

Uncovering the genetic diversity of yams (*Dioscorea* spp.) in China by combining phenotypic trait and molecular marker analyses

Tianxu Cao¹, Jingyu Sun¹, Nan Shan¹, Xin Chen¹, Putao Wang¹, Qianglong Zhu¹, Hongyu Zhang¹, Qinghong Zhou¹, and Yingjin Huang¹

¹Jiangxi Agricultural University

November 21, 2020

Abstract

China is one of the native places of yams with abundant representative local varieties. However, the genetic differences between these local varieties remains unclear, thus considerably inhibiting their utilization and development. In this study, 26 phenotypic traits of 112 accessions from 21 provinces in China were evaluated, and 24 simple sequence repeat (SSR) and 29 sequence-related amplified polymorphism (SRAP) markers were used for the genetic diversity analysis. A total of 56 variations of 20 quality traits were detected in 112 accessions, and the most diverse phenotypic trait was stem color. Meanwhile, seven principal components were obtained from 26 phenotypic traits, with a cumulative contribution rate of 69.39%, and all the accessions were divided into six groups by clustering the phenotypic traits. Unweighted pair-group method with arithmetic means and principal coordinate analysis based on SSR-SRAP marker data showed that 112 accessions were also divided into six groups, similar to the result of phenotypic traits but with a slight difference among few accessions. Results of genetic structure analysis showed that 112 accessions could be divided into two groups; one group composed of the accessions of *D. opposita*, the most abundant resource in China, and the other group was a collection of the other accessions. In addition, the analysis of the origin and genetic relationship of yam also indicated that the species of *D. opposita* may have originated in China. These results clarified the genetic differences in yam in China, thereby providing a basis for the identification, evaluation, and conservation of yam resource.

Uncovering the genetic diversity of yams (*Dioscorea* spp.) in China by combining phenotypic trait and molecular marker analyses

Tianxu Cao ^{a,b,#}, Jingyu Sun^{a,#}, Nan Shan^a, Xin Chen^a, Putao Wang^a, Qianglong Zhu^a, Hongyu Zhang^a, Qinghong Zhou^{a*}, Yingjin Huang ^{a,c*}

^a Agronomy College, Jiangxi Agricultural University, Nanchang 330045, P.R.China

^b School of Advanced Agriculture and Bioengineering, Yangtze Normal University, Chongqing 408100, China

^c Key Laboratory of Crop Physiology, Ecology and Genetic Breeding (Jiangxi Agricultural University), Ministry of Education of China, Nanchang 330045, P.R.China

*Correspondence ZHOU Qing-hong, Tel: +86-0791-83828220, E-mail: qinghongzhou@126.com; HUANG Ying-jin, E-mail: yjhuang.cn@126.com

These authors contributed equally to this study

Uncovering the genetic diversity of yams (*Dioscorea* spp.) in China by combining phenotypic trait and molecular marker analyses

Abstract:

China is one of the native places of yams with abundant representative local varieties. However, the genetic differences between these local varieties remains unclear, thus considerably inhibiting their utilization and development. In this study, 26 phenotypic traits of 112 accessions from 21 provinces in China were evaluated, and 24 simple sequence repeat (SSR) and 29 sequence-related amplified polymorphism (SRAP) markers were used for the genetic diversity analysis. A total of 56 variations of 20 quality traits were detected in 112 accessions, and the most diverse phenotypic trait was stem color. Meanwhile, seven principal components were obtained from 26 phenotypic traits, with a cumulative contribution rate of 69.39%, and all the accessions were divided into six groups by clustering the phenotypic traits. Unweighted pair-group method with arithmetic means and principal coordinate analysis based on SSR-SRAP marker data showed that 112 accessions were also divided into six groups, similar to the result of phenotypic traits but with a slight difference among few accessions. Results of genetic structure analysis showed that 112 accessions could be divided into two groups; one group composed of the accessions of *D. opposita*, the most abundant resource in China, and the other group was a collection of the other accessions. In addition, the analysis of the origin and genetic relationship of yam also indicated that the specie of *D. opposita* may have originated in China. These results clarified the genetic differences in yam in China, thereby providing a basis for the identification, evaluation, and conservation of yam resource.

Keywords:Yams; Genetic diversity; Phenotypic trait; Molecular markers; Population structure.

1. Introduction

Yam is considered the common name for plants, with over 600 species of *Dioscorea* genus. The species of yams widely cultivated in the world are *D. bulbifera*, *D. panthaica*, *D. esculenta*, *D. japonica*, *D. trifida*, *D. pentaphylla*, and *D. rotundata* (Lebot, 2009). Yams are also one of the 10 most important edible tuber and root plants in the world and next to potato, cassava, and sweet potato in yield (Shewry, 2003). They play an important role in sustaining many livelihoods in the tropics and subtropics (Tschannen et al., 2005; Wu et al., 2019). In China, yam is known as a medicinal and edible crop with high nutritional and medicinal value; its tuber is rich in starch, protein, and medicinal ingredients (e.g., allinogenin, diosgenin, and dehydroepiandrosterone; Lebot, et al., 2019).

In China, yam resources are extremely rich, with a total of 93 species. *D. opposita*, *D. alata*, *D. persimilis*, *D. fordii*, and *D. quinqueloba* are widely cultivated in China nowadays (Huang, 2011). Yams are cultivated in other provinces, except Qinghai and Tibet, including a large number of local varieties. However, they have long been regarded as “orphan” or “neglected” crops despite their considerable edible and medicinal value and received little attention or investment from researchers (Tamiru et al., 2017). Moreover, *Dioscorea* is mainly dioecious, rarely flowering, and has difficult forming mature seeds (Bressan et al., 2011). The development of medicinal ingredients from a few species of yam has long been emphasized in China (Li et al., 2018; Lebot et al., 2018; Cheng et al., 2020), but the analysis of the resource types and genetic diversity of yam is insufficient. Moreover, several studies only focused on *D. alata*, and the method of resource evaluation is relatively simple (Siqueira et al., 2014; Arnau et al., 2017; Wu et al., 2019). In the long-term cultivation and domestication process, the variation types of yam are complex, and a single classification method is difficult to identify, thereby causing confusion in various records and nomenclatures and even in the classification of some species. These factors seriously hinder the resource conservation and further utilization of yam.

Therefore, studying the genetic diversity, genetic variation, and population structure of yam is considerably important to its origin, distribution, resource utilization, parental selection, and development (Mignouna et al., 2003). So far, phenotypic traits, isozyme, karyotype analysis and DNA diversity have been used to describe the genetic diversity of germplasm (Sartie et al., 2012; Nemorin et al., 2013; Kouam et al., 2018; Ghorbani et al., 2020; Cao et al., 2020). However, research on the genetic diversity of yams has focused on a few species in Africa and the Americas. Anokye et al. (2014) used morphological descriptors to dissect the phenotypic diversity of *D. alata* from Ghana and Nigeria and divided these materials into different clusters independent of geographic origin. Mignouna et al. (2002) evaluated the genetic variation of 45 cultivated yams (*D. cayenensis* / *D. rotundata* complex) from Cameroon by using morphology and isozyme, and the

clustering results of the two were relatively consistent. Silva et al. (2016) used 12 microsatellites and four morphological markers to analyze the genetic diversity and structure of 89 yam (*D. alata*) in Brazil; a high diversity is found in Brazilian yam, but it is not related to geographical location. At present, despite some reports available on the genetic diversity of Chinese yam local varieties, the research methods were too single, the number of molecular markers was small, and the representation of germplasm source regions was poor.

In the present study, large-scale yam resources were collected in 21 provinces, including the main species used in cultivation to fully identify and evaluate the genetic information of yams in China. The genetic diversity, genetic relationship, and population structure of yams were evaluated using the combination analysis of phenotypic traits and two molecular markers (SRAP and SSR). The results could provide a basis for yam classification, breeding, germplasm innovation, utilization, and variety protection.

2. Materials and methods

2.1. Plant materials

A total of 112 accessions were planted in yam germplasm resource garden of Jiangxi Agricultural University (Nanchang City, Jiangxi Province, Supplementary Table 1). Landrace varieties were collected from farmers' fields, institutions, and markets in China, and wild resources were acquired from mountainous regions. Experimental planting was arranged in ridges on April 10, 2019, with 20 cm distance for each individual plant and 1.2 m distance between ridges. Tuber segments (80–120 g) were used as seeds. Individual plants were supported by bamboo stakes. Standard weeding and agronomic measures were applied regularly to provide adequate plant growth conditions.

2.2. Phenotype assessment

A total of 26 phenotypic trait data (Table 1) were evaluated during the vegetative growth of the plants and when tubers were harvested, including 20 qualitative traits and six quantitative traits. The observed values of 20 qualitative traits were repeated six times, and the measured values of six quantitative traits were observed six times (Table 1). All traits were investigated in accordance with the description specification and data standard of Chinese yam germplasm resources (Huang and Huang, 2013; Wang and Shen, 2014).

2.3. DNA extraction

DNA samples of all the accessions were isolated from the young leaves by using a plant genomic DNA kit. The DNA samples were quantified using a Nano Drop 2000 (Wilmington, USA) spectrophotometer and checked on 1.0% (w/v) agarose gels stained with ethidium bromide. On the basis of concentration estimations, all samples were diluted to 20 ng/μL and stored at -20 °C.

2.4. SSR and SRAP genotyping

Twenty-four primers (Supplementary Table 2) selected from recently published 53 SSR primers were used in this study (Narina et al., 2011; Nemorin et al., 2013; Loko et al., 2016). The primers were synthesized by Shanghai Sangon Biotech, Co., Ltd. PCR amplification reactions were performed using a master mix solution of 10 μL containing 5 μL of 2 × Master Mix Blue (TSINGKE, China), 0.25 μL of each primer (10 mM), and 0.75 μL of DNA template (20 ng/μL), and the remaining volume was supplemented with ddH₂O. The following cycling parameters were used in the amplification reaction: first pre-denaturation at 94 °C for 5 min, followed by 40 cycles of 30 s at 94 °C, annealing for 30 s at 54 °C, 30 s at 72 °C, and a final extension of 10 mins at 72 °C. The amplified PCR products were detected on an 8% non-denaturing polyacrylamide gel. Silver nitrate staining was employed, and the image was captured for analysis.

Forty-nine different SRAP primer combinations (Supplementary Table 3), seven forward primers (Li and Quiros, 2001), and seven reverse primers were employed. Twenty-nine primer combinations with good repeatability and high polymorphism were also selected. Each 14 μL of the PCR reaction mixture consisted of 7 μL of 2 × Master Mix Blue, 0.35 μL of each primer (10 mM), and 1.4 μL of DNA template (20 ng/μL), and the remaining volume was supplemented with ddH₂O. PCR amplification was performed under the following

conditions: denaturation at 94 °C for 5 min, five cycles of three steps: denaturation at 94 °C for 1 min, annealing at 35 °C for 1 min, and elongation at 72 °C for 1 min. In the following 30 cycles, the annealing temperature was increased to 56 °C, with the final extension step of 10 min at 72 °C. The amplified products were analyzed through 3% agarose gel electrophoresis prepared in 1 × TBE buffer. The gels were then visualized in a UV transilluminator (Bio-Rad GeL Doc XR+, USA) and photo-documented.

2.5. Data analysis

The 20 qualitative traits were classified, and different values were assigned in accordance with the survey results. The distribution frequency of each classification was also calculated. Then, Shannon diversity index (I) was calculated in accordance with the distribution frequency as follows:

$$I = -\sum_{i=1}^n (p_i)(\ln p_i),$$

where p_i represents the relative frequency of the i th phenotypic class of a trait (Kouam et al., 2018). The maximum, minimum, average, standard deviation (SD), and coefficient of variation (CV) of six quantitative traits were calculated using SPSS 25.0 software. Then, in accordance with the overall average (\bar{x}) and SD (σ), the quantitative trait data were divided into 10 levels, from the first level [$\bar{x} - 2\sigma$] to the 10th level [$\bar{x} + 2\sigma$], and each 0.5σ was a level. Principal component (PC) analysis was carried out with 26 phenotypic indices on SPSS 25.0 software. In accordance with the phenotypic trait survey data, a matrix (1,0) was constructed, and the registration at the i th level of a trait was 1; otherwise, it was 0.

The bands of SRAP and SSR markers were scored for each primer as presence (1) and absence (0) for each locus, and the binary matrix was constructed and statistically analyzed. The allele number (N_a), effective number of alleles (N_e), allele frequency, Nei's (1973) gene diversity index (H), and I of each primer were calculated on POPGENE software version 1.32 (Yeh et al., 1999). The genetic similarity coefficient was evaluated and the principal coordinate analysis (PCoA) was conducted using NTSYS-pc software version 2.10e (Rohlf, 2000). Cluster analysis of the unweighted pair-group method with arithmetic means of the phenotypic traits and molecular markers was performed on MEGA software version 4.1 (Tamura et al., 2007).

The combined data of SRAP and SSR were analyzed via Bayesian model on STRUCTURE software version 2.3.1 to analyze population structure (Pritchard et al., 2000). K (number of clusters) was estimated to be in the range of 2–10, and the software was run three times to determine this value. STRUCTURE HARVESTER (Earl and Vonholdt, 2012), which determines the best K on the basis of the probability of data given K and ΔK (Evanno et al., 2005), was used to estimate the most likely number of clusters (K).

3. Results

3.1. Phenotypic diversity analysis

3.1.1. Analysis of qualitative and quantitative traits

A total of 56 variations of the 20 qualitative traits were detected in 112 accessions, and the average variation of each trait was 2.8. I ranged from 0.09 to 1.03, with an average value of 0.650. The I s of the traits (leaf shape, petiole color, and stem color) were greater than 1. The trait with the highest diversity was stem color ($I = 1.030$), and 59.82% of the accessions were purple green, followed by purple (25.89%), green (8.93%), and brownish green (5.36%). The I s of stem thorn and stem rotation were the lowest, except for CY-256 and CY-257. All accessions had no stem thorn, and the stem rotated to the right. Bulbil was observed among 35.96% of the accessions. Only two accessions had a round leaf shape, while most were triangular oval (66.07%). Three colors of leaf were observed, and dark green was the most common (68.75%). The leaf apex shape of the yam was acuminate or acute, and the proportion of acute shape was 59.82%. In 112 accessions, 77% of the yam had smaller auricle distance, 21.43% had greater auricle distance, and 2% had an auricle distance of 0. Green and purple were observed in the leaf margin of all accessions, with frequency distributions of 41.07% and 58.93%. Petiole color varied from purple (15.18%), green (24.11%),

greenish purple (1.79%), and purplish red (58.93%). The leaf vein color was mainly yellow green, with frequency distribution of 68.75%, and 95.61% of yam leaf veins was observed to be seven. Only 36.61% of the accessions were observed to have stems and wings. Three tuber shapes were observed in all accessions, but cudgel was the most common state (57.14%), followed by massive (33.04%) and columnar (9.82%). Most yam accessions had a low density of root hairs distributed from top to bottom of the tuber. The scarfskin color of most yams was brown, the endodermis color of the tuber was beige, and the tuber flesh color was white, accounting for 70.54%, 73.21%, and 75.89% of all accessions, respectively (Table 2). As shown in Fig. 1, yams have high diversity of phenotypic traits in leaves, stems, and tubers.

The CV of six quantitative traits ranged from 20.81% to 76.11%, and the CVs of the tuber fresh weight, tuber length, tuber diameter, and leaf length were all more than 30%, with the highest being tuber fresh weight and the lowest being length-to-width ratio. The range of tuber fresh weight changed the most (1643.73 g), followed by tuber length (89.92 cm), tuber diameter (15.75 mm), leaf length (13.39 cm), and leaf width (8.88 cm; Table 3). This finding indicated that the quantitative traits of yam tuber and leaf significantly differed in China, the variation range of tuber-related traits in the underground part was larger than that in the aboveground part, and the genetic diversity was rich.

3.1.2. PC analysis and evaluation of phenotypic traits

PC analysis was carried out with 26 phenotypic indices, and eigenvalues greater than 1 were selected to obtain a total of seven PCs (Table 4). The contribution rates of these seven comprehensive indicators were 28.13%, 11.44%, 7.932%, 6.85%, 5.91%, 4.79%, and 4.34%, respectively, with a cumulative contribution rate of 69.39%. The results showed that these traits contained the most genetic information among all phenotypic traits, and they could be used for the comprehensive evaluation of genetic resources. The feature vectors of the first PC1 were mainly stem wing (0.124), leaf length (0.118), tuber shape (0.114), tuber endodermis color (0.105), bulbil or not (0.103), and flowering (0.100). PC2 was a comprehensive response of stem, including stem thorn (-0.290) and stem rotation (-0.290). PC3 was correlated with leaf margin color, leaf apex shape, and leaf color, with eigenvector values of 0.284, 0.218, and -0.220, respectively. The feature vectors of PC4 was connected to leaf vein (0.318), petiole color (0.293), leaf width (0.216), and length-to-width ratio (-0.260). These two PCs were comprehensive responses of leaf. The tuber flesh weight (0.369), tuber diameter (0.253), and leaf vein color (0.265) of PC5 had large eigenvector values. PC6 was associated with leaf vein color (0.457), leaf vein (0.300), and root hair distribution (0.383). PC7 was correlated with root hair density (0.556) and root hair distribution (0.386). In accordance with the contribution rate and eigenvalues of the PCs, leaf color, leaf vein, leaf margin color, leaf vein color, leaf apex shape, petiole color, stem wing, and stem thorn were the main factors that caused the phenotype difference in yam germplasm resources. The 20 traits could be regarded as the main phenotypic indicators for the evaluation of yam germplasm resources in the future. Therefore, these traits could also be regarded as the main phenotypic traits in the evaluation of Chinese yam germplasm resources in the future.

3.1.3. Cluster analysis based on phenotypic traits

A cluster dendrogram of 112 accessions was created from 20 phenotypic traits obtained from PC analysis (Fig. 2). The 112 accessions were divided into seven major groups. The G1 group only contained two accessions of CY-256 and CY-257 from Hainan Province, characterized by no flower, no bulbil, and no stem wing but spiny and turning left. The leaves were nearly round, and the tubers were small and nearly round. The G2 group consisted of 29 accessions, and the main features were stem wings, purple-green stems, purple tuber flesh color or purple inner epidermis color, no flowers nor bulbils, and purple-green petiole. The G3 group included 13 accessions and was very similar to the G2 group. The biggest difference was that the tuber endodermis color was mostly beige, while the tuber flesh color was mostly white. The G4 group contained nine accessions. CY-206 and CY-209 belonged to *D. fordii*, while the rest of the accessions belonged to *D. opposita*. The G5 group was composed of 40 accessions. The typical characteristics were flowers, bulbils, dark-green leaf color, yellow-green vein, columnar-white tuber flesh color, and a variation of length-to-width leaf ratio ranging from 1.2 to 1.5. The G6 group contained 19 accessions characterized by no flowers, no bulbils, oval leaves, purple petioles, yellow-green veins, and brown tuber skin.

3.2. Polymorphism analysis of molecular markers

3.2.1. Marker efficiency analysis

A total of 104 bands and 99 polymorphic bands were amplified from 24 SSR primers, and the percentage of polymorphic loci was 95.19%. Ninety-six pairs of alleles were amplified, the largest N_{as} (eight) were amplified using primers YM06 and YM08, with an average of four alleles per primer. The effective N_a ranged from 1 to 6.429, with an average of 2.488. The mean I , observed heterozygosity (H_o), and expected heterozygosity (H_e) were 0.864, 0.422, and 0.441, respectively. The H ranged from 0 to 0.845, with an average of 0.439 (Table 5).

From the 49 SRAP primers, 29 were selected to produce distinct and stable bands, and 215 bands were amplified, in which 212 were polymorphic, accounting for 98.60% of the total bands. The number of bands amplified by each primer ranged from two to 11, with an average of 7.4. The polymorphic bands amplified by primers F1R2, F3R7, and F7R4 were up to 11. The number of available alleles ranged from 1.072 to 1.519, with an average of 1.342, and the mean H was 0.216. I varied from 0.145 to 0.472, with an average of 0.344 (Table 6).

3.2.2. Cluster analysis based on SSR and SRAP markers

On the basis of SSR-SRAP molecular marker data, 112 accessions of yam were divided into six groups (Fig. 3-A). The G1 group included CY-256 and CY-257 from Hainan Province, and they belonged to *D. esculenta*. A total of 51 accessions were assigned to the G2 group, with CY-40 belonging to *D. opposita*. The G3 group contained 14 accessions, with CY-179 and CY-251 belonging to *D. persimilis*, and the similarity coefficient between CY-244 and CY-247 reached 0.98, suggesting that they may be of same variety. Seven accessions were categorized into the G4 group, and the accession of CY-117 belonged to *D. opposita*, while the remaining belonged to *D. alata*. The G5 group was a collection of *D. fordii* and *D. alata*, in which 29 accessions were *D. alata*, and the remaining (CY-193, CY-201, CY-206, and CY-209) were *D. fordii*. In the G6 group, CY-227 and CY-23 were from *D. opposita*, and the remaining accessions were from *D. alata*.

The results of PCoA was basically consistent with those of molecular marker cluster analysis (Fig. 3-B). The 112 accessions could be divided into three groups (G1, G2, and G3). The G1 group composed of *D. persimilis*, the G2 group was *D. opposita*, and the G3 group consisted of *D. alata* and *D. fordii*.

3.2.3. Population structure analysis

According to the result output from STRUCTURE HARVESTER, when ΔK was at a maximum, the optimal K value was 2 (Fig. 4-A). At $K = 2$, 112 accessions were divided into two subgroups, with red representing the first subgroup (63 accessions) and green representing the second subgroup (49 accessions). The first subgroup was a collection of *D. alata*, *D. persimilis*, *D. fordii*, and *D. esculenta*, the second subgroup was mainly derived from *D. opposita*, which accounted for 43.75%. At $K = 3$, the accessions of *D. opposita* and *D. persimilis* were gathered separately, whereas the remaining species were gathered together. At $K = 4$, the accessions of *D. fordii* and *D. esculenta* were mixed together, whereas those of the other three species were gathered separately (Fig. 4-B).

4. Discussion

4.1 Genetic differences of local Yam Varieties in China were discovered by the combining analyses of phenotypic traits investigation and molecular marker identification

Phenotypic diversity is the external manifestation of genetic diversity, and it is the most basic method for germplasm selection and genetic background research (Mignouna et al., 2002; Sartie et al., 2012; Zhang et al., 2019). In this study, the phenotype diversity of 112 local varieties of yam was high, and the I of leaf shape, petiole color, and stem color of yam were all greater than 1, the local resources of Chinese yam were clustered into 6 groups based on phenotypic traits variations, which was consistent generally with the classification of classical biology (Pei and Ding, 1985). In addition, the typical phenotypic traits were screened for identifying special yam resources. For instance, the stem wings can be effectively used for

identification of *D. alata* (Bressan et al., 2011), and the leaf morphology is the most notable feature of *D. opposita* (Pei and Ding, 1985;). Additionally, the typical characteristic of *D. esculenta* is the stem spines and the stem rotates to the left. (Chair et al. 2005).

Additionally, flowering is a very important breeding requirement in any crop, but the entire genus *Dioscorea* is characterized by dioecy, and most important yam varieties are cultivated for their edible tubers and do not bloom (Renner, 2014; Girma, 2017). In the present study, only 43 accessions (38.39%) were bloomed, 26 of which produced female flowers, and the accessions of *D. esculenta*, *D. alata* and *D. persimilis* did not bloom. In recent years, Tamiru et al., (2017) identified a genomic region and W locus related to sex determination by whole-genome resequencing of white Guinea yam, and a diagnostic marker for sex identification could be developed for yam breeding. The bud bursting stage of yam is generally just at the early stage of expansion of underground tuber. Moreover, the flowering period of the male plants is relatively short (50 days), the flowers are prone to self-withering, and the impact on the underground tubers is also small. Its yield and quality are higher than the female plants (et al., 2011).

Molecular markers have been widely used in species genetic diversity analysis (Yan et al., 2019; Saini et al., 2019; Nicholas et al., 2019). In order to clarify the genetic differences of local Chinese yam varieties at the molecular level, In this study, SSR and SRAP and markers were used to the genetic differences of all accessions, The results of phenotypic and molecular analyses showed good consistency, and the difference was that molecular markers could successfully identify the accessions of *D. opposita* and *D. persimilis*, while the accessions of *D. fordii* and *D. alata* could be identified by phenotypic trait. Moreover, the combination analysis can also identify local varieties (Siqueira et al., 2014; Nicholas et al., 2019). For instance, CY-3 (ZhuGaoShu) is a local variety cultivated for 500 years, in Jiangxi Province, its leaves looks like *D. persimilis*, but the tuber grows just like *D. opposita*, so it is not clear for its species," Zhugaoshu" yam now is identified as *D. persimilis* based on our results. Furthermore, we collected CY-76, CY-79, CY-80, CY-81 and CY-194 in Ruichang County, Jiangxi Province, which have been considered as *D. opposita* before, but their phenotypic characteristics are different from this species. Combining with current study and chloroplast genome sequencing analysis (data not shown), we preliminarily speculated that it may be a new species or a new variant of *D. opposita*.

4.2 The possible origin and domestication of *D. opposita* and *D. alata* were speculated by genetic difference and population structure analysis of China yam resources

Yam had gradually formed the cultivation and domestication centers of Asian group, African group and American group in the long-term evolution process. Africa group, including species of *D. rotundata*, *D. alata*, *D. cayenensis* and *D. dumetorum*, has become the main regional belt of yam production in the world, mainly distributed in Ghana, Togo, Benin and other West African regions, Central Africa and Western Congo (Coursey, 1976; Martin and Sadik, 1977). *D. trifida* was the earliest domesticated variety in South America and was widely cultivated, mainly distributed in Brazil, Venezuela, Paraguay and other regions (Coursey, 1976). Asia was also the main distribution center of yam. The main cultivated and domesticated species are *D. opposita*, *D. alata*, *D. esculenta*, *D. japonica*, *D. bulbifera*, *D. hispida* and *D. quinqueloba* (Gong et al., 2004). China is one of the important domestication centers of yams, with records about yams from 4000 years ago. Thus far, the species of yams widely cultivated in China are *D. opposita*, *D. alata*, *D. persimilis*, *D. fordii*, and *D. esculenta* (Li et al., 2009).

At present, the origin of yam could be traced back to the late Cretaceous (Maurin et al., 2016). *Dioscorea* is considered to be a monophyletic group originating from a common ancestor (Wu, 2012), which represents an early-diverging lineage of monocots just internal to Acorus (Hansen, 2007). However, there are still many different arguments about its origin, evolution process and domestication process. In this study, a total of 112 cultivars, landraces and wild varieties of yam were collected in 21 provinces (cities) in China, and its genetic diversity was comprehensively evaluated. *D. opposita* was the most important with the largest cultivation area and widest distribution in China. It was distributed from Heilongjiang Province in the north, Hainan Province in the south, Shanghai in the east, and Sichuan Province in the west. And its growth environment was also more complicated, including plains, mountainous region and coastal areas. Therefore, this may be

the reason for the larger genetic difference compared with other species to adapt to their respective growth environment and climate change and climatic conditions. Moreover, *D. opposita* had the shortest growth period and the smallest leaves compared with other species, indicating that the aboveground biomass was small. In addition, a large number of wild resources of *D. opposita* were distributed in the northern and southern provinces of China. In the present study, wild resources were collected in Henan, Hebei, Guizhou, Sichuan, and Jiangxi Provinces. This may be a strong evidence that *D. opposita* may have originated in China and domesticated by wild species.

Forty-one accessions of *D. alata* were collected in eight provinces in Southern China, accounting for 36.61% of the total. The most typical features of this species were stem wings, strong growth potential, and long growth period, and these characteristics were similar to the previous results (Bressan et al., 2011). This may be evidence to infer that it may have originated in tropical or subtropical areas. With regards to previous report, *D. alata* may originate in the north and east of the Bay of Bengal and spread to Southeast Asia, Malaysia, Pacific tropical islands, Africa, and America (Nemurin, 2013). Some authors have proposed that the ginseng potato was domesticated in India or Yunnan Province in China (Coursey, 1976; Chair et al., 2016; Wu et al. 2019). Moreover, *D. alata* is considered a heterozygous species and inferred as a cross between the wild relative *D. hamiltonii* and *D. persimilis*.

The wild resources of *D. persimilis* was previously reported to be distributed in Hunan, Guangdong, Guangxi, Guizhou and Yunnan provinces (Pei, 1985). And we also collected the accessions of *D. persimilis* in Fujian and Jiangxi provinces, and it has a long history of cultivation. In addition, *D. persimilis* and *D. opposita* were closely related, and some accessions were interpenetrated on phenotypic traits. The rDNA internal transcribed spacer (ITS) sequences also showed that *D. persimilis* and *D. opposita* were closely related (Wu et al., 2013; Liu et al., 2001). It is also speculated that *D. persimilis* is considered to be an internal mutation of *D. opposita* (Liu et al., 2001).

The wild resources of *D. fordii* are distributed in Zhejiang, Guangdong, Guangxi, Fujian and Hunan provinces. It has been widely cultivated for its high yield and good resistance to stress and has been cultivated for more than 200 years so far. *D. fordii* may be formed by long-term domestication of wild species. In the current study, CY-193, CY-201, CY-206 and CY-209 from *D. fordii* were dispersed in *D. alata* (Figs. 3-A and B), indicating that *D. alata* and *D. fordii* were the closest species genetically. Similar results were obtained by Lei et al. (2013) and Li et al. (2016) on genetic diversity and Wu et al. (2014) on ITS. However, research on the origin and evolution of *Dioscorea* requires additional genomic information and increased number of species for further study.

ACKNOWLEDGEMENTS

The authors would like to express their deep thanks to Associate Researcher Hua Shumei (Sanming Academy of Agricultural Sciences), Professor Huo Xiuwen (Inner Mongolia Agricultural University), Researcher Zhang Peitong (Jiangsu Academy of Agricultural Sciences), and Associate Professor Wu Wenqiang (Hainan University) for their support in collecting yam germplasm resources.

Funding

This work was supported by the earmarked fund for Jiangxi Agriculture Research System (JXARS-19) and Jiangxi Provincial Key Research and Development Project (20192BBF60005, 20201BBF61002).

Conflict of interest

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are included within the supplementary additional document.

References:

- Anokye, M., Tetteh, J.P., Otoo, E., 2014. Morphological characterization of some water yam (*Dioscorea alata* L.) germplasm in Ghana. *Journal of Agricultural Science and Technology* 4, 518–532.
- Arnau, G., Bhattacharjee, R., Mn, S., Chair, H., Malapa, R., Lebot, V., K, A., Petro, D., Penet, L., Pavis, C., 2017. Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PLoS ONE* 12(3): e0174150. <https://doi.org/10.1371/journal.pone.0174150>.
- Bressan, E.A., Briner Neto, T., Zucchi, M.I., Rabello, R.J., Veasey, E.A., 2011. Morphological variation and isozyme diversity in *Dioscorea alata* L. landraces from Vale do Ribeira, Brazil. *Scientia Agricola* 68, 494-502. <http://dx.doi.org/10.1590/S0103-90162011000400016>.
- Cao, T.X., Zhu, Q.L., Chen, X., Wang, P.T., Shan, N., Zhou, Q.H., Huang, Y.J., 2020. The complete chloroplast genome sequence of the *Dioscorea persimilis* Prain et Burkill (Dioscoreaceae). *Mitochondrial DNA Part B*, 5:1, 451-452. <http://dx.doi.org/10.1080/23802359.2019.1704645> .
- Chair, H., Perrier, X., Agbangla, C., Marchand, J.L., Dainou, O., Noyer, J.L., 2005. Use of cpSSRs for the characterisation of yam phylogeny in Benin. *Genome* 48:674–684
- Cheng, J., Chen, J., Liu, X., Li, X., Zhang, W., Dai, Z., Lu, L., Zhou, X., Cai, J., Zhang, X., Jiang, H., Ma, Y., 2020. The origin and evolution of diosgenin biosynthetic pathway in Yam, *Plant Communications* <https://doi.org/10.1016/j.xplc.2020.100079>.
- Coursey, D. G., 1976. The origins and domestication of yams in Africa. In *Origin of African plant domesticates*. Edited by Harlan, J. R., J. M. J. DE WET and A. B. L. Stemler, The Hague 383–408.
- Earl, D.A., VonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>.
- Evanno, G., Regnaut, S., Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14(8), 2611-2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Ghorbani, S., Esmaili, H., Ebrahimi, S.N, Palazón, J., Sonboli, A., Mirjalili, M.H., 2020. Genetic structure, molecular and phytochemical analysis in iranian populations of *Ruscus hyrcanus* (Asparagaceae). *Industrial Crops and Products* 154, 112716. <https://doi.org/10.1016/j.indcrop.2020.112716>.
- Girma, G., Bhattacharjee, R., Lopez-Montes, A., Abberton, M., 2017. Re-defining the yam (*Dioscorea* spp.) core collection using morphological traits. *Plant Genetic Resources: Characterization and Utilization* 16(3), 1-8. <https://doi.org/10.1017/S1479262117000144>.
- Gong, Q.P., Cheng, G., Yuan, L.G., 2004. Standardized cultivation techniques of Chinese yam. China agricultural press, Beijing, China.
- Hansen, D.R., Dastidar, S.G., Cai, Z.Q., Penafior, C., Kuehl, J.V., Boore, J.L., Jansen, R.K., 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular phylogenetics and evolution*, 45(2), 547-563. <https://doi.org/10.1016/j.ympev.2007.06.004>.
- Huang, D.Y., and Huang, X.L., 2013. Descriptors and data quality control criterion for *Dioscorea opposita* Thunb. Science Press, Beijing, China.
- Huang, W.H., 2005. Nonpollution and Standardization Cultivation for Yam. China Agriculture Press, Beijing, China.
- Huang, Y.X., Huang, S., Liang, K.J., Ma, H.B., Lei, F.G., Hua, S.M., Xu, X.M., 2011. Genetic Diversity of Germplasm Resources on *Dioscorea* by SRAP Markers. *Chinese Wild Plant Resources*, 30(6), 48-54.

- Kouam, E.B., Avana-Tientcheu, M.L., Lekeumo, V.D., Akitio, H.M., Khasa, D.P., Pasquet, R.S., 2018. Agro-ecological distribution of the phenotypic diversity of aerial yam (*Dioscorea bulbifera* L.) in Cameroon using multivariate analysis: prospect for germplasm conservation and improvement. *Open Agriculture* 3(1), 190-206. <https://doi.org/10.1515/opag-2018-0020>.
- Lebot, V., 2009. Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams and Aroids. CABI, Wallingford, UK. <https://doi.org/10.1017/S0014479709007832>.
- Lebot, V., Faloye, B., Okon, E., Gueye, B., 2019. Simultaneous quantification of allantoin and steroidal saponins in yam (*Dioscorea* spp.) powders. *Journal of Applied Research on Medicinal and Aromatic Plants* 13. <https://doi.org/10.1016/j.jarmap.2019.02.001>.
- Lebot, V., Malapa, R., Molisálé, T., 2018. Development of HP-TLC method for rapid quantification of sugars, catechins, phenolic acids and saponins to assess Yam (*Dioscorea* spp.) tuber flour quality, *Plant Genetic Resources* 17(1):62-72. <https://doi.org/10.1017/S1479262118000333>
- Lei, F.G., Hua, S.M., Shu, M., Tu, Q.C., He, P.Z., Li, Q.X., Cao, Y.Y., 2013. Genetic Relationship of *Dioscorea Polystachya* Turcz. Resources Based on ISSR Markers. *Fujian Journal of Agricultural Sciences* 28, 27-32.
- Li, G., Quiros, C.F., 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: Its application to mapping and gene tagging in Brassica. *TAG Theoretical and Applied Genetics* 103(2-3), 455-461. <https://doi.org/10.1007/s001220100570>.
- Li, J., Liang, Q., Li, C.F., Liu, M.D., Zhang, Y.S., 2018. Comparative Transcriptome Analysis Identifies Putative Genes Involved in Dioscin Biosynthesis in *Dioscorea zingiberensis*, *Molecules* 23(2), 454. <https://doi.org/10.3390/molecules23020454>.
- Li, L.H., Hua, S.M., Chen, Z.H., Lei, F.G., Cao, Y.Y., 2016. Phenotypic Genetic Diversity Research on *Dioscorea* Landraces of Fujian Province. *Journal of Yunnan Agricultural University (Natural Science)* 31(2), 257-262. [https://doi.org/10.16211/j.issn.1004-390X\(n\).2016.02.010](https://doi.org/10.16211/j.issn.1004-390X(n).2016.02.010).
- Li, Q.X., Hua, S.M., Lei, F.G., Chen, Z.H., He, P.Z., Cao, Y.Y., Zhou, J.J., 2013. Comparison of ISSR and SRAP Markers on the Genetic Diversity Analysis of *Dioscoreae* . *Fujian Journal of Agricultural Sciences* 28(9), 876-883. <https://doi.org/10.19303/j.issn.1008-0384.2013.09.009>.
- Liu, Y.P., He B.Z., Cao, H., 2001. Application of gene technology in quality control of Chinese materia medica II. Identification of Chinese Rhizoma *Dioscoreae* by DNA sequencing, *Chinese traditional and Herbal Drugs* 32(11), 1026-1030.
- Loko, Y.L., Bhattacharjee, R., Agre, A.P., Dossou-Aminon, I., Orobiyi, A., Djedatin, G.L., Dansi, A., 2016. Genetic diversity and relationship of Guinea yam (*Dioscorea cayenensis* Lam.-*D. rotundata*Poir. complex) germplasm in Benin (West Africa) using microsatellite markers. *Genetic Resources and Crop Evolution* 64(6), 1205-1219. <https://doi.org/10.1007/s10722-016-0430-z>.
- Martin, F.W.; Sadik, S., 1977, Tropical yams and their potential. Part 4. *Dioscorea rotundata* and *Dioscorea cayenensis* . *USDA Agriculture Handbook* (502): iv; 36, 1977
- Maurin, O., Muasya, A.M., Catalan, P., Shongwe, E.Z., Viruel J., Wilkin, P., van der Bank M., 2016. Diversification into novel habitats in the Africa clade of *Dioscorea* (*Dioscoreaceae*): erect habit and elephant's foot tubers. *BMC Evolutionary Biology* 16(1), 238. <https://doi.org/10.1186/s12-862-016-0812-z>.
- Mignouna, D., Mank, A., Ellis, N., Van Den Bosch, N., Asiedu, R., Abang, M., Peleman, J., 2002. A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. *Theoretical and Applied Genetics*, 105(5), 726-735. <https://doi.org/10.1007/s00122-002-0912-6>.
- Mignouna, H.D., Abang, M.M., Fagbemi, S.A., 2003. A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea ratundata*) germplasm characterization. *Annals of*

Applied Biology 142(3), 269-276. <https://doi.org/10.1111/j.1744-7348.2003.tb00250.x>.

Mignouna, H.D., Dansi, A., Zok, S., 2002. Morphological and isozymic diversity of the cultivated yams (*Dioscorea cayenensis* / *Dioscorea rotundata* complex) of Cameroon. Genetic Resources and Crop Evolution 49(1), 21-29. <https://doi.org/10.1023/A:1013805813522>.

Narina, S.S., Buyyarapu, R., Kottapalli, K.R., Sartie, A.M., Ali, M.I., Robert, A., Hodeba, M.J., Sayre, B.L., Scheffler, B.E., 2011. Generation and analysis of expressed sequence tags (ESTs) for marker development in yam (*Dioscorea alata* L.). BMC Genomics 12, 100. <https://doi.org/10.1186/1471-2164-12-100>.

Nemorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., Arnau, G., 2013. Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. Annals of botany 112(5), 811- 819. <https://doi.org/10.1093/aob/mct145>.

Nicholas, N.D., Frederick, J.A., Brian, D., Francisca, A.F., Godfree, C., Richard, O.F., Michael, T.B., 2019. Genetic diversity, population structure and key phenotypic traits driving variation within soyabean (*Glycine max*) collection in Ghana. Plant Breeding, 138(5), 577-587.

Pei, J., Ding, Z.Z., 1985. Flora of China: 16 volumes. Science Press, Beijing, China.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155(2), 945-959.

Renner, S.S., 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. American journal of botany 101, 1588-96.

Rohlf, F.J., 2000. NTSYSpc, numerical taxonomy and multivariate analysis system. Version 2.1x. New York: Exeter Publication Ltd Setauket.

Saini, P., Kamboj, D., Yadav, N.R., 2019). SRAPs and EST-SSRs provide useful molecular diversity for targeting drought and salinity tolerance in Indian mustard, 46(1), 1213-1225. <https://doi: 10.1007/s11033-019-04590-4>.

Sartie, A., Asiedu, R., Franco, J., 2012. Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). Genetic Resources and Crop Evolution 59(8), 1753-1765. <https://doi.org/10.1007/s10722-012-9797-7>.

Shewry, P.R., 2003. Tuber Storage Proteins. Annals of Botany 91(7), 755-769. <https://doi.org/10.1093/aob/mcg084>.

Silva, L.R.G., Mezette, T.F., Nascimento, W.F., Silva, E.F., Veasey, E.A., 2016. Spatially structured morphological and molecular diversity among *Dioscorea cayenensis* and *D. rotundata* yam accessions. Plant Genetic Resources 15(4), 296-309. <https://doi.org/10.1017/S147926211-5000647>.

Siqueira, M. V. B. M., Bonatelli, M. L., Gunther, Torsten, Gawenda, I., Schmid, K. J., Pavinato, V. A. C., 2014. Water yam (*Dioscorea alata* L.) diversity pattern in brazil: an analysis with SSR and morphological markers. Genetic Resources and Crop Evolution 61(3), 611-624. <https://doi.org/10.1007/s10722-013-0063-4>.

Tamiru, M., Natsume, S., Takagi, H., White, B., Yaegashi, H., Shimizu, M., Yoshida, K., Uemura, A., Oikawa, K., Abe, A., Urasaki, N., Matsumura, H., Babil, P., Yamanaka, S., Matsumoto, R., Muranaka, S., Girma, G., Lopez-Montes, A., Gedil, M., Bhattacharjee, R., Abberton, M., Kumar, P.L., Rabbi, I., Tsujimura, M., Terachi, T., Haerty, W., Corpas, M., Kamoun, S., Kahl, G., Takagi, H., Asiedu, R., Terauchi, R., 2017. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biology 15(1), 86. <https://doi.org/10.1186/s12915-017-0419-x>.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA 4:molecular evolutionary genetics analysis(MEGA)software version 4.0. Molecular Biology and Evolution 24(8), 1596-1599. <https://doi.org/10.1093/molbev/msm092>.

- Tamiru, M., Becker, H.C, Maass, B.L., 2011. Comparative Analysis of Morphological and Farmers' Cognitive Diversity in Yam landraces (*Dioscorea* spp.) from Southern Ethiopia. *Tropical Agriculture and Development* 55(1), 28-43.
- Toth, G., Gaspari, Z., Jurka, J., 2000. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome research* 10(7), 967-981. <https://doi.org/10.1101/gr.10.7.967>.
- Tschannen, A.B., Escher, F., Stamp, P., 2005. Post-harvest treatment of seed tubers with gibberellic acid and field performance of yam (*Dioscorea cayenensis-rotundata*) in ivory coast. *Experimental Agriculture*, 41(2), 175-186. <https://doi.org/10.1017/S001447970400-2455>.
- Wang, H.P., and Shen, D., 2014. Descriptors and data standard for yam (*Dioscorea* spp.). China Agriculture Press, Beijing, China.
- Wu, W.Q., Chen, C., Zhang, Q., Ahmed, J.Z., Xu, Y., Huang, X.L., Xie, J., Xia, W., Huang, D.Y., 2019. A comparative assessment of diversity of greater yam (*Dioscorea alata*) in China. *Scientia Horticulturae* 243, 116-124. <https://doi.org/10.1016/j.scienta.2018.08.016>.
- Wu, Z.G., Fan, C.Y., Bao, X.Q., Tao, Z.M., Jiang, W., 2014. Sequence analysis on rDNA-ITS region of germplasm resources from *Dioscoreae Rhizoma*. *Chinese Traditional and Herbal Drugs*, 45(8), 1136-1142.
- Wu, Z.G., Li, X.X., Fan, C.Y., Tao, Z.M., 2013. Investigation of Genetic Diversity of Yam Cultivars by SRAP Markers. *Chinese Pharmaceutical Journal* 48(9), 681-685. <https://doi.org/10.11669/cpj.2013.09.006>.
- Wu, Z.G., Ning, W., Wang, Y.S., Lv, S.S., Zhao, Y.B., Zhao, X., 2011. Genetic diversity of taraxacum germplasm revealed by sequence-related amplified polymorphism (SRAP) analysis. *African Journal of Biotechnology* 10(47), 9557-9562. <https://doi.org/10.5897/ajb10.2611>.
- Yan, W., Li, J, Zheng, D., Friedman, C., Wang, H., 2019. Analysis of genetic population structure and diversity in *Mallotus oblongifolius* using ISSR and SRAP markers. *PeerJ* 7:e7173 DOI 10.7717/peerj.7173.
- Yeh Francis, C., Yang, R.C., Boyle Timothy, B.J., Ye, Z.H., Mao Judy, X., 1999. POPGENE Version 1.32, the User-friendly Shareware for Population Genetic Analysis. Molecular Biology and Biotechnology Centre. University of Alberta, Canada.
- Zhang, W.J., Chen, J.Y., Liu, B.C., Zhao, Y.Q., Huang, Y.Z., Chen, Y., 2019. Genetic Variations on Major Traits of 37 Chinese Yam Germplasms [J]. *Fujian Journal of Agricultural Sciences*, 34(11), 1246-1254.

Table 1 Descriptors used for phenotypic assessment of yam accessions in this study.

Traits	Classes (Codes)
Flowering (FL)	1 = Absent, 2 = Male, 3 = Female.
Bulbil or not (BN)	1= Bulbil, 2 = No bulbil.
Leaf shape (LS)	1 = Heart triangle, 2 = Triangular ovate, 3 = Lanceolate, 4 = Round, 5 = Shape of ha
Leaf color (LC)	1 = Yellow-green, 2 = Greenish-gray, 3 = Dark green.
Leaf apex shape (LAX)	1 = Acuminate, 2 = Acute.
Auricle distance (AD)	1 = Small, 2 = Big, 3 = No.
Leaf margin color (LMC)	1 = Green, 2 = Purple.
Petiole color (PC)	1 = Purple, 2 = Green, 3 = Greenish purple, 4 = Purplish red.
Leaf vein color (LVC)	1 = Yellow-green, 2 = Green, 3 = Purple.
Leaf vein (LV)	1 = Five, 2 = Seven veins, 3 = Nine veins.
Stem wing (SW)	1 = No wing, 2 = Wing.
Stem color (SC)	1 = Green, 2 = Green with purple, 3 = Brownish-green, 4 = Purple.
Stem thorn (ST)	1 = Absence, 2 = Presence.
Rotation of stem (RS)	1 = Right, 2 = Left.
Tuber shape (TS)	1 = Cudgel, 2 = Columnar, 3 = Massive.

Traits	Classes (Codes)
Root hair density (RHD)	1 = Low, 2 = High
Root hair distribution (RHDB)	1 = All, 2 = Upper and Middle
Scarfskin color (SSC)	1 = Brown, 2 = Black, 3 = Gray.
Endodermis color of tuber (ECT)	1 = Beige, 2 = Purple.
Flesh color (FC)	1 = White, 2 = Yellow, 3 = Purple, 4 = Purple with white.
Leaf length (LL)	Average leaf length of six mature leaves (cm).
Leaf width (LW)	Average leaf width of six mature leaves (cm).
Length to width ratio (L/W)	Average leaf length / average leaf width.
Tuber length (TL)	Average tuber length of six plants (cm).
Tuber diameter (TD)	Average tuber diameter of six plants (mm).
Tuber flesh weight (TFW)	Average yield of six plants (g).

Table 2 The distribution frequency (%) and Shannon's Diversity Index (I) of each phenotype in qualitative traits.

Traits	Distribution of grade /%	Distribution of grade /%	Distribution of grade /%	Distribution of grade /%	Distribution of grade /%
	1	2	3	4	5
FL	61.61	15.18	23.21	-	-
BN	35.96	64.04	-	-	-
LS	6.25	66.07	5.36	1.79	20.5
LC	24.11	7.14	68.75	-	-
SST	59.82	40.18	-	-	-
AD	76.79	21.43	1.79	-	-
LMC	41.07	58.93	-	-	-
PC	15.18	24.11	1.79	58.93	-
LVC	68.75	25.00	6.25	-	-
LV	3.51	95.61	0.88	-	-
SW	36.61	63.39	-	-	-
SC	8.93	59.82	5.36	25.89	-
ST	98.21	1.75	-	-	-
RS	98.21	1.75	-	-	-
TS	57.14	9.82	33.04	-	-
RHD	85.71	14.29	-	-	-
RHDB	81.25	18.75	-	-	-
SSC	70.54	20.54	8.93	-	-
BC	73.21	26.79	-	-	-
FC	75.89	4.46	12.50	7.14	-
Mean	-	-	-	-	-

Table 3 Statistical analysis of quantitative characteristics.

Characters	Mean	Maximum	Minimum	Rang	SD	CV (%)
LL (cm)	11.30	19.27	5.88	13.39	3.46	30.60
LW (cm)	6.91	11.73	2.85	8.88	1.73	25.03
L/W	1.65	3.25	0.95	2.30	0.34	20.81
TL (cm)	33.62	97.17	7.25	89.92	16.26	48.36
TD (mm)	8.19	17.50	1.75	15.75	3.59	43.79

Characters	Mean	Maximum	Minimum	Rang	SD	CV (%)
TFW (g)	500.34	1649.33	5.60	1643.73	380.79	76.11

SD standard deviation, CV coefficient of variation.

Table 4 Principal component analysis of 26 single indexes of *Dioscorea* ssp.

Phenotype trait	Eigenvector of the principal component PC1	Eigenvector of the principal component PC2	Eig
FL	-0.100	0.023	0.0
BN	-0.103	0.012	-0.0
LS	-0.068	0.127	0.1
LC	-0.066	0.057	-0.0
SST	0.021	0.080	0.2
AD	-0.032	0.194	0.0
LMC	-0.026	0.142	0.2
PC	0.026	0.010	0.1
LVC	0.013	0.091	0.0
LV	0.018	-0.064	-0.0
SW	0.124	0.053	0.0
SC	-0.071	0.017	0.0
ST	0.004	-0.290	0.0
RS	0.004	-0.290	0.0
TS	0.114	-0.040	0.0
RHD	-0.020	0.018	-0.0
RHDB	0.028	0.042	-0.0
SSC	-0.077	-0.063	0.1
BC	0.105	0.045	0.1
FC	0.093	0.055	0.1
LL	0.118	0.044	-0.0
LW	0.089	-0.075	-0.0
L/W	0.067	0.135	-0.0
TL	-0.085	0.111	-0.0
TD	0.083	0.058	-0.0
YPP	0.060	0.120	-0.0
Eigen value	7.315	2.974	2.0
Contributive ratio (%)	28.13	11.44	7.9
Cumulative contributive ratio (%)	28.13	39.57	47.0

Table 5 Genetic differentiation parameters on 112 yam varieties as revealed by 24 polymorphic SSR markers.

Marker	N_a	N_e	I	H_o	He	H
YM02	7	3.619	1.447	0.356	0.728	0.724
YM03	2	1.894	0.665	0.619	0.474	0.472
YM06	8	6.429	1.954	0.667	0.850	0.845
YM07	7	5.069	1.745	0.578	0.807	0.803
YM09	7	4.115	1.625	0.411	0.761	0.757
YM12	2	1.574	0.551	0.360	0.367	0.365
YM13	5	1.743	0.880	0.474	0.429	0.426

Marker	N_a	N_e	I	H_o	H_e	H
YM15	1	1	0	0	0	0
YM17	2	1.625	0.573	0.520	0.387	0.385
YM19	4	2.385	1.094	0.765	0.584	0.581
YM21	2	1.280	0.377	0.250	0.220	0.219
YM24	2	1.932	0.675	0.574	0.485	0.482
YM27	1	1	0	0	0	0
YM30	2	1.271	0.369	0.202	0.214	0.213
YM32	8	4.726	1.797	0.722	0.793	0.788
YM33	4	1.957	0.928	0.437	0.491	0.489
YM35	5	3.036	1.246	0.768	0.674	0.671
YM37	7	4.675	1.695	0.891	0.790	0.786
YM41	2	1.127	0.227	0.100	0.113	0.113
Dab2D08	2	1.032	0.081	0.0312	0.031	0.031
Da1A01	1	1	0	0	0	0
Da1D08	7	2.747	1.332	0.747	0.640	0.636
Da1F08	2	1.044	0.105	0	0.043	0.043
SSR-17	6	3.439	1.358	0.662	0.714	0.709
Mean	4	2.488	0.864	0.422	0.441	0.439
Total	96	59.717	20.724	10.134	10.593	10.535

N_a : observed number of alleles, N_e : effective number of alleles, I : Shannon's diversity index, H_o : observed heterozygosity, H_e : expected heterozygosity, H : Nei's gene diversity.

Table 6 Genetic differentiation parameters on 112 yam varieties as revealed by 29 polymorphic SRAP markers.

Prime	Polymorphic Loci	N_a	N_e	H	I
F1R1	8.00	2.00	1.332	0.205	0.322
F1R2	11.00	2.00	1.335	0.215	0.346
F1R3	4.00	2.00	1.495	0.291	0.435
F1R6	5.00	2.00	1.784	0.430	0.619
F2R1	5.00	2.00	1.335	0.241	0.402
F2R2	2.00	2.00	1.178	0.151	0.284
F2R6	8.00	2.00	1.197	0.138	0.237
F3R1	2.00	2.00	1.072	0.067	0.149
F3R2	9.00	2.00	1.332	0.200	0.319
F3R3	8.00	2.00	1.327	0.207	0.333
F3R6	7.00	2.00	1.346	0.231	0.378
F3R7	11.00	2.00	1.263	0.169	0.279
F4R2	7.00	2.00	1.395	0.252	0.400
F4R4	8.00	2.00	1.403	0.251	0.400
F4R5	8.00	2.00	1.519	0.312	0.472
F4R7	9.00	2.00	1.379	0.249	0.391
F5R1	8.00	2.00	1.233	0.137	0.220
F5R2	6.00	1.83	1.500	0.283	0.418
F5R3	6.00	1.83	1.497	0.285	0.423
F5R4	8.00	2.00	1.137	0.083	0.145
F5R6	10.00	2.00	1.404	0.267	0.426
F6R2	7.00	2.00	1.414	0.279	0.443
F6R3	5.00	2.00	1.250	0.156	0.256

Prime	Polymorphic Loci	N_a	N_e	H	I
F6R4	9.00	2.00	1.233	0.153	0.255
F6R5	8.00	2.00	1.198	0.134	0.226
F6R6	7.00	2.00	1.253	0.165	0.275
F7R1	9.00	2.00	1.440	0.272	0.424
F7R3	9.00	2.00	1.338	0.207	0.335
F7R4	11.00	1.91	1.344	0.219	0.353
Mean	7.40	1.99	1.342	0.216	0.344
Total	215.00	57.58	38.929	6.251	9.961

N_a : observed number of alleles, N_e : effective number of alleles, H : Nei's gene diversity, I : Shannon's diversity index.

Figure legends Fig. 1. The rich phenotypic characters of yam germplasm resources. **A** . Different leaf morphology, **B** . Different stem morphology, **C** . Different tuber morphology, Bar= 1 cm.

Fig. 2. Cluster dendrogram based on the phenotypic traits. The letter "G" stands for different groups. The letter "G" is an abbreviation for Group.

Fig. 3. The UPGMA Cluster dendrogram (A) and PCoA analysis (B) of 112 yam accessions based on the combined SSR-SRAP data. The letter "G" is an abbreviation for Group, and G1-G6 stands for different groups respectively.

Fig. 4. Population structure analysis of 112 yam accessions (**A**) ΔK values for different numbers of populations assumed (K), (**B**) Population genetic structure at K = 2, 3, 4.







