BIOCELL
2020 44(3): 381-388

Tech Science Press

Construction of castor functional markers fingerprint and analysis of genetic diversity

TING YANG[#]; JIANNONG LU[#]; AKWASI YEBOAH; SHUAILEI GU; DONGNA LI; YUZHEN SHI; XUEGUI YIN*

Department of Crop Breeding and Genetics, College of Agricultural Sciences, Guangdong Ocean University, Zhanjiang, 524088, China

Key words: Castor, FMs (Functional Markers), Fingerprint, Genetic diversity

Abstract: In order to provide a molecular basis for selecting good hybrid combinations for the identification of castor bean germplasm resources, fingerprint and genetic diversity analysis of 52 castor bean materials from 12 regions in 5 countries were constructed by using the Functinal Markers (FMs) associated with fatty acid metabolism-related genes. A total of 72 alleles were amplified by 29 pairs of FMs with an average of 2.483 per marker and the polymorphic information content was 0.103–0.695. Shannon's information index (I), observed heterozygosity (Ho) and expected heterozygosity (He) were 0.699, 0.188 and 0.436 respectively. The clustering results indicated that the castor germplasm could be divided into two groups with the genetic similarity coefficient of 0.59. The genetic similarity of 12 regions ranged from 0.518 to 0.917 and the genetic distance was between 0.087 and 0.658. A total of 5 pairs of core primers were screened to construct a digital fingerprint of different castor germplasm resources, which could distinguish all 52 germplasms. This study provides a scientific basis for screening high-quality castor germplasm resources and broadening the genetic basis of castor breeding at the molecular level.

Introduction

Castor (*Ricinus communis* L., 2n = 2x = 20), a dicotyledonous annual or perennial shrub belonging to the family Euphorbiaceae, is one of the world's top ten oil crops with high social value (Costa *et al.*, 2006). It originated in East Africa (Vavilov, 1951) and has been now widely cultivated in tropical and subtropical regions (Govaerts *et al.*, 2000) with India, Brazil and China as the main producing countries of castors (Downey *et al.*, 1989). The oil content of castor seeds ranks among the best in seed oil crops, reaching as high as 46% to 55%. Castor oil features unique physiochemical properties and has been widely used for machinery, aerospace, pharmaceutical, paint, soap, cosmetics, lubricant, textile, printing, dyeing, energy and environmental protection, composite materials and phytoremediation purposes (Brigham, 1993; Ogunniyi, 2006).

Castor is deemed as an economically important oilseed crop with 3–5% increase in demand annually (Anjani, 2012). However, the lower genetic diversity of castor and relatively lagging genetic research make castor lack high-yield and high-quality varieties, and farmers lack enthusiasm for planting, resulting in a significant reduction in castor planting area. The growing demand for castor products and the decreasing supply of castor have caused a serious supply-demand disequilibrium in the market, e.g., the castor oil and its derivatives in developed countries such as the United States is mainly acquired through import (Roetheli *et al.*, 1991). After 2014, the import dependence rate of castor raw materials from China, as the main producing country, has also reached more than 90% (Information from 2014 to 2018 Annual Meeting of the Chinese Academy of Agricultural Engineering Castor Technology and Economics Branch); therefore, it is of great significance to broaden the genetic background of castor germplasm, improve the utilization rate of germplasm resources, enhance the selection efficiency and improve the scientific nature of castor breeding for cultivating high-quality, stable and resistant castor varieties.

The construction of fingerprints and study of genetic diversity enable it possible to effectively utilize castor germplasm resources and cultivate high-yield and high-quality varieties. And molecular marker technology plays an important role in the study of biological genetic diversity (Kumar *et al.*, 2009) and kinship. At present, the molecular marker methods commonly adopted on castor include the amplified fragment length polymorphism marker (AFLP) (García-Zambrano et al., 2018), the related sequence

*Address correspondence to: Xuegui Yin, yinxuegui@126.com
[#]These authors contributed equally to this work
Received: 12 November 2019; Accept: 13 April 2020

Doi: 10.32604/biocell.2020.09112 www.techscience.com/journal/biocell

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

amplification polymorphism (SRAP) (Allan *et al.*, 2008), the random primer polymorphism marker (RAPD) (Tantasawat *et al.*, 2018), the simple sequence repeat marker (ISSR) (Kallamadi *et al.*, 2015; Wang *et al.*, 2013), the start codon targeting polymorphic primer (SCOT) (Kallamadi *et al.*, 2015), the microsatellite (also known as simple sequence repeat, SSR) (Allan *et al.*, 2008; Bajay *et al.*, 2009; Seo *et al.*, 2011; Senthilvel *et al.*, 2017), EST-SSR (Qiu *et al.*, 2010; Thatikunta *et al.*, 2016; Wang *et al.*, 2016) and SNP markers (Foster *et al.*, 2010).

Functional Markers (FMs) are molecular markers developed on the basis of polymorphisms in gene sequences. Different allelic variations of these genes are directly related to phenotypes (Andersen and Lübberstedt, 2003). With the enrichment of a large number of gene EST sequences in public databases, more FMs have been developed to a greater extent, which provide guidance for accurate evaluation and efficient use of genetic information (Gupta and Rustgi, 2004). It has become a powerful means for germplasm resource evaluation, hybrid genetic purity detection and genetic diversity identification (Fjellstrom *et al.*, 2004; Kumar *et al.*, 2014; Liu *et al.*, 2012; Simões *et al.*, 2017). Several studies have reported on the application of FMs in different plant species. Fjellstrom *et al.* (2004) developed three rice blast-resistant functional genes for screeenig new rice varieties. Also a large number of FMs related to agronomic traits of processing quality and disease resistance have been developed in wheat, which plays an important role in the molecular selective breeding of wheat (Liu *et al.*, 2012). In addition, Kumar *et al.* (2014) used TRAP (a type of functional marker) to study the genetic diversity of 263 native species of chickpea preserved in the USA-ARS Western Regional Plant Introduction Station. At present, few reports are available on castor fatty acid metabolism-related functional markers. Only Simões *et al.* (2017) published an article on TRAP marker development of casto, but no report has been found regarding the systematic marker diversity analysis and fingerprint construction of castor oil related genes. The castor fatty acid metabolism-related genes are single-copy (Chan *et al.*, 2010) and it is more effective to use such single-copy genes as FMs for development and related fingerprint mapping and genetic diversity studies.

On the basis of the published castor genome sequence (Chan *et al.*, 2010), this study randomly selected several genes in relation to fatty acid metabolism and screened 29 pairs of FMs for 52 castor germplasm resources (from 12 countries or regions) for genetic diversity analysis and fingerprinting. Upon identification of the genetic background differences of castor germplasm resources, it provides an effective way to protect castor varieties and a theoretical basis for breeding castor varieties and broadening the genetic basis of castor breeding.

Materials and Methods

The 52 materials used in the test (Tab. 1) originated from 5 countries and 12 regions, containing wild materials and varieties. The wild materials are representative varieties selected from the molecular breeding laboratory system of the Agricultural College of Guangdong Ocean University in South China. The remaining materials are representative varieties selected from various countries and regions. All these materials were planted in the experimental field of the Agricultural College of Guangdong Ocean University randomly in 2008 in the order of the field ranks, 5 repeats per material, the plant spacing of 0.8 m, the line spacing of 1 m, three repetitions, randomly arranged.

Experimental methods

DNA was extracted from young leaves by using the modified CTAB method (Couch and Fritz, 1990). Referring to the EST sequence of castor published in 2010 (Chan *et al.*, 2010), the primer of castor fatty acid metabolism-related genes was designed and synthesized by Shanghai Sheng Gong Co., Ltd. From the developed primers, 29 pairs of primers with good amplification effect, clear and stable bands were used to PCR-amplify the genomic DNA of 52 castor materials. PCR amplification was performed in a 20 μl reaction system: 1.5 μl template DNA (20 ng/μl), 0.4 μl Taq enzyme (3U/μl), 2 μl 10 × PCR Buffer, 0.2 μl dNTP (10 mmol/L), 2 μl primer (2 μmol/L) and 13.9 μl double distilled water under optimized experimental conditions according to the following reaction procedures, speciaifcally, initial-denaturation at 94°C for 5 min, denaturation at 94°C for 30 s, annealing at 55°C for 50 s, extension at 72°C for 1 min, 35 cycles and extension at 72°C for 5 min. The amplified products were separated and detected by 6% non-denaturing polyacrylamide gel electrophoresis.

Data analysis

Clear amplicons were scored by a read indicating “1” with no read “0,” and “0, 1” matrix is thus established. The number of observed alleles (Na), number of effective alleles (Ne), observed heterozygosity (Ho), Shannon’s information index (I) and the inheritance between 12 inter groups were calculated by using Pop Gen version 1.32 (Tehrani *et al.*, 1998) software. PIC values and genotypes were calculated by using Power marker Version 3.25 (Liu and Muse, 2005) software. The genetic similarity coefficients were calculated by the method of Nei and Li (Nei and Li, 1979). The similarity matrix was clustered by UPGMA method (Sneath and Sokal, 1973), and the NT sys software was used to construct the clustering diagram (Rohlf, 1993).

Results

SSR marker polymorphism analysis

A total of 52 pairs of FM primers were used for polymorphism detection of 52 castor materials. The amplification results (Tab. 2) showed that 29 pairs of primers amplified 72 alleles in 52 germplasm resources, and each pair of primers was amplified. The number of alleles ranged from 1 to 4 with an average of 2.48 per locus. The average number of effective alleles (Ne) was 1.951 with OM19 having the lowest effective allele at 1.000 and the highest effective alleles in OM3 at 3.113. The Shannon’s Information Index (I) ranged from 0.000 to 1.240 with an average of 0.687. The observed heterozygosity (Ho) of the primer had the maximum and minimum value of 0.941 and 0 respectively. The average polymorphism information content was 0.397, ranging from 0.103 to 0.695.

TABLE 1

The castor materials

Code	Material name	Origin	Type
1	GX-1	Guangxi, China	Wild material
2	GX-2	Guangxi, China	Wild material
3	GX-3	Guangxi, China	Wild material
4	GX-4	Guangxi, China	Wild material
5	GX-5	Guangxi, China	Wild material
6	GX-6	Guangxi, China	Wild material
7	GD-1	Guangdong, China	Wild material
8	GD-2	Guangdong, China	Wild material
9	GD-3	Guangdong, China	Wild material
10	GD-4	Guangdong, China	Wild material
11	GD-5	Guangdong, China	Wild material
12	GD-6	Guangdong, China	Wild material
13	GD-7	Guangdong, China	Wild material
14	GD-8	Guangdong, China	Wild material
15	GD-9	Guangdong, China	Wild material
16	GD-10	Guangdong, China	Wild material
17	GD-11	Guangdong, China	Wild material
18	HN-1	Hainan, China	Wild material
19	HN-2	Hainan, China	Wild material
20	HN-3	Hainan, China	Wild material
21	HN-4	Hainan, China	Wild material
22	YN-1	Yunnan, China	Variety
23	YN-2	Yunnan, China	Variety
24	YN-3	Yunnan, China	Variety
25	YN-4	Yunnan, China	Variety
26	YN-5	Yunnan, China	Variety
27	SD-1	Shandong, China	Variety
28	SD-2	Shandong, China	Variety
29	SD-3	Shandong, China	Variety
30	SD-4	Shandong, China	Variety
31	SD-5	Shandong, China	Variety
32	SD-6	Shandong, China	Variety
33	SD-7	Shandong, China	Variety
34	SD-8	Shandong, China	Variety
35	SD-9	Shandong, China	Variety
36	SD-10	Shandong, China	Variety
37	SX-1	Shanxi, China	Variety
38	SX-2	Shanxi, China	Variety
39	TW-1	Taiwan, China	Variety
40	NM-1	Inner Mongolia, China	Variety
41	NM-2	Inner Mongolia, China	Variety
42	NM-3	Inner Mongolia, China	Variety
43	TG-1	Thailand	Variety
44	TG-2	Thailand	Variety
45	FG-1	France	Variety
46	FG-2	France	Variety
47	FG-3	France	Variety
48	FG-4	France	Variety
49	MLXY-1	Malaysia	Variety
50	MLXY-2	Malaysia	Variety
51	MLXY-3	Malaysia	Variety
52	BJST-1	Pakistan	Variety

(Continued)

Table 1 (continued).

Construction of DNA fingerprinting

According to the amplification results of 29 pairs of FM primers, the genotypes were analyzed after the primer diversity (Tab. 2). Considering the size of the primer PIC and the statistical difficulty of the allele number, the primers with the highest number of polymorphisms were selected and used to distinguish the 52 castor materials. The number of genotypes distinguished by all primers ranged from 2 to 7, as shown in Tab. 2; therefore, all germplasm resources were distinguished by primer combinations including OM3, ACC3, PEPC10, PEPC6 and ACC12 respectively. A total of 52 genotypes were involved in all tested materials. Based on the primer combination sequence of OM3, ACC3, PEPC10, PEPC6 and ACC12, the result of corresponding amplified fragment size of each variety was converted into a binary, which was connected into a series of numbers. Thus a unique code of each variety was obtained, namely the fingerprint (Tab. 3). The 5 pairs of highly polymorphic primers obtained in this study can provide a convenient and fast recording method for DNA fingerprinting at the molecular level for identification of castor materials.

Construction of DNA fingerprinting

According to the amplification results of 29 pairs of FM primers, the number of genotypes was analyzed upon analysis of the diversity of primers (Tab. 2). Considering the size of PIC of primers, the number of alleles and the difficulty of banding statistics, the primers with the most genotypes were selected. For these 29 tobacco germplasms, if not all identified as one primer at a time, all materials were separated. Tab. 2 shows that the number of genotypes that all primers can distinguish is 2–7, so all germplasm resources are distinguished by the primer combinations including OM3, ACC3, PEPC10, PEPC6 and ACC12 respectively in this paper. A total of 52 genotypes were involved in all tested materials. In this study, according to the primer combination order of OM3, ACC3, PEPC10, PEPC6 and ACC12, the result of the size of each amplified fragment corresponding to each variety was then converted into a binary, which was connected into a series of numbers. Thus a unique code of each variety were obtained as the "Fingerprint," shown in Tab. 3.

Cluster analysis

The cluster analysis was performed by using the UPCMA method (unweighted averaging method) to map the phylogenetic tree (Fig. 1). The results showed that the genetic similarity coefficients of 52 materials ranged from

TABLE 2

Genetic diversity of 52 castors based on 29 FM markers

Primer	PIC	Genotypes	Na	Ne	I	Ho	He
ACC3	0.695	6	4	3.103	1.240	0.935	0.678
BCCP1	0.226	2	2	1.352	0.429	0.000	0.260
FAD2-3	0.575	4	3	2.863	1.073	0.019	0.651
BCP2	0.526	5	3	2.408	0.949	0.000	0.585
BCCP4	0.395	3	2	1.940	0.677	0.000	0.484
BCP4	0.451	5	2	2.000	0.693	0.000	0.500
BCCP3	0.423	4	3	1.664	0.723	0.000	0.399
PEPC2	0.379	4	2	1.926	0.674	0.020	0.481
ACC4	0.583	4	3	2.818	1.067	0.000	0.645
OM3	0.621	7	4	3.113	1.240	0.808	0.679
OM20	0.483	4	3	1.922	0.815	0.000	0.480
ACC11	0.326	2	2	1.696	0.601	0.000	0.411
OM19	0.206	2	1	1.000	0.000	0.000	0.000
PEPC9	0.382	3	3	1.985	0.727	0.019	0.496
PEPC10	0.439	6	3	2.185	0.850	0.404	0.542
FAD2-2	0.206	3	2	1.304	0.395	0.038	0.233
PEPC5	0.510	4	2	1.953	0.681	0.844	0.488
ACC2	0.245	2	2	1.401	0.461	0.000	0.286
BCP1	0.316	2	2	1.649	0.582	0.000	0.393
PEPC1	0.484	5	3	2.093	0.890	0.471	0.522
ACC13	0.103	2	2	1.122	0.221	0.000	0.109
OM17	0.275	4	2	1.368	0.440	0.000	0.269
FAD2-16	0.226	2	2	1.352	0.429	0.000	0.260
BCP14	0.206	2	2	1.304	0.395	0.000	0.233
PEPC7	0.415	4	3	2.030	0.739	0.941	0.507
BCCP15	0.347	3	2	1.807	0.639	0.019	0.447
ACC12	0.620	5	3	2.946	1.090	0.060	0.661
PEPC8	0.262	2	2	1.451	0.490	0.000	0.311
PEPC6	0.592	5	3	2.836	1.069	0.863	0.647
Mean	0.397	4	2.483	1.951	0.699	0.188	0.436

0.59 to 0.94 and were divided into two large groups at a genetic similarity coefficient of 0.59 and 0.67 respectively. Not all materials from the same region could be clustered into the same group, therefore, the genetic grouping of materials was not entirely affected by the region. Nonetheless, an obvious trend was available that the groups were closely related to the regions. The materials from Taiwan and Malaysia were grouped separately, namely Group A and D, and the materials from Hainan were also distributed in small subgroups of Group G and J. Group C, F, G and H include all materials from the neighboring Guangdong Province and Guangxi Province.

Population distance, genetic identity and UPCMA cluster analysis of different populations

The results of genetic distance and genetic identity distribution of germplasm from 12 different regions in 5 countries including China, Malaysia and Pakistan were compared. The results revealed that the genetic distance of the 12 regions was between 0.087 and 0.658 (Tab. 4), and the materials from Taiwan, Malaysia and Hainan recorded the highest genetic distance between 0.594 and 0.658. The genetic distance of materials from Shandong and France (0.087), materials from Guangxi and Guangdong (0.113) was small; the genetic consistency between 12 castor populations ranged between 0.518 and 0.917. The materials from Taiwan were inferior to those from Malaysia and Hainan (0.518, 0.552), and the relationship between materials from France and Shandong (0.917), Guangxi and Guangdong (0.893) was far apart. The higher genetic coherence was, the higher frequency of genetic communication between them would be. The UPCMA clustering results between the 12 populations (Fig. 2) were consistent with the results of the cluster analysis (Fig. 1).

TABLE 3

DNA fingerprints code of subset of 52 castors with 5 primers

Material name	Numbered fingerprint	Material name	Numbered fingerprint
GX-1	00100000010101100	**SD-1**	10100101010011001
GX-2	10101100110101010	**SD-2**	10100101100011010
GX-3	11001100110101100	**SD-3**	11000101010100100
GX-4	10011010110011010	**SD-4**	11000110010011001
GX-5	10101100110010010	**SD-5**	10100101010101100
GX-6	10010101100011001	**SD-6**	11000101010101100
GD-1	10100101100101100	**SD-7**	11000101100100000
GD-2	10101100100011110	**SD-8**	10100101010101001
GD-3	10010110100011001	**SD-9**	11000101010101010
GD-4	10010101110011110	**SD-10**	10100101100101010
GD-5	10100010110011001	**SX-1**	10010101010011010
GD-6	00011100100100110	**SX-2**	10100110011011010
GD-7	10010101110101001	**TW-1**	00100101110000001
GD-8	10100110110011010	**NM-1**	10101100010101100
GD-9	10100000110100100	**NM-2**	10100010010101100
GD-10	10100110100011010	**NM-3**	10100110011011100
GD-11	10100010100011010	**TG-1**	10100000100101010
HN-1	10101100100011100	**TG-2**	10100101110101010
HN-2	10100000100011000	**FG-1**	00100110100101100
HN-3	10100000001011100	**FG-2**	10100101100011001
HN-4	10100000100011010	**FG-3**	00010110101101001
YN-1	10010101110011010	**FG-4**	10100101100101100
YN-2	10100101110100001	**MLXY-1**	01000110110100010
YN-3	00100101110101001	**MLXY-2**	01000110110101010
YN-4	10010101010101001	**MLXY-3**	01000110110101100
YN-5	10010101010101010	**BJST-1**	10000101110011001

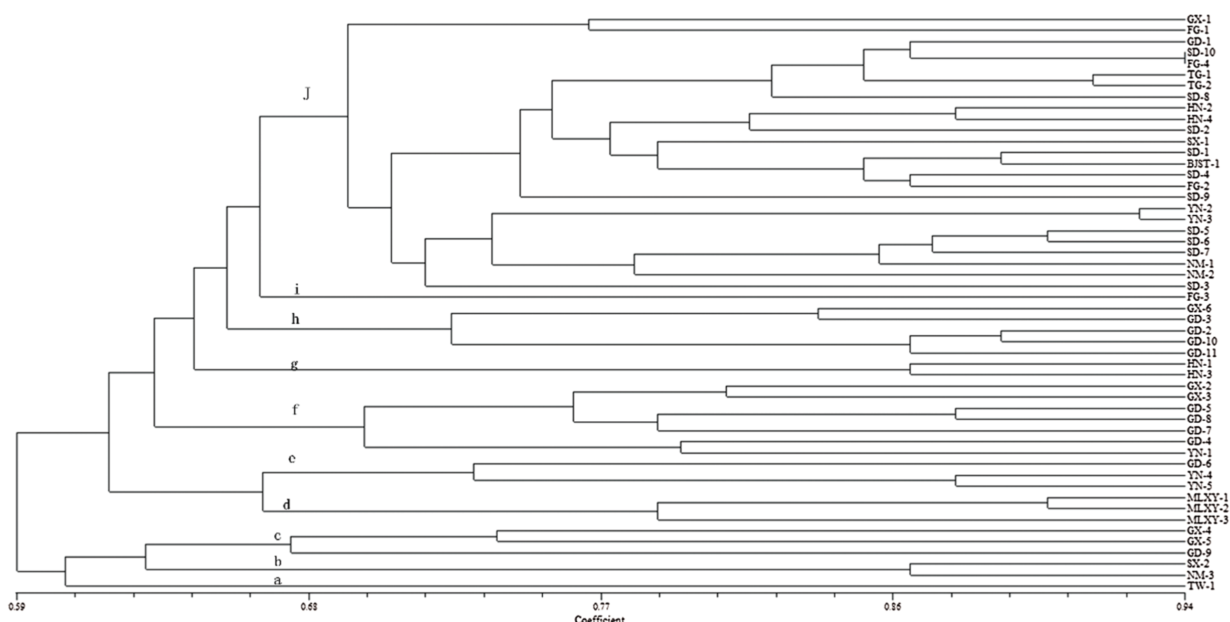

FIGURE 1. UPCMA clustering analysis for 52 castor resources with 29 FM markers.

TABLE 4

Genetic consistency (above diagonal) and genetic distance (below diagonal) among 12 castor populations

	GX	GD	HN	YN	SD	SX	TW	NM	TG	FG	MLXY	BJST
GX		0.893	0.813	0.825	0.859	0.802	0.623	0.788	0.789	0.798	0.739	0.766
GD	0.113		0.807	0.890	0.868	0.774	0.674	0.761	0.834	0.886	0.731	0.770
HN	0.207	0.214		0.780	0.835	0.748	0.552	0.753	0.726	0.789	0.641	0.756
YN	0.193	0.116	0.249		0.886	0.818	0.687	0.782	0.811	0.811	0.741	0.840
SD	0.152	0.141	0.180	0.121		0.833	0.653	0.797	0.821	0.917	0.728	0.874
SX	0.221	0.256	0.291	0.201	0.182		0.578	0.827	0.806	0.688	0.658	0.767
TW	0.473	0.394	0.594	0.376	0.427	0.548		0.640	0.612	0.714	0.518	0.712
NM	0.239	0.273	0.284	0.246	0.227	0.190	0.447		0.711	0.702	0.693	0.676
TG	0.237	0.182	0.321	0.210	0.197	0.215	0.490	0.341		0.780	0.635	0.688
FG	0.225	0.121	0.236	0.210	0.087	0.374	0.337	0.354	0.248		0.701	0.819
MLXY	0.303	0.314	0.445	0.300	0.318	0.419	0.658	0.366	0.454	0.356		0.598
BJST	0.266	0.261	0.279	0.174	0.134	0.266	0.340	0.391	0.374	0.200	0.514	

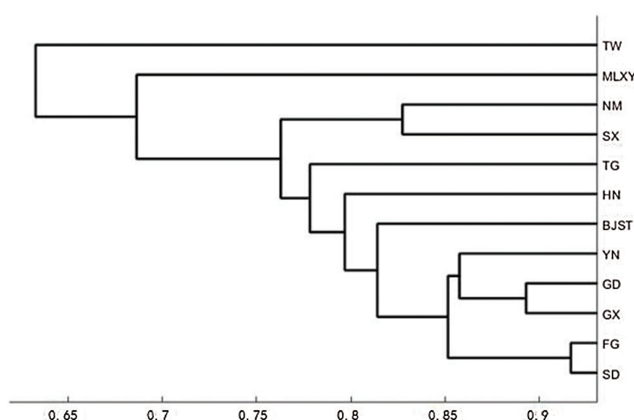

FIGURE 2. UPGMA clustering analysis for 12 castor populations.

Discussion

DNA-fingerprinting, as a powerful tool for testing the authenticity and purity of varieties, highlights the advantages of fastness and accuracy. It has been used for resource diversity and purity identification in many species including corn (Wang *et al.*, 2011) and watermelon (Zhang *et al.*, 2012). The development of molecular marker technology has enriched the identification method of varieties. FM markers are molecular markers developed based on the polymorphism of gene sequences. Different allelic variations of these genes are directly related to the phenotype (Andersen and Lübberstedt, 2003). The fatty acid metabolism-related genes in castor is a single copy (Chan *et al.*, 2010), therefore the FM marker associated with the castor oil content is one of the ideal markers for identification of castor materials.

The 52 experimental materials involved in this study have a wide range of sources, covering 12 regions in 5 countries. 21 of these materials are wild materials from South China. Studies have pointed out that Chinese castor materials are likely to originate in South China, and the genetic diversity of wild materials is higher than that of cultivated materials (Fan *et al.*, 2019). The genetic diversity of wild materials in South China is slightly higher than that from other places (Wang *et al.* 2019). Therefore, the castor materials used in this study feature rich genetic backgrounds and can represent the majority castor materials. Since the five pairs of primer combinations as selected can completely separate the 52 germplasms of this experiment, it is highly possible to distinguish the majority castor materials that have appeared so far. However, with the continuous enrichment of germplasm resources, the identification ability of these five pairs of primers may decrease, so the number of primers may increase or the primers may be replaced according to specific conditions.

Cluster analysis of 52 different types of castor germplasm resources was carried out by using FM molecular markers. When the genetic similarity coefficient was 0.59, 52 materials were divided into 2 groups, and then into several groups when the genetic similarity coefficient was 0.67. Although not all the materials from the same region were clustered in the same group, an obvious trend was found that the groups were related to the regions to a great extent, which was consistent with the results of some genetic diversity analyses of castor as reported. Allan *et al.* (2008) used 16 pairs of AFLP markers and 9 pairs of SSR markers for genetic diversity analysis of 200 castor materials from 41 regions in 35 countries of 5 continents. Senthilvel (Senthilvel *et al.*, 2017) used 45 SSR markers for genetic analysis of 144 castor inbred lines. Also, Kallamadi (Kallamadi *et al.*, 2015) used RAPD, ISSR and SCOT markers to analyze 35 castor materials from 7 regions of the world. Recently, Agyenim-Boateng (Agyenim-Boateng *et al.*, 2019) used SRAP markers to analyze the genetic diversity of 473 castor-bean materials from South China.

The genetic distance and genetic consistency of materials from 12 regions showed that Taiwan is far away from Hainan and Malaysia. This may be attributed to the differences in geographical location, which significantly reduced their inter-regional genetic communication and separated them,

as identified by the cluster analysis (Figs. 1 and 2). French materials are closely related to Shandong materials perhaps because the breeders used materials from one region with the parent materials from the other. For the materials from Guangdong, Guangxi and Hainan, it can be seen that the relationship between Guangdong and Hainan materials is the farthest, followed by Guangxi and Hainan, Guangdong and Guangxi, which is consistent with the findings of Wang *et al.* (2019).

Conclusion

As shown in this study, with the application of the fatty acid metabolism-related functional marker technology, five pairs of primers, OM3, ACC3, PEPC10, PEPC6 and ACC12, can be used to distinguish 12 parts of castor materials from 12 regions in 5 countries. As these 52 parts of castor material are from a wide range of sources, they can represent the majority castor materials; therefore, it can be predicted that the five primers selected in this study can distinguish most of the castor germplasm resources in the world. This study provides a technical basis for identifying the castor materials and protecting the germplasm resources.

Acknowledgement We wish to appreciate everyone who supported us in preparing this manuscript.

Funding Statement: This study was funded by National Natural Science Foundation of China, (31271759); Guangdong Provincial Science and Technology Projects (2013B060400024, 2014A020208116 and 2016A020208015) (China); Project of Enhancing School with Innovation of Guangdong Ocean University (GDOU2013050206) (China).

Conflicts of Interest: The authors declare there are no conflicts of interest.

References

Agyenim-Boateng KG, Lu J, Shi Y, Zhang D, Yin X (2019). SRAP analysis of the genetic diversity of wild castor (*Ricinus communis* L.) in South China. *PloS One* **14**: e0219667. DOI 10.1371/journal.pone.0219667.

Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2008). Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. *Genetic Resources and Crop Evolution* **55**: 365–378. DOI 10.1007/s10722-007-9244-3.

Andersen JR, Lübberstedt T (2003). Functional markers in plants. *Trends in Plant Science* **8**: 554–560. DOI 10.1016/j.tplants.2003.09.010.

Anjani K (2012). Castor genetic resources: a primary gene pool for exploitation. *Industrial Crops and Products* **35**: 1–14. DOI 10.1016/j.indcrop.2011.06.011.

Bajay MM, Pinheiro JB, Batista CEA, Nobrega MBM, Zucchi MI (2009). Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. *Conservation Genetics Resources* **1**: 237–239. DOI 10.1007/s12686-009-9058-z.

Brigham RD (1993). *Castor: return of an old crop*. New Crops. New York: Wiley, 380-383.

Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G (2010). Draft genome sequence of the oilseed species Ricinus communis. *Nature Biotechnology* **28**: 951–956. DOI 10.1038/nbt.1674.

Costa MD, Pereira WE, Bruno Rd LA, Freire EC, Nóbrega MDM, Milani M, Oliveira Ad (2006). Divergência genética entre acessos e cultivares de mamoneira por meio de estatística multivariada. *Embrapa Algodão-Artigo em Periódico Indexado (ALICE)* **41**: 1617–1622.

Couch JA, Fritz PJ (1990). Isolation of DNA from plants high in polyphenolics. *Plant Molecular Biology Reporter* **8**: 8–12. DOI 10.1007/BF02668875.

Downey RK, Robbelen G, Ashri A (1989). Oil crops of the world: their breeding and utilization. *Quarterly Review of Biology* **34**: 349–360.

Fan W, Lu J, Pan C, Tan M, Lin Q, Liu W, Li D, Wang L, Hu L, Wang L (2019). Sequencing of Chinese castor lines reveals genetic signatures of selection and yield-associated loci. *Nature Communications* **10**: 1–11. DOI 10.1038/s41467-018-07882-8.

Fjellstrom R, Conaway-Bormans CA, McClung AM, Marchetti MA, Shank AR, Park WD (2004). Development of DNA markers suitable for marker assisted selection of three Pi genes conferring resistance to multiple Pyricularia grisea pathotypes. *Crop Science* **44**: 1790–1798. DOI 10.2135/cropsci2004.1790.

Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010). Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis* L). *BMC Plant Biology* **10**: 13. DOI 10.1186/1471-2229-10-13.

García-Zambrano EA, Ojeda-Zacarías MC, Gutiérrez-Soto JG, Sinagawa-García SR, Gutiérrez-Díez A, Martínez-de la Cerda J, Pedroza-Flores JA, Gallegos-Vázquez C, García-Zambrano JI (2018). Molecular differentiation of Xoconostle (Opuntia spp.) using AFLP molecular markers. *Phyton, International Journal of Experimental Botany* **87**: 87–93.

Govaerts R, Frodin DG, Radcliffe-Smith A, Carter S (2000). *World checklist and bibliography of Euphorbiaceae (with Pandaceae)*. Royal Botanic Gardens, Kew (eds), Royal Botanic Gardens, New York.

Gupta P, Rustgi S (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Functional & Integrative Genomics* **4**: 139–162.

Kallamadi PR, Nadigatla VGR, Mulpuri S (2015). Molecular diversity in castor (*Ricinus communis* L.). *Industrial Crops and Products* **66**: 271–281. DOI 10.1016/j.indcrop.2014.12.061.

Kumar P, Gupta V, Misra A, Modi D, Pandey B (2009). Potential of molecular markers in plant biotechnology. *Plant Omics* **2**: 141.

Kumar Y, Kwon SJ, Coyne CJ, Hu J, Grusak MA, Kisha TJ, McGee RJ, Sarker A (2014). Target region amplification polymorphism (TRAP) for assessing genetic diversity and marker-trait associations in chickpea (Cicerarietinum L.) germplasm. *Genetic Resources and Crop Evolution* **61**: 965–977. DOI 10.1007/s10722-014-0089-2.

Liu K, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**: 2128–2129. DOI 10.1093/bioinformatics/bti282.

Liu Y, He Z, Appels R, Xia X (2012). Functional markers in wheat: current status and future prospects. *Theoretical and Applied Genetics* **125**: 1–10. DOI 10.1007/s00122-012-1829-3.

Nei M, Li WH (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76**: 5269–5273. DOI 10.1073/pnas.76.10.5269.

Ogunniyi DS (2006). Castor oil: a vital industrial raw material. *Bioresource Technology* **97**: 1086–1091. DOI 10.1016/j.biortech.2005.03.028.

Qiu L, Yang C, Tian B, Yang JB, Liu A (2010). Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biology* **10**: 278. DOI 10.1186/1471-2229-10-278.

Roetheli JC, Glaser LK, Brigham RD (1991). *Castor: assessing the feasibility of US production: summary of workshop held at Plainview*. Texas: Growing Industrial Material Series.

Rohlf F (1993). *Numeric taxonomy and multivariate analysis system*. NTSYS-pc. ScienceOpen, Inc New York.

Senthilvel S, Shaik M, Anjani K, Shaw RK, Kumari P, Sarada C, Kiran BU (2017). Genetic variability and population structure in a collection of inbred lines derived from a core germplasm of castor. *Journal of Plant Biochemistry and Biotechnology* **26**: 27–34. DOI 10.1007/s13562-016-0356-8.

Seo KI, Lee GA, Ma KH, Hyun DY, Park YJ, Jung JW, Lee SY, Gwag JG, Kim CK, Lee MC (2011). Isolation and characterization of 28 polymorphic SSR loci from castor bean (*Ricinus communis* L.). *Journal of Crop Science and Biotechnology* **14**: 97–103. DOI 10.1007/s12892-010-0107-7.

Simões K, Silva S, Machado E, Brasileiro H (2017). Development of TRAP primers for *Ricinus communis* L. *Genetics and Molecular Research* **16**: 100–113.

Sneath PH, Sokal RR (1973). Numerical taxonomy. *Principles and Practice of Numerical Classification* **12**: 190–199.

Tantasawat PA, Lunwongsa K, Linthaisong T, Wirikitgul P, Campatong N, Talpolkrung N, Tharapreuksapong A, Poolsawat O, Khairum A, Sorntip A, Kativat C (2018). Genetic diversity of water primrose (Ludwigia hyssopifolia) in Thailand based on morphological characters and RAPD analysis. *Phyton, International Journal of Experimental Botany* **87**: 162–170.

Tehrani MS, Mardi M, Saeidi H, Gharehyazi B, Assadi M, Balfourier F, Charmet G, Ravel C, Balfourier F, Imbert C (1998). The user-friendly shareware for population genetic analysis. *International Journal of Botany* **4**: 100–110.

Thatikunta R, Sankar AS, Sreelakshmi J, Palle G, Leela C, Rani CVD, Shankar VG, Lavanya B, Reddy PN, Dudhe M (2016). Utilization of in silico EST-SSR markers for diversity studies in castor (Ricinus communis L.). *Physiology and Molecular Biology of Plants* **22**: 535–545. DOI 10.1007/s12298-016-0367-x.

Vavilov NI (1951). The origin, variation, immunity and breeding of cultivated plants. *LWW* **72**: 482.

Wang C, Li GR, Zhang ZY, Peng M, Shang YS, Luo R, Chen YS (2013). Genetic diversity of castor bean (*Ricinus communis* L.) in Northeast China revealed by ISSR markers. *Biochemical Systematics and Ecology* **51**: 301–307. DOI 10.1016/j.bse.2013.09.017.

Wang FG, Tian HL, Zhao JR, Yi HM, Wang L, Song W (2011). Development and characterization of a core set of SSR markers for fingerprinting analysis of Chinese maize varieties. *Maydica* **56**: 1693.

Wang M, Dzievit M, Chen Z, Morris JB, Norris J, Barkley NA, Tonnis B, Pederson GA, Yu J (2016). Genetic diversity and population structure of castor (*Ricinus communis* L.) germplasm within the US collection assessed with EST-SSR markers. *Genome* **60**: 193–200. DOI 10.1139/gen-2016-0116.

Wang YF, Lu JN, Shi YZ, Chen S, Li WX, He XL, Yin XG, Yuan C (2019). Genetic diversity analysis of south China wild castor oil plant based on SSR markers. *Molecular Plant Breeding* **17**: 910–916.

Zhang H, Wang H, Guo S, Ren Y, Gong G, Weng Y, Xu Y (2012). Identification and validation of a core set of microsatellite markers for genetic diversity analysis in watermelon, Citrullus lanatus Thunb. Matsum. & Nakai. *Euphytica* **186**: 329–342. DOI 10.1007/s10681-011-0574-z.