Genetic diversity and population structure of sweet potato landraces based on phenotypic traits, carotenoid content, and SSR molecular markers

Yihang Chen
  College of Grassland Agriculture, Northwest A&F University
Chaochen Tang
  Crops Research Institute, Guangdong Academy of Agricultural Sciences
Rong Zhang
  Crops Research Institute, Guangdong Academy of Agricultural Sciences
Jingwei Jin (✉ jjw@ms.iswc.ac.cn)
  College of Grassland Agriculture, Northwest A&F University
Zhangying Wang
  Crops Research Institute, Guangdong Academy of Agricultural Sciences

Abstract

Germplasm resources are an important basis for genetic breeding, and the study of genetic diversity is beneficial to the discovery and improvement of germplasm resources. In this study, 132 local germplasm of sweet potato were used as test materials, and 20 phenotypic traits were analyzed for genetic diversity using distribution frequency, coefficient of variation, and Shannon's diversity index; carotenoid contents in sweet potato tubers were measured by HPLC, and their genetic diversity was analyzed using the coefficient of variation, mean, and standard deviation. Meanwhile, the genetic diversity, as well as population structure, were analyzed based on 10 pairs of SSR molecular markers using Nei's genetic diversity index and Shannon's diversity index. The results showed that the 132 sweet potato landraces had high genetic diversity based on phenotypic traits, SSR molecular markers, and carotenoid content, respectively; based on population structure analysis, the 132 sweet potato landraces could be divided into five subgroups. Analysis of molecular variation (AMOVA) showed that 84% of the variation existed within populations and 16% between populations. This study provides a basis for the selection of parental material for the genetic breeding improvement of sweet potato and for the mining of superior genes.

Background

Sweet potato (*Ipomoea batatas* [L.] Lam), an annual or perennial dicotyledonous plant of the genus sweet potato in the family Convolvulaceae, is an important food, animal feed, and industrial raw materials crop worldwide. China, a major producer of sweet potato, produced 49,195,561 tons of sweet potato annually in 2020, accounting for 54.97% of the total annual global sweet potato production (FAO, 2022). Sweet potato originated from America and was introduced into China at the end of the 16th century[1, 2]. Due to its extensive environmental adaptability, it has formed rich germplasm resources in China's changeable climate and complex geographical environment.

Carotenoids are colored pigments that are widely distributed in nature. In sweet potato, some varieties have orange and yellow flesh due to the presence of carotenoids[3]. Carotenoids are an important source of vitamin A[4] and play an important role in the treatment of vitamin A deficiency (VAD). In addition, carotenoids have various physiological functions such as antioxidant, eye and skin protection, and prevention of cardiovascular diseases[4-8].

Genetic diversity is a fundamental part of biodiversity and is the basis for ecological and species diversity. It usually refers to the sum of genetic variation among different groups of species or among individuals within a group[9, 10]. There is a wide range of genetic variations in nature, which can be influenced not only by natural factors but also by interventions from humans. The domestication of natural species by human selection has resulted in the conservation of species with beneficial agricultural production and the neglect of others. This behavior has drastically reduced the species diversity of many species.

The study of genetic diversity includes morphological level and molecular level, etc. The morphological level of genetic diversity mainly uses phenotypic characteristics to distinguish different individuals, and this method is also the most intuitive. Compared with phenotypes, the greatest advantage of molecular markers is that they are not affected by the natural environment and can reflect the differences between materials and their relatedness more truly than phenotypes. Among many molecular markers, SSR molecular markers have the advantages of multiple alleles, co-dominance, and high polymorphism[11] and are widely used in genetic diversity analysis. Genetic diversity is usually not analyzed using phenotypic or molecular markers alone, but rather a combination of phenotypic and molecular markers, etc. are analyzed along with each other, as reported in other crops[12-14].

The purpose of this study was to evaluate the genetic diversity and population structure of 132 sweet potato landraces preserved in the National Sweet Potato Germplasm Resources Garden (Guangzhou) based on phenotypic and SSR molecular markers.

Materials And Methods

Plant materials

This study used 132 sweet potato landraces preserved in the National Sweet Potato Germplasm Resource Nursery (Guangzhou) as experimental materials. The materials were planted in Baiyun Experimental Base of Guangdong Academy of Agricultural Sciences in Baiyun District, Guangzhou, which is located at about 23 ° 39 ' north latitude and 113 ° 44 ' east longitude, with an average annual temperature of 22 °C. In each resource, 20 plants of the first stage of tender and strong seedlings with good growth without diseases and pests were selected for planting, double row planting, row spacing 110 cm, plant spacing 20 cm, ridge height 30 cm, and normal field management.

Phenotypic data collection

The phenotypic traits were investigated according to the Descriptors and Data Standard for Sweet potato [*Ipomoea batatas*(L.)Lam.][15]. The aboveground traits were investigated 40-50 days after transplanting, and the underground traits were investigated 130-140 days after transplanting. Phenotypic data were collected by the field observation method. The color of top leaf was measured by Konica Minolta CM-700d / 600d spectrophotometer. The leaf color and the predominant flesh color were measured by UNIS Uniscan M1 Fast Plate Scanner.

Determination of carotenoid content

0.5g cold-dried powder of sweet potato storage root and 5mL extract (n-hexane/acetone/ethanol (1:1:1, v/v/v) +0.01% BHT) were added to a 15 mL centrifuge tube. The mixture was fully oscillated at room temperature and centrifuged at 4 °C, 4000rpm / min for 5min. The supernatant was transferred to a new centrifuge tube; 5 mL extract was added again and centrifuged, took the supernatant to the same centrifuge tube; finally, added 4 mL extract, and repeated the previous operation. The centrifuge tube containing the supernatant was dried by nitrogen blowing instrument. After the completion of nitrogen blowing, added 1 mL of complex solution (methyl tert-butyl ether), shaken well, and ultrasonic 10 min assisted dissolution. After standing for 5 min, the supernatant was extracted and filtered with 0.22 μm filter membrane into a 1.5 mL brown storage bottle.

Carotenoids were detected by HPLC on an Agilent 1260 Infinity II liquid chromatograph with a DAD diode array detector and data analysis system (Agilent, USA). The column used for the detection was a YMC C30 column (2 mm×100 mm, 3 μm, YMC, Japan). The mobile phase A was methanol: acetonitrile (1:3, v/v) + 0.01% BHT, and the mobile phase B was methyl tert-butyl ether + 0.01% BHT. The elution procedures were: 0 min, 85% A, 15% B; 1 min, 80% A, 20% B; 2 min, 70% A, 30% B; 3 min, 55% A, 45% B; 4 min, 40% A, 60% B; 5 min, 25% A, 75% B; 5.10 min, 85% A, 15% B; 8 min, 85% A,15% B. The column temperature was 23 °C; the flow rate was 0.8 mL min^{-1}; the injection volume was 1 μL, and the wavelength was 450 nm.

DNA extraction and capillary electrophoresis

The leaf samples of each sweet potato accession were collected 45 days after planting, and the genomic DNA of the samples was extracted by the CTAB method[16]. The SSR primer sequences GDAAS0911, GDAAS0819, GDAAS0922, GDAAS0782, GDAAS0338, GDAAS0940, GDAAS0871, GDAAS0694 developed by the research group[17], and SPGS2 and SPGS3 developed by Meng et al. were used as primers[18]. SSR-PCR reaction system was 20 μL, containing 14.8 μL of ddH$_2$O, 0.4 μL of dNTP, 2 μL of buffer, 0.3 μL of forward and reverse primers each, 2 μL of DNA template, 0.2 μL of Taq. SSR-PCR amplification procedure: initial denaturation at 94℃ at 5 min, followed by denaturation at 94℃ for 30 s, renaturation at 54℃ for 35 s, and extension at 72℃ for 40 s. After 35 cycles, the reaction was terminated with the final extension time at 72℃ for 3 min. The capillary electrophoresis method: after mixing the formamide and the internal standard with a volume ratio of 100:1, 15 μL was added to the sample plate, and 1 μL of 10-fold diluted PCR product was added. Capillary electrophoresis was performed using a 3730 XL sequencer.

Data analysis

The Fragment (Plant) analysis software in Genemarker was used to analyze the original data obtained by the sequencer. The position of the molecular weight internal standard in each lane was compared with the position of the peak value of each sample. Using the 0/1 assignment method, the amplified bands at the same migration location were marked as 1, and the non-amplified bands were marked as 0. A binary metadata matrix of 0 and 1 was constructed. The distribution map of germplasm resources was drawn by ArcMap 10.7 software. The mean, variance, range, coefficient of variation (CV), and Shannon's diversity index (H') of 20 phenotypic traits were calculated using Microsoft Excel 2019 software. The correlation analysis was completed by the software Origin 2021.

The Euclidean distance between each germplasm of phenotypic traits was calculated by SPSS 26 software. The Euclidean distance matrix was imported into MEGA 11 software to obtain the NJ (Neighbor-Joining) clustering results of phenotypic traits, and the clustering map was drawn. The SSR molecular marker data were processed by NTSYSpc 2.10 software to obtain Nei's distance matrix of SSR molecular markers. The Nei's distance matrix was imported into MEGA 11 software to obtain the NJ (Neighbor-Joining) clustering results of phenotypic traits, and the clustering map was drawn. The average number of alleles (Na), the average number of effective alleles (Ne), Nei's genetic diversity index (He), Shannon's diversity index (I), and other genetic diversity indexes of SSR molecular markers were obtained through software POPGENE32.

STRUCTURE ver2.3.4[19] based on the Bayesian model-based clustering algorithm was used to explore the population structure of 132 sweet potato location accession based on 10 SSR markers. The data were tested from K min = 2 to K max = 20. The testing used 10 replications and a burn-in period of 10,000, followed by 100,000 Markov chain Monte Carlo (MCMC) iterations for each run. The optimal K capturing the major structure in the sweet potato data was determined using Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/)[20, 21]. The AMOVA analysis was done by GenALEx 6.5.

Results

Distribution of Chinese sweet potato landraces

132 sweet potato landraces were from different provinces in China. As shown in Fig. 1, the largest number of materials came from Guangdong Province, with a total of 87 materials, 15 materials from Hainan Province, 6 materials each from Guangxi Zhuang Autonomous Region and Guizhou Province, 5 materials from Yunnan Province, 4 materials each from Fujian Province and Sichuan Province, 1 material each from Jiangsu Province, Taiwan Province, Zhejiang Province, Hunan Province, and Anhui Province.

Diversity Analysis of the phenotypic traits

Most traits of sweet potato landraces showed significant phenotypic and physiological differences under investigation. The genetic diversity of 20 phenotypic traits of 132 Chinese sweet potato landraces was analyzed, and different phenotypic traits showed different diversity. As shown in Table 1, the color of top leaf, the color of top bud, and the leaf color accounted for the largest proportion of green, especially the leaf color, with a frequency as high as 79.55%, indicating that the overall color of plants was mainly green. The shape of top leaf and the shape of leaf were mostly incised. The main vein pigmentation color and the pigmentation of basic leaf vein were mostly purple, the side vein pigmentation color was mainly green, and the leaf apex shape was mainly acute. The petiole predominant color and the pigmentation of basic petiole were mainly investigated. Green was the dominant color of the petiole predominant color, with a frequency of 80.30%, while the pigmentation of basic petiole was mostly green and purple. Four traits of the vine were investigated, including the color of vine, the predominant color of vine, the secondary color of vine, and the vine tip pubescence. Among them, green was the highest in the color of vine and the predominant color of vine, both frequencies were close to 80%. The secondary color of vine was mainly purple with a small amount of none. The vine tip pubescence distribution was relatively uniform, with the highest frequency of none, followed by little. There were mainly two types of plant types, semi-erect and prostrate, mainly prostrate. Three traits were investigated for the storage root: the storage root shape, the skin color of storage root, and the predominant flesh color. The storage root shape was mainly elliptic, and the other storage root shapes were distributed except curve. The skin color of storage root was mauve, followed by white. The dominant color of predominant flesh color was white, followed by light yellow.

The coefficient of variation of 20 traits ranged from 92.25% to 17.83%, with a mean value of 45.84%. The highest coefficient of variation (92.25%) was found for stem end velvet and the lowest coefficient of variation (17.83%) was found for plant type. The Shannon diversity index of 20 traits ranged from 2.05 to 0.61, with an average value of 1.16. 8 traits were exceeding the average value, among which the Shannon diversity index of the skin color of storage root was the highest (2.05). Although mauve was the main color, other colors were also distributed, indicating that the genetic diversity of this trait was high. The Shannon diversity index of the leaf apex shape was the smallest (0.61), indicating that the genetic diversity of this trait was low.

Correlation analysis of the 20 phenotypic traits showed that out of 190 pairs, 43 pairs showed significant correlations ($P<0.05$) and 31 pairs showed highly significant correlations ($P<0.01$). There were 35 pairs of traits with significant positive correlations, including 24 pairs with highly significant positive correlations, and 8 pairs of traits with significant negative correlations, including 3 pairs with highly significant negative correlations. The highest correlation coefficient (0.83) was found for the shape of the terminal leaf and leaf traits, indicating a very high correlation between these two traits (Fig. 2).

Table 1 Phenotypic traits diversity of sweet potato landraces

Trait	0	1	2	3	4	5	6	7	8	9	10	S d in
Color of top leaf		Light green	Green	Purple green	Brown green	Light purple	Purple	Brown	Golden yellow	Red		1
		11.36	41.67	3.79	11.36	5.30	18.18	7.58	0.76	0.00		
Shape of top leaf		Round	Reniform	Cordate	Acuminate-cordate	Triangular	Incised					1
		0.00	0.00	30.30	12.12	5.30	52.27					
Color of top bud		Light green	Green	Light purple	Purple	Dark purple	Brown					1
		15.15	50.76	7.58	13.64	5.30	7.58					
Leaf color		Light green	Green	Purple green	Brown green	Light purple	Purple	Brown	Golden yellow	Red		0
		9.85	79.55	6.82	0.76	0.00	2.27	0.00	0.76	0.00		
Shape of leaf		Round	Reniform	Cordate	Acuminate-cordate	Triangular	Incised					0
		0.00	0.76	28.79	6.82	3.03	60.61					
Main vein pigmentataion		Light green	Green	Yellow	Light purple	Purple	Purple speckle					1
		9.85	17.42	0.76	18.18	38.64	15.15					
Side vein pigmentation		Light green	Green	Yellow	Light purple	Purple	Purple speckle					1
		9.85	31.82	0.76	17.42	28.03	12.12					
Pigmentation of basic leaf vein		Light green	Green	Light purple	Purple	Dark purple						1
		0.00	16.67	13.64	56.82	12.88						
Leaf apex shape	Absent	Acute	Blunt									0
	0.00	69.70	30.30									
Petiole predominant color		Light green	Green	Light purple	Purple	Dark purple						0
		2.27	80.30	7.58	6.82	3.03						
Pigmentation of basic petiole		Light green	Green	Light purple	Purple	Dark purple						1
		0.76	38.64	17.42	35.61	7.58						
Color of vine		Light green	Green	Brown	Light purple	Purple	Dark purple					0
		1.52	78.79	5.30	0.00	8.33	6.06					
Predominant color of vine		Light green	Green	Mauve	Light purple	Purple	Dark purple	Brown				0
		1.52	77.27	2.27	0.00	12.88	3.79	2.27				
Secondary color of vine	Absent	Green	Mauve	Purple	Brown							1
	32.58	16.67	1.52	38.64	10.61							
Vine tip pubescence	None	Little	Moderate	More								0

	35.61	25.76	18.94	19.70							
Plant type	Erect	Semi-erect	Prostrate	Scramble							0
	0.00	34.09	65.91	0.00							
Vigour of plant	Strong	Intermediate	Weak								0
	43.18	46.97	9.85								
Storage root shape	Rotundity	Short elliptic	Elliptic	Long elliptic	Obovate	Ovate	Rectangle	Curve	Anomaly		1
	1.52	3.03	43.94	10.61	10.61	21.97	6.82	0.00	1.52		
Skin color of storage root	White	Light yellow	Brown yellow	Yellow	Brown	Pink	Red	Mauve	Purple	Dark purple	2
	18.94	6.82	4.55	9.85	4.55	6.82	12.12	26.52	9.09	0.76	
Predominant flesh color	White	Light yellow	Brown yellow	Yellow	Brown	Pink	Red	Mauve	Purple	Dark purple	1
	32.58	25.00	19.70	13.64	4.55	0.00	0.76	0.00	3.79	0.00	

CV: the coefficient of variation.

Diversity of carotenoids

The content of 13 carotenoid monomers and the total content (the sum of 13 carotenoid monomers) of 132 sweet potato landraces in China were determined. The carotenoids in different sweet potato varieties were different. As shown in Table 2, 2 out of 13 carotenoid monomers were not detected in all materials and no carotenoids were detected in 38 materials. Among the 11 detected monomers, the average content of β-carotene was the highest, and the average content of β-cryptoxanthin was the lowest. The average total content of carotenoids was 6.33 µg/g, while the average content of β-carotene was 4.32 µg/g. The variety with the highest total carotenoid content (59.90 µg/g) also has very high β-carotene content (55.54 µg/g). After counting, the 55 materials had the highest total carotenoid content of β-carotene, indicating that β-carotene may be the predominant carotenoid in the sweet potato. Carotenoid monomer content and total content had a high coefficient of variation, indicating that there were significant differences in carotenoid content between Chinese sweet potato landraces, and the carotenoid content of landraces was rich in diversity.

Correlation analysis of 11 carotenoid monomers and total content showed that among the 66 pairs of combinations, 27 pairs of carotenoid monomers showed a significant positive correlation ($P<0.05$), of which xx pairs of carotenoid monomers showed a highly significant positive correlation ($P<0.01$); 10 carotenoid monomers showed a significant positive correlation between monomeric substances and total content, of which 9 carotenoid monomers showed a significant positive correlation between monomeric substances and The correlation between 10 carotenoid monomers and the total content was highly significant, and 9 carotenoid monomers and the total content were highly significant. The highest correlation coefficients among carotenoid monomer substances were lutein and α-cryptoxanthin (0.99), indicating a very high correlation between these two monomer substances; the highest correlation coefficient with the total content was β-carotene (0.99), indicating a very high correlation between β-carotene and the total content (Fig. 2).

Table 2 Carotenoid content diversity of sweet potato landraces

	Mean	Standard Deviation	Max	Min	CV (%)
Violaxanthin	0.802	0.715	3.244	0.000	89.174
Neoxanthin	0.253	0.545	2.850	0.000	215.294
Antheraxanthin	0.884	0.988	4.140	0.000	111.734
Lutein	0.006	0.056	0.638	0.000	1009.910
Zeaxanthin	0.004	0.036	0.342	0.000	822.049
α-Cryptoxanthin	0.007	0.079	0.916	0.000	1107.100
ε-Carotene	0.016	0.082	0.515	0.000	518.563
β-Cryptoxanthin	0.002	0.011	0.096	0.000	683.719
α-Carotene	0.000	0.001	0.016	0.000	1144.552
β-Carotene	4.321	10.360	57.417	0.000	239.763
6R-δ-Carotene	ND	ND	ND	ND	ND
γ-Carotene	0.037	0.129	1.151	0.000	348.041
Lycopene	ND	ND	ND	ND	ND
All	6.331	11.470	59.904	0.000	181.170

CV: the coefficient of variation. ND: not detected.

SSR markers diversity

A total of 135 gene loci were amplified by PCR using 10 pairs of SSR primers. In all loci, the average number of alleles (Na) was 1.68, the average effective number of alleles (Ne) was 1.25, there were 92 polymorphic loci, the percentage of polymorphic loci was 68.15%, Nei's genetic diversity index (He) was 0.15, and Shannon's diversity index (I) was 0.24. Among single primers, the minimum value of Na was 1.44 (SPGS2) and the maximum value was 1.91 (GDAAS0922). The minimum value of Ne was 1.12 (SPGS2) and the maximum value was 1.42 (GDAAS0782). The minimum of H ' is 0.08 (SPGS2) and the maximum is 0.25 (GDAAS0782). The minimum value of I ' was 0.13 (SPGS2) and the maximum value was 0.38 (GDAAS0782) (Table 3).

Population structure analysis

According to the results of STRUCTURE, the highest value of ΔK was obtained at K=5 (Fig. 3a). Based on this result, 132 sweet potato landraces were divided into 5 sub-populations (Fig. 3c). Sub-population I consisted of 37 materials, accounting for 28.03 %, which were from Guangdong (23), Hainan (5), Sichuan (3), Guangxi (2), Guizhou (2), Fujian (1), and Yunnan (1). There are 27 materials in sub-population II, accounting for 20.45%, respectively from Guangdong (23), Hainan (3), and Guizhou (1). Guangdong materials account for the highest proportion (85.19%) in this sub-population Sub-population III had 14 materials, accounting for 10.61 %, respectively from Guangdong (9), Hainan (3), Yunnan (1), and Anhui (1). There were 17 materials in sub-population IV, accounting for 12.88 %, which were from Guangdong (12), Hainan (2), Guangxi (2), and Hunan (1). There were 37 materials in sub-population V, accounting for 28.03 %, which were from Guangdong (20), Guizhou (3), Fujian (3), Yunnan (3), Guangxi (2), Hainan (2), Zhejiang (1), Taiwan (1), Jiangsu (1) and Sichuan (1). As shown in the phylogenetic tree, the distance between sub-population I and sub-population V was the closest, indicating that sub-population I and sub-population V had a similar genetic relationship, and the distance between sub-population III and sub-population IV was the farthest, indicating that the genetic relationship between the two was very far (Fig. 3d).

The mean fixation index (Fst) of 5 sub-populations was determined. The Fst value of sub-population III was the highest (0.73), followed by sub-population IV (0.49), sub-population II (0.32), and sub-population V (0.21), and the Fst value of sub-population I was the lowest (0.01) (Table 4). The Fst value of sub-population I was between 0-0.05, indicating that its genetic differentiation was very small. The Fst value of sub-population V was between 0.05 and 0.25, indicating that there was moderate genetic differentiation. The Fst values of sub-population II, III, and IV were all greater than 0.25, indicating that these three sub-populations had great genetic differentiation. AMOVA analysis by GenALEx showed that the genetic variation in this species was mainly within populations, accounting for 84% of the total variation, and between populations accounted for 16% of the total variation (Fig. 3b).

Table 3 Genetic diversity analysis of 10 pairs of SSR molecular markers in sweet potato landraces

Primer name	Number of samples	Total number of strips	Number of polymorphic loci	The percentage of polymorphic loci (PIC)	The average number of alleles (Na)		The average number of effective alleles (Ne)		Nei's genetic diversity index (He)		Shannon's diversity index (I)	
					Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
GDAAS0338	132	21	17	80.95%	1.81	0.39	1.20	0.33	0.12	0.18	0.19	0.25
GDAAS0694	132	11	5	45.45%	1.45	0.50	1.19	0.31	0.11	0.18	0.17	0.26
GDAAS0782	132	10	8	80.00%	1.80	0.40	1.42	0.36	0.25	0.19	0.38	0.26
GDAAS0819	132	11	7	63.64%	1.64	0.48	1.29	0.34	0.18	0.18	0.27	0.26
GDAAS0871	132	8	7	87.50%	1.88	0.33	1.26	0.34	0.16	0.19	0.25	0.26
GDAAS0911	132	18	13	72.22%	1.72	0.45	1.30	0.35	0.18	0.19	0.28	0.26
GDAAS0922	132	17	12	70.59%	1.71	0.46	1.29	0.36	0.17	0.20	0.26	0.27
GDAAS0940	132	11	10	90.91%	1.91	0.29	1.25	0.29	0.17	0.16	0.27	0.22
SPGS2	132	23	10	43.48%	1.43	0.50	1.11	0.21	0.08	0.13	0.13	0.20
SPGS3	132	5	3	60.00%	1.60	0.49	1.37	0.38	0.22	0.20	0.32	0.28
All	132	135	92	68.15%	1.68	0.47	1.25	0.33	0.15	0.18	0.23	0.26

Table 4 The mean fixation index (Fst) of 5 sub-populations

Sub-population	Fst
	0.01
	0.32
	0.73
	0.49
	0.21

Fst: the mean fixation index.

Discussion

Genetic diversity of sweet potato based on phenotypic traits

In agricultural breeding, the use of germplasm resources is closely related to phenotypic traits, which can provide useful pre-breeding data in agriculture[14]. Compared to molecular markers, phenotypic traits have the advantage of being more intuitive and convenient for germplasm resource evaluation. Liu et al. evaluated the genetic diversity of 215 sugar beet germplasm resources for phenotypic traits by correlation analysis, principal component analysis, and cluster analysis[22]; Zhang et al. used 39 fruit phenotypic traits from 570 pear materials to identify the genetic diversity of pears[23]. In this study, a total of 132 landraces of sweet potato were investigated for 20 phenotypic traits, including the color of top leaf, the shape of top leaf, the color of top bud, the leaf color, the shape of leaf, the main vein pigmentation color, the side vein pigmentation color, the pigmentation of basic leaf vein, the leaf apex shape, the petiole predominant color, the pigmentation of basic petiole, the color of vine, the predominant color of vine, the secondary color of vine, the vine tip pubescence, the plant type, the vigour of plant, the storage root shape, the skin color of storage root, and the predominant flesh color. The pair of traits with the highest correlation were the shape of top leaf and the shape of leaf, and these two traits had the same evaluation criteria. Therefore, we compared the specific trait statistics of the shape of top leaf and the shape of leaf in 132 sweet potato landraces and found that 103 of the 132 sweet potato landraces had the same trait assignment. From this phenomenon, it can be assumed that in sweet potato germplasm, the shape of the leaf can be roughly inferred from the shape of the top leaf, and the shape of the top leaf may influence the shape of the leaf. The traits significantly correlated with the predominant flesh color were the color of vine, the predominant color of vine, the vine tip pubescence, the plant type, the vigour of plant, and the skin color of storage root, among which the color of vine, the predominant color of vine, the vine tip pubescence and the skin color of storage root were significantly positively correlated and the color of vine was highly significantly positively correlated; the plant type and stem and the vigour of plant were significantly negatively correlated and the vigour of plant was highly significantly negatively correlated. It can be assumed that the darker and purpler the flesh color, the darker and closer to purple the color of vine, the predominant color of vine, and the skin color of storage root, and the more erect the plant type, the stronger the vigour of plant, the more pubescence at the vine tip.

Carotenoids are an important nutrient. Among sweet potatoes, those with orange flesh tend to contain higher levels of carotenoids. Also, sweet potatoes with orange flesh are rich in β-carotene[5, 24]. This explains that in the carotenoid monomer, there is a very high correlation coefficient between β-carotene and total content. The unique chemical structure of beta-carotene makes it the best pro-carotenoid for vitamin A[7]. The results of the correlation analysis revealed a

very high correlation coefficient between lutein and α-cryptoxanthin as well as β-carotene and total content of carotenoid monomeric substances, with a highly significant positive correlation. In the synthetic pathway, α-cryptoxanthin is synthesized by the action of β-carotene hydroxylase (CHYB), and when there is a large amount of α-cryptoxanthin present, the presence of CYHB makes lutein also present in large amounts. β-carotene mainly appears orange and is the main carotenoid in orange flesh sweet potatoes[25]. Among 132 local varieties of sweet potatoes, 38 varieties had no carotenoids detected, and the main color of the flesh of these varieties was mostly purple or white; 19 varieties had a total carotenoid content of more than 10 μg·g^{-1} and three of them had a total carotenoid content of more than 50 μg·g^{-1}, and the main color of the flesh of these varieties was usually orange; 75 varieties had a small number of carotenoids detected These varieties are mostly yellow or light yellow. Among the 132 landraces, the proportion of orange-fleshed varieties with high carotenoid content was low and the proportion of yellow and white-fleshed varieties with low carotenoid content was high.

Genetic diversity and population structure of sweet potato landraces based on SSR markers

Although phenotypic traits can visually reflect the differences between materials, they are easily influenced by the natural environment, while the analysis of germplasm resources by molecular markers can truly reflect the differences between materials, avoiding the influence of the environment. SSR molecular markers are widely used in the study of genetic diversity; Karcı et al. used 92 SSR molecular markers to analyze the genetic diversity of 66 Pistachio varieties and genotypes from different geographical sources[26]; Kimaro et al. used 33 SSR molecular markers to assess the genetic diversity and genetic relatedness of 48 Pigeonpea entries[27]; Zhu et al. used 45 SSR markers from the sorghum genome to genotype 140 sorghum accessions and assess the genetic diversity of sorghum[28]. In this study, 132 local varieties of sweet potato were analyzed for genetic diversity by 10 pairs of SSR molecular markers. 135 loci were examined, and 92 polymorphic loci were detected, accounting for 68.15%. the percentage of polymorphic loci for the 10 pairs of SSR molecular markers ranged from a minimum of 43.48% to a maximum of 90.91%. the 10 pairs reported by Meng et al. and Luo et al. The PIC of these 10 pairs of SSR molecular markers reported by Meng et al. and Luo et al. were 82.82%, 62.35%, 86.46%, 91.57%, 72.78%, 95.93%, 91.12%, 81.22%, 81.12%, and 58.24%, respectively[17, 18]. The *t*-test results indicated that the PICs of the 10 pairs of SSR molecular markers detected by Meng et al. and Luo et al. were not significantly different from those in this study.

Based on population structure analysis, the 132 sweet potato local varieties were divided into five subpopulations, which contained 37, 27, 14, 17, and 37 materials, respectively, with the majority of materials from Guangdong in each subpopulation, and it can be assumed that these local varieties may have evolved from the local varieties in Guangdong Province. Sweet potato has been introduced to China for about 400 years since the late 16th century, and Guangdong was one of the first provinces to cultivate sweet potato[29]. By comparing these five subpopulations on the phylogenetic tree, it was found that subpopulations I and V were the closest relatives, followed by subpopulation II, while subpopulation III was the most distantly related. This result was consistent with the Fst of each subpopulation, and the more distant the affinity of the subpopulation, the higher the degree of genetic differentiation, and the closer it was, the lower the degree of genetic differentiation. Based on the results of AMOVA analysis, the genetic variation within populations (84%) of these five subpopulations was much higher than the genetic variation among populations (16%), indicating that the genetic differences between different groups were smaller and the intra-group genetic differences were larger. Therefore, genetic improvement of sweet potato local varieties can be done by considering the selection of materials among populations.

Conclusion

Genetic diversity analysis of 20 phenotypic traits was performed by distribution frequency, coefficient of variation, and Shannon's diversity index using 132 sweetpotato local germplasm as test materials, and the average coefficient of variation was 45.84% and the average Shannon's diversity index was 1.16, with a wide distribution of most traits. 11 carotenoid monomers as well as the total content had high coefficients of variation, and the average total content was 6.33 μg/g. In the genetic diversity results of SSR molecular markers, the overall Nei's genetic diversity index of sweetpotato local germplasm was 0.15 and the overall Shannon's diversity index was 0.23. In conclusion, the 132 sweetpotato local germplasm had good genetic diversity. Based on population structure analysis, the 132 sweetpotato local germplasm could be divided into five subgroups. Analysis of molecular variation (AMOVA) showed that 84% of the variation existed within populations and 16% between populations.

Declarations

Acknowledgments

Authors' contributions

CYH and WZY designed the experiments, analyzed the data, and wrote the main manuscript text. CYH, TCC, ZR, and JJW performed the experiments. All authors reviewed and approved the manrscript.

Funding

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article or are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

In this study, we collected fresh sweet potato plant material. The plant samples were identified by the team of researcher Wang Zhangying at the Institute of Crop Research Institute, Guangdong Academy of Agricultural Sciences. The study complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] College of Grassland Agriculture, Northwest A&F University, Yangling 712100, China.

[2] Crops Research Institute, Guangdong Academy of Agricultural Sciences/Guangdong Provincial Key Laboratory of Crop Genetic Improvement, Guangzhou 510640, China.

References

1. O'Brien PJ: **The Sweet Potato: Its Origin and Dispersal**. *AM ANTHROPOL* 1972, **74**(3):342-365.
2. Fang BP, Zhang XJ, Chen JY, An K: **The history and status of sweet potato germplasm research in China**. *Guangdong Agricultural Sciences* 2004(z1):3-5.
3. Stange C: **Carotenoids in Nature**: Springer International Publishing; 2016.
4. Pasquariello R, Anipchenko P, Pennarossa G, Crociati M, Zerani M, Brevini TA, Gandolfi F, Maranesi M: **Carotenoids in female and male reproduction**. *PHYTOCHEMISTRY* 2022, **204**:113459.
5. Islam SN, Nusrat T, Begum P, Ahsan M: **Carotenoids and β-carotene in orange fleshed sweet potato: A possible solution to vitamin A deficiency**. *FOOD CHEM* 2016, **199**:628-631.
6. Gong X, Draper C, Allison G, Marisiddaiah R, Rubin L: **Effects of the Macular Carotenoid Lutein in Human Retinal Pigment Epithelial Cells**. *ANTIOXIDANTS-BASEL* 2017, **6**(4):100.
7. Drapal M, Fraser PD: **Determination of carotenoids in sweet potato (Ipomoea batatas L., Lam) tubers: Implications for accurate provitamin A determination in staple sturdy tuber crops**. *PHYTOCHEMISTRY* 2019, **167**:112102.
8. Przybylska S: **Lycopene – a bioactive carotenoid offering multiple health benefits: a review**. *International Journal of Food Science & Technology* 2019, **55**(1):11-32.
9. Wilson EO: **The Biological Diversity Crisis**. *BIOSCIENCE* 1985.
10. Magurran AE: **Ecological Diversity and Its Measurement**: Biometrics; 1988.
11. Rajput SG, Santra DK: **Evaluation of Genetic Diversity of Proso Millet Germplasm Available in the United States using Simple-Sequence Repeat Markers**. *CROP SCI* 2016, **56**(5):2401-2409.
12. Belalia N, Lupini A, Djemel A, Morsli A, Mauceri A, Lotti C, Khelifi-Slaoui M, Khelifi L, Sunseri F: **Analysis of genetic diversity and population structure in Saharan maize (Zea mays L.) populations using phenotypic traits and SSR markers**. *GENET RESOUR CROP EV* 2019, **66**(1):243-257.
13. Golkar P, Nourbakhsh V: **Analysis of genetic diversity and population structure in Nigella sativa L. using agronomic traits and molecular markers (SRAP and SCoT)**. *IND CROP PROD* 2019, **130**:170-178.
14. Sharma D, Nanjundan J, Singh L, Parmar N, Singh KH, Verma KS, Thakur AK: **Genetic Diversity and Population Structure Analysis in Indian Mustard Germplasm Using Phenotypic Traits and SSR Markers**. *PLANT MOL BIOL REP* 2022, **40**(3):579-594.
15. Zhang YG, Fang BP: **Descriptors and Data Standard for Sweetpotato [Ipomoea batatas(L.)Lam.]**. Beijing China: China Agriculture Press; 2006.
16. Yu D, Zhang J, Tan G, Yu N, Wang Q, Duan Q, Qi X, Cheng M, Yan C, Wei Z *et al*: **An easily-performed high-throughput method for plant genomic DNA extraction**. *ANAL BIOCHEM* 2019, **569**:28-30.
17. Luo ZX, Fang BP, Li R, Wang ZY, Huang LF, Cheng JY, Zhang XJ, Li YJ, Chen XL: **Construction of DNA Fingerprint Database Based on EST-SSR Markers for Sweet Potato Germplasm**. *Journal of Plant Genetic Resources* 2014, **15**(4):810-814.
18. Meng YS, Zhao N, LI H, Zhai H, He SZ, Liu QC: **SSR fingerprinting of 203 sweetpotato (Ipomoea batatas (L.) Lam.) varieties**. *J INTEGR AGR* 2018, **17**(1):86-93.
19. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations**. *AM J HUM GENET* 2000.
20. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software structure: a simulation study**. *MOL ECOL* 2005, **14**(8):2611-2620.
21. Earl DA, VonHoldt BM: **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method**. *CONSERV GENET RESOUR* 2012, **4**(2):359-361.

22. Liu D, Wang X, Li W, Li J, Tan W, Xing W: **Genetic Diversity Analysis of the Phenotypic Traits of 215 Sugar Beet Germplasm Resources**. *SUGAR TECH* 2022, **24**(6):1790-1800.

23. Zhang Y, Cao Y, Huo H, Xu J, Tian L, Dong X, Qi D, Liu C: **An assessment of the genetic diversity of pear (Pyrus L.) germplasm resources based on the fruit phenotypic traits**. *J INTEGR AGR* 2022, **21**(8):2275-2290.

24. Rios-Romero EA, Ochoa-Martínez LA, Bello-Pérez LA, Morales-Castro J, Quintero-Ramos A, Gallegos-Infante JA: **Effect of ultrasound and steam treatments on bioaccessibility of β-carotene and physicochemical parameters in orange-fleshed sweet potato juice**. *HELIYON* 2021, **7**(4):e6632.

25. Jia R, Zhang R, Gangurde SS, Tang C, Jiang B, Li G, Wang Z: **Integrated analysis of carotenoid metabolites and transcriptome identifies key genes controlling carotenoid compositions and content in sweetpotato tuberous roots (Ipomoea batatas L.)**. *FRONT PLANT SCI* 2022, **13**.

26. Karcı H, Paizila A, Güney M, Zhaanbaev M, Kafkas S: **Revealing genetic diversity and population structure in Pistachio (Pistacia vera L.) by SSR markers**. *GENET RESOUR CROP EV* 2022, **69**(8):2875-2887.

27. Kimaro D, Melis R, Sibiya J, Shimelis H, Shayanowako A: **Analysis of Genetic Diversity and Population Structure of Pigeonpea [Cajanus cajan (L.) Millsp] Accessions Using SSR Markers**. *Plants* 2020, **9**(12):1643.

28. Zhu M, Chen J, Yuyama N, Luo L, Xiao X, Lv Y, Liu Y, Cai H: **Genetic Diversity and Population Structure of Broomcorn Sorghum Investigated with Simple Sequence Repeat Markers**. *TROP PLANT BIOL* 2020, **13**(1):62-72.

29. Shao K, Bu FX: **Crops Introduction and Spreading in Ming and Qing Dynasties**. *Jounal of Anhui Agricultural Sciences* 2007(22):7002-7003.
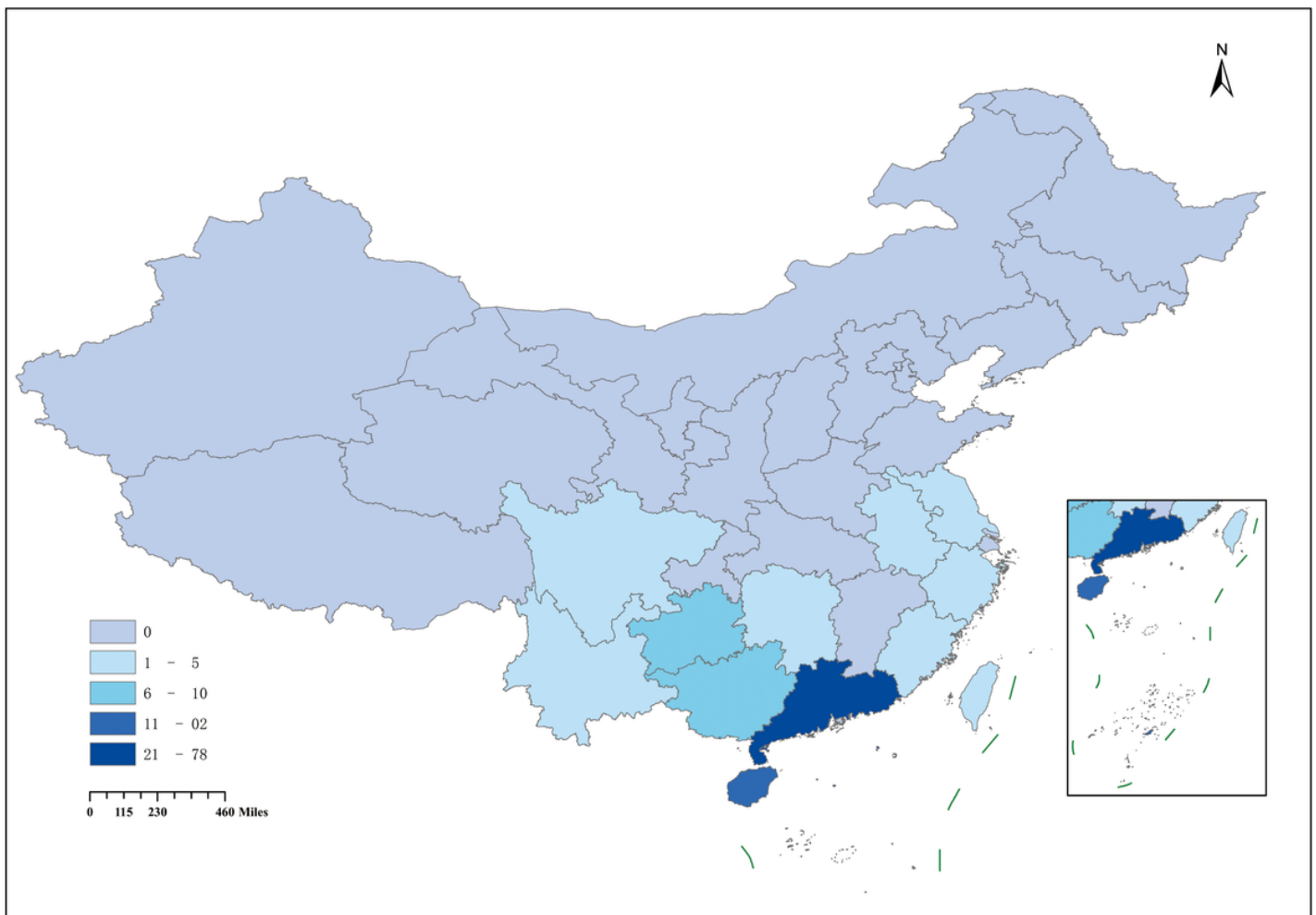
Figures

Figure 1

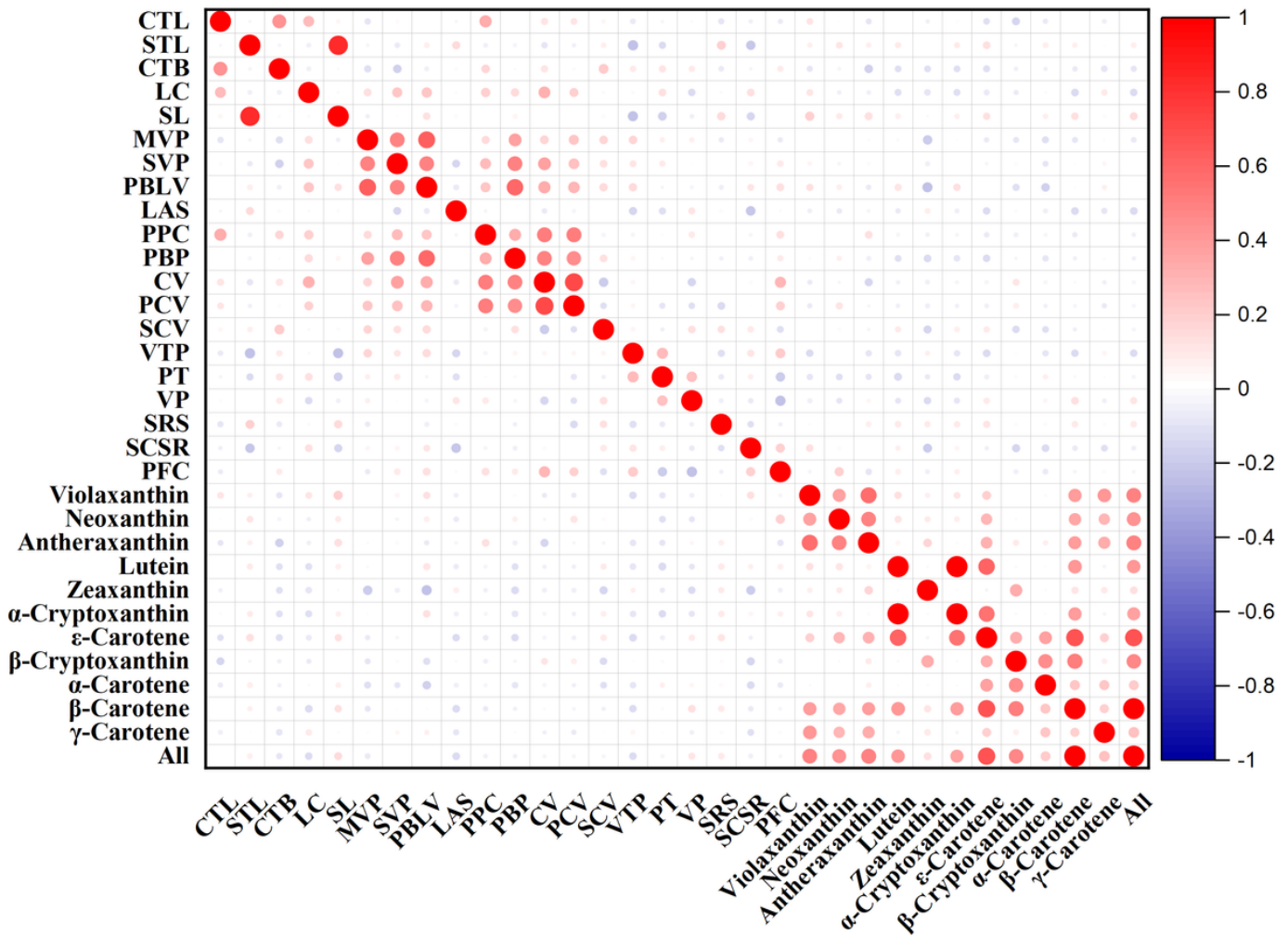Source information of sweet potato landraces

Figure 2

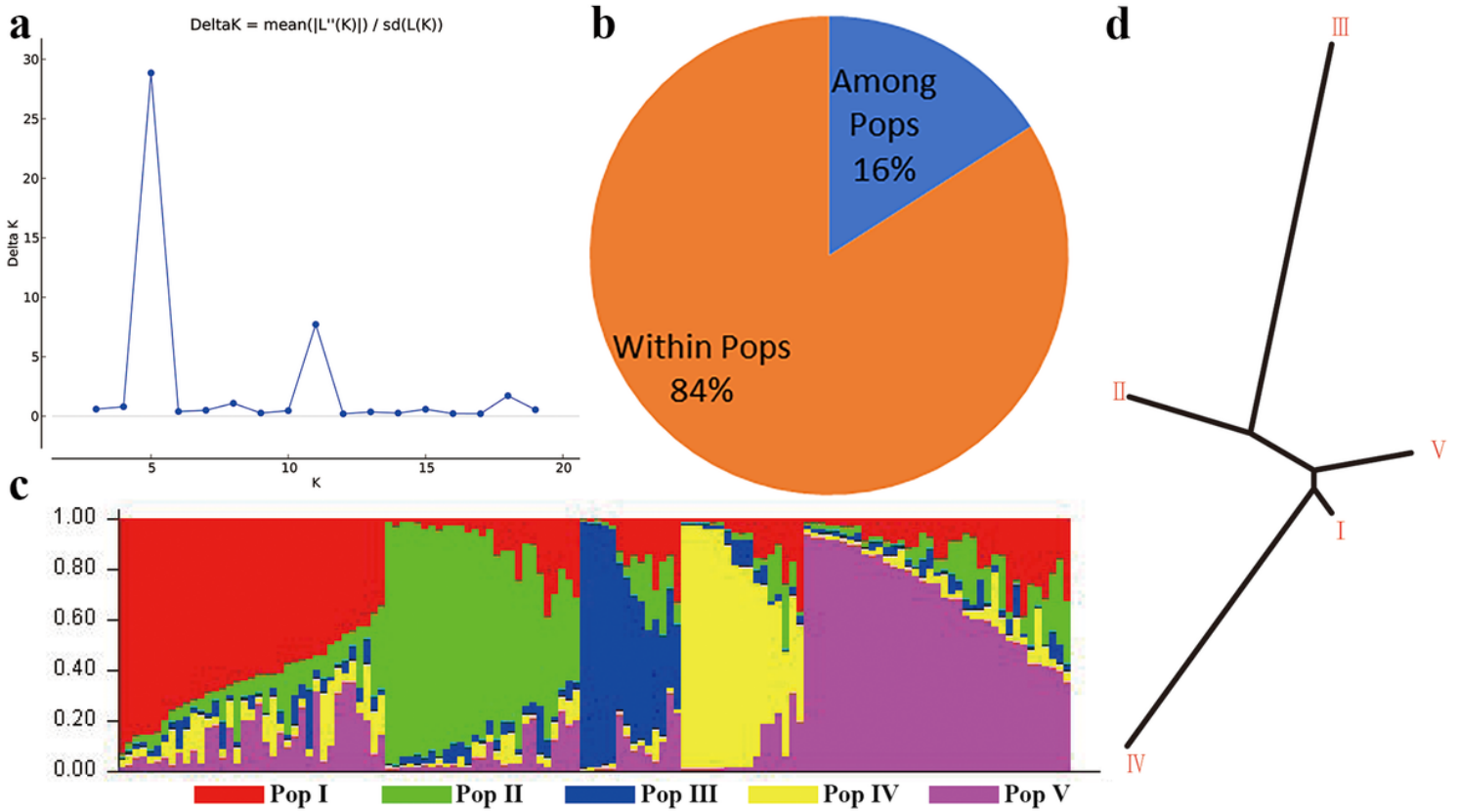Correlation between phenotypic traits and carotenoid content

Figure 3

Population structure analysis of 132 sweet potato landraces. A, distribution of ΔK. K, the optimal number of genetic groups. B, AMOVA analysis of 5 sub-populations. C, the population structure of sweet potato landraces. D, the evolutionary tree of 5 sub-populations.