

Genetic, Biochemical, and Morphological Diversity of the Legume Biofuel Tree *Pongamia pinnata*

Qunyi Jiang¹, Shang-Heng Yen¹, Jiri Stiller², David Edwards^{2,3}, Paul T. Scott^{1,#}, and Peter M. Gresshoff^{1,2,*}

¹ ARC Centre of Excellence for Integrative Legume Research, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia; ² School of Agriculture and Food Science, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia; ³ Australian Centre for Plant Functional Genomics, The University of Queensland, St Lucia, Brisbane, QLD 4072, Australia.

Received: May 7, 2012 / Accepted: June 10, 2012

Abstract

Pongamia pinnata is regarded as a sustainable biofuel feedstock of the future because of its abundant production of oil-rich seeds, tolerance to abiotic stress, and ability to undergo biological nitrogen fixation (minimizing nitrogen inputs). However, it needs extensive domestication through selection and genetic improvement. Owing to its outcrossing nature, *Pongamia* displays large phenotypic diversity, which is advantageous for selection of desirable phenotypes but problematic for plantation management. In this study, variation was evaluated for seed mass, oil content, and oil composition. To evaluate genetic diversity and to lay the basis for a molecular breeding approach we developed second generation sequencing (2GS)-derived ISSR markers (*Pongamia* Inter-Simple Sequence Repeats; PISSR). The special feature of PISSRs is that the number of nucleotide repeats and the 5' and 3' nucleotide extensions were not arbitrarily chosen, but were based on *Pongamia* genomic sequences obtained from a NGS (Illumina®) database. Amplification products were resolved by polyacrylamide gel electrophoresis and silver staining or automated capillary electrophoresis to yield distinct and reproducible profiles. Polymorphic bands were excised from polyacrylamide gels and sequenced to reveal similarity to DNA sequences from other legumes. We demonstrated: 1) an abundance of nucleotide core repeats in the *Pongamia* genome, 2) large genetic and phenotypic diversity among randomly sampled *Pongamia* trees, 3) restricted diversity in progeny derived from a single mature tree; 4) sta-

bility of PISSR markers in *Pongamia* clones; and 5) genomic DNA sequences within PISSR markers. PISSRs provide a valuable biotechnology tool for assessment of genetic diversity, gene tagging and molecular breeding in *Pongamia pinnata*.

Keywords: Biofuel, *Millettia*, molecular marker, PISSR, plant oil, *Pongamia*.

Abbreviations: CE: Capillary electrophoresis; PAGE: polyacrylamide gel electrophoresis; SS: silver staining, PISSR: *Pongamia* Inter-Simple Sequence Repeat; 2GS: Second generation sequencing.

Introduction

Pongamia pinnata (L.) Pierre (also known as *Millettia pinnata* or Indian Beech, and hereafter referred simply referred to as *Pongamia*) is a perennial, fast-growing, leguminous tree, widely distributed on the Indian subcontinent, south-east Asia, Oceania, northern Australia, the East-African coast and southern China (Murphy et al., 2012). In addition *Pongamia* has been introduced to other parts of the world, including the United States (Hawaii, Florida; Scott et al., 2008; Biswas et al., 2011; Kazakoff et al., 2011). As a nodulating and nitrogen-fixing legume the nitrogen (N) fertilizer requirements of *Pongamia* are potentially minimal or eliminated. This is of advantage for a biofuel

* Corresponding author: p.gresshoff@uq.edu.au.

The last two authors made equal contribution to the research.

crop as N fertilizer inputs detract from the net energy gain, a fact often overlooked in choosing biofuel crops (Hill et al., 2006; Sutton et al., 2011). Nitrogen fertilizer is increasingly costly as it is linked to energy inputs through industrial synthesis, transportation, and application cost, and is a contributor to greenhouse gas (GHG, mainly N₂O) emissions (Crutzen et al., 2007; Fransen et al., 2011; Jensen et al., 2011) as well as environmentally significant ground- and surface-water pollution (Rockström et al., 2009; Sutton et al., 2011).

Pongamia distinguishes itself from other crop/forage legumes (e.g., soybean, pea, lupin, clover and alfalfa) and bioenergy/food crops (e.g., maize, sugarcane, jatropha, oil palm and canola) as it is not a food or animal feed/forage crop; the tree grows in multiple geo-climatic conditions, ranging from humid tropical/sub-tropical regions (such as coastal Queensland) to cooler and semi-arid zones (500-600 mm annual rainfall, 2200 mm annual evaporation; e.g., Roma, central Queensland). Pongamia survives winter temperatures of -5°C during its dormancy period. Seeds mature about 10-11 months after flowering. It requires a long warm summer (30-40°C day temperatures with warm nights (minimally 17°C), and is able to grow on low agricultural productivity (LAP) land, thus minimising competition with

food crops or related fertilizer/water/land resources needed for food and fodder production (Odeh et al., 2011). It thus satisfies all 'sustainability criteria' expected from modern second and third generation biofuel crops. Pongamia plants tolerate soil salinity of up to 20 dS.m⁻¹, will grow well at 10 dS.m⁻¹, increasing its cultivation range into saline water-fed arid regions (C. Wilkinson, CILR, UQ, unpublished data).

Recent interest in Pongamia has centered on its abundant production of large (about 1.3-2.6 g/seed average mass), oil-rich seeds (Brisbane area data). The mature seed contains about 30-40% oil, which itself contains about 55% oleic acid (a critical fatty acid for high quality biofuel needed for transportation; Scott et al., 2008; Kazakoff et al., 2011), with potential annual oil yields of 3-5 tons per hectare. As a result of the recognition that Pongamia has the capacity to produce large, oil-rich seeds it is now widely acknowledged that it offers a potential solution as a highly desirable oil feedstock for transport biofuel production (Azam et al., 2005; Karmee and Chadha 2005; Meher et al., 2006; Scott et al., 2008; Graham et al., 2011; Kazakoff et al., 2011).

However, currently available Pongamia trees have undergone minimal amount of domestication. Perhaps some selection

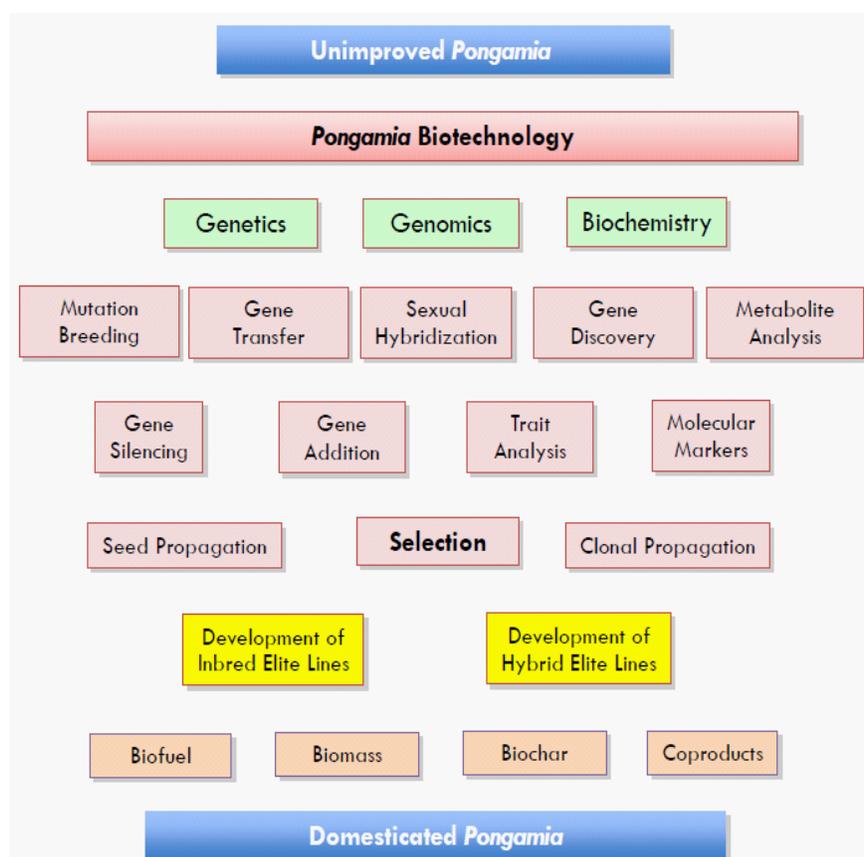


Figure 1. Proposed pathway for Pongamia domestication using modern biotechnology. Core disciplines, though interacting, are shown in 'green'. These comprise specific areas of analysis ranging from classical genetics to metabolite analysis (in pink). 'Selection' under specific environments, recognizing the classical Genotype X Environment (GXE) interaction is essential. The successful combination of these approaches will lead to the development of elite tree lines, producing a range of endproducts (in beige).

for better growth occurred in India during its long period of rural application as a low technology biofuel crop, but no directed crop improvement program has been undertaken. Domestication of *Pongamia* now requires the use of modern biotechnology, coupling biology, genetics, genomics and biochemistry (Figure 1). A range of genetic approaches is possible, reflecting translational biology, where techniques and concepts developed for established crops (i.e., soybean and canola) are applied to this newly emerging crop. The advent of fast and low cost DNA sequencing and associated bioinformatics opens the possibility of rapid gene discovery as well as molecular marker development. These advances allow the coupling of structural components to functional traits, accelerating selection and crop improvement.

Although the current attributes provide a competitive advantage to *Pongamia* over many other biofuel feedstock species, the out-crossing nature of *Pongamia* prevents the establishment of uniform tree cultivars with elite phenotypes, other than by vegetative propagation involving grafting, rooted cuttings, or tissue culture propagules. Thus, raising plants from seeds results, among many, in variability of growth, plant structure, time of flowering, seed ripening and the quality of seed for biofuel production (Biswas et al., 2011).

Rapid and reliable molecular tools for the evaluation of diversity in *Pongamia* germplasm are needed, especially as many traits are only assayable at later plant age. Quantification of such diversity will be aided by DNA markers for desired traits. For example, SSR (Simple Sequence Repeat) and ISSR (Inter-Simple Sequence Repeat) markers have been used in a wide range of organisms for diverse applications, including linkage map development, diversity tests to assess breeding material, protection of Plant Variety Rights (PVR), gene isolation and discovery, and association tagging for quantitative trait loci (QTLs) (Gupta and Varshney, 2000; González et al., 2005; Cobos et al., 2009; Sharma et al., 2009; El Aabidine et al., 2010; Lin et al., 2010; Pérez-Vega et al., 2010; Tanya et al., 2010).

An advantage of ISSRs is that in their simplest form, no prior genomic knowledge is required to design arbitrary sequence oligonucleotide primers (Gonzalez et al., 2005). Such primers, targeting different repeats of a nucleotide core sequence (e.g., GAGAGAGAGAGA), anchored either at their 3' or 5' termini by a 2 to 4 nucleotide extension, are used to amplify genomic sequences flanked by two inversely oriented nucleotide core repeats (Zietkiewicz et al., 1994). In the present study we improved on this approach by designing primers based on multiple NGS sequence tags with their 'inherent' extensions, amplifying genomic DNA of individuals then separating and detecting the amplification products (called PISSRs) by thin-layer polyacrylamide gel electrophoresis (PAGE) and by silver staining (Bassam et al., 1991; Bassam and Gresshoff, 2007), or capillary electrophoresis (CE) with fluorescent labelling, respectively. This resolved robust genetic differences in *Pongamia*.

With increasing interest in *Pongamia* as a future biofuel crop, molecular approaches to its genetic characterization, specifically of germplasm from the Indian sub-continent, have been recently published (Rout et al., 2009; Sujatha et al., 2009; Kesari et al., 2010; Sahoo et al., 2010; Sharma et al., 2011). These studies have utilized RAPD (random amplification of polymor-

phic DNA), AFLP (amplified fragment length polymorphism), and ISSR-based methodologies for applications in genetic diversity and variability assessment of natural populations as well as clonally propagated germplasm. With particular reference to the studies assessing genetic diversity via ISSRs, primers were based principally on arbitrarily derived di-nucleotide repeats and detection in a low-resolving agarose matrix. In all of these studies polymorphisms were easily detected with relatively few primers required. The level of polymorphism was high, reflecting the outcrossing nature of *Pongamia*, with the general consensus that polymorphism was positively correlated with the geographical distance between populations of trees. Studies did not evaluate diversity of single seeds and individuals, but focused on populations.

Here we describe a further advance on previous studies, namely the use of specifically anchored, *Pongamia* 'deep DNA sequence'- derived ISSR markers, thus representing actual *Pongamia* genome sequence. Our study focused on individual seeds/trees and their progeny rather than pooled populations. The so-called PISSR (*Pongamia* Inter-Simple Sequence Repeat) primers were designed following analysis of our DNA sequence database that was constructed using Illumina® Solexa GAllx deep sequencing (Marshall et al., 2010) of a single *Pongamia pinnata* tree located in Brisbane, Australia.

As with all PCR-based DNA marker techniques, the basic steps for the PISSR technology involved a) extraction of *Pongamia* DNA; b) design of PISSR primers; c) PCR and electrophoretic separation of DNA products; d) detection of DNA markers or polymorphisms, and subsequent analysis for genetic relatedness. In contrast with the other published studies, we combined this approach with polyacrylamide gel electrophoresis (PAGE; Caetano-Anollés et al., 1991) and DNA silver staining (Bassam et al., 1991; Bassam and Gresshoff, 2007) to make PISSR robust and information-rich. In addition, the capability of capillary electrophoresis (CE) to resolve small molecular weight PISSR markers was examined. CE has advantages over the conventional slab PAGE in the detection of genetic polymorphisms. In terms of efficiency, CE requires only small quantities of DNA sample, short separation time, and has high sample throughput performance via an array system. The visualization of peaks is viewed either by semi-quantitative peak height or quantitative peak area (Petersen et al., 2003).

Materials and Methods

Plant Material and DNA Extraction

Plant material was collected from different locations in south-east Queensland (Australia) and the Kuala Lumpur region (Malaysia). To detect seed diversity, the seeds were germinated with 1:1 sand/soil in the glasshouse (18/6 h day/night cycle and 28°C/20°C day/night temperature regime). Young leaf tissues visually clean and unaffected by pathogens were collected for DNA extraction from seedlings two months after germination. Genomic DNA extraction was performed by a CTAB procedure (Murray et al., 1980; Doyle and Doyle 1987; Singh et al., 1999). The quality and quantity of the extracted DNA were

confirmed by measurements with a ND-1000 Spectrophotometer (NanoDrop Products, USA).

PISSR (Pongamia Inter-Simple Sequence Repeat) Primer Design and PCR

The approach utilized a Pongamia DNA sequence database recently generated via Illumina® Solexa GAllx deep DNA sequencing technology at UQ. The database was based on a total genomic DNA library from a single Brisbane tree constructed from fragments of average 3 kb size, resulting in paired end reads each of 75 bp. The presence of dinucleotide repeats (e.g., CAn, GAn, ATn, and CTn) in the Pongamia genome was determined by BLAST analysis (Altschul et al., 1990) of the database. From the paired end reads (75 bp), primers were designed on the basis of sequences containing eight repeats of dinucleotide core units with addition of the adjacent two or three nucleotides either at the 5' or 3' of the repeat (Table 5). PISSR primers were synthesized by Sigma-Aldrich®. PCRs were performed in a MJ Research thermal cycler with a thermal cycling profile consisting of denaturation for 3 min at 94°C, then 35 cycles of 45 s at 94°C, 30 s at the specific annealing temperature (this temperature varied depending on the %GC of the primer), and 1.5 min at 72°C, and a final extension cycle of 10 min at 72°C. Each PCR contained 1 unit of Taq DNA polymerase (Invitrogen, Carlsbad, USA), PCR buffer (20 mM Tris-HCl, pH8.4; 50 mM KCl), 0.2 mM dNTPs, 1.5 mM MgCl₂, 0.5 μM primers and 50 ng template DNA.

PCR Product Detection by PAGE and Recovery of DNA Markers

PCR amplification products were separated by polyacrylamide gel electrophoresis (PAGE) using a Mini-Protean II cell (Bio-Rad, Hercules, USA) and visualized following silver staining (Bassam et al., 1991; Bassam and Gresshoff, 2007). Separation was in 0.45 mm thick, 7.5 x 10 cm vertical slab gels of 5% polyacrylamide backed on GelBond PAG polyester film (Lonza, Rockland, USA) in TBE buffer, until the dye front reached the end of the gel. The polyester-backed polyacrylamide gels were air-dried and stored in a photo album as a "molecular archive", whereupon DNA fragments of interest were extracted, re-amplified, cloned and sequenced.

Small pieces of polyacrylamide gel containing the desired DNA fragment were carefully excised from dry or fresh gels with a sterile scalpel. In the case of dry gels each gel piece was cleaned by soaking in 95% ethanol. A scalpel was used to sharply delimit the desired DNA fragment, and the excised gel piece was then rehydrated (if needed) in 10 μl of sterile water. The gel segment was next placed in 20 μl of PCR reagents with the same primer as was used to generate the relevant DNA marker. Re-amplified PCR products were separated by PAGE and visualized by silver staining, as previously. Purified PCR products were then further characterized by DNA sequence analysis.

In addition, capillary electrophoresis (CE) was done by a MegaBACE™ 1000 capillary system (GE Healthcare Life Science, Piscataway, USA).

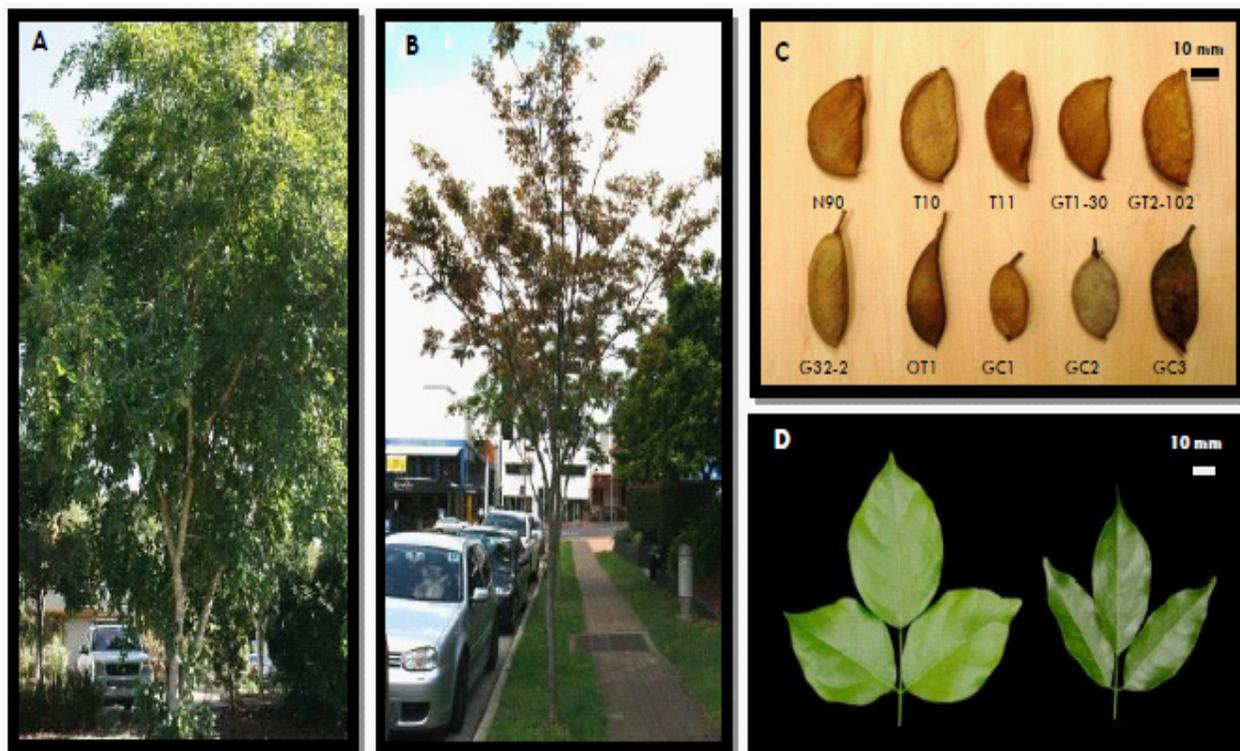


Figure 2. Some Pongamia trees displaying different phenotypes. A: T10 (7.5 m high, stronger body shape); B: GC2 (3.6 m high, thinner body shape, similar to Pongamia trees found in northern Queensland close to Cooktown); C: seed size and shape from 10 local trees including T10 and GC2; D: leaf shape (left: T10; right: GC2).

Table 1. Variation of seed mass, seed oil and oleic acid content in Pongamia trees.

No.	*sample ID	Seed mass (g)	% seed oil / seed mass	% oleic acid / seed oil
1	T10-6	1.34	50.5	51.6
2	T11	1.5	33.2	47.9
3	N90	1.37	32.6	39.9
4	G32-2	1.31	35.3	43.8
5	OT1	0.62	33.4	25.4
6	GC1	0.41	33.3	43.9
7	GC2	0.61	42.0	54.2
8	GC3	0.75	37.1	38.8
9	GT1-30	1.02	19.7	45.6
10	GT2-164	0.92	35.7	39.2
AVE±STD		0.99±0.4	35.3±7.8	43.1±8.1

*The seed samples were collected from Taringa, Milton, Gatton and Ascot (Brisbane, QLD) in December 2010.

Table 2. Variation of seed oil and oleic acid in progeny from a single tree.

No.	sample ID	Seed mass (g)	% seed oil / seed mass	% oleic acid / seed oil
1	T10-1	0.97	46.4	51.7
2	T10-2	1.07	52.3	60.8
3	T10-3	1.29	40.3	53.2
4	T10-4	1.34	42.5	56.9
5	T10-5	1.37	47.4	68.3
6	T10-6	1.34	50.5	51.6
AVE±STD		1.2±0.2	46.6±4.4	57.2±6.5

T10-1 to T10-6 are single seeds derived from mother tree T10.

Analysis of Genetic Similarity

PISSR polymorphic markers were scored manually using a binomial '1' and '0' matrix for their presence and absence, respectively. The level of genetic similarity among Pongamia individuals was established by clustering method UPGMA (unweighted pair-group arithmetic average) with the SHAN subroutine, through the software NTSYS-pc version 2.0. Dendrograms were used to represent the genetic relationship among the 29 local Pongamia trees.

Analysis of Seed Oil Content and Composition

Seed oil was extracted by the chloroform/methanol extraction procedure (Schmid 1973; Christie 1993) using finely chopped individual seeds. Fatty acids were analysed using gas chromatography (Shimadzu GC-17A, Japan) on a DB-23 60 m x 0.25 mm x 0.25 µm capillary column with GC-FID (Shimadzu Co., Japan) by Analytical Services, School of Agriculture and Food Science, UQ.

Bioinformatics Analysis with DNA Sequence of the PISSR Markers

DNA sequencing was performed at the Australian Genome Research Facility (AGRF), The University of Queensland. Bioinformatics analysis of the DNA sequences from PISSR markers was performed using public databases such as NCBI (<http://www.ncbi.nlm.nih.gov>); Gene Indices (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/Blast/index.cgi>); Lotus japonicus EST index (<http://est.kazusa.jp/en/plan>) or Phytozome (<http://www.phytozome.net/soybean>). From these databases, a BLAST-search of DNA sequences amplified by PISSR markers identified putative Pongamia genes. The DFCI gene indices database

Table 3. Selected PISSR primers used for DNA marker analysis by PAGE/SS.

Primer	Primer sequence	Total number of bands	Range of marker sizes (bp)	Number of polymorphic markers*
PISSR1	5'(GAGAGAGAGAGAGAGA)AT3'	20-23	250-1600	15
PISSR2	5'(GAGAGAGAGAGAGAGA)AA3'	18-21	300-1700	12
PISSR3	5'(GAGAGAGAGAGAGAGA)CG3'	16-22	400-1850	11
PISSR4	5'(GAGAGAGAGAGAGAGA)TG3'	16-21	350-1800	11
PISSR5	5'(GAGAGAGAGAGAGAGA)TA3'	13-16	350-1700	10
PISSR6	5'(GAGAGAGAGAGAGAGA)GA3'	18-21	600-1700	9
PISSR7	5'CA(GAGAGAGAGAGAGAGA)3'	15-19	300-1650	0
PISSR8	5'GT(GAGAGAGAGAGAGAGA)3'	17-21	400-1900	0
PISSR13	5'(CACACACACACACA)AAC3'	15-17	550-1700	7
PISSR14	5'(CACACACACACACA)ATG3'	17-19	400-1900	8
PISSR17	5'(CACACACACACACA)TAG3'	10-15	650-1700	9
PISSR18	5'(CACACACACACACA)ATT3'	17-20	700-1700	13

* Total number of polymorphic markers generated from 29 Pongamia samples.

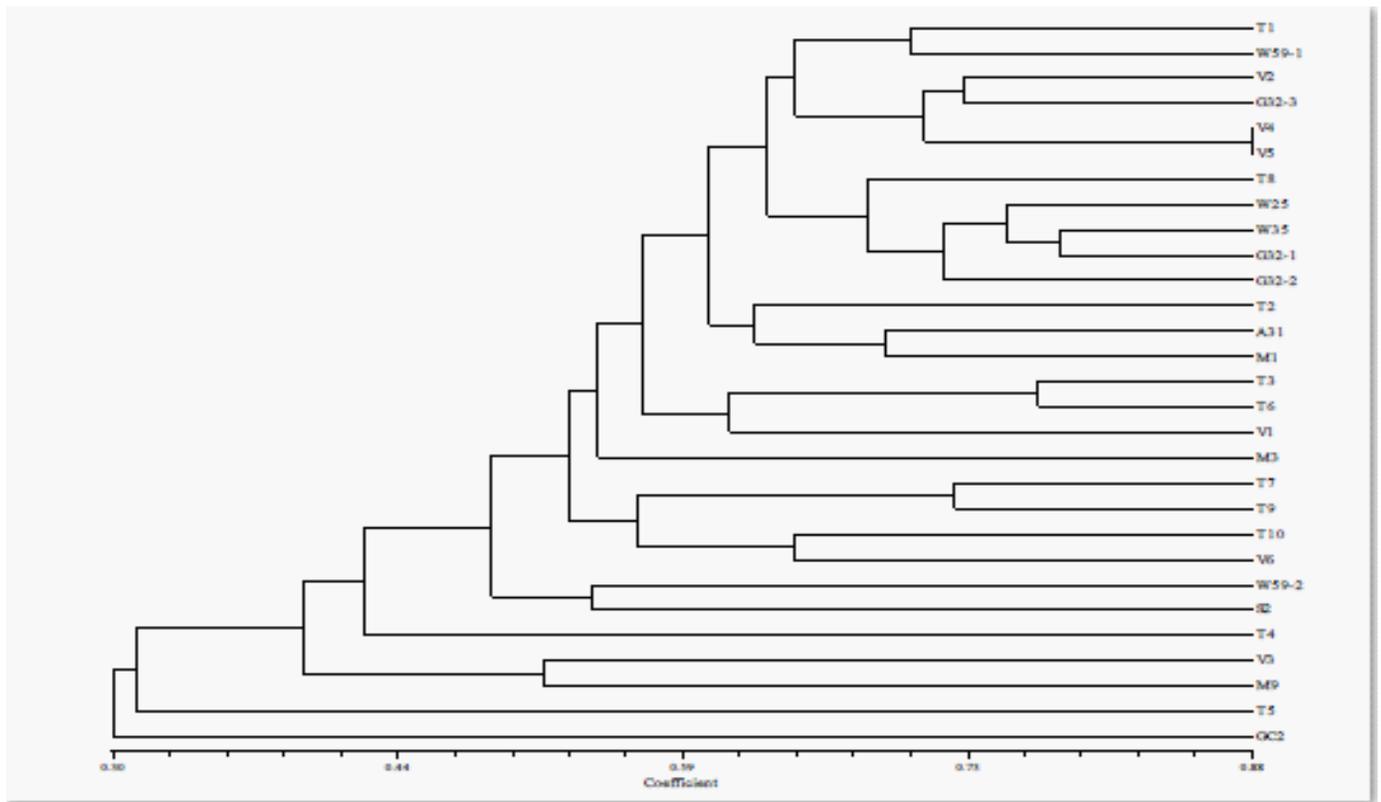


Figure 3. Genetic similarities of the individual Pongamia trees from South-east Queensland and Malaysia based on PISSR markers.

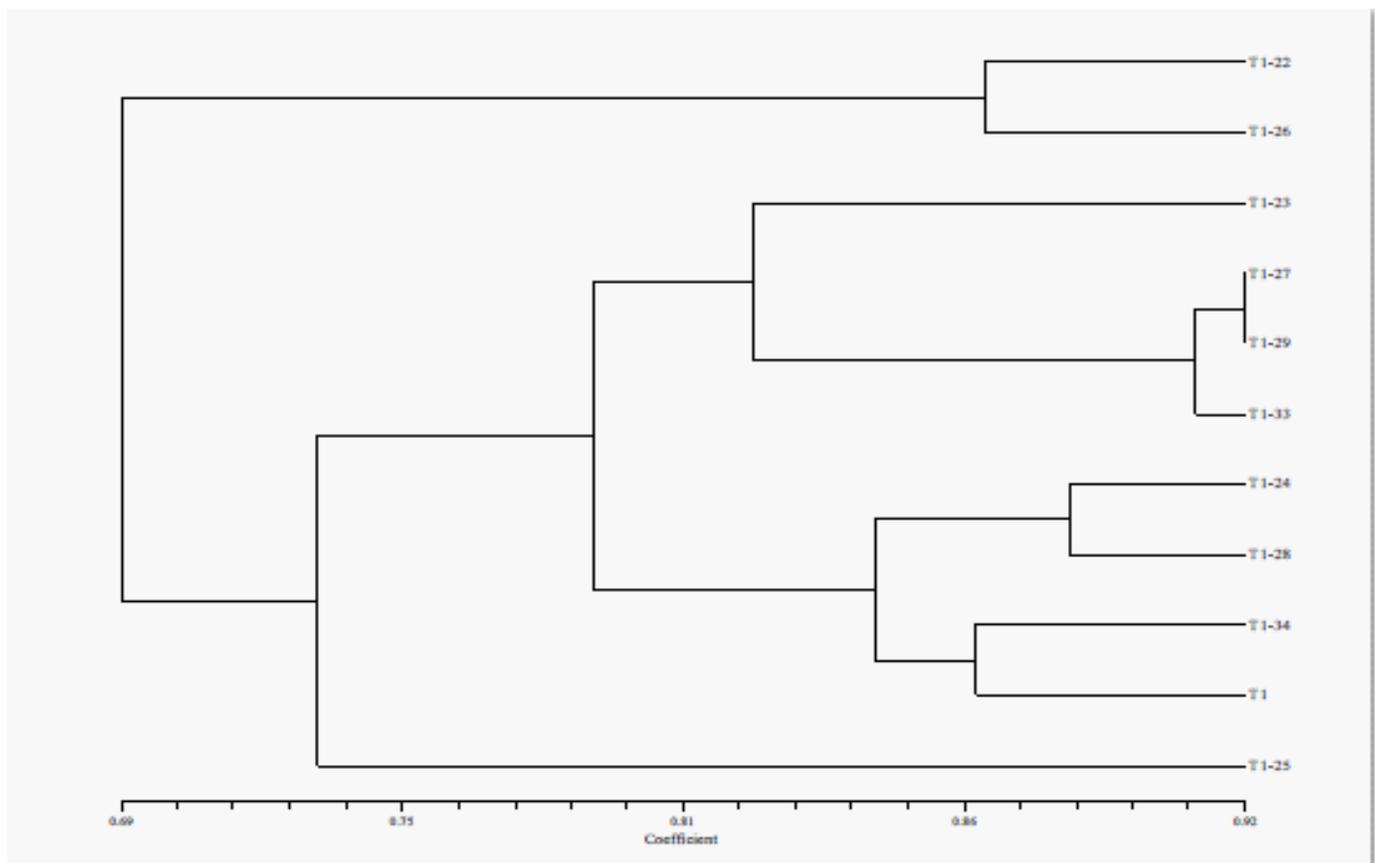


Figure 4. Genetic similarity using the progeny derived from a single Pongamia mother tree (T1) based on multiple PISSR markers.

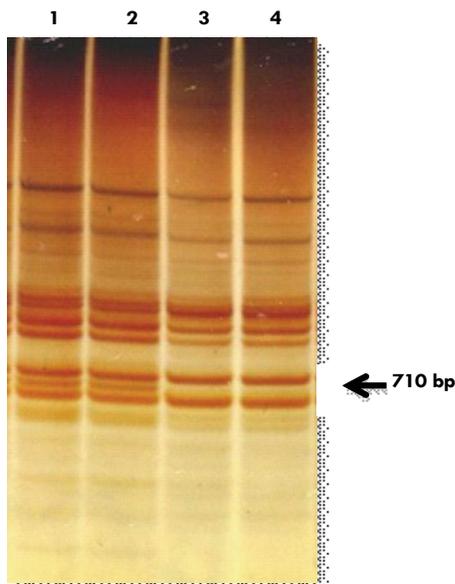


Figure 5. Reproducibility of PISSR markers derived from PISSR6 using clonal *Pongamia* trees. 1 = mother tree W35; 2 = clonal duplicate of W35; 3 = mother tree W25; 4 = clonal duplicate of W25.

provided access to the UniProtKB/Swiss-Prot database (<http://www.uniprot.org>) to allow for more insight of functional similarities if markers were related with protein-encoding sequences.

Results

Phenotypic Diversity of *Pongamia*

The genetic diversity in a randomly chosen set of *Pongamia* trees was reflected in distinct phenotypic differences at the gross level. As an example, Figure 2 displays these differences in whole tree architecture and leaf morphology between south-east Queensland street trees T10 and GC2 (Figures 2A, B, and D). Furthermore, seed-derived *Pongamia* trees, planted for a life cycle analysis of *Pongamia* at the UQ Gatton campus, showed diversity for flowering time with 6% of trees flowering and setting seed precociously by 15 months of age. Significant differences in seed size, shape and weight were also observed (Figure 2C, Tables 1 and 2). Seed oil analysis showed variation of oil content and composition between trees and between progeny seeds of a single parent tree, T10 (Tables 1 and 2). For individual seeds from 10 randomly selected *Pongamia* trees, the seed mass, oil content and oleic acid/oil content varied from 0.41 - 1.5 g, 19.7 - 50.5 % and 25.4 - 54.2 %, respectively. Recently, genetic material with mature seed mass of 2.6 g/seed have been discovered in South-East Queensland, though these values may still be lower than those reported in an unsubstantiated way from India and Sri Lanka. The potential of using such selected germplasm in breeding cannot be underestimated.

The lower values appear to be derived from a set of distinct *Pongamia* trees (OT1, GC1, GC2, GC3; see Table 1), possibly belonging to an as-yet-undefined Australian sub-species. In contrast, six progeny seeds of tree T10 (a high performer) showed

less variation for seed mass, oil content and oleic acid/oil content (0.97-1.37 g/seed, 40.3 - 52.3 % per seed and 51.6 - 68.3 % oleic acid content). The results from seed oil analysis suggested that the variations of seed oil content are larger between seeds from different trees than between seeds from the same parent tree (Tables 1 and 2).

PISSR Primers Generated Extensive Polymorphic Bands

A total of 27 PISSR primers were tested in this study. All primers were based on a (GA)₈ or (CA)₈ motif, with an additional 5' or 3' di- or tri-nucleotide extension. These extensions were based on flanking DNA sequences from the paired-end Illumina® reads to limit the number of amplicons for diversity scoring and assessment. Although not utilized in this study, many 1- to 3-nucleotide core unit tandem repeats were identified in the NGS database. Figure S1 displays the DNA sequence of a selection of 75 bp reads and illustrates the typical di-nucleotide repeat sequences found in the *Pongamia* genome. Of the 27 primers tested, 23 had a 3' extension and four had a 5' extension. Importantly, all primers successfully enabled the reliable amplification of numerous DNA fragments (e.g., up to 23 bands for primer PISSR1; Table 3). Interestingly, all ISSR primers tested that had a 5' extension, were able to generate only common bands with no polymorphic markers able to discriminate between accessions, whereas all PISSR primers with a 3' extension were able to generate polymorphic markers. As an example, Figure S2a demonstrates the typical profile of common and polymorphic amplicons, in this case derived from nine local *Pongamia* trees using the primer PISSR4 ((GA)₈GTG). Resolution of amplicons by PAGE and silver staining enabled the routine scoring of bands in the size range of 250 to 1,900 bp.

To test the robust application of the methodology described above for the assessment of *Pongamia* genetic diversity, DNA was extracted from 29 trees, 26 from south-east Queensland and three from Malaysia. Genetic relatedness was determined following PCR with 12 PISSR primers (Table 3). The DNA profiles obtained by PAGE and silver staining were highly reproducible with clearly definable bands being scored as either conserved or polymorphic markers. Of the 12 PISSR primers used in this part of the study, 10 primers produced 105 conserved and polymorphic DNA fragments with apparent sizes from 250 bp to 1.9 kb. The number of reproducibly visible bands ranged from 10 to 23 for each primer (Table 3). From the pool of conserved and polymorphic amplification products, 7 to 15 polymorphic markers were generated per PISSR primer. The highest level of polymorphism (i.e., 75%) was detected with the primers PISSR1 (GA)₈AT and PISSR18 (GA)₈ATT. The number and size of the amplicons suggested that the PISSR primers were able to generate markers with a wide distribution and location in the genome of *Pongamia*.

CE was able to resolve fragments optimally in the size range of 80 to 400 bp. Detailed resolution of PISSR markers using CE was demonstrated (Figure S3). Table 4 lists the number of common and polymorphic DNA markers amplified from genomic DNA of 22 selected field samples with 8 PISSR primers. Due to the higher resolving power of CE, the laser detection enabled



Figure 6. Pongamia plantation growth and seed development. Upper panel: closely spaced trees at UQ Gatton campus (70 km west of Brisbane, Australia) showing juvenile branching as well as plant vigour at 15 months after germination. Average plant height was 2.5 m. Lower left: close-up of precociously formed pods with viable seeds. Lower right: Abundant plant biomass and foliage on precociously flower forming (PFF) Pongamia variant. About 6% of trees at Gatton showed the PFF trait.

identification of markers with a minimal difference of 1-2 bp (Figure S3). Thus CE offered maximal resolving power with more polymorphisms compared to PAGE/SS, but over a smaller size range. As an example, 53 polymorphic markers were generated from 22 Pongamia samples with primer PISSR22 by CE (Table 4), being almost three times more than those generated from PAGE/SS, even though the effective size detection range was restricted in CE. With primers PISSR 14, 17 and 18, the marker size ranged from 80-400 bp and 400 -1,900 bp for CE and PAGE/SS, respectively. These results suggest that a combination of both approaches is able to obtain more extensive polymorphic markers.

Genetic Similarity Analysis

Binomial scoring for the presence (1) or absence (0) of the 105 polymorphic markers generated a quantitative assessment of genetic similarity/diversity for 29 randomly selected trees.

The Jaccard's similarity coefficient ranged from 0.30 to 0.88 (Figure 3). UPGMA cluster analysis indicated that there was no correlation between the location of Pongamia trees and genetic similarity (Figure 3). For example, three Malaysian trees (M1, M3 and M9) were genetically interspersed amongst the remaining South-east Queensland trees. Malaysian tree M1 and Queensland tree A31 were in a cluster with a coefficient value of 0.67, while trees M9 and V3 were in another cluster with a coefficient value of 0.51. The reproductive origin of these trees is not known, but it is likely that each tree was grown from a seed, in part an explanation for their wide genetic diversity. Despite the relatively wide diversity amongst the tested accessions, this analysis generated a 'single rooted phylogenetic tree' (Figure 3), suggesting a common origin for Pongamia.

The PISSR approach demonstrated the outcrossing nature of Pongamia and the subsequent genetic variation between a parent plant and its seed-derived progeny. Four PISSR primers were used to characterize a single mature tree (T1) and ten progeny

Table 4. Selected PISSR primers used for DNA marker analysis by CE.

Primer	Sequence	Range of marker sizes (bp)	Maximal number of markers in one sample	Total number of polymorphic markers*
PISSR 14	5'(CA)8ATG3'	86-396	13	51
PISSR 17	5'(CA)8TAG3'	80-374	14	24
PISSR 18	5'(CA)8ATT3'	81-371	14	16
PISSR 20	5'(CA)8TCA3'	82-396	15	49
PISSR 22	5'(CA)8GTC3'	87-386	29	53
PISSR 24	5'(CA)8GCA3'	82-393	12	22
PISSR 25	5'(CA)8CTC3'	81-398	14	35
PISSR 26	5'(CA)8CGA3'	81-394	11	26

*Total number of polymorphic markers generated from 22 *Pongamia* samples.

Table 5. Nucleotide sequence relatedness of cloned PISSR markers extracted from PAGE/SS gels to three model legumes.

Primer	Marker	Sequence length (bp)	<i>Lotus japonicus</i>		Soybean (<i>Glycine max</i>) HSP (high scoring segment pairs)		<i>Medicago truncatula</i>	
			Length (bp)	Accession	Length (bp)	Accession	Length (bp)	Accession
PISSR1	QJ7	768	482	BW594779				
PISSR15	SH41	219	1237	TC57311	460	DB985567	438	BQ140302
PISSR16	SH23	1089			672	HO760602		
PISSR17	SH21	1028			1476	TC486715	1139	ES466846
	SH22	1054			672	HO760602	1229	EC366194
PISSR18	SH19	565	240	GE117618				
PISSR19	SH24	955			1477	TC486716	1230	EC366194
	SH34	478	260	G0034876	241	FG993806	254	EX528031
PISSR20	SH12	770			175	GD716163	819	ES613526
	SH49	135	712	TC62522	820	TC438781		
PISSR21	SH50	184	475	FS324643	742	TC437835	1024	TC189785
PISSR24	SH53	347					668	TC182451
PISSR26	SH38	184	588	TC78985	340	TC482151	532	TC188978
PISSR27	SH40	224			549	TC435357		

*Nucleotide similarities results via BLAST in CBI/GenBank and DFCI gene indices databases.

saplings; forty-six polymorphic markers were generated. The similarity coefficients for the parent tree and its progeny ranged from about 0.69 to 0.92 (Figure 4). More specifically, sapling T1-34 was most closely related to the parent tree T1 (similarity coefficient value 0.86). Saplings T1-24 and T1-28 were similarly highly related (0.88), while T1-27, T1-28 and T1-33 showed the highest value of greater than 0.92). T1-25 was the most distantly related sapling (0.73). Despite this level of genetic variation the parent tree T1 and its progeny were overall more closely related than tree T1 was to the other 28 *Pongamia* accessions described above (Figure 3).

Vegetative propagation through grafting, cuttings and/or tissue culture is an effective way to expand the numbers of elite *Pongamia* lines for large-scale, broad acre plantings (Biswas et al., 2011). To confirm the reliability of the PISSR marker approach, two clonal trees from rooted cuttings of parent trees W35 and W25 were tested. W25 and W35 DNA differed in one distinct polymorphic marker at 710 bp. As expected, this polymorphic marker together with other conserved DNA fragments were maintained in clonal replicates (Figure 5).

Annotation and Association of PISSR Markers

DNA fragments generated by PISSR primers were physically recovered from silver stained polyacrylamide gels and purified for DNA sequencing. Comparative analysis of the derived DNA sequences was performed relative to sequences deposited in the public databases NCBI (<http://www.ncbi.nlm.nih.gov/>), UniProtKB/Swiss-Prot (<http://www.uniprot.org/>), Phytozome (<http://www.phytozome.net/soybean>), Lotus EST index (<http://est.kazusa.or.jp/en/plant/lotus/EST/index.html>), or the Gene Index Project (<http://compbio.dfci.harvard.edu/tgi/tgipage.html>). Table 5 shows the nucleotide sequence relatedness of PISSR markers to genomic sequences of *L. japonicus*, *G. max* (soybean) and *M. truncatula*. These data were selected and tabulated from the greatest high-scoring segment pairs (HSP) searched within the genome of each species. These similar sequences came mainly from soybean, secondly in *M. truncatula*, and least from *L. japonicus* genomes (reflecting the levels of complete genome sequence determination of these legumes). This result suggested that it is possible to BLAST-search public DNA

sequences from PISSR markers at the levels of DNA, cDNA and amino acid sequence to search for potential gene similarities in *Pongamia*. Those marker sequences (shown in Table 5) were further analysed to infer their possible functional annotation related to sequences in *L. japonicus*, *G. max*, and *M. truncatula* (Table 6).

Discussion

Phenotypic diversity was easily determined in *Pongamia pinnata* plants. Here we demonstrated quantitatively the degree of variation in terms of seed size, seed oil content and seed oil composition (as indicated by oleic acid, C18:1). In parallel, molecular marker technologies were developed to give clear and more direct information of the genetic polymorphisms distinguishing particular accessions of *Pongamia*.

Advantages of many molecular marker techniques are that (i) no prior genomic sequence information is required, (ii) markers are stable, (iii) they are detectable at all developmental stages of an organism, and (iv) they are not cell specific (Agarwal et al., 2008). We advanced on these positive attributes of molecular marker technology as the Australian Centre for Plant Functional Genomics (ACPGF) and the ARC Centre of Excellence for Integrative Legume Research (CILR) exclusively created a *Pongamia* DNA SOLEXA-GALL database. The database provided 29,474,558 *Pongamia* 75 bp reads. This database provided the impetus to design PISSR primers using the abundant genetic information from these short reads. Thus, the special feature of PISSR primers is that the number of repeats of nucleotide core units and anchored 5' and 3' nucleotide residues of PISSR primers represented real *Pongamia* genome sequence. Therefore, PISSR primers are significantly distinguished from arbitrary ISSR primers, as reported by Zietkiewicz et al. (1994). We conducted a BLAST search of our *Pongamia* GALL database for different nucleotide core units (GA; CA; TA: AT) with different numbers of repeats and then designed PISSR primers according to needs.

The separation of complex DNA samples with high resolution by polyacrylamide gel electrophoresis (PAGE) has broad application. DNA silver staining has proven a very effective visualization method offering superior clarity and sensitivity (Bassam and Gresshoff, 2007). Amplifications with ISSR primers were usually resolved by agarose gel electrophoresis and ethidium bromide (EB) staining (Wolfe et al., 1998; Sahoo et al., 2010) or resolved by PAGE and visualized by autoradiography (Zietkiewicz et al., 1994). However, Gonzalez et al., (2005) used large acrylamide gels (380 x 320 mm) and silver staining to separate and visualize ISSR amplification products, allowing the distinction of sympatric wild and domesticated populations of common bean. Here we used both 'mini-PAGE' (100 mm x 80 mm) and silver staining methods to separate the PCR products amplified by PISSR primers. The advantages of PAGE/SS over agarose gels and ethidium bromide staining were obvious, as PAGE/SS displayed clear and sharp images, and highly sensitive visualization on polyacrylamide gels (Figure S2). Thus PAGE/SS was selected as a part of PISSR marker detection, as it allowed robust PISSR detection as well as subsequent band sequence determination.

Zietkiewicz et al. (1994) stated that the 3' anchored arbitrary ISSR primers of (CA)8RG or (CA)8RY, in which R stands for either purine and Y for either pyrimidine, resulted in marker sizes from 200 to 2,000 bp in various eukaryotic species. Table 3 showed that PISSR primers produced numerous markers with a similar size range. For example, primers (GA)8AT and (GA)8AA produced PISSR markers ranging from 250 to 1,900 bp. To expand the PISSR primer range (with a sequence of (GA)8 and two nucleotide extensions; ((GA)8+2), primers carrying a (CA)8 core unit and a three nucleotide extension at their 3' termini [(CA)8+3'] were generated and produced abundant markers. For '(CA)8+3' primers the smallest reliably detected fragments were 400 bp (instead of the 250 bp seen for [(GA)8+2'] and the percentage of polymorphic markers was fractionally lower than the average for the former set of primers. This suggests that there is a more stringent (locus specific) PCR amplification when using PISSR primers with longer nucleotide extensions at the 3' terminus.

The phylogenetic tree diagrams (Figures 3 and 4) were made on the basis of the presence or absence of the markers identified in ISSR amplicons. The dendrogram exhibited at least nine clusters with different coefficient values from 0.3 to 0.88, suggesting large genetic variation of the individual *Pongamia* trees from South-east Queensland and Kuala Lumpur based on 105 PISSR markers (Figure 3). As three Malaysian samples were classified to three clusters with some Queensland *Pongamia* trees, there is no evidence, at least from this study, indicating a correlation between geographic location and genetic similarity. However, we make this conclusion in the knowledge that we cannot describe in detail the ancestry of these tested *Pongamia* trees.

As described, the Jaccard's similarity coefficient ranged from 0.30 to 0.88 among the 29 *Pongamia* trees (Figure 3). In contrast, coefficient values for DNA products from progeny saplings T1 ranged from just 0.69 to 0.91 (Figure 4). The coefficient value range in T1 seeds demonstrated the closer kinship between T1 seeds and its parent than the relatedness between randomly selected trees (Figure 3). We conclude that PISSR polymorphisms occurred frequently among *Pongamia* individuals, but less so between related progeny, factually supporting presumed outcrossing breeding in *Pongamia*.

Previously reported ISSR analyses utilized DNA amplification primers that were arbitrarily designed to have nucleotide sequence repeats, with 1-3 nucleotides on the 3' or 5' termini, to enable randomly amplifying "inter-repeat" genomic sequences. A study described by Sahoo et al. (2010) used inter-sequence simple repeat (ISSR) analysis to examine genetic diversity between pooled samples from trees of different geographic locations in India. These amplified genomic sequences were used to assess genetic diversity between pooled Indian tree populations, but there was no attempt to correlate genotype with phenotype. However, our analysis correlated the extreme (outlier) genotype of Queensland *Pongamia* tree GC2 with its unique phenotypic characteristics (oil content and composition, leaf shape, seed shape, growth habit). This result means that PISSR amplification profiles from GC2 reveal polymorphic markers in concert with its phenotypic traits.

The PAGE and CE are specialized in separation of DNA

products with different size range, in this study 250 to 1,900 bp and 80 to 400 bp, respectively. CE offered higher resolving power than that of PAGE, but the range of marker size was more limited. Throughout the analysis of PAGE and CE, DNA markers generated by the majority of PISSR primers generated a reasonably even distribution across the range of sizes for both PAGE and CE (Tables 3, 4). Hence both approaches provide future opportunities to discover more informative DNA markers over an extensive range.

The development of molecular markers in the biofuel tree *Pongamia* opens the possibility for further crop improvement and domestication. These processes are slow and are especially hindered in a tree crop where key phenotypic traits, such as oil content or seed yield, are only expressed in a mature form. Finding molecular markers, which can be easily assessed at a juvenile stage, combined with low level (6%) precocious flowering as we observed in *Pongamia* (see Fig. 6), permits the generation of hybrid material of elite selected tree lines. This can form the basis for further breeding by hybridization, or clonal propagation using either organ culture, grafting or root cuttings. Such association mapping and development of molecular linkage maps for both single gene traits and QTLs are now possible in the near future. PISSR regions are likely to yield conventional SSR markers for clearer and faster association to traits. Together, these molecular genetic approaches will help advance the biotechnological improvement of *Pongamia pinnata*.

Acknowledgements

The authors gratefully acknowledge funding support from the Australian Research Council, The University of Queensland, The Global Change Institute (UQ), and BioEnergy Research Pty Ltd. We thank Dr Tee Chong Siang for sending *Pongamia* DNA samples from Malaysia and Dr Jennifer Waanders for oil analysis.

References

- Agarwal M, M Shrivastava, and H Padh (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 27 (4): 617–631.
- Azam MM, A Waris, NM Nahar (2005) Prospects and potential of fatty acid methyl esters of some non-traditional seed oils for use as biodiesel in India. *Biomass and Bioenergy* 29: 293–302.
- Altschul SF, W Gish, W Miller, EW Myers, and DJ Lipman (1990) Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Ambus P, U Skiba, K Butterbach-Bahl, and MA Sutton (2011) Reactive nitrogen and greenhouse gas flux interactions in terrestrial ecosystems. *Plant Soil* 343: 1–3.
- Bassam BJ, G Caetano-Anollés, and PM Gresshoff (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* 196 (1): 80–83.
- Bassam BJ, and PM Gresshoff (2007). Silver staining DNA in polyacrylamide gels. *Nature Protocols* 2: 2649–2654.
- Biswas B, PT Scott, and PM Gresshoff (2011) Tree Legumes as Feedstock for Sustainable Biofuel Production: Opportunities and Challenges. *Journal of Plant Physiology*, available online: doi:10.1016/j.jplph.2011.05.015.
- Caetano-Anollés G, BJ Bassam, and PM Gresshoff (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *BioTechnology* 9 (6): 553–557.
- Christie WW (1993) *Advances in Lipid Methodology - Two*, pp. 195–213.
- Cobos MJ, P Winter, M Kharrat, JI Cubero, J Gil, T Millan and J Rubio (2009) Genetic analysis of agronomic traits in a wide cross of chickpea. *Field Crops Res.* 111: 130–136.
- Crutzen PJ, AR Mosier, KA Smith and W Winiwarter (2007) N₂O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmos. Chem. Phys. Discuss.* 7: 11191–11205.
- CSIRO (2011) Flight path to sustainable aviation. Commonwealth Scientific and Industrial Research Organization (CSIRO).
- Doyle JJ and JL Doyle (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- El Aabidine AZ, J Charafi, C Grout, A Doligez, S Santoni, A Moukhli, C Jay-Allemand, C El Modafar, and B Khaadari, (2010) Construction of a genetic linkage map for the olive based on AFLP and SSR markers. *Crop Sci.* 50 (6): 2291–2302.
- Fransen T, P Bhatia and A Hsu (2007) *The Greenhouse Gas Protocol, Measuring to Manage: A Guide to Designing GHS Accounting and Reporting Programs – Full report 2007.* World Resource Institutes <http://www.ghgprotocol.org/files/measuring-to-manage.pdf>
- Gonzalez A, A Wong, A Delgado-Salinas, R Papa and P Grepts (2005) Assessment of Inter Simple Sequence Repeat Markers to differentiate sympatric wild and domesticated populations of common bean. *Crop Science* 45 (2): 606–615.
- Graham PW, LJ Reedman, L Rodrigues, J Raison, A Braid, V Haritos, P Adams, TS Brinsmead, JA Hayward, J Taylor and D O’Connell (2011) Sustainable aviation fuels roadmap; data assumptions and modeling, Commonwealth Scientific and Industrial Research Organization (CSIRO).
- Gupta PK and RK Varshney (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113: 163–185.
- Hill J, E Nelson, D Tilman, S Polasky and D Tiffany (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc. Nat. Acad. Sci.* 103 (30): 11206–11210.
- Jensen ES, MB Peoples, RM Boddey, PM Gresshoff, H Hauggaard-Nielsen, BJR Alves and MJ Morrison (2011) Legumes for mitigation of climate change and feedstock in a bio-based economy – A review, *Agronomy & Sustainable Development.* DOI 10.1007/s13593-011-0056-7.
- Kazakoff SH, PM Gresshoff, PT Scott, (2011) *Pongamia pinnata*, a sustainable feedstock for biodiesel production. In: Halford NG, Karp A, editors. *Energy Crops*, Cambridge, UK: Royal Society for Chemistry, pp. 233–258.
- Karmee, SK and A Chadha (2005) Preparation of biodiesel from crude oil from *Pongamia pinnata*. *Bioresources Tech.* 96 (13): 1425–1429.
- Kesari V, VM Sathyanarayana, A Parida and L Rangan (2010) Molecular marker-based characterization in candidate plus trees of *Pongamia pinnata*, a potential biodiesel legume, *AoB Plants* 2010: plq017.
- Lin KH, WL Yeh, HM Chen and HF Lo (2010) Quantitative trait loci influencing fruit-related characteristics of tomato grown in high-temperature conditions. *Euphytica* 174: 119–135.
- Marshall DJ, A Hayward, D Eales, M Imelfort, J Stiller, PJ Berkman, T Clark, M McKenzie, K Lai, C Duran, J Batley and D Edwards (2010) Targeted identification of genomic regions using TAGdb. *Plant Methods* 6, 19. doi:10.1186/1746-4811-6-19.
- Meher LC, SD Vidya, SN Naik (2006) Optimization of alkali-catalyzed transesterification of *Pongamia pinnata* oil for production of biodiesel. *Bioresources Tech.* 97 (12): 1392–1397.

- Murphy HT, DA O'Connell, G Seaton, RJ Raison, LC Rodriguez, AL Braid, DJ Kriticos, T Jovanovic, A Abadi, M Betar, H Brodie, M Lamont, M McKay, G Muirhead, J Plummer, NL Arpiwi, B Ruddie, PT Scott, C Stucley, B Thistlethwaite, B Wheaton, P Wylie, PM Gresshoff (2012) A common view of the opportunities, challenges and research actions for *Pongamia* in Australia. *BioEnergy Research* DOI 10.1007/s12155-012-9190-6
- Murray MG, WF Thompson (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8 (19): 4321-4325.
- Odeh IOA, DKY Tan and T Ancev (2011) Potential suitability and viability of selected biodiesel crops in Australian marginal agricultural lands under current and future climates. *Bioenergy Research* 4: 167-179.
- Pérez-Vega E, A Pañeda, C Rodríguez-Suaréz, A Campa, R Giraldez and JJ Ferreira (2010) Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 120: 1367-1380.
- Petersen JR, AO Okorodudu, A Mohammad, DA Payne (2003) Capillary electrophoresis and its application in the clinical laboratory. *Clinica Chimica Acta* 330 (1): 1-30.
- Rout GR, DP Sahoo and S Aparajita (2009) Studies on inter and intra-population variability of *Pongamia pinnata*: a bioenergy legume tree. *Crop Breed. Appl. Biotechnol.* 9: 268-273.
- Rockström J, M Falkenmark, L Karlberg, H Hoff, S Rost, D Gerten (2009) Future water availability for global food production: the potential of green water for increasing resilience to global change, *Water Resources Research*.45: W00A12. doi: 10.1029/2007WR006767.
- Scott PT, L Pregelj, N Chen, JS Hadler, MA Djordjevic and PM Gresshoff (2008) *Pongamia pinnata*: An untapped resource for the biofuels industry of the future. *Bioenergy Res.* 1: 2-11.
- Sahoo DP, S Aparajita and GR Rout (2010) Inter and intra-population variability of *Pongamia pinnata*: a bioenergy legume tree. *Plant Syst. Evol.* 285 (1-2): 121-125.
- Sharma SN, V Kumar and S Mathur (2009) Comparative analysis of RAPD and ISSR markers for characterization of sesame (*Sesamum indicum* L) genotypes. *J. Plant Biochem. Biotechnol.* 18 (1): 37-43.
- Sharma SS, MS Negi, P Sinha, K Kumar and SB Tripathi (2010) Assessment of genetic diversity of biodiesel species *Pongamia pinnata* accessions using AFLP and three endonuclease-AFLP. *Plant Molecular Biology Reporter* 29 (1): 12-18.
- Schmid P (1973) Extraction and purification of lipids: II. Why is chloroform-methanol such a good lipid solvent? *Physiol. Chem. Phys.* 5: 141-150.
- Singh M, Bandana, PS Ahuja (1999) Isolation and PCR amplification of genomic DNA from market samples of dry tea. *Plant Molecular Biology Reporter* 17 (2): 171-178.
- Stemmer WPC (2002) Molecular breeding of genes, pathways and genomes by DNA shuffling. *Journal of Molecular Catalysis B: Enzymatic* 19-20: 3-12.
- Sujatha K, AV Rajwade, VS Gupta and L Hazra (2010) Assessment of *Pongamia pinnata* (L.) – a biodiesel producing tree species using ISSR markers. *Curr. Sci.* 99 (10): 1327-1329.
- Tanya P, P Taepayoon, Y Hadkam and P Srinives (2010) Genetic diversity among *Jatropha* and *Jatropha*-related species based on ISSR markers. *Plant Molecular Biology Reporter* 29 (1): 252-264.
- Sutton MA, CM Howard, JW Erisman, G Billen, A Bleeker, P Grennfelt, HV Grinsven and B Grizzetti (2011) (eds) *The European Nitrogen Assessment* (Cambridge Univ. Press); available at <http://go.nature.com/5n9lsq>.
- Wolfe A, Q Xiang and S Kephart (1998) Assessing hybridization in natural populations of *Penstemon* (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Molecular Ecology* 7: 1107-1125.
- Zietkiewicz E, A Rafalski and D Labuda (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176-183.

Table S1. Nucleotide sequences of PISSR primers.

PISSR primers	Primer sequence 5' -> 3'
PISSR1	(GA) ₈ AT
PISSR2	(GA) ₈ AA
PISSR3	(GA) ₈ CG
PISSR4	(GA) ₈ TG
PISSR5	(GA) ₈ TA
PISSR6	(GA) ₈ CA
PISSR7	CA(GA) ₈
PISSR8	GT(GA) ₈
PISSR9	AA(GA) ₈
PISSR10	TC(GA) ₈
PISSR11	TA(GA) ₈
PISSR12	AG(GA) ₈
PISSR13	(CA) ₈ AAC
PISSR14	(CA) ₈ ATG
PISSR15	(CA) ₈ AGA
PISSR16	(CA) ₈ ACT
PISSR17	(CA) ₈ TAG
PISSR18	(CA) ₈ TTT
PISSR19	(CA) ₈ TGC
PISSR20	(CA) ₈ TCA
PISSR21	(CA) ₈ GAG
PISSR22	(CA) ₈ GTC
PISSR23	(CA) ₈ GGT
PISSR24	(CA) ₈ GCA
PISSR25	(CA) ₈ CTC
PISSR26	(CA) ₈ CGA
PISSR27	(CA) ₈ CCT

- 1) GAGGGAA**GAGAGAGAGAGAGAGAA**AGAGAGGTGTGGGTGTGTGGATGAAGGAGGGGAAGAAGGGTATTTAGGT
- 2) TAATAATAATAATATGTTTTGTTCTAAAA**GAGAGAGAGAGAGAGAGAC**GATGACGTTTATTGTAAATTATAAATTG
- 3) TTGGGAAGTT**GAGAGAGAGAGAGAGAG**AGGTA AAA GAA AAG AAGT GAGAGACAGAGAAATATAATAACGTACAAG
- 4) GAGGG**AGAGAGAGAGAGAGAGAA**AGAGAGGTGTGGGTGTGTGGATGAAGGAGGGGAAGAAGGGTATTTAGGT
- 5) TATTTAATGTATAAAATTTAAA**ATATATATATATATAT**CTTGACCGGTTCCGACCCTGGTTGAACCACTAAAC
- 6) GTTGCAAAAACCTATTTCTGTGTTCTGTTTATT**ATATATATATATATAT**CTGCTGCCACCTAAACCATGCT
- 7) CACAACCTCAATTGCATTCAATTA**ACACACACACACACA**AAAGCTTATTAGTTGACATACCTTTATAGA
- 8) GCGAACGT**ACACACACACACACACAT**AGAGA AATATAA AAA AATCTTTTTTAAAAAACGTTTGAGAGTTTGT
- 9) GGCTGCATCACCTATCATTCTCAAACCTAG**CTCTCTCTCTCTCT**CATCTCTCTCAAAAACCTTTACAAG
- 10) CCTCTCTTCTCACTTGACAC**CTCTCTCTCTCTCT**GGCCACCATTGATTTTCCAACCTCTTTAAGAAGTTTG

Figure S1. Selected SOLEXA 75 bp reads picked for PISSR primers design. The sequences in red targeted different repeats of nucleotide core units (GA, AT, CA, and CT), which are anchored either at the 3' or 5' termini of the repeats by a 2 to 3 nucleotide extension (in green).

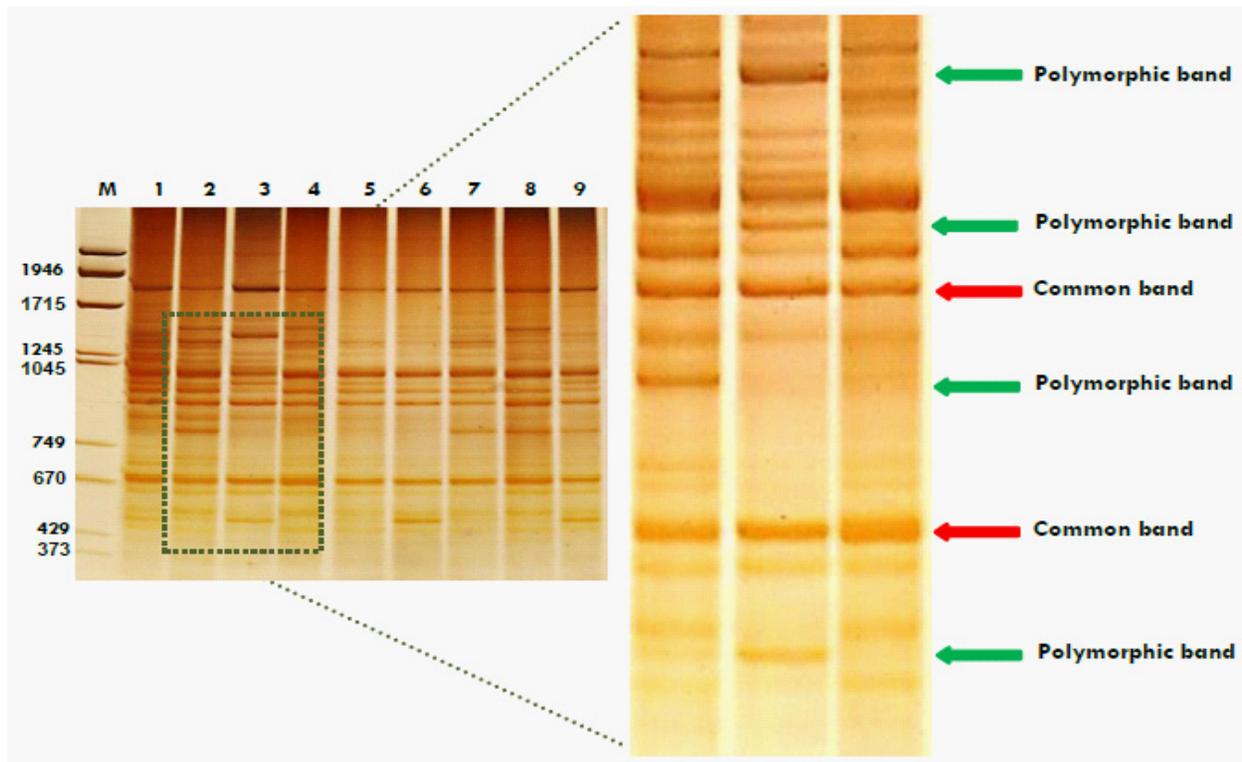


Figure S2. Molecular diagnostics of PISSR markers using PAGE and silver staining. Left: original silver-stained polyacrylamide gel, M, molecular weight marker (bp); 1-9, individual Pongamia trees; Right: Partially enlarged polyacrylamide gel, clearly displaying polymorphic and conservative bands. The PCR products were amplified with primer PISSR4. The PISSR marker size ranges from 350 to 1,800 bp.

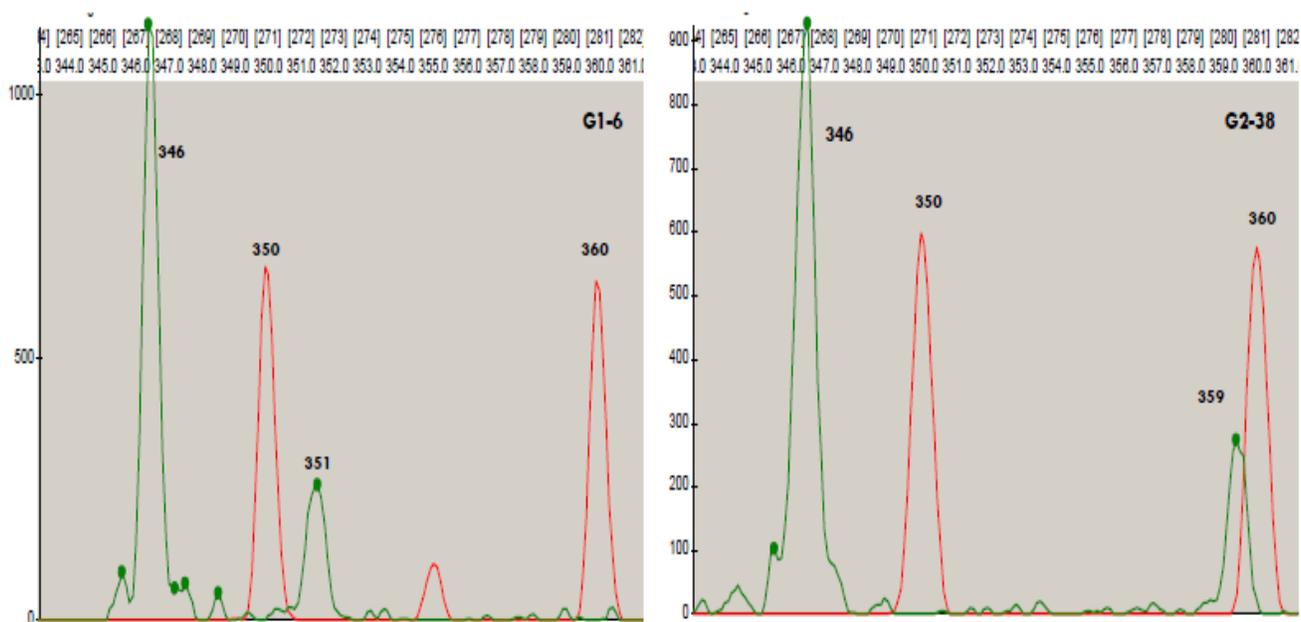


Figure S3. Molecular diagnostics of PISSR markers using capillary electrophoresis. This method is able to resolve fragments optimally in the size range of 80 to 400 bp by tagged fluorescent label HEX. Primer PISSR22 was used for displaying the genetic differences. The visualization of peaks is viewed in either a manner of semi-quantitative peak height or quantitative peak area. Position of red-coloured peak (ladder, from left to right): 350 bp; 360 bp. Position of green-coloured peak (from left to right): 346 bp, derived from both Pongamia trees G1-6 and G2-38 as conservative peak; 351 bp, Polymorphic peak in G1-6; 359 bp, Polymorphic peak in G2-38.