Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Archives of Applied Science Research, 2014, 6 (4):209-218
(http://scholarsresearchlibrary.com/archive.html)

ISSN 0975-508X
CODEN (USA) AASRC9

Development and polymorphism of simple sequence repeats (SSRs) in Kelampayan(*Neolamarckia cadamba*–Rubiaceae) using ISSR suppression PCR method

Seng-Loi Phui[1], Wei-Seng Ho[1*], Shek-Ling Pang[2] and Julaihi Abdullah[2]

[1]*Forest Genomics and Informatics Laboratory (fGiL), Department of Molecular Biology, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak*
[2]*Applied Forest Science and Industry Development (AFSID), Sarawak Forestry Corporation, Kuching, Sarawak*

ABSTRACT

Simple sequence repeat (SSR) marker is a polymerase chain reaction (PCR)-based marker system which has become a marker of choice for understanding plant genetic diversity and a powerful tool in addressing genetic resources questions. However, the availability of SSR markers especially for forest tree species is limited thus far due to the high development cost, labour-intensive and time-consuming. The present study aimed to develop an array of SSR markers for Neolamarckia cadamba using inter-simple sequence repeat (ISSR) suppression PCR method and further assessed the polymorphisms and transferability of the markers to other species.In total, 15 out of 31 SSR markers specific for N.cadamba were successfully developed and further characterized and validated by using 30 individuals of N. cadamba. The markers exhibited a considerable high level of polymorphism across the tested N. cadamba genotypes whereby 66 alleles were detected with an average of four alleles per locus. Most of the SSRloci analyzed showed high polymorphism as indicated by their PIC value which was above 0.5. The most polymorphic lociwere: NCAC11 (PIC=0.849), NCAC12 (PIC=0.722) and NCAG01 (PIC=0.712).The transferability rate was ranging from 26.7% to 73.3% among the four cross-genera species tested. The present study is the first report on the development of SSR markers in N. cadamba. These markers provide valuable genomic resources that could pave the way for exploiting SSR genotype data for effective selection of plus trees, provenance trials and establishment of forest seed production areas (SPAs) of N. cadamba in the selected forest areas dedicated to planted forest development, and molecular breeding of N. cadamba and other indigenous tropical tree species in future.

Keywords: SSRs, *Neolamarckia cadamba*, kelampayan, ISSR suppression PCR method, genetic diversity, molecular breeding

---

INTRODUCTION

Understanding the genetic variation of forest tree species is a prerequisite for the appropriate utilisation of forest genetic resources. This information will be a basis for establishing tree improvement programme and for management or conservation of natural communities. Our previous findings have shown that populations with little genetic variation are more vulnerable to the arrival of new pests or diseases, pollution, changes in climate and habitat destruction due to human activities or other catastrophic events [16, 19, 46]. Such genetic information of a forest tree species is assessable either through measurement of morphological characters in the field or through the study of molecular markers in the laboratory. However, those morphological characters can be influenced by environmental factors meanwhile molecular markers avoid many of the complications of environmental factors acting upon characters by looking directly at variation controlled by genes or by looking at the genetic material itself [14].

The development of an ideal molecular DNA marker system which is genetically co-dominant and highly polymorphic allowing precise discrimination even of closely related individuals is becoming a major concern due to the genetic complexity of breeder's populations and high levels of heterozygosity in individual genotypes [12]. As such microsatellites or simple sequence repeat (SSR) marker is the ideal marker system of choice for the tree breeders. The nature of SSR markers itself also provide them a number of advantages over other molecular markers such as RAPD, DAMD or ISSR markers[19, 22, 42]. Because of these attributes, SSRs loci are currently the excellent markers of choice for individual genotyping and studies of gene flow, mating system, genetic diversity and structure, cultivar identification, marker-assisted selection and pedigree analysis in forest tree species. For instance, four SSR loci were successfully used to determine genetic relatedness among selected mother trees of *Shorealeprosula* and *Dipterocarpuscornutus*[46]. They found that four and three selected mother trees of *S. leprosula* and *D. cornutus*, respectively were not closely related and therefore, could be used as potential seed sources for an advanced breeding programme and tree plantations.

Simple sequence repeats (SSRs)are tandemlyrepeated motifs of 1-6 nucleotides found in all prokaryotic and eukaryotic genomes. The uniqueness and value of microsatellites arises from their multiallelic nature, codominant transmission, relative abundance and extensive genome coverage[35].They provide a powerful tool for addressing genetic questions related to breeding where genetic discrimination is paramount. These markers can be used to gain more information on the level and distribution of genetic variation which is particularly a prerequisite for tree improvement and conservation programmes for forest trees[29]. Although SSRs are powerful tools for studies of populations, the development of species-specific primers could be a time-consuming and expensive process. Conventionally, isolation of SSR loci involves the construction of restricted DNA library with enrichment step and screening with SSR-specific probes. Such isolation procedures are usually labor-intensive, time-consuming and expensive. To overcome this disadvantage, an alternative protocol proposed by Lian et al. [26], Inter-simple sequence repeat (ISSR) suppression-PCR method was used to isolate SSR loci. This method is relatively simple, rapid without enrichment and screening, and requires only basic skill in molecular biology.

Neolamarckia cadamba(Roxb.)Bosser or locally known as kelampayan, belongs to the Rubiaceae, is becoming one of the most frequently planted trees in the tropics. *N.cadamba* wood is mainly used for pulp, producing low- and medium quality paper. The wood can be used for indoors' light construction since it is perishable in contact with ground. According to Joker[23], *N.cadamba* is also suitable for reforestation in watersheds and eroded areas and for windbreaks in agroforestry systems due to its fast growing property. It can be used as a shade tree for dipterocarp line planting while its leaves and bark can be applied in medicine. The dried bark can be used to relieve fever and as a tonic while an extract of the leaves can serve as a mouth gargle [47]. Despite *N. cadamba* is important in ecosystems and tropical forestry, but genetic information about the species is limited compared to other economically important tropical trees.In fact, this species has been selected as one of the important species for planted forest development in Malaysia, especially in Sarawak [17]. The main goal of this study was to develop simple sequence repeat (SSR) markers specific for genotyping *N. cadamba* trees with the intention of determining the genetic diversity and structure of *N. cadamba*. In addition, cross species transferability of the newly developed SSR markers was also tested in the present study.

MATERIALS AND METHODS

Plant materials and DNA extraction:
The fresh young leaves of *N. cadamba*seedlings were sampled from the nursery of Sarawak Forest Tree Seed Bank, Sarawak Forestry Corporation (SFC), Semengok, Sarawak. On the other hand, plant materials used for SSR marker validation were collected from the KelampayanProvenance Trial Plot established at the Landeh Nature Reserve, Semengok, Sarawak. The trial plot was established with seeds collected from the selected individuals naturally grown in three forest seed production areas (SPAs), namely Pasai Bon, Niah and Lawas in Sarawak. The sampling sites were B3L5, B3L6, B3L7 and a total of 30 samples were collected. Total genomic DNA was extracted from fresh leaves by using a modified CTAB DNA extraction protocol [10].

Cloning and sequencing of amplified fragments of inter-simple Sequence Repeats:
Three SSR primers, namely $(AC)_{10}$, $(AG)_{10}$ and $(GTG)_6$ were used to amplify the ISSR fragments flanked by two microsatellite sequences arranged in opposite orientations from the DNA of *N. cadamba*. The ISSR-PCR amplification was carried out in a Mastercycler Gradient PCR (eppendorf, Germany). PCR was performed in a volume of 25 µl containing 50 ng of genomic DNA, 10 pmol of the primer, 0.2 mM of each dNTP, 1x PCR buffer, 2.5 mM $MgCl_2$ and 0.5 U of *Taq* DNA polymerase (Invitrogen, USA). The PCR profile was 94°C for 2 min followed by 39 cycles of denaturation at 94°C for 30 sec, annealing at 57.6°C for $(AC)_{10}$, 62°C for $(AG)_{10}$ and 65°C $(GTG)_6$ for 30 sec and extension at 72°C for 1 min. A final extension at 72°C for 10 min was included. The amplified fragments with high concentration and size ranged from 500 bp to 3 kb later were selected for gel

purification by using QIAquick Gel Extraction Kit (QIAGEN, Germany). The purified fragments were then ligated into vectors by using pGEM®-T Easy Vector Systems (Promega, USA) according to the manufacturer instructions. Subsequently, the plasmids were transformed into *E.coli* JM109 competent cell and the transformed cultures were spread onto LB/amphicilin/X-Gal/IPTG agar plates, which act as selective media. Screening was performed on the transformed culture by blue-white detection of recombinant clones. In addition, colony PCR using M13 primer was also carried out to further verify the insert of the fragment into the vector from recombinant clones. Plasmid DNA was then extracted and purified from the selected positive clones by using Wizard® Plus SV Minipreps DNA Purification System (Promega, USA) and eventually sequenced by Applied Biosystems 3730XL DNA Analyzer using BigDye® Terminator v 3.1 CycleSequencing Kits (ABI, USA). It was sequenced bidirectionallyby using M13 forward (5'-GTAAAACGACGGCCAGT-3') and M13 reverse (5'-CAGGAAACAGCTATGAC-3') sequencing primers. Consequently, an initiating primer (IP1) was designed from the region flanking the SSR sequence. For nested PCR, another initiating primer (IP2) from the sequence between IP1 and the SSR sequence was also designed.

DNA libraries construction:
Adaptor-ligated, restricted DNA libraries were constructed according to Siebert et al. [40]. DNA was separately digested with *Eco*RV, *Ssp*I, *Alu*I, *Rsa*I and *Hae*III restriction enzymesat 37°C for 4 to 5 hours. The fragments were then ligated to a specific blunt adaptor (consisting of a 48-mer:5'-GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT-3' and an 8-mer with the 3-end capped by an amino residue: 5'-ACCAGCCC-NH_2-3'). As adaptor primers for nested PCR, AP1 (5'-CCATCCTAATACGACTCACTATAGGGC-3') and AP2 (5'-CTATAGGGCACGCGTGGT-3') also were prepared.

Determination of the sequence beyond the determined ISSR Sequences:
Genome walking method was used to determine the sequence beyond the determined ISSR sequence. Two steps of PCR amplification were carried out. The primary nested PCR reaction was conducted with each constructed DNA library by using the adaptor primer AP1 and ISSR-specific primer IP1. The reaction was conducted in a 25 µl reaction mixture containing 1.0 µl of the adaptor-ligated DNA, 0.2 mM of each dNTP, 5.0 pmol of adaptor primer AP1 and ISSR-specific primer IP1, 1x PCR buffer, 1.5 mM $MgCl_2$ and 0.5 U of *Taq* DNA polymerase (Invitrogen, USA) with a Mastercycler Gradient PCR (eppendorf, Germany). The PCR profile started with 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 1 min, annealing at 62°C for 30 sec and extension at 72°C for 2 min. Then, it was followed by another one cycle of denaturation at 94°C for 1 min, annealing at 62°C for 30 sec and extension at 72°C for 5 min. A final extension at 72°C for 10 min was included. The PCR products were then electrophoresed in a 1.5% agarose gel together with 100 bp DNA ladders (Promega, USA). The gel was stained with ethidium bromide and de-stained by using ddH_2O for at least 30 minutes. It was observed on a UV transilluminator and then the result was documented by using a Gel Documentation System (Perkin Elmer, USA). The secondary reaction was conducted with 1.0 µl of a 100-fold dilution of the primary PCR product using the adaptor primer AP2 and ISSR-specific primer IP2. The same reaction mixture and PCR conditions as in the primary PCR were used, except that the annealing temperature and cycle number were reduced to 60°C and 29 cycles. The PCR products were examined on a 1.5% agarose gel electrophoresis to check whether they formed a single band. Subsequently, PCR products which were forming a single band were purified, cloned and then sequenced by Applied Biosystems 3730XL DNA Analyzer using BigDye® Terminator v 3.1 Cycle Sequencing Kits (ABI, USA). Primer IP3 was designed for each locus from the newly identified sequence between the AP2 binding site and the SSR core region. Primer pairs IP1/IP3 or IP2/IP3 which flanking the SSR regions were used as SSR markers.

SSR search:
The sequences generated from ISSR fragments and nested PCR products were searched for the presence of SSRsby using SSR Finder software (http://geboc.org/ssr/ssr.html). The software was configured to locate a minimum of 12 bp SSRs. Motifs that were searched included perfect di-, tri-, tetra- and pentanucleotides repeats as well as compound repeats composed of di-, tri- and tetranucleotides repeats and imperfect repeats. In the perfect microsatellites, the minimum number of repeat units for dinucleotides is six, four repeat units for trinucleotides, three repeat units for both tetra- and pentanucleotides. In the compound and imperfect repeats, the minimum length of di-, tri- and tetranucleotides repeats is three units. Imperfect repeat is defined as having no more than one disruptive element of length less than 5 bp.

PCR validation with newly developed SSR primers:
To investigate characteristics of isolated SSR loci, template DNA was extracted from the fresh leaves of 30 *N. cadamba* trees. PCR was carried out in a 25 µl reaction mixture containing 50 ng of template DNA, 0.2 mM of each dNTP, 5.0 pmol of each designed primer pair, 1x PCR buffer, 2.0 mM of $MgCl_2$ and 0.5U of *Taq* DNA polymerase (Invitrogen, USA). The PCR profile started with 94°C for 2 min followed by 35 cycles of denaturation at 94°C for 1

min, annealing at 60°C for 30 sec and extension at 72°C for 1 min. A final extension at 72°C for 10 min was included. The PCR products were separated on 3.5% Metaphor agarose gels (FMC) andDNA size markers, which were 25 bp and 100 bp DNA ladders (Promega, USA) were ran simultaneously. The result was documented by using a Gel Documentation System (Perkin Elmer, USA).

Cross species amplification of newly developed SSR primers:
The SSR markers isolated from *N. cadamba* were checked for its success in cross species amplification with *Coffeacanephora (coffee, Rubiaceae), Duabangamoluccana* (*sawih*, Sonneratiaceae), *Shoreaparvifolia* (*merantisarangpunai*, Dipterocarpaceae) and *Canariumodontophyllum* (*dabai*, Burseraceae). The same reaction mixture and PCR conditions as describe above were used. Amplification products were scored as positive only if a sharp and reproducible band was observed.

SSR data analysis:
Standard genetic diversity parameters such as major allele frequency, number of allele, gene diversity and PIC value were calculated using Power Marker software[27]. In addition, deviation from Hardy-Weinberg equilibrium was analyzed using Chi-square test (*P*<0.05) using POPGENE Version 1.32 software [49].

RESULTS AND DISCUSSION

Cloning of amplified ISSR fragments and DNA library construction:
PCR reactions were carried out based on the optimum thermal cycling profiles to amplify ISSR fragments from *N. cadamba*. In total, 13 amplified ISSR fragments were successfully cloned and sequenced. Those ISSR sequences were used to design nested PCR primers (IP1 and IP2) for determining the unknown flanking regions (**Figure 1**). Primer IP1 was designed from the sequenced region flanking the SSR core region and for nested PCR, another primer IP2 was designed based on the sequence between IP1 and the SSR core region.

Figure 1: Nested PCR primers(IP1 and IP2) flanking the SSR core region from one of the amplified ISSR fragments by using $(AC)_{10}$ primer

Five adaptor-ligated, restricted DNA libraries were constructed according to Siebert et al. [40]. Genomic DNA was separately digested with restriction enzyme, *Eco*RV, *Ssp*I, *Alu*I, *Rsa*I and *Hae*III. Smeary band was observed from the digested DNA samples (**Figure 2**), indicating that the DNA samples were completely digested by each enzyme. All digested DNA were then purified and ligated to a blunt-end adaptor. It is a universal practice to use different restriction enzymes in the genome walking method to make different libraries in order to identify and isolate the desired upstream elements of variable size from some of the libraries. In most cases, the information on the distribution of restriction enzyme sites in a region of interest is unavailable, thus, combinations of several different enzymes must be tried in order to increase the probability of generating convenient DNA fragments for a successful genome walk using PCR [37]. Restriction enzymes such as *Bam*HI, *Hind*III that do not generate blunt ends were not suitable for the present study. In those cases different adaptor sequences have to be used to create ligated libraries since each restriction enzyme has its own sequence recognition pattern. This probably will cause the method more complicated and problematic. Furthermore, as explained by Rishi et al. [38], by using other restriction enzymes that

do not generate blunt ends, the strategy may not work in some cases where, the location of the restriction site is very far away beyond 10 kb upstream of the gene specific primer. Such conditions cannot be amplified under the experimental conditions adopted.

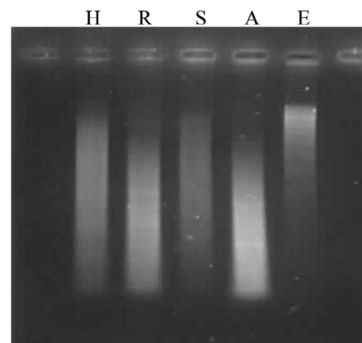

Figure 2: Digested DNA samples obtained on 0.8% agarose gel. H: *Hae*III; R: *Rsa*I; S: *Ssp*I; A: *Alu*I and E: *Eco*RV

Determination of the sequence beyond the determined ISSR sequence:
Genome walking method is a systematic identification of unknown regions flanking a known DNA sequence. The traditional approach to isolate such unknown flanking regions is to screen the genomic library by using a specific probe, however, this strategy is laborious and time-consuming [40]. Some PCR-based techniques such as inverse PCR, vectorette PCR and adaptor-specific PCR have been developed and have become popular since the method is efficient, fast and without the need to construct and screen libraries. For all PCR-based techniques, genomic DNA is digested with restriction enzyme and then ligated (to self, or to vectorette, or to adaptor, respectively). The ligated product later is used as a template for amplifying flanking regions using PCR. In present study, adaptor-specific PCR was carried out as genome walking method by using locus- and adaptor-specific nested primers to amplify flanking regions. The technique is more accurate due to the use of two-round PCR, wherein nested locus-specific primer is used in the second PCR, thereby increasing the probability of amplification of the desired flanking DNA sequence [38].

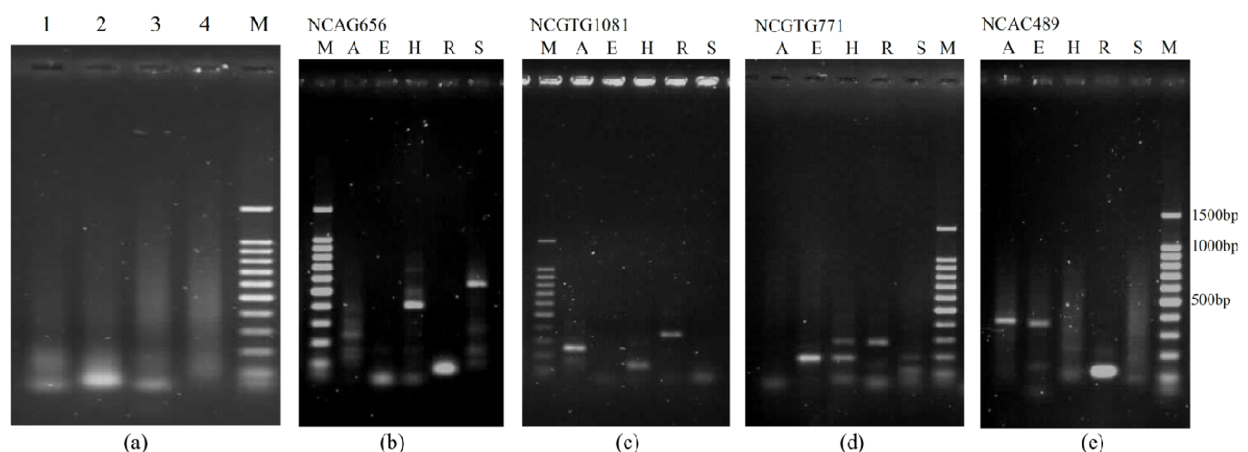

Figure 3: PCR products obtained using genome walking method. (a). Primary PCR products amplified with the adaptor primer APl and ISSR-specific primer IPl. (b) – (e). Second nested PCR products amplified with different ISSR-specific primers from different restricted DNA libraries. Lane A: *AluI* restricted library; Lane E: *EcoRV* restricted library; Lane H: *HaeIII* restricted library; Lane R: *RsaI* restricted library; Lane S: *SspI* restricted library; Lane M: 100 bp ladder of DNA size markers

Nested PCR amplifications from different restricted DNA libraries were performed by using the primers prepared based on sequences of each ISSR fragment (IPl and IP2) and the longer strand of the adaptor (APl and AP2). The primary amplification using the IPl and APl primer pair usually yielded smeared banding patterns throughout agarose gel electrophoresis. However, the second amplification using the IP2 and AP2 primer pair produced more specific single band from DNA libraries as shown in **Figure 3**. Nested PCRs for 13 ISSR-specific primers produced a total of 22 single banded profiles from most of the DNA libraries. However, some amplification of SSR sequence produced either no clear band or small fragments from some restricted DNA libraries. No amplification in certain libraries may be due to an absence of a restriction site close to the ISSR-specific primer while the amplification of the small fragments could be due to the presence of a restriction site in the genome close to the designed primer

[38]. Single-banded fragments were later cloned and sequenced. A primer from the newly defined flanking sequence (IP3) was designed for amplification of the region containing a SSR in combination with IP2.

SSR search and primer design:
The nucleotide sequences were initially searched for the presence of SSR motifsby using SSR Finder software (http://geboc.org/ssr/ssr.html) (**Figure 4**). A total of 39 SSR regions had been identified from the 22 nucleotide sequences, however, only 31 (79.5%) of the detected SSR regions were found useful for primer design.Based on Weber's criterion [45]of repeated structure, among the 31 SSR loci, seven (22.6%) SSRs were classified into the perfect type, eleven (35.5%) belonged to the imperfect type and thirteen (41.9%) belonged to the compound type. The most abundant SSR class recovered was simple perfect dinucleotide repeats with five occurrences and followed by compound perfect trinucleotide repeats with four occurrences. These represented 16.1% and 12.9% of all SSRs recovered in the present study. In imperfect repeats, the interrupting sequence varied from 1 to 5 nucleotides in length. The majority of imperfect repeats were represented by compound trinucleotides repeats with a short disruptive element. The number of repeats varied between 5 and 21 among all identified simple SSRs. The most abundant repeat number found for simple SSRs was five with six occurrences. On the other hand, the calculation of repeat length for compound SSR sequences was based on the longest uninterrupted repeat. This was determined by evidence that length variations preferentially affect the longest repeat in compound and interrupted microsatellite sequences [28]. In the present study, the number of repeats varied between 2 and 8 among all identified compound microsatellites. The most abundant repeat number found for compound SSRs was three which covered 50% of all detected compound SSRs.

The most abundant dinucleotide motif found in *N. cadamba*was TG/CT. However, most of the surveys in plantshave shown that the most frequently occurring dinucleotide repeats are $(AT)_n$, with $(AG)_n$and $(AC)_n$as second andthird most frequent [6, 11, 25, 32, 44]. Surveys of SSR sequences in forest trees have generally foundthe $(AG)_n$motif to be more abundant than $(AC)_n$, for example, *Quercusmacrocarpa*(83% $(AG)_n$) [9] and *Melaleucaalternifolia*(43% $(AG)_n$) [39]. The factors responsible for differences in the abundanceof SSR motifs among plant genera arenot well understood;however, for hybridisation surveys they may simply reflectdifferences in cloning, enrichment or screening procedures [2].There is evidence that the repeat type and numberis influenced by the restriction enzyme used to size-fractionatethe genome when constructing the library [15]. In addition,Condit and Hubbell [6] found that smaller inserts gave higher estimates for $(AC)_n$ abundance and suggested that poly-AC sites are not randomly or uniformly spread through the genome of tropical trees but are highly clustered in certain regions.

Table 1: Primer sequences, complexity, type, size (bp) and genetic diversity parameters of each identified SSR repeat motif

ISSR	Locus	Repeat Motif	Complexity	Type	Primer sequence (5'-3')	EPS (bp)	MAF	A	h	PIC
NCAC489	NCAC02	$(TG)_4T(TG)_4$	Simple	Imperfect	F: GTGCAATGGCAGAGAGGAAG R:TCACAGAGGATCGTTCACAGAG	133	0.70	3	0.464	0.419
	NCAC03	$(TG)_5$	Simple	Perfect	F: TGCCTTGAAGTGCCTGTACTC R: CACATAAGGAGGGGTGATCG	147	0.45	4	0.659	0.597
NCAC748	NCAC05	$(TGG)_3TTA(TGG)_2$	Simple	Imperfect	F: TGGTGGTGCGTGATAGATTC R: CTCATGACGGCCAATCAAC	205	0.63	3	0.527	0.467
	NCAC09	$(GTG)_3GTAAT(GATTG)_2$	Compound	Imperfect	F: TTGGCCTTTTGGTAGTGACG R: GCACCACCAGATCCATAACC	204	0.57	3	0.571	0.499
NCAC839	NCAC10	$(CT)_{21}$	Simple	Perfect	F: CGATCTCGCTCTCCCTCT R: TCGATGGAGAAACCAAGC	182	0.37	4	0.668	0.598
	NCAC11	$(CT)_8(CCCT)_3$	Compound	Perfect	F: GTTGGCTTGGTTTCTCCATC R: ACATTACCCACAACCCCATC	187	0.20	9	0.864	0.849
	NCAC12*	$(TGG)_3TTGTC(GTT)_3$	Compound	Imperfect	F: GCTTGGTTAGGTGGTGATGG R: CAACCCTCTCCCAAACTGTC	162	0.32	5	0.762	0.722
	NCAC13	$(TGG)_2(TGT)_2(GGT)_2$	Compound	Perfect	F: TGTTGGAGGAGACAGTTTGG R: CACAACCTCCATCTCTTCTACC	122	0.47	4	0.633	0.565
	NCAC14	$(TG)_4AA(TG)_4$	Simple	Imperfect	F: GTTGTCAGAGAATGAGACAGG R: ACTCTACACGCACAATACACG	166	0.43	5	0.701	0.653
	NCAC15	$(GT)_3(GA)_2(GT)_3$	Compound	Perfect	F: TGCGTGTATTGTGCGTGTAG R: CTAACACTTAGGGCCGCTTG	297	0.50	5	0.673	0.613
NCAG849	NCAG01	$(CT)_8$	Simple	Perfect	F: CATTCCGCGTACGTGTTACC R: TTGAGAGATCCCGACGACAC	190	0.38	6	0.749	0.712
	NCAG02	$(ATC)_2(TTTA)_3$	Compound	Perfect	F: TTTACCGACCAGAGAACGTG R: ACCCATGGATTCTGATCCTC	248	0.35	4	0.709	0.655
	NCAG03*	$(GT)_{10}$	Simple	Perfect	F: TGAGTGTGTGTGTGTGTGTGTG R: CTTGGGCTCATGTCGATTG	239	0.67	4	0.496	0.442
NCAG724	NCAG05	$(CAC)_5$	Simple	Perfect	F: CCACCACGTGACTAGCATTG R: CATCACTTCCCCTTCATTCC	177	0.67	3	0.480	0.412
NCGTG749	NCGTG11	$(GGC)_2C(GTG)_3$	Compound	Imperfect	F: CAGGGAGGAGATCGTGGA R: GCAACGCCAGTCACCTTC	120	0.45	4	0.669	0.612
						Mean	0.48	4	0.642	0.598

*F = Forward; R = Reverse; EPS (bp) = Expected product size (bp); MAP = Major allele frequency; A = Number of alleles; h = Gene diversity; PIC = polymorphism information content;*not significant departure from HWE ($P < 0.05$)*

SSR marker validation:
All designed primer pairs were tested for their functionality where primer pairs that failed to amplify or amplified poorly were eliminated. Based on the results, 15 out of 31 primer pairs were able to produce an amplified fragment of the expected size and later selected for SSR marker validation test. Total genomic DNA was extracted from adult leaf tissue following the protocol described earlier. The selected primer pairs were standardized for PCR amplification and then validated using the extracted DNA samples (**Table 1**). As described by Kirst et al. [24], selected SSR loci should have the following criteria: (1) ease of interpretation of the amplified alleles to allow widespread use by several labs that do not necessarily have access to automated sequencers; (2) minimum stutter of PCR products; (3) high polymorphism based on a preliminary characterization and (4) absence of null alleles.

```
1     ACACACACACACACACACACACACACACTCTCTCTCTCTCTCTCTCTCTC    50
51    TCTCTCTCTCTCCCTCCCCAAATCGAACACCCTCTATTCTCTATCGATCT    100
101   CGCTCTCCCTCTCTTTCTCCTTCCCTCTGTTTTCTATCAATCTCTCTCTC    150
                          (CT)21
151   TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTCTTCCTCCCTTTGTT    200
201   CTCTCTCTCTAACTAGTGGCTGTGGGCTGGGGAGTTGTTGGGCAGTTGGT    250
251   GATTGTTGGCTTGGTTTCTCCATCGATCTCTCTTTCCCTTCCTCCCTCCC    300
                                        (CT)8(CCCT)3
301   TCTGTTCTCTATCAATCTCTCTCTCTCTCTCTCCCTCCCTCCCTCTCTAA    350
351   CTGGTGGCTGTGGGCTGGAGAGTTGTTGGGCAGTTGGTGATTGTTGGCTT    400
401   GGTTAGGTGGTGATGGCCGTGGATGGGGTTGTGGGTAATGTTGGAGATGG    450
                 (TGG)3TTGTC(GTT)3
451   ATTTGGTGGTGGTTGTCGTTGTTGTTGCTGGTTGTGTGATTATAGTGGTG    500
501   GTTGAATGGTGGTGTGGTGGCAGCTGGTTGTTGGAGGAGACAGTTTGGGA    550
                                             (TGG)2(TGT)2(GGT)2
551   GAGGGTTGTGTGGTTATAATGGTGGTTGTATGGTGGTGTTGTGGTGGTTG    600
601   GTTGTCAGAGAATGAGACAGGTAGTATTGGTAGAAGAGATGGAGGTTGTG    650
                (TG)4AA(TG)4
651   GACTGTGTGTGAATGTGTGTGACTGCGTGTGCTTTGTGTGAGTGTTACTG    700
701   TGTGTGCTGTGTGTAGATGTGTGTGTTACAGTGTGTGTAAGTGCGTGTAT    750
                                            (GT)3(GA)2(GT)3
751   TGTGCGTGTAGAGTGTGAACTGTGTAACAGTGTGTGAGAGTGTGTGAATT    800
801   TACTGTGTGTATTAATATATGTTGTGTGTGAAAATTATGCTAGTGTGTAT    850
851   TCTTGGTGTTCTTGGTGTTTTATGGACAAAAATGAAATTTAATTAAATTT    900
901   CATTCTAATACACCCATCTAACCTATATGTGAACCAAGCAAAGCTATTGG    950
951   AATGAAAAAAATTCATTTCAATTTAGGGTGCATGATTTTATTTCTAATTC    1000
1001  CGATTCTAATTCACGAACCAAGCGGCCCTAAGTGTTAGTGATAAATATGT    1050
1051  CCCTTATGACATAAAACAAGAATCGAATAAAGGAAATTAGAAAGGTAACT    1100
1101  ATTAAAAAAAT                                           1150
```

Figure 4: SSR search by SSR Finder software. A primer from the newly defined flanking sequence (IP3) is designed for amplification of the region containing a SSR in combination with IP2

Some of the DNA samples did not amplify the expected fragment, indicating a possible sequence divergence in the complementary nucleotide sequence of one of the two primers [3] and the presence of null allele in the sample [30]. On the other hand, analysis of the PCR-amplified SSRs typically shows several additional signals besides the expected PCR product. They are either known as shadow bands, stutter bands or spurious products. Their occurrence may complicate genotyping and in extreme cases can lead to ambiguous results.In the present study, stutter bands were detected from some of the SSR amplifications. According to Fazekas et al. [13], among the factors shown to influence the proportion of stutter bands to the main product are the repeat number, length of the repeat motifs, PCR cycle number and the DNA template concentration. The contribution from stutter bands is shown to increase with

the repeat number while shorter repeated units usually gave rise to stronger stutter products. Besides, length of flanking sequences correlate positively with the shadow band contribution and the possibility of stutter bands generation is increased with the use of low DNA template concentration [33].

SSR data analysis:
As mentioned by Choroszy et al. [4], number and frequency of alleles identified at individual loci, the effective number of alleles per locus, gene diversity and the PIC value are among the basic indicators most often used for analyzing the polymorphism of SSR DNA sequences in different populations (**Table 1**). Considerable allelic amplifications were obtained for 15 SSR markers across the tested genotypes. In general, the newly developedSSR markers revealed medium to high allelic diversity. A total of 66 alleles were identified at 15 loci with an average of 4 per locus. Most of the SSRlocishowed high polymorphism as indicated by the number of alleles detected at particular lociand by the PIC value, which was above 0.5. The most polymorphic among the 15 lociwere: NCAC11 (9 alleles with sizes ranging from 150 to 250 bp, PIC=0.849), NCAC12 (5 alleles ranging from 160 to 185 bp, PIC=0.722) and NCAG01 (6 alleles ranging from 200 to 250 bp, PIC=0.712). Overall, the number of alleles at a SSR locus is directly correlated with the number of repeats present, with longer repeats being more polymorphic than shorter ones [20, 45]. However, this is not a universal observation[30]. According to Jarvis et al. [21], most polymorphic markers had repeat lengths greater than 20 bp. In the present study, locus NCAC11 became the most polymorphic marker with the repeat length of 28 bp. This confirmed the conclusions of Mason et al. [31] who suggests that the future development of SSR markers should focus on the identification of markers with repeat length of >20 bp in order to maximize polymorphism.

Deviation from Hardy-Weinberg equilibrium (HWE) was analyzed using Chi-square test ($P<0.05$). Among 15 loci, two of them, namely NCAC12 and NCAG03 were followed the HWE. The deviation from HWE may cause by the nature of the tested genotypes which individual samples that were originated from the same mother trees tend to have low genetic diversity[18, 43]. In addition, deviation from HWE may also due to the presence of null alleles or an excess of homozygotes.Null alleles are thought to arise as a consequence of divergence in the SSR flanking sequences since deletions or point mutations in the primer-annealing site would be expected to inhibit or prevent primer binding and hence amplification[41]. Although null alleles are rare among human SSRs, they are relatively common in barley and in various polyploidy species, particularly in wheat [7, 8].

Cross species amplification of SSR markers:
SSRs have become one of the most widely used molecular markers for genetic studies in recent years. As compared to other DNA marker systems, SSRs have several advantages such as locus specificity, co-dominant nature, high reproducibility and substantial size polymorphism. However, *de novo* isolation of SSRs is required for the species that need to be studied for the first time[34]. Such isolation requires extraction, cloning, sequencing and characterization of SSR loci, thus making the procedures highly time consuming and expensive. To minimize the cost, cross species amplification of SSRs has beenapplied for species with very little or no information on the DNA sequence, through the screening of primers from different sources. Potential transferability of SSR primers across species of the same genus [2, 5] and across genera of the same family was reported in Leguminaceae[36], Myrtaceae[50] and Fagaceae[1]. Such transferable markers later can be utilized for the marker assisted breeding programs for the improvement of the cross amplified taxa [34, 51].

In the present study, transferability of SSR primers across genera of the same family, *Coffeacanephora (coffee)* and across genera of the different family, *Duabangamoluccana* (sawih), *Shoreaparvifolia* (merantisarangpunai) and *Canariumodontophyllum* (dabai) was checked for all the 15 microsatellites isolated from *N. cadamba*. Successful marker amplification was determined by comparing the expected fragment size and the percentage of marker amplified as surrogates of transfer success. In total, 12 markers (80%) showed cross species amplification andtwo markers, namely NCAC02 and NCAC03 showed cross species amplification in all species studied while NCAC05, NCAG01 and NCAG03 did not show any cross species amplification (**Table 2**). Among the species studied, dabai gave cross species amplification with highest number of markers (11 markers or 73.3%) followed by coffee, merantisarangpunai and sawih showed cross amplification with least number of markers (4 markers or 26.7%). The amplified bands were distinct with either high or low intensity and no stuttering was observed.

Transferability of *N. cadamba* SSR markers across the genera of same and different families was confirmed in this study. This indicates that the sequences flanking the SSR regions in *N. cadamba* are highly conserved across genera. Although the transferability of SSRs indicate low success rate among plant species, some sequences are associated with highly conserved regions [36, 51]. Aldrich et al.[1] showed the successful amplification of *Quercusrubra*SSR loci in *Castaneamollissima*, a distantly related species of Fagaceae. Similarly, eucalyptus SSRs amplified across its unrelated members belonging to the family Casuarinaceae and furthermore lead to the development of SSR loci in *Casuarina*species [48]. Hence, the newly developed SSR markers show good cross species amplification efficiency

and can be useful for the studies of intra-species diversity, population genetic structure and their molecular characterization.

Table 2: Potential cross-species amplification of *N. cadamba* SSR loci in *Coffeacanephora* (coffee), *Duabangamoluccana* (sawih), *Shoreaparvifolia*(merantisarangpunai) and *Canariumodontophyllum* (dabai)

Locus	*C. canephora*	*D. moluccana*	*C. odontophyllum*	*S. parvifolia*
NCAC02	+	+	+	+
NCAC03	+	+	+	+
NCAC05	x	x	x	x
NCAC09	+	x	+	+
NCAC10	x	x	+	x
NCAC11	+	+	+	x
NCAC12	+	+	+	x
NCAC13	+	x	+	+
NCAC14	+	x	+	+
NCAC15	+	x	+	+
NCAG01	x	x	x	x
NCAG02	x	x	x	+
NCAG03	x	x	x	x
NCAG05	+	x	+	+
NCGTG11	+	x	+	+
Percentage (%)	66.7	26.7	73.3	60.0

"+" indicates expected amplifiable products and "x" indicates no amplification

CONCLUSION

This study has demonstrated that ISSR suppression-PCR method is a relatively cost-effective, simple without enrichment and screening procedures for developing SSR markers in *N. cadamba*. A total of 15 SSR markers were successfully developed and characterized by using ISSR suppression-PCR method. To the best of our knowledge, this is the first report on the development of SSR markers in *N. cadamba* and we hope these newly developed SSR markers could pave the way for exploiting genotype data for comparative genetic and genomic analyses, individual genotyping and studies of gene flow, mating system, genetic diversity and structure, cultivar identification, marker-assisted selection and molecular breeding of *N. cadamba* and other indigenous tropical tree species in future.

Acknowledgements
The authors would like to thank all the lab assistants and foresters involved in this research project for their field assistance in sample collection. This work is part of the joint Industry-University Partnership Programme, a research programme funded by the Sarawak Forestry Corporation (SFC), Sarawak Timber Association (STA) and Universiti Malaysia Sarawak under grant no. 02(DPI09)832/2012(1), RACE/a(2)/884/2012(02) and GL(F07)/06/2013/STA-UNIMAS(06).

REFERENCES

[1] P.R. Aldrich, M. Jagtap, C.H. Michler and J. Romero-severson, *Silvae Genetics*, **2003**, 52: 176–179.
[2] P.A. Butcher, S. Decroocp, Y. Gray and G.F. Moran, *Theoretical and Applied Genetics*, **2000**, 101: 1282-1290.
[3] F. Carriero, G. Fontanazza, F. Cellini and G. Giorio, *Theoretical and Applied Genetics*, **2002**, 104: 301-307.
[4] B. Choroszy, A. Janik, Z. Choroszy and T. Zabek, *Animal Science Papers and reports Supplement 2*, **2006**, 24: 71-77.
[5] G. Cipriani, G. Lot, W.G. Huang, M.T. Marrazzo, E. Peterlunger and R. Testolin, *Theoretical and Applied Genetics*, **1999**, 99: 65-72.
[6] R. Condit and S.P. Hubbell, *Genome*, **1992**, 34: 66-71.
[7] K.M. Devos, G.J. Bryan, A.J. Collins, P. Stephenson and M.D. Gale, *Theoretical and Applied Genetics*, **1995**, 90: 247-252.
[8] P. Donini, P. Stephenson, G.R. Bryan and R.M.D. Koebner, *Genetic Resources and Crop Evolution*, **1998**, 45: 415-421.
[9] B.D. Dow, M.V. Ashley and H.F. Howe, *Theoretical and Applied Genetics*, **1995**, 91: 137-141.
[10] J. Doyle and L. Doyle, *Focus*, **1990**, 12: 13-15.
[11] C.S. Echt and P. May-Marquardt, *Genome*, **1997**, 40: 9-17.
[12] S.J. Elizabeth, J.H. Leonie, C.D. Michelle, T.A. Michael, P.T. Terry, C.B. Michaelson-Yeates and W.F. John, *Plant Science*, **2003**, 165: 531-539.
[13] A.J. Fazekas, R. Steeves and S.G. Newmaster, *BioTechniques*, **2010**, 48(4): 277–285.
[14] R.Wickneswari, W.S. Ho, K.S. Lee and C.T. Lee,*Forest Genetics*, **2003**, 11(3-4): 193-201.
[15] M.B. Hamilton and R.C. Fleischer, *Journal of Heredity*, **1999**, 90(5): 561-563.
[16] W.S. Ho, R. Wickneswari, M.C. Mahani and M.N. Shukor,*Forest Ecology and Management*, **2004**, 203: 209-228.

[17] W.S. Ho, S.L. Pang and A. Julaihi, A. *Physiology and Molecular Biology of Plants*, **2014**, 20(3): 393-397.
[18] B.L. Tchin, W.S. Ho, S.L. Pang and J. Ismail, *Biotechnology*, **2012**, 11(6): 307-317.
[19] W.S. Ho, R. Wickneswari, M.C. Mahani and M.N. Shukor, *Journal of Tropical Forest Science*, **2006**, 18(1): 557-565.
[20] H. Innan, R. Terauchi and T.N. Miyashita, *Genetics*, **1997**, 146: 1441-1452.
[21] D.E. Jarvis, O.R. Kopp, E.N. Jellen, M.A. Mallory, J. Pattee, A. Bonifacio, C.E. Coleman, M. Stevens, D.J. Fairbanks and P.J. Maughan, *Journal of Genetics*, **2008**, 87: 39-51.
[22] E. Jewell, A. Robinson, D. Savage, T. Erwin, C.G. Love, G.A.C. Lim, X. Li, J. Batlet, G. Spangenberg and D. Edwards, *Nucleic Acids Research*, **2006**, 34: 656-659.
[23] D. Joker, SEED LEAFLET*Neolamarckia cadamba*(Roxb.) Bosser(*Anthocephaluschinensis*(Lam.) A. Rich. ex Walp.) (http://curis.ku.dk/portal-life/files /20648324/ neolamarckia_cadamba_int.pdf), **2000**.
[24] M. Kirst, C.M. Cordeiro, G.D.S.P. Rezende and D. Grattapaglia, *Journal of Heredity*, **2005**, 96(2): 161-166.
[25] U. Lagercrantz, H. Ellegren and L. Andersson, *Nucleic Acids Research*, **1993**, 21: 1111-1115.
[26] C.L. Lian, Z.H. Zhou and T. Hogetsu, *Journal of Plant Research*, **2001**, 114: 381-385.
[27] K. Liu and S.V. Muse, *Bioinformatics*, **2005**, 21(9): 2128-2129.
[28] K. Loridon, B. Cournoyer, C. Goubely, A. Depeiges and G. Picard, *Theoretical and Applied Genetics*, **1998**, 97: 591-604.
[29] K.S. Liew, W.S. Ho, S.L. Pang and A.Julaihi, A. *Physiology and Molecular Biology of Plants*, **2014**, *in press*
[30] P. Masi, P.L. Spagnoletti and P. Donini, *Molecular Breeding*, **2003**, 11: 303-313.
[31] S.L. Mason, M.R. Steven, E.N. Jellen, A. Bonifacio, D.J. Fairbanks and C.E. Coleman, *Crop Science*, **2005**, 45: 1618-1630.
[32] M. Morgante and A.M. Olivieri, *The Plant Journal*, **1993**, 3(1): 175-182.
[33] M. Olejniczak and W.J. Krzyzosiak, *Electrophoresis*, **2006**, 27: 3724-3734.
[34] D.V.N. Pamidimarri, G.S. Sudheer, H.R. Mastan, P. Ravi, S. Sweta and P.R. Muppala, *Molecular Biology Reporter*, **2010**, 38(2): 1383-1388.
[35] Y. Park, J.K. Lee and N. Kim, *Molecules*, **2009**, 14: 4546-4569.
[36] R. Peakall, S. Gilmore, W. Keys, M. Morgante and A. Rafalski, *Molecular Biology and Evolution*, **1998**, 15: 1275-1287.
[37] P.S. Reddy, S. Mahanty, T. Kaul, S. Nair, S.K. Sopory and M.K. Reddy, *Analytical Biochemistry*, **2008**, 381: 248–253.
[38] A.S. Rishi, N.D. Nelson and A. Goyal, *Journal of Biotechnology*, **2004**, 111: 9-15.
[39] M. Rossetto, R. Slade, P. Baverstock, R. Henry and L. Lee, *Molecular Ecology*, **1999**, 8(4): 633-643.
[40] P.D. Siebert, A. Chenchik, D.E. Kellogg, K.A. Lukyanov and S.A. Lukyanov, *Nucleic Acids Research*, **1995**, 23(6): 1087-1088.
[41] M.J.M. Smulders, G. Bredemeijer, W. Rus-Kortekaas, P. Arens and B. Vosman, *Theoretical and Applied Genetics*, **1997**, 97: 264-272.
[42] K. Tamura, M. Nishioka, M. Hayashi, Z.C. Zhang, C.L. Lian, T. Hougetsu and K. Harada, *Breeding Science*, 2005, 55: 247-252.
[43] S.Y. Tiong, S.F. Chew, W.S. Ho and S.L. Pang, *Advances in Applied Science Research*, **2014**, 5(3): 458-463.
[44] Z. Wang, J.L. Weber, G. Zhong and S.D. Tanksley, *Theoretical and Applied Genetics*, **1994**, 88: 1-6.
[45] J.L. Weber,*Genomics*, **1990**, 7: 524-530.
[46] R. Wickneswari and W.S. Ho, *TROPICS*, **2003**, 13(2): 139-149.
[47] World Agroforestry Centre, A tree species reference and selection guide: *Anthocephaluscadamba*. (http://www.worldagroforestrycentre.org/sea/Products/ AFDbases/af/asp/SpeciesInfo.asp?SpID=17933), **2004**.
[48] R. Yasodha, M. Ghosh, R. Sumathi and K. Gurumurthi, *ActaBotanicaCroatica*, **2005**, 64(1): 115-120.
[49] F.C. Yeh, R.C. Yang and T. Boyle, POPGENE Version 1.32: Microsoft Windows–based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada, **1999**.
[50] M.I. Zucchi, R.P.V. Brondani, J.B. Pinheiro, C. Brondani and R. Vencovsky, *Molecular Ecology Notes*, **2002**, 2:512.
[51] P.S. Lai, W.S. Ho and S.L. Pang, *Biotechnology*,**2013**, 12(6): 225-235.