziensis* (Poaceae) detected with ISSR markers. *Genetics and molecular research : GMR* **2011**, *10*.
7. Shilin, C.; Xiaohui, P.; Hui, Y.; Jianping, H.; Kun, L. Identification System and Perspective for DNA Barcoding Traditional Chinese Materia Medica. *World ence & Technology* **2011**.
8. Poyraz; Ismail. Comparison of ITS, RAPD and ISSR from DNA-based genetic diversity techniques. *Comptes Rendus Biologies* **2016**.
9. Lou, S.K.; Wong, K.L.; Li, M.; But, P.H.; Tsui, K.W.; Shaw, P.C. An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). *Bmc Genomics* **2010**, *11*, 402.
10. Shu, W.J.; Fen, Y.J.; Ting, Z.R.; Pharmacy, S.O. Identification of *Amomum villosum* Lour. and Its Adulterants Based on *rbcl* Gene. *Lishizhen Medicine and Materia Medica Research* **2018**.
11. Jian-Ping, H.; Mei-Ni, L.I.; Lin-Chun, S. Identification of *Amomi fructus* and its adulterants based on ITS2 sequences. *Global Traditional Chinese Medicine* **2011**.
12. Gui-Fang, Z.; Zhi-Min, Z.; Song, H.; Xiao-Ping, L. Identification and Clustering Analysis of Zingiberaceae with *matK* Barcode. *Lishizhen Medicine and Materia Medica Research* **2018**.
13. Xin, C. PERSPECTIVE OF MOLECULAR BIOLOGICAL TECHNIQUES APPLIED IN INSECT SYSTEMATICS. *Acta Zootaxonomica Sinica* **2000**.
14. Hebert, P.D.N.; Ratnasingham, S.; De Waard, J.R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc Biol* **2003**, *270*, S96.
15. Heng-Guo, H.E. Application of Molecular Marker Techniques in Insect Population Genetics. *Journal of China West Normal University(Natural ences)* **2008**.
16. Mitochondrial DNA Indicates Late Pleistocene Divergence of Populations of *Heteronympha merope*, an Emerging Model in Environmental Change Biology. *Plos One* **2009**, *4*, e7950.
17. Ye, F.Y.; Zhu, C.D.; Yefremova, Z.; Liu, W.X.; Guo, J.Y.; Wan, F.H. Life history and biocontrol potential of the first female-producing parthenogenetic species of *Diglyphus* (Hymenoptera: Eulophidae) against agromyzid leafminers. *Scientific Reports* **2018**, *8*, 3222.
18. Vere, N.D.; Rich, T.C.G.; Trinder, S.A.; Long, C. DNA Barcoding for Plants. *Methods in Molecular Biology* **2015**, *1245*, 101-118.

19. Kress, W.J.; Wurdack, K.J.; Zimmer, E.A.; Weigt, L.A.; Janzen, D.H. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences* **2005**, *102*, 8369-8374.
20. Hollingsworth, M.L.; Clark, A.A.; Forrest, L.L.; Richardson, J.; Pennington, R.T.; Long, D.G.; Cowan, R.; Chase, M.W.; Gaudeul, M.; Hollingsworth, P.M. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* **2010**, *9*, 439-457.
21. Hollingsworth, P.M.; Forrest, L.L.; Spouge, J.L.; Hajibabaei, M.; Little, D.P. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* **2009**, *106*, 12794-12797.
22. Hebert, P.D.N.; Ryan, G.T. The promise of DNA barcoding for taxonomy. *Systematic Biology* **2005**, 852-859.
23. Keller, A.; Schleicher, T.; Schultz, J.R.; Müller, T.; Dandekar, T.; Wolf, M. 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene* **2009**, *430*, 50-57.
24. Ankenbrand, M.J.; Alexander, K.; Matthias, W.; Jörg, S.; Frank, F. ITS2 Database V: Twice as Much. *Molecular Biology & Evolution*, 3030.
25. Sarwat, M. DNA barcoding, microarrays and next generation sequencing: recent tools for genetic diversity estimation and authentication of medicinal plants: Critical Reviews in Biotechnology: Vol 36, No 2. *Critical Reviews in Biotechnology* **2014**.
26. Phillips, J.D.; Gillis, D.J.; Hanner, R.H. Incomplete estimates of genetic diversity within species: Implications for DNA barcoding. *Ecology and Evolution* **2019**.
27. Wei, X.; Mingfang, Z.; Guixia, J.; Xuqing, C.; Qian, W.; Xiuhai, Z.; Ran, D.; University, J.A. ISSR Analysis on Genetic Diversity of Lily Varieties with Different Hybrid Lines. *Molecular Plant Breeding* **2019**.
28. Hai, H. The Application of Molecular Marker Technique to the Studies of Germplasm Resources of Dendrobium. *Biotechnology Bulletin* **2010**.
29. Moradkhani, H.; Mehrabi, A.A.; Etmiran, A.; Pour-Aboughadareh, A. Molecular diversity and phylogeny of Triticum-Aegilops species possessing D genome revealed by SSR and ISSR markers. **2015**.
30. Yu; Ye-Bing; Guo; Qiao-Sheng; Zhao; Wei-Hong; Zhang; Ming-Ming; Shen; Wen-Biao. Analysis of the genetic diversity and population structure of Perinereis aibuhitensis in China using TRAP and AFLP markers. *Biochemical Systematics and Ecology* **2015**.
31. Evolution and the Genetics of Populations. 4: Variability within and among natural populations: By S. Wright. 1978. Chicago & London: The University of Chicago Press. A vol. in-8<sup>th</sup>, x + 582 pp., figs, tabs, bibl., index. Bound £26.25. *Academic Press* **1980**, *9*.
32. Slatkin. Gene flow and the geographic structure of natural populations. *Ence* **1987**, *236*, 787-792.
33. Jinxing, M.A.; Zhuanzhuan, Y.; Jiyu, Z.; Qisheng, F.; Tiemei, W.; Xinshi, L.U.; University, B.F. Analysis of Genetic Diversity of 20 wild Alfalfa Germplasms in Xinjiang with RAPD and ISSR Markers. *Journal of Yunnan Agricultural University(Natural ence)* **2018**.
34. Qiao-Yan, X.; Xi-Yang, H.; Hong, L.I.; Shui-Yuan, J. ISSR Analysis on Genetic Diversity of Medically Important Zanthoxylum nitidum(Roxb.)DC.in Guangxi. *Guangxi ences* **2014**.

## Figures

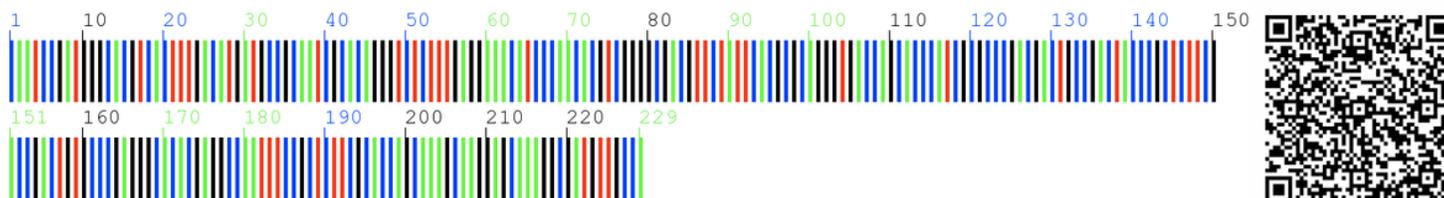


Figure 1

ITS2 sequence of *A. villosum*

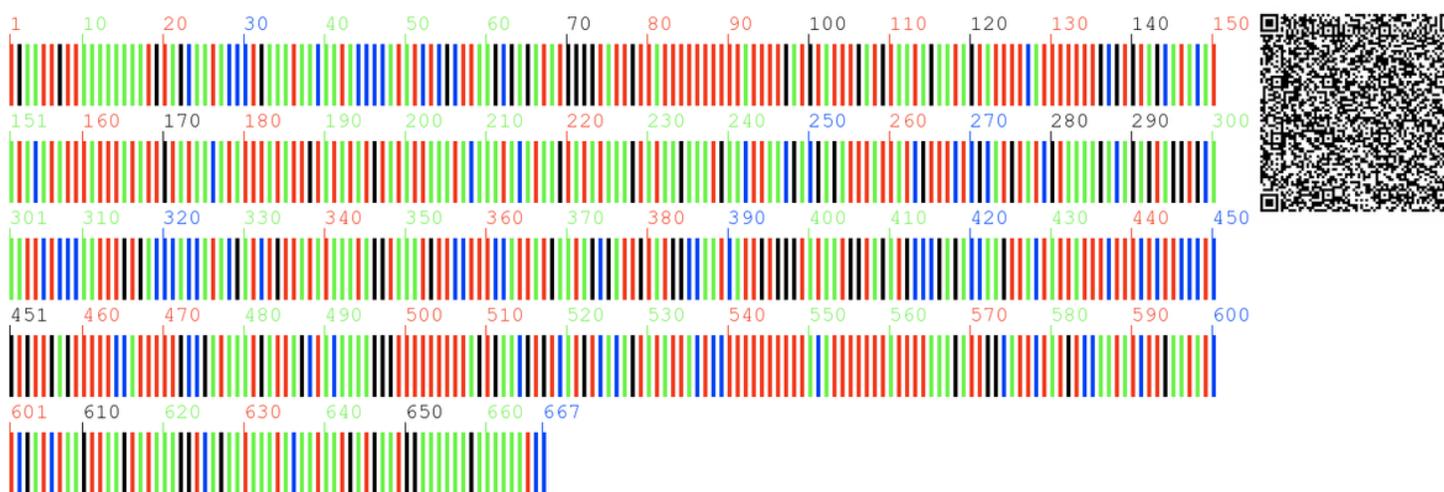


Figure 2

psbA-trnH sequence of *A. villosum*

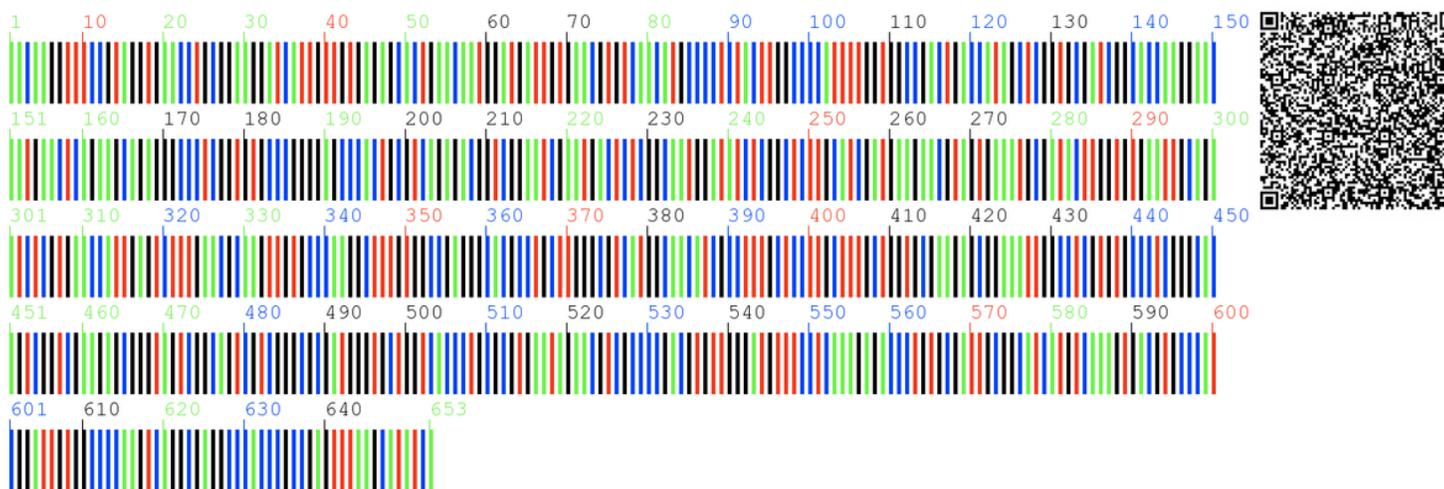


Figure 3

### ITS sequence of *A. villosum*

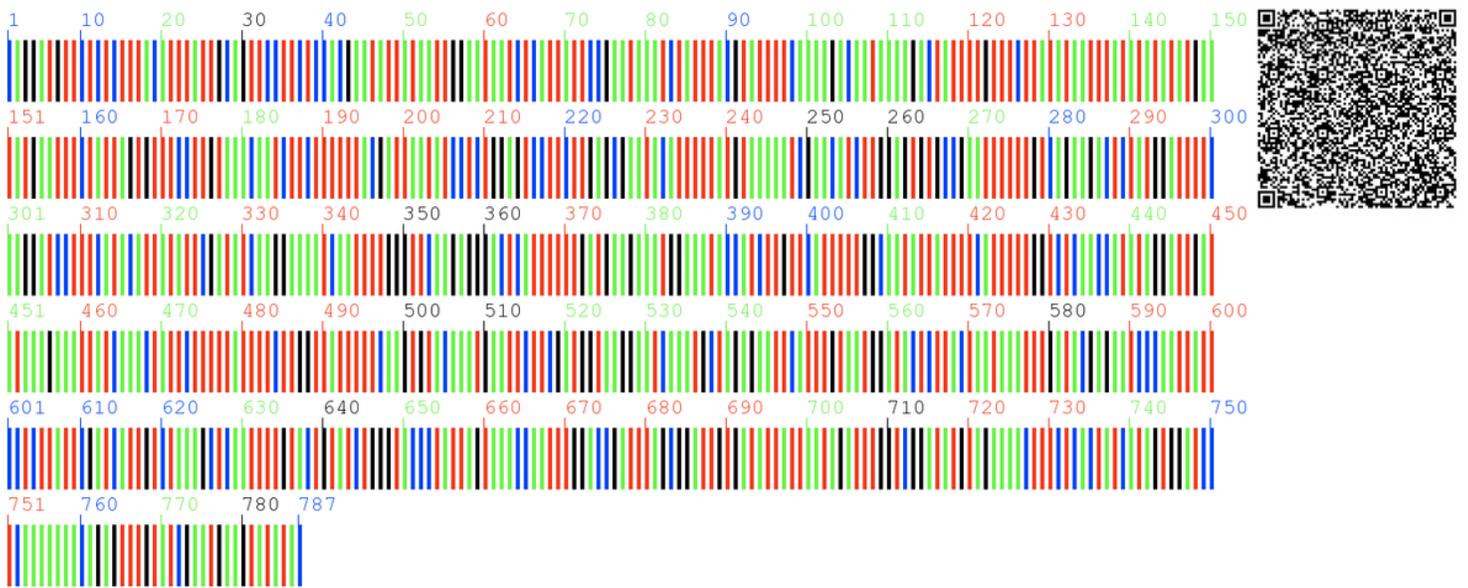


Figure 4

### matK sequence of *A. villosum*



Figure 5

### rbcl sequence of *A. villosum*

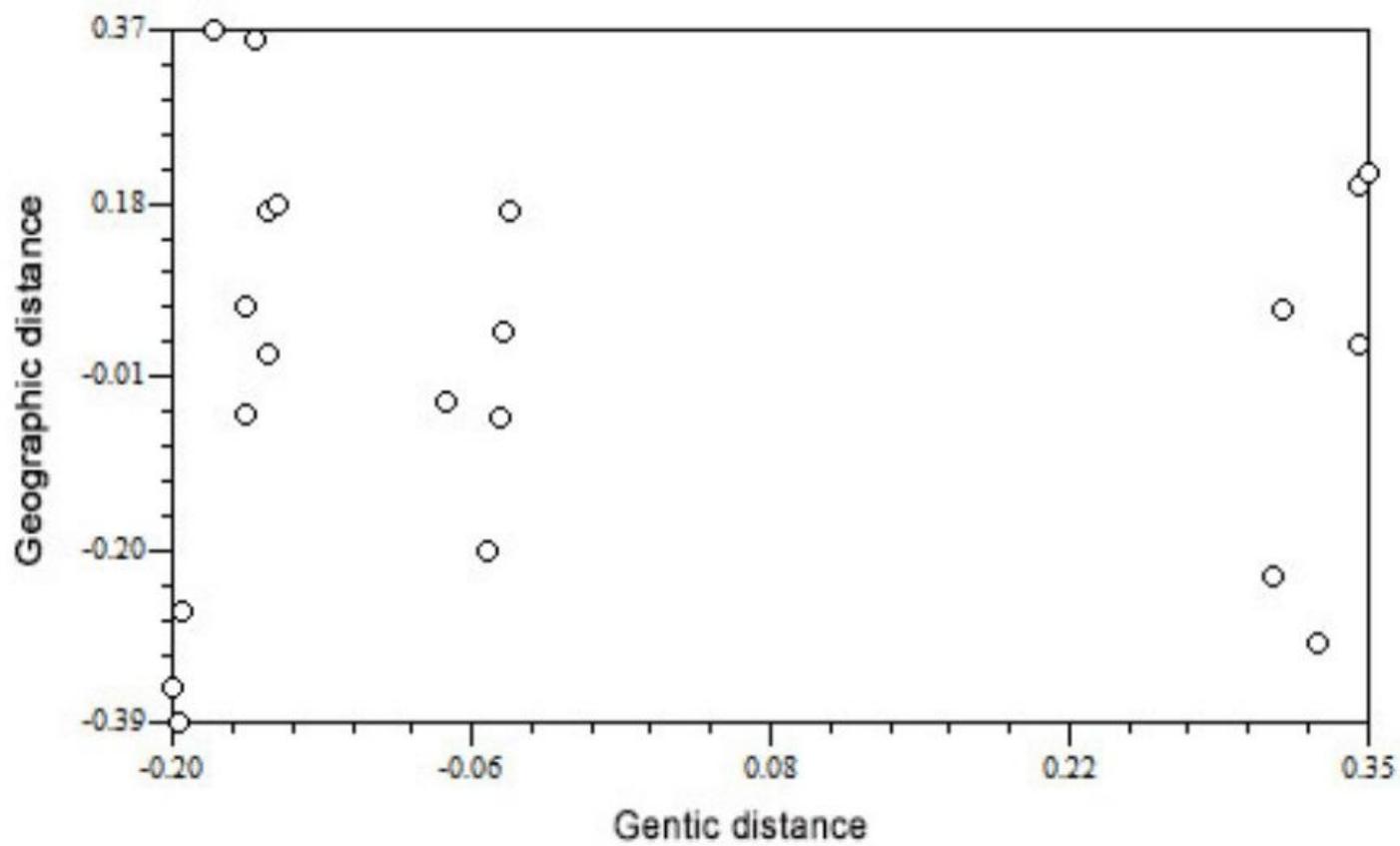


Figure 6

Correlation of geographic distance and genetic distance

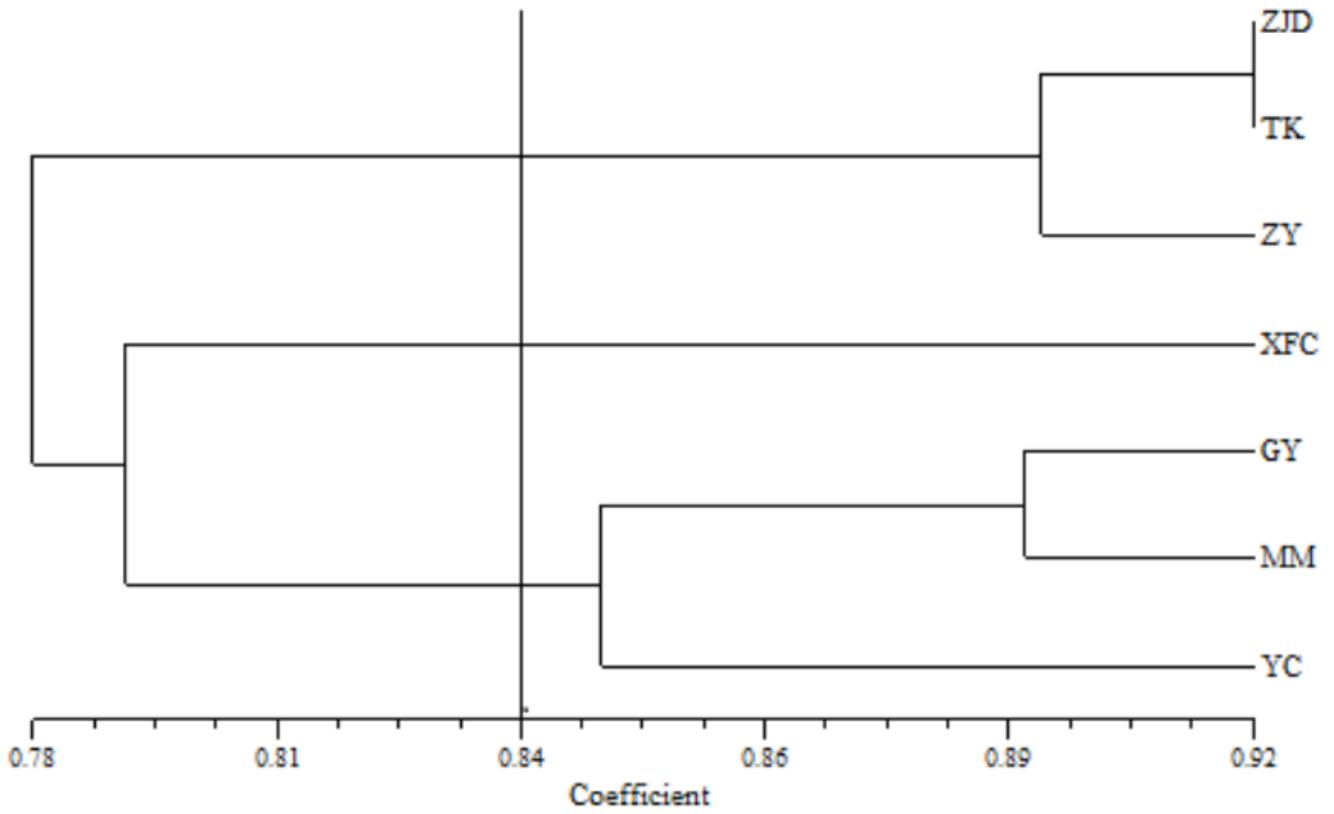


Figure 7

UPGMA clustering map of *A. villosum*

### Principal Coordinates (PCoA)

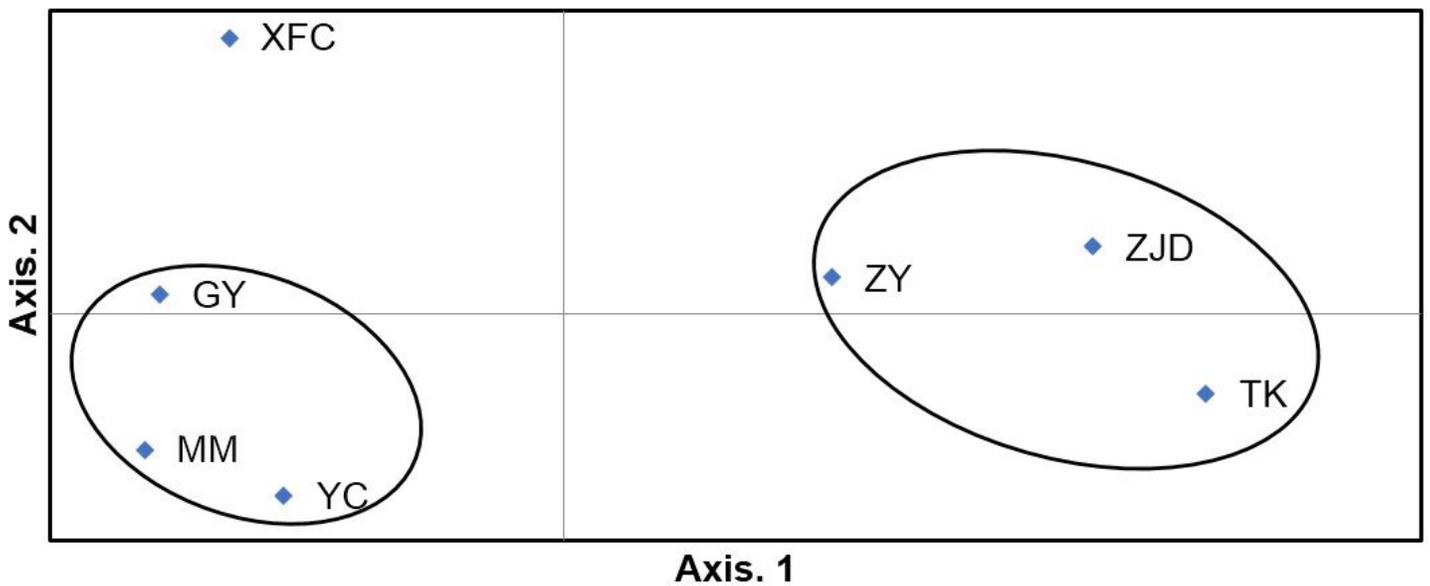
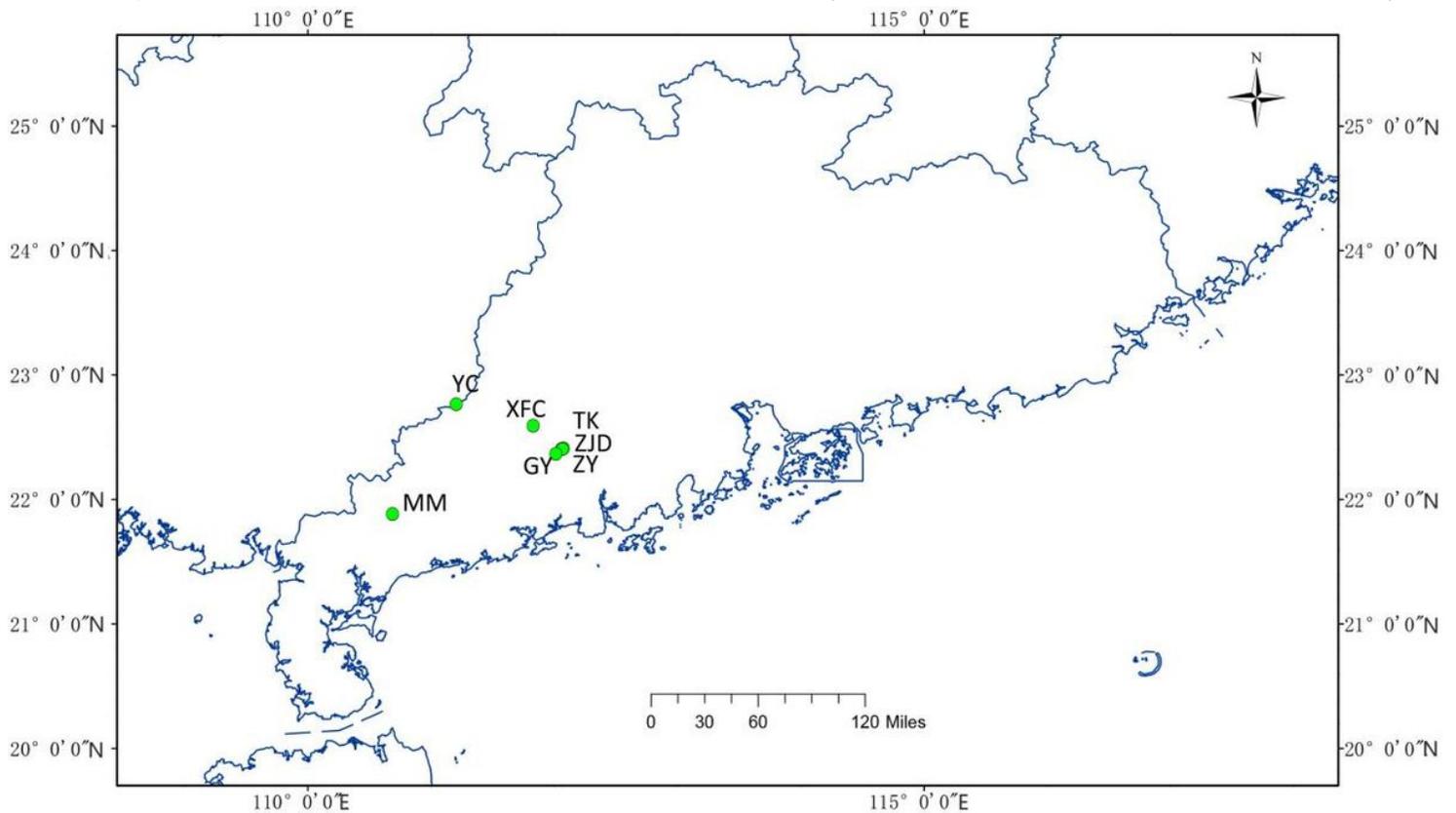


Figure 8

PCoA map of *A. villosum* based on ISSR molecular markers (coord. 1 = 48.04% and coord. 2 = 26.74%)



**Figure 9**

Geographical distribution of collected *A. villosum* populations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

## Supplementary Files

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

- [supplementarymaterial.rar](#)