Doctoral Dissertation

Genetic analyses of the biofuel crop *Jatropha curcas* L. by molecular markers

Haiyan Li

December 2017

Laboratory of Dynamic Cell Biology
Department of Biotechnology
Graduate School of Engineering
OSAKA UNIVERSITY

Contents

List of abbreviations

EST-SSR	Expressed sequence tag- Simple sequence repeat
F	kinship coefficient
$\boldsymbol{F_{IS}}$	Inbreeding coefficient
FLK	Flanking sequence
$\boldsymbol{F_{st}}$	Fixation index
GWAS	Genome wide association studies
$\boldsymbol{H_E}$	Expected heterozygosity
$\boldsymbol{H_O}$	Observed heterozygosity
LTR	Long terminal repeat
MAS	Marker assisted selection
MTA	Marker trait association
QTL	Quantitative trait locus
RBIP	Retrotransposon-based insertion polymorphism
RTN	Retrotransposon
SSR	Simple sequence repeat

Chapter 1

General introduction

Climate change has driven a worldwide recognition of the importance of biofuels. Biofuels are derived directly from plants or plant-derived materials. They are renewable and carbon neutral, thus considered as clean and sustainable energy sources to substitute the fossil fuel.

Energy crops such as sugarcane, cassava, sweet sorghum, corn and soybeans are plants with high energy content that can be grown specifically as biomass feedstock. Corn and soybean are the major feedstock for biodiesel production in USA nowadays (Elbehri et al. 2013). However, using these crops, vast amounts of land and fresh water would be needed to produce enough oil to completely replace fossil fuel usage, and the cultivation will result in competition with the crop cultivation (The royal society 2008). It is reported that the diversion of food crops to biofuel production would cause food crisis problems (Tenenbaum 2008). Growing perennial plants for biofuel production on arid or semi-arid lands is preferable compared with growing annual crops, like ordinary corn, sugarcane, or soybeans, on arable lands, because it can minimize initial carbon debts incurred by placing native habitats under cultivation (Fargione et al. 2008; Searchinger et al. 2008). Besides, wide cultivation for biofuel on degraded and/or fallow lands could increase the financial income for native people, especially women farmers (Ouattara et al. 2014). These situations make drought-resistant perennial biofuel crops

becoming more and more important.

1.1 *Jatropha curcas* L.: a promising source of biofuel production

Jatropha curcas L., commonly known as Jatropha or physic nut, is a non-edible, oil-bearing shrub of the family Euphorbiaceae. Jatropha is a semi-domesticated plant, and wild Jatropha seems to be almost extinct. Phorbol esters in the plant are the principal chemical compounds that cause toxicity. Jatropha oil ranges from 30% to 50% of its seed weight, and can be easily converted to biodiesel with US and European standards (Azam 2005), thus has been projected as a promising source of biofuel. Jatropha has been traditionally used as hedges or medicinal plants, and is now widely distributed in tropical and semi-tropical zones of the world (Heller 1996). Moreover, it is believed that Jatropha can be grown on low-rainfall areas with poor soil (Yue et al. 2013).

Recently, many investments were made for commercial cultivations of Jatropha in African and Asian countries (Fairless 2007; Von Maltitz et al. 2013). However, all of the trials were ended up with failures by limited productivity there (Sanderson 2009; Von Maltitz et al. 2014). Low genetic variability in African and Asian Jatropha might have restricted their phenotypic performance, thus lead to the failures in breeding and

cultivation (Tatikonda et al. 2009; Divakara et al. 2010; Montes-Osorio et al. 2014; Montes et al. 2014).

Improved productivity of Jatropha oil can be achieved by increasing seed yield, which is influenced by the number of inflorescences and female flowers of Jatropha plant, as well as increasing the fruit yield, seed size and so on. Since 2008, researchers have started breeding programs of Jatropha to improve the productivity, but they remain in infancy and genetic analysis of Jatropha used for the basis of yield improvement is still limited (Divakara et al. 2010). In order to increase the yield of Jatropha and therefore meet the increasing demand of biofuel worldwide, efforts to increase the genetic basis of Jatropha, especially that of yield, are required.

1.2 Genetic diversity and characterization of the center of origin of Jatropha

The conservation and use of plant genetic resources are essential to the sustainable maintenance and improvement of agricultural production. An assessment of the genetic diversity is crucial for breeders in many ways: to understand the evolutionary and genetic relationships among different genotypes, to select germplasms in systemic and effective ways, and to develop strategies to incorporate useful diversity in their breeding programs. Generally speaking, the high genetic variation tends to exist

in the center of origin and domestication. Diverse Jatropha varieties from the origin will be valuable resources for improving African and Asian Jatropha with low genetic diversity, and then generating elite cultivars.

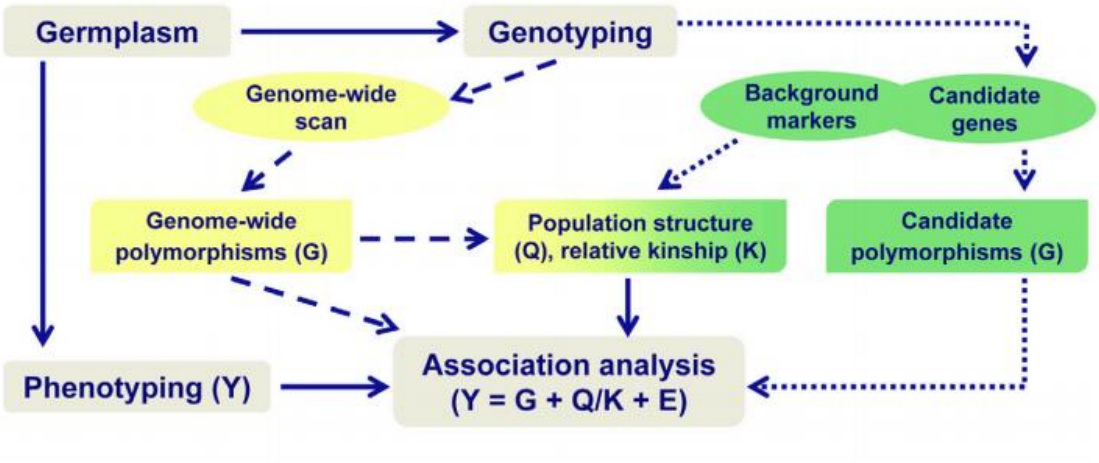
Characterization of the center of origin for Jatropha by evaluating the genetic variation has long been a hot topic among researchers. It has been widely accepted that Jatropha has its center of origin in Mesoamerica (Heller 1996). Although researchers keep on narrowing the center of origin of Jatropha, there is no consensus on this issue mainly because of lacking the comprehensive genetic study. Low genetic diversity of African and Asian Jatropha was reported in many studies, while the study on genetic and historical reasons to explain that phenomenon is still not enough (Montes-Osorio et al. 2014).

1.3 Association study: a powerful tool for understanding genetic basis of yield of Jatropha

In plants, QTLs or functional genes have been mapped by bi-parental crosses, but they were restricted in allelic diversity and limited genomic resolution. Association study directly examines many common genetic variants in different individuals to see if any variant revealed by molecular markers is associated with a trait. That is to say, association study aims to check if the genetic change at any loci can change the trait

performance. Scanning for responsible DNA regions on a genome wide scale using genome-wide molecular markers is called genome-wide association study (GWAS, Figure1.1, Zhu et al. 2008), and it has been applied in many crops and plants (Sharma and Chauhan 2012; Yang et al. 2015; Zhang et al. 2016). Association study overcomes the limitation of traditional gene mapping or QTL mapping by (i) providing higher resolution, often to the gene level, and (ii) directly examining the statistical association between genetic markers and phenotypes in a broader germplasm. It enables us to analyze genes or genomic regions responsible for complex traits, and has become more and more popular recently.

An important drawback of the association study is that false positive marker-trait associations due to population structure and familial relatedness could arise if unaccounted for (Yu et al. 2006). Consequently, association studies in plants will employ statistical approaches that take population structure and familial relatedness into account. Kinship matrix is generated to show the population relatedness. Analysis model for association study that incorporated kinship matrix as a fixed effect in the model is call mixed liner model (MLM). Statistical power was gained by using markers that not in linkage disequilibrium. Therefore, kinship matrix between individuals and linkage disequilibrium between markers would be indispensible in the association study.



(Zhu et al. 2008)

Figure 1.1 Schematic diagram of genome-wide association study. The inclusion of population structure (Q), relative kinship (K), or both in final association analysis depends on the genetic relationship of the association mapping panel and the divergence of the trait examined. E stands for residual variance. Genome-wide polymorphisms and candidate polymorphisms are genetic polymorphisms detected by molecular markers.

Selecting suitable breeding materials is a step that limits the speed of breeding. The marker loci that are tightly linked to major genes can be used for selection, and marker-assisted selection (MAS) is sometimes more efficient than direct selection by target phenotypes. Phenotypes are often controlled by two or more genes, and the phenotypic values are also modified by environmental factors. Association study has been widely applied in plant species, to provide markers for MAS. Identification of a specific gene by selection with markers, and then transfer it to the specific line can lead to improvement of the trait (Si et al. 2016).

1.4 Simple sequence repeat (SSR) and Retrotransposon-Based

Insertion Polymorphism (RBIP) markers

Molecular markers are the biological features that are determined by allelic forms of genes or genetic loci. They can be transmitted from one generation to another, and thus can be used as experimental probes or tags to keep track of an individual, a tissue, a cell, a nucleus, a chromosome or a gene. Selection of appropriate and reliable molecular markers is the most initial and important step for successful diversity analysis, association study, and MAS (Jiang 2013).

SSRs, also called microsatellites, are used for polymerase chain reaction (PCR)-based markers. SSR markers are characterized by their hyper-variability, reproducibility, co-dominant nature, locus-specificity, and random genome-wide distribution. As shown in Figure 1.2, the repeat number of each SSR can vary among individuals, which is one of the sources of the genetic polymorphism in plants. Primers specific for conserved DNA sequences surrounding the core sequence are designed to use in the PCR reaction.

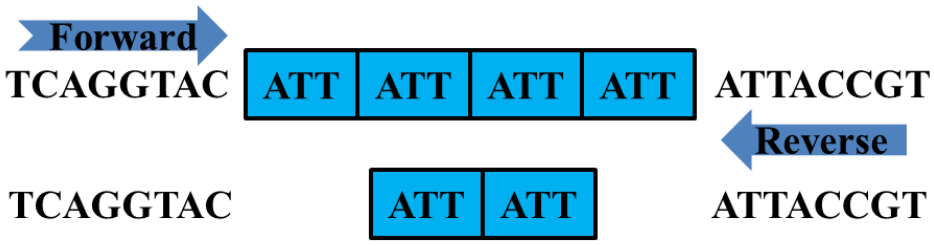
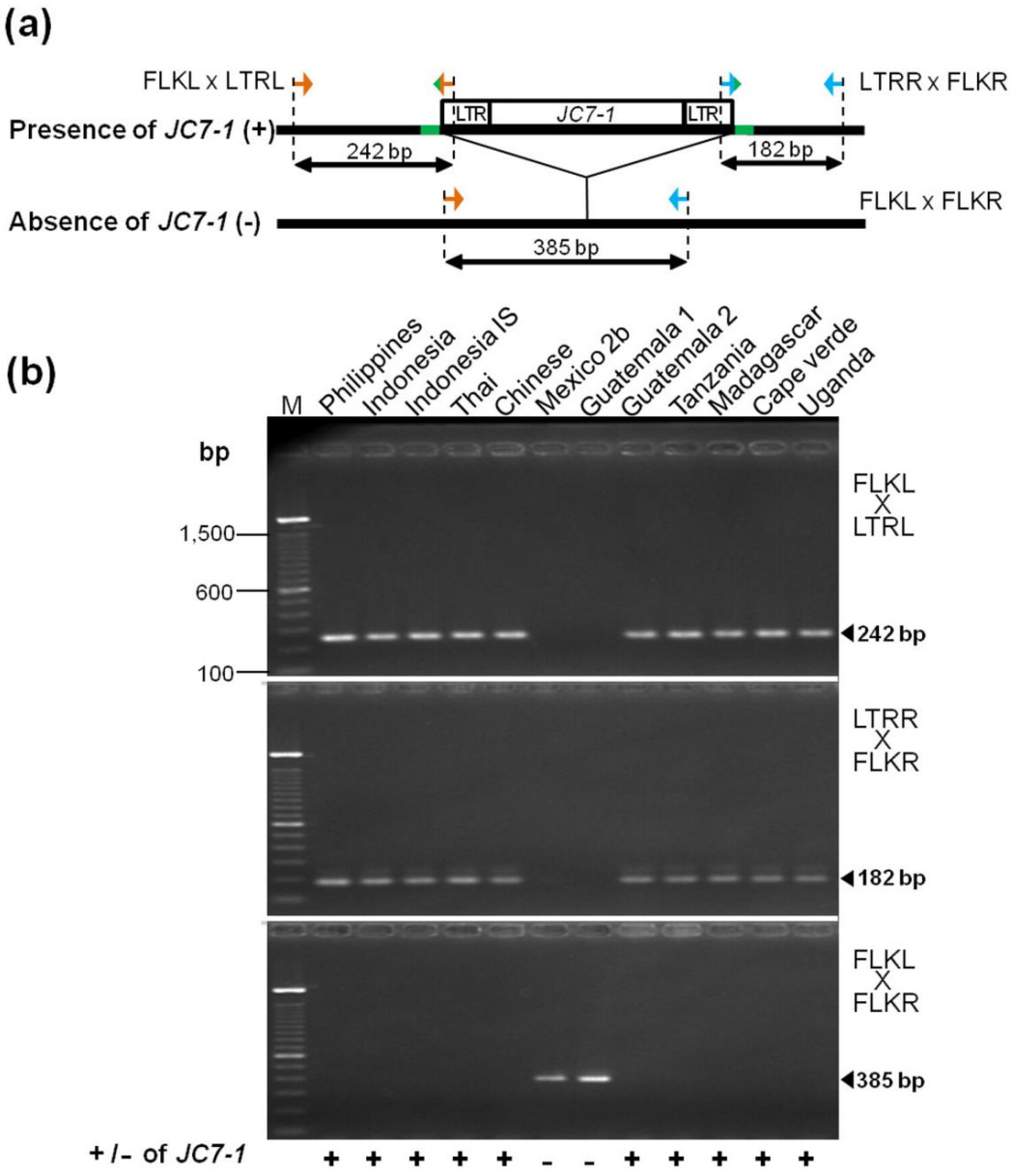


Figure 1.2 An example to show the principle of SSR. Forward and reverse mean forward and reverse primers, respectively.

The whole genome sequencing project of Jatropha has been completed by using a combination of the conventional Sanger method and new-generation multiplex sequencing methods (Sato et al. 2011; Hirakawa et al. 2012). The total length of the non-redundant sequences obtained was 285,858,490 bp consisting of 120,586 contigs and 29,831 singlets. Based on these sequences, SSRs were designed for the marker analysis in this study.

LTR (long terminal repeat) retrotransposons are ubiquitous in plant genomes. They include copia-type, gypsy-type, and other types, which are classified by similarity of sequences, as well as order of the genes in retrotransposons. The presence and absence of retrotransposons at genomic locations are sometimes different among lines, which depend on when and in which line the retrotransposition took place. Retrotransposition is irreversible because new copies are produced via transcription, not excision. For their ubiquity, diversity and irreversibility, retrotransposons have a great

potential to be genetic markers for the plant genome, especially in biodiversity and phylogenetic analyses. Using primers derived from LTR and its flanking DNA region to detect individual retrotransposon insertions by PCR, different allelic status at a locus can be revealed (Figure 1.3). The first report that used LTR retrotransposons as potential DNA markers, which is called retrotransposon-based insertion polymorphism (RBIP) markers, to examine the diversity in Jatropha was done by Alipour et al. (2013).



(Alipour et al. 2013)

Figure 1.3 PCR detection of the retrotransposon insertion polymorphism.

(a) Schematic illustration of RBIP primer design. Arrows indicate the position and direction of the primers, and target site insertion is represented in green. FLKL X LTRL: A primer set corresponding to the left flanking sequence (forward) and left LTR (reverse); LTRR X FLKR: A primer set corresponding to the right LTR (forward) and the right flanking sequence (reverse); FLKL X FLKR: A primer set corresponding to the left and the right flanking sequences. (b) Presence or absence of the JC7-1 restrotransposon marker in twelve Jatropha accessions. PCR products are indicated by arrowheads.

1.5 Objective and overview of the study

The objective of the study is to extend the genetic basis for molecular breeding of Jatropha by exploring wide genetic resources from its center of origin, and to uncover molecular markers that associate with the yield. The information could be directly used in the breeding program, and finally would help us to fight against the climate change.

The present research is composed of 4 chapters. Firstly, general introduction is presented in Chapter 1. In Chapter 2, the worldwide genetic distribution of Jatropha was evaluated, later the center of origin and the voyage from the origin to the world were identified, by employing 59 SSR markers and eight RBIP markers, as well as a worldwide collection of 246 Jatropha accessions from Mesoamerica, Africa and Asia. Results in Chapter 2 are important to select genetic resources to improve Jatropha. In Chapter 3, association study in Jatropha was conducted and molecular markers that associated with yield-related traits were characterized, by combining 16 yield-related traits of 110 Mexican accessions and their genotypic data. Finally, the entire research works and future prospect are summarized and discussed in Chapter 4.

Chapter 2

Worldwide genetic characterization reveals the center of origin for Jatropha

2.1 Introduction

Jatropha (*Jatropha curcas* L.) is generally believed to have an origin in Mesoamerica (Ambrosi et al. 2010; Lin et al. 2010; Parawira 2010; Umamaheswari et al. 2010; Montes et al. 2014). Within Mesoamerica, Mexico is the most probable candidate where the origin of Jatropha situates. Not only does Mexico have toxic Jatropha varieties like in other countries, but also it has non-toxic ones unique to this country (Dias et al. 2012). Basha et al. (2009) demonstrated that, among 72 accessions of Jatropha from 13 countries, the Mexican accessions are unique in terms of phorbol ester content (absent or trace presence) and of molecular profile (characterized with RAPD and ISSR markers). Still, the most primitive species of the genus *Jatropha* and related taxa also occur in Mexico (Dias et al. 2012).

The possible center of origin for Jatropha has been recently narrowed down to the state of Chiapas, the southern-most state of Mexico that bordering Guatemala, which has the highest genetic variation of Jaropha among Mexican states (Pecina-Quintero et al. 2011 and 2014). In these reports, the genetic variation of Jatropha was examined by using amplified fragment length polymorphism (AFLP) markers. A further confirmation of this conclusion is necessary by using more reliable co-dominant markers. To support the idea that Chiapas is the origin, it was in this state that wild-like Jaropha trees were

observed recently (Ando, personal communication). The state of Chiapas, which contains nine sub-regions, has complicated geographic and climatic conditions, and it is separated into different geographic and biological ecological regions by highland and mountains (Figure 2.1). Genetic and phenotypic distribution of Jatropha within Chiapas remains unclear. Since accurate and precise regional genetic analysis in Chiapas is expected to narrow the center of origin to a smaller area and thus help breeders to identify elite breeding materials quickly and efficiently, a detailed identification of the genetic distribution within Chiapas is quite necessary.

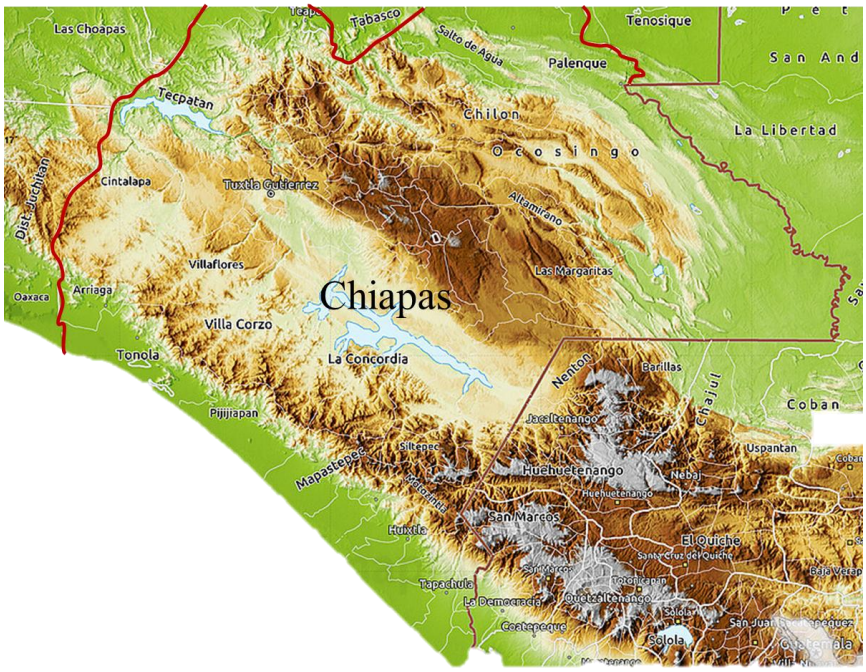


Figure 2.1 A map of the state of Chiapas, Mexico. URL: http://www.maphill.com/mexico/chiapas/maps/physical-map/.

It is presumed that, during or after the Age of Explorations, Portuguese navigators brought Jatropha from Mesoamerica to the Cape Verde islands in the Atlantic Ocean, and later it was brought to Africa and then Asia, resulting in a wide distribution of Jatropha all over the Old World (Heller 1996). It is reasonable to speculate that this historical fact might result in the low genetic diversity in Africa and Asia nowadays. To improve the productivity of African and Asian Jatropha, to clarify a series of issues, such as from where the Mesoamerican Jatropha were brought to Africa and Asia, or the reason for the narrow genetic base of African and Asian Jatropha, would be as important as to characterize the center of origin. Hybridization of local Jatropha in Africa and Asia with genetically-distinct Jatropha from the center of origin would be one of the accessible approaches. This approach is expected to breed elite varieties and therefore solve the low productivity in African and Asian countries. As African and Asian Jatropha was transmitted from Mesoamerica, thus to identify Mesoamerican genetic resources genetically distinct from African and Asian Jatropha, it requires characterization of the ancestral genotype of African and Asian Jatropha in Mesoamerica.

In this chapter, 59 SSR markers and eight RBIP markers were applied to assess the genetic diversity of 246 Jatropha accessions from countries of Africa and Asia,

Guatemala and eight Mexican states (Chiapas, Guerrero, Michoacan, Morelos, Oaxaca, Tabasco, Veracruz and Yucatan). Genetic variability of Jatropha accessions of each of nine regions in Chiapas was evaluated and the center of origin for Jatropha was identified. Mesoamerican accessions that were likely to share the ancestors with African and Asian accessions were characterized, which suggested the route of the voyage of Jatropha from Chiapas to the Old World. Human selection causing the low genetic basis in Africa and Asia was also discussed.

2.2 Materials and Methods

Workflow of the analysis in Chapter 2 was shown in Figure 2.2. After genomic DNA was extracted from leaves provided by collaborators, genotyping and scoring were performed. To identify the center of origin of Jatropha, the genetic variation, as well as the phenotypic variation characterized by Mexican collaborators, was used. To characterize the transmission route of African and Asian Jatropha from Mesoamerica, Mesoamerican accessions that were the genetically closest to African and Asian accessions were identified by population structure analysis, phylogenetic tree construction and principle coordinate analysis. Finally, by integrating with the number of retrotransposons, the transmission route from Mesoamerica to Africa and Asia was identified.

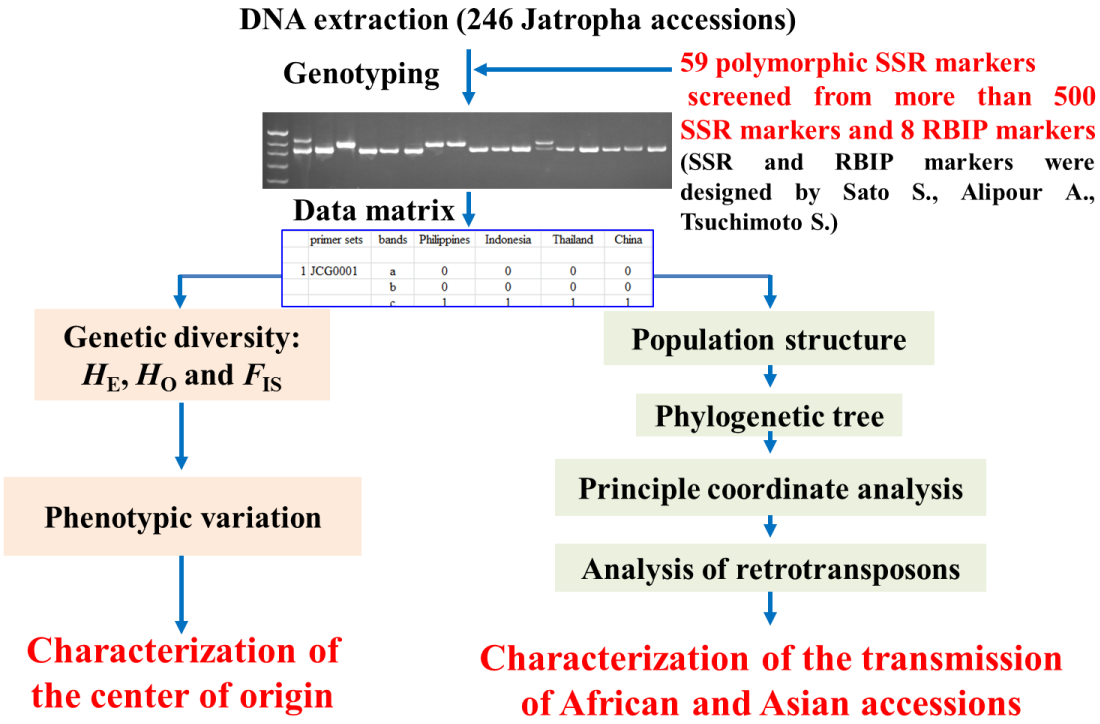


Figure 2.2 Workflow of the analysis conducted in Chapter 2.

2.2.1 Plant materials

A worldwide collection of 246 Jatropha accessions in this study was consisted of 207 accessions from Mesoamerica (198 from Mexico and nine from Guatemala), seven from Africa, and 32 from Asia (Table 2.1). The Mexican accessions and seven Guatemalan accessions were obtained from the collection of the Instituto Nacional de Investigaciones Forestales, Agropecuarias (INIFAP, Mexico). INIFAP provided leaves the accessions under the academic exchange agreement with Tottori University. Seven Philippine accessions originating in four provinces of Luzon and Mindanao islands were

obtained from the collection of University of the Philippines, Los Baños (UPLB) under MOA with Osaka University. Twenty-one Vietnamese accessions from plants grown in Quang Tri Province were provided by Prof. Yoshiaki Kitaya of Osaka Prefecture University. Sudanese and Egyptian accessions were obtained from the University of Khartoum and University of Sadat City, respectively, under the collaboration. Other accessions were the same as those used in the previous study in our laboratory by Sato et al. (2011).

Table 2.1 Sources of the Jatropha accessions

Sources				Sample number
Continent	Country	State	Region	
America	Mexico	Chiapas	Centro	27
			Altos	5
			Fronteriza	14
			Frailesca	15
			Sierra	13
			Norte	2
			Selva	3
			Soconusco	34
			Istmo-Costa	5
		Guerrero		11
		Michoacan		9
		Morelos		10
		Oaxaca		9
		Tabasco		7
		Veracruz		26
		Yucatan		8
	Guatemala			9
Africa	Cape Verde			1
	Egypt			1
	Madagascar			1
	Sudan			2
	Tanzania			1
	Uganda			1
Asia	China			1
	Indonesia			2
	Philippines			7
	Thailand			1
	Vietnam			21

2.2.2 DNA markers

More than 500 SSR markers were developed by Dr. Shusei Sato by designing primer pairs surrounding SSR sequences identified from the genomic database of an Asian accession (Sato et al. 2011; Hirakawa et al. 2012). Eight RBIP markers employed in this study were from members of the copia-type families identified in the genomic database by Alipour et al. (2013). Primers were designed for flanking sequences as well as for LTRs (Alipour et al, 2013; Tsuchimoto, personal communication). Retrotransposon insertion at each locus was observed by combining primers designed from the flanking sequences at both sides and primers designed from LTR sequences.

2.2.3 DNA extraction

Genomic DNA was extracted from 100 mg Jatropha leaves with the DNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer's instructions. The DNA samples were diluted to a final concentration of 0.35 ng/μl and stored at −20℃ until use.

2.2.4 Genotyping and scoring

PCR for SSR markers was performed in a 10-μl total volume solution containing 1.4 ng of DNA template, 0.16 μl of SSR primer mix (25 μM each), 1× PCR buffer, 0.03 μM $MgCl_2$, 0.8 μl dNTPs (2.5 mM), 0.08 U BIOTAQ polymerase (Bioline, UK, 5

units/μL), and Milli Q water. Touch-down amplification was performed as follows: 3 min hold at 94 ℃, followed by 3 cycles at 94 ℃ for 30 s, and 68 ℃ for 30 s, with the annealing temperature reduced by 2 ℃ until 64 ℃ every 3 cycles. Continuing with 3 cycles at 94 ℃ for 30 s, 62 ℃ for 30 s, and 72 ℃ for 30 s, the annealing temperature was reduced by 2 ℃ until 58 ℃ every 3 cycles. A further 30 cycles of amplification was performed at 94 ℃ for 30 s, 55 ℃ for 30 s, and 72 ℃ for 30 s. The PCR products were stored at −20 ℃ until separation by polyacrylamide gel electrophoresis (PAGE). Amplified bands were stained with ethidium bromide. For markers that showed ambiguous bands, experiments were repeated by changing experimental conditions until clear bands were observed.

For RBIP markers, PCR was performed in a 5 μl total volume solution containing 0.7 ng of DNA template, 0.08 μl of primer mix (50 μM each), 1× PCR buffer, 0.03 μM $MgCl_2$, 0.4 μl dNTPs (2.5mM), and 0.08 U BIOTAQ polymerase (Bioline, UK, 5 units/μL), made to the volume with Milli Q water. PCR reaction was started with denaturation at 94 ℃ for 2 min, followed by 35 cycles of 45 s at 94 ℃, 45 s at 55 ℃ and 2 min at 72 ℃, continued with a final elongation step at 72 ℃ for 10 minutes. Amplified bands were stained with ethidium bromide. The PCR products were stored and analyzed in the same way as the SSR markers.

Screening of SSR markers was conducted by using representative Mexican accessions in which high polymorphism were expected. SSR markers that showed high polymorphism were chosen to amplify all accessions. For all the SSR markers and RBIP markers, all bands were individually scored (presence 1, absence 0). All data analyses were performed based on the genotypic data matrix.

2.2 5 Number of retrotransposons

Presence or absence of each retrotransposon was examined by amplifying bands using primers for left and right flanking sequences, which should show no amplification if there is a retrotransposon (Fig 1.3). It was also examined by using a primer of one of flanking sequences and one of LTR primers, which show amplification if there is a retrotransposon (Fig 1.3). According to the results, the numbers of retrotransposons present in each accession was counted.

2.2.6 Genetic heterozygosity

Expected heterozygosity (H_E) and observed heterozygosity (H_O) show the portion of heterozygotes in populations. They are measures of the extent of genetic variation in a population. H_E is heterozygosity under Hardy-Weinberg equilibrium (HWE) by random mating of individuals (Nei 1987). The basic Hardy-Weinberg is $p^2 + 2\,pq + q^2 = 1$, p and q are the frequencies of their corresponding allele, where the heterozygosity is

given by $2pq$ and ($p^2 + q^2$) is the homozygosity. Both statistics were calculated with GenALEx 6.5 (Peakall and Smouse 2006). Comparisons of the significances of H_E and H_O among populations were performed using Kruskal-Wallis one-way analysis of variance. Differences between H_E and H_O are often caused by inbreeding as a result of selection or the small population size. The larger differences between H_E and H_O, the higher extent of inbreeding there is in a population. Inbreeding coefficient (F_{IS}) is a measure of the degree of inbreeding in a population, it equals to (H_E and H_O)/ H_E. F_{IS} and its P values were performed with FSTAT version 2.9.3.2 (Goudet 2001).

2.2.7 Population structure

Model-based clustering analysis was performed using Mesoamerican accessions or all accessions (including African and Asian accessions) with STRUCTURE 2.3.4 (Pritchard et al. 2000). This analysis was conducted to know the optical population number (K) based on the genetic differentiation among individuals. The analysis was performed using the Markov Chain Monte Carlo (MCMC) method based on 1×10^5 iterations followed by a burn-in period of 5×10^4 iterations. To determine the optimal number of K to show the differentiation in the population at the biggest extent, firstly the K was set to vary from 1 to 10, then after 10 independent runs for each K, ΔK was calculated based on the rate of change of the log likelihood for K values between

successive K values by following Evanno et al. (2005). Finally the optical K value, the ΔK of which have a clear peak, was chosen.

Principal coordinates analysis (PCoA analysis) of all individuals was performed with GenALEx 6.5. The final results were shown as a three-dimensional plot (X, Y and Z represent the three principal coordinates) of all the accessions.

2.2.8 Phylogenetic relationship

A neighbor-joining (NJ) tree of individuals was constructed on the basis of genetic distances (Nei et al. 1983) with POPULATION 1.2.32 (Langella 2002). The tree was drawn with Treeview 1.6.6 (Page 1996) and Figtree 1.4.2 (Rambaut and Drummond 2014). Based on the genetic information revealed by the markers, genetic distances (Nei's DA distance) between groups were also calculated with POPULATION 1.2.32.

2.2.9 Phenotypic variation

Four representative yield-related traits (the number of inflorescences per plant, number of female flowers per plant, ratio between female and male flowers, seed yield per plant) of 110 accessions from seven Mexican states [Chiapas (61), Guerrero (5), Michoacan (5), Morelos (4), Oaxaca (7), Veracruz (23) and Yucatan (5)] were measured individually at the Rosario Izapa Experimental Farm, INIFAP (Chiapas, Mexico).

INIFAP offered these phenotypic data under the academic exchange agreement with Tottori University, Japan. Chiapas accessions were further sub-classified into those from six regions [Centro (15), Frailesca (8), Fronteriza (4), Sierra (4), Selva (2) and Soconusco (28)]. To compare the variation among accessions from central Chiapas (Centro, Fronteriza, Frailesca, and Sierra), peripheral areas of Chiapas (Selva and Soconusco), and other Mexican states, the mean value and standard deviation (SD) of the four traits in each area were calculated.

2.3 Results

2.3.1 Genetic diversity in Mesoamerica

Screening of SSR markers (>500) was conducted by using representative Mexican accessions in which high polymorphism were expected. Among all the SSR markers, 59 SSR markers which showed clear polymorphism were selected for further PCR amplification of all the accessions.

Eight RBIP markers (Alipour et al. 2013; Tsuchimoto, personal communication) were used in the study. Among the eight RBIP markers, six full-length members with right and left LTRs showing 100% sequence identity with each other, and two solo LTRs displaying 100% sequence identity with the consensus sequence of the families (Alipour et al. 2013; Tsuchimoto, personal communication). All of the retrotransposons

in this study were expected to have retrotransposed recently. Compared with old retrotrasposons, newly-retrotransposed ones have a better ability to differentiate Jatropha accessions. . Example of SSR and RBIP markers that show polymorphism among Jatropha accessions was presented in Figure 2.3. Primers for polymorphic SSR and RBIP markers used in the study were listed in Table 2.2 and 2.3.

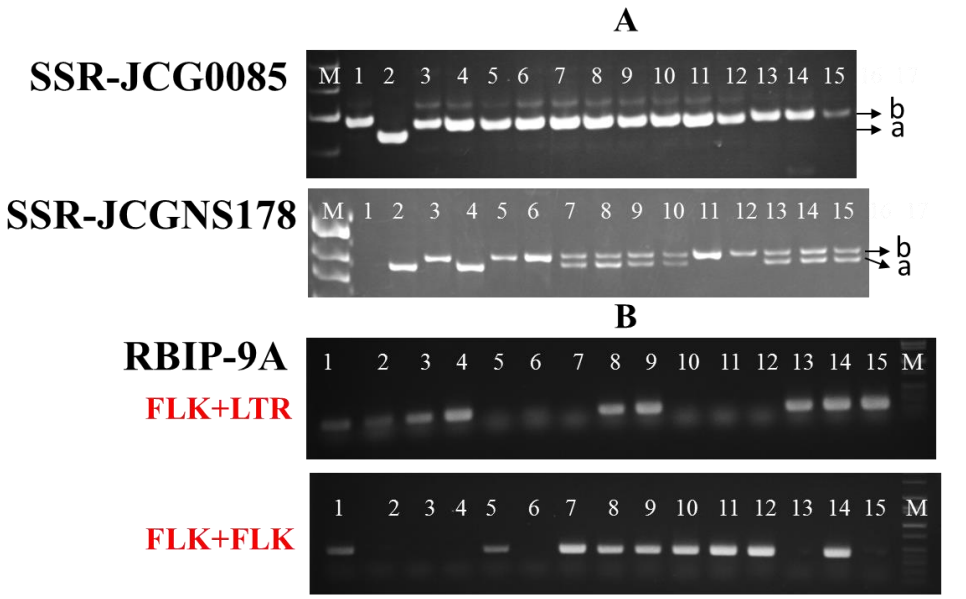
Table 2.2 List of the SSR markers and primers used in the study

SSR marker	Forward primer (5´-3´)	Reverse primer (5´-3´)
JCG0001	AGCACACCACCATTAAAGCC	GGCTTTGCATAACACCATCA
JCG0007	AATCGAAGAGCAGGTGCAGT	GCAAAATCAAGCCAAATCGT
JCG0009	TGGGCGATTGAGCTTACTTT	CCCTTCAATGGGTTCTTTCA
JCG0027	ATGATACGAGGACGAATGGG	GAATGTGGCTGCAGGGTATT
JCG0028	AGGTAACCAACAGTGACCCCT	CCGTCCCCTTTTATTGGATT
JCG0034	AGGTTGCACAGATACGGGAC	GCACAGTGACAAATTGTGCTT
JCG0038	GCATGCATTTTGGTTGTGAG	TGCAATGTAAGCCCATTCAA
JCG0040	TTGGGCTATAGTACCGCAGG	TTGAACTGAGGTGGAAACCC
JCG0041	TCCGTTGTTTGGTTGCAATA	GATGGTGGTTGCTCCAATCT
JCG0050	TCCTCGCCTTCAAGTGGTAA	GCTGCAGCAAGCACAATAAA
JCG0054	TCGACAAAACCCCAATTATCA	TCAGTTCACCCAACAGTCCA
JCG0057	AGCTTGCCTGAAAAACCAGA	TTTCAGCCACTACCTCAGCC
JCG0061	TCACTCTTCTGTCTCGCTTCC	CGCAAAATTTGTGTTGGAAT
JCG0063	CTTTGATGAGGCTTCCCTTG	CATCGCATCATCCAGTTGTC
JCG0066	TTTCCTGGTGTTTTGCATTG	TTCATCCCTTTCTGGTGGAG
JCG0085	CTCCCTTGTTCTTTGCTTCG	CAGCATATGACACCTGCACC
JCG0087	TGCTAAGGTCCCTCTTGGAA	TGAAGATGATGGGAAAAGGG
JcGNS101	CCTTTCTGTCATCTCCGAGG	CACGACACCATAAGCACTGG
JcGNS105	GACCCTCTGGCTGAAGTTTG	GCTGTCTGCAGCTCTGAATG
JcGNS106	CCTTCCCTCTTCCTCCAATC	CCTTCTCCTATACGCCATCG
JcGNS0109	TCGATGTCAGTACAGGCTGC	TTCTCCCTCTGTATGGTCGC
JcGNS0112	GGAAGACTGGAACTGCAAGC	TAGCCTACTGCAACAAGCCC
JcGNS0113	GGAGCGTAAGGATCTGCAAC	CTCTCTAAACCCTTGCAGGC
JcGNS0117	CTGCTGTCTCTGCTTGGTTG	GCACATTCACCCTCTCCTTC
JcGNS0126	GTGCCTGAAGTAAGGGCAAG	TCAGTGAGAGGCACAGTTGC
JcGNS0127	ATCCGTTCGACCTACACTGG	TTCACTAAGCCTCCGACCAC
JcGNS0131	GAAAGAGGCGAACAGTAGCG	GTTTCTTACGTCGCGTCCTC
JcGNS0134	TGGGTTTCAGCAGAATAGGG	TCTGCAGAGGCATCTGTGTC
JcGNS0143	ATGGTGCCATTGCTGATACC	AACTTGCTCCTAGGGCTTCC
JcGNS0144	AGAAGCCCATTGATGGTGTC	TCCTGCGTACTAATCCCAGC
JcGNS0153	CTTGGAAAGAGCTTCCATGC	ATGGGTGACTCTTCGTCTGC
JcGNS0158	CTTCGCTCGCTTTCTTGTTC	GATAGCTGCGGTGGAAAGAG

JcGNS0165	GTTGAAAACCCACAAGGACG	GAGCGAATGGCTAGGTTACG
JcGNS0178	TCAACAACCGCAACAACAAC	GGAAGAAAGCCAGCAGTGTC
JcGNS0180	ATTGGTTGTTGTTGTTGGGC	GGGAATAGAGAAAGCCGGAG
JcGNS0182	AAATGAGCTTGCAGCCTTTG	GTCGGGGGCCTTAAACTTAC
JcGNS0183	CCATTTTCGCTGTGTTGATG	CTAACCACTTGGGCCAGTTC
JcGNS0187	ACAACCCCAATGAAATGAGC	TATAGAGCTGGCATACCGCC
JcGNS0189	ACCCTTTTTGCCTTGTCATC	ACCTTCCAGGGAGACATTCC
JcGNS0191	TGCTCATCATCAAGCAATCC	GAAGTGGGTCTCCTTCAACG
JcGNS0192	TTATCCAAGCAACCAAACGG	TAGCCGCTTCTCTCTCTTCG
JcGNS0193	TCATGATGAGAATCTTGCGG	CCAAGCGAGACAAAGGAGAG
JcGNS0196	TCTTGAATGGAGAAATGGGC	CCCCTCTCTATTTTCCCCAG
JcGNS0204	CATGGCAACCAAGAACATTG	TGGGATAGGCAGAGTCCAAC
JcGNS0214	CTTCTTCTGTCCAAGGGAGC	TTTGAGATGCCCAAGAGAGG
JcGNS0217	TATACCTGCCTCTATGGGCG	AAAGGGTAAGGTTGAACGGG
JcGNS0224	CGATGGACGTAAAGCTCACC	TTGCCCTTAAACTGCCTCAC
JcGNS0225	GTGACAGGGTATGTGCATGG	ATTGTGTTGGAGGCTGAAGG
JcGNS0228	GCTCAAAGCTCACCATCTCC	ATGCTATCCATTCTGCTGGC
JcGNS0233	GGGAGCCTGAGGAAAGAATC	GGCAAAAACCCAACTTAGCC
JcGNS0241	GGACGTGGTTCAGTCATGTG	ACTTCCATCCTTCGCCTTTC
JcGNS0245	CCAACGAGGACCGATGTTAG	AAGTTAACCGGTGATGTCGC
JcGNS0248	CTGCTCCTGATTCCTTCTGG	TTTCCAAGTCCACTTGCTCC
JcGNS0257	GGATCTGCCTTTGATTCAGC	TCACCCCTCTTTCACCAATC
JcGNS0261	TGCACTTGGACTTGGTCTTG	CGTTTCACCAGTTTGACGAG
JcGNS0270	ATGTGCGGTCCTTTACCTTG	ATCTACAACCATTGGGCAGC
JcGNS0279	AGATTTGTGGGTTGGTCCAG	TGCGGTATAGTTGCATCCAG
JcGNS0293	CCGAGAAGTACAAACAGCAGAC	AACCCTCGTCGAACGATATG
JcGNS0299	GAGTTCCAACTGGGAAATGG	CCGAACAGGAATTCAACCAC

Table 2.3 List of the RBIP markers and primers used in the study.

RBIP marker		LTR primer(5´-3´)		FLK primer(5´-3´)
5S	Left	GAGTGGGCACCGCACAACAA	Left	TGGAGAATTTGGGTTTGGTC
	Right	TTATTGCCGGGGCCTAACAC	Right	CTCGAGACCTCTCAACGAAC
8A	Left	GAAAATTAAATCCAACAATA	Left	GAACCAGGATCACGTTCAAC
	Right	TGAGATTAATTCTTACATAT	Right	TCGCCCCACTTACTTTCTTG
8B	Left	GAAAATTAAATCCAACATGT	Left	CAAAGCACACGAGGATTCAG
	Right	GAGATTAATTCTTAACAGAA	Right	CAGGTCCAAATCTCCTCGTG
9A	Left	TTTCTCTTCATCCGACAAAA	Left	AGATGCTGATAGGGTTGGTG
	Right	ATTTTTCCCTCTGTGAAACAG	Right	CAGCACGGCCTCGTTTATAG
9B	Left	TTTCTCTTCATCCGACATGG	Left	GTGGGATCTTGAAGGACCAG
	Right	TTTCCCTCTGTGAAACACCC	Right	TGTTGAGAAACATGGTCAAGC
9C	Left	TTTCTCTTCATCCGACAAAA	Left	TTGCCCAAATTTCACTTCATC
	Right	TTTCTCTCTGTAAAACATCT	Right	CCGAATTTTGAGCCAGCTTG
9D	Left	TTTCTCTTCATCCGACATTA	Left	CTTACTGACTTCATTAATTG
	Right	TTTCCCTCTGTGAAACAGAT	Right	CACCCACCCTCTTCTTCATC
11SA	Left	ACAACTAATCCTTAACAAAC	Left	CACTCCTAAAATGCGGCTAAC
	Right	CCCACATTAAATCTACAAGG	Right	TTCCACTGCTATTGTTTAATTCAT



C

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
JCG0085	a	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	b	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
JCGNS178	a	0	1	0	1	0	0	1	1	1	1	0	0	1	1	1
	b	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1
9A	FLK+LTR	1	1	1	1	0	0	0	1	1	0	0	0	1	1	1
	FLK+FLK	1	0	0	0	1	0	1	1	1	1	1	1	0	1	0

Figure 2.3 Example of polymorphic SSR (A), RBIP markers (B) and their scoring dataset (C). Fifteen accessions were shown. M: size markers. a and b: alleles.

To know the genetic variability worldwide, H_E and H_O of Mesoamerican, African and Asian accessions were calculated, respectively. The higher H_E and H_O there are, the higher genetic diversity the population has (see Materials and Methods). Significant high H_E and H_O were detected in Mesoamerican accessions ($P < 0.001$), while negligible H_E and H_O were observed in African and Asian accessions (Figure 2.4A). Because the center of genetic variation should have the highest genetic diversity. These results

indicated that the genetic-variation center of Jatropha should be in Mesoamerica, which is consistent with the previous conclusions (Dias et al. 2012). Mesoamerican accessions were then divided into those from Guatemala and eight Mexican states, and a comparison of H_E and H_O was conducted among them. The highest H_E and H_O were found in the state of Chiapas ($P < 0.001$), and interestingly, a clear decreasing cline of H_E from Chiapas and its bordering states (Tabasco, Veracruz and Oaxaca) to non-bordering states (Michoacan, Morelos, Yucatan and Guerrero) was observed (Figure 2.4B).

To further narrow down the candidate region for the center of genetic variation, as well as to characterize the genetic distribution in Chiapas, H_E and H_O were calculated in each of the nine regions of Chiapas. Five regions including Centro, Frailesca, Fronteriza, Sierra and Altos were grouped as central Chiapas, while other four regions, Norte, Selva, Soconusco and Istmo-Costa, were assigned as northern and southern peripheral areas of Chiapas. As a result, a significant higher H_E in central Chiapas than in peripheral areas of Chiapas was detected ($P < 0.001$, Figure 2.4C).

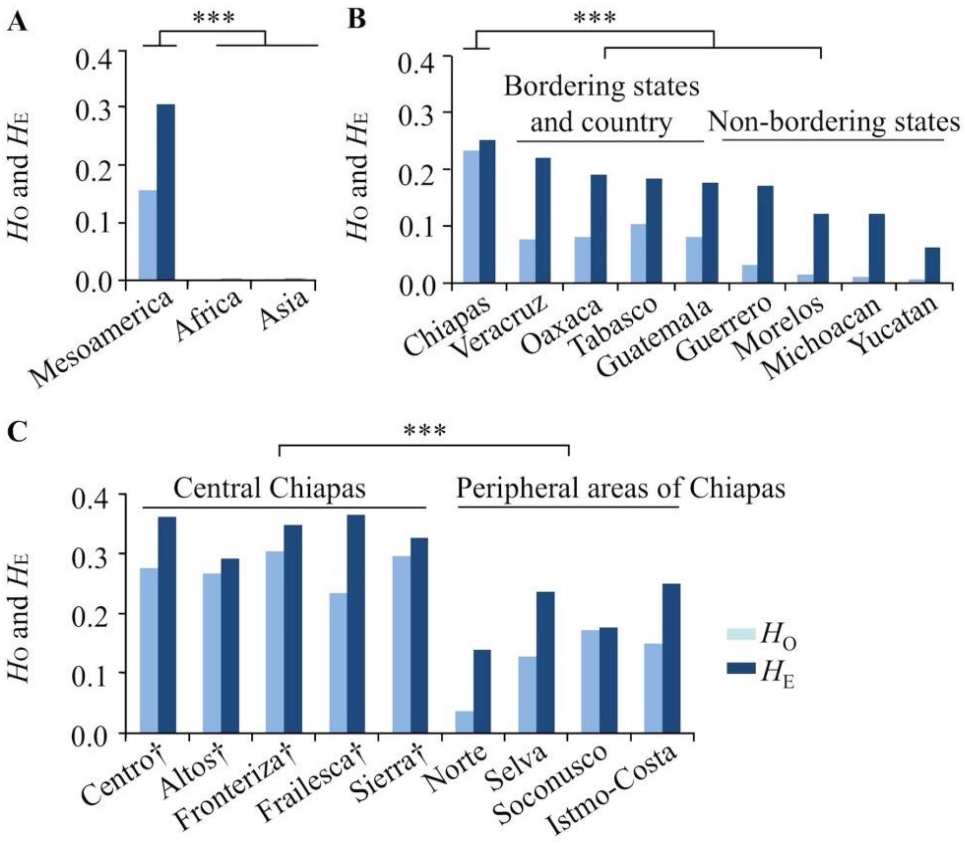


Figure 2.4 Genetic divergence of Jatropha in Mesoamerica. (A) Observed (H_O) and expected (H_E) heterozygosity in Mesoamerica, Africa, and Asia. (B) H_O and H_E in Mexican states and Guatemala. (C) H_O and H_E of regions in Chiapas states. Regions in central Chiapas covering the Chiapas Central Depression are shown by †. ***P<0.001.

Differences between H_E and H_O are often caused by inbreeding as a result of selection or the small population size. The larger differences between H_E and H_O, the higher extent of inbreeding there is in a population (see the Materials and Methods). Slight differences between H_E and H_O were observed among the accessions from central Chiapas, whereas large differences between these statistics were detected among the accessions from most of the peripheral areas of Chiapas, other Mexican states, and Guatemala (Figure 2.4 B, C). Higher F_{IS} were observed in the peripheral areas of Chiapas, except for Soconusco, other Mexican states, and Guatemala than in central Chiapas (Table 2.4), further indicating central Chiapas is the likely center of genetic variation. The cline of F_{IS} may be due to artificial selection and/or decrease of effective population size.

The 59 SSR markers yielded 221 polymorphic bands ranging from two to seven with an average of 3.75 alleles per marker. In Table 2.4, 161 alleles were specific to each Mexican state or Guatemala, whereas no allele was specific to African or Asian countries. The state of Chiapas had 47 specific alleles, which was the highest number among Mexican states and Guatemala. In Chiapas, 30 alleles were specific to the central area (Centro, Altos, Fronteriza, Frailesca, and Sierra), while only one allele was specific to the peripheral areas (Norte, Selva, Soconusco, and Istomo-Costa).

Table 2.4 F_{IS} and number of specific alleles of Jatropha from different geographical origins.

Origin	Sample number	F_{IS}	Number of specific alleles
Africa	7	-	0
Cape Verde	1	-	0
Egypt	1	-	0
Madagascar	1	-	0
Sudan	2	-	0
Tanzania	1	-	0
Asia	32	-0.020	0
China	1	-	0
Indonesia	2	-	0
Philippines	7	-	0
Thailand	1	-	0
Vietnam	21	-0.031	0
Mesoamerica	207	0.487***	161
Guatemala	9	0.582***	0
Mexico	198	0.464***	129
Chiapas	118	0.310***	47
Central Chiapas	74	0.280***	30
Centro†	27	0.253***	10
Altos†	5	0.205***	3
Fronteriza†	14	0.161***	0
Frailesca†	15	0.388***	2
Sierra†	13	0.131***	1
Peripheral area of Chiapas	44	0.251***	1
Norte	2	0.857***	0
Selva	3	0.600***	0
Soconusco	34	0.043***	0
Istmo-Costa	5	0.492***	1
Guerrero	11	0.824***	2
Michoacan	9	0.915***	0
Morelos	10	0.879***	0
Oaxaca	9	0.613***	1

Tabasco	7	0.498***	0
Veracruz	26	0.657***	23
Yucatan	8	0.894***	0

† Central Chiapas； *** $P < 0.001$

2.3.2 Genetic groups in Mesoamerica

The genetic constitutions of all Mesoamerican accessions were examined by population structure analysis, and Δ (see Materials and Methods) showed a clear peak when K equaled to three, indicating that there were three genetic groups in Mesoamerica. Mesoamerican accessions were classified into three genetic groups: A, B, and C (Figure 2.5). The numbers of accessions assigned to each group in Guatemala, Mexican states, and each Chiapas region are presented in Figure 2.6 A, B. The distribution of the groups revealed a distinct geographic cline. Accessions from central Chiapas were mostly in Group A (orange). Accessions from Group B (purple) were mainly distributed in the peripheral areas of Chiapas and in neighboring states and country, whereas accessions from Group C (blue) were mainly distributed in states distant from Chiapas (Guerrero, Michoacan, and Morelos). The highest genetic variation and the lowest F_{IS} among the three groups were observed in Group A (Table 2.5), which was as expected because Group A mainly overlapped with the central Chiapas where the highest genetic diversity and the lowest inbreeding were observed.

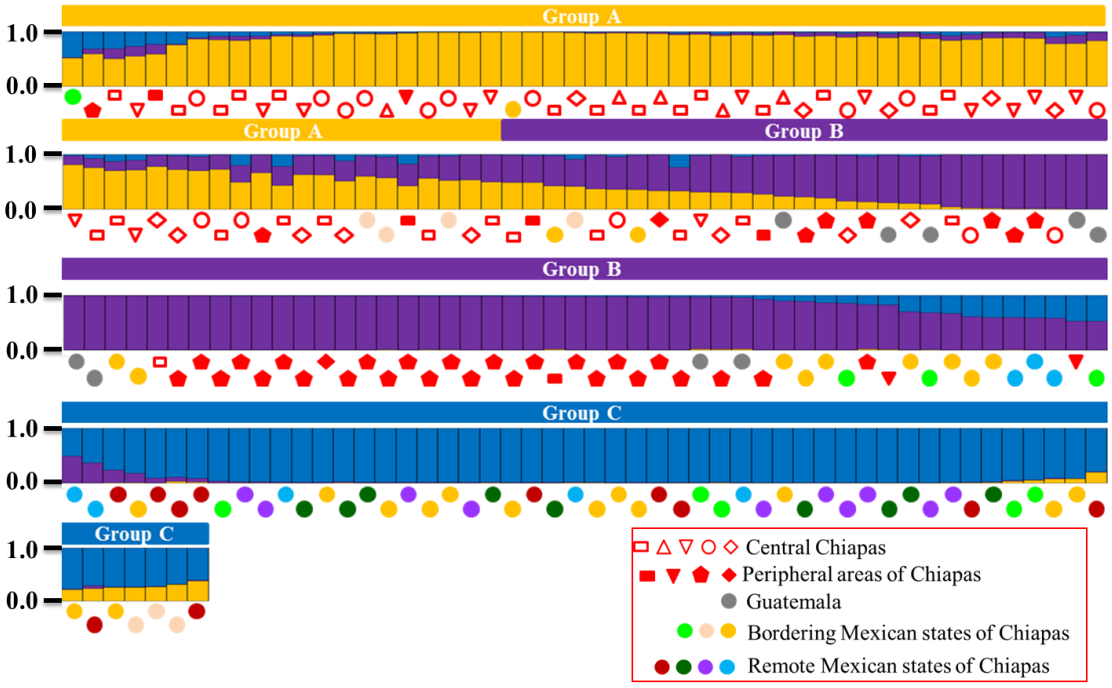


Figure 2.5 Model based clustering (K=3) of Mesoamerican accessions and their distribution of accessions assigned to Groups A, B and C. Three genetic populations, Groups A, B and C, are indicated in orange, purple, and blue, respectively. Each bar represented an accession, different colors in the bar indicated the likelihood of Groups A, B and C. Colored symbols under the bars are collection sites of accessions.

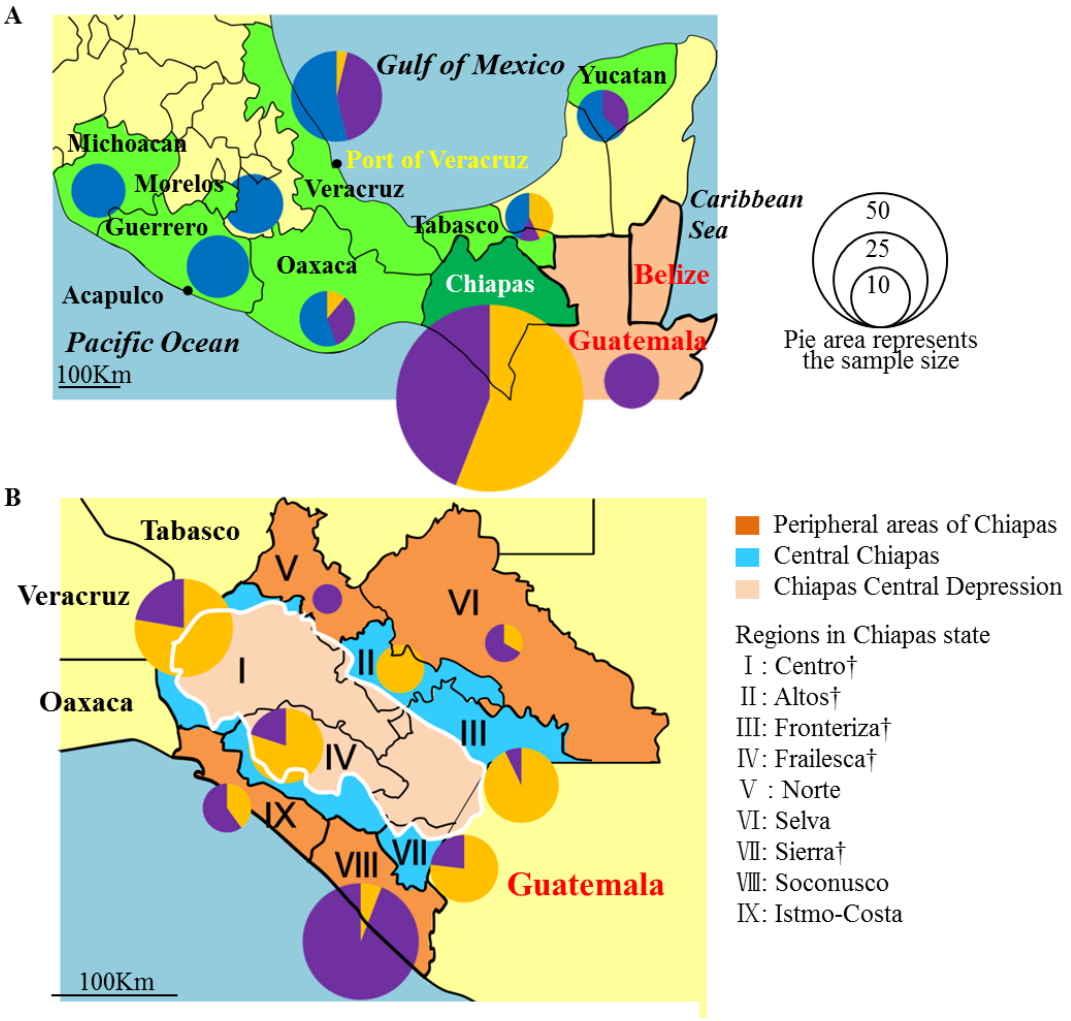


Figure 2.6 Distribution of accessions assigned to Groups A, B, and C in Mexican states, Guatemala (A) and Chiapas regions (B) illustrated as pie graphs. Sizes of the circles correspond to the sample sizes. Mexican states were colored green. The ratio of accessions in Groups A, B and C were shown by pies in orange, purple, and blue, respectively. Regions in central Chiapas covering the Chiapas Central Depression are shown by †.

Table 2.5 H_O, H_E, and F_{IS} of Groups A, B, and C

Group	H_O	H_E	F_{IS}
A	0.290	0.405	0.289***
B	0.120	0.210	0.432***
C	0.042	0.170	0.754***

*** $P < 0.001$.

2.3.3 Phenotypic variation in Mesoamerica

Four yield-related traits including the number of inflorescences per plant, number of female flowers per plant, ratio between female and male flowers, and seed yield (g) per plant were evaluated and compared among collection sites for the 110 Mexican accessions. The mean values and SD of the accessions from central Chiapas were higher than those from the peripheral areas of Chiapas and other Mexican states (Table 2.6). Mean values and SDs of phenotypic variations in Groups A, B and C were also calculated and compared. Accessions from Group A had higher mean values and variations of all the four traits than those from Groups B and C (Table 2.6). Phenotypic comparisons among the three geographic areas and genetic groups indicated that accessions from central Chiapas or Group A had the highest phenotypic variation among the Mexican accessions.

Table 2.6 Phenotypic comparisons of four agronomic traits.

Yield related traits	Geographic region						Genetic groups					
	Central Chiapas		Peripheral areas of Chiapas		Other Mexican states		A		B		C	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Number of inflorescences per plant	25.35	34.27	14.93	11.88	8.35	7.65	24.40	34.9	14.45	12.95	7.62	6.59
Number of female flowers per plant	103.23	141.61	66.17	50.41	42.29	40.39	98.08	144.30	63.94	53.19	41.38	42.72
Ratio between female and male flowers	22.61	42.48	10.06	30.49	10.29	30.56	17.25	38.42	12.82	33.71	11.88	32.71
Seed yield (g) per plant	94.61	114.86	61.50	73.23	43.86	46.17	88.13	106.89	64.89	78.62	38.48	42.51

2.3.4 Ancestral genotype of African and Asian Jatropha in Mesoamerica

To examine the relationship between accessions from Africa and Asia and those from Mesoamerica, and also to specify the Mesoamerican accessions that are genetically close to African and Asian ones, population structure analysis including Mesoamerican, African and Asian Jatropha was performed. As a result, ΔK showed a clear peak when K equaled to three, meaning that all accessions could be divided into three groups based on their genetic information. All accessions including African and Asian accessions were classified into three genetic populations: Groups I, II, and III (Figure 2.7). Group II mostly corresponds to Mesoamerican Group A (see section 2.3.2),

whereas Group III is composed of Mesoamerican accessions of Groups B and C, further indicating the distinct differences between Group A in central Chiapas and other accessions. Group I included all African and Asian Jatropha and interestingly, eight Mesoamerican accessions: five Guatemalan, two Veracruz and one Centro, Chiapas accessions. All these accessions were grouped into Mesoamerican Group B, as shown by black crosses (**+**) in Figure 2.5, suggesting that African and Asian accessions have a closer relationship with Group B, than with Groups A and C. Genetic distances between African and Asian accessions and Groups A, B and C also support the idea (Table 2.7). These results suggest that Group B is the ancestral genotype of Asian and African Jatropha before shipped from Mesoamerica.

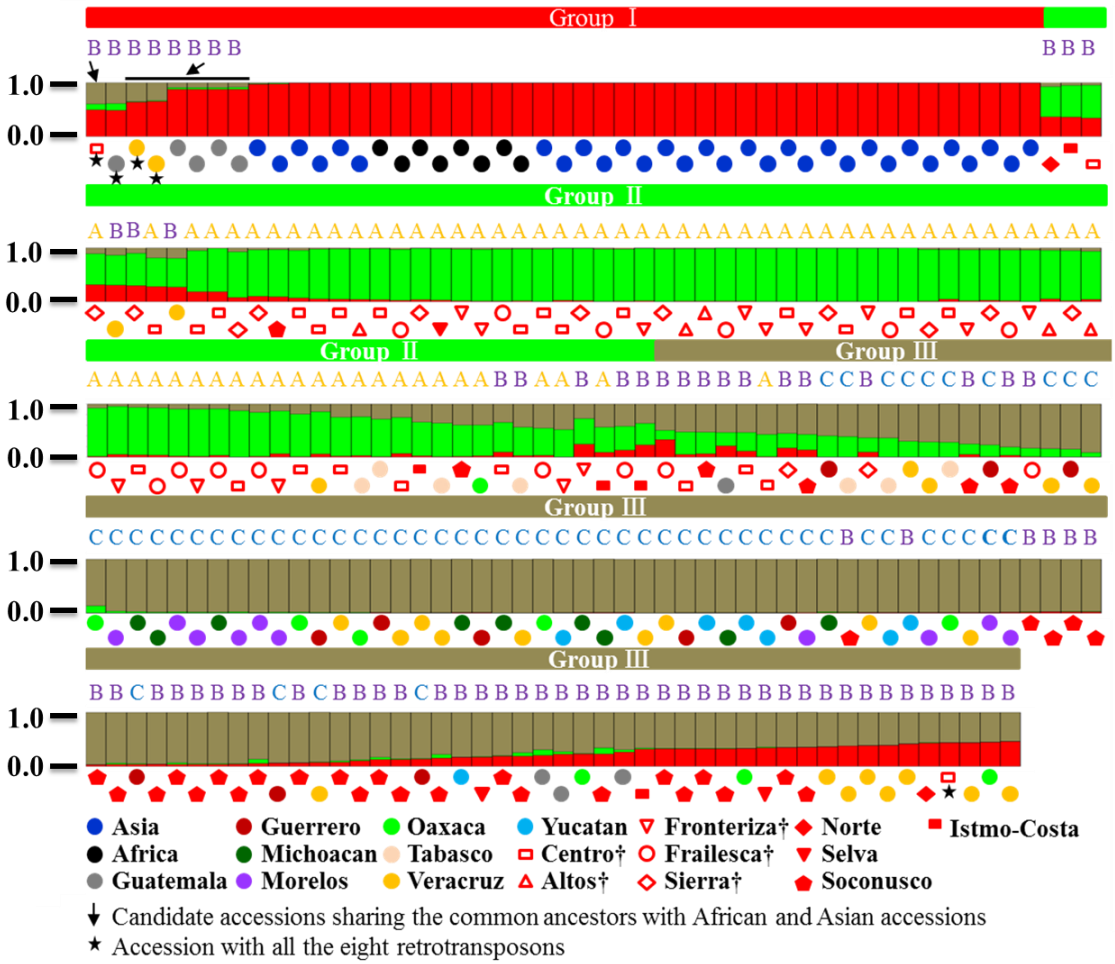


Figure 2.7 Model-based clustering (K = 3) of Mesoamerican, African and Asian accessions. Three Groups I, II, and III are indicated in red, green and grey colors, respectively. Each bar represented an accession, different colors in the bars indicated the likelihood of Groups I, II, and III. Results of the grouping in Mesoamerican accessions (see Figure 2.5) are also shown above the bars as A, B, and C. ★ Accessions carrying with all the eight retrotransposons. ↓ Accessions that genetically closest to African and Asian accessions (see Figures 2.8 and 2.9). †Central Chiapas, covering the Chiapas Central Depression.

Table 2.7 The genetic distance (Nei's D_A distance) between Groups A, B, C, and African and Asian accessions

Group	Africa	Asia
A	0.24	0.23
B	0.13	0.12
C	0.23	0.22

2.3.5 Mesoamerican accessions genetically closest to African and Asian Jatropha

Out of the eight Mesoamerican accessions that grouped together with African and Asian group, there were seven Mesoamerican accessions in the same branch as African and Asian accessions in the phylogenetic tree of all the accessions (Figure 2.8). Four of the seven accessions were from Guatemala, two from Veracruz and one from Centro, Chiapas. These seven accessions were included in the eight accessions of Group I (see section 2.3.4), and remaining one Guatemalan Group I accession was in a different branch. Close genetic relationship between the seven accessions and African and Asian ones were also shown in PCoA analysis (Figure 2.9). Therefore, these seven accessions would be genetically the closest to African and Asian accessions, and would be candidates that share the common ancestors with African and Asian Jatropha.

Large and stable insertions in the genome caused by retrotransposons are usually

irreversible, thus the presence of retrotransposons is reliable for determining parental lineages. Eight retrotransposons from an Asian accession (Alipour et al. 2013) were shared by all the African and Asian accessions. Among the seven accessions, three accessions, one from Chiapas and two from Veracruz, harbored all the eight members (Table 2.8, also see Figures 2.5, 2.7 and 2.8). On the other hands, the remaining four accessions, all from Guatemala, lacked one of the eight retrotransposons. This observation indicates that the genotypes of the three accessions (accession numbers 127, 210, and 354; see Table 2.9) would be genetically the closest to African and Asian Jatropha, and probably sharing the same ancestral genotypes of African and Asian Jatropha.

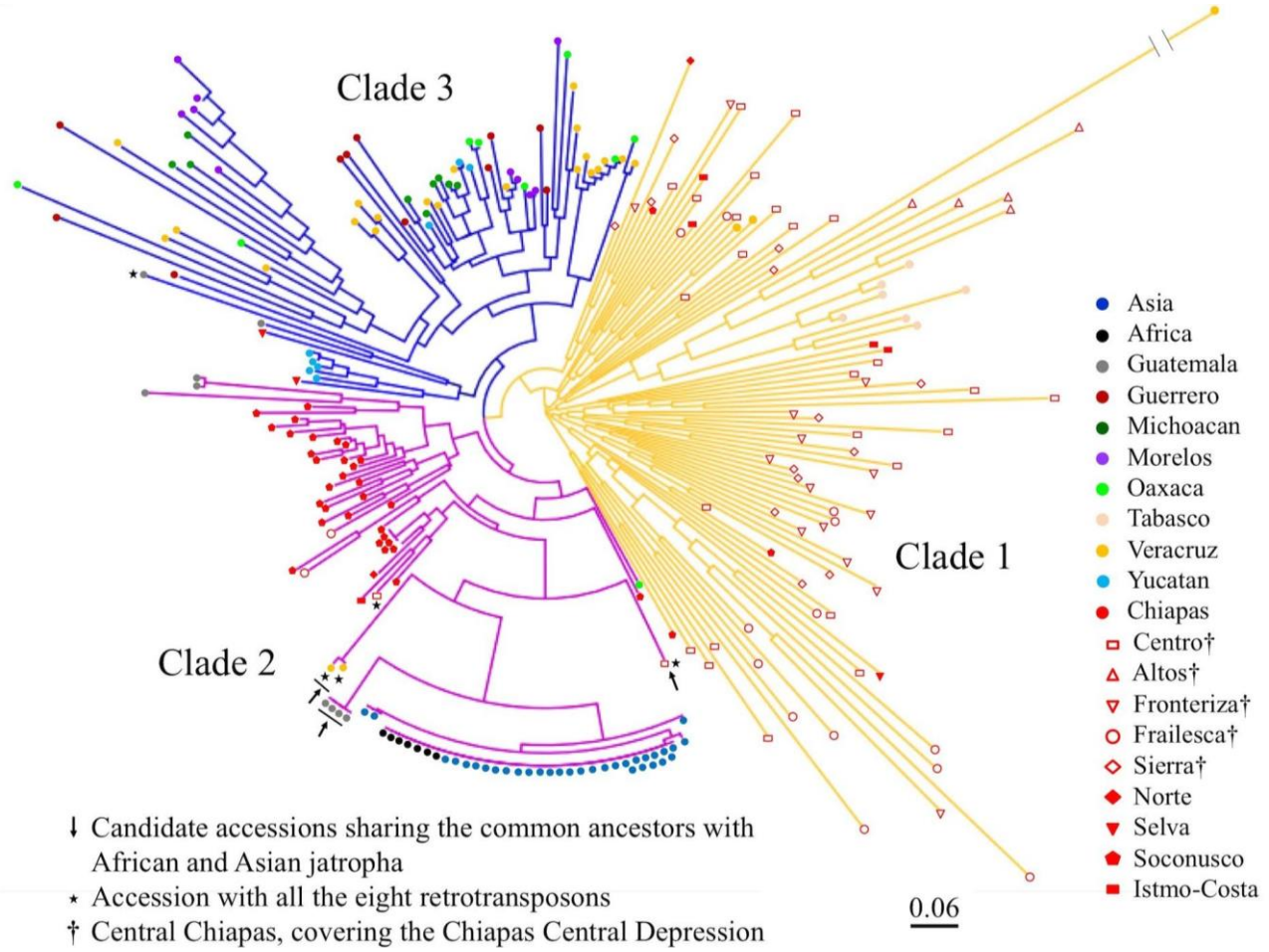


Figure 2.8 The phylogenetic tree of all the accessions including African and Asian Jatropha.

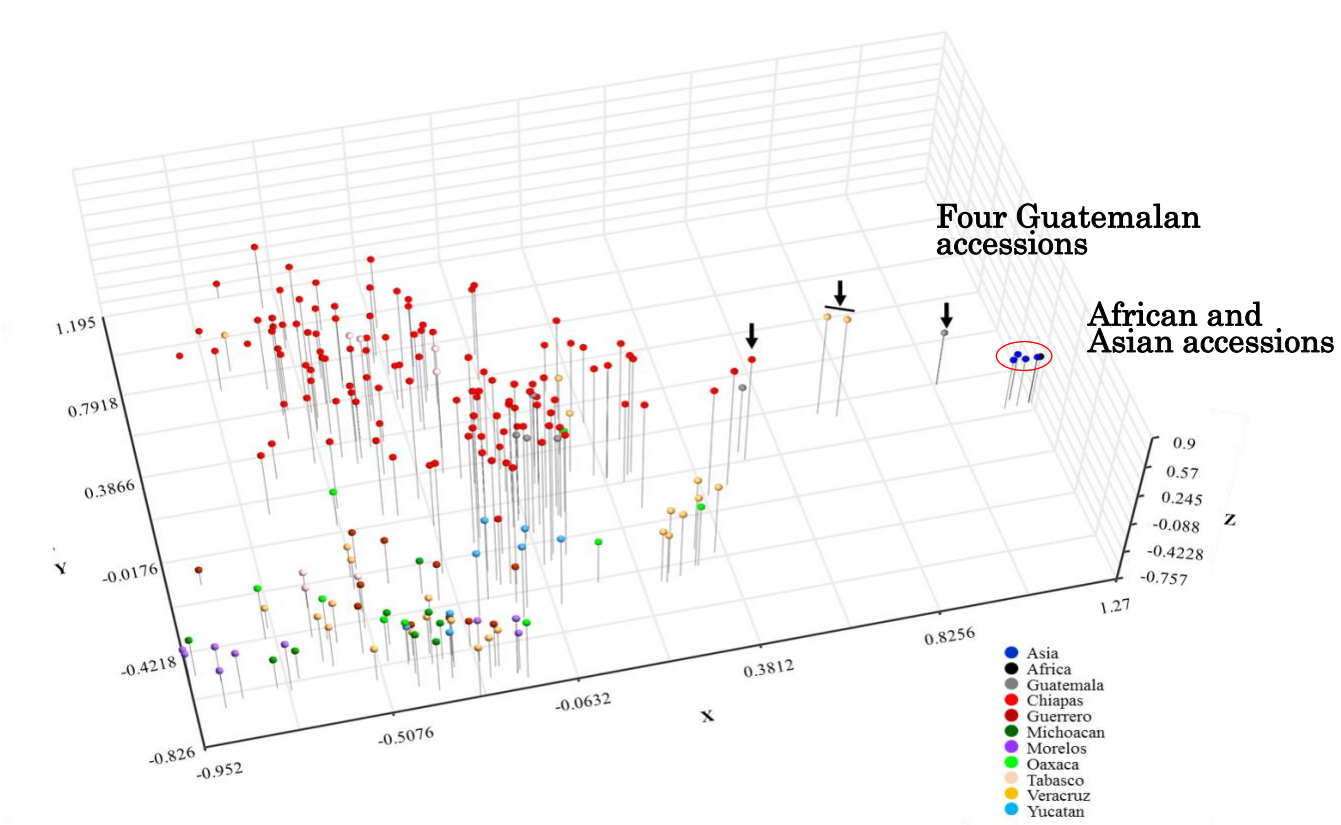


Figure 2.9 Scatter plot for all accessions based on principal coordinate analysis (PCoA), in which 19.82% was explained by x axis, 10.12% was explained by y axis and 5.72% was explained by z axis. The black arrows indicate seven candidate accessions that share common ancestors of African and Asian accessions. Four Guatemalan accessions are merged as one grey dot because of the same genotype. A red circle indicates African and Asian accessions.

Table 2.8 Numbers of accessions from different collection sites, characterized by the number of retrotransposons (RTN).

Number of RTNs	Number of Accessions										
	Africa	Asia	Guatemala	Chiapas	Guerrero	Michoacan	Morelos	Oaxaca	Tabasco	Veracruz	Yucatan
8	7	34	1	2	0	0	0	0	0	2	0
7	0	0	6	5	0	0	0	2	0	7	0
6	0	0	0	32	2	0	0	2	0	5	0
5	0	0	0	36	5	2	5	4	2	8	5
4	0	0	0	21	2	5	0	1	3	4	0
3	0	0	0	8	0	0	3	0	0	0	0
2	0	0	0	3	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0
Total	7	34	7	107	9	7	8	9	5	26	5
Average	8.00	8.00	7.14	5.02	5.00	4.29	4.25	5.56	4.4	5.81	5.00

Table 2.9 Three accessions that share a common ancestor with modern African and Asian accessions.

Accession no.	Collection site (Region, State)
127	Soledad de Doblado, Veracruz
210	Suchiapa, Cento, Chiapas
354	Entrada a Independencia, Veracruz

2.4 Discussion

2.4.1 Genetic distribution in Chiapas

Distinct from Jatropha accessions from central Chiapas, those in peripheral area of Chiapas including northern and southern regions of Chiapas were grouped together and had a significant lower genetic variation than central Chiapas ones. Five regions in Central Chiapas covered the Chiapas Central Depression, which is an ecological region defined by its vegetative and geographic features (Conrad 2007). The Chiapas Central Depression, which extends over five regions in central Chiapas, is composed of hot and semiarid lowlands dominated by deciduous shrubs. It is sandwiched by the northern highlands, central plateau and the southern Sierra Madre Mountains, and it is thus isolated from the northern and southern areas of Chiapas, both of which have hot and humid climates (http://www.backyardnature.net/chiapas/veg-map.htm;http://www.inafed.gob.mx/work/enciclopedia/EMM07chiapas/). An east-west running river divides the Depression and a lake forms in the Depression. It is reasonable that the mountains and highlands surrounding the central Chiapas might have restricted natural propagation of Jatropha from there, making accessions of northern and southern parts genetically distinct from the central.

2.4.2 The center of origin of Jatropha

In this study, the center of diversity for Jatropha in central Chiapas was identified, both genetically and phenotypically. The overlap between the centers of genetic and phenotypic variations, as well as the pronounced geographical differentiation in vegetation of Chiapas Central Depression (Figure 2.1), indicates that the Chiapas Central Depression is most likely the center of origin of Jatropha. The isolated geographical situation and the semiarid climate of the Depression may have favored evolution of this drought-adapted plant species.

Population structure analysis using Mesoamerican accessions and that using all the accessions were performed. The former analysis characterized three genetic populations (Groups A-C) in Mesoamerica. Group A was the major group in central Chiapas (see section 2.3.2). This suggests that it represents the original genotypes of Jatropha. Specific genotypes of Group A might have spread from the Chiapas Central Depression to bordering areas, and subsequently formed distinct genotypes that comprise Groups B and C. Elucidation of the center of origin of Jatropha supports the proposition that Jatropha germplasm from the central Chiapas or Group A should be collected for use in breeding programs to improve the Jatropha oil production.

The latter analysis identified eight Group B Mesoamerican accessions that

genetically close to African and Asian accessions. This suggested that African and Asian accessions were derived from Group B, genetically distinct from those of Group A in the center of origin. Hybridization between African and Asian Jatropha and Group A accessions would be preferable to broaden genetic basis. Progenies not only can adapt the local climate in Africa and Asia, but also can expect hybrid vigor caused by crossing between plants with different genetic backgrounds. Cultivars suitable for mass commercial cultivation can be produced by vegetative propagation of elite progenies. Because Group A accessions are genetic resources of Mexico, collaboration with the Mexican government is necessary for the breeding.

2.4.3 The voyage from Chiapas, Mexico to the Old World

Three Mesoamerican accessions that are genetically the closest to the African and Asian accessions were characterized. Interestingly, the two Veracruz accessions (127 and 354) that carry all of the eight retrotransposons were collected in places within 40 km and 100 km southwest of the Port of Veracruz, respectively, and one Chiapas accession (210) was collected in the Chiapas Central Depression (Figure 2.10). Veracruz has been a major port facing the Gulf of Mexico since 1519, and there was no comparable port on the Caribbean coast of Mexico and Guatemala during or after the Age of Exploration. Moreover, the routes from Veracruz to other places in Mexico were gradually developed

by the conquistadors as the importance of the port increased. During the 16th and 17th centuries, Veracruz was one of the three major American ports for transatlantic trade. There was a fixed route between the Cape Verde Islands and Veracruz for slave transport from Africa (Thomas 2006; Bryant 2014). It is thus reasonable to speculate that Jatropha was taken to the Cape Verde Islands and then to Africa via Port of Veracruz. It is also noteworthy that these three accessions are in Group B, which is not a major genotype in those areas of central Chiapas and Veracruz, where Groups A and C, respectively, represent the major genetic backgrounds of Jatropha (Figure 2.6). The absence of any genetically similar accessions in other locations in Chiapas and Veracruz excludes a gradual natural expansion by seeds, but favors the human-mediated spread of Jatropha from Chiapas to Veracruz. These findings indicate that there were selections when people brought Jatropha from Chiapas to Veracruz. One important question is as follows: who first selected the ancestral African and Asian Jatropha? The conquistadors normally did not enter the Chiapas Central Depression, where no mineral resources were expected (Baudez and Picasso 1987). Therefore, it is probably the indigenous people of Mexico that used the old route to select and carry a few Jatropha seeds or cuttings in central Chiapas and transported them to Veracruz, and the ancestors of accessions 127 and/or 354 were then transported to the Cape Verde Islands by Portuguese sailors from the Port of

Veracruz (Figure 2.10).

Figure 2.10 Jatropha's voyage from the Chiapas Central Depression to the Old World. The map shows the altitudes of Chiapas and its surrounding areas. The background map is available from https://commons.wikimedia.org/wiki/File:Mexico_topographic_map-blank.svg. The stars represent the sampling sites of three accessions that share the same ancestor of African and Asian Jatropha. The red line represents a current road between Veracruz and Chiapas.

2.4.4 Genetic and historical reason for narrow genetic basis in Africa and Asia

Since a limited number of genotypes would have been brought to the Old World by

Portuguese galleons, and crosses could have occurred only among small populations in the Old World plantings, yielding very low genetic variation. This breeding bottleneck represents a typical example of the founder effect and is probably the major cause of the failure of Jatropha breeding in African and Asian countries. In addition, accumulation of mildly deleterious genes exhibited in the homozygous state by inbreeding might show inbreeding depression. In fact, the Veracruz accession (127), which has been identified to share a common ancestor with African and Asian accessions, had low values for yield-related traits (number of inflorescences per plant: 3; number of female flowers per plant: 18; ratio of female to male flowers: 0.06 and seed yield per plant: 15.10g), indicating that the low seed yield of African and Asian accessions is genetically determined.

Mesoamerican people seemed not to have placed any value on the seed-yielding potential of Jatropha as they selected and brought Jatropha seeds or cuttings from Chiapas to Veracruz for use as live fences and medicines (Duke 1985; Prasad et al. 2012; Agbogidi et al. 2013), and these applications are still widespread among Veracruz inhabitants even nowadays. Fences and medicines are the likely reasons why the Portuguese shipped the plant to Africa and Asia (Sabandar et al. 2013). Therefore, seed production did not have a high priority in Africa and Asia until recently.

2.5 Summary

A comprehensive genetic study was conducted with a worldwide collection of Jatropha accessions by using SSR markers and RBIP markers. As a result, the most diverse and various genetic resources in central Chiapas were found. Considering the genetic distribution in Chiapas combined with Chiapas topography, the center of origin for Jatropha should be in the Chiapas Central Depression. Identification of one Chiapas and two Veracruz accessions that are genetically the closest to African and Asian Jatropha allowed to speculate that the voyage from Mexico to the Old World started from central Chiapas, transmitited from Port of Veracruz, and finally reached Africa and Asia. African and Asian Jatropha was derived from Mesoamerican Group B.

Heller (1996) described the center of origin in Mesoamerica and the transmission from Mesoamerica to Africa and later Asia. My study is the first one to characterize the center of origin for Jatropha in Chiapas Central Depression and to clarify the route from the center of origin to the port of Veracruz. I also indicated that human selection during the transmission from the origin to the Old World caused a low genetic basis in Africa and Asia nowadays. Yield improvement of African and Asian Jatropha could be achieved by hybridization with Mesoamerican Group A, which harbors the highest genetic diversity. This approach is expected to help in breeding program in Africa and

Asia, therefore helps in the biofuel production and fight against the climate change.

Chapter 3

Association study of seed-yield related traits for *Jatropha curcas* L. in Mexico

3.1 Introduction

Jatropha (*Jatropha curcas* L.) is a diploid species with a chromosome number (2n) of 22 (Dahmer et al. 2009) and a genome size of 418 Mb (Carvalho et al. 2008). Whole genome sequence analysis of Jatropha was firstly performed by Sato et al. (2011), and the genomic sequence obtained was 297.7 Mb in length and contained 30,203 complete and partial protein-encoding genes (Hirakawa et al. 2012). There are few studies on linkage mapping and QTL mapping of Jatropha so far. Wang et al. (2011) generated an interspecific linkage map of Jatropha by using the cross between *J. curcas* and *J.integerrima*, and this map was then used for whole-genome scan for QTL and expression QTL (eQTL) that affected the seed oil traits (Liu et al. 2011). QTL mapping on growth and seed yield traits was also performed by this map, and two QTLs controlling seed yield were found in two QTL clusters harboring several yield-related traits in different linkage groups, respectively (Sun et al. 2012). Later King et al. (2013) developed an intraspecific linkage map of Jatropha. Recently, Wu et al. (2015) reported the draft genome sequences anchored to the interspecific linkage map.

There is only one trial of association study so far performed by Sharma and Chauhan (2012). Using of Jatropha genotypes categorized by oil content and DNA markers developed from candidate genes responsible for fatty acid biosynthesis, they

found that amplicon of the stearoyl desaturase gene showed association with oil content. However, a large scale's association study of yield-related traits in Jatropha has not been reported yet. The whole genomic sequence of Jatropha allowed to develop SSR and RBIP markers. This study was conducted to find markers that significantly associated with yield of Jatropha, which can be used in the selection of elite Jatropha. Elite Jatropha can be used in the mass plantation worldwide, especially in Africa and Asia where low productivity was found.

3.2 Materials and Methods

Workflow of the analysis conducted in Chapter 3 was shown in Figure 3.1. After genomic DNA was extracted, genotyping and scoring were performed. Genotypic data and phenotypic data, as well as kinship matrix and LD, were integrated in the association analysis. After running association analysis in TASSEL, the results were output with *P* values between each marker and each trait. Significant associations were selected by setting a threshold of *P* value.

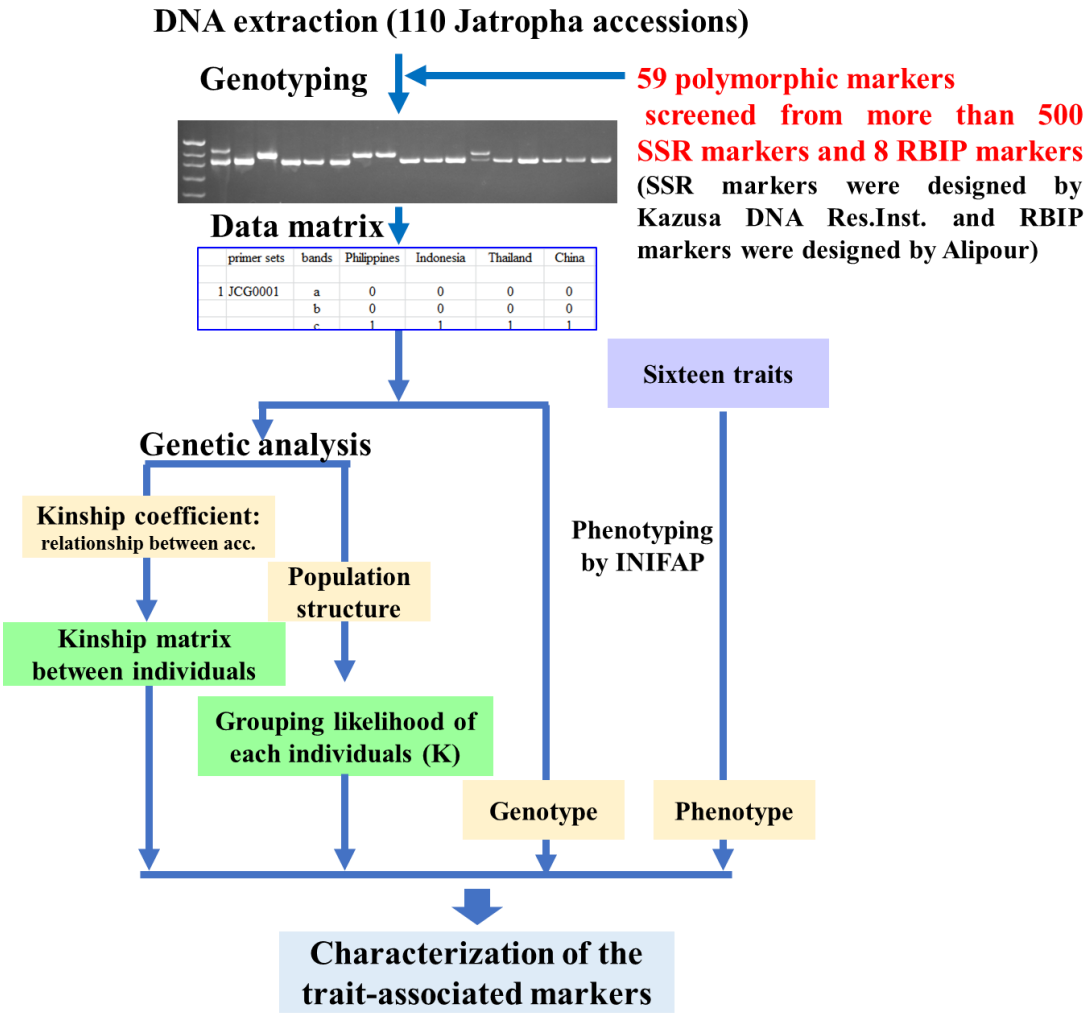


Figure 3.1 Workflow of the analysis conducted in Chapter 3.

3.2.1 Accessions and agronomic traits

One hundred and ten Mexican Jatropha accessions were used in this study. They were composed of 61 from Chiapas, five from Guerrero, five from Michoacan, four from Morelos, seven from Oaxaca, 23 from Veracruz and five from Yucatan. All the Mexican accessions were from the collection of INIFAP, Mexico. These accessions were grown in the Rozario Izapa station (Chiapas, Mexico). They were at least 3 years

old and the distances between plants were 3 x 2 m. The NPK fertilizers with urea (N), triple superphosphate (P) and potassium chloride (K) were used. In the first year of planting, the NPK formula 60-40-20 was used in 2 applications; the first application was carried out 30 days after sowing and the second at 6 months post-pruning training. In the second year and following, the NPK formula 80-40-20 was used and was distributed in 2 applications per year: the first at the beginning of the rain season and the second at 6 months later (after pruning). The experiment was conducted under rainfall conditions without irrigation. The average annual rainfall was 4.7 m per year during an 8-month rainy season.

Sixteen agronomic traits (Tables 3.1 and 3.2) were measured in the Rozario Izapa station. INIFAP offered the agronomic traits under the academic exchange agreement with Tottori University, Japan. Four of traits were also examined in Chapter 2. Mean value and standard deviation (SD) for sixteen traits, as well as correlation coefficient (R) between pairwise traits, were calculated prior to the association study. It is assumed that phenotypes follow a normal distribution in the mixed linear model (MLM). So the traits that were not normally distributed were transformed by using natural logarithmic function, and Shapiro-Wilk Test (Shapiro and Wilk 1965) was performed to confirm the normality.

3.2.2 DNA markers

More than 500 SSR markers were developed by Dr. Shusei Sato by designing primer pairs surrounding SSR sequences identified from the genomic database of an Asian accession (Sato et al. 2011; Hirakawa et al. 2012). Eight RBIP markers employed in this study were from members of the copia-type families identified in the genomic database by Alipour et al. (2013). Primers were designed for flanking sequences as well as for LTRs (Alipour et al. 2013; Tsuchimoto, personal communication). Retrotransposon insertion at each locus was observed by combining primers designed from the flanking sequences at both sides and primers designed from LTR sequences.

3.2.3 DNA extraction

Genomic DNA was extracted from 100mg Jatropha leaves with the DNeasy Plant Mini Kit (Qiagen, USA) according to the manufacturer's instructions. The DNA samples were diluted to a final concentration of 0.35 ng/μl and stored at −20℃ until use.

3.2.4 Genotyping and data scoring

PCR for SSR markers was performed in a 10-μl total volume solution containing 1.4 ng of DNA template, 0.16 μl of SSR primer mix (25 μM each), 1× PCR buffer, 0.03 μM $MgCl_2$, 0.8 μl dNTPs (2.5 mM), 0.08 U BIOTAQ polymerase ((Bioline, UK, 5

units/μL), and Milli Q water. Touch-down amplification was performed as follows: 3 min hold at 94 ℃, followed by 3 cycles at 94 ℃ for 30 s, and 68 ℃ for 30 s, with the annealing temperature reduced by 2 ℃ until 64 ℃ every 3 cycles. Continuing with 3 cycles at 94 ℃ for 30 s, 62 ℃ for 30 s, and 72 ℃ for 30 s, the annealing temperature was reduced by 2 ℃ until 58 ℃ every 3 cycles. A further 30 cycles of amplification was performed at 94 ℃ for 30 s, 55 ℃ for 30 s, and 72 ℃ for 30 s. The PCR products were stored at −20 ℃ until separation by polyacrylamide gel electrophoresis (PAGE). Amplified bands were stained with ethidium bromide. For markers that showed ambiguous bands, experiments were repeated by changing experimental conditions until clear bands were observed.

For RBIP markers, PCR was performed in a 5 μl total volume solution containing 0.7 ng of DNA template, 0.08 μl of primer mix (50 μM each), 1× PCR buffer, 0.03 μM $MgCl_2$, 0.4 μl dNTPs (2.5mM), and 0.08 U BIOTAQ polymerase ((Bioline, UK, 5 units/μL), made to the volume with Milli Q water. PCR reaction was started with denaturation at 94 ℃ for 2 min, followed by 35 cycles of 45 s at 94 ℃, 45 s at 55 ℃ and 2 min at 72 ℃, continued with a final elongation step at 72 ℃ for 10 minutes. Amplified bands were stained with ethidium bromide. The PCR products were stored and analyzed in the same way as the SSR markers.

Screening of SSR markers was conducted by using representative Mexican accessions in which high polymorphism were expected. SSR markers that showed high polymorphism were chosen to amplify all accessions. For all the SSR markers and RBIP markers, all bands were individually scored (presence 1, absence 0). All data analyses were performed based on the genotypic data matrix.

3.2.5 Kinship and Linkage Disequilibrium (LD) analysis

The kinship coefficients (F) of all pairwise accessions were calculated using software package SPAGeDi (Hardy and Vekemans 2002), by outputting a kinship matrix. Negative values between accessions were set to 0, as described by Yu et al. (2006). LD was estimated as the squared allele frequency correlation (r^2) between all pairs of DNA markers by using TASSEL2.1 (Bradbury et al. 2007).

3.2.6 PCA and structure analysis

Principal component analysis (PCA) based on the genotype data was performed in R (R Development Core Team 2012), to compare the relatedness among all accessions. Final results were displayed by a two-dimensional figure (x and y represent the first two principal components).

Model-based clustering analysis was performed using Mesoamerican accessions or all accessions (including African and Asian accessions) with STRUCTURE 2.3.4

(Pritchard et al. 2000). This analysis was conducted to know the optical population number (K) in a population based on the genetic differentiation among individuals. The analysis was performed using the Markov Chain Monte Carlo (MCMC) method based on 1×10^5 iterations followed by a burn-in period of 5×10^4 iterations. To determine the optimal number of K to show the differentiation in the population at the biggest extent, firstly the K was set to vary from 1 to 10, then after 10 independent runs for each K, ΔK was calculated based on the rate of change of the log likelihood for K values between successive K values by following Evanno et al. (2005). Finally the optical K value, the ΔK of which have a clear peak, was chosen. The cluster membership coefficient from STRUCTURE was integrated to get a Q matrix.

3.2.7 Association study

Association study was performed in TASSEL2.1 using the MLM model by integrating Q matrix and kinship matrix, which controls spurious results arising from population stratification and relative kinship. The distribution observed in –log (P value) of marker-traits associations (MTAs) for each trait and marker was compared by quantile-quantile plot (Q-Q plot), and the departure of observed distribution and the expected was used to assess reliability of the model used. Significant associations were shown by Q-Q plots as well, and were tested based on threshold $p = 6.58 \times 10^{-4}$, a

stringent Bonferroni correction calculated by dividing 0.05 by the total DNA marker number used in this analysis.

3.2.8 Alignment of markers

Alignments were performed to check if markers that found to significantly associate with agronomic traits, were in any published QTLs (Liu et al. 2011; Sun et al. 2012). To do this, both markers and published markers in the QTLs (Liu et al. 2011; Sun et al. 2012) were aligned to the genomic database of Jatropha, and later checked the distance between markers, and then confirmed if the markers were in the QTLs or not.

3.3 Results

3.3.1 Mean value, SD and correlation coefficient of the agronomic traits

Mean value and SD for each of the 16 agronomic traits were calculated within 110 Mexican Jatropha accessions (Table 3.1). Huge variations were detected in all the agronomic traits. We also evaluated correlation among the traits by correlation coefficients, and found different extents of correlation coefficients with each other (Table 3.2). The correlation between traits suggests that their genetically-responsible regions are overlapped. A strong positive correlation (R = 0.78 or above) was observed among six traits; number of inflorescences per plant, number of female flowers per plant, number of fruits harvested per plant, fruits weight (g) per plant, number of seed per

plant and seed yield (g) per plant. Another strong correlation (R = 0.80 or above) was witnessed among five traits; fruit weight (g), seed weight (g), seed length (cm), seed width (cm) and seed thickness (cm). The third strong correlation (R = 0.99) was between proportion of female flowers to male flowers and ratio of female and hermaphrodite flowers. Low or no correlations were detected among other traits. Fourteen traits that did not follow normal distribution, except for fruit weight (g) and seed weight (g), were transformed and then applied for the association analysis, as described in Materials and Methods.

Table 3.1 Mean value and SD for sixteen agronomic traits.

	Mean value	SD
Number of inflorescences per plant	14.94	20.90
Number of female flowers per plant	65.97	86.95
Ratio between female and male flowers	0.08	0.08
Number of flowers per inflorescence	71.93	44.92
Ratio of female and hermaphrodite flowers	0.17	0.30
Number of fruits per cluster	3.53	2.45
Fruit weight (g)	11.35	4.87
Seed weight (g)	0.73	0.29
Seed length (cm)	1.60	0.58
Seed width (cm)	0.97	0.38
Seed thickness (cm)	0.71	0.28
Number of fruits harvested per plant	29.24	37.77
Fruiting rate	0.46	0.50
Fruit weight (g) per plant	378.50	479.14
Number of seeds per plant	76.03	97.24
Seed yield (g) per plant	62.85	80.29

Table 3.2 Correlation coefficient of pairwise agronomic traits.

	PHE1	PHE2	PHE3	PHE4	PHE5	PHE6	PHE7	PHE8	PHE9	PHE10	PHE11	PHE12	PHE13	PHE14	PHE15	PHE16
PHE1	**1.00**	**0.96[1]**	-0.10	0.07	-0.03	0.00	0.17	0.22	0.19	0.26	0.24	**0.79[1]**	-0.05	**0.79[1]**	**0.78[1]**	**0.79[1]**
PHE2		1.00	0.00	0.08	0.02	0.17	0.22	0.26	0.21	0.28	0.25	**0.84[1]**	-0.04	**0.85[1]**	**0.82[1]**	**0.84[1]**
PHE3			1.00	-0.35	**0.99[2]**	0.31	0.17	0.07	0.07	0.04	0.05	0.04	0.02	0.07	0.05	0.05
PHE4				1.00	-0.55	0.14	0.11	0.20	0.24	0.21	0.20	0.12	0.04	0.10	0.12	0.11
PHE5					1.00	0.17	0.22	0.16	0.13	0.18	0.19	0.00	0.06	0.03	0.00	0.01
PHE6						1.00	0.27	0.28	0.26	0.23	0.24	0.22	0.07	0.24	0.21	0.22
PHE7							1.00	**0.92[3]**	**0.87[3]**	**0.81[3]**	**0.80[3]**	0.26	0.37	0.34	0.27	0.31
PHE8								1.00	**0.93[3]**	**0.88[3]**	**0.88[3]**	0.26	0.30	0.32	0.26	0.32
PHE9									1.00	**0.90[3]**	**0.90[3]**	0.24	0.34	0.26	0.24	0.26
PHE10										1.00	**0.97[3]**	0.29	0.10	0.32	0.29	0.32
PHE11											1.00	0.24	0.09	0.27	0.24	0.27
PHE12												1.00	0.20	**0.98[1]**	**1.00[1]**	**0.98[1]**
PHE13													1.00	0.20	0.20	0.19
PHE14														1.00	**0.98[1]**	**0.99[1]**
PHE15															1.00	**0.98[1]**
PHE16																1.00

PHE1: Number of inflorescences per plant; PHE2: Number of female flowers per plant; PHE3: Ratio between female and male flowers; PHE4: Number of flowers per inflorescence; PHE5: Ratio of female and hermaphrodite flowers; PHE6: Number of fruits per cluster; PHE7: Fruit weight (g); PHE8: Seed weight (g); PHE9: Seed length (cm); PHE10: Seed width (cm); PHE11: Seed thickness (cm); PHE12: Number of fruits harvested per plant; PHE13: Fruiting rate; PHE14: Fruit weight (g) per plant; PHE15: Number of seeds per plant; PHE16: Seed yield (g) per plant. Three highly-correlated trait groups are shown; PHE1, 2, 12, 14, 15 and 16 (marked with [1]); PHE3 and 5 (marked with [2]); PHE7, 8, 9, 10 and 11 (marked with [3])

3.3.2 Polymorphism of markers, kinship coefficient and linkage disequilibrium

Screening of SSR markers among all the SSR markers was conducted by using representative Mexican accessions in which high polymorphism are expected. Markers which showed clear polymorphism among them were selected for further PCR amplification of all the accessions. As a result, 59 SSR markers and eight retrotransposon markers that showed polymorphisms were employed in the study. The eight RBIP markers consists of six full-length members with right and left LTRs showing 100% sequence identity with each other and two solo LTRs displaying 100% sequence identity with the consensus sequence of the families (Alipour et al. 2013; Tsuchimoto, personal communication). All of these were expected to have retrotransposed more recently than other members. The SSR markers yielded 228 polymorphic bands, ranged from two to seven with an average of 3.2 in the 110 accessions examined.

Genetic relatedness of all accession pairs was shown by kinship coefficients. A total number of 3526 (58.82%) pairs showed no kinship with each other, and kinship coefficients of 1920 (32.03%) pairs ranged from 0.0 to 0.2, and the average kinship coefficient of any two accessions was 0.054 (Figure 3.2). The pattern of genetic relatedness as revealed in the analysis showed that most pairs have no or weak

relationship.

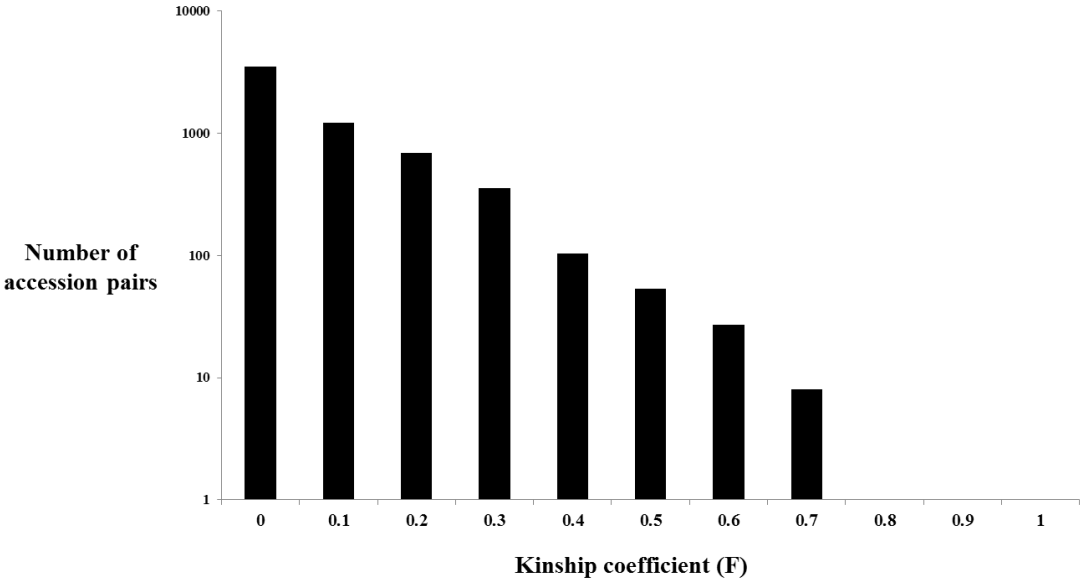


Figure 3.2 Distribution of pairwise kinship coefficient between accessions

Different degree of linkage disequilibrium (LD) among the DNA markers was detected in the analysis panel, with the mean pairwise r^2 equaled to 0.032, and the biggest and smallest values were 0.450 and 0.0001, respectively (Figure 3.3). It was obvious that most markers used in the panel showed little or no genetic linkage between each other.

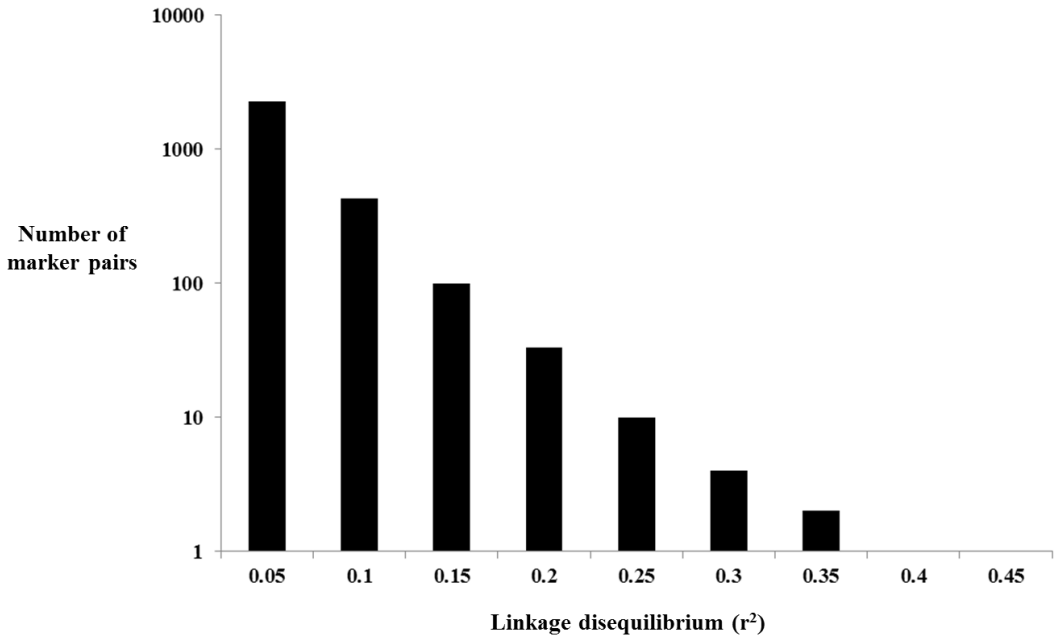


Figure 3.3 Distribution of pairwise LD between markers.

3.3.3 PCA and population structure

PCA on the basis of the genetic data from DNA markers, which could explain 32.53% of the total variation, indicated that there was a relationship between the

genotype and the geographical origin (Figure 3.4). Accessions from the state of Chiapas and those from other Mexican states were distributed in different positions. Accessions from Chiapas had a higher level of genetic discrimination among individuals over those in other Mexican states.

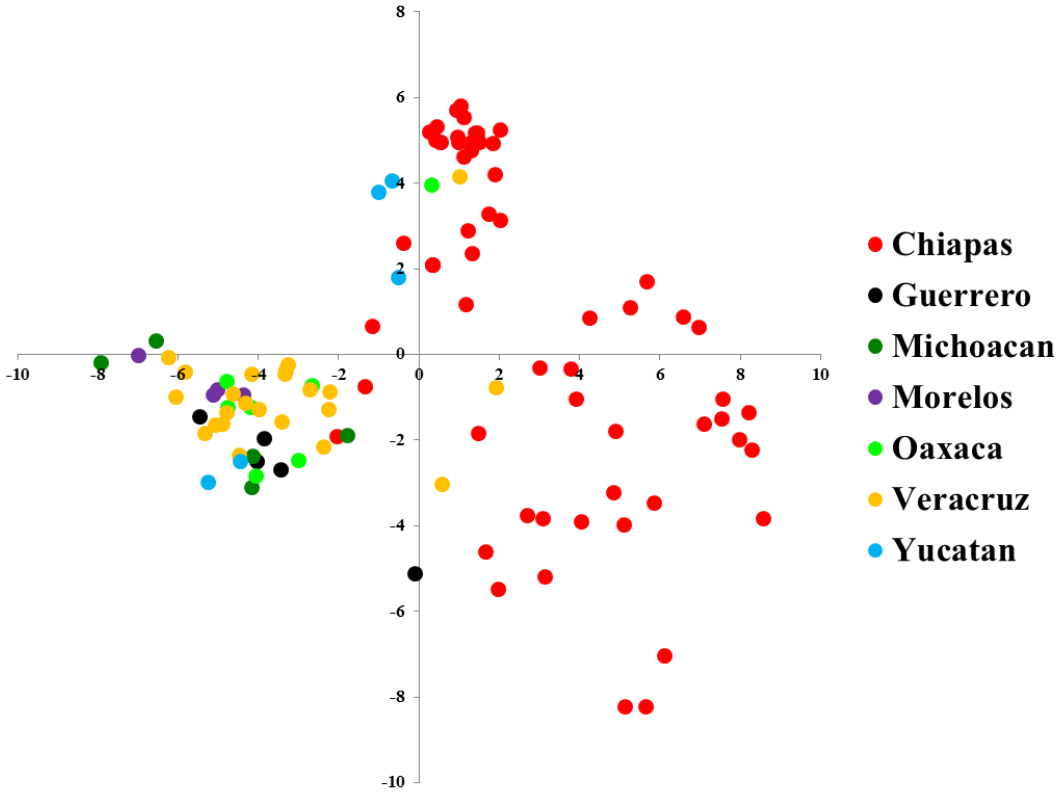


Figure 3.4 PCA scatter plot for Jatropha accessions based on the genetic data. X and Y axes are the first two principal components.

One hundred and ten accessions were separated into two distinct groups in structure analysis (Figure 3.5). Group I consisted of 32 accessions (30 accessions from

Chiapas and the other two from Veracruz). Group II was composed of 78 accessions, in which 31 accessions from Chiapas, 21 from Veracruz, and all the accessions from other five Mexican states (Guerrero, Michoacan, Morelos, Oaxaca, and Yucatan) were included.

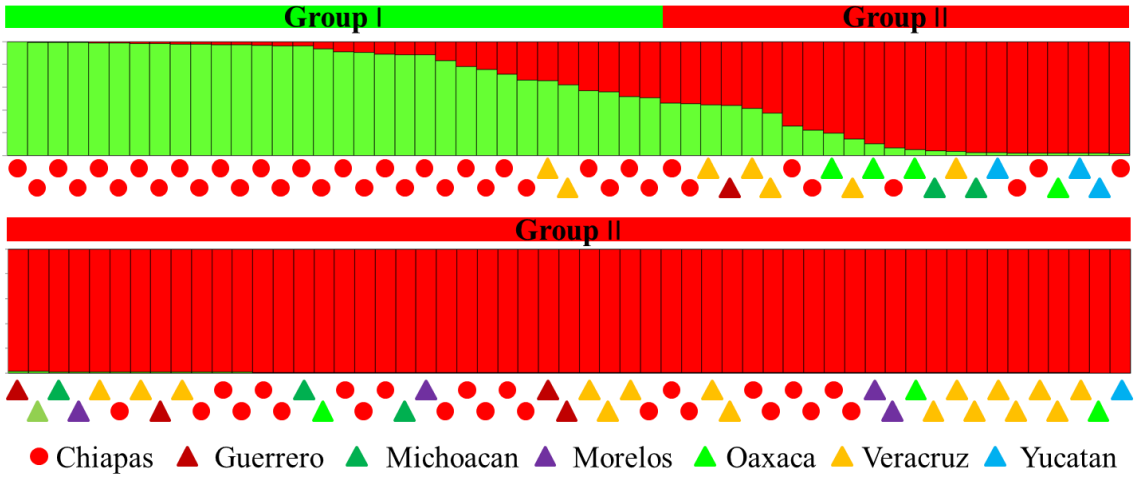


Figure 3.5 Population structure analysis of Jatropha accessions (K=2).

Both results showed an obvious population structure in the panel. Genetic relatedness among accessions was observed. It suggested that control of Q matrix was necessary in the model in the analysis in order to avoid spurious MTA.

3.3.4 Association analysis

Integrated with kinship matrix and Q matrix, sixty-seven markers and sixteen transformed or untransformed agronomic traits (see 3.2.1 in Materials and Methods) were used in the association study. The result of the association analysis contained the *P*

values of associations between each trait and each marker. *P* value showed the reliability of associations. The arbitrary threshold of –log (*P* value) = 3.18 was set to minimize the chance of detecting false positives. Then eight significant MTAs were detected. They included six traits (all passed the Shapiro-Wilk Test for normal distribution) and six SSR markers (Table 3.3). All the significant associations were clearly displayed in Q-Q plot (Figure 3.6).

Table 3.3 Significant association signals of seed-yield related traits of Jatropha.

Trait	*P* value [a]	Marker	R^2 [b]
Number of inflorescences per plant	2.04E-04	8B	0.1565
	2.13E-04	JcGNS0217	0.1352
Number of female flowers per plant	9.50E-06	JcGNS0101	0.2043
	3.01E-04	JcGNS0261	0.2129
Number of flowers per inflorescences	1.89E-04	JcGNS0101	0.2855
Number of fruits per cluster	1.81E-04	JcGNS0192	0.1499
Fruit weight (g)	3.01E-08	JCG0061	0.4391
Seed yield (g) per plant	2.57E-06		0.3151

a: *P* value for each MTA; b: R^2, the percentage of phenotypic variation explained by each MTA.

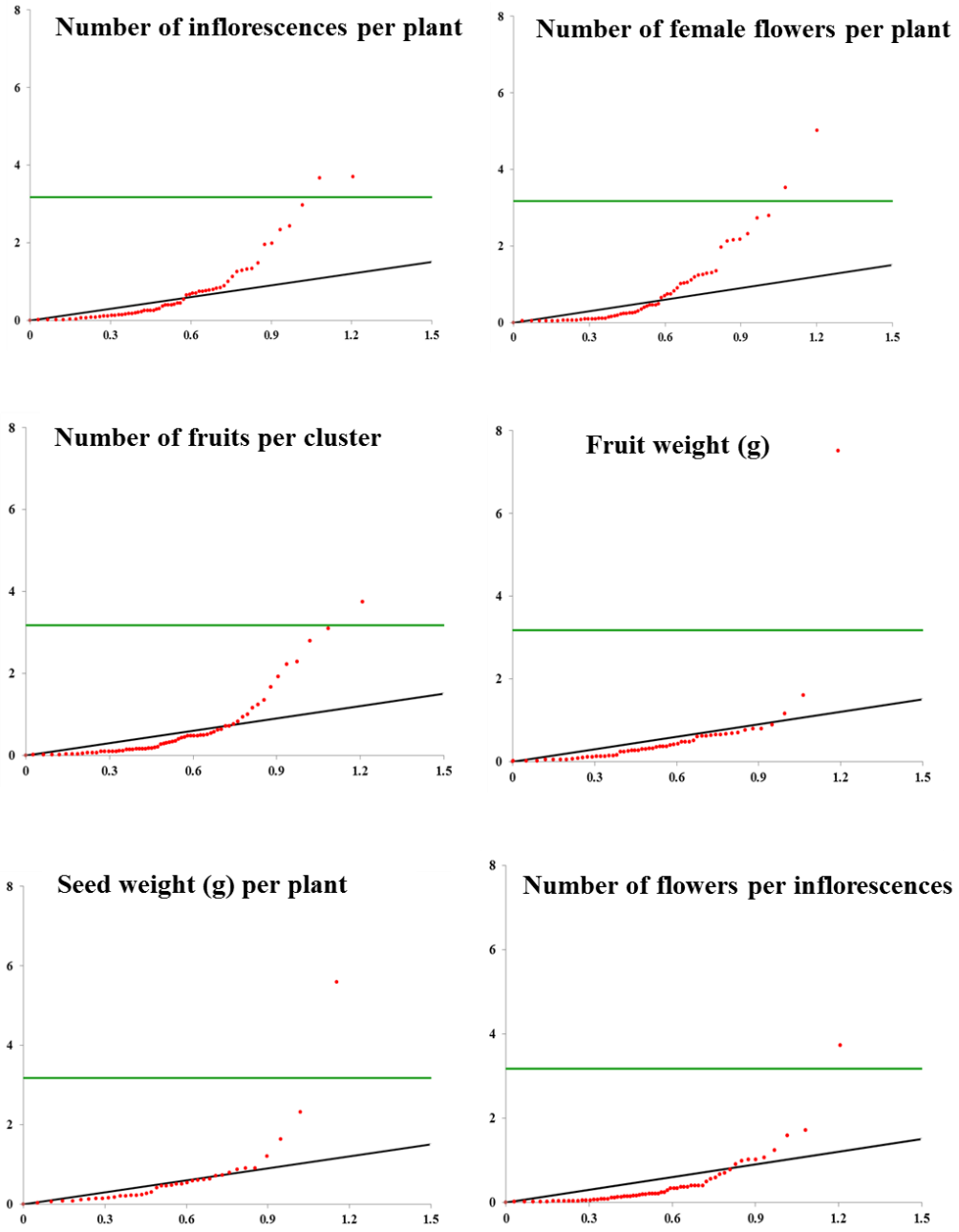


Figure 3.6 Q-Q plot of association analysis of the six associated traits. Each dot represents a MTA. The vertical axes indicate observed –log(*P*) for association and the horizontal axes indicate expected –log(*P*) under the null hypothesis of no association. Black lines indicate observed and expected –log(*P*) under the null distribution. Horizontal green lines represent the Bonferroni-corrected significance threshold (3.18) of observed –log(*P*).

Two significant associations were detected between number of inflorescences per plant and two markers (8B and JcGNS0217). Both of these two associations had high R^2 values (0.1565 and 0.1352 respectively), indicating high portion of variation of inflorescences number can be explained by these two markers. Two significant associations were observed between number of female flowers per plant and two markers (JcGNS0101 and JcGNS0261), and these two markers could represent 0.2043 and 0.2129 of the variation of the female flower number, respectively. Markers JcGNS0101 and JcGNS0192 were involved in association with number of flowers per inflorescences and number of fruits per cluster, respectively, with R^2 values equaled to 0.2855 and 0.1499. Note that JcGNS0101 was associated with two traits. Marker JCG0061 was also found to be correlated with two traits; fruit weight (g) and seed yield (g) per plant with high contributions (0.4391 and 0.3151, respectively). “The elite allele” of each MTA, that shows the highest mean value of each trait, was determined. Presence and absence of each of the elite alleles was examined in Chiapas and other Mexican states (Table 3.4). As a result, accessions from Chiapas had all the six elite alleles, while other Mexican states only had half of them.

Table 3.4 Presence (+) and absence (-) of the elite alleles of associated markers in Chiapas and other Mexican states.

Regions	Sample number	Associated marker					
		8B	JCG0061	JcGNS0101	JcGNS0192	JcGNS0217	JcGNS0261
Chiapas	61	+	+	+	+	+	+
Other Mexican states	49	+	-	+	+	-	-

Marker JCG0061 was associated with two traits including the most important trait in breeding, seed yield (g) per plant, and showed the highest contribution to phenotypic variations. These suggested that this marker was the most useful for molecular breeding of Jatropha among the markers examined in this study. In order to improve the seed yield, elite alleles of this marker should be useful for the MAS of Jatropha.

3.3.5 Alignment of the significant markers against the genome sequences

Genome-wide alignment of all 67 markers was performed against the Jatropha genomic sequences published by Sato et al. (2011) and those published by Wu et al. (2015). All markers could be successfully identified in scaffolds, and 59 markers could be mapped to 11 linkage groups of the Jatropha genome (data not shown). Markers were

distributed to all of the 11 linkage groups. Then It was examined that whether there were known genes which play important roles in seed yield in Jatropha plants in the flanking sequences of the six associated markers or not, but no annotated yield-related gene was found in the scaffolds where the markers existed. A further investigation of these markers was conducted to see whether they were involved in interspecific QTLs (Liu et al. 2011; Sun et al. 2012). As a result, only the marker JCG0061 was found to be located within a QTL that contributes to the female flower number. Loci of other five markers did not reside within any QTL.

3.4 Discussion

Efforts made to improve the seed yield are the most important in breeding programs of seed-oil producing crops. By utilizing oils as biofuels, it would exhibit great significance in alleviating current energy crisis, processing and reducing the greenhouse effect gases, and finally preventing the world climate change. However, the seed yield is a complex trait controlled by many genes with major or minor effects (Moncada et al. 2001). Breeding of seed yield remains one of the most challenging tasks to date, especially in the new emerging crop like Jatropha. Previous efforts to map useful traits of Jatropha were performed either by the interspecific QTL mapping or the candidate gene association study for one trait with limited number of markers. Limit of

the number of markers and accessions probably restricted the association study. This study is the first attempt to perform association analysis in Jatropha on the whole genome level (59 markers distributed in 11 genetic linkage groups), and six markers were found significantly associated with trait, these markers could be used as scientific basis for further molecular breeding and MAS.

Origin characterization of Jatropha has long been a hot topic for researchers. Recent reports identified that the Chiapas might be the origin, as its higher genetic diversity over that in other Mexican state (Pecina-Quintero et al. 2011 and 2014). Furthermore, in Chapter 2 of this thesis, the center of origin of Jatroph had been narrowed to the Chiapas Central Depression. In this chapter, Chiapas contained more elite alleles than that in other Mexican states (Table 3.5), which was consistent with the findings above.

Association study can disclose genetic variations related to traits, and also can give hints of candidate genes causal for phenotypic variations of the traits. It is a better approach than gene expression analysis as changes in the genetic architecture are inheritable and stable, or QTL mapping as it saves time and resources because it does not require bi-parental progenies. The first attempt to perform association study in Jatropha was done by Sharma and Chauhan (2012). They designed markers from genes

that are responsible for fatty acid biosynthesis in seeds of castor bean, and among twelve SSR markers they examined, only one was detected to have polymorphism. The polymorphism ratio (8.3%) was lower than that in this study (up to 13.4%). In their case, one significant MTA was detected for the oil content with SNPs in an exon-intron junction of the stearoyl desaturase gene. Lower polymorphism and less number of MTAs than that in this study were probably caused by genotypes that they used were from India where has been shown to have no or low level of polymorphism in Jatropha (Montes-Osorio et al. 2014). In comparison, samples in this study were collected from Mexico, and the sample size of accessions from Chiapas was over half of the total size. Because Jatropha varieties in Mexico, especially those in the state of Chiapas, were shown to have higher genetic diversity than others (Pecina-Quintero et al. 2011and 2014; see also Chapter 2), thus the association panel in this study is the one that possesses genetic and phenotypic diversity high enough to detect several MTAs.

The marker JCG0061, which was associated with both of individual fruit weight (g) and seed yield (g) per plant, was located in QTL of female flower number. It indicated that polymorphism in JCG0061 can cause changes in female flower number. In this study, although JCG0061 was not associated with female flower number per plant, seed yield (g) per plant was highly correlated with female flower number per

plant (Table 3.2). These suggest that JCG0061 is probably linked to the gene(s) that control female flower number, which should affect seed number, and then would contribute to the seed-yield variation in the Jatropha population. Interestingly, individual fruit weight was highly correlated with individual seed weight, which may also affect seed yield. Besides, although the other five associated markers were not located in any published QTL, it is worth noting that these QTLs were found in the population derived by interspecific crossing, thus these QTLs may not reliable. It is also revealed that a considerable high level of genetic variability was found in Chiapas, so in order to perform QTL mapping with high accuracy, it was also accessible to use Chiapas accessions which are expected to get intraspecific populations with high genetic and phenotypic variations.

Further investigation to explain the variation of the six traits of Jatropha by the associated markers was performed, but no annotated yield-related gene was found in the scaffolds that the markers existed. This may be due to the small length of the scaffolds (208 kb at maximum), or the causal genes that locating in different chromosome. In order to find the linked candidate genes, it is necessary to expand scaffolds or explore more DNA markers.

Breeding study of Jatropha is still in infancy; further association mapping is

critical to guarantee a successful breeding. Although association mapping is receiving major attention for genetic studies for QTL in almost all major crops, the level of success in association mapping for crop development is not compatible to that in the area of human health care for diagnosis of complex human diseases (Gupta et al. 2014). There are still concerns and issues to be addressed; type of materials, population structure, model selection, rare marker alleles, rare genetic variants, multiple testing corrections, background genotypes, and environmental effects. As well as genetic factors, initial emphasis should be given on environmental effects. Epigenetic variations, which include DNA methylation, post-translational modifications of the histone proteins and small RNAs, are also important in order to address the non-genetic factors. Note that environmental factors may have a considerable effect on the different performance of epigenetic modifications. To quantify differences in the epigenetic level by using kinds of biological techniques is necessary to address some phenomena that cannot be fixed by genetic reasons. Epigenetic changes have been found to affect many traits in plants including flowering time, leaf development photosynthetic characteristics and stress tolerance (Ci et al. 2016, Song et al. 2015). Especially in case of Jatropha, it is believed that the epigenetic marker can be an efficient tool to figure out the considerable phenotypic variation detected in African and Asian accession, because genetic variance

in Asia and Africa is very narrow (Sato et al. 2011, Montes-Osorio et al. 2014, also see Chapter 2). This may help breeders to uncover the obstacle that block the success in recent breeding projects in Africa and Asia.

This study could address associations between genetic and phenotypic variations of Jatropha. It would be because all the accessions were sampled in Mexico where the genetic variation is high, and all the phenotypic data used in this study was scored at the same experimental station, both of which would have reduced the effect of environmental factors. These should have made the phenotypic variance in the study largely caused by genetic factors. Results in this study will offer important basis for further breeding in Jatropha, by using the markers to select progenies with the improved seed yield, especially for crossing between Asian and African accessions and Mexican ones. Potential next step will be identification of more numbers of associated markers and responsible genes for the seed yield by doing a larger scale of the analysis.

3.5 Summary

This study is the first trial using Mexican accessions to explore yield-associated markers of Jatropha in a large scale. Six markers that significantly associated with yield-related traits were identified, and they could be applied in the MAS. It was found that one of the markers, JCG0061, located in an agriculturally-important QTL, which

controls the female flower number of Jatropha. Identification of responsible genes that linked to the six associated markers would be important. A more comprehensive association study of yield-related traits, which covers a huge number of DNA markers and a wide collection of Jatropha and phenotypic data, would be required to identify more numbers of useful markers and responsible genes.

Chapter 4

General conclusion

Jatropha is a promising biofuel producing plant. However, commercial cultivations of Jatropha were mainly failed in Africa and Asia and breeding program there always ended up with unsuccess. To support the breeding program in Africa and Asia and finally improve the seed yield of Jatropha, a worldwide genetic diversity assessment and an association study were conducted. In Chapter 2, significant high genetic and phenotypic variations were found in central Chiapas and three genetic groups were identified in Mesoamerica. Combining with local topography and climate of Chiapas, it was concluded that the Chiapas Central Depression was the center of origin, and had a diverse genetic resource of Jatropha that can be used in the breeding program. Three Mesoamerican accessions (one from Chiapas and two from Veracruz) that were the genetically closest to African and Asian Jatropha were found. These results allowed to presume that the voyage of Jatropha from Chiapas to the Old World by transmission from the Port of Veracruz. Human selection during transmission might cause the low yield of Jatropha in Africa and Asia nowadays. In Chapter 3, 110 Mexican accessions were applied in the association study. Eight significant MTAs between six markers and six yield-related traits were identified. Those markers would be effective tools in the marker-assisted selection.

Future prospect

It is concluded that elite Jatropha varieties could be developed by combining results of the genetic diversity analysis and association study. The following strategies for genetic improvement in African and Asian Jatropha were proposed. First, make wide crosses between elite lines from each country in Africa and Asia, which should have their epigenetic adaptations to the climatic conditions, and accessions of Group A from central Chiapas, which are genetically distant from African and Asian Jatropha and have the largest genetic diversity. Not only hybrid vigor, but also some extent of phenotypic variations which are derived from the highest heterozygosity of Group A, would be expected in F_1 progenies. Second, select elite F_1 or F_2 progenies by trait-associated markers in their young age. Finally, use vegetative propagation (Behera et al., 2014) of selected progenies to preserve excellent lines in the heterozygous status. Such elite Jatropha will contribute to the biofuel production and help in the mass production of biofuel in Africa and Asia, and finally help in the fight against climate change.

In association study, five trait associated markers were not found in published QTLs of Jatropha. No annotated genes near the associated markers, which were responsible for yield, were found. To further delimit the genome regions underlying yield-related traits of Jatropha and to reveal new QTL, a wide collection of DNA

markers, large number of accessions, and phenotypic data collection across multiple years and locations should be required. High-throughput genotyping, as well as correct and robust phenotyping are also necessary. With these resources, association analysis with candidate genes including those for development of inflorescence and fruit, vegetative phase change and seed size could be performed. Broader collection of traits, such as oil amount and constitution, are also necessary for a further comprehensive association study.

Reference

Agbogidi OM, Akparobi SO, Eruotor PG. 2013. Health and environmental benefits of *Jatropha curcas* linn. App. Sci. Report 1: 36-39.

Alipour A, Tsuchimoto S, Sakai H, Ohmido N, Fukui K. 2013. Structural characterization of copia-type retrotransposons leads to insights into the marker development in a biofuel crop, *Jatropha curcas* L.. Biotechnol. Biofuels 6: 129.

Ambrosi DG, Galla G, Purelli M, Barbi T, Fabbri A et al. 2010. DNA Markers and FCSS analyses shed light on the genetic diversity and reproductive strategy of *Jatropha curcas* L. Diversity (Basel) 2: 810-836.

Azam MM, Waris A, Nahar NM. 2005. Prospects and potential of fatty acid methyl esters of some non-traditional seed oils for use as biodiesel in India. Biomass Bioenergy 29: 293-302.

Basha SD, George F, Makkar HPS, Becher K, Sujatha M. 2009. A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. Plant Sci. 176: 812-823.

Baudez C, Picasso S. 1987. *Les Cités perdues des Mayas*. Gallimard, Paris. (Ochiai K, 1991, *Maya civilization, A Second of Lost Cities* (In Japanese), Tokyo.)

Behera LK, Nayak MR, Nayak D, Jadeja DB. 2014. In vitro mass multiplication of Jatropha (*Jatropha curcas* L.) through axillary bud culture. Journal of Applied & Natural Science 6: 189-192.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y et al. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Bryant SK. 2014. *Rivers of Gold, Lives of Bondage: Governing through Slavery in Colonial Quito*. University of North Carolina Press, USA, pp. 128.

Carvalho CR, Clarindo WR, Pracxa MM, Araujo FS, Carels N. 2008. Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. Plant Sci. 174: 613-617.

Ci D, Song Y, Du Q, Tian M, Han S et al. 2016. Variation in genomic methylation in natural populations of *Populus simonii* is associated with leaf shape and photosynthetic traits. J. Exp. Bot. 67: 723-37.

Conrad J. 2007. Vegetation zones of Chiapas. Backyard nature, Mexico.

(http://www.backyardnature.net/chiapas/veg-map.htm)

Dahmer N, Schifino-Wittmann MT, Dias LAS. 2009. Chromosome numbers of *Jatropha curcas* L.: an important agrofuel plant. Crop Breed. Appl. Biotechnol. 9: 386-389.

Dias LAS, Missio RF, Dias DCFS. 2012. Antiquity, botany, origin and domestication of *Jatropha curcas* (Euphorbiaceae), a plant species with potential for biodiesel production. Genet. Mol. Res. 11: 2719-2728.

Divakara BN, Upadhyaya HD, Wani SP, Gowda CLL. 2010. Biology and genetic improvement of *Jatropha curcas* L.: a review. Appl. Energy 87: 732-742.

Duke JA. 1985. Medicinal plants. Science 229: 1036.

Elbehri A, Segerstedt A, Liu P. 2013. Biofuels and sustainability challenges: a global assessment of sustainability issues, trends and policies for biofuels and related feedstocks. Food and Agricultural Organization of United Nations, Rome.

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.

Fairless D. 2007. Biofuel: The little shrub that could – maybe. Nature 449: 652-655.

Fargione J, Hill J, Tilman D, Polasky S, Hawthorne P. 2008. Land clearing and the biofuel carbon debt. Science 319: 1235-1238.

Goudet J. 2001. Fstat, Version 2.9.3. University of Lausanne, Switzerland. (http://www2.unil.ch/popgen/softwares/fstat.htm)

Gupta PK, Kulwal PL, Jaiswal V. 2014. Association mapping in crop plants: opportunities and challenges. Adv. Genet. 85: 109-147.

Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2: 618-620.

Heller J. 1996. *Physic Nut (Jatropha curcas L.): Promoting the Conservation and Use of Underutilized and Neglected Crops*. Institute of Plant Genetics and Crop Plant Research, Rome.

Hirakawa H, Tsuchimoto S, Sakai H, Nakayama S, Fujishiro T et al. 2012. Upgraded genomic information of *Jatropha curcas* L. Plant Biotechnol. 29: 123-130.

Jiang GL. 2013. Molecular markers and marker assisted breeding in plants. Andersen SB (eds) Plant breeding from laboratories to fields (Rijeka: INTECH), 45-83.

King AJ, Montes LR, Clarke JG, Affleck J, Li Y et al. 2013. Linkage mapping in the oilseed crop *Jatropha curcas* L. reveals a locus controlling the biosynthesis of phorbol esters which cause seed toxicity. Plant Biotechnol. J. 11: 986-96.

Langella O. 2002. POPULATIONS, Version 1.2.29. CNRS, France. (http://bioinformatics.org/~tryphon/populations/)

Lin J, Zhou X, Wang J, Jiang P, Tang K. 2010. Purification and characterization of curcin, a toxic lectin from *Jatropha curcas*. Pre. Biochem. Technol. 40: 107-118.

Liu P, Wang CM, Li L, Sun F, Liu P et al. 2011. Mapping QTLs for oil traits and eQTLs for oleosin genes in Jatropha. BMC Plant Biol. 11: 132.

Moncada P, Mart ńez CP, Borrero J, Chatel M, Gauch Jr H et al. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* × *Oryza rufipogon* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. 102: 41-52.

Montes JM, Technow F, Martin M, Becker K. 2014. Genetic diversity in *Jatropha curcas* L. assessed with SSR and SNP markers. Diversity 6: 551-566.

Montes-Osorio LR, Torres Salvador AF, Jongschaap RE, Azurdia Perez CA, Berduo Sandoval JE et al. 2014. High level of molecular and phenotypic

biodiversity in *Jatropha curcas* from Central America compared to Africa, Asia and South America. BMC Plant Biol. 14: 77.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19: 153-170.

Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ouattara B, Ndir KN, Gueye MC, Diedhiou I, Barnaud Adeline et al. 2014. Genetic diversity of *Jatropha curcas* L. in Senegal compared with exotic accessions based on microsatellite markers. Genet. Resour. Crop Evol. 61: 1039-1045.

Page RDM. 1996. TreeView: An application to display phylogenetic trees on personal computers, Bioinformatics 12: 357-358.

Parawira W. 2010. Biodesel production from *Jatropha curcas*: A review. Sci. Res. Essay 5: 1796-1808.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288-295.

Pecina-Quintero V, Anaya-López JL, Zamarripa A, Montes N, Nuñez CA et al. 2011. Molecular characterisation of *Jatropha curcas* L. genetic resources from

Chiapas, México through AFLP markers. Biomass Bioenergy 35: 1897-1905.

Pecina-Quintero V, Anaya-López JL, Zamarripa-Colmenero A, Solís-Bonilla JL. Jiménez-Becerril MF et al. 2014. Genetic structure of *Jatropha curcas* L. in Mexico and probable center of origin. Biomass Bioenergy 60: 147-155.

Prasad DMR, Izam A, Khan MMR. 2012. *Jatropha curcas*: plant of medical benefits. J. Med. Plant Res. 6: 2691-2699.

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

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rambaut A, Drummond A. 2014. FigTree, Version 1.4.2. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. (http://tree.bio.ed.ac.uk/software/figtree/)

Sabandar CW, Ahmat N, Jaafar FM, Sahidin I. 2013. Medicinal property, phytochemistry and pharmacology of several Jatropha species (Euphorbiaceae): a review. Phytochemistry 85: 7-29.

Sanderson K. 2009. Wonder weed plans fail to flourish. Nature 461: 328-329.

Sato S, Hirakawa H, Isobe S, Fukai E, Watanabe A et al. 2011. Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. DNA Res. 18: 65-76.

Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A et al. 2008. Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land use change. Science 319: 1238-1240.

Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). Biometrika 52: 591-611.

Sharma A, Chauhan RS. 2012. Identification and association analysis of castor bean orthologous candidate gene-based markers for high oil content in *Jatropha curcas*. Plant Mol. Biol. Rep. 30: 1025-1031.

Si L, Chen J, Huang X, Gong H, Luo J. 2016. *OsSPL13* controls grain size in cultivated rice. Nat. Genet. 48: 447-56.

Song Y, Liu L, Feng Y, Wei Y, Yue X et al. 2015. Chilling- and freezing-induced alterations in cytosine methylation and its association with the cold tolerance of an alpine subnival plant, *Chorispora bungeana*. PLoS One. 10: e0135485.

Sun F, Liu P, Ye J, Lo LC, Cao S et al. 2012. An approach for Jatropha improvement using pleiotropic QTLs regulating plant growth and seed yield. Biotechnol. Biofuels. 5: 42

Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK et al. 2009. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L. a biofuel plant. Plant Sci. 176: 505-513.

Tenenbaum DJ. 2008. Food vs. fuel: diversion of crops could cause more hunger. Environ. Health Perspect. 116: A254-A257.

The royal society. 2008. Sustainable biofuels: prospects and challenges. (https://royalsociety.org/~/media/Royal_Society_Content/policy/publications/2008/7980.pdf)

Thomas H. 2006. *The Slave Trade: the Story of the Atlantic Slave Trade.* London: Phoenix.

Umamaheswari D, Paramathma M, Manivannan N. 2010. Molecular genetic diversity analysis in seed sources of Jatropha (*Jatropha curcas* L) using ISSR markers. Elect. J. Plant Breed. 1: 268-278.

Von Maltitz G, Gasparatos A, Fabricius C. 2014. The rise, fall and potential resilience benefits of Jatropha in Southern Africa. Sustainability 6: 3615-3643.

Von Maltitz GP, Setzkorn KA. 2013. A typology of Southern African biofuel feedstock production projects, Biomass Bioenergy 59: 33-49.

Wang CM, Liu P, Yi C, Gu K, Sun F, Li L et al. 2011. A first generation microsatellite- and SNP-based linkage map of Jatropha. PLoS One. 6: e23632.

Wu P, Zhou C, Cheng S, Wu Z, Lu W et al. 2015. Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. Plant J. 81: 810-21.

Yang W, Guo Z, Huang C, Wang K, Jiang N et al. 2015. Genome-wide association study of rice (*Oryza sativa* L.) leaf traits with a high-throughput leaf scorer. J. Exp. Bot. 66: 5605-15.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203-208.

Yue GH, Sun F, Liu P. 2013. Status of molecular breeding for improving *Jatropha curcas* and biodiesel. Renewable and Sustainable Energy Reviews 26: 332-343.

Zhang J, Song Q, Cregan PB, Jiang GL. 2016. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). Theor. Appl. Genet. 129: 117-30.

Zhu C, Gore M, Buckler E, Yu J. 2008. Status and prospects of association mapping in plants. The Plant Genome 1: 5-20.

List of publications

Haiyan Li, Suguru Tsuchimoto, Kyuya Harada, Masanori Yamasaki, Hiroe Sakai, Naoki Wada, Atefeh Alipour, Tomohiro Sasai, Atsushi Tsunekawa, Hisashi Tsujimoto, Takayuki Ando, Hisashi Tomemori, Shusei Sato, Hideki Hirakawa, Victor Pecina-Quintero, Alfredo Zamarripa, Primitivo Santos, Adel Hegazy, Abdalla Mohamed Ali, and Kiichi Fukui. Genetic tracing of *Jatropha curcas* L. from its Mesoamerican origin to the world. Front. Plant Sci. 8: 1539.

Haiyan Li, Suguru Tsuchimoto, Kyuya Harada, Masanori Yamasaki, Hiroe Sakai, Naoki Wada, Atefeh Alipour, Tomohiro Sasai, Atsushi Tsunekawa, Hisashi Tsujimoto, Takayuki Ando, Hisashi Tomemori, Shusei Sato, Hideki Hirakawa, Victor Pecina-Quintero, Alfredo Zamarripa, and Kiichi Fukui. Association study of seed-yield related traits for *Jatropha curcas* L. in Mexico (Tropical Agriculture and Development, in press)

Haiyan Li, Suguru Tsuchimoto, Kyuya Harada and Kiichi Fukui. 2017. The Genome-Wide Association Study. The Jatropha Genome, Tsuchimoto S (eds), 159-176, Springer.

Acknowledgements

I would like to express my deepest and sincere gratitude to Prof. Kiichi Fukui. Thanks very much for his offering me an opportunity to pursue my PhD in Japan and for his guidance, support and patience throughout my entire PhD study.

I am very grateful to my advisor, Associate Professor Suguru Tsuchimoto for his patience, encouragement, guidance and support during the past three years. He kindly offered me lots of advices and help. I am grateful to Dr. Kyuya Harada for offering me lots of scientific help and advices. I would also express my sincere thanks to Assistant Professor Hiroe Sakai and Assistant Professor Naoki Wada for providing me valuable help in my research work. I also want to express my thanks to Professor Masanori Yamasaki of Kobe University for his kindly advices and help on the research plan of this study.

I would like to acknowledge all the coauthors. Thanks to Professor Atsushi Tsunekawa, Professor Hisashi Tsujimoto, and Dr. Hisashi Tomemori for their collaborations in organizing this project and supporting the donation lab. Thanks to Professor Takayuki Ando for his help in offering us various useful informations about Mexican Jatropha. Thanks to Dr. Satoshi Tabata, Professor Shusei Sato and Dr. Hideki Hirakawa for their support for the marker analysis. I also want to give thanks to Dr.

Victor Pecina-Quintero, Dr. Alfredo Zamarripa for offering us Mexican accessions and phenotypic data. Thanks to Dr. Primitivo Santos from University of the Philippines Los Banos for providing us Jatropha samples from Philippines, and Professor Yoshiaki Kitaya for providing Jatropha samples from Vietnam. Thanks to Professor Adel Hegazy and Professor Abdalla Mohamed Ali for their offering us Egyptian and Sudanese accessions, respectively.

I am particular thankful to Prof. Toshiya Muranaka, Prof. Kazuhito Fujiyama, Prof. Hajime Watanabe for their insightful scientific inputs, criticisms and valuable suggestions.

I would like to express my thanks to Associate Professor Susumu Uchiyama, Assistant Professor Hideaki Takata and all members in Laboratory of Dynamic Cell Biology. With your help, support and companion, I spent a pleasant time in the lab.

I would like to acknowledge the contributions of all of collaborators in this project, especially Corporate Social Responsibility (CSR) Foundation of the Sumitomo Electric Industries (SEI), Ltd and Arid Land Research Center, Tottori University, Japan.

I would like to give my special thanks to my beloved husband, Mr. Minglong Xu who always support, encourage and help me. My particular appreciation will give to my parents for their continuous encourage and support.