C. R. Biologies xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Molecular biology and genetics/Biologie et génétique moléculaires

Genetic variation and phylogenetic relationship analysis of *Jatropha curcas* L. inferred from nrDNA ITS sequences

Guo-Ye Guo [a,b], Fang Chen [a,*], Xiao-Dong Shi [a], Yin-Shuai Tian [a], Mao-Qun Yu [b], Xue-Qin Han [c], Li-Chun Yuan [c], Ying Zhang [d]

[a] *Key Laboratory of Bio-resources and Eco-environment, Ministry of Education, College of Life Science, Sichuan University, 610064 Chengdu, Sichuan, PR China*
[b] *Chengdu Institute of Biology, Chinese Academy of Sciences, 610041 Chengdu, Sichuan, PR China*
[c] *Tropical Eco-agriculture Institute, Yunnan Academy of Agricultural Sciences, 651300 Yuanmou, Yunnan, PR China*
[d] *College of life science, Hainan Normal University, 571158 Haikou, Hainan, PR China*

ARTICLE INFO

Article history:
Received 14 March 2016
Accepted after revision 12 June 2016
Available online xxx

Keywords:
Jatropha curcas
nrITS
Genetic variation
Phylogenetic relationship
Migration route

ABSTRACT

Genetic variation and phylogenetic relationships among 102 *Jatropha curcas* accessions from Asia, Africa, and the Americas were assessed using the internal transcribed spacer region of nuclear ribosomal DNA (nrDNA ITS). The average G + C content (65.04%) was considerably higher than the A + T (34.96%) content. The estimated genetic diversity revealed moderate genetic variation. The pairwise genetic divergences (GD) between haplotypes were evaluated and ranged from 0.000 to 0.017, suggesting a higher level of genetic differentiation in Mexican accessions than those of other regions. Phylogenetic relationships and intraspecific divergence were inferred by Bayesian inference (BI), maximum parsimony (MP), and median joining (MJ) network analysis and were generally resolved. The *J. curcas* accessions were consistently divided into three lineages, groups A, B, and C, which demonstrated distant geographical isolation and genetic divergence between American accessions and those from other regions. The MJ network analysis confirmed that Central America was the possible center of origin. The putative migration route suggested that *J. curcas* was distributed from Mexico or Brazil, via Cape Verde and then split into two routes. One route was dispersed to Spain, then migrated to China, eventually spreading to southeastern Asia, while the other route was dispersed to Africa, via Madagascar and migrated to China, later spreading to southeastern Asia.

1. Introduction

Jatropha curcas L. (Euphorbiaceae) is a perennial small tree or large shrub that is distributed across semiarid tropical and subtropical regions of the Americas, Africa, and Asia [1,2]. *J. curcas* is a monoecious plant with unisexual flowers and has a high rate of self-pollination [3]. *J. curcas* has wide adaptability to a range of ecological habitats and climatic conditions and high potential for greening and eco-rehabilitation of wastelands [3]. *J. curcas* has recently become popular due to its potential economic value, as it has a relatively high content of seed oil, which can be used as biodiesel [4].

The geographic center of origin of *J. curcas* remains controversial. Botanists have hypothesized that *J. curcas* was distributed by Portuguese seafarers from Central America, via Cape Verde to Africa and Asia [3]. The plant has been widely disseminated and has become naturalized in many tropical and subtropical regions. Its current geographical distribution is vague and insufficient for

* Corresponding author.
E-mail address: 890chenfang@sina.com (F. Chen).

http://dx.doi.org/10.1016/j.crvi.2016.06.004
1631-0691/

recovering genetic structure due to the interference of forces like domestication and potential genotype- environment interactions [5]. A better understanding of the geographic origin, genetic structure, intraspecific relatedness, and differentiation among *J. curcas* germplasms would provide insights into population structure, breeding programs, and conservation of genetic resources. Molecular approaches are powerful for tracking geographic origins and introduction histories of exotic species, reconstructing evolutionary relationships, and assessing genetic variation in the colonization process [6].

Previous studies on the genetic variation in *J. curcas* germplasms have been conducted using various types of molecular markers such as AFLPs [7,8], ISSRs [9], RAPDs [10], and SSRs [11]. These studies demonstrated that genetic variation in Central America is much higher than that in other regions [12–14]. However, few studies have reported the genetic variation and structure of *J. curcas* on all continents simultaneously. The internal transcribed spacer region (ITS1-5.8S-ITS2) of nuclear ribosomal DNA (nrDNA) is one of the most popular and effective genetic markers for the inference of phylogenetic relationships and evolutionary studies in plants [15].

In the present study, we sampled 102 accessions of *J. curcas* from 11 countries that are representative of the plant's distribution across Asia, Africa, and the Americas. Using ITS sequencing, we had the following aims:

- evaluate the genetic variation and genetic diversity level within accessions and among haplotypes;
- determine phylogenetic relationships and estimate genetic divergence of *J. curcas* accessions;
- explore the center of origin of *J. curcas* and its potential dispersal route.

2. Materials and methods

2.1. Plant materials

A total of 102 accessions of *J. curcas* were collected from the following 11 countries: 72 from China, 7 from India, 7 from Vietnam, 5 from Burma, 4 from Thailand, 2 from Indonesia, 1 from Laos, 1 from Zambia, 1 from Burkina Faso, 1 from Mali, and 1 from Brazil. The accessions from China are highly representative of within-country dispersal, with 69 accessions collected from 6 provinces in southwestern and southern China, including 3 hybrid varieties. These materials were kindly provided by the Tropical Eco-Agriculture Institute of the Yunnan Academy of Agricultural Sciences and Institute of Tropical Bioscience and Biotechnology, CATAS. The ITS spacer region of 102 *J. curcas* accessions were sequenced in this study, and 20 additional sequences (from plants in India, Madagascar, Cape Verde, Spain, Brazil, and Mexico) were obtained from GenBank (National Center for Biotechnology Information; NCBI). The sample code, location, latitude, longitude, altitude, country, haplotype, and GenBank accession number of all accessions are listed in Table 1. *Jatropha gossypifolia* was used as an outgroup for phylogenetic analysis [1].

2.2. DNA extraction, PCR amplification, and sequencing

Total genomic DNA was extracted with a Plant Genomic DNA Kit (TIANGEN Biotech, Beijing, China). The nrITS spacer regions were amplified using external primers of the following sequences: ITS1 (TCCGTAGGTGAACCTGCGG) and ITS4 (TCCTCCGCTTATTGATATGC) [16]. PCR was conducted in a 50-μL mixture reaction volume containing 5.0 μL 10 × Taq Buffer, 1 μL dNTP Mix (10 mM each), 0.5 μL Taq DNA polymerase (5 U/μL), 3.0 μL 25 mM $MgCl_2$, 2.0 μL of each primer (10 μM), 2.0 μL of genomic template DNA (2.5 ng/μL), and additional ddH_2O to the final volume (Vazyme Biotech, Nanijing, China). The PCR amplification procedure for nrITS was 3 min at 94 °C for predenaturation followed by 35 cycles of 1 min at 94 °C for denaturation, 1 min annealing at 56 °C, and 1 min at 72 °C for primer extension; this was followed by a final primer extension of 10 min at 72 °C on a BIO-RAD S1000^{TM} Thermal cycler. Purified samples were sequenced by BGI China. All sequences generated in this study were deposited in GenBank under accession numbers: KP190940 _ KP191040 and KP191042 (Table 1).

2.3. Molecular variability and demographic analysis

Multiple sequences were aligned using the Clustal W algorithm in MEGA 6.06 [17] and refined by manual adjustment with BioEdit version 7.2.5 [18]. The homogeneity of the base composition was detected for Id-test, nucleotide substitutions, and transition/transversion ratio and was calculated with MEGA version 6.06 [17]. Substitution saturation was estimated and transversions were calculated under the F84 model using the software DAMBE version 5.3.8 [19]. DnaSP 5.10 [20] was used to estimate several molecular diversity indices including haplotype diversity (*Hd*), nucleotide diversity (π), Watterson's diversity of segregating sites (θ_w), and the average of nucleotide differences (k). Estimates of evolutionary divergence were calculated using the Maximum Composite Likelihood Model. Selection neutrality was tested to detect historical demographic expansions by Tajima's D, Fu and Li's D, and Fu's Fs methods. The demographic history was assessed by the distribution of pairwise sequence differences (mismatch distribution) and site frequency spectra using the program DnaSP 5.10 [21].

2.4. Phylogeny reconstruction

Bayesian inference (BI) analysis was performed using MrBayes version 3.1.2 [22]. The GTR + G model was identified as the best-fit model using MrModel Test 2.3 [23]. A four-chain Markov Chain Monte Carlo (MCMC) was run for 1,000,000 generations and two simultaneous analyses were performed. Trees were sampled every 100 generations. The first 5000 trees were discarded as burn-in, and the remaining trees were used to construct a 50% majority rule consensus phylogram. Maximum parsimony (MP) analysis was implemented using PAUP version 4.0b10[24]. The following were in effect: heuristic searches with 1000 random addition sequence replicates,

Please cite this article in press as: G.-Y. Guo, et al., Genetic variation and phylogenetic relationship analysis of *Jatropha curcas* L. inferred from nrDNA ITS sequences, C. R. Biologies (2016), http://dx.doi.org/10.1016/j.crvi.2016.06.004