

Genetic diversity of *Jatropha curcas* collections from different islands in Indonesia

Tantri Dyah Ayu Anggraeni^{1,2}, Dani Satyawan^{1,3}, Yang Jae Kang¹, Jungmin Ha^{1,4}, Moon Young Kim^{1,4}, Annapurna Chitikineni⁵, Rajeev K. Varshney⁵ and Suk-Ha Lee^{1,4*}

¹Department of Plant Science and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea, ²Indonesian Sweetener and Fiber Crops Research Institute, IAARD, Malang 65100, Indonesia, ³Indonesian Center for Agricultural Biotechnology and Genetic Resources and Development, IAARD, Bogor 16111, Indonesia, ⁴Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Republic of Korea and ⁵Centre of Excellence in Genomics, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Andhra Pradesh, India

Received 13 July 2017; Accepted 8 December 2017 – First published online 5 March 2018

Abstract

Jatropha curcas L. is a potential bioenergy crop but has a lack of improved cultivars with high yields and oil content. Therefore, increasing our understanding of *J. curcas* germplasm is important for designing breeding strategies. This study was performed to investigate the genetic diversity and population structure of Indonesian *J. curcas* populations from six different islands. To construct a reference, we *de novo* assembled the scaffolds (N50 = 355.5 kb) using 182 Gb Illumina HiSeq sequencing data from Thai *J. curcas* variety Chai Nat. Genetic diversity analysis among 52 Indonesian *J. curcas* accessions was conducted based on yield traits and single nucleotide polymorphism (SNP) markers detected by mapping genotyping-by-sequencing reads from Indonesian population to Chai Nat scaffolds. Strong variation in yield traits was detected among accessions. Using *J. integerrima* as an outgroup, 13,916 SNPs were detected. Among *J. curcas* accessions, including accessions from other countries (Thailand, the Philippines and China), 856 SNPs were detected, but only 297 SNPs were detected among Indonesian *J. curcas* populations, representing low genetic diversity. Through phylogenetic and structural analysis, the populations were clustered into two major groups. Group one consists of populations from Bangka and Sulawesi in the northern part of Indonesia, which are located at a distance of 1572.59 km. Group two contains populations from islands in the southern part: Java, Lombok-Sumbawa, Flores and Timor. These results indicate that introduction of diverse *J. curcas* germplasms is necessary for the improvement of the genetic variation in the Indonesian collections.

Keywords: genetic diversity, genotyping-by-sequencing, Indonesia, islands, *Jatropha curcas*, migratory rout

Introduction

Jatropha curcas L. is a perennial shrub in the Euphorbiaceae family that produces oil that can be used as a substitute for diesel. The oil characteristics are

competitive with those of diesel oil from fossil fuels (Heller, 1996; de Oliveira *et al.*, 2009). The toxic components contained in the oil make it non-edible oil, therefore preventing competition between foods versus fuel. *J. curcas* has several advantages for the environment compared with other bioenergy sources, as this crop is drought resistant and widely adapted to different agroclimates and marginal lands, allowing it to be used to control erosion and for

*Corresponding author. E-mail: sukhalee@snu.ac.kr

land reclamation (Fearless, 2007). In addition, the by-product of the seeds contains high nitrogen, phosphorous, and potassium levels and can thus be used as fertilizer. Therefore *J. curcas* has attracted much attention as a new bioenergy crop that also has a positive impact on the environment.

J. curcas is believed to be native to Mexico, Central America, and South America, where it grows in the forests of coastal regions. This plant has been widely disseminated and has become naturalized in many parts of the tropical and subtropical world. Portuguese traders are believed to have distributed this species from the Caribbean (via the Cape Verde Islands) to countries in Africa and Asia, including Indonesia, in the 16th century (Heller, 1996).

Indonesia, a developing country and the fourth most populous country worldwide, announced in its National Energy Policy in 2006 that targeted development of biodiesel plants, including *J. curcas*, to fulfill the demands of oil consumption is continually increasing (Silitonga *et al.*, 2011). However, crop improvement of this plant is at an early stage. *J. curcas* remains undomesticated, and improved plant varieties with desirable characteristics are not yet available. To develop new varieties of *J. curcas* with desirable characteristics, comprehensive characterization of genetic variation of this species is required based on molecular markers as well as morphological traits. Even though genetic diversity has been recently studied in a small number of Indonesian *J. curcas* accessions using molecular markers (Wen *et al.*, 2010; Osorio *et al.*, 2014; Maghuly *et al.*, 2015), detailed analyses of genetic diversity, relationships, and differentiation among Indonesian accessions from different islands have not yet been performed. Therefore, in the current study, we investigated the genetic diversity, genetic differentiation level, and genetic relationships among *J. curcas* collections from six islands in Indonesia based on yield traits and single nucleotide polymorphism (SNP) markers generated by genotyping-by-sequencing (GBS).

Materials and methods

Plant materials, field experiments and analysis of yield trait variation

For genome assembly of *J. curcas*, we used a Thai variety 'Chai Nat' (hereafter called 'CN') which is one of the mapping parents used in a previous study to construct a linkage map (Amkul *et al.*, 2016). CN produces toxic seeds possessing high phorbol esters (Amkul *et al.*, 2016). The plant materials used for genetic diversity study included 52 *J. curcas* accessions from six different islands in Indonesia derived from Indonesian Sweetener and Fiber Crops Research Institute (ISFCRI) germplasm collections (Table S1 and Fig. S1). The islands of origin were Bangka

and Sulawesi (located in the northern part of the Indonesian archipelago) and Java, Lombok-Sumbawa, Flores, and Timor (located in the southern part of the archipelago) (Fig. S1). The accessions were planted at Asembagus Experimental Garden in Situbondo, East Java in 2005 as living collections. The annual rainfall is 920 mm/year, and the altitude is 5.5 m above sea level. The plant materials were derived from vegetative cuttings and planted in July 2011. The experiment was arranged in a completely randomized block design with five replications. Yield trait observations were conducted for fruit number, seed yield and oil content. Fruit number and seed yield were measured yearly for every tree in the first to third year of the study (2012–2014). Oil contents were measured from seeds harvested in 2012 with two replications. Seed oil was extracted using Soxhlet extraction method (Amalia Kartika *et al.*, 2013) and oil contents were measured as a percentage of oil weight to seed weight on dry basis. The variation between accessions was analysed using analysis of variance (ANOVA) using R software.

Genomic DNA isolation, whole-genome sequencing and GBS

Genomic DNA was extracted from leaves using an MGTM Plant Genomic DNA Extraction SV Miniprep kit (MacroGen Inc., Seoul, South Korea) and quantified using an ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). DNA quality was confirmed using gel electrophoresis. For whole genome sequencing of CN, a paired-end library with insert sizes of 200 bp and a mate-pair library with insert sizes of 5 and 10 kb were constructed from nuclear DNA according to the manufacturer's protocol (Illumina Inc., San Diego, CA, USA).

GBS was performed as follows. Genomic DNA was digested using the restriction enzyme ApeKI. Digested DNA fragments were ligated to the P1 and P2 adapters at both primer sites, and the end sites were repaired via the addition of A. The P1 adapter contained a forward amplification primer site, an Illumina sequencing primer site, and a barcode sequence. After ligation, the fragments were PCR amplified with P1- and P2-specific primers according to the GBS protocol as described in Elshire *et al.* (2011). The library was validated on an Agilent Technologies 2100 Bio-analyser and the ABI StepOnePlus Real-Time PCR system. The libraries were pooled and sequenced using Illumina TrueSeq Version 3.0 single end sequencing chemistry with reading lengths of 100 bp in one lane on HiSeq 2000 Platform.

Genome assembly and SNP discovery

After trimming of raw read sequences from CN, trimmed reads were assembled by using the software package of

All-PATHS-LG with default parameters (Gnerre *et al.*, 2011). The assembled scaffolds were used as a reference to detect nucleotide variations between CN and the *J. curcas* collections. Sequencing data generated from GBS were filtered and sorted based on the barcode adaptor to assign the sequence reads to each sample. The sorted sequences were trimmed from the barcode adaptor and processed via sequential steps to obtain variation data. As a comparison, *Jatropha* sequences from other countries were used, which were available in a database, including (1) whole-genome sequence DNA from a *J. curcas* line from Palawan Island, the Philippines (Sato *et al.*, 2011); (2) transcriptome sequence of a physic nut inbred line from China; and (3) transcriptome sequence of *J. integerrima* (Wu *et al.*, 2015). These sequences were downloaded from the DNA Data Bank of Japan (DDBJ/GenBank/EMBL, <http://www.ddbj.nig.ac.jp/>).

First, GBS sequence data from every sample were aligned to the scaffold sequences of CN using Bowtie2 software (Langmead and Salzberg, 2012). The generated output, a Sequence Alignment Map (SAM) format file, was then converted to a Binary Alignment Map (BAM) format file and sorted. Finally, SAMtools were used to generate a pileup of reading bases using the alignments to a reference sequence, and BCFtools were used to call variants from the output of the SAMtools mpileup command. The variant calling results were stored in a Variation Calling Format (VCF) file.

Genetic diversity parameter and phylogenetic tree analyses

The genetic diversity parameters were measured for allele number (Na), an effective number of alleles (Ae), Shannon information index (I), expected heterozygosity (He), and fixation index (F). Genetic differentiation among populations was estimated using analysis of molecular variation (AMOVA). To estimate the presence of isolation by distance, a Mantel test was conducted to calculate the correlation between genetic distances and geographical distances among Indonesian *J. curcas* collections. These analyses were performed using GenALEX 6.5 (Peakall and Smouse, 2012).

Population structure was analysed using STRUCTURE 2.3.4. (Pritchard *et al.*, 2000). The estimated number of populations (K) was set from 1 to 10, with a 100,000 step burn-in and 100,000 MCMC iterations. Allele frequencies were assumed to be correlated among populations, and an admixture model was chosen. This process was repeated 10 times for each K. The true K value was determined using log likelihood and the ΔK method (Evanno *et al.*, 2005), which were calculated using STRUCTURE HARVESTER (Earl and vonHoldt, 2012).

A phylogenetic tree was constructed using Nei's genetic distance (Nei *et al.*, 1983) with PowerMarker 3.25 (Liu and Muse, 2005), and an unrooted neighbour-joining tree with 1000 bootstrap replications was drawn using Mega 6.0 (Tamura *et al.*, 2013) and visualized using iTOL (Letunic and Bork, 2011).

Results

Variations in Indonesian accessions based on yield traits

We detected highly significant variations between Indonesian accessions based on yield traits (Tables S2 and S3). Fruit number and seed yield in the *J. curcas* plantation decreased from the first to the second year but increased from the second to the third year. The variations were also indicated by the high coefficient of variation values. Accession code (Acc) 33 from Timor Island and Acc 35 from Sulawesi Island showed the highest yields, with relatively constant yields over 3 years (Table S2). However, these two accessions had different performance in terms of oil content. Acc 33 had low oil content (28.4%), whereas Acc 35 had higher oil content (37.7%). Moreover, Acc 13 from Java Island had the highest oil content (43.5%) and high seed yields in the third year (Table S2).

We constructed a phylogenetic tree based on yield traits, which grouped the Indonesian accessions into nine clusters (Fig. S2). There were admixtures where accessions from different islands were grouped in the same cluster. Clusters I–VI are dominated by accessions from Sulawesi Island mixed with some accessions from other islands, except Bangka Island, whereas all accessions from Bangka Island were grouped into cluster IX.

J. curcas genome assembly and SNP detection

Our genome assembly of *J. curcas* (var. CN) was performed using approximately 182 Gb short-read sequences with a depth of 437 \times generated from standard paired-end and 5 to 10 Kb mate-pair libraries on Illumina HiSeq 2000 (Table S4 and Fig. S3). A total number of the produced scaffolds is 3710, which corresponds to a total length of 319.04 Mb covering 76.7% of the *J. curcas* genome (Fig. S3). The N50 scaffold size is 355.5 kb. The largest and smallest scaffolds are 2.14 Mb and 920 bp, respectively. This assembled sequence can be downloaded from <http://plantgenomics.snu.ac.kr/>.

The number of processed reads (filtered and trimmed) for each accession generated by the GBS method ranged from 1,450,218 (Acc 153 from Timor Island) to 5,724,830 (Acc 65 from Java Island), with an average 3,259,792 reads (Table S5). GBS reads were mapped to the CN

scaffolds as the reference sequence. Alignments to the CN scaffold sequence showed a low alignment rate for all accessions. The average alignment rate per accession was 47.8%. Most reads were aligned only once to the CN scaffold sequences, with an average alignment rate of 41.7% per accession, and only 6.1% of the reads were aligned more than once. This low alignment rate of GBS reads may be attributed to unassembled or misassembled genome sequences of *J. curcas*, those of which are mostly repetitive sequences.

The variation-calling step generated 13,916 SNPs, which possessed a minimum read depth of three and a minimum Phred-scale quality score of 30. The average coverage of the SNPs was 2.1% of the genome, and the average number of bases per SNP was 485 (Table S5). Among the 13,916 SNPs, only 6% were polymorphic within *J. curcas* accessions from Indonesia, China, Thailand, and the Philippines, and 297 SNPs were polymorphic among Indonesian collections.

Genetic diversity parameters of Indonesian *J. curcas* collections

The genetic diversity parameters among *J. curcas* populations are shown in Table 1. A population from Sulawesi Island showed the highest N_a (1.808), whereas a population from Bangka Island showed the lowest N_a (1.549). The number of samples might be correlated with the allele numbers that could be detected in the populations. The number of alleles increased with increasing sample number. Collections from Timor Island showed the highest A_e , I and H_e values, indicating that the Timor population had the highest diversity among Indonesian collections. Fixation indices were negative in all populations, indicating that there was an excess of heterozygosity (open pollination). Overall, the Indonesian *J. curcas* collections showed a low level of H_e (0.271), implying that, at a single locus, the probability that any two alleles will be different

from each other is only 27.1%. This result is in contrast with the high variation level found in yield traits (Table S3).

Genetic differentiation

We conducted AMOVA to reveal genetic differentiation in the Indonesian *J. curcas* collections. The results show significant genetic differentiation ($P < 0.0016$), even though only low genetic differentiation is indicated by the value of $F_{ST} = 0.04$ (Table 2). Wright (1978) categorized F_{ST} of 0–0.05 values as showing low genetic differentiation, which was also indicated by the low contribution of variation among a population to the total variation within the population (4%). To investigate whether isolation by distance contributed to the genetic differentiation in the *J. curcas* populations, we conducted Mantel analysis to reveal the relationship between genetic distances (F_{ST}) and geographic distances (km) among subpopulations (islands). The results show a non-significant value of R_{xy} , indicating a lack of isolation by distance among the Indonesian populations (Fig. S4).

Population structure analysis

The bar chart in Fig. S5 shows the grouping of populations and the proportions of alleles in each genetic group (different colours) for each individual (upper vertical bar) per geographical population. Based on the delta K method, the true K is 2. At $K = 2$, there was mixed clustering of the *J. curcas* accessions from different islands. However, we detected separation of accessions located in the northern versus southern parts of the archipelago. In particular, the accessions from Java, Lombok-Sumbawa, Flores and Timor Islands (located in the southern part) were not placed in the same cluster with accessions from Bangka and Sulawesi Islands (in the northern part). The exception is two

Table 1. Genetic diversity parameters among Indonesian *J. curcas* collections

Island of origin	No. of accessions	No. of alleles (N_a)	Effective number of alleles (A_e)	Shannon information index (I)	Gene diversity/expected heterozygosity (H_e)	Fixation index/inbreeding coefficient (F)
Bangka	4	1.549	1.451	0.345	0.241	−0.623
Java	9	1.673	1.449	0.368	0.250	−0.332
Lombok-Sumbawa	5	1.694	1.512	0.411	0.283	−0.443
Flores	8	1.717	1.508	0.408	0.280	−0.390
Timor	8	1.744	1.519	0.422	0.289	−0.368
Sulawesi	18	1.808	1.506	0.415	0.281	−0.334
Total	52	1.698	1.491	0.395	0.271	−0.404
Standard deviation	–	0.011	0.011	0.007	0.005	0.014

Table 2. Analysis of molecular variance (AMOVA) of Indonesian *J. curcas* collections

Source	Degree of freedom	Sum of squares	Mean square	Est. Var.	%	F statistic ^a	P value
Among islands	5	369.52	73.90	1.84	4	$F_{ST} = 0.04^*$	<0.016
Within island	98	4229.01	43.15	43.15	96	–	–
Total	103	4598.53	–	45.03	100	–	–

^aMeasure of genetic differentiation among subpopulations to the total populations.

accessions located in Sumbawa Island, which were grouped in the same cluster with accessions from Sulawesi Island. Furthermore, we observed the same patterns of mixed ancestry in every individual from $K=2$ to 7. There was very little admixture in collections from Bangka, Java and Sulawesi Island, whereas higher levels of admixture were found in collections from Lombok-Sumbawa, Flores and Timor Island.

Genetic relationship of *J. curcas*

In the unrooted neighbour-joining tree, the populations are grouped into four clusters (Fig. 1(a)). The first cluster consists of populations from Sulawesi Island and two accessions from Lombok-Sumbawa Island. The second cluster consists of all accessions from Bangka Island and three accessions from Sulawesi Island, which is located in the southernmost portion of the islands. The third cluster consists of all accessions from Java Island and two accessions from Timor Island. Finally, the fourth cluster consists of accessions from Lombok-Sumbawa, Flores, and Timor Island. The phylogenetic tree shows a clear separation between accessions from Sulawesi and Bangka Island and accessions from Java, Flores and Timor Island. However, accessions from Lombok-Sumbawa Island are divided into two major clusters. Interestingly accessions from Sulawesi and Bangka are grouped together, even though these two islands are separated by 1572.59 km. These results confirm the STRUCTURE analysis results (Fig. S5).

In addition, we constructed a phylogenetic tree showing the genetic relationships among Asian *J. curcas* accessions from Indonesia, the Philippines, Thailand and China and between *J. integerrima*, a related species in the genus *Jatropha*: this rooted neighbour-joining tree was drawn based on Nei's genetic distances (Fig. 2). *J. integerrima* which served as an outgroup clustered separately from *J. curcas* accessions, with far genetic distances. Asian *J. curcas* accessions clustered geographically; *J. curcas* accessions from China (East Asia) are in separate clusters from *J. curcas* accessions from Indonesia, Thailand and the Philippines (Southeast Asia), which are in the same cluster. Indonesian accessions clustered separately from accessions from Thailand and the Philippines.

Discussion

In this study, we found very high variability in yield traits, including fruit number, seed yield, and oil content, among Indonesian *J. curcas* accessions from six different islands. Several other studies also showed high variability in phenotypic and yield traits among *J. curcas* collections (Yi et al., 2010; Saadaoui et al., 2015). These results can help facilitate *J. curcas* breeding efforts to create high-yielding varieties. The variability in yield traits can provide material for selection. Some accessions with consistently high yields and oil contents that were analysed in this study represent valuable parents for use in hybridization.

Surprisingly, we detected low genetic diversity among collections, unlike the strong variation in yield traits. Using *J. integerrima* as an outgroup, we successfully generated a relatively high number of SNP markers. However, among SNPs, only 6% were polymorphic in the *J. curcas* accessions. Similar studies with other plant species also showed reduced numbers of polymorphic SNPs with related species, but the reduction was not as high as in the current study. There were 21.9% polymorphic SNPs in pigeon pea (*Cajanus cajan*) compared with its wild relative species (Kassa et al., 2012), and 31.5% polymorphic SNPs in soybean (*Glycine max*) compared with its wild relative, *Glycine soja* (Wang et al., 2016). The strong reduction in SNP numbers among *J. curcas* accessions may reflect the low genetic variation in this plant gene pool (Gupta et al., 2012). This low genetic diversity was also revealed by a low *He* value (0.271); similar results were obtained in other studies conducted on populations in Asia and Africa (Machua et al., 2011; Na-ek et al., 2011; Tanya et al., 2011), with values lower than those of populations from Chiapas, Mexico (*He* = 0.34–0.54; Sanou et al., 2015).

Phylogenetic trees based on yield traits and molecular markers showed an admixture of the members of clusters. However, molecular marker-based clustering showed a clear separation between accessions from islands located in the northern versus southern part of the archipelago (Fig. 1(a)). Many studies suggest that the low genetic diversity in Asian and African accessions was caused by the narrow genetic base of the introduced germplasm. In the present study, the grouping appeared only to reveal the actual origin of the germplasm rather than new genetic

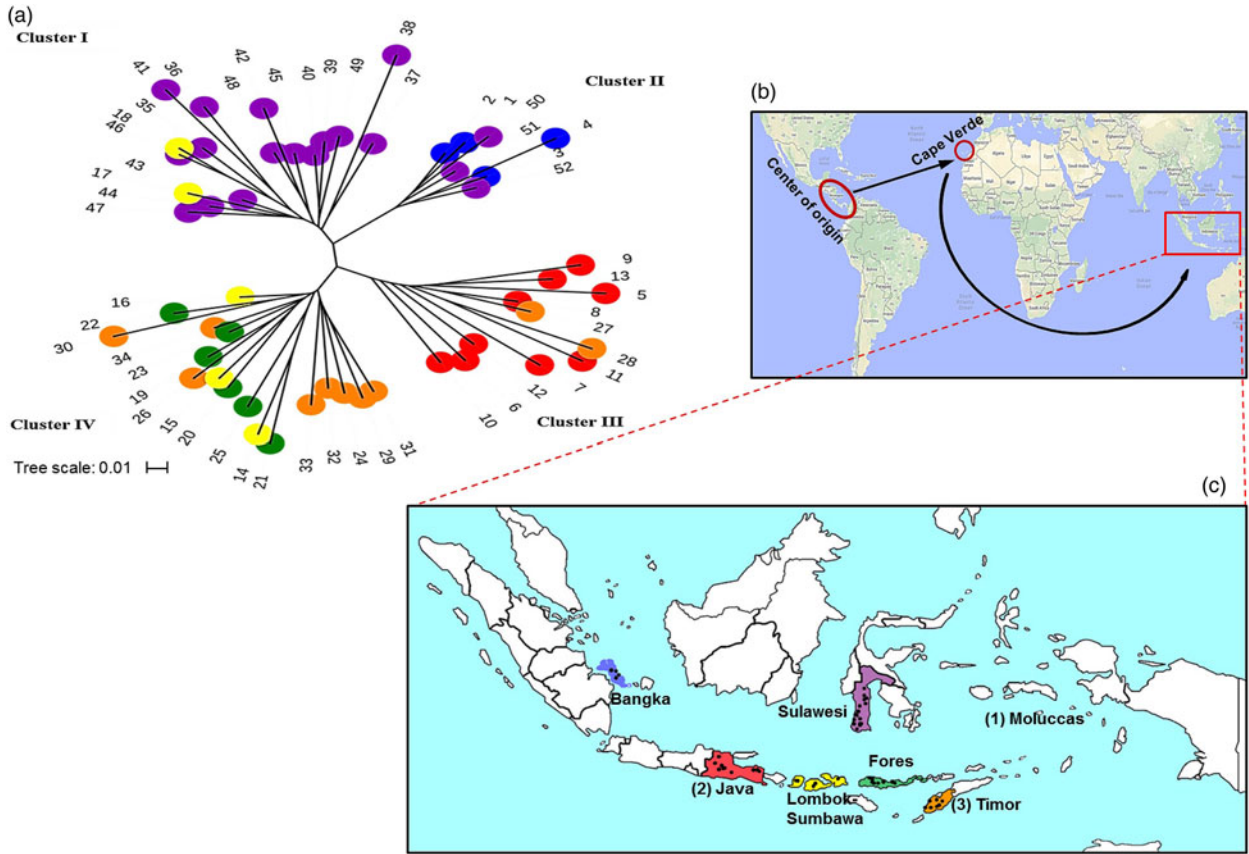


Fig. 1. Genetic relationships among Indonesian *J. curcas* collections and migratory route of *J. curcas*. (a) Unrooted tree showing genetic relationships among Indonesian *J. curcas* collections. The tree was constructed based on Nei’s genetic distance with 1000 bootstrap replications. Each colour represents the island of origin: Blue = Bangka, red = Java, yellow = Lombok-Sumbawa, green = Flores, orange = Timor, and purple = Sulawesi. (b) Possible migratory route of *J. curcas* into Indonesia based on the previous studies (Heller, 1996; Pamidimarri and Reddy, 2014). (c) Possible introductory locations of *J. curcas* in Indonesia based on Portuguese seafarer entrance location during the 16th century through (1) Moluccas Island, (2) Java Island and/or (3) Timor Island.

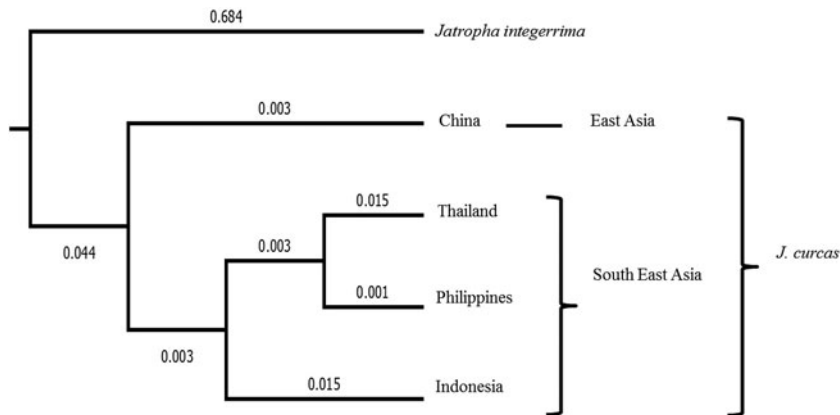


Fig. 2. Rooted neighbour-joining tree based on Nei’s genetic distances among *J. curcas* and *J. integrerrima*. Numbers above the branches show branch lengths.

diversity that has arisen due to geographical barriers, since the time since introduction is still not sufficient to generate new distinctiveness. Accessions from the northern part of the archipelago (Sulawesi Island) are separated from accessions from islands located in the southern part (Java, Flores and Timor). This result suggests that accessions from Sulawesi and the southern part are distinct. However, some accessions from Bangka (northern part) and Sumbawa (southern part) are similar to accessions from a particular location in Sulawesi, indicating that there was dispersion from Sulawesi to these two islands. Another study (Satyawan and Tasma, 2011) also suggested that most of the Indonesian *J. curcas* diversity probably reflects diversity in the ancestral population of the introduced germplasms than local differentiation.

To confirm this notion, we tried to estimate the possible introduction and dispersion history by examining the history of Indonesia itself. Heyne (1927), a Dutch botanist, recorded several local names for *J. curcas*, indicating that, at that time, this plant species had already spread in Indonesia. This confirms the assumption that the Portuguese brought *J. curcas* from its centre of origin to other countries in Africa and Asia through the Cape Verde Islands during the 16th century (Heller, 1996; Brittain and Lutaladio, 2010) (Fig. 1(b)). The Portuguese came to Indonesia before the Dutch, in 1512–1515, through 2–3 locations: (1) the northern part: Moluccas Island; (2) the southern part: Java Island; and/or (3) the southern part: Timor Island (Poesponegoro and Notosusanto, 1984), as shown in Fig. 1(c). STRUCTURE and phylogenetic tree analyses separated accessions from Sulawesi (located near Moluccas Island) from Java and Timor Island, suggesting that *J. curcas* germplasms might have been introduced to two to three locations, which were the entry points for Portuguese traders in this archipelago.

Furthermore, the mixed grouping between accessions from Sulawesi, Bangka, and Lombok suggests the presence of dispersion via human activity. Beginning in the 14th century, Bugis people (Sulawesi) are thought to have established trade routes across Southeast Asia (Flores, Timor, Moluccas, Borneo, Manila, Cebu, Cambodia, Malacca, Siam, Bali, Java and Sumatera-Bangka) and to have had excellent navigational knowledge (Ammarel, 1999; Lampe, 2012). However, even though Sulawesi Island is near Moluccas Island, and there is historical evidence that the Bugis people had a trade route to the Moluccas, the possibility that Sulawesi accessions were derived from Moluccas Island must be confirmed with future research, as there were no accessions from the Moluccas in the current study. Studies on genetic diversity analysis of *J. curcas* germplasm from India also indicated that only two distinct germplasms were introduced to India and distributed throughout this country through human activity (Pamidimarri and Reddy, 2014).

The low level of genetic diversity found in the present study and another study contrasts with the fact that *J. curcas* is an outcrossing species. Outcrossing species exhibit heterozygosity, since their mating pattern is autogamous, which can create high genetic diversity. However, in the case of *J. curcas*, excess heterozygosity in some populations did not positively contribute to high variation. Some studies have shown that *J. curcas* is an outcrossing species, but with a high capability of producing fruits through selfing (Raju and Ezradanam, 2002; Negussie *et al.*, 2014). In general, selfing leads to increased levels of homozygosity and limited gene flow. Therefore, the low level of genetic variation in the *J. curcas* collections outside the centre of origin might have been caused by the narrow germplasm origins and genetic bottlenecks generated in the early dispersion process, combined with the ability of the species to self-pollinate (Achten *et al.*, 2010; Sanou *et al.*, 2015). In addition, *J. curcas* is easily propagated by cutting and grafting. The distribution of cuttings and grafting by farmers might also have contributed to the spreading of the same genotypes to other locations, which could reduce the genetic diversity level of the introduced germplasm, which already had a narrow genetic base (Ouattara *et al.*, 2014). The high variation in yield traits could be caused by environmental effects, various agricultural practices, and epigenetic effects (Yi *et al.*, 2010). However, these assumptions require further study.

In summary, even though we found high variation in yield traits, we detected low levels of genetic diversity in the Indonesian *J. curcas* collections, perhaps due to the narrow genetic diversity of the introduced germplasms. Therefore, the genetic variation in the Indonesian collections must be increased, which could be achieved by (1) crossing with accessions from other countries and from the species origin (Mexico) and (2) crossing with related species. These strategies are expected to positively contribute to breeding and conservation efforts in *J. curcas* plantations in Indonesia.

Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262117000387>.

Acknowledgements

This work was supported by a grant from the Next Generation BioGreen 21 Program (Code No. PJ01326101), Rural Development Administration, Republic of Korea and also funded by the Indonesian Agency for Agricultural Research and Development (IAARD), Ministry of Agriculture through the Sustainable Management of Agricultural Research and Technology Dissemination (SMARTD) project. The authors gratefully acknowledge

Dr Mastur, Dr Rully Dyah Purwati, and Dr Sesanti Basuki (ISFCRD), IAARD, Indonesia, for providing access to the *J. curcas* germplasm and the laboratory facility for DNA extraction.

References

- Achten WMJ, Nielsen LR, Aerts R, Lengkeek AG, Kjaer ED, Trabucco A, Hansen JK, Maes WH, Graudal L, Akinnifesi FK and Muys B (2010) Towards domestication of *Jatropha curcas*. *Biofuels* 1: 91–107.
- Amalia Kartika I, Yani M, Ariono D, Evon P and Rigal L (2013) Biodiesel production from *Jatropha* seeds: solvent extraction and in situ transesterification in a single step. *Fuel* 106: 111–117. doi: 10.1016/j.fuel.2013.01.02.
- Amkul K, Panngam M, Tanya P, Srinives P and Laosatit K (2016) Pollen viability and seed set of interspecific hybrids between *Jatropha curcas* x *Jatropha integerrima*. *Genomics and Genetics* 9: 50–55.
- Ammarel G (1999) *Bugis Navigation*, vol. 2. New Haven: Yale University Southeast Asia Studies.
- Brittaine R and Litaladio NB (2010) *Jatropha curcas* L. In: Brittain R and Litaladio NB (eds) *Jatropha: A Smallholder Bioenergy Crop, the Potential for Pro-poor Development*. Rome: FAO, pp. 13–26.
- de Oliveira JS, Leite PM, de Souza LB, Mello VM, Silva EC, Rubima JC, Meneghetti SMP and Suarez PAZ (2009) Characteristics and composition of *Jatropha gossypifolia* and *Jatropha curcas* L. oils and application for biodiesel production. *Biomass and Bioenergy* 3: 449–453. doi: 10.1016/j.biombioe.2008.08.006.
- Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359–361. doi: 10.1007/s12686-011-9548-7.
- Elshire J, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler E and Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379. doi: 10.1371/journal.pone.0019379.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14: 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.
- Fearless D (2007) Biofuel: the little shrub that could – maybe. *Nature* 449: 652–655.
- Gnerer S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES and Jaffe DB (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1513–1518.
- Gupta P, Idris A, Mantri S, Asif MH, Yadav HK, Roy JK, Tuli R and Mohanty CS (2012) Discovery and use of single nucleotide polymorphic (SNP) markers in *Jatropha curcas* L.. *Molecular Breeding* 30: 1325–1335. doi: 10.1007/s11032-012-9719-6.
- Heller J (1996) Origin and centre of diversity. In: Heller J (ed.) *Promoting the Conservation and use of Underutilized and Neglected Crops. 1. Physic Nut. Jatropha curcas L.* Rome: International Plant Genetic Resources Institute, pp. 13–15.
- Heyne K (1927) *De Nuttige Planten Van Nederlandsch-Indie II (in Dutch)*. Buitenzorg (Bogor), Indonesia: Departement van Landbouw, Nijverheid en Handel.
- Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, Sarma BK, Datta S, Upadhyaya HD, Varshney RK, von Wettberg EJB and Cook DR (2012) Genetic patterns of domestication in pigeon pea (*Cajanus cajan* (L.) Millsp.) and wild *Cajanus* relatives. *PLoS ONE* 7: e39563. doi: 10.1371/journal.pone.0039563.
- Lampe M (2012) Bugis-Makassar seamanship and reproduction of maritime cultural values in Indonesia. *Humaniora* 24: 121–132.
- Langmead B and Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359. doi: 10.1038/nmeth.1923.
- Letunic I and Bork P (2011) Interactive tree of life v2. Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 39: 475–478. doi: 10.1093/nar/gkr201 PMID:21470960.
- Liu K and Muse SV (2005) Powermarker: integrated analysis environment for genetic diversity in core collection accessions of wild barley, *Hordeum vulgare* ssp. *Spontaneum*. *Hereditas* 136: 67–73.
- Machua J, Muturi G, Omondi SF and Gicheru J (2011) Genetic diversity of *Jatropha curcas* L. Populations in Kenya using RAPD molecular markers: implication to plantation establishment. *African Journal of Biotechnology* 10: 3062–3069. doi: 10.5897/AJB10.1990.
- Maghuly F, Jankowicz-Cieslak J, Pabinger S, Till BJ and Laimer M (2015) Geographic origin is not supported by the genetic variability found in a large living collection of *Jatropha curcas* with accessions from three continents. *Biotechnology Journal* 10: 536–551. doi: 10.1002/biot.201400196.
- Na-ek Y, Wongkaew A, Phumichai T, Kongsiri N, Kaveeta R, Reewongchai T and Phumichai C (2011) Genetic diversity of physic nut (*Jatropha curcas* L.) revealed by SSR markers. *Journal of Crop Science and Biotechnology* 14: 105–110. doi: 10.1007/s12892-011-0008-4.
- Negussie AM, Achten WMJ, Verboven H, Hermy M and Muys B (2014) Floral display and effects of natural and artificial pollination on fruiting and seed yield of the tropical biofuel crop *Jatropha curcas* L. *Global Change Biology Bioenergy* 6: 210–218. doi: 10.1111/gcbb.12072.
- Nei M, Tajima F and Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 19: 153–170.
- Orosio LRM, Salvador AFT, Jongschaap REE, Perez CAA, Sandoval JEB, Trindade LM, Visser RGF and van Loo EN (2014) High level of molecular and phenotypic biodiversity in *Jatropha curcas* from Central America compared to Africa, Asia and South America. *BMC Plant Biology* 14: 77. doi: 10.1186/1471-2229-14-77.
- Ouattara B, Ndir KN, Gueye MC, Diedhiou I, Adeline B, Fonceka D, Cisse N, Akpo EL and Diouf D (2014) Genetic diversity of *Jatropha curcas* L. in Senegal compared with exotic accessions based on microsatellite markers. *Genetic Resources and Crop Evolution* 61: 1039–1045. doi: 10.1007/s10722-014-1006-5.
- Pamidimarri DVNS and Reddy MP (2014) Phylogeography and molecular diversity analysis of *Jatropha curcas* L. and the dispersal route revealed by RAPD, AFLP and nrDNA-ITS analysis. *Molecular Biology Reports* 41: 3225–3234. doi: 10.1007/s11033-014-3185-7.
- Peakall R and Smouse PE (2012) Genalex 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537–2539. doi: 10.1093/bioinformatics/bts460.
- Poesponegoro MD and Notosusanto N (1984) *Sejarah Nasional Indonesia IV (in Indonesian)*. Jakarta: Balai Pustaka.

- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Raju AJS and Ezradanam V (2002) Pollination ecology and fruiting behaviour in a monoecious species *Jatropha curcas* L. (Euphorbiaceae). *Current Science* 83: 1395–1398.
- Saadaoui E, Martin JJ, Bouazizi R, Romdhane CB, Grira M, Abdelkadir S, Khouja ML and Cervantes E (2015) Phenotypic variability and seed yield of *Jatropha curcas* L. introduced to Tunisia. *Acta Botánica Mexicana* 110: 119–134.
- Sanou H, Angulo-Escalante MA, Martinez-Herrera J, Kone S, Nikiema A, Kalinganire A, Hansen JK, Kjaer ED, Graudal L and Nielsen LR (2015) Loss of genetic diversity of *Jatropha curcas* L. through domestication: implications for its genetic improvement. *Crop Science* 55: 749–759.
- Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A, Kato M, Kawashima K, Minami C, Muraki A, Nakazaki N, Takahashi C, Nakayama S, Kishida Y, Kohara M, Yamada M, Tsuruoka H, Sasamoto S, Tabata S, Aizu T, Toyoda A, Shin-i T, Minakuchi Y, Kohara Y, Fujiyama A, Tsuchimoto S, Kajiyama S, Makigano E, Ohmido N, Shibagaki N, Cartagena JA, Wada N, Kohinata T, Atefeh A, Yuasa S, Matsunaga S and Fukui K (2011) Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Research* 18: 65–76. doi: 10.1093/dnares/dsq030.
- Satyawan D and Tasma M (2011) Genetic diversity analysis of *Jatropha curcas* provenances using randomly amplified polymorphic DNA markers. *Journal Agrobiogen* 7: 47–55.
- Silitonga AS, Atabania AE, Mahlia TMI, Masjukia HH, Badruddina IA and Mekhilefe S (2011) A review on prospect of *Jatropha curcas* for biodiesel in Indonesia. *Renewable & Sustainable Energy Reviews* 15: 3733–3756. doi: 10.1016/j.rser.2011.07.011.
- Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.
- Tanya P, Taeprayoon P, Hadkan Y and Srinives P (2011) Genetic diversity among *Jatropha* and *Jatropha*-related species based on ISSR markers. *Plant Molecular Biology Reporter* 29: 252–264. doi: 10.1007/s11105-010-0220-2.
- Wang J, Chu S, Zhang H, Zhu Y, Cheng H and Yu D (2016) Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Scientific Reports* 6: 20728. doi: 10.1038/srep20128.
- Wen M, Wang H, Xia Z, Zou M, Lu C and Wang W (2010) Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha curcas* L. *BMC Research Notes* 3: 42. doi: 10.1186/1756-0500-3-42.
- Wright S (1978) *Evolution and the Genetics of Populations: Variability within and among Natural Populations*. Chicago: University of Chicago Press.
- Wu P, Zhou C, Cheng S, Wu Z, Lu W, Han J, Chen Y, Chen Y, Ni P, Wang Y, Xu X, Huang Y, Song C, Wang Z, Shi N, Zhang X, Fang X, Yang Q, Jiang H, Chen Y, Li M, Wang Y, Chen F, Wang J and Wu G (2015) Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant Journal* 81: 810–821. doi: 10.1111/tpj.12761.
- Yi C, Zhang S, Liu X, Bui HTN and Hong Y (2010) Does epigenetic polymorphism contribute to phenotypic variances in *Jatropha curcas* L.? *BMC Plant Biology* 10: 259. doi: 10.1186/1471-2229-10-259.