Tree Genetics & Genomes (2015) 11:7
DOI 10.1007/s11295-014-0825-y

ORIGINAL PAPER

Population genetic analysis of *Xylia xylocarpa* (Fabaceae—Mimosoideae) in Thailand

Tanat Wattanakulpakin · Siriluck Iamtham · Kunsiri Chaw Grubbs · Hugo A. Volkaert

Received: 3 February 2014 / Revised: 11 November 2014 / Accepted: 2 December 2014

Abstract The genetic diversity of *Xylia xylocarpa* (Roxb.) W. Theob. var. *kerrii* I. C. Nielsen in Thailand was studied using molecular markers. The results revealed a large diversity within and moderate differentiation between populations that could be correlated with geography. Samples were collected from individual trees in 16 populations. Eight single-locus nuclear gene markers were used to identify different alleles using PCR–single-strand conformation polymorphism (SSCP) combined with sequencing. For easier detection of some alleles, 10 SNPs were converted into cleaved amplified polymorphic sequence (CAPS) assays and 5 SNPs were converted to allele-specific PCR (AS-PCR) assays. From 10 to 33 alleles were identified for the different loci. The effective number of alleles was much lower, ranging between 1.1 and 7.6, as several of the alleles were found at low frequency in just one or a small number of populations. Several loci showed deviation from the Hardy–Weinberg equilibrium. A high level of gene flow was estimated. The largest genetic distances calculated according to Nei (G_{st}) and Jost (D_{est}) were found between populations from the northeast and the west of Thailand. STRUCTURE analysis revealed support for population differentiation in two or three clusters, though genetic distances between the clusters were small. Four populations from western Thailand separated from the others, as did the three populations from northeastern Thailand. The large number of low-frequency alleles would make conservation of genetic resources in this species a difficult task as a large number of populations and individuals will need to be protected.

Keywords *Xylia xylocarpa* · Genetic diversity · SSCP · SNP · AS-PCR · CAPS

Communicated by W. Ratnam

Electronic supplementary material The online version of this article (doi:10.1007/s11295-014-0825-y) contains supplementary material, which is available to authorized users.

T. Wattanakulpakin · S. Iamtham · K. C. Grubbs
Faculty of Liberal Arts and Science, Kasetsart University Kamphaengsaen Campus, Nakhon Pathom 73140, Thailand

T. Wattanakulpakin · S. Iamtham · H. A. Volkaert
Center for Advanced Studies for Agriculture and Food, KU Institute for Advanced Studies, Kasetsart University (CASAF, NRU-KU), Bangkok 10900, Thailand

H. A. Volkaert (✉)
Center for Agricultural Biotechnology, Kasetsart University Kamphaengsaen Campus, Nakhon Pathom 73140, Thailand
e-mail: hugo.v@ku.ac.th

H. A. Volkaert
Center of Excellence in Agricultural Biotechnology Postgraduate Education and Research Development Office Commission on Higher Education, Ministry of Education, Bangkok 10900, Thailand

H. A. Volkaert
NSTDA-BIOTEC, Plant Research Unit, Thailand Science Park, Pathumthanee 12120, Thailand

K. C. Grubbs
Department of Biology, Winthrop University, Rock Hill, SC 29733, USA

Introduction

Forest cover in Thailand has decreased dramatically in recent times. According to an official estimate, the current forest cover is 26 % of the country, down from 53 % in 1961 (Bhumibhamon 1986; Charuppat 1998). In addition, most of the remaining forest fragments have been partly degraded. Forest loss and degradation have been blamed for occasional severe flash floods in mountainous areas in the south and north of the country. Concern about the decrease in forest cover led to the issuance of a complete logging ban in 1989.

Published online: 24 January 2015

More recently, the awareness about the threat of global warming has heightened the interest in preservation of the remaining forests and in replanting trees. For any attempt at conservation or reforestation, it would be useful to have knowledge about the genetic diversity of indigenous forest tree species, their adaptation to the local environment and their responses to changing climate conditions.

Given the high rates of degradation and fragmentation in the deciduous forests in Thailand in the recent past, tree populations are (locally) experiencing contractions in size and may be more affected by demographic stochasticity (i.e. genetic drift) than at any other time in their histories. Altered patterns of gene flow and genetic structure may have implications for mating systems, levels of inbreeding, and the maintenance of genetic diversity. Hamrick (2004) has argued that because of their long-distance pollen flow, long lifespan, and high genetic diversity, forest tree species in general may be less vulnerable to anthropogenic changes than other organisms. However, this may not apply to tropical tree species that are highly dependent on animals for pollination and/or seed dispersal (Pacheco and Simonetti 2000; Wang et al. 2007).

Xylia xylocarpa (Roxb.) W. Theob. var. *kerrii* I. C. Nielsen is a long-lived, medium- to large-sized tree growing scattered in the monsoon deciduous forests in South and Southeast Asia. The species belongs to the family Fabaceae (Leguminosae)—Mimosoideae. Its flowers are heavily visited by a wide variety of pollinating insects, while its seeds are dispersed during the explosive dehiscence of the pods. In Thailand, it can be found in lowland and lower montane deciduous forests north of the Isthmus of Kra. It occurs mixed with other broadleaf hardwoods such as *Tectona grandis*, *Pterocarpus*, *Afzelia*, and *Lagerstroemia*.

The reddish brown wood is hard and durable and makes an excellent material for all kinds of construction, furniture, turnery, and household implements. Even though there are potential economic uses for the species, there have been almost no attempts to commercialize the production of *Xylia*. Since it can be established well in plantations and is relatively fast growing and free of pests and diseases, forest managers in Malaysia have begun to take an interest in growing the species (Josue 2004).

Molecular analysis techniques are now widely used to study genetic diversity in living organisms (Karp et al. 1997; Neale and Kremer 2011). As genetic diversity is the basis for short-term adaptation and long-term evolution, information about genetic diversity and its distribution among natural populations is of fundamental importance for the conservation and utilization of genetic resources (Van Inghelandt et al. 2010). Hence, this study aimed to develop DNA markers and use them to analyse the genetic diversity of *X. xylocarpa* within and among populations in its natural range in Thailand. As no genetic studies of *X. xylocarpa* had ever been done, a primary objective was to investigate the presence of population genetic structure in this species, which should be taken into account when drafting approaches for utilization and conservation of genetic resources.

Materials and methods

Plant materials

X. xylocarpa young leaves, or in a few cases inflorescences, were collected from mature trees growing in their natural habitat in 14 national parks and 2 community forests (Fig. 1 and Table 1). The tissues were air dried. The number of trees sampled in each population ranged from 11 to 44 for a total of 538 trees. The location of each tree was recorded by GPS and sampled trees were at least 100 m apart. In addition, 10 individual trees not associated with a population were sampled from 7 other locations.

DNA extraction

An area of dried leaf corresponding to about 60-mg fresh weight (or an equivalent amount of flowers) was put in a tube with a stainless steel ball and crushed to a very fine powder in a mixer mill (Retsch®, Germany) for 3 min at 30 Hz. Genomic DNA was extracted from the powdered tissue using the Genomic DNA Mini Kit (Plant) (Geneaid, Taiwan), following the manufacturer's protocol. The DNA concentration of the samples was estimated on agarose gel and by NanoDrop 8000 (Thermo Scientific).

Development of gene-based markers

Primers for PCR amplification of specific genes were designed to match conserved codons among genes sequenced in a multitude of other plant species. These primers (Table 2A) contained one or more positions where mixtures of nucleotides were incorporated during synthesis (degenerate PCR primers). The genes chosen to develop the primers were low-copy-number gene families mostly coding for metabolic enzymes, i.e. catalase (*CAT*), isopentenyldiphosphate isomerase (*IDI*), aspartate aminotransferase (*AAT*, *GOT*), sucrose synthase (*SUS*), isocitrate dehydrogenase (*IDH*), leucine aminopeptidase (*LAP*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), isocitrate lyase (*ICL*), cysteine proteinase inhibitor (*CPI*), abscisic acid insensitive 3 (*ABI3*), and *Arabidopsis* auxin response factors 6 and 8 (*ARF6/8*). Between 35 and 147 sequences from 26 to 63 plant species, including genomic sequences, cDNAs, and ESTs, were retrieved from publicly accessible sequence repositories to construct the alignments.

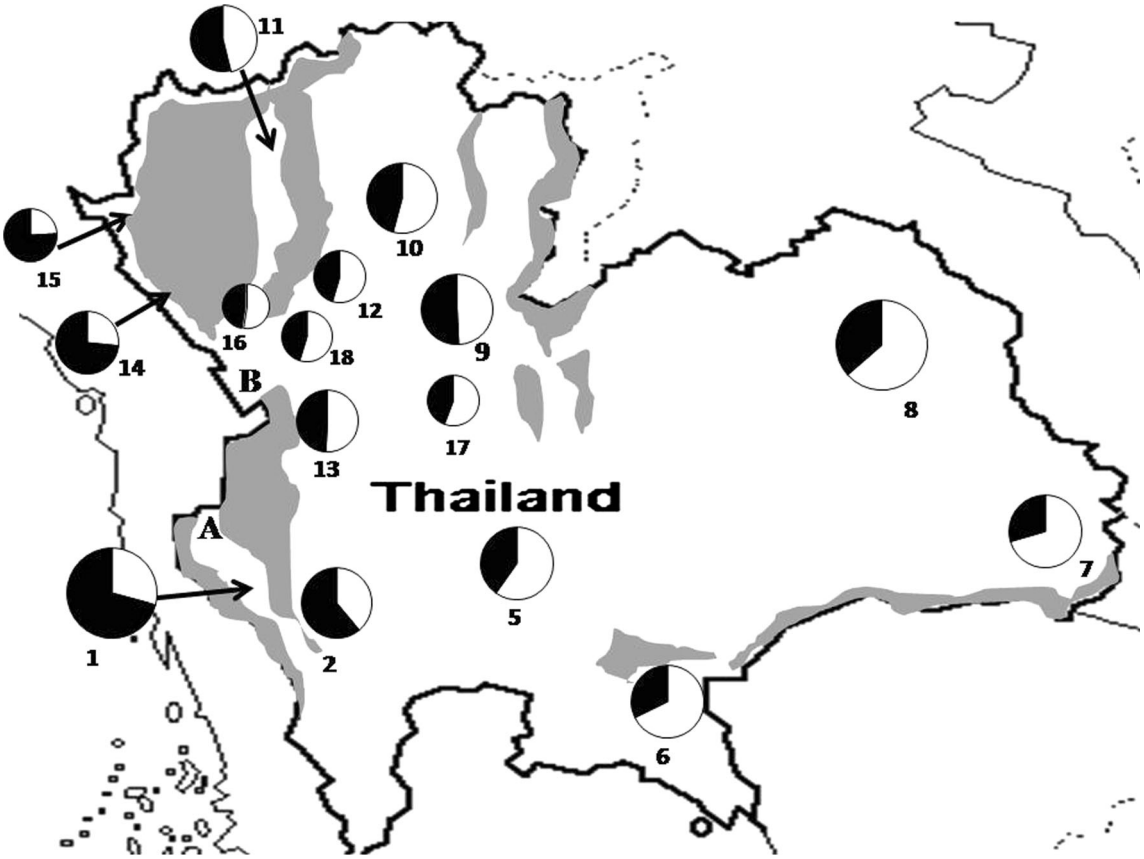

Fig. 1 Location of the forest areas where *Xylia* trees were sampled. The population numbers correspond to those in Table 1. The pie charts correspond to the population-wise cluster assignments in STRUCTURE assuming two clusters. The size of the circles is proportional to the number of individuals in the sampled populations. Areas of more than 1000 m a.s.l. have been indicated by *shading*. *A* Three Pagodas Pass (300 m a.s.l.), *B* Mae Sot gap (650 m a.s.l.)

DNA fragments were PCR amplified from two or three samples using the consensus primer pairs in a total reaction mixture of 25 μl, containing 200 μM dNTPs, 5 pmol of each primer, 1× PCR buffer (50 mM KCl, 10 mM $(NH_4)_2SO_4$, 10 mM Tris·HCl pH 8.3, 0.1 mg/ml bovine serum albumin (BSA); RBC, Taiwan) with $MgCl_2$ adjusted to 2 mM, 0.3 units of Taq DNA polymerase (Fermentas, Lithuania), and 10 ng of genomic DNA template. Amplification was carried out at 94 °C for 3 min, followed by 40 cycles of 45 s at 94 °C, 45 s at the appropriate annealing temperature (44–51 °C), 1 min 30 s at 72 °C, and a final extension at 72 °C for 5 min. PCR products were visualized on agarose gel.

PCR reactions showing reproducible fragment bands were selected and purified using a PCR fragment cleanup kit (Geneaid, Taiwan). The purified fragments were then cloned into a pGEM-T® plasmid vector (Promega, USA). Blue–white colony selection was used to identify transformants. Individual colonies were picked and checked for the presence of an insert by PCR amplification using the degenerate primer pairs. The remainder of the same colony was grown overnight, and plasmids were extracted using High-Speed Plasmid Mini Kit (Geneaid, Taiwan) following the manufacturer's instruction and then sent for sequencing by a service provider (1Base, Malaysia).

Sequence analysis and design of specific primers

The sequences were compared with those from other plant species to check that the targeted gene had been obtained. Intron–exon boundaries (GT–AG) were determined in the *X. xylocarpa* sequences by comparison with cDNA sequences from other species. Specific primer pairs for the amplification of unique DNA fragments from *X. xylocarpa* were then designed based on the obtained sequences from each gene family in such a way that different loci in each gene family could be amplified separately from the *X. xylocarpa* genome (Table 2B).

Single-strand conformation polymorphism

For the single-strand conformation polymorphism (SSCP) assay, DNA fragments were PCR amplified in a total reaction mixture of 10 μl, containing 200 μM dNTPs, 5 pmol of each primer, 1× PCR buffer (50 mM KCl, 10 mM $(NH_4)_2SO_4$,

Table 1 Location of the sampled populations of *Xylia xylocarpa*

Number	Name, province	Region	*N*	Location		m a.s.l.
1	Lam Khlong Ngu National Park, Kanchanaburi	West	41	14° 53′ 26″ N	98° 49′ 52″ E	555–625
2	Erawan National Park, Kanchanaburi	West	35	14 23′ 14″ N	99° 04′ 57″ E	556–661
5	Vichienburi, Phetchabun	Central	23	15° 39′ 03″ N	101° 01′ 35″ E	129–338
6	Pang Sida National Park, Sakaew	Northeast	41	14° 00′ 52″ N	102° 34′ 30″ E	123–171
7	Phu Jong Nayoi National Park, Ubon Ratchathani	Northeast	30	14° 23′ 26″ N	105° 08′ 41″ E	130–413
8	Phupha Lek National Park, Sakonnakhon	Northeast	37	17° 16′ 07″ N	103° 27′ 34″ E	253–603
9	Khlong Tron National Park, Uttaradit	North	44	17° 36′ 10″ N	100° 33′ 31″ E	170–244
10	Thamphatai National Park, Lampang	North	41	18° 36′ 51″ N	99° 53′ 50″ E	330–376
11	Pha Daeng National Park, Chiangmai	North	37	19° 30′ 51″ N	98° 58′ 35″ E	431–755
12	Doi Jong National Park, Lampang	North	41	17° 55′ 03″ N	99° 15′ 00″ E	265–389
13	Khlong Wang Jao National Park, Kamphaeng Phet	Northwest	36	16° 29′ 55″ N	99° 10′ 04″ E	216–414
14	Mae Ngao National Park, Mae Hong Son	Northwest	36	17° 50′ 34″ N	97° 58′ 15″ E	150–217
15	ThaTaFang area, Mae Hong Son	Northwest	11	18° 02′ 49″ N	97° 45′ 57″ E	125–578
16	Mae Ping National Park, Lampang	North	28	17° 35′ 59″ N	98° 51′ 22″ E	333–679
17	Kaeng Jed Kwae National Park, Phitsanulok	North	25	17° 12′ 25″ N	100° 24′ 50″ E	89–182
18	Wiang Kosai National Park, Phrae	North	32	17° 42′ 32″ N	99° 19′ 49″ E	265–354
Total			538			89–755
Other locations			10	–	–	12–380

N number of individuals sampled, *m a.s.l.* meters above average sea level

10 mM Tris·HCl pH 8.3, 0.1 mg/ml BSA; RBC, Taiwan) with $MgCl_2$ adjusted to 2 mM, 0.3 units of Taq DNA polymerase (Fermentas, Lithuania), and 10 ng of genomic DNA template. Amplification was carried out at 94 °C for 3 min, followed by 40 cycles of 45 s at 94 °C, 45 s at the appropriate annealing temperature (46–51 °C, Table 2B), 1 min at 72 °C, and a final extension at 72 °C for 5 min. Non-denaturing polyacrylamide gels (6–8 % polyacrylamide gels with a ratio of 99:1 acrylamide/bis-acrylamide) were prepared in 0.6× Tris–borate–EDTA (TBE) buffer. The PCR product (1.5 μl) was directly mixed with four volumes of a loading dye (95 % formamide, 0.025 % bromophenol blue, 0.025 % xylene cyanol, and 10 mM NaOH). The double-stranded DNA was denatured by heating at 95 °C for 10 min and cooled on ice to stabilize single strands. Aliquots (2.5 μl) were loaded on the gel. The gels were run in a refrigerator (4–8 °C) at a constant power of 10–20 W for 12–24 h depending on the size of the PCR product with 1× TBE as a running buffer. The DNA fragments were revealed by silver staining (Bassam et al. 1991).

Allele-specific assays

The PCR amplification products of one to three representative samples of the different alleles identified on SSCP gels and from some fragments that did not give clear SSCP genotyping results were sequenced without further purification using an internal primer (Table 2C) by a sequencing service provider (1Base, Malaysia). The haplotypes deduced from the sequence data were aligned and SNPs were detected. Based on sequence data, five SNPs were developed into allele-specific PCR (AS-PCR) amplification assays using the BatchPrimer3 website (http://probes.pw.usda.gov/batchprimer3/) incorporating an extra mismatching nucleotide at the −3 or −4 position in addition to the allele-specific 3′ terminal nucleotide (Newton et al. 1989). The allele-specific DNA fragments were amplified by PCR in two separate reactions each with one allele-specific primer and a common primer. The annealing temperature for the AS-PCR was optimized for each SNP assay. Details of the primers are presented in Table 3.

Additionally, cleaved amplified polymorphic sequence (CAPS) assays (Konieczny and Ausubel 1993) were designed to differentiate 10 SNPs. The dCAPS Finder 2.0 website (http:helix.wust.edu/dcaps/dcaps.html) was used to determine appropriate restriction enzyme assays (Table 3). The CAPS–PCR reaction mixtures and conditions were the same as those used for the PCR–SSCP protocol. For the restriction enzyme digestion, 1.2-μl buffer and 0.1–0.15-U enzyme were added to 6-μl PCR product in a total volume of 12 μl, incubated at an appropriate temperature. The digestion products were separated on 2–3 % agarose gels.

Population genetic analysis

The genotypes obtained from PCR–SSCP, allele-specific tests, and DNA sequencing were scored as multi-allelic co-dominant data. Genetic diversity parameters within and between populations were estimated using GenAlEx 6.501 (Peakall and Smouse 2006, 2012). The programme

Table 2 Nucleotide sequences of the consensus (A) and locus-specific (B) PCR primers and primers used for sequencing the amplified fragments (C)

	Primer name	T_A	Product size (bp)	Sequence (5′–3′)
A: Gene consensus primers				
Isocitrate dehydrogenase	IDH-F2	50	900	AAGAGTGAAGGAGGNTAYGTNTGG
	IDH-R2		1047	GCATCAATGAAYTCYTCNGTRTT
Isopentenyldiposphate isomerase	IDI-F2	48	557	ACAAACACTTGYTGYAGYCAYCC
	IDI-R2			CCTTTCTCTACRTGRTCCCACCA
Isocitrate lyase	ICL-F2	51	467	ACCAAGAAGTGTGGNCAYATGGC
	ICL-R2			CTTGGCAAGTCCCARTCCCARAA
	ICL-F3	50	414	TTCAACTGGGATGCTTCNGGNATG
	ICL-R3			TTGAATTGCTCTTCTGTNACNCCT
Sucrose synthase	SUS-F2	51	544, 553	GATCCCAAGTTCAACATTGTNTCNCC
	SUS-R2		652	CCAAATGCTTCATARAANGCNGG
Leucine aminopeptidase	LAP-F	49	1770	GTTGCAGCTTGTGARAAYATGAT
	LAP-R			CCTGATTTCATTGAYTCCCARTA
Catalase	CAT-F	46	1037	GGTTTCTTTGARGTYACNCAYGA
	CAT-R		1095	TGATGAGCACAYTTTGGNGCRTT
Aspartate aminotransferase	AAT-F	46	1273	CATGCTTGTGCTCATAAYCCNACYGG
	AATcc-R			AAAGTGAACATTCCDATYTGYTT
Cysteine proteinase inhibitor	CPI-F	44	490	GCCGTCGATGARCAYAAYAARAA
	CPI-R			CCATGGYTTTWCCCANACYTT
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH-F	46	757	ACTCAGAAGACTGTTGAYGGNCC
	GAPDH-R		815	TAACCCCATTCRTTRTCRTACCA
Abscisic acid insensitive 3	ABI3-F	46	511	AAGGTGTTGAAGCARAGYGAYGT
	ABI3-R			AT CACTATGAARTCTCCYTCYTG
Auxin response factor 6 and 8	ARF-F	48	788	CACACTCGTGTTTCTGTTGGNATG
	ARF-R		1075	GTTGTTAATGGCTCDATYTCCCA
B: *Xylia xylocarpa* locus-specific primers				
Isocitrate dehydrogenase 1	Xx-IDH1-F	56	629–686	AGTGATTTCTTAGCCCAAGGTG
	Xx-IDH1-R			AAGTCCAACAATCTTGCATTTCC
Isopentenyldiphosphate isomerase 1	Xx-IDI1-F	55	449	CTCATTGAAGAGCAAGCCCTCG
	Xx-IDI1-R			AACAGCTCCTTCAATTGATCACG
Sucrose synthase 1	Xx-SUS1-F2	56	513–515	AGACTGACTACCCTGCACGGTTCA
	Xx-SUS1-R2			GCACCTTTAGTGTCTGCTATATAG
Sucrose synthase 2	Xx-SUS2-F2	56	502	AGACTCGCTCAATTTCACCCTGCC
	Xx-SUS2-R2			GCTCCCTTTGTGTCAGCAATGCA
Leucine aminopeptidase 1	Xx-LAP1-F	57	905–934	ATGCGGCCTGGAGATATTGTCAC
	Xx-LAP1-R			GTGTTGTTAACCTATACAAGAGG
Catalase 1	Xx-CAT1-F	57	880	CCTGTTATTGTCCGCTTCTCTAC
	Xx-CAT1-R			AATGTTTCTATTCAAGACCAGGCG
Catalase 2	Xx-CAT2-F	57	845–847	CCTGTCATTGTCCGTTTCTCGAC
	Xx-CAT2-R			AATGTTCTTATTCAAGACCAAACG
Aspartate aminotransferase (cytoplasmic isoform)	Xx-AATcy-F1	54	519–527	AGCTGATAAGATCAAAAGCTCT
	Xx-AATcy-R1			GAATACATGGGCCTGATCACTAG
C: Primers used for sequencing of amplified fragments				
Isocitrate dehydrogenase 1	Xx-IDH1-seqR	43	–	CCATTTCAACTGTTCCAA
Isopentenyldiphosphate isomerase 1	Xx-IDI-seqR	45	–	CCAAGGAGACAGTTTCAC
Sucrose synthase 1, 2	Xx-SUS-seqR2	48	–	AGCGGTAGAGCTCACCATT
Leucine aminopeptidase 1	Xx-LAP1-seqF	47	–	AACAGGTATGCGGCCTGG
Catalase 1, 2	Xx-CAT-seqR	46	–	GATGTCCTCAGGCCAAGT
Aspartate aminotransferase (cytoplasmic isoform)	Xx-AATcy-seqF	44	–	GCTGAAACTAGTGATCAG

Table 2 (continued)

	Primer name	T_A	Product size (bp)	Sequence (5′–3′)
	Xx-AATcy-seqR	46	–	GATTACCGGTCCTTGAGG

Nucleotides Y, R, W, D, and N according to the IUPAC nucleotide ambiguity code

T_A annealing temperature; *F, R* PCR primer used as forward and reverse primer, respectively; *seqF*, *seqR* primer used for sequencing of the fragments in forward or reverse orientation, respectively

GENEPOP (version 4.2) (Raymond and Rousset 1995; Rousset 2008) at the http://genepop.curtin.edu.au/ website was used to test for deviations from the HW equilibrium and linkage equilibrium. Nei's unbiased genetic distances were estimated using GenAlEx. The harmonic mean of Jost's measure of genetic differentiation across all loci, D_{est} (Jost 2008), and Hedrick's (2005) standardized measure of genetic differentiation were calculated using the programme SMOGD (version 1.2.5) at the website http://www.ngcrawford.com/django/jost/ (Crawford 2010). Co-ancestry-based population genetic distances (Reynolds et al. 1983) were calculated with the programme GENDIST from the PHYLIP package version 3.695 (Felsenstein 1989, 2005). Based on the calculated matrices of population pairwise genetic distances (Nei G_{st}, Jost D_{est}, and Reynolds *D*), neighbour joining phylogenies were calculated using the programme NEIGHBOR in PHYLIP. An

Table 3 PCR primers and reaction conditions for SNP detection by AS-PCR and CAPS assays

Primer	T_A	Sequence (5′–3′)	Assay type	SNP position
Xx-LAP1-F Xx-LAP1-Tsp-R	60	ATGCGGCCTGGAGATATTGTCAC TATTCAACCGGTTcTGCAGCAG	CAPS *Taa*I (CAN^GT)	225
Xx-IDH1-CAPS-F1 Xx-IDH1-CAPS-R1	56	AGTGATTTCGTAGCCCAAGGT AATCCTGCAAAGGGATAGGGA	CAPS *Dde*I (C^TNAG)	46
Xx-IDH1-CAPS-F2 Xx-IDH1-CAPS-R2	54	ATGCATGGGTTGGAAAATGT GACCAAGCAAAAATGGATGC	CAPS *Dde*I (C^TNAG)	204
Xx-IDH1-CAPS-F2 Xx-IDH1-CAPS-R2	54	ATGCATGGGTTGGAAAATGT GACCAAGCAAAAATGGATGC	CAPS *Bcc*I (GGTAGN^)	353
Xx-IDH1-508-F Xx-IDH1-508-R-T	57	AGCTTGAGATGCTTCATTTGGT GAAGCCTAAAACAACAAAAGTA	AS-PCR	508
Xx-IDH1-508-R-A		GAAGCCTAAAACAACAAAAGTT		
Xx-IDH1-558-F-A Xx-IDH1-558-F-C	56	TCCCCATTTGCTCCCAAAGAA TCCCCATTTGCTCCCAAAGAC	AS-PCR	558
Xx-IDH1-RL		TTTGGTCCATGAACAAGAAGCG		
Xx-IDH1-F Xx-IDH1-253-R-T	58	AGTGATTTCTTAGCCCAAGGTG CAAAGAAAAGGAAGAATGCCA	AS-PCR	253
Xx-IDH1-253-R-C		CAAAGAAAAGGAAGAATGCCG		
Xx-CAT1-709-F-C Xx-CAT1-709-F-T	53	CAGTCATGCAACTCAAGACATC CAGTCATGCAACTCAAGACATT	AS-PCR	709
Xx-CAT1-709-R		ATCATCTGAGTAGTATATTCCAGG		
Xx-CAT1-F Xx-CAT1-AluI-R	54	CCTGTTATTGTCCGCTTCTCTAC CAATTAGGCTTGTTTCCCTAAG	CAPS *Alu*I (AG^CT)	131
Xx-CAT2-Tas-F Xx-CAT2-Tas-R	52	AAGTTCCCCATTTTGACTAAGG ATTTTACATAGTGTGCTTTCCC	CAPS *Tas*I (^AATT)	477
Xx-CAT2-Tsp-F Xx-CAT2-Tsp-R	62	ATCCAGGAGAACTGGAGGATCC CTGTGGTTGGATCCTCCAACCC	CAPS *Taa*I (CAN^GT)	597
Xx-CAT2-Nla-F Xx-CAT2-Nla-R	58	GGGTGTCCCACAAGATTACAGG ATGGTCAGGATCCATTGTCTGG	CAPS *Nla*III (CATG^)	657
Xx-CAT2-610-F Xx-CAT2-610-R-G	53	ATCTTCTGGATTCTGGTCAGTG GCCTCCTCCTCCAATAGAAAC	AS-PCR	610
Xx-CAT2-610-R-A		GCCTCCTCCTCCAATAGAAAT		
Xx-IDI1-F Xx-IDI1-R	56	CTCATTGAAGAGCAAGCCCTCG AACAGCTCCTTCAATTGATCACG	CAPS *Vsp*I (AT^TAAT)	78
Xx-IDI1-F Xx-IDI1-R	56	CTCATTGAAGAGCAAGCCCTCG AACAGCTCCTTCAATTGATCACG	CAPS *Taq*I (T^CGA)	56

T_A annealing temperature

analysis of molecular variance (AMOVA) was performed in GenAlEx 6.501 with or without grouping the 16 populations into two or three regions. Of the dataset, 9999 permutations were done to test for significance of the variances. To test for association between population genetic distances and population geographic distances, simple Mantel tests were done using the programme zt (Bonnet and Van de Peer 2002). A principal component analysis was done using the programme DARwin (version 5) (Perrier et al. 2003). DARwin5 calculates a genetic distance based on the number of shared alleles among diploid individuals. This genetic distance information was also used to construct a phylogenetic tree using weighted neighbour joining.

Population genetic structure was analysed using the programme STRUCTURE version 2.3.4 (Pritchard et al. 2000) for the number of clusters (K) ranging from 1 to 6, using the model with admixture and correlated allele frequencies. The data were analysed both without and with using the sampling location as a prior (Hubisz et al. 2009). Ten replicate analyses were done each with 250,000 steps for burn-in and 750,000 steps of the MCMC sampling after burn-in. The obtained likelihood values were further analysed following the method of Evanno et al. (2005) as implemented in the STRUCTURE HARVESTER programme (Earl and vonHoldt 2012). The programme CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) was used to summarize the STRUCTURE results and visualized with distruct1.1 (Rosenberg 2004).

Results

Analysis of specific nuclear DNA markers

The consensus PCR primers (Table 2A) complementary to the conserved regions of some metabolic enzyme-coding genes could amplify one or a small number of fragments in *X. xylocarpa*. Between 3 and 6 clones containing inserts were isolated and sent for sequencing. The results indicated that the fragment inserts corresponded to 18 putative loci (AAT: 1, CAT: 2, IDI: 1, IDH: 2, ICL: 2, ABI3: 1, GAPDH: 2, LAP: 1, SUS: 3, ARF6/8: 2, CPI: 1). The sequences have been submitted to EMBL with accession numbers LM651377–LM651394. With adjustment of PCR amplification conditions, these consensus primers can be used to specifically amplify fragments in a wide range of plant species (data not shown).

Based on the obtained sequences, specific primer pairs were designed for 6 enzyme systems encoded by 8 loci: SUS (2 loci), AAT (1 locus), IDH (1 locus), LAP (1 locus), CAT (2 loci), and IDI (1 locus). On agarose gel, single bands ranging in size from 350 to 930 bp were visible but different alleles could not be distinguished. Separation of the fragments by SSCP identified multiple alleles in all populations for all primer pairs. Although distinct SSCP banding patterns were detected for most alleles, genotypes could not conclusively be assigned to some individuals. This was particularly the case for the IDH1 and IDI1 loci. To resolve these uncertain genotypes, PCR products were amplified from trees presumably homozygous for the different alleles and from several of the trees with ambiguous SSCP assay results. These PCR-amplified fragments were sent to be sequenced using an internal sequencing primer (Table 2C). Based on the obtained DNA sequence information, allele-specific assays were then designed using the AS-PCR or CAPS approach. Assays were designed for a total of 15 SNPs from 5 loci: IDH1 (6), LAP (1), CAT1 (2), CAT2 (4), and IDI1 (2). Ten SNPs were detected by digestion of a PCR product with restriction enzymes (CAPS). Five nucleotide polymorphisms were converted into AS-PCR tests incorporating an extra mismatching nucleotide at the −3 position (Table 3).

Genetic parameters

The genotypes at each locus were deduced from the combined information of SSCP, CAPS, AS-PCR, and direct sequencing. Almost all loci were polymorphic in all populations, with only four populations monomorphic for one locus each. The 8 loci appeared to be unlinked with no locus pairs having a significant linkage disequilibrium (LD) in the overall analysis and only 11 pairs of loci indicating LD in one to three of the 16 populations. Although the total number of haplotypes ranged from 10 (CAT2) to 33 (IDH1), the effective number of alleles ranged from 1.118 (AATcy) to 7.586 (IDH1), which was almost threefold higher than that for the second highest, CAT1 (2.706) (Table 4). This was due to the presence of a large number of rare alleles. Seventy-one private alleles were found (2 to 11 per population) among 83 trees with AATcy and IDH1 having 17 and 15 private alleles in 9 and 10 of the populations, respectively. All of the private alleles were low-frequency alleles (<5 %), also in the population where they were found, with 57 of them occurring only once (singletons). The observed heterozygosity was moderately high for 6 loci (ranging from 0.420 to 0.819) but low for the AATcy (0.091) and LAP (0.289) loci with a mean of 46.3 %. CAT1, CAT2, IDI1, and AATcy had positive inbreeding coefficients (F) in the total sample (0.210, 0.200, 0.157, and 0.120, respectively) indicating heterozygote deficits. However, the average inbreeding coefficient per population was moderately large only for CAT1 and CAT2 (0.185 and 0.146, respectively). The inbreeding coefficient at the IDI and AATcy loci dropped substantially when analysed per population. SUS1 and IDH1 had the lowest inbreeding coefficients. The estimated number of migrants (Nm) was higher than 1 for all loci, with a mean of 4.350, indicating a high level of gene flow among the populations. All eight loci showed significant deviations from the Hardy–Weinberg equilibrium in the total sample, with six of

them extremely so (Table 5). When considered per population, the SUS1 locus was in the Hardy–Weinberg equilibrium (HWE) in all populations, the AATcy locus deviated from the HWE in just one population, and the CAT1 locus had the largest number (8) of populations that were not in the HWE (Supplementary Table S1).

Genetic distances and AMOVA

Pairwise population genetic distances were calculated according to Nei's unbiased genetic distance (G_{st}), Jost D_{est}, and Reynolds et al.'s D. All three measures showed a similar pattern with the largest genetic distances between the ThaTaFang area and Mae Ngao National Park (NP) on one hand and Pang Sida and Phu Jong Nayoi National Parks on the other (Supplementary Table S2). The neighbour joining phenetic trees calculated based on these distances showed the corresponding clustering among the 16 populations (Fig. 2). Overall though, the genetic distances were quite small as even the geographically most distant populations (Lam Khlong Ngu, Erawan, Mae Ngao, ThaTaFang from western Thailand vs. Pang Sida, Phu Jong Nayoi, Phupha Lek from northeastern Thailand) shared several genotypes. A Mantel test based on population genetics and Euclidean geographical distances showed a moderate but significant correlation (R_{xy}=0.336, 0.339, and 0.368 with P=0.034, 0.029, and 0.028) based on the Nei, Jost, and Reynolds distance matrices.

An AMOVA for the 538 *Xylia* trees from 16 populations indicated that for all loci combined, F_{st}=0.061, F_{is}=0.075, and F_{it}=0.132 when considering three regions (Table 6). Just a small amount of the total variation was distributed among populations (2 %) and among regions (4 %), with the remainder within populations (7 % among individuals within populations and 87 % within individuals). For the individual loci, the variance due to regions ranged from 0 % (CAT1) to 13 % (IDI). The variance for all sources was significant (P<0.01). The pairwise F_{st} and F'_{st} (Meirmans 2006) were mostly low (0.002–0.075) except again for the comparisons of the western populations with those from the northeast (0.106–0.193), but almost all were highly significant (Supplementary Table S3).

Clustering

The programme DARwin5 was used for a principal component clustering and neighbour joining tree calculation. DARwin5 calculates a genetic distance among individuals based on the number of shared alleles in a diploid organism. Neither the principal component analysis of all *Xylia* genotypes nor the unweighted neighbour joining tree revealed a clustering among the 548 individuals, including the 10 individual trees sampled from different locations in Thailand (Supplementary Figure 1). The STRUCTURE analysis of all 548 trees without using population origin as part of the estimation of the prior likelihood distribution lent weak support for a division of all *X. xylocarpa* trees into two groups. When analysing the 538 trees sampled from 16 populations and using their population origin as a prior to assist the clustering (LOCPRIOR option), the population division becomes much clearer. When assuming two clusters (K=2), the trees from western Thailand (Lam Khlong Ngu, Erawan, Mae Ngao, and ThaTaFang) separate from the other populations. When assuming three clusters (K=3), also the populations from northeastern Thailand could be separated from the others (Fig. 3).

Table 4 Summary of population means of genetic parameters per locus for the 16 *Xylia xylocarpa* populations, 538 individuals

Locus	*N*	*A*	*Ne*	H_o	UH_e	*F*	*Nm*	H_T	G_{st}	G'_{st}	D_{est}
SUS1	33.563	3.813	2.287	0.555	0.564	0.005	4.113	0.589	0.042	0.100	0.061
SUS2	33.625	4.313	2.227	0.503	0.547	0.063	4.860	0.565	0.034	0.077	0.045
AATcy	33.625	3.125	1.144	0.100	0.117	0.060	4.987	0.121	0.033	0.037	0.005
CAT1	33.625	5.563	2.637	0.496	0.622	0.186	7.746	0.632	0.016	0.044	0.028
CAT2	33.563	4.688	2.163	0.411	0.481	0.146	2.783	0.517	0.068	0.135	0.072
LAP1	33.563	5.500	1.506	0.311	0.314	0.024	4.601	0.326	0.036	0.054	0.018
IDI1	33.563	5.438	2.492	0.513	0.537	0.057	1.740	0.605	0.112	0.250	0.156
IDH1	33.625	11.375	5.762	0.822	0.832	−0.005	3.918	0.871	0.045	0.283	0.249
Mean	33.594	5.477	2.527	0.464	0.502	0.066	4.344	–	–	–	–

N average number of individuals genotyped per population, *A* average number of alleles observed per population, *Ne* average effective number of alleles per population, H_o average observed heterozygosity per population, UH_e average unbiased expected heterozygosity per population (Nei 1973), *F* average inbreeding coefficient per population, *Nm* migration rate $[(1/F_{st})-1]/4$, H_T nearly unbiased estimator of total subpopulation heterozygosity (Nei 1973), G_{st} standardized measure of relative differentiation (Nei 1973), G'_{st} standardized measure of genetic differentiation (Hedrick 2005), D_{est} estimator of actual differentiation (Jost 2008)

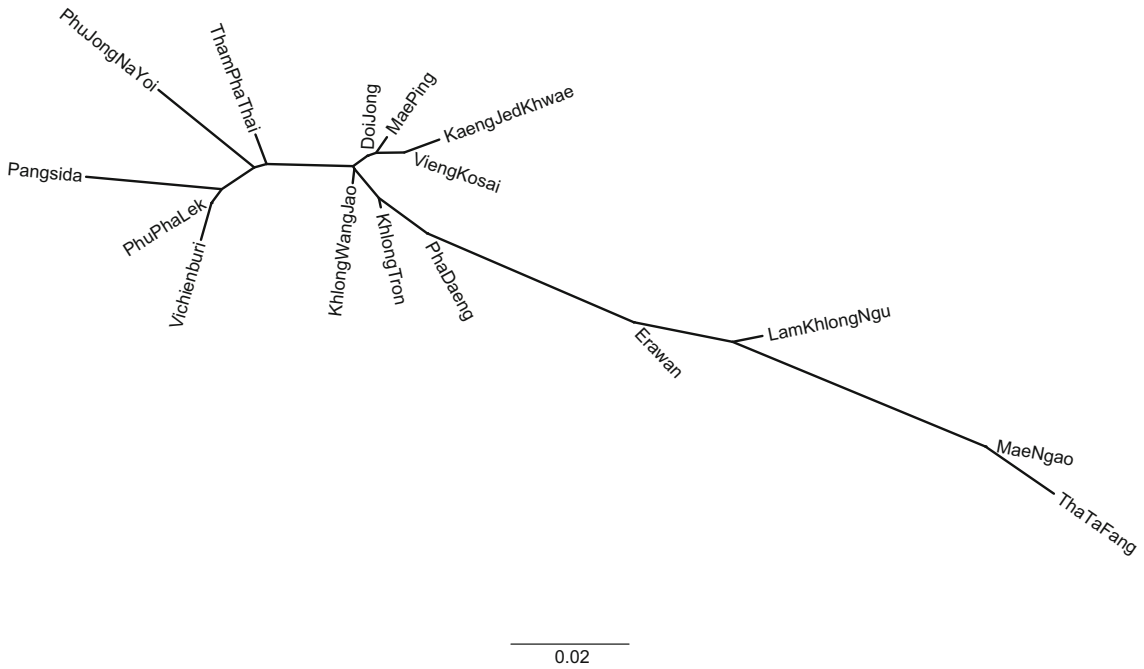

Fig. 2 Neighbour joining tree based on pairwise distances between all pairs of *Xylia xylocarpa* populations. The tree was annotated with the programme FigTree (Rambaut 2009)

Discussion

Analysis of single-copy nuclear DNA markers

For the study of population genetic diversity, selectively neutral markers are preferred as any deviation from expectations can then be interpreted as resulting from stochastic population events (Hedrick 2011). Various molecular marker systems have been used in population genetic studies ranging from isozymes (Brown 1978; Ritland et al. 2005), dominant PCR-amplified presence–absence markers such as RAPDs and AFLPs (Cao et al. 2006), to co-dominant PCR-amplified markers such as repeat length variants (Ismail et al. 2012). Detection of polymorphisms in introns has been used as a gene-based marker system in conservation genetics and phylogeography of plants (Garrick et al. 2008). This approach is broadly based on the exon-primed intron-crossing PCR (EPIC-PCR) developed by Palumbi and Baker (1994). The genetic loci that were chosen in this study are mostly genes essential for cellular processes, similar to the genes that have been the preferred choice for population genetic analysis using isozyme assays which have been shown to behave as mostly neutral genetic markers. In all plant species from which sequence information was used in the alignment for the selection of consensus primers, these genes were probably low-copy-number gene families, even when only a single sequence could be retrieved from the databases. Indeed, after cloning and sequencing of amplified fragments, two or three loci were recovered for several of the targeted genes though the genome may harbour still some more.

Table 5 Summary of genetic parameters per locus for all *Xylia xylocarpa* samples

Locus	*N*	*A*	N_e	H_o	UH_e	*F*	*I*	HWE
SUS1	547	12	2.436	0.559	0.590	0.051	1.028	0.0069
SUS2	548	12	2.267	0.513	0.559	0.082	1.030	0.0000
AATcy	548	21	1.118	0.093	0.106	0.120	0.326	0.2396
CAT1	548	23	2.712	0.498	0.632	0.211	1.276	0.0000
CAT2	547	10	2.130	0.424	0.531	0.200	1.121	0.0000
LAP1	547	21	1.461	0.298	0.316	0.056	0.836	0.0000
IDI1	547	11	2.463	0.501	0.595	0.157	1.268	0.0000
IDH1	548	33	7.586	0.819	0.869	0.056	2.315	0.0000
Mean	547.5	17.875	2.772	0.463	0.525	0.117	1.150	0.0000

N number of individuals genotyped, *A* total number of alleles observed, N_e effective number of alleles, H_o observed heterozygosity, UH_e unbiased expected heterozygosity, *F* inbreeding coefficient, *I* Shannon information index, *HWE* probability of observed allele frequency distribution if all alleles are in the Hardy–Weinberg equilibrium

Based on the nucleotide sequences obtained after cloning DNA fragments amplified using the consensus primers, specific primer sets were designed in such a way that single loci could be amplified (Table 2B). Because each amplification product was designed to include at least one intron, and introns have been found to be more polymorphic than exons (Graur and Li 2000), a high level of polymorphism could be anticipated. Single-copy amplification products can then be screened for polymorphisms by one of many techniques for detection of genetic variants at the molecular level. In this

Table 6 Analysis of molecular variance of 538 *Xylia* trees from 16 populations with probabilities based on 9999 permutations

A: Overall AMOVA											
Source	*df*	SS	MS	Est. var.	Percentage						
Among 3 regions	2	67.476	33.738	0.087	4						
Among 16 pops	13	65.763	5.059	0.044	2						
Among indiv	522	1127.516	2.160	0.151	7						
Within indiv	538	999.500	1.858	1.858	87						
Total	1075	2260.256		2.139	100						
B: Derived *F* statistics per locus and for all loci combined											
	AATcy	SUS1	SUS2	CAT1	CAT2	LAP	IDI1	IDH1	All		
F_{st}	0.027 (0.000)	0.055 (0.000)	0.045 (0.000)	0.014 (0.002)	0.072 (0.000)	0.048 (0.000)	0.149 (0.000)	0.049 (0.010)	0.061 (0.000)		
F_{is}	0.101 (0.002)	0.012 (0.269)	0.052 (0.066)	0.196 (0.000)	0.162 (0.000)	0.031 (0.115)	0.061 (0.012)	0.014 (0.198)	0.075 (0.000)		
F_{it}	0.126 (0.000)	0.066 (0.028)	0.095 (0.003)	0.207 (0.000)	0.222 (0.000)	0.077 (0.002)	0.201 (0.000)	0.062 (0.000)	0.132 (0.000)		

study, three simple assays, PCR–SSCP, CAPS, and AS-PCR that are inexpensive, reliable, and sensitive for mutation detection, were applied for genotyping the samples. SSCP polyacrylamide gel electrophoresis (Orita et al. 1989), which can detect conformational differences of single-stranded DNA fragments in a non-denaturing polyacrylamide gel, was chosen because it is technically easy to perform and has a high sensitivity to detect polymorphisms even in sequences of identical length (Plomion et al. 1999; Sato and Nishio 2004). SSCP does not require prior knowledge of the particular substitution, insertion, or deletion that was responsible for the observed polymorphism. The high sensitivity of the SSCP approach has the additional advantage that any amplification product of a paralogous locus in an individual would be easily detected since it would be expected to migrate through the polyacrylamide matrix at a different speed. No such spurious bands were observed when scoring the gels. All very-low-frequency or unique alleles were sequenced, and in all cases, the results indicated that the sequence was derived from the same locus. Thus, we have great confidence that all genotype data were derived from orthologous loci across all individuals.

The SSCP analysis of the amplified DNA fragments revealed a large diversity of bands on the silver-stained gels, many of which could be tentatively assigned to alleles. PCR products were then sequenced from those trees with different identifiable alleles and from several samples where the SSCP patterns were ambiguous. The haplotypes deduced from the sequencing analysis indicated the presence of additional alleles among the samples that gave the unclear SSCP patterns. To reveal those alleles in the natural populations, some of the detected SNPs were converted into allele-specific PCR or CAPS assays. Primers were designed for a total of 15 SNPs. Five SNPs were developed into allele-specific PCR assays and 10 SNPs were converted into CAPS assays. The CAPS approach had been developed in *Arabidopsis* (Konieczny and Ausubel 1993) and has been widely used for assaying SNPs in many crop species such as barley (Kota et al. 2008), rice (Komori and Nitta 2005), and soybean (Shu et al. 2010). The results of the SSCP were thus confirmed by sequencing, and the presence of additional SNPs discovered by sequencing was in turn confirmed by the allele-specific detection methods, thus giving us high confidence that the detected alleles are true biological variants and not technical artifacts.

The combination of SSCP with sequencing of a limited number of samples, followed by AS-PCR and CAPS screening, greatly sped up the analysis of a large number of individuals for population genetic diversity study in a tree species for which no previous studies have been reported: the SSCP assay provides information on the most common alleles while the specific CAPS and AS-PCR assays revealed the presence of particular variants that were more difficult to differentiate otherwise. This resulted in the detection of a very high level of diversity with all loci polymorphic in almost all

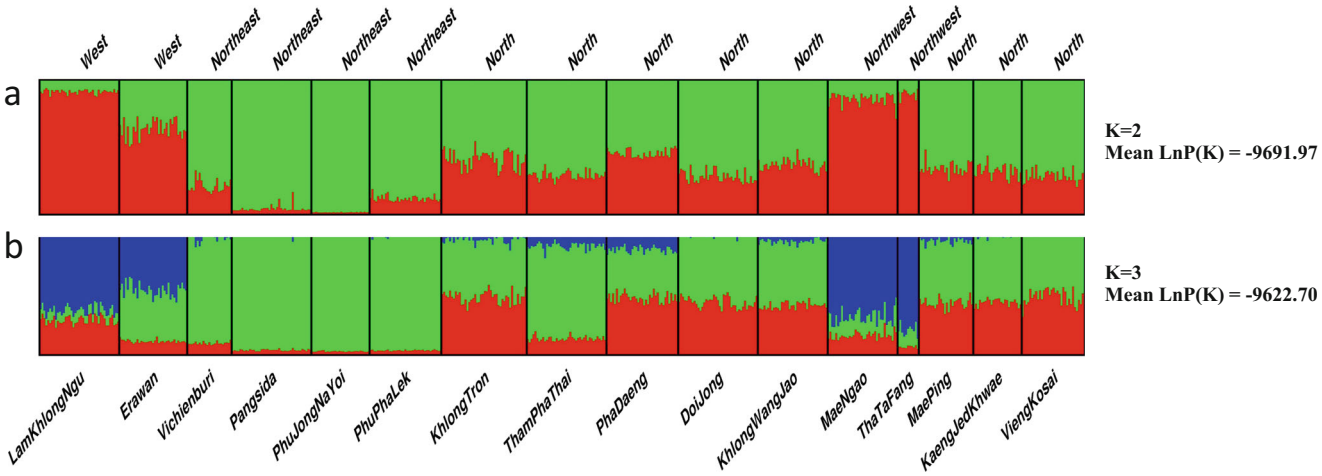

Fig. 3 Bayesian clustering analysis results for $K=2$ (**a**) and $K=3$ (**b**) with admixture and correlated allele frequencies and using population origin in the prior

populations. Out of 548×8 genotyping assays, only 4 individual genotypes remained missing. The level of null alleles is expected to be very low since the loci that were genotyped were all assumed to be part of the gene-coding regions and the primers for the PCR were located in exons.

Population genetic analysis

Between 10 and 33 alleles were observed at the different loci in the *X. xylocarpa* genome. However, many alleles occurred in very low frequency, and thus, the effective number of alleles was much lower, between 1.1 and 7.6. Such a high number of alleles have never been observed before for gene-linked markers (isozymes or intron polymorphisms) except for genes involved in self-incompatibility systems in plants (Lawrence 2000). Studies using gene-based markers at the DNA level are still very rare in plant population genetics and phylogeography. Four amplicons derived from two catalase genes revealed up to 108 distinct SSCP banding patterns among a collection of 119 poplar clones belonging to 4 species and interspecific hybrids (Racchi et al. 2011). Sequencing of gene regions putatively involved in the adaptation of flowering time in *Arabidopsis* has revealed large numbers of alleles (Caicedo et al. 2004; Méndez-Vigo et al. 2011). Even though *Arabidopsis* is an annual species that is mostly selfing, a large number of rare alleles were observed in a Europe-wide collection of accessions (Nordborg et al. 2005). Thus, the high number of alleles found in the present study may turn out to be not so unusual for a long-lived outcrossing tropical tree species. More studies using similar markers in other tree species may put the allele frequencies observed in *X. xylocarpa* better in perspective.

Deviations from the Hardy–Weinberg Equilibrium per locus were highly significant for all loci when all trees were considered together, but when considered per population, only some loci showed deviations for a number of populations. CAT1 had the highest number of populations (8) with significant deviations from the HWE, while none were observed for the SUS1 marker. This would indicate that some of the overall deviation from the HWE would be due to population subdivision effects (Wahlund effect), which could be responsible for the observed region-wide population structure.

CAT1, CAT2, and IDI1 had high inbreeding coefficients (F=0.210, 0.200, and 0.157, respectively; Table 5) indicating heterozygote deficits (UH_e-Ho=0.133, 0.108, and 0.094, respectively), while SUS1, LAP1, and IDH1 had the lowest F (0.051–0.056). However, the average inbreeding coefficient per population was moderately large only for CAT1 and CAT2 (0.185 and 0.146, respectively). The population-wise inbreeding coefficient at the IDI1 locus showed a large drop compared to the values as single population, again indicating that the heterozygote deficit at the IDI1 locus may be due to the population structure. The drop in the value of inbreeding coefficient at the AATcy locus was most likely due to the presence of a large number of private alleles at this locus. Ten populations had from one to five private alleles at the AATcy locus.

The largest pairwise population genetic distances were observed between the ThaTaFang area and Mae Ngao National Park on one side and the Pang Sida and Phu Jong Nayoi National Parks. These populations are also geographically most distant, from the Thai–Myanmar border in the West to the Thai–Cambodian border in the East. It has to be remarked that only 11 trees were collected in the ThaTaFang area. Although the largest genetic distances were found between populations that are geographically very distant, a Mantel test to detect correlations between population geographical and genetic distances showed only a moderate correlation (R_{xy} about 0.34, significant at 5 %). A principal component analysis in DARwin5 was not able to reveal identifiable clusters among the total set of 548 trees (Figure S1). There was a slight trend for the trees from western

Thailand (ThaTaFang area, Mae Ngao National Park, Lam Khlong Ngu NP, and Erawan NP) to cluster in one area of the PCA plot and neighbour joining tree, but several individuals from other areas in Thailand were still intermingled with them. The Bayesian clustering algorithm implemented in STRUCTURE gave only a weak indication for population differentiation in two or maybe three clusters, but the clustering became very clear when the population origin information was included in the prior likelihood distribution (Hubisz et al. 2009). The rather sharp contrast in allele frequencies between the populations in the Salween river watershed (ThaTaFang and Mae Ngao National Park) and the nearby populations in northern Thailand (e.g. Mae Ping National Park) indicates that the population structure is real and that the high mountains (over 1000 m a.s.l.) separating the Salween River watershed from the Ping River watershed are a contemporary barrier to gene flow for *X. xylocarpa*, which rarely grows at elevations over 800 m.

The population structure analysis indicated a split of the *X. xylocarpa* populations into three clusters in Thailand, based on geographical location, including physiographic differences, and genetic differences. One zone would comprise the western mountain region of Thailand, west of the Thanon Thong Chai mountain range and further south. Another area would be the rest of northern Thailand, and the third region would be the populations on the Khorat plateau in northeastern Thailand. The Khorat plateau has its own physical geography, being drier than the other regions.

The populations in Kanchanaburi, to the south of the Thanon Thong Chai mountains, appear to belong to the same cluster as the Salween populations. Apparently, it may have been possible for *X. xylocarpa* and similar plant species to cross from the Salween watershed into the watersheds further east and south through gaps in the mountain ranges such as at the Three Pagodas Pass (300 m a.s.l.) area or Mae Sot area (650 m a.s.l.).

The limited differentiation of genetic diversity among geographically separate populations within Thailand might indicate a high level of gene flow over long distances. The estimated number of effective migrants (Nm) averaged for all loci was 4.291 and as high as 7.399 for the CAT1 locus. Probably the pollination syndromes and seed dispersal mechanisms can explain this. *Xylia* flowers are small, pale yellow, in dense spherical heads. The flowers are hermaphroditic or male. The main flowering is normally around the time of leaf flushing. At the time of flowering, the flowers are heavily visited by various insects, among them social and solitary bees and stingless bees. However, the foraging distances travelled by these insects are unknown. The seed dispersal in *Xylia* is probably rather limited. When pods dry, they tend to open suddenly and the seed can be flung away for some distance. However, the seed does not have any adaptation for further dispersal through the air or once on the ground.

The large number of low-frequency alleles and private alleles at these protein-coding gene regions might indicate that they play a role in *Xylia* adaptation and evolution. With 71 private alleles found among 83 trees and with all populations having 2 to 9 private alleles, it means that almost 17 % of the individuals (ranging from 4.9 to 43.5 % per population) had a private allele at one of the 8 loci. Since all populations harboured private alleles, it is highly likely that additional alleles can still be found when more populations or individuals in a population are sampled. Such a high level of rare alleles has not been reported in any other tropical forest tree species, though in a study of diversity in *Koompassia malaccensis* (Fabaceae—Caesalpinioideae), up to 10 allozyme variants were found with many alleles at very low frequency, though only a single private allele (Lee et al. 2007). The occurrence of such low-frequency alleles would imply that sizeable populations at various locations need to be protected for the effective conservation of the genetic diversity in *X. xylocarpa*.

The large number of rare alleles that has been found here could indicate that the gene pool of this long-lived tree species may not (yet) have been affected too much by logging and forest fragmentation, as speculated by Hamrick (2004). The flowers are pollinated by a variety of insects, and though not much is known about effective pollen dispersal distances, there may not be a limitation in gene flow.

Although *Xylia* wood is used frequently for parquet and furniture, nothing much is known about historical levels of exploitation. However, in all locations, there was ample evidence of past logging and teak and *Xylia* were the most commonly logged species. Of all the populations sampled, the Vichienburi population is the most clearly affected by anthropogenic disturbances. The population is on a hill, almost completely surrounded by agricultural fields and villages. As the area has no formal protection status, the forest is heavily exploited by local villagers for bamboo and other forest products. Not a single large-sized *Xylia* tree was left. Still, many rare and private (10) alleles were found in this population. The historical genetic diversity of this species may still have been larger than what has remained until the present.

The gene-based markers developed in this study were successfully used to investigate the population genetic structure of *X. xylocarpa* in Thailand. Evidence for population genetic division into three geographic regions was found, which could be used as a basis for utilization and conservation of the genetic resources for this tree species. While at the same time a large number of private alleles were detected, the significance of this observation will need more studies of the molecular genetic diversity in *X. xylocarpa* and other tree species to clarify whether this is a unique feature of *X. xylocarpa* or a more general phenomenon. The comparison of genetic diversity among the mature trees with that among seed or seedlings may reveal whether these low-frequency alleles are in danger of being lost when populations are being fragmented.

Acknowledgments A part of this research was funded by a grant from the Center for Advanced Studies for Agriculture and Food (NRU_CASAF), Kasetsart University, to SI. Publication of the manuscript was supported by the Kasetsart University Graduate School. Permission to collect tissue samples was obtained from the Department of Wildlife and Plant Conservation (permission numbers 0907.1/4792 and 0907/15941).

Data archiving statement The genotyping data and sample location data are provided as supplementary online materials. The data have also been submitted to the TreeGenes database (http://dendrome.ucdavis.edu/treegenes/) under accession number TGDR021.

The sequences have been assigned EMBL/GENBANK/DDBJ accession numbers LM651377–LM651394.

References

Bassam BJ, Anollés GC, Gresshoff PM (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80–83

Bhumibhamon S (1986) The environmental and socio-economic aspects of tropical deforestation: a case study of Thailand. Department of Silviculture, Faculty of Forestry, Kasetsart University, Bangkok

Bonnet E, Van de Peer Y (2002) zt: a sofware tool for simple and partial Mantel tests. J Stat Softw 7(10)

Brown AHD (1978) Isozymes, plant population genetic structure and genetic conservation. Theor Appl Genet 52:145–157

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interactions between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal clime in a life history trait. PNAS 101:15670–15675

Cao C-P, Finkeldey R, Siregar IZ, Siregar UJ, Gailing O (2006) Genetic diversity within and among populations of *Shorea leprosula* Miq. and *Shorea parvifolia* Dyer (Dipterocarpaceae) in Indonesia detected by AFLPs. Tree Genet Genomes 2:225–239

Charuppat T (1998) Forest situation in the past 37 years (1961–1998). Royal Forest Department, Bangkok (in Thai)

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10:556–557

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

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

Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Garrick RC, Meadows CA, Nicolas AN, Nason JD, Dyer RJ (2008) A set of polymorphic nuclear intron markers for conservation genetics and phylogeography of *Euphorbia* species (Pedilanthus clade). Conserv Genet 9:1673–1676

Graur D, Li H-S (2000) Fundamentals of molecular evolution, 2nd edn. Sinauer Associates Inc., USA

Hamrick JL (2004) Response of forest trees to global environmental changes. For Ecol Manag 197:323–335

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Hedrick PW (2011) Genetics of populations, 4th edn. Jones and Bartlett Publishers, Sudbury

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Ismail SA, Ghazoul J, Ravikanth G, Uma Shaanker R, Kushalappa CG, Kettle CJ (2012) Does long-distance pollen dispersal preclude inbreeding in tropical trees? Fragmentation genetics of *Dysoxylum malabaricum* in an agro-forest landscape. Mol Ecol 21:5484–5496

Jakobsson M, Rosenberg NA (2007) *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jost L (2008) G_{st} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

Josue J (2004) Some wood properties of *Xylia xylocarpa* planted in Sabah. Sepilok Bull 1:1–15

Karp A, Edwards KJ, Bruford M, Funk S, Vos B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D, Hewitt GM (1997) Molecular techniques for biodiversity evaluation: opportunities and challenges. Nat Biotechnol 15:625–628

Komori T, Nitta N (2005) Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers. Breed Sci 55:93–98

Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J 4:403–410

Kota R, Varshney R, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics 8:223–233

Lawrence MJ (2000) Population genetics of homomorphic self-incompatibility polymorphisms in flowering plants. Ann Bot 85(suppl1):221–226

Lee CT, Lee SL, Ng KKS, Siti Salwana H, Norwati M, Saw LG (2007) Allozyme diversity of *Koompassia malaccensis* (Leguminosae) in Peninsular Malaysia. J Trop For Sci 19:73–78

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:1299–1402

Méndez-Vigo B, Picó FX, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C (2011) Altitudinal and climatic adaptation is mediated by flowering traits and FRI, FLC, and PHYC genes in *Arabidopsis*. Plant Physiol 157:1942–1955

Neale DB, Kremer A (2011) Forest tree genomics: growing resources and applications. Nat Rev Genet 12:111–122

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A 70:3321–3323

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503–2516

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Zhao K, Kalbfleish T, Schulz V, Kreitman M, Bergelson J (2005) The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol 3:e196

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874–879

Pacheco LF, Simonetti JA (2000) Genetic structure of a mimosoid tree deprived of its seed disperser, the spider monkey. Conserv Biol 14:1766–1775

Palumbi SR, Baker CS (1994) Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11:426–435

Peakall R, Smouse PE (2006) GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peakall R, Smouse PE (2012) GENALEX6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28:2537–2539

Perrier X, Flori A, Bonnot F (2003) Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC (eds) Genetic diversity of cultivated tropical plants. Enfield Science Publishers, Montpellier, pp 43–76

Plomion C, Hurme P, Frigerio J-M, Ridolfi M, Pot D, Pionneau C, Avila C, Gallardo F, David H, Neutelings G, Campbell M, Canovas FM, Savolainen O, Bodénès C, Kremer A (1999) Developing SSCP markers in two *Pinus* species. Mol Breed 5:21–31

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

Racchi ML, Turchi A, Caparrini S, Camussi A (2011) SSCP intron marker system is a convenient tool for clonal fingerprinting of poplar (*Populus*) cultivars of different species and interspecific hybrids. Genet Resour Crop Evol 58:507–518

Rambaut A (2009) Molecular evolution, phylogenetics and epidemiology: FigTree version 1.4.0. Available Source: http://tree.bio.ed.ac.uk/software/figtree/

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248–249

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767–779

Ritland K, Meagher LD, Edwards DGW, El-Kassaby YA (2005) Isozyme variation and the conservation genetics of Garry oak. Can J Bot 83: 1478–1487

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Res 8:103–106

Sato Y, Nishio T (2004) Mutation detection in rice waxy mutants by PCR-RF-SSCP. Theor Appl Genet 107:560–567

Shu Y, Li Y, Zhu Z, Bai X, Cai H, Ji W, Guo D, Zhu Y (2010) SNPs discovery and CAPS marker conversion in soybean. Mol Biol Rep 38:1841–1846

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Wang BC, Sork VL, Leong MT, Smith TB (2007) Hunting of mammals reduces seed removal and dispersal of the Afrotropical tree *Antrocaryon klaineanum* (Anacardiaceae). Biotropica 39:340–347