Research Paper

Assessment of genetic diversity in *Camellia oleifera* Abel. accessions using morphological traits and simple sequence repeat (SSR) markers

Zhilong He[1,2,3], Caixia Liu[1,2], Xiangnan Wang[1,2], Rui Wang*[1,2], Yongzhong Chen[1,2] and Yun Tian[3]

[1] *Hunan Academy of Forestry*, Shaoshan South Road, No. 658, Changsha 410004, China
[2] *National Engineering Research Center for Oil Tea Camellia*, Changsha 410004, China
[3] *Hunan Agricultural University*, Changsha 410128, Hunan, China

Camellia oleifera Abel. (*C. oleifera*) is a cultivable plant with important economic value. It is very helpful for the scientific utilization, cultivation and preservation of germplasm resources through evaluating the genetic diversity. In this study, we estimated the genetic relationship of 150 accessions of *C. oleifera* using morphological and economic traits, as well as SSR molecular marker. Through the variation and cluster analysis of 17 morphological and economic traits, the germplasm was divided into a candidate core breeding group with higher economic traits and a core breeding group with higher morphological traits. The genetic similarity coefficients of SSR markers ranged from 0.05 to 0.91, and the germplasm materials were divided into five groups. The results demonstrated that *C. oleifera* germplasms perform a rich genetic variation. This is the first report to evaluate the genetic diversity of different *C. oleifera* germplasms using the morphological and economic traits, together with SSR molecular marker, and the results allow us to find evidence for the origin of varieties, establish core breeding populations and its fingerprint.

Key Words: *Camellia oleifera*, genetic diversity, phenotypic traits, SSR.

Introduction

Camellia oleifera Abel. (*C. oleifera*), belonging to genus *Camellia* in the family *Theaceae*, originates from China and is one of the world's four major oil tree species (together with oil palm, olive and coconut) (Chen *et al.* 2015). It has been cultivated in south-central and southern China for more than 2,000 years (Zhu *et al.* 2020), due to its major product (nearly 50% of dry kernel weight), tea oil, is an edible oil known as "eastern olive oil" with high nutritional value and health benefits. The crop covers a huge area of approximately 4.4 million hectares with abundant resources and an annual output of over 2.4 million tons of seeds, and its oil yield is more than 0.6 million tons, and is distributed in the Hunan, Jiangxi, Guangxi, Fujian, Hainan and provinces (Liu *et al.* 2018). Recently, *C. oleifera* has drawn growing attention because of its highly commercial, medic, cosmetic and ornamental value. In addition, *C. oleifera* species are viewed as important germplasm resources with valuable gene pool for the development of new varieties of oil tree species.

Due to the diversity of *C. oleifera* germplasm resources,

a variety of excellent individual plants, clones and families have been bred and cultivated in China. Phenotypic diversity is a comprehensive manifestation of the interaction between biological genetics and environmental factors, which is one of the major topics of genetic diversity research (Costa *et al.* 2017). Detecting genetic variation from phenotypic traits is a simple and easy method, which can reveal the magnitude of genetic variation to a certain extent, and the assessment of phenotypic variation of different individuals also will be beneficial to the management and utilization of germplasm (Zhou *et al.* 2015).

Generally, there is a lot of redundancy of *C. oleifera* germplasms in the circumstance that the morphology of *C. oleifera* varies under the cooperation of genetic background and environment, which due partly to same varieties with different names. Moreover, the absence of flowers and fruits for the majority of the year in *C. oleifera*, especially slight changes of leaf shape exhibits. makes it difficult to distinguish the varieties based on the morphology of leaves, flowers and fruits.

Commonly, the genetic diversity studies of plant germplasm resources are capable of reducing the redundancy of germplasm conservation and promoting the construction of core germplasm resources, which is important for the effective utilization of genetic resources in plant breeding (Verma *et al.* 2019). Nowadays, the techniques in genetic diversity research at the DNA level is contributed

to understand the genetic relationship and distribution. Although the studies on molecular biology of *C. oleifera* has been brought for nearly twenty years, there is still few reports about it. The random amplified polymorphic DNA (RAPD) markers were firstly applied to analysis of genetic diversity of 32 superior hybrid varieties and wild type varieties of *C. oleifera* in 2003 (Zhang 2003). In 2007, sixteen the inter-simple sequence repeat (ISSR) primers were developed and used for analysis of genetic diversity of 10 superior clone of *C. oleifera* (Zhang *et al.* 2007). The earliest report about the application of Sequence-related amplified polymorphism (SRAP) markers in *C. oleifera* identified the genetic difference and evaluated the relationship of 48 individuals from 5 different regions (Zhang *et al.* 2011). Among these genetic techniques, the simple sequence repeat (SSR) is one of the powerful molecular marker technologies established in 1985 (Jeffreys *et al.* 1985). Over the past decade, the researches that SSR molecular marker used in *Camellia genus* have been reported. Liu *et al.* (Liu *et al.* 2011) optimized the SSR-PCR reaction system and proved the inter-species versatility of microsatellite loci in *Camellia* genus. Huang first applied SSR markers to study the genetic structure and inter-species hybrid introgression between *C. oleifera* and *C. meiocarpa* (Huang 2013). However, the limited genetic studies have been performed on the genetic evaluation of *C. oleifera* species over 100 accessions from various series using SSR markers.

At present, to the best of our knowledge, the combined phenotypic characteristics and SSR markers has not been applied in *C. oleifera* species for its genetic diversity assessment. Therefore, in this study, the genetic diversity analysis of *C. oleifera* germplasm materials was performed using phenotypic characteristics and SSR markers. The results indicate that comprehensive utilization of phenotypic characteristics and SSR markers can be efficient in the analysis of genetic diversity in *C. oleifera* for plant breeding.

Materials and Methods

Plant materials

A collection of 150 accessions of *C. oleifera* was used in this study (**Table 1**). All these accessions were planted in the experimental field at the National *Camellia Oleifera* Germplasm Resource Preservation Center, Hunan, China. The germplasm materials from H001 to H029 were ancient *C. oleifera* trees that living for over three hundred years. The rest germplasm materials were either collected or bred by Hunan Academy of Forestry then were grafted to scions and planted in field for about fifteen years.

Measurement of quantitative traits

Nine branches and leaves of each *C. oleifera* germplasm materials were applied to the measurement of morphological traits. Fifteen fruits together with their seeds of each

C. oleifera germplasm materials were applied to the measurement of morphological and economic traits.

DNA extraction and PCR analysis

Genomic DNA was extracted from the young leaves of *C. oleifera* by a modified cetyltrimethyl ammonium bromide (CTAB) method (Huang *et al.* 2018). A set of 50 pairs of SSR primers (Chen *et al.* 2016, Zhou *et al.* 2017) were initially screened, and 9 pairs of primers with good, clear banding patterns were selected for analysis of genetic diversity (**Table 4**). SSR amplification was performed as follows: the reaction conditions were 50 ng of template DNA, 0.8 μL of each primer at 10 μM and 10 μL 2 × Taq PCR Master Mix in a total reaction volume of 20 μL. The reaction mixture was initially denatured at 94°C for 4 min, followed by 35 cycles of 94°C for 30 s, annealing temperature according to **Table 2** for 45 s, and 72°C for 2 min; and a final extension at 72°C for 10 min. Fluorescently labeled PCR products were analyzed concurrently with the GeneScan™-500LIZ™ Size Standard on an ABI 3730XL sequencer, and sizes were determined with GENEMAPPER version 4.0. Due to the putatively polyploid character of *C. oleifera*, allele frequencies, polymorphism information content (PIC) were calculated using the methods described by Huang (2013).

Data analysis

SSR amplified bands in the gel were manually scored as present (1) or absent (0). Only the consistently reproducible bands were scored, and the smeared and weak bands were excluded. The resulting 1/0 data matrix was used to assess the level of genetic diversity of *C. oleifera*. The obtained genetic distance matrix was then used to perform the cluster analysis and construct the un-weighted pair-group method with arithmetic average (UPGMA) dendrogram using NTsys-pc 2.1 software (Rohlf 2000).

Results

Variation of quantitative traits

Quantitative traits of morphological and economic characteristics such as size and weight of fruits and seeds, as well as productions of fruits, seeds and oil were measured and analyzed in this study (**Fig. 1**). As shown in **Table 2**, the seventeen quantitative traits of 150 *C. oleifera* germplasm materials showed relatively high variation levels, due to the change fold of maximum compare to minimum in each quantitative traits ranged from 1.95 to 8.72, while their coefficient of variation (CV) ranged from 12.20% to 40.11%.

The heatmap was established with normalized quantitative traits (**Fig. 2**). The clustering results showed that 150 germplasm materials could be divided into two groups. The first group consists of 55 germplasms with relatively high levels of economic traits than the other 95 germplasms from the second group, as well as relatively small sizes of

Table 1. Information of *C. oleifera* germplasm materials in this study

No.	Name	Origin	Location	No.	Name	Origin	Location
H001	YC	Hengnan, Hunan	E 112°68′, N 26°74′	H076	XL197	Yongzhou, Hunan	E 113°35′, N 27°00′
H002	YC	Hengnan, Hunan	E 112°68′, N 26°74′	H077	XL210	Changsha, Hunan	E 112°99′, N 28°26′
H003	YC	Shimen, Hunan	E 111°38′, N 29°58′	H078	XL218	Miluo, Hunan	E 113°07′, N 28°81′
H004	YC	Shimen, Hunan	E 111°38′, N 29°58′	H079	XL228	Yongzhou, Hunan	E 113°35′, N 27°00′
H005	YC	Shimen, Hunan	E 111°38′, N 29°58′	H080	XL258	Miluo, Hunan	E 113°07′, N 28°81′
H006	YC	Shimen, Hunan	E 111°38′, N 29°58′	H081	XL349	Youxian, Hunan	E 113°35′, N 27°00′
H007	YC	Shimen, Hunan	E 111°38′, N 29°58′	H082	XL352	Miluo, Hunan	E 113°07′, N 28°81′
H008	YC	Shimen, Hunan	E 111°38′, N 29°58′	H083	XL371	Miluo, Hunan	E 113°07′, N 28°81′
H009	YC	Shimen, Hunan	E 111°38′, N 29°58′	H084	XL374	Miluo, Hunan	E 113°07′, N 28°81′
H010	YC	Shimen, Hunan	E 111°38′, N 29°58′	H085	XL376	Miluo, Hunan	E 113°07′, N 28°81′
H011	YC	Shimen, Hunan	E 111°38′, N 29°58′	H086	XL380	Miluo, Hunan	E 113°07′, N 28°81′
H012	YC	Shimen, Hunan	E 111°38′, N 29°58′	H087	GY11	Changsha, Hunan	E 112°99′, N 28°26′
H013	YC	Shimen, Hunan	E 111°38′, N 29°58′	H088	GY12	Changsha, Hunan	E 112°99′, N 28°26′
H014	YC	Yiyang, Hunan	E 112°37′, N 28°58′	H089	GY13	Changsha, Hunan	E 112°99′, N 28°26′
H015	YC	Yiyang, Hunan	E 112°37′, N 28°58′	H090	GY14	Changsha, Hunan	E 112°99′, N 28°26′
H016	YC	Yiyang, Hunan	E 112°37′, N 28°58′	H091	DZ1	Pingjiang, Hunan	E 113°58′, N 28°70′
H017	YC	Zhangjiajie, Hunan	E 110°55′, N 29°35′	H092	ChaL466	Chaling, Hunan	E 113°54′, N 26°78′
H018	YC	Zhangjiajie, Hunan	E 110°55′, N 29°35′	H093	ChaL74-14	Chaling, Hunan	E 113°54′, N 26°78′
H019	YC	Zhangjiajie, Hunan	E 110°55′, N 29°35′	H094	ChaL75-1	Chaling, Hunan	E 113°54′, N 26°78′
H020	YC	Zhangjiajie, Hunan	E 110°55′, N 29°35′	H095	ChaL75-3	Chaling, Hunan	E 113°54′, N 26°78′
H021	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H096	ChaL77-6	Chaling, Hunan	E 113°54′, N 26°78′
H022	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H097	DP7101	Yongzhou, Hunan	E 111°59′, N 26°46′
H023	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H098	GuiY7810	Yongzhou, Hunan	E 111°59′, N 26°46′
H024	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H099	JT22	Yongzhou, Hunan	E 111°59′, N 26°46′
H025	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H100	XT11	Yongzhou, Hunan	E 111°59′, N 26°46′
H026	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H101	YJT1	Yongzhou, Hunan	E 111°59′, N 26°46′
H027	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H102	Tian117	Miluo, Hunan	E 113°07′, N 28°81′
H028	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H103	TG151	Miluo, Hunan	E 113°07′, N 28°81′
H029	YC	Zhuzhou, Hunan	E 113°17′, N 27°86′	H104	HH1	Huaihua, Hunan	E 110°04′, N 27°58′
H030	XL1	Changsha, Hunan	E 112°99′, N 28°26′	H105	HH2	Huaihua, Hunan	E 110°04′, N 27°58′
H031	XL3	Hengyang, Hunan	E 112°74′, N 27°23′	H106	HH3	Huaihua, Hunan	E 110°04′, N 27°58′
H032	XL4	Youxian, Hunan	E 113°35′, N 27°00′	H107	HH4	Huaihua, Hunan	E 110°04′, N 27°58′
H033	XL5	Youxian, Hunan	E 113°35′, N 27°00′	H108	HH6	Huaihua, Hunan	E 110°04′, N 27°58′
H034	XL9	Changsha, Hunan	E 112°99′, N 28°26′	H109	JY7805	Jiangyong, Hunan	E 111°34′, N 25°27′
H035	XL11	Changsha, Hunan	E 112°99′, N 28°26′	H110	Zao7722	Chenzhou, Hunan	E 113°01′, N 25°78′
H036	XL12	Huaihua, Hunan	E 110°04′, N 27°58′	H111	Zao7730	Chenzhou, Hunan	E 113°01′, N 25°78′
H037	XL13	Yongzhou, Hunan	E 111°59′, N 26°46′	H112	Zao7801	Chenzhou, Hunan	E 113°01′, N 25°78′
H038	XL14	Changsha, Hunan	E 112°99′, N 28°26′	H113	G6	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H039	XL15	Ningyuan, Hunan	E 111°94′, N 25°57′	H114	G68	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H040	XL19	Yongzhou, Hunan	E 111°59′, N 26°46′	H115	G70	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H041	XL26	Miluo, Hunan	E 113°07′, N 28°81′	H116	GS84-8	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H042	XL27	Yongzhou, Hunan	E 111°59′, N 26°46′	H117	GW1	Nanchang, Jiangxi	E 115°94422, N 28°54538
H043	XL32	Changsha, Hunan	E 112°99′, N 28°26′	H118	GW2	Nanchang, Jiangxi	E 115°94422, N 28°54538
H044	XL67	Changsha, Hunan	E 112°99′, N 28°26′	H119	GW5	Nanchang, Jiangxi	E 115°94422, N 28°54538
H045	XL69	Youxian, Hunan	E 113°35′, N 27°00′	H120	GW11	Nanchang, Jiangxi	E 115°94422, N 28°54538
H046	XL75	Miluo, Hunan	E 113°07′, N 28°81′	H121	GW16	Nanchang, Jiangxi	E 115°94422, N 28°54538
H047	XL78	Changsha, Hunan	E 112°99′, N 28°26′	H122	GW83-4	Nanchang, Jiangxi	E 115°94422, N 28°54538
H048	XL81	Changsha, Hunan	E 112°99′, N 28°26′	H123	GX46	Xingguo, Jiangxi	E 115°36314, N 26°33779
H049	XL82	Changsha, Hunan	E 112°99′, N 28°26′	H124	GX48	Xingguo, Jiangxi	E 115°36314, N 26°33779
H050	XL97	Youxian, Hunan	E 113°35′, N 27°00′	H125	GY6	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H051	XL102	Yiyang, Hunan	E 112°37′, N 28°58′	H126	CL3	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H052	XL103	Miluo, Hunan	E 113°07′, N 28°81′	H127	CL4	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H053	XL104	Miluo, Hunan	E 113°07′, N 28°81′	H128	CL21	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H054	XL105	Miluo, Hunan	E 112°37′, N 28°58′	H129	CL23	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′

Table 1. (continued)

No.	Name	Origin	Location	No.	Name	Origin	Location
H055	XL106	Yiyang, Hunan	E 110°04′, N 27°58′	H130	CL27	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H056	XL107	Huaihua, Hunan	E 113°07′, N 28°81′	H131	CL40	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H057	XL108	Yiyang, Hunan	E 112°37′, N 28°58′	H132	CL53	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H058	XL109	Yiyang, Hunan	E 112°37′, N 28°58′	H133	CL55	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H059	XL110	Yiyang, Hunan	E 112°37′, N 28°58′	H134	CL156	Jiangxi	E 113°34′–118°28′, N 24°29′–30°04′
H060	XL115	Miluo, Hunan	E 113°07′, N 28°81′	H135	HB39	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H061	XL117	Miluo, Hunan	E 113°07′, N 28°81′	H136	HB361	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H062	XL123	Yiyang, Hunan	E 112°37′, N 28°58′	H137	HB424	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H063	XL131	Miluo, Hunan	E 113°07′, N 28°81′	H138	EY1	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H064	XL133	Yiyang, Hunan	E 112°37′, N 28°58′	H139	EY2	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H065	XL134	Yiyang, Hunan	E 112°37′, N 28°58′	H140	EY5	Hubei	E 108°47′–114°15′, N 24°38–30°08′
H066	XL141	Hengyang, Hunan	E 112°74′, N 27°23′	H141	CR2	Cenxi, Guangxi	E 110°9949, N 22°91828
H067	XL152	Miluo, Hunan	E 113°07′, N 28°81′	H142	CR3	Cenxi, Guangxi	E 110°9949, N 22°91828
H068	XL156	Hengyang, Hunan	E 112°74′, N 27°23′	H143	CR11	Cenxi, Guangxi	E 110°9949, N 22°91828
H069	XL163	Miluo, Hunan	E 113°07′, N 28°81′	H144	DBS1	Anhui	E 114°54′–119°37′, N 29°41′–34°38′
H070	XL169	Hengyang, Hunan	E 112°74′, N 27°23′	H145	DBS2	Anhui	E 114°54′–119°37′, N 29°41′–34°38′
H071	XL171	Miluo, Hunan	E 113°07′, N 28°81′	H146	DBS3	Anhui	E 114°54′–119°37′, N 29°41′–34°38′
H072	XL183	Changsha, Hunan	E 112°99′, N 28°26′	H147	DBS4	Anhui	E 114°54′–119°37′, N 29°41′–34°38′
H073	XL185	Chaling, Hunan	E 113°54′, N 26°78′	H148	Min43	Fujian	E 115°40–120°30, N23°30–28°20
H074	XL190	Hengdong, Hunan	E 112°95′, N 27°08′	H149	Min48	Fujian	E 115°40–120°30, N23°30–28°20
H075	XL192	Zhuzhou, Hunan	E 113°17′, N 27°86′	H150	Min60	Fujian	E 115°40–120°30, N23°30–28°20

Table 2. Quantitative traits of *C. oleifera*

Quantitative traits	Mean ± sd	Range	CV
Shoot length/cm	15.01 ± 3.59	6.22–26.94	23.90%
Shoot diameter/mm	2.85 ± 0.42	1.45–4.07	14.65%
Leaf area/cm^2	12.87 ± 3.01	6.77–24.23	23.40%
Leaf length/cm	5.39 ± 0.83	3.64–7.64	15.45%
Leaf average width/cm	2.32 ± 0.29	1.46–3.18	12.44%
Leaf maximum width/cm	3.14 ± 0.41	2.00–4.32	12.93%
Fruit height/cm	37.17 ± 4.92	25.63–50.9	13.24%
Fruit diameter/cm	36.53 ± 5.44	23.65–50.56	14.89%
Fruit weight/g	26.83 ± 10.63	7.52–62.10	39.60%
Peel thickness/mm	5.07 ± 1.06	2.72–8.21	20.97%
Seed height/cm	23.84 ± 2.91	17.69–34.41	12.20%
Seed diameter/cm	19.87 ± 2.71	12.26–27.38	13.66%
Fresh seed rate/%	42.31 ± 6.85	24.74–56.56	16.18%
Dry seed rate/%	21.62 ± 4.68	12.08–34.11	21.66%
Dry kernel rate/%	13.06 ± 3.89	5.06–23.89	29.80%
Dry kernel oil content/%	44.54 ± 6.52	23.10–57.68	14.63%
Fresh fruit oil producrion rate/%	5.98 ± 2.40	1.49–12.91	40.11%

Fig. 1. Pictures of the shape variation of leaves, fruits and seeds among accessions.

leave, fruits and seeds.

We conducted a correlation analysis with 17 quantitative traits among 150 *C. oleifera* germplasms, the results showed that the correlation of these quantitative traits divided into three clusters (**Fig. 3**). Cluster I consisted with 5 economic characteristics that refer to the fruits, seeds and oil yield. Cluster II included 6 morphological characteristics described the sizes of fruits and seeds. The rest 6 morphological characteristics that involved in the size of leaves and growth of branches constituted Cluster III.

The principal component analysis (PCA) was carried out with seventeen quantitative traits among 150 *C. oleifera* germplasms. The results showed that the first three principal component (PCs) explained 68.33% of the variations in total (**Table 3**).

PCA also showed that the size of leaves and fruits, together with economic traits had a significant contribution of the total variation. However, the germplasm materials could not be grouped clearly based on principal component in three-dimensional space (**Fig. 4**).

Detection of polymorphisms

By using 9 pairs of SSR primer among 150 *C. oleifera* germplasms, as shown in **Table 4**, a total of 145 alleles (100% polymorphic alleles) were detected, ranging 5 to 28
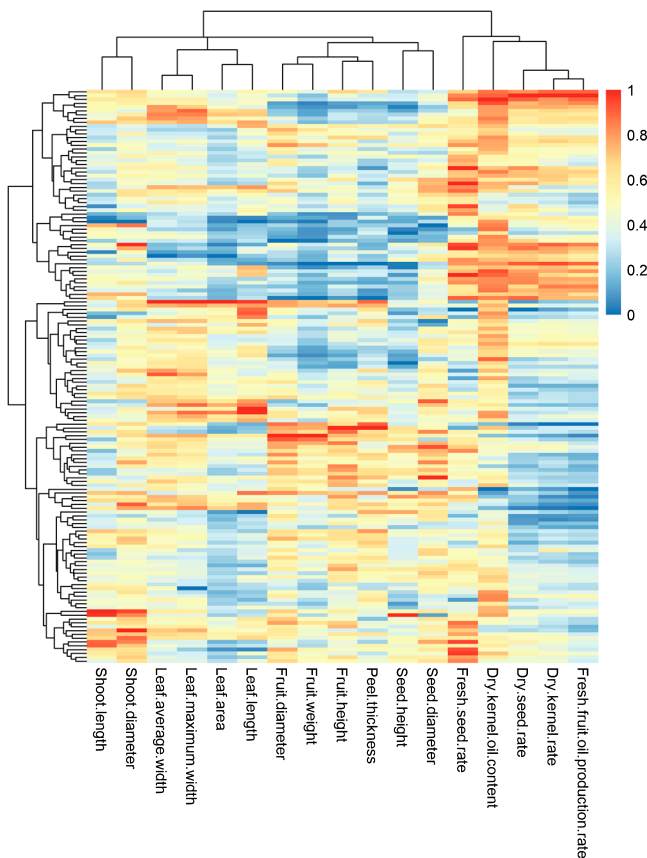
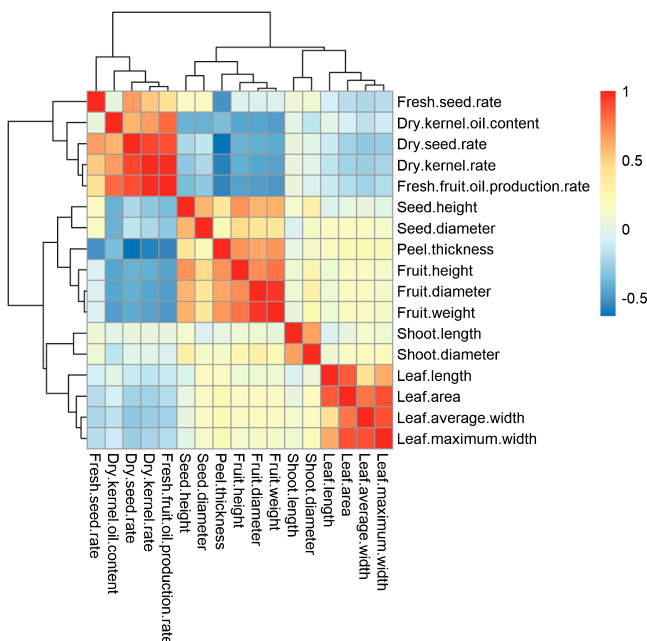**Fig. 2.** The heatmap of normalized quantitative traits of 150 *C. oleifera* germplasms.

Fig. 3. The correlation analysis of quantitative traits of 150 *C. oleifera* germplasms.

Table 3. The principal component analysis in the *C. oleifera* germplasm collection

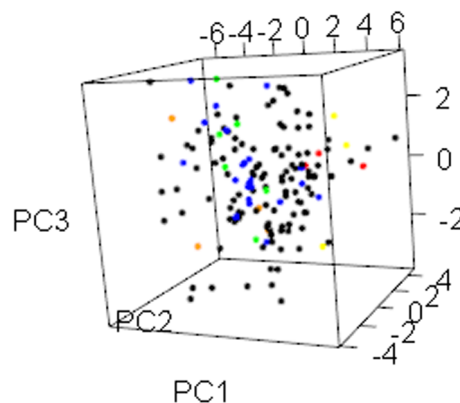| Principal components	PC1	PC2	PC3
Standard deviation	6.25	3.05	2.31
Proportion of variance	36.77%	17.96%	13.60%
Cumulative proportion	36.77%	54.73%	68.33%
Traits		Eigen vectors	
Shoot length	0.05	0.01	0.44
Shoot diameter	0.25	−0.01	0.57
Leaf area	0.44	0.84	0.23
Leaf length	0.30	0.70	0.19
Leaf average width	0.45	0.72	0.18
Leaf maximum width	0.42	0.83	0.20
Fruit height	0.76	−0.32	0.29
Fruit diameter	0.76	−0.31	0.29
Fruit weight	0.79	−0.32	0.26
Peel thickness	0.77	−0.09	−0.13
Seed height	0.57	−0.43	0.43
Seed diameter	0.47	−0.18	0.37
Fresh seed rate	−0.37	−0.25	0.64
Dry seed rate	−0.80	−0.04	0.52
Dry kernel rate	−0.82	0.01	0.46
Dry kernel oil content	−0.69	0.23	0.12
Fresh fruit oil production rate	−0.84	0.08	0.40

Fig. 4. Principal component analysis of 150 *C. oleifera* germplasms in three-dimensional space based on the quantitative traits. Colors of points reflects the origin sources of individuals. Black – Hunan province, blue – Jiangxi province, green – Hubei province, yellow – Guangxi province, orange – Anhui province, red – Fujian province.

NJFUC251. Alleles with low frequencies (<5%) accounts for 66.21% (96/145), and the specific alleles (only observed in one accession) account for 8.97% (13/145). The PIC (polymorphism information content) values varied from 0.588 to 0.937 with an average of 0.749, the highest PIC value was observed at NJFUC243 and the lowest was observed at NJFUC273 and NJFUC601.

Genetic diversity and cluster analysis

Based on the Dice genetic similarity coefficient between two different samples, a clustering map of genetic relationships among *C. oleifera* germplasms was constructed using

per locus with an average of 16.11 (**Table 4**). The highest NA (number of alleles) was observed at NJFUC243 and

Table 4. SSR primers used in this study

Primer	Forward sequence	Reverse sequence	Repeat unit	Product size (bp)	Number of alleles/ polymorphic alleles	Number of genotypes	Polymorphism information content
NJFUC53	TGCCCTAAGTGTCATTC	CAGGGATGATATTGTTTCT	(AAAAT)5	157–217	10/10	41	0.782
NJFUC57	ATAGGTCTTTGTCTGGTT	ATGTAGAGGAAGACTGGA	(TC)10	223–275	25/25	55	0.623
NJFUC243	TGTATGGTTTGGCTCG	GGTTGGCAAGATGAGA	(AGA)10	161–251	28/28	138	0.937
NJFUC251	AATCAACCAAGCGTAC	AGATCCTCCAAACTCC	(TAT)7	93–174	28/28	112	0.913
NJFUC273	ATCTGTAGCTTAATTCTAG	ATTTTCTGGAGCATCT	(AAAAC)7	290–350	12/12	78	0.588
NJFUC600	GTCTTGGCTATCATTTT	TTTCCTATTGACCTCC	(AT)7	222–250	11/11	23	0.588
NJFUC653	ATCGACCTTTGTTGGG	ATTGAAGCTGGCATTT	(TCGGC)3	180–210	5/5	15	0.677
NJFUC787	GTGGCTCAATAAGGAT	CATTACACCGTCTTCAT	(AT)7	140–232	15/15	77	0.887
NJFUC833	GTGGGTTACGGGTTTA	CGGGACAAGTTCAGTT	(AAT)5	213–252	11/11	32	0.746

UPGMA method. The Dice genetic similarity coefficients of 150 germplasm materials ranged from 0.05 to 0.91, which performed the rich genetic diversity.

A very complicated phylogenetic tree based on the SSR cluster analysis results of 150 germplasms was showed in **Fig. 5**. When the genetic similarity coefficient was 0.31, 150 *C. oleifera* germplasms could be divided into five groups. Cluster I consists of only one germplasm, namely 'ChaL75-1' (short for 'Chaling No. 75-1'), from Hunan Province. Its average genetic similarity with other germplasms was 0.24, indicated that there was a significant genetic difference between the identified germplasm material and other materials. Cluster II consists of four germplasms, namely 'XL108', 'ChaL77-6' and two ancient germplasms, all from Hunan province. Cluster III consists of four germplasms which were 'XL3', 'XL210', and 'ChaL75-3' from Hunan province while 'EY1' from Hubei province. Cluster IV consists of four germplasms which were 'XL141' and 'XL197' from Hunan province while 'G6' and 'DBS3' from Jiangxi and Anhui province, respectively. The rest germplasms constituted Cluster V which divided into five subcluster. Cluster Va consists of five germplasms which were 'XL134', 'XL376', 'JY7805' and 'Zao7730' from Hunan province while 'EY2' from Hubei province. Cluster Vb consists of six germplasms which were 'XL26', 'XL109', 'XL192', 'ChaL74-14' and 'Zao7722' from Hunan province while 'CL53' from Jiangxi province. Cluster Vc consists of eight germplasms which were six ancient germplasms from Hunan province while 'CL40' and 'HB361' from Jiangxi and Hubei province. Cluster Vd consists of nine germplasms which were 'XL5', 'XL12', 'XL32', 'XL81', 'DP7101' and two ancient germplasms from Hunan province while 'GS84-8' and 'GW83-4' from Jiangxi province. Cluster Ve consists of the rest 109 germplasms with an average genetic similarity of 0.40.

Discussion

Interactions of genetic diversity and environmental back-
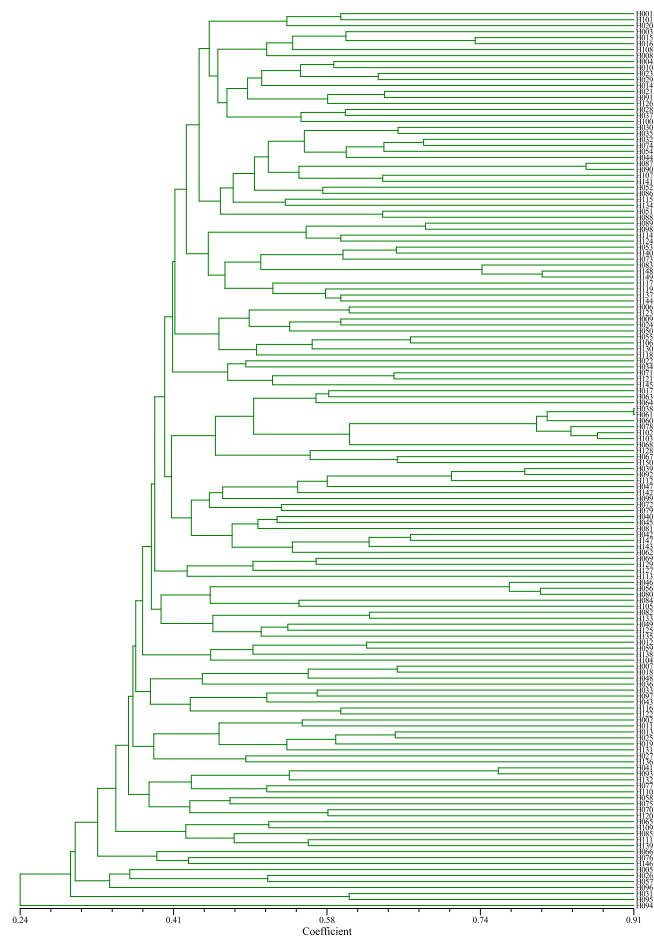

Fig. 5. Phylogenetic tree of 150 *C. oleifera* germplasms based on SSR data.

ground resulted in phenotypic diversity (Backouchi *et al.* 2015). The variations of quantitative traits were mainly under regulation of multiple genes, and they also differentiated by adapting to different environmental conditions (Zhou *et al.* 2015). *Camellia oleifera* is a perennial economic tree species that has a long history of cultivation for more than 2,000 years. Since its characteristics of

cross-pollinating, the phenotypic variations of *C. oleifera* germplasm were highly enriched under the environmental stress and genetic selection (Zhang *et al.* 2019). Although DNA molecular markers had been applied to study the genetic diversity of *C. oleifera* germplasm for a long time, it was still reliable to carry out phenotypic studies on the genetic diversity of the same species of *C. oleifera* with the same habitat and age. In this study, phenotypic variations were analyzed within 29 ancient *C. oleifera* trees and 121 unique *C. oleifera* seedlings that bred by human. The co-efficient of variation (CV) can evaluate the degree of dispersion of quantitative traits to a certain extent. The germplasm materials applied to this study exhibit significant genetic variation for all the seventeen morphological and economic traits, since the CV of each quantitative traits was higher than 12.20%. The heatmap showed that the germplasm materials cluster into two groups. Cluster I consisted with 55 *C. oleifera* germplasms with relatively high economic traits, whose fresh seed rate, dry seed rate, dry kernel rate, dry kernel oil content, fresh fruit oil production rate were 17.26%, 36.00%, 49.23%, 13.93% and 69.11% higher than those of Cluster II, respectively. In contrast, the leaf area, fruit height, fruit diameter, fruit weight and peel thickness of Cluster II were 12.80%, 13.17%, 14.43%, 44.52% and 31.12% higher than those of Cluster I, respectively. Together with the result of correlation analysis of 17 quantitative traits, negative correlations between economic traits and morphological traits were existed, which provided a new strategy for *C. oleifera* breeding of high oil content.

Previously, various DNA markers, including simple sequence repeat (SSR), inter-simple sequence repeat (ISSR), random amplified polymorphic DNA (RAPD) and sequence-related amplified polymorphism (SRAP) have been used in the studies of the genetic diversity of *C. oleifera* (Chen *et al.* 2019, Jiang *et al.* 2018, Zhai 2018, Zhang *et al.* 2018). In this study, 29 ancient *C. oleifera* trees living for over three hundred years collected from different areas in Hunan province and 121 human bred varieties from the main cultivation areas were used for the analysis of their genetic relationship and origins. As shown in **Fig. 2** and **Fig. 5**, it was showed that all the *C. oleifera* materials are divided into different groups in molecular level, respectively. These estimates of genetic diversity could be biased by the choice of phenotypic or molecular data (Vinu *et al.* 2013). Thus, both phenotypic traits and molecular markers should be considered to achieve unbiased genetic diversity estimation. The genetic similarity coefficients ranged from 0.05 to 0.91, which indicated that these SSR primers used in this study could distinguish all the germplasm materials completely. This result might due partly to the long history of cultivation and characteristics of cross-pollination. With the help of cross-pollination characteristics, the genetic resources of *C. oleifera* constantly exchanged with each other during the long-term cultivation, resulted in a high level of genetic diversity of germplasm materials. *C. oleifera* germplasms from six provinces were analyzed, 112 of them were from Hunan province while 22, 6, 3, 4 and 3 were from Jiangxi, Hubei, Guangxi, Anhui and Fujian province, respectively. The uneven distribution of materials might affects the analysis of genetic diversity, but it was showed that some of the germplasm materials with the same geographic origin or similar genetic background can be clustered in the same group, while the other germplasm materials with geographical origin and inconsistent genetic background also were clustered in the same group.

Generally, *C. oleifera* is a hexaploid specie with 90 chromosome (Ye *et al.* 2020), studies using SSR molecular markers to evaluate the genetic diversity of *C. oleifera* could hardly cover whole genome. Furthermore, *C. oleifera* has no reference genome data, SSR molecular markers used in the studies of *C. oleifera* had been made were selected from the prediction of transcriptomic data. In this experiment, 9 pairs of SSR primers were applied under the consideration of the purpose of genetic relationship evaluation of germplasm resources and the practical feasibility of completing the analysis process. In the result, the genetic similarity coefficients between *C. oleifera* germplasm materials ranged from 0.05 to 0.91, indicated that these SSR primers used in this study could distinguish all the germplasm materials completely. Among these 150 accessions, there were 29 ancient *C. oleifera* trees which lives for at least three hundred years those were the special precious germplasm resources used in this study, they were collected from different district in Hunan Province by the authorship due to their remaining abilities of normal flowering and fruiting. They were chosen in this study helping explore the genetic origin of artificial bred cultivar. For the rest 121 accessions, there were superior cultivars recommended by the National Forestry and Grassland Administration of China, and candidate breeding germplasm resources with great potential in improving production of *C. oleifera*. Firstly, Since these SSR molecular markers can distinguish all germplasm materials, these SSR molecular markers can be used to construct fingerprints of superior cultivars. Secondly, the genetic relationship between candidate breeding resources provide the study on affinity optimization of hybridization combination for fundamentals. Thirdly, the assessment of genetic relationship between the ancient *C. oleifera* trees and candidate breeding resources provide evidences for the studies on the origin of germplasm materials. There are still a numerous of candidate *C. oleifera* breeding resources which are left need to been analyzed in our future project to establish the phenotype and genetic characteristics database of *C. oleifera* core breeding population. During the process, more SSR and maybe some other molecular markers would be applied to meet its need. Altogether, the results of this study indicated that phenotypic traits and SSR markers can be efficiently integrated in the exploration of their genetic diversity for improvement and germplasm utilization.

Literature Cited

Backouchi, I.Z., M. Aouida, N. Khemiri and M. Jebara (2015) Genetic diversity in Tunisian populations of faba bean (*Vicia faba* L.) based on morphological traits and molecular markers. Genet. Mol. Res. 14: 7587–7596.

Chen, X., Y. Yun, Y. Wu, H. Qi, L. Yang, J. Chen and D. Zheng (2019) Genetic diversity analysis of *Camellia oleifera* resources based on SRAP markers in Hainan Island. Journal of Tropical and Subtropical Botany 27: 659–668 (in Chinese).

Chen, Y., B. Wang, J. Chen, X. Wang, R. Wang, S. Peng, L. Chen, L. Ma and J. Luo (2015) Identification of Rubisco rbcL and rbcS in *Camellia oleifera* and their potential as molecular markers for selection of high tea oil cultivars. Front. Plant Sci. 6: 189.

Chen, Y., X. Dai, J. Hou, H. Guan, Y. Wang, Y. Li and T. Yin (2016) DNA fingerprinting of oil camellia cultivars with SSR markers. Tree Genet. Genomes 12: 7.

Costa, J.C., G.G.M. Fracetto, F.J.C. Fracetto, M.V.F. Santos and M.A. Lira Júnior (2017) Genetic diversity of *Desmanthus* sp accessions using ISSR markers and morphological traits. Genet. Mol. Res. 16.

Huang, G., T. Guan, Y. Zhao, M. Chen and H. Liu (2018) A rapid and efficient method for extracting DNA from *Camellia oleifera* leaves. Mol. Plant Breed. 16: 4350–4354 (in Chinese).

Huang, Y. (2013) Population genetic structure and interspecific introgressive hybridization between *Camellia meiocarpa* and *C. oleifera*. Chinese Journal of Applied Ecology 24: 2345–2352 (in Chinese).

Jeffreys, A.J., V. Wilson and S.L. Thein (1985) Individual-specific 'fingerprints' of human DNA. Nature 316: 76–79.

Jiang, D., Y. Fang, X. Xiao, H. Cheng, X. Zhang, J. Cheng and L. Li (2018) Identifying *Camellia oleifera* germplasm accessions with inter simple sequence repeat markers. Hubei Agricultural Science 57: 119–125 (in Chinese).

Liu, B., L. Jin, C. Cao, Q. Wang, M. Zhou, S. Wu, Q. Shu and L. Zhang (2011) Establishment and optimization of SSR-PCR reaction system for oil tea (*Camellia oleifera* Abel.). Journal of Anhui Agricultural University 38: 858–862 (in Chinese).

Liu, C., L. Chen, W. Tang, S. Peng, M. Li, N. Deng and Y. Chen (2018) Predicting potential distribution and evaluating suitable soil condition of oil tea *Camellia* in China. Forests 9: 487.

Rohlf, F. (2000) Ntsys-pc numerical taxonomy and multivariate analysis system version 2.1. owner manual.

Verma, H., J.L. Borah and R.N. Sarma (2019) Variability assessment for root and drought tolerance traits and genetic diversity analysis of rice germplasm using SSR markers. Sci. Rep. 9: 16513.

Vinu, V., N. Singh, S. Vasudev, D.K. Yadava, S. Kumar, S. Naresh, S.R. Bhat and K.V. Prabhu (2013) Assessment of genetic diversity in *Brassica juncea* (*Brassicaceae*) genotypes using phenotypic differences and SSR markers. Rev. Biol. Trop. 61: 1919–1934.

Ye, T., Y. Li, J. Zhang, Q. Gong, D. Yuan and S. Xiao (2020) Optimization of chromosome mounting technique and karyotype analysis of *Camellia oleifera*. Journal of Nanjing Forestry University (Natural Sciences Edition). DOI: 10.3969/j.issn.1000-2006. 202003022 (in Chinese).

Zhai, Q. (2018) Genetic variation of *Camellia oleifera* in Anhui province based on AFLPs. Dissertation, Anhui Agricultural University (in Chinese).

Zhang, G., W. Zhong, Y. Wu, X. Tan and T. Du (2007) Identification of oil tea (*Camellia oleifera*) superior clones by ISSR molecular marker. Forest Research 20: 278–282 (in Chinese).

Zhang, M., H. Zhang, X. Cai, H. Wang, Y. Ye and Y. Wu (2018) SSR analysis of genetic relationship between *Camellia* germplasm resources. Nonwood Forest Research 36: 130–134 (in Chinese).

Zhang, T., S. Liu, H. Mei and Y. Dong (2011) SRAP analysis of genetic diversity of *Camellia oleifera* resources in Hubei province. Journal of Huazhong Normal University (Nat. Sci.) 45: 477–479 (in Chinese).

Zhang, W., Y. Zhao, G. Yang, J. Peng, S. Chen and Z. Xu (2019) Determination of the evolutionary pressure on *Camellia oleifera* on Hainan Island using the complete chloroplast genome sequence. PeerJ 7: e7210.

Zhang, Y. (2003) Analysis of genetic diversity and genetic characters by RAPD markers on oil *Camellia*. Dissertation, Fujian Normal University (in Chinese).

Zhou, R., Z. Wu, X. Cao and F.L. Jiang (2015) Genetic diversity of cultivated and wild tomatoes revealed by morphological traits and SSR markers. Genet. Mol. Res. 14: 13868–13879.

Zhou, W., W. Qiang, J. Yang, J. Wang, L. Xv and L. Xv (2017) Establishment of DNA fingerprints and cluster analysis for *Oil Camellia* cultivars based on SSR markers. Mol. Plant Breed. 15: 238–249 (in Chinese).

Zhu, G., H. Liu, Y. Xie, Q. Liao, Y. Lin, Y. Liu, Y. Liu, H. Xiao, Z. Gao and S. Hu (2020) Postharvest processing and storage methods for *Camellia oleifera* seeds. Food Rev. Int. 36: 319–339.