Plant Omics Journal POJ

POJ 6(3):193-200 (2013) ISSN:1836-3644

Development of microsatellite markers in *Cocos nucifera* and their application in evaluating the level of genetic diversity of *Cocos nucifera*

Yong Xiao[1#], Yi Luo[1#], Yaodong Yang[1*], Haikuo Fan[1,3], Wei Xia[1], Annaliese S. Mason[5], Songlin Zhao[1], Ross Sager[4], Fei Qiao[2]

[1]Hainan Key Laboratory of Tropical Oil Crops Biology/Coconuts Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, Hainan 571339, P.R. China
[2]Key Laboratory of Crop Gene Resources and Germplasm Enhancement in Southern China, Ministry of Agriculture/Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, 571101, China
[3]College of Landscape Architecture, Sichuan Agriculture University, Yaan, Sichuan, 625000, P.R. China
[4]Delaware Biotechnology Institute, University of Delaware, USA, Newark, Delaware, 19711, USA
[5]School of Agriculture and Food Sciences and Centre for Integrative Legume Research, the University of Queensland, Brisbane, Australia

***Corresponding author: xiaoyong1980714@gmail.com**
#The authors have equal contribution to the manuscript

Abstract

Cocos nucifera (coconut) is an economically important tropical crop, but opportunities for molecular breeding are limited by lack of DNA sequence information for this species. Previous assessments of coconut germplasm have been conducted based solely on phenotypic data for agronomic and quality traits, due to lack of available molecular markers. In this study, we developed 30 novel microsatellite markers from Illumina transcriptome sequence data, and used these markers to evaluate coconut genetic diversity in 30 individuals representing accessions from China (12 samples) and Southeast Asia (18 samples). The microsatellite markers displayed low to high genetic polymorphism across the population: observed heterozygosity varied from 0.06 to 0.79, with an average of 0.39 ± 0.15. Our results indicated that the Southeast Asian population had a significantly higher number of alleles ($p = 0.02$), but not significantly different ($p<0.05$) observed heterozygosity ($p = 0.08$) or Shannon's diversity index score ($p = 0.12$) compared to the population from China. Population structure analysis showed that Chinese coconuts were genetically similar to a subset of coconut germplasm from Southeast Asia, suggesting that Chinese coconuts did not evolve independently of the Southeast Asian populations. Combining population structural analyses and historic information, we have devised a possible explanation for coconut dispersal patterns from Southeast Asia to China: sea currents could have carried coconuts into the Hainan province, while human dispersal from Southeast Asia may have brought coconuts to the Yunnan province.

Keywords: *Cocos nucifera*, SSR, Genetic diversity, Population Stucture, Illumina sequencing.
Abbreviations: SSR-Simple Sequence Repeats; Ta-optimal annealing temperature; CTAB-Cetyltrimethylammonium bromide; H_o-Observed heterozygosity; *H'*-Shannon diversity index.

Introduction

Coconut palm (*Cocos nucifera* L.), a member of the monocotyledonous family Aracaceae (Palmaceae), is a tropical oil crop and is widely cultivated in tropical regions due to its extensive applications in agriculture and industry. Currently, *C. nucifera* is distributed across 89 tropical countries, including Central and South America, East and West Africa, Southeast Asia and the Pacific islands, and accounts for over 12 million hectares of land (Batugal et al., 2005). Evidence has shown that these coconuts can be divided into two major subpopulations due to independent origins of domestication: the Pacific Ocean and the Indo-Atlantic Ocean populations (Teulat et al., 2000; Perera et al., 2003; Gunn et al., 2011). In the Pacific Ocean, an area extending from the Malay Peninsula to New Guinea seems to be the most likely evolutionary origin of the coconut (Harries, 1990; Larson et al., 2005). Coconuts can be classified into two main categories: "tall" and "dwarf", based on morphological characteristics and breeding habits. Tall coconuts are slow to mature (flowering 8-10 years after planting), can grow to a height of about 20-30 m, and have medium to large nut sizes (Menon and Pandalai, 1958). Tall coconuts are hardy and can adapt to a wide range of environment conditions. Dwarf coconuts are early-flowering (4-6 years after planting), can grow to a height of about 10-15 m, and can produce a number of small nuts (Harries, 1978). Tall coconuts are generally outcrossing and therefore mostly considered heterozygous, while dwarf coconuts are autogamous. Based on morphological characteristics of coconut fruit, tall coconuts can be classified into two main types: "Niekafa" which have oblong or triangular fruit, and which may have resulted from natural selection via ocean dispersal, and "Nievai", which have rounded fruit, and may have evolved as a result of human cultivation and dissemination (Whitehead, 1966; Harries, 1981; Baudouin and Lebrun, 2009). In China, coconuts are widely grown in the Hainan and Yunnan provinces (located in southern and southwestern China, respectively), and have over 2000 years of cultivation history. Compared to coconuts in the Malay Peninsula, coconuts in southern and southwestern China have adapted to different environmental

conditions, including higher latitudes and lower temperatures in winter. Currently, most studies of coconut germplasm in China focus on investigating morphological and agronomic characteristics. However, these assessments may not be very accurate, as many characteristics exhibit complex inheritance and are influenced by external environmental factors as well as genetics. The advent of molecular marker techniques has allowed researchers to begin to evaluate coconut germplasmat the DNA level, using RAPD (Randomly Amplified Polymorphic DNA), RFLP (Restriction Fragment Length Polymorphism), and AFLP (Amplified Fragment Length Polymorphism) markers (Ashburner et al., 1997; Rajesh et al., 2013; Lebrun et al., 1998; Perera et al., 1998). Despite the economic importance of coconut in tropical regions, coconut cultivation is in relative decline in many areas due to low productivity and devastating diseases. To improve coconut agronomic traits, it is necessary to evaluate the genetic diversity of available coconut germplasm. Microsatellites are thought to be a good tool for molecular breeding and diversity assessment (Morgante and Olivieri, 1993; Powell et al., 1996; Tabkhkar et al., 2012; Reusch et al., 2010). However, very few short tandemly-repeated markers have been developed in coconuts. In this study, we describe the development of thirty novel coconut microsatellite primers based on Illumina sequence data. Using these markers, we evaluated the level of genetic diversity in coconut germplasm from China and Southeast Asia, and analyzed genetic relationships between these accessions.

Results

Development and characterization of thirty microsatellite markers in Cocos nucifera

Using MSATCOMMANDER software, 57,304 unigenes were scanned for di-, tri-, tetra- and hexa-nucleotide repeat sequences, with a minimum set of 6 repeats. A total of 4796 microsatellite loci were found. Of these, 76 microsatellite loci were stochastically selected for subsequent polymorphism analysis: 27 dinucleotide repeats, 42 trinucleotide repeats and 7 tetranucleotide motif repeats. Primer pairs were designed according to the flanking sequences of the 76 loci, and used to amplify DNA obtained from 10 *Cocos nucifera* individuals. Based on the PCR amplification results, 46 microsatellite loci were eliminated due to poor amplification or monomorphism. The remaining 30 microsatellite markers had easily scorable, polymorphic bands in the ten coconut individuals; detailed information for the thirty primer pairs is listed in Table 1. Of the 30 SSR loci, 12 (40%) had a dinucleotide-repeat motif, 17 had a trinucleotide-repeat motif (57%) and one had a tetranucleotide motif(3%). The proportions of polymorphic di-, tri- and tetra-nucleotide repeats out of the 76 loci screened were 44.4%, 40.5% and 14.3%, respectively. The size of the amplification products resulting from these primer pairs ranged from 110 to 620 bp. To further characterize polymorphism in these SSR markers, an additional 20 *Cocos nucifera* individuals to make 30 total were tested (Fig 1). A total of 91 alleles were detected from the 30 microsatellite loci, with an average of 3.03 alleles per locus. Across the 91 microsatellite alleles, 43 alleles were from di-nucleotide motifs (average 3.58 alleles per locus), and 45 alleles were from tri-nucleotide motifs (average 2.65 alleles per locus). Observed heterozygosity ranged from 0.06 to 0.79 across markers, with an average of 0.39 ± 0.15, indicating a moderate level of polymorphism. Observed heterozygosity in the 12 dinucleotide-motif markers averaged 0.50, and observed heterozygosity in the 17 trinucleotide-motif markers averaged 0.38. We subsequently tested if the thirty polymorphic markers from *Cocos nucifera* could be applied to other palm species. The primer pairs corresponding to the polymorphic microsatellite markers were used to amplify DNA obtained from 8 accessions of *Elaeis guineensis* and 8 accessions of *Areca catechu*. In *Elaeis guineensis*, 56.7% (17) of these SSR could be amplified and showed polymorphism; the number of alleles per locus ranged from 2 to 4, and observed heterozygosity ranged from 0.22 to 0.73 (Table 2). However, in *Areca catechu*, only 33.3% (10)of these SSRs could be amplified and 8 showed polymorphism; the number of alleles per locus ranged from 1 to 2 and observed heterozygosity ranged from 0 to 0.66.

Genetic diversity of Cocos nucifera

To analysis the genetic diversity of *Cocos nucifera* based on the 30 developed SSR markers in this study, thirty individuals representing accessions from two main coconut growing regions(China (12) and Southeast Asia (18)) were collected. Coconut plants from China were collected from the Hainan and Yunnan provinces (16-23°N), located in southern and southwestern China, respectively, and representing the most productive coconut-growing areas in China. Coconuts from Southeast Asia were sampled from Vietnam (8-23°N), Myanmar (9-28°N), Thailand (7-20°N), Malaysia (1-7°N), and Indonesia (12°S-7°N). Using the program Popgene 3.2, we calculated the number of alleles (*A*), observed heterozygosity (H_o), and Shannon's diversity index (*H'*) for the two populations (Table 3). Our results indicated that the Southeast Asian population had a significantly higher number of alleles ($p = 0.02$), but not significantly different observed heterozygosity ($p = 0.08$) or Shannon's diversity index score ($p = 0.12$) compared to the population from China (Table 3).

Genetic relationships

Southeast Asia is one of the most productive regions for coconut cultivation worldwide, with optimal temperature and soil conditions. Coconuts are also widely grown in southern and southwestern China, for example in the Hainan and Yunnan provinces. However, due to comparatively low temperatures in these districts of China, Chinese coconuts may have greater tolerance of cold stress than those from different countries in Southeast Asia. To gain a better understanding of how Southeast Asian and Chinese coconuts may have evolved genetic differences over time due to their separate environments, we estimated the genetic divergence between the two coconut populations from each region. In total, 101 alleles were detected for the 30 SSR loci. Of these alleles, 82 were shared in common between the two populations. Four alleles at two SSR loci (WCYZ1721 and WCYZ8939) were specific to the Chinese population, and 15 alleles at 11 SSR loci (WCYZ29240, WCYZ28231, WCYZ25080, WCYZ28277, WCYZ28374, WCYZ21753, WCYZ28635, WCYZ23544, WCYZ27691, WCYZ28904 and WCYZ25609) were specific to the population from Southeast Asia. In order to characterize the extent of the genetic differences between the Chinese and Southeast Asian coconut populations, the average genetic distance was examined between and within populations. Similar genetic distance (0.319) was detected between the two populations from China and Southeast Asia, within the Chinese population (0.298) and within the Southeast Asian population (0.296). Genetic differences between the two populations from China and Southeast Asia were also investigated using Principal Component Analysis (PCA) (Fig 2): the total variation explained by the first and second principal components were 70.23% and 4.75% respectively, supporting

Table 1. Characteristics of the 30 microsatellite loci of *Cocos nucifera*, showing for each primer pair the forward (F) and reverse (R) primer sequences, repeat motif, allele size range, optimal annealing temperature (Ta), GenBank accession number, and observed heterozygosity.

Primer codes	Primer sequences (5'-3')	Repeat	Size (bp)	Ta[1]	GenBank Accession NO.[2]	Observed heterozygosity
WCYZ2470	F: GGGAATAACAAAGCAGCAAAAC R: CCCCTACTAATGTCCACCAGAA	(AG)6	250-260	53	JX275987	0.55
WCYZ9240	F: TATAAGTCGCGTGCCAAGAAA R: GAGAGAGTCGTCAGGAAGAGGA	(AG)8	250-410	54.7	JX275988	0.68
WCYZ8097	F: GAGGAGAGAGAGAAAGGAAGGG R: CATGGTTGATGAGAAACACCTC	(AGA)8	110-280	51.5	JX275989	0.56
WCYZ3075	F: GGAATGGGACTTGGAATTGATA R: TGAACTTACAACAATGAGCGGA	(AT)6	190-210	51.5	JX275990	0.48
WCYZ8231	F: GTGCCAGGTCATGTGAGGATA R: TCGGTATTGTGCTTGGTTGATA	(AT)6	120-270	53	JX275991	0.56
WCYZ2928	F: TCCTCACATTCCTCTTCGTCTT R: ATTCCTCCAGCGATGTACTTGA	(AT)7	480-510	54.7	JX275992	0.79
WCYZ2116	F: GAAGAAAGATGAGAATGGCTCG R: GTGGGGAGATTACGAGTACGAC	(ATC)6	260-270	56.1	JX275993	0.44
WCYZ5080	F: CTGAGGAGACCGCCTACAAG R: GGTGATGATTGCGATGGTAGT	(CAC)6	376	53.2	JX275994	0.12
WCYZ4825	F: TTCTTTTGCTATCGAGTACCGC R: AGCTAGGTGAGGAAGTTGGTGA	(CAC)6	161-560	56.1	JX275995	0.47
WCYZ8483	F: AGGCAGACAGGTTTATCCAGAA R: CATTGTCACCTCCTCCTCCTC	(CAG)6	210-220	53.2	JX275996	0.57
WCYZ8277	F: GCTCCAGTATCTCTGTCCCCT R: TGATGAATCGGTTGACTATCCA	(CAT)7	310	54.7	JX275997	0.17
WCYZ6681	F: AAAGAAAGAAAGACGCATGGAC R: GGCTCGTATTAGGGTTAGGGTT	(CCA)6	350-400	54.5	JX275998	0.56
WCYZ8374	F: ACGTCAACCAGAGTTTATTGGC R: ATGGTAATGGTAATGGCAAAGG	(CAC)7	250-620	56.1	JX275999	0.52
WCYZ1753	F: AATAACTAGCGGCAGAAGAAGC R: CTTGCAGAAAAGGGATTTGAAG	(CCG)7	420	48.6	JX276000	0.06
WCYZ9119	F: GAGATGAGGTAGAGGAGGAGCA R: GAGTAATAGAGGCGATGGCAAG	(CCG)7	80	57.6	JX276001	0.27
WCYZ8635	F: AACCAAACTCCTCTCCAACCTC R: GATGAAGAAGCGGGTGAAGTC	(CCT)6	210	54.7	JX276002	0.21
WCYZ8558	F: ACTTCAGTTACCTCTATCCGGC R: AATAAAGCCCTCTCACTCCCTT	(CCT)6	216	54.7	JX276003	0.35
WCYZ1448	F: TGTACGAAGAGAAAGGAGGAGG R: ACGGAGTAGCTTGAACACCTTG	(CGG)6	180	56.1	JX276004	0.45
WCYZ2158	F: CATCGGCCTCTTCAAACAAT R: TGTCAGGTGCATAAAGGTAAGG	(CT)6	318-390	56.1	JX276005	0.17
WCYZ1521	F: GTCTGTTTTCTGCTCTTCGCTT R: TCCTTTTCCCATCTCTCTGAAC	(CT)7	150	54	JX276006	0.30
WCYZ8624	F: TACGTTTCCAATATCGTCGTCA R: GGAAGGAATTTAGGTTTGTGGG	(CT)8	116-140	56.1	JX276007	0.59
WCYZ8939	F: GTTAAAAGAGGAGGTCTCGGGT R: TCCCAGGAGGAACTAAAGACAC	(CT)9	144-210	55	JX276008	0.51
WCYZ8947	F: AGCTTAGATTTCATGGCTACGG R: AGATTGGATTTTGATGGGAGG	(CTC)7	298-350	55	JX276009	0.51
WCYZ3554	F: CCCGGCTCTCTCTCTAAACTCT R: CTCCTCCTGCTCCTTGTTGAG	(GA)8	264-500	54.7	JX276010	0.43
WCYZ8774	F: ACGAAGAAAATCTCAAACAGGC R: GGACTAGGCAATGGATTCAGAG	(GCA)7	373-470	54.7	JX276011	0.51
WCYZ7691	F: TTTCACTCCCGTCTTCATCTTT R: TTGCTGCTGTTCTTCTTGTTGT	(GGA)6	280-305	54.7	JX276012	0.26
WCYZ8904	F: GCAATCGACAGCTTTTAGCC R: TCTCTCCTCGTTCTTCTTCGTC	(GGTC)6	113-120	54.7	JX276013	0.43
WCYZ8794	ATATGCTCCATTCCTCCATCTC ACCCTTCCCCTTATCGTACATT	(TA)8	202-210	54.7	JX276014	0.42
WCYZ5609	TATATCGGGCATCAAATGTCG CATGAAAGACCTGCAACGAATA	(TC)8	260-275	54.7	JX276015	0.49
WCYZ1721	ACCGTACTTACATCCTCACCCA CACTCTTCCCTTTGCTCTTCAC	(TCA)6	190-630	54.7	JX276016	0.45

[1] Optimal annealing temperature. [2] Accession number of SSR-containing sequences used for primer design.

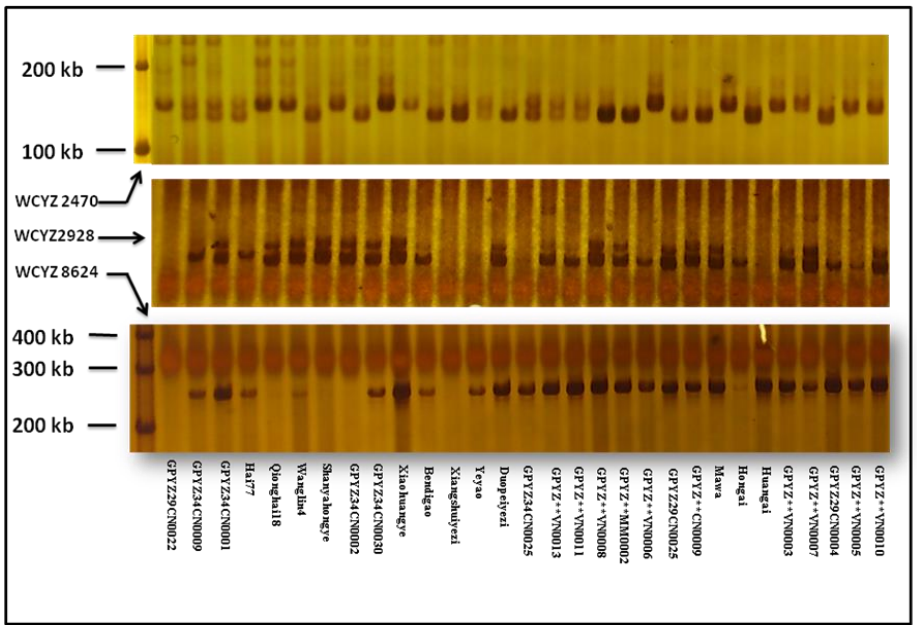

Fig 1. PCR products and polymorphic characteristics of three gene-based microsatellite markers across 30 coconut accessions.

genetic distinctiveness between the two populations. However, there were some exceptions: two coconut accessions from Vietnam were clustered together with the China population, and one accession from Vietnam showed significant genetic divergence from the other 29 coconut accessions examined. This particular coconut accession, called "Makapuno" lacks a liquid endosperm, a trait obtained through continuous artificial selection from natural coconut mutants.

Origin of the Chinese coconut

When the method of Evanno et al. (2005) was applied to identify the most likely number of 'true populations', we found K = 3 genetic groups (Fig 3). The STRUCTURE assignment procedure showed that the largest genetic group comprised 22 coconut individuals from Yunnan and Hainan in China, Vietnam, Malaysia, Thailand, and Myanmar. The second-largest genetic group included seven coconut individuals: two from Malaysia, four from Vietnam, and one from Indonesia. Finally, the smallest group only contained one individual—the Makapuno coconut. Compared to the Principal Component Analysis (PCA), STRUCTURE analysis did not show a clear separation between coconut germplasm from Southeast Asia and China: Chinese coconuts were genetically similar to some coconuts from Southeast Asia (33.3% of Vietnam coconuts, all Myanmar and Thailand coconuts, and 50% of Malaysia coconuts). This analysis indicated that coconuts from China were not an independent evolutionary population compared to coconuts from Southeast Asia. In addition, an UPGMA dendrogram was designed according to Nei-Li's method, and generated seven different clusters (Fig 4). Three genetic clusters (cluster I, cluster V and cluster VII), each of them comprising two to eight coconut accessions, were found to exclusively contain coconut accessions from Southeast Asia. Cluster VI exclusively contained two Chinese coconut accessions. However, other clusters (cluster II, cluster III and cluster IV) were represented by coconut individuals from both China and Southeast Asia. The dendrogram analysis had good agreement with the STRUCTURE analysis, again validating the genetic similarity between Chinese coconuts and some coconuts from Southeast Asia.

Discussion

Coconut is an important tropical economic crop, but only a few studies have been carried out in relation to coconut molecular biology and marker-assistance breeding. Due to a lack of available molecular markers, previous assessment and characterization of coconut germplasms had relied mostly on morphological and agronomic traits (Akpan 1994; Sugimura et al., 1997). However, these traits do not generally provide an accurate measure of genetic diversity, because many show complex inheritance and are easily influenced by the external environment. To solve this problem, Rivera et al. (1999) described the development and characterization of 38 microsatellite markers for *Cocos nucifera* using genomic-SSR hybrid screening, but following their study, almost no related research was reported. In our research, we used computational methods to utilize sequence information from next-generation sequencing (Fan et al., 2013) to develop SSR markers. Of 76 SSR primer pairs designed to screen identified SSRs for polymorphism, 30 (39.5%) exhibited polymorphic bands in a panel of ten coconut individuals. PCR amplification of these 30 SSR primer pairs in an additional 20 *Cocos nucifera* individuals, 8 *Elaeis guineensis* individuals and 8 *Areca catechu* individuals showed that these SSR markers were stable and conserved. Our study has helped validate the idea that using next-generation sequencing technology to develop SSR markers are as feasible for *Cocos nucifera* as it is for other species with more available genomic information. Previous analyses of coconut using RFLP, microsatellite and AFLP markers has found two genetically dissimilar populations, corresponding broadly to the Pacific Ocean basin and the Indian and Atlantic Oceans (Teulat et al., 2000; Perera et al., 2003; Gunn et al., 2011). The possible genetic origin of Pacific Ocean coconuts is within a region extending from the Malay Peninsula to New Guinea, including some countries from Southeast Asia, such as Malaysia, Indonesia, Southern Thailand and Myanmar (Harries, 1990; Larson et al., 2005). It has been proposed that Southeast Asia was likely a center of artificial domestication and cultivation for coconuts in the Pacific Ocean basin (Hurles et al., 2005); supporting this theory, the coconuts in this region have rich phenotypic diversity and population heterogeneity (Gunn et al., 2011).On the other hand, China may

Table 2. Polymorphic information of microsatellite markers suitable for *Areca catechu* and *Elaeisguineensis*

Locus	*Areca catechu*		*Elaeisguineensis*	
	Observed allele number	Observed heterozygosity	Observed allele number	Observed heterozygosity
WCYZ2470	2	0.375	2	0.375
WCYZ9240	3	0.5938	-	-
WCYZ8097	4	0.7344	1	0
WCYZ3075	2	0.4688	-	-
WCYZ2116	3	0.5938	2	0.3047
WCYZ8277	2	0.4688	2	0.375
WCYZ6681	3	0.6562	-	-
WCYZ8374	3	0.6562	-	-
WCYZ1753	2	0.5	-	-
WCYZ8635	2	0.375	-	-
WCYZ8558	2	0.375	2	0.375
WCYZ1521	3	0.5391	-	-
WCYZ8939	3	0.5938	-	0.2188
WCYZ3554	3	0.4062	1	0
WCYZ8774	2	0.2188	2	0.4688
WCYZ7691	2	0.4688	2	0.2188
WCYZ8904	2	0.4688	2	0.2188

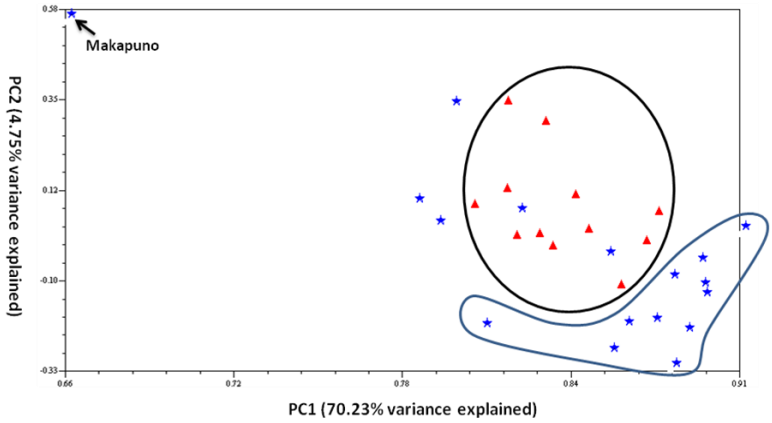

Fig 2. Diagram of association between coconut germplasm from China and Southeast Asia using Principal Component Analysis. Coconuts from China and Southeast Asia are represented by red triangles and blue stars, respectively. The top-left star indicated by the arrow represents a special coconut germplasm called Makapuno. Two genetic groups are indicated: one group in the circle containing both red triangles and blue stars, and the other group comprising only blue stars.

not have originally had coconut crops until approximately 2000 years ago during the Han Dynasty, when they were introduced to China from Southeast Asia (Tang et al., 2006). In China, coconuts were grown in the Hainan and Yunnan provinces as economic and ornamental plants, though due to the limited number of areas where they were cultivated, Chinese coconuts show a low level of genetic diversity and genetic differences between lines (Liu et al., 2011). In this study, we compared the level of genetic diversity within coconut germplasm from southern and southwestern China and Southeast Asia. The results showed that coconut germplasm from Southeast Asia had higher genetic diversity, in accordance with the idea of Southeast Asia as the possible genetic origin of *Cocos nucifera* in the Pacific Ocean region. STRUCTURE analysis of coconut germplasm from Southeast Asia showed a mixture of diverse genotypes; these coconuts can be classified into three groups. However, the population structure of coconut germplasm from China is simpler; all 12 coconuts were classified into one single group. Interestingly, there is not a clear separation between coconut germplasm from China and coconut germplasm from Southeast Asia. This genetic similarity indicates the likelihood that coconuts in southern and southwestern China came originally from Southeast Asia. Based upon historic and geographic information, we may theorize as to how coconuts first came to southern and southwestern China. The Hainan province in southern China is an island in the Pacific Ocean, located on the northern border of the tropics ($N16^0$-22^0). It seems likely that the coconut nut could have been introduced into Hainan around two thousand years ago, after being floated by sea currents from Southeast Asia. However, such natural means of transport could not explain the presence of coconuts in the Yunnan province, which is located in southwestern China at an elevation of 1000-2000 meters. Therefore, the coconut must have spread to Yunnan province due to human activity. In ancient times, it is known that the Yunnan province in China and some Southeast Asian countries traded between each other using a route along the Mekong River. Our results show that coconut germplasm from Yunnan appears to be genetically homogeneous with germplasm from some Southeast Asia countries, especially Myanmar and Thailand. This indicates that coconuts were likely introduced into Yunnan province via commercial contact between these regions along the Mekong River—perhaps not as a commodity at first, but rather because they were brought along from Southeast Asia as an easily portable source of food and water.

Table 3. Statistics of *Cocos nucifera* genetic diversity from different ecological regions.

Groups	NO. of individuals	Observed number of alleles	Observed heterozygosity	Shannon diversity index
China	12	2.733	0.394	0.635
South Asia	18	3.100	0.453	0.750

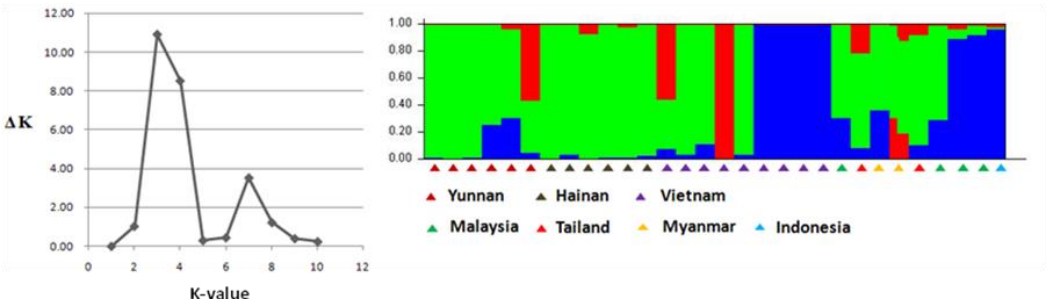

Fig 3. Inferred population structure of coconut germplasm in China and Southeast Asia. Each coconut individual is represented by a single vertical line. Each color represents one cluster. The length of the colored segment indicates the proportion of an individual assigned into one genetic group. The left diagram indicates the true number of genetic clusters.

Table 4. Geographic distribution and description of thirty coconut accessions.

Sample number	Accession name	Height	Fruit Color	Geographic origin
1	GPYZ29CN0022	High	Red	Hainan, China
2	GPYZ34CN0009	High	Blue	Hainan, China
3	GPYZ34CN0001	High		Yunnan, China
4	Hai 77	High	Red	Yunnan, China
5	Qionghai 18	High	Red	Hainan, China
6	Wanglin 4	High	Red	Hainan, China
7	Shanyahongye	High	Red	Hainan, China
8	GPYZ34CN0002	High	-	Yunnan, China
9	GPYZ34CN0030	High	-	Yunnan, China
10	Xiaohuangye	Dwarf	Yellow	Malaysia
11	Bendigao	High	Blue	Hainan, China
12	Xianghuiyezi	Dwarf	Blue	Thailand
13	Yeyao	Dwarf	Blue	Hainan, China
14	Duopeiyezi	Dwarf	-	-
15	GPYZ34CN0025	High	-	Yunnan, China
16	GPYZ**VN0013	Dwarf	-	Vietnam
17	GPYZ**VN0011	-	-	Vietnam
18	GPYZ**VN0008	-	Blue	Vietnam
19	GPYZ**MM0002	Dwarf	-	Myamar
20	GPYZ**VN0006	Dwarf	Red	Vietnam
21	GPYZ29CN0025	High	-	Thailand
22	GPYZ**VN0009	High	-	Vietnam
23	Mawa	High	Blue	Malaysia
24	Hongai	Dwarf	Red	Malaysia
25	Huangai	Dwarf	Yellow	Malaysia
26	GPYZ**VN0003	Dwarf		Vietnam
27	GPYZ**VN0007	Dwarf	Yellow	Vietnam
28	GPYZ29CN0004	Dwarf	Red	Indonesia
29	GPYZ**VN0005	Dwarf	-	Vietnam
30	GPYZ**VN0010	Dwarf	-	Vietnam

Materials and methods

Development of SSR markers

In our previous research, 57,304 unigenes were obtained through sequencing RNA mixture from leaves and fruit flesh of Hainan tall coconut (Fan et al., 2013). Using the program MSATCOMMANDER (Faircloth, 2008), microsatellite loci in the 57,304 unigenes were scanned for, using a minimum repeat number cut-off of six. After that, 76 microsatellite loci were stochastically selected for SSR marker primer design. Primer pairs were designed using Primer 3 from sequence flanking the 76 SSRs: 27 di-, 42 tri- and 7 tetra-nucleotide motif SSRs (Rozen and Skalesky 2000).

Plant materials and PCR amplification

To characterize the 76 microsatellite loci, 10 individuals of *Cocos nucifera* from China and Southeast Asia were used to test the polymorphism of the microsatellite markers. Genomic DNA was extracted from young leaves using CTAB mini protocol (Neal Stewart and Via, 1993). PCR amplifications were performed in 10 µl reaction mixtures containing 100 ng of genomic DNA, 1×PCR buffer, 2 mM $MgCl_2$, 1 U TaqDNA polymerase (TaKaRa, China), 0.5 µM of each primer, and 0.2 mM dNTP mix, with the following program: denaturation for 4 minutes at 94°C, 30 cycles of 94°C for 30 seconds, 30s at 48-57°C and 30 s at 72°C, with a final extension of 7 minutes

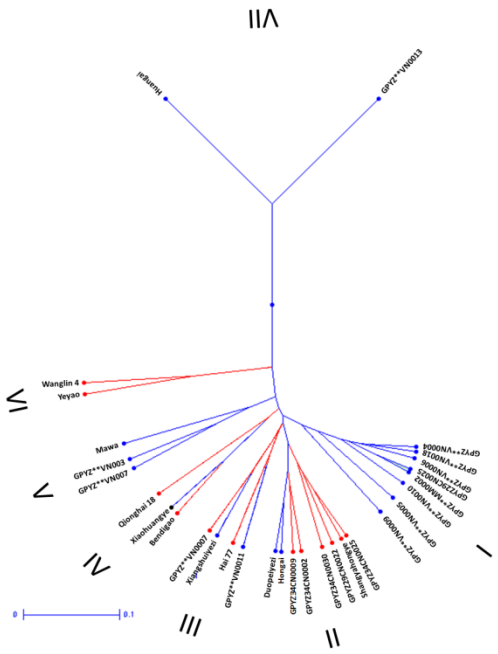

Fig 4. UPGMA dendrogramcreated with DARwin 5 based on genotypic information of 30 microsatellite markers in 30 coconut accessions from China and Southeast Asia. The red lines represent the coconut accessions from China; blue lines represent the coconut accessions from Southeast Asia.

at 72°C. PCR products were electrophoretically separated on 8% polyacrylamide denaturing gels and visualized by silver staining. Product sizes were determined by comparison to a 100 bp DNA ladder.

Genetic structure and Principal Component Analysis

Thirty polymorphic markers selected from the 76 microsatellite loci were used to evaluate the genetic diversity and population structure of 30 coconut individuals from China (six from Yunnan in southwest China and six from Hainan in southern China) and Southeast Asia (nine from Vietnam, two from Myanmar, three from Thailand, three from Malaysia and one from Indonesia) (Table 4). Bayesian clustering was employed to analyze the coconut population structure using the STRUCTURE software (Pritchard et al., 2000) in its revised version 2.2 (Falush et al., 2003). The length of burn-in time and replication number were both set up to 100,000 in each run. Maximum likelihood was applied to assign every sampled line

to clusters with a probability cut-off of 0.6. The most probable number of populations (K) was identified by plotting ΔK values of K, from 1 to 10 replicate runs for each K and corresponding to the peak of the ΔK graph. The Principal Component Analysis was performed using the NTSYS-PC program (Rohlf, 1997).Before clustering, genetic dissimilarity was calculated with DARWIN (Perrier and Jacquemoud-Collet, 2006), set to 100 bootstraps. The phylogenetictree was constructed using the Unweighted Pair Group Method (UPGMA).

Conclusions

Using data from Illumina sequencing, we developed and characterized 30 gene-based microsatellite markers in a sample of 30 *Cocos nucifera* individuals from China and Southeast Asia. This method for SSR marker development is efficient and the developed markers are suitable for ongoing study of *Cocos nucifera*, including analyses of population structure and genetic diversity. Some specific SSR alleles related to geographical origin will be valuable for the study of genetic divergence caused by geographical isolation and cultivar identification. Moreover, 13 microsatellite markers also have potential applicability to other species in the Palmaceae family. Based on genotypic data from the 30 microsatellite markers, we compared the genetic diversity in coconut: higher genetic diversity was found in coconut accessions from Southeast Asia compared to Chinese coconut accessions. Population structure analyses of the 30 coconut accessions showed that some coconut accessions from Southeast Asia have genetic similarity to Chinese coconuts. By combining molecular data and historic information, we propose two methods for the spread of coconuts from Southeast Asia to China: natural dispersal by sea currents to the Hainan province, and transport by human traders from Southeast Asia to the Yunnan province.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31101179), Special Fund for Agro-scientific Research in the Public Interest (200903026), The Project for High-Level Returned Scholar of Ministry of Human Resources and Social Security of P.R.China (2011) and the National Nonprofit Institute Research Grant of CATAS (1630032012044). ASM is supported by an Australian Research Council Discovery Early Career Researcher Award (DE120100668).

References

Akpan EEJ (1994) Evaluation of tall coconut (*Cocos nucifera* L.) genotypes within the Nigerian coconut germplasm bank. Oleagineux. 49: 13-20.

Ashburner GR, Thompson WK, Halloran GM (1997) RAPD analysis of South Pacific coconut palm populations. Crop Sci. 37: 992-997.

Batugal P, Rao V, Oliver J, eds (2005) Coconut Genetic Resources. IPGRI-Regional Office for Asia, the Pacific and Oceania, Serdang, Malaysia.

Baudouin L, Lebrun P (2009) Coconut (*Cocos nucifera* L.) DNA studies support the hypothesis of an ancient Austronesian migration from Southeast Asia to American. Genet Resour Crop Ev. 56: 257-262.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 14: 2611-2620.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol. 8: 92-94.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus data: linked loci and correlated allele frequencies. Genetics. 164: 1567-1587.

Fan H, Xiao Y, Yang Y, Xia W, Mason AS, Xia Z, Qiao F, Zhao S, Tang H (2013) RNA-Seq Analysis of *Cocos nucifera*: Transcriptome sequencing and *De Novo* assembly for Subsequent Functional Genomics Approaches. PLOS ONE. 8: e59997.

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. PLOS ONE. 6: e21143.

Harries HC (1978) Evolution, dissemination and classification of *Cocos nucifera* L. Bot Rev. 44: 265-320.

Harries HC (1981) Germination and taxonomy of the coconut. Ann Bot. 48: 873-883.

Harries HC (1990) Malesian origin for a domestic: *Cocos nucifera*. In: Bass P, Kalkman K, Geesink R, eds. The Plant Diversity of Malesia. Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 351-357.

Hurles M, Skyes B, Jobling M, Forster P (2005) The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal line ages. Am J Hum Genet. 76: 894-901.

Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Stewart L, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A (2005) Worldwide phyloeography of wild boar reveals multiple centers of pig domestication. Science. 307: 1618-1621.

Lebrun P, N'Cho YP, Seguin M, Grivet L, Baudouin L (1998) Genetic diversity in coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. Euphytica. 101: 103-108.

Liu X, Tang H, Li D, Hou L (2011) Genetic diversity of coconut cultivars in China by microsatellite (SSR) markers. Mole Plant Breed. 2: 83-91.

Menon KPV,Pandalai KM (1958) The coconut. A monograph. Indian central coconut committee. Government press, Madra, India. pp 86-102.

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markets in plants. Plant J. 3: 175-182.

Neal Stewart C, Via LE (1993) A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. Biotechniques. 14: 748-749.

Perera AU, Russell JR, Provan J, McNicol JW, Powell W (1998) Evaluating genetic relationships between indigenous coconut (*Cocos nucifera* L.) accessions from Sri Lanka by means of AFLP profiling. Theor Appl Genet. 96: 545-550.

Perera I, Russell JR, Provan J, Powell W (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. Eupytica. 132: 212-128.

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwinciradfr/darwin.

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215-222.

Pritchard JK, Stephen M, Donnelly (2000) Inference of population structure using multilocus genotype data. Genetics. 155: 945-959.

Rajesh MK, Jerard BA, Preethi P, Thomas RJ, Fayas TP, Rachana KE (2013) Development of a RAPD-derived SCAR marker associated with tall-type palm trait in coconut. Sci Hortic. 150: 312-316.

Reusch TB, Bolte S, Sparwel M, Moss AG, Javidpour J (2010) Microsatellites reveal origin and genetic diversity of Eurasian invasions by one of the world's most notorious marine invader, *Mnemiopsis leidyi* (Ctenophora). Mol Ecol. 19: 2690-2699.

Rivera R, Edwards KJ, Barker JHA, Arnold GM, Ayad G, Hodgkin T, Karp A (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. Genome. 42: 668-675.

Rohlf FJ (1997) NTSYS-PC 2.1. Numerical taxonomy and multivariate analysis system. Exeter Software, Setauket, N.Y.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In S. Krawetz and S. Misener [eds.], Bioinformatics methods and protocols: Methods in molecular biology, 365-386. Humana Press, Totowa, New Jersey, USA.

Sugimura Y, Itano M, Salud CD, Otsuji K, Yamaguchi H (1997) Biometric analysis on diversity of coconut palm: cultivar classification by botanical and agronomical traits. Euphytica. 98: 29-35.

Tabkhkar N, Rabiei B, Sabouri A (2012) Genetic diversity of rice cultivars by microsatellite markers tightly linked to cooking and eating quality. Aust J Crop Sci. 6: 980-985.

Tang BN, Tang M, Chen CF, Qiu PH, Liu Q, Wang M, Li CE (2006) Characteristics of soil fauna community in the dongjiao coconut plantation ecosystem in Hainan, China. Acta Ecologica Sinica, 26: 26-32.

Teulat B, Aldam C, Trehin R, Lebrun P, Barker JHA, Arnold GM, Karp A, Baudouin L, Rognon F (2000) An analysis of genetic diversity of the coconut (*Cocos nucifera*) population from across the geographic range using sequence-tagged microsatellite (SSRs) and AFLPs. Theor Appl Genet. 100: 764-771.

Whitehead RA (1966) Sample survey and collection of coconut germplasm in the Pacific island (30 May – 5 September 1964). London, UK; HMSO, Ministry of Overseas Development.