HIGHLY VARIABLE STR MARKERS OF *NEOBALANOCARPUS HEIMII* (DIPTEROCARPACEAE) FOR FORENSIC DNA PROFILING

LH Tnah[1], SL Lee[1,] *, KKS Ng[1], QZ Faridah[2] & I Faridah-Hanum[2]

[1]*Forest Research Institute Malaysia, 52109 Kepong, Selangor Darul Ehsan, Malaysia*
[2]*Universiti Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia*

Received November 2009

TNAH LH, LEE SL, NG KKS, FARIDAH QZ & FARIDAH-HANUM I. 2010. Highly variable STR markers of *Neobalanocarpus heimii* (Dipterocarpaceae) for forensic DNA profiling. *Neobalanocarpus heimii*, locally known as chengal, is an important timber species in Peninsular Malaysia. Owing to the high demand for its valuable timber, *N. heimii* is subjected to illegal logging and this species may become endangered in the near future. The present study was designed to identify a set of highly polymorphic short tandem repeat (STR) markers for timber tracking of *N. heimii*. An extensive evaluation of 51 STRs developed for Dipterocarpaceae managed to identify 12 STR loci (*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018*, *Hbi161*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04* and *Shc07*), which showed specific amplification, high polymorphism, single-locus mode of inheritance, absence of null alleles and absence of mononucleotide repeat motifs in *N. heimii*. These loci can be readily used to establish a linkage between the evidentiary sample and the source, thus providing a useful set of markers for individual identification in *N. heimii*.

Keywords: Chengal, forensic science, individual identification, tropical tree species, illegal logging

TNAH LH, LEE SL, NG KKS, FARIDAH QZ & FARIDAH-HANUM I. 2010. Penanda jujukan ulangan pendek (STR) variasi tinggi bagi *Neobalanocarpus heimii* (Dipterocarpaceae) untuk pemprofilan DNA forensik. *Neobalanocarpus heimii* atau nama tempatannya chengal merupakan spesies balak yang penting di Semenanjung Malaysia. Disebabkan permintaan besar terhadap kayu balaknya yang bernilai tinggi, spesies ini terdedah kepada aktiviti pembalakan haram dan dijangka akan terancam dalam masa terdekat. Kajian ini bertujuan untuk mengenal pasti penanda jujukan ulangan pendek (STR) yang berguna dalam sistem pengesanan kayu balak chengal. Penilaian terperinci ke atas 51 STR yang dijana untuk Dipterocarpaceae mengenal pasti 12 lokus yang sesuai diguna pakai untuk *N. heimii* (*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018*, *Hbi161*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04* dan *Shc07*). Kesemua lokus ini menunjukkan ciri-ciri berikut: amplifikasi yang spesifik, ketiadaan alel nol, pola pewarisan mengikut Hukum Mendel dan ketiadaan jujukan ulangan mononukleotida. Lokus-lokus ini sedia digunakan untuk pengenalpastian individu chengal dengan mewujudkan kesinambungan antara sampel bukti dengan sumber.

INTRODUCTION

While one usually thinks of human DNA when considering the admissibility of DNA evidence in court cases, the analysis of DNA from non-human has been evaluated for forensic botany. Although the application involving plant materials has been fairly limited, there are primarily three areas being investigated. The first is linking of botanical materials to suspects or victims in order to make an association with a particular area where a crime is committed (Mestel 1993, Yoon 1993, Siver *et al.* 1994, Korpelainen & Virtanen 2003, Craft *et al.* 2007). The second is linking *Cannabis sativa* specimens to aid in forensic drug investigations (Coyle *et al.* 2001, Coyle *et al.* 2003, Gilmore *et al.* 2003, Zinnamon & Keim 2003, Howard *et al.* 2009) and the third is linking illegally harvested tree stumps to suspected stolen wood (White *et al.* 2000). These applications involve the use of DNA markers such as randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and short tandem repeat (STR), which aid the forensic community identify botanical samples and determine the provenance of seized samples.

STRs are short (1–6 bp in length) tandemly repeated DNA sequences. They comprise simple mononucleotide to hexanucleotide repeats, varying from a few tens of bases up to

**Author for correspondence. E-mail: leesl@frim.gov.my*

typically one hundred and are spread randomly throughout the genomes of animals and plants. The number of tandemly repeated units has been shown to be highly polymorphic between individuals and this is thought to be due to slippage of the DNA polymerase during the synthesis and mismatched repair (Levinson & Gutman 1987). As a whole, the popularity of STR markers stems from a unique combination of several important advantages, namely, their codominant inheritance, high genomic abundance, high degree of polymorphism, high reproducibility, enormous extent of allelic diversity and the ease of assessing size variation by PCR with pairs of flanking primers (Weising *et al.* 2005), which make them suitable for DNA profiling.

Neobalanocarpus heimii, or locally known as chengal, is endemic and widely distributed in Peninsular Malaysia. It produces a naturally, highly durable wood and is among the strongest timbers in the world. Under the International Union for Conservation of Nature and Natural Resources (IUCN) Red List of Threatened Species version 3.1 criteria, it was assigned under the vulnerable category due to a decline in the area of its distribution, the extent of occurrence and/or quality of habitat, and actual or potential levels of exploitation (Chua 1998). Owing to the high demand for its valuable timber, *N. heimii* is subjected to illegal logging and this species may become endangered in the near future. With the aim of establishing DNA profiling database to identify the illegal source of *N. heimii*, the present study was designed to identify a set of highly polymorphic STR markers for *N. heimii*, and subsequently to evaluate the quality of these markers for specific amplification, single-locus mode of inheritance, absence of null alleles and easy interpretation of genotype.

MATERIALS AND METHODS

Sample collection and DNA extraction

Forty-eight unrelated individuals of *N. heimii* collected from Pasoh Forest Reserve (FR), Negeri Sembilan, Malaysia, together with three half-sib families (each family comprising 48 seeds) collected from Lenggor and Labis FRs, Johore were used for STR loci screening. The samples were collected in the form of inner bark or leaf tissues. Genomic DNA was extracted using the procedure described by Murray and Thompson (1980) with modifications. Approximately 5 g of leaf or inner bark tissues were mixed with 20 ml of prewarmed (60 °C) CTAB (2%) extraction buffer (20 mM EDTA; 100 mM Tris-HCl, pH 8.0; 1.4 M NaCl; 1% PVP-40; 1.5% β-mercapthethanol) in a 50 ml tube. The homogenate was incubated at 60 °C for 30 min. An equal volume of chloroform–isoamyl alcohol (24:1) was added and mixed gently for about 15 min, followed by centrifugation at 3000 rpm for 10 min. The extraction step was repeated with equal volume of chloroform–isoamyl alcohol (24:1). The supernatant was then transferred to a new 50 ml tube and the DNA was precipitated from the aqueous phase with 0.6 volume of cold isopropanol. The mixture was pelleted by centrifuging at 3000 rpm for 10 min. Subsequently, the supernatant was removed and the DNA pellet was transferred into a 1.5 ml tube containing 1 ml of wash buffer (76% ethanol, 10 mM NH_4OAc) and the tube was then centrifuged at 13 000 rpm for 10 min. The supernatant was removed and the DNA pellet was dried using a DNA Mini Speed Vacuum (Heto, Denmark) for 15 min and redissolved in TE buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA). The extracted DNA was further purified using High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH).

PCR amplifications and electrophoresis condition

PCR amplifications were performed in 10 µl reaction mixture, consisting of approximately 5 ng of template DNA, 50 mM KCl, 20 mM Tris-HCl (pH 8.0), 1.5 mM $MgCl_2$, 0.4 µM of each primer, 0.2 mM of each dNTP, and 0.5 U of *Taq* DNA polymerase (Promega). The reaction mixture was subjected to amplification using a GeneAmp PCR System 9700 (Applied Biosystems) for an initial denaturing step of 3 min at 94 °C, 40 cycles of 94 °C for 1 min, 45–52 °C annealing temperature for 30 s and 72 °C for 30 s, followed by 7 min at 72 °C. The PCR products were electrophoresed along with GeneScan ROX 400 (Applied Biosystems) internal size standard on an ABI PRISM 377 Automated Genetic Sequencer (Applied Biosystems). Allele sizes were assigned against the internal size standard and individuals were genotyped using GENESCAN and GENOTYPER softwares (Applied Biosystems).

Selection of STR loci

Six native STR primers of *N. heimii* (Iwata *et al.* 2000) and 45 non-native STR primers developed for *Hopea bilitonensis* (Lee *et al.* 2004a), *Shorea leprosula* (Lee *et al.* 2004b), *S. lumutensis* (Lee *et al.* 2006b) and *S. curtisii* (Ujino *et al.* 1998) were selected for primer screening (Table 1). Selection of the STR loci was primarily based on the screening protocol described by Selkoe and Toonen (2006) with modification. Specifically, the primers were screened for specific amplification, level of polymorphism, single-locus mode of inheritance, presence of null alleles, size homoplasy and variation in repeat motifs.

Specific amplification and level of polymorphism

As *N. heimii* is a diploid, any case of a locus displaying more than two alleles in one individual that was not traceable to cross-contamination of samples was categorised as non-specific amplification. In addition, a locus was considered as polymorphic when two or more alleles were enumerated.

Single-locus mode of inheritance

Since a full-sib family was not available for *N. heimii*, the mode of inheritance was examined using three half-sib families. A locus was said to follow a single-locus mode of inheritance if the qualitative observation showed each progeny possessed at least one maternal allele.

Presence of null alleles

The presence of null alleles was tested using the program MICRO-CHECKER version 2.2.1 (Oosterhout *et al.* 2004). This program was performed to discriminate between Hardy–Weinberg deviations caused by null alleles or inbreeding and Wahlund effects using a Monte Carlo simulation. The theory of Hardy–Weinberg equilibrium was used to calculate the expected allele frequencies and the frequency of any detected null alleles.

Size homoplasy and variation in repeat motifs

For loci showing specific amplification, high level of polymorphism, single-locus mode of inheritance and absence of null alleles, direct sequencing was performed to assess the effect of size homoplasy, variation in repeat motifs and conservativeness of the STR sequence. Three homozygous individuals with identical fragment size and two with different sizes were sequenced to examine the variations in repeat numbers and their flanking regions. For this purpose, the PCR products were separated on 2% agarose gels which were run at 100 V for 45 min in 1 × TAE buffer and stained with ethidium bromide. The target fragments were purified using MinElute PCR Purification Kit (Qiagen) and sequenced in both directions using BigDye Terminator Sequencing Kit (Applied Biosystems) based on the standard dideoxy-mediated chain termination method. The sequencing thermal

Table 1 Six native STRs of *Neobalanocarpus heimii* and 45 STRs developed for *Hopea bilitonensis*, *Shorea leprosula*, *S. lumutensis* and *S. curtisii* used for primer screening in *N. heimii*

Origin	STR
Neobalanocarpus heimii	*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018*, *Nhe019*
Hopea bilitonensis	*Hbi016*, *Hbi019*, *Hbi022*, *Hbi055*, *Hbi116*, *Hbi159a*, *Hbi160*, *Hbi161*, *Hbi204*, *Hbi221*, *Hbi247*, *Hbi303a*, *Hbi316*, *Hbi325a*, *Hbi329*
Shorea leprosula	*Sle074a*, *Sle079*, *Sle105*, *Sle111a*, *Sle118*, *Sle216*, *Sle267*, *Sle271a*, *Sle280*, *Sle290*, *Sle291a*, *Sle294*, *Sle303a*, *Sle384*, *Sle392*, *Sle475*, *Sle562*, *Sle605*
S. lumutensis	*Slu044a*, *Slu057*, *Slu110*, *Slu124*, *Slu175*, *Slu315*
S. curtisii	*Shc01*, *Shc02*, *Shc03*, *Shc04*, *Shc07*, *Shc09*

profile was 25 cycles at 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min on a GeneAmp PCR System 9700. Sequencing reactions were purified using ethanol precipitation and run on the ABI PRISM 377 Automated Genetic Sequencer. The resulting sequenced data were then edited using the Sequencing Analysis Software 3.7 (Applied Biosystems) and multiple alignments were performed using the program CLUSTALW (Thompson *et al.* 1994).

RESULTS

Specific amplification and polymorphism

In the initial screening of 45 primer pairs developed for *Hopea* and *Shorea* species, 15 failed to amplify, one was monomorphic and 17 produced multiple banding patterns. The remaining 12 (*Hbi016*, *Hbi161*, *Hbi329*, *Sle111a*, *Sle280*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04*, *Shc07* and *Shc09*) produced specific PCR products within the expected size range, which showed repeatable banding patterns and were polymorphic.

Single-locus mode of inheritance

The 12 specific amplified and polymorphic loci, together with 6 native loci of *N. heimii* (*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018* and *Nhe019*), were further evaluated for their modes of inheritance using three half-sib families. The results showed that, except for *Hbi329*, the amplified fragments of all loci segregated in a codominant manner and each of the progeny possessed at least one maternal allele, which supported the postulation of a single-locus mode of inheritance. For instance, in *Slu044a*, a heterozygote mother tree with alleles 147 and 151 bp transmitted at least one of the two alleles to the progenies (Figure 1a). In *Hbi329*, however, progenies 1 and 4 did not possess any maternal allele (Figure 1b). Hence, *Hbi329* was excluded from subsequent evaluations.

Presence of null alleles

The 17 specific amplified and polymorphic loci, with single-locus mode of inheritance were further evaluated for the presence of null alleles. Qualitative observations indicated the presence of null alleles in *Sle280*, as this locus repeatedly failed to be amplified in some individuals while all other loci were amplified normally. For the remaining 16 loci, a quantitative approach using the program MICRO-CHECKER revealed the absence of null alleles (Table 2).

Variation in repeat motifs and size homoplasy

Multiple alignments of partial sequences of the 16 STR loci are illustrated in Figure 2. The result showed that the tandemly repeated sequences were present in all the 16 loci although the repeat motifs and the number of repeats varied considerably. The six native loci showed either perfect (*Nhe004*, *Nhe005* and *Nhe011*) or compound repeats (*Nhe015*, *Nhe018* and *Nhe019*) although *Nhe018* and *Nhe019* were previously reported as perfect repeats. Most of the 10 non-native loci showed imperfect repeats (*Hbi016*, *Sle605*, *Slu044a*, *Shc04* and *Shc07*) compared with perfect (*Sle111a*, *Shc03* and *Shc09*) and compound repeats (*Hbi161* and *Sle392*). The most notable feature of the non-native loci was the changing of perfect repeats to compound (*Hbi161* and *Sle392*) and imperfect repeats (*Hbi016*, *Sle605*, *Slu044a* and *Shc04*). Specifically, the flanking and repeat regions of these non-native loci contained

Table 2 Estimated null allele frequencies of 16 STRs for *N. heimii*

Locus	Estimated null allele frequency
Nhe004	-0.0617
Nhe005	-0.0388
Nhe011	0.014
Nhe015	-0.0241
Nhe018	-0.048
Nhe019	-0.0082
Hbi016	-0.0628
Hbi161	-0.0942
Sle111a	-0.0487
Sle392	-0.0257
Sle605	0.0282
Slu044a	-0.0974
Shc03	-0.0008
Shc04	-0.003
Shc07	-0.0107
Shc09	-0.0574

The presence of null alleles is significant if the null allele frequency is greater than 0.05.

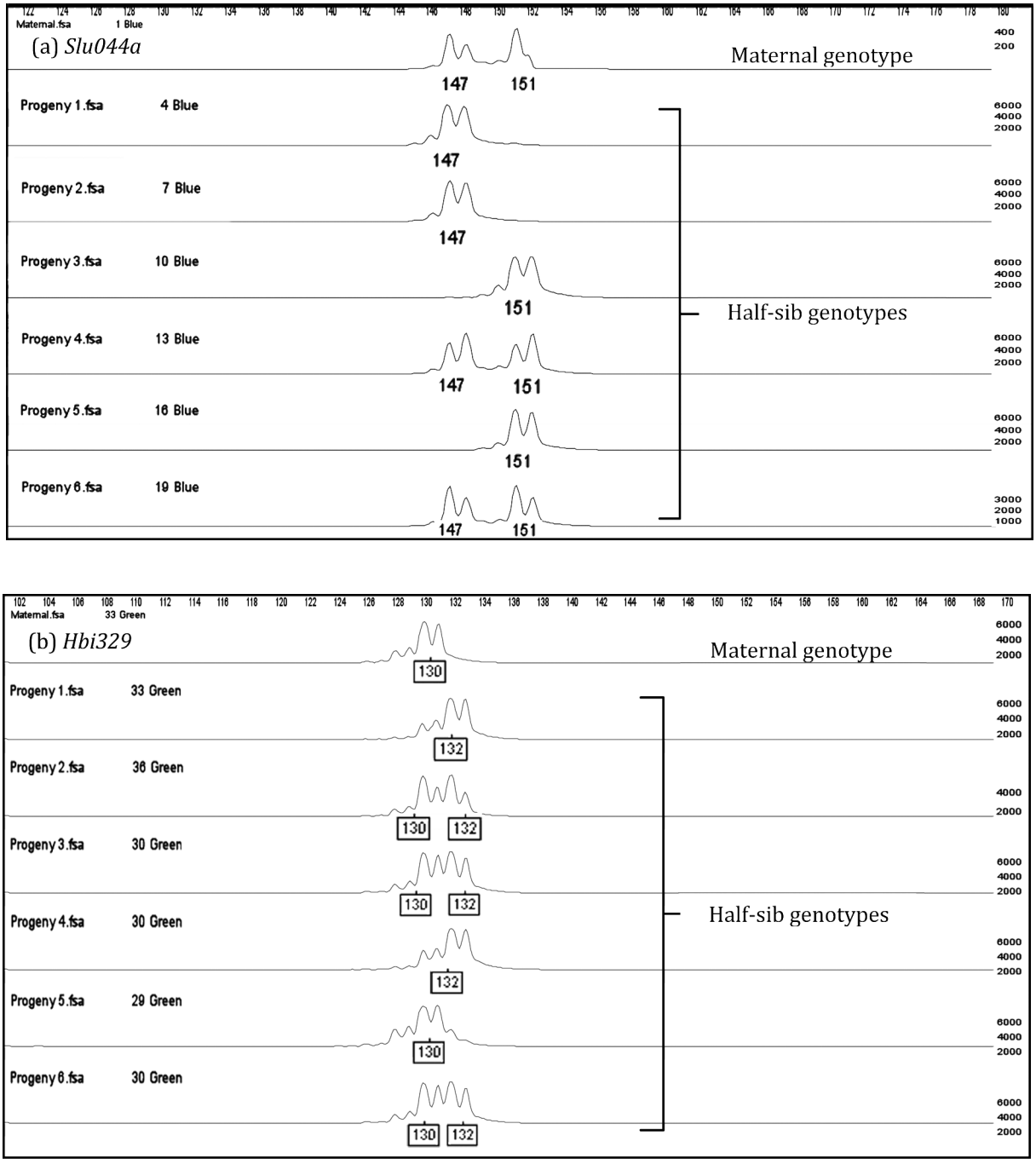
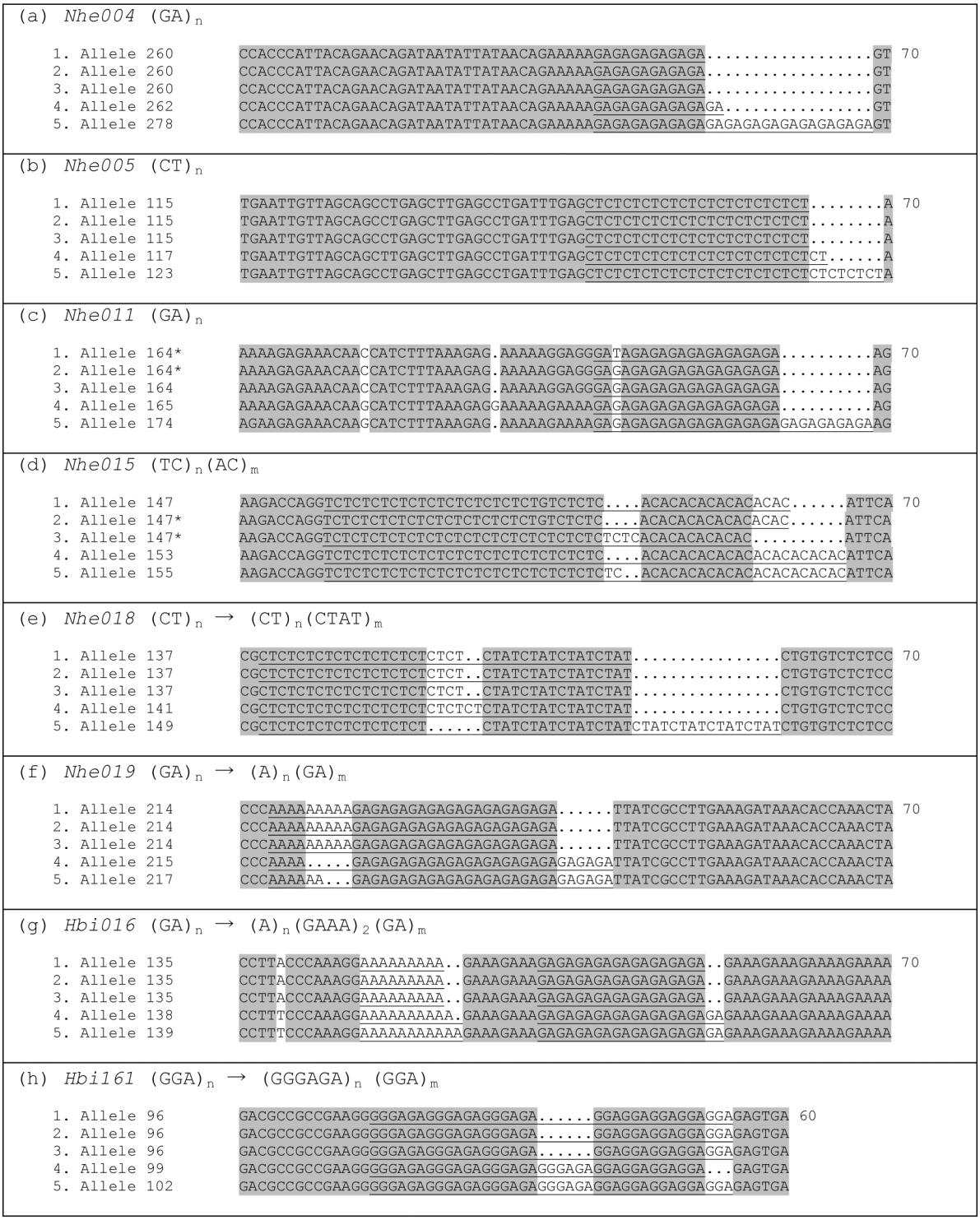


Figure 1 Qualitative observations showing that each progeny possessed at least one maternal allele to support the postulation of single-locus mode of inheritance in *Slu044a* (a) but not in *Hbi329* (b)

numerous imperfections resulting from indels or substitutions. In addition, two non-native loci were changed from dinucleotide perfect repeats to mononucleotide perfect repeats (*Sle111a* and *Shc09*).

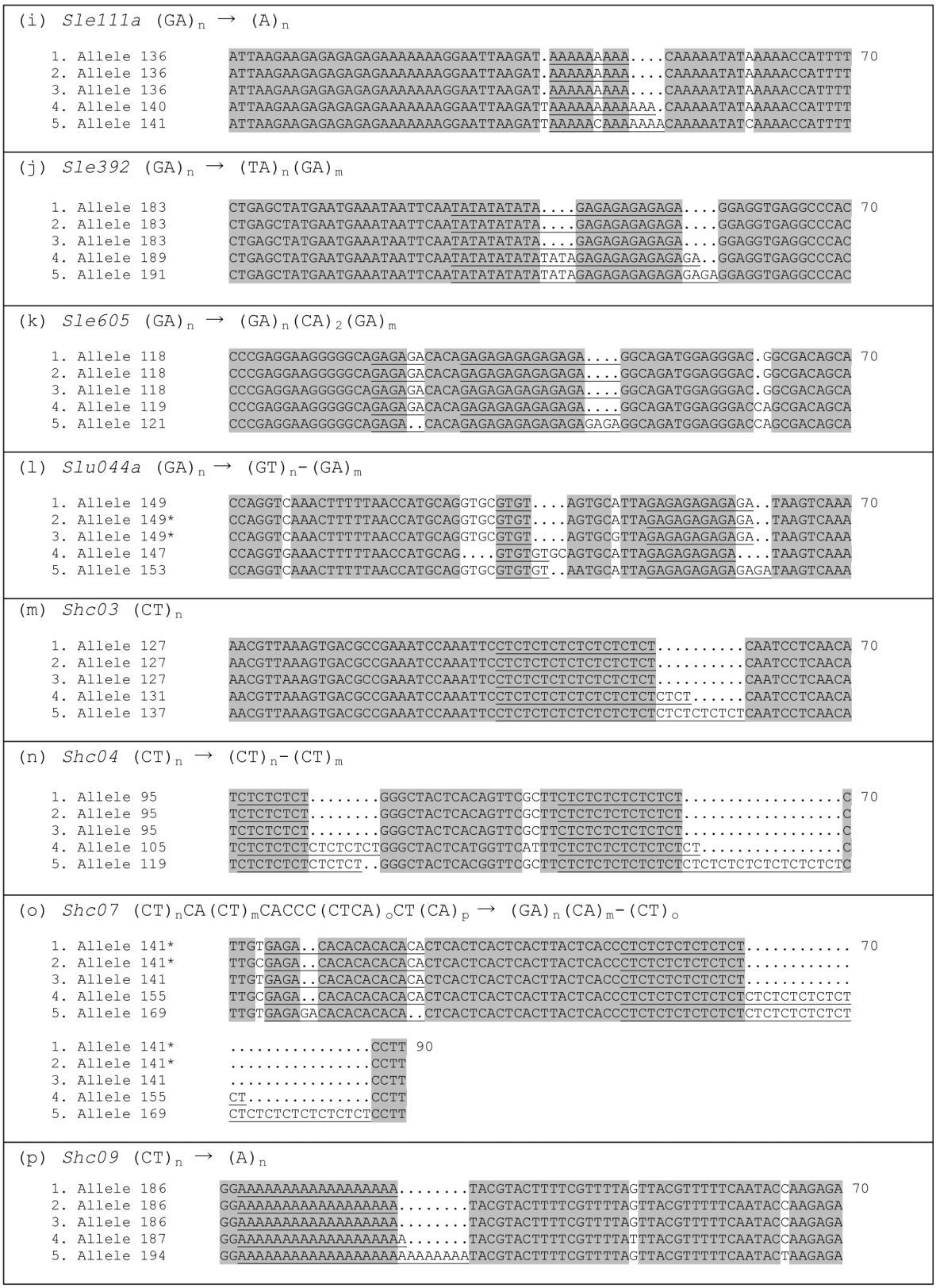
In *Hbi161*, the repeat motif of $(GGA)_n$ was reported in the native species *H. bilitonensis*. However, when the locus was cross-amplified to *N. heimii*, the repeat motif was modified to $(GGGAGA)_n(GGA)_m$. Similarly, in *Hbi016*, the repeat motif was changed from $(GA)_n$ to $(A)_n(GAAA)_2(GA)_m$. In addition, the length variations observed in loci *Nhe019* and *Hbi016* were not solely caused by the number of variations in the primary tandem repeat unit but mononucleotide runs in sequences immediately flanking the di/tetranucleotide repeat motif. For instance, in *Hbi016*, although both the 138 and the 139 bp



(continued)

Figure 2 Multiple alignments of partial sequences of 16 STRs in *Neobalanocarpus heimii* (a–p). For each of the locus, three homozygous individuals with identical allele size (1–3) and two homozygous individuals with different allele sizes (4–5) were sequenced to examine the variations in repeat number and their flanking regions. Stops indicate gaps, underlined regions denote STR repeat regions and shadings highlight the base conservation in all the five individuals. Size homoplasy was observed in allele size with asterisk (*).

Figure 2 (continued)



alleles contained a similar number of dinucleotide repeats $(GA)_{10}$, the 138 bp allele contained 9 poly(A) repeat units whereas the 139 bp allele consisted of 10 poly(A) repeats. This also means that the size variation was solely caused by the indel in the mononucleotide repeat.

Size homoplasy was observed in both the native and non-native loci, i.e. in 4 out of 16 loci—either due to single-base substitution (*Nhe011*, *Slu044a* and *Shc07*) or variation in the number of tandem repeats (*Nhe015*). The single-base substitution in the flanking region appeared to be the common factor contributing to size homoplasy. In most loci, variation in fragment sizes corresponded to variation in the number of tandem repeats. For example, in *Nhe004*, *Nhe005*, *Nhe015*, *Nhe018*, *Hbi161*, *Sle392* and *Shc03*, the variation in fragment sizes were solely reflected by the variation in the repeat regions. Similarly, in *Hbi016*, *Shc07* and *Shc09*, although point mutations were apparent in the flanking regions, the differences in fragment length were still reflected by the variation in the repeat regions. However, the differences in fragment lengths in *Nhe011*, *Sle111a*, *Sle605* and *Slu044a* were caused by variation in both the flanking and repeat regions.

As a whole, out of the 16 loci evaluated, four loci (*Nhe019*, *Hbi016*, *Sle111a* and *Shc09*) comprised mononucleotide repeat motifs. Size homoplasy was observed at four loci (*Nhe011*, *Nhe015*, *Slu044a* and *Shc07*). Since size homoplasy does not correspond to a significant problem for DNA profiling (Gill *et al.* 1997), and mononucleotide repeat motifs often cause complexity in allele sizing and instability during PCR amplifications (Hancock 1999), loci *Nhe019*, *Hbi016*, *Sle111a* and *Shc09* were excluded. Thus, the following 12 STR loci were finally selected to generate a DNA profiling database for *N. heimii*: *Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018*, *Hbi161*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04* and *Shc07* (Figure 3).

DISCUSSION

The ultimate goal of DNA profiling analysis is to reliably distinguish unrelated individuals from one another. For this reason, DNA markers such as STRs, which possess numerous observed alleles and show high levels of heterozygosity and high polymorphic information content, are desirable to generate DNA profiling databases for individual identification. In the general consensus, inclusion of many polymorphic STRs will increase the reliability of the data set. However, before STRs can be incorporated into any DNA profiling work, the quality of the marker (such as specific amplification, single-locus mode of inheritance, absence of null alleles and easy interpretation of genotype) needs to be

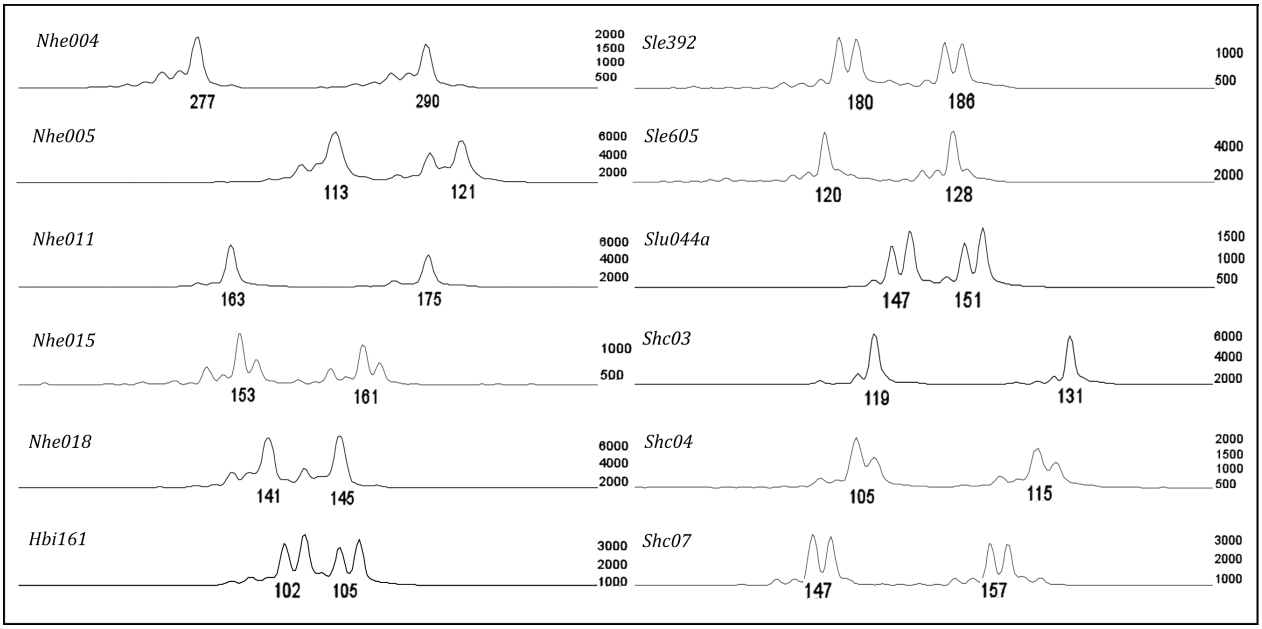


Figure 3 Electropherograms of the 12 selected STR loci (*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015*, *Nhe018*, *Hbi161*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04* and *Shc07*) in *N. heimii*

evaluated, since the inclusion of low quality loci can affect both the precision and accuracy of estimates (Selkoe & Toonen 2006).

A major impediment to STR analysis in the past was the need to develop PCR primers for every species. However, many studies have shown that STR primers developed in one species can be cross-amplified in related taxa (White & Powell 1997, Stacy *et al.* 2001, Guicking *et al.* 2006, Lee *et al.* 2006a). Molecular phylogenetic analysis of Dipterocarpaceae reveals close affinities among *Shorea*, *Hopea* and *Neobalanocarpus* (Tsumura *et al.* 1996, Kajita *et al.* 1998, Dayanandan *et al.* 1999). *Neobalanocarpus heimii* is monotypic (Ashton 1982) and has been postulated to be derived via hybridisation between the ancestral lineages of *Shorea* and *Hopea* (Kamiya *et al.* 2005). However, in this study, only 12 of the 45 primer pairs developed for *H. bilitonensis* (Lee *et al.* 2004a), *S. leprosula* (Lee *et al.* 2004b), *S. lumutensis* (Lee *et al.* 2006b) and *S. curtisii* (Ujino *et al.* 1998) were successfully amplified in *N. heimii*. The low success rate of transferability among genera of the same family has also been reported in Mimosaceae (Dayanandan *et al.* 1997), Fagaceae (Steinkellner *et al.* 1997), Brassicaceae (Van Treuren *et al.* 1997), Asteraceae (Whitton *et al.* 1997), Fabaceae (Peakall *et al.* 1998), Oleaceae (De la Rosa *et al.* 2002) and Leguminosae (Lee *et al.* 2006b). This may indicate that the success of cross-species amplification appear to be associated with taxonomic classification at the genus level only.

Qualitative observations indicated the presence of null alleles in *Sle280* as this locus repeatedly failed to be amplified in some individuals while all other loci were amplified normally. Allele dropout may occur due to mutations at the 3'-end of a primer and thus produces little or no extension during PCR (Butler 2005). If a base pair change occurs in the DNA template at the PCR primer binding region, the hybridisation of the primer can be disrupted resulting in a failure to amplify and, therefore, failure to detect an allele that exists in a template DNA. The potential of null alleles is not a problem within a laboratory that uses the same primer set to amplify a particular STR marker. However, with the emergence of national and international DNA databases, which store only the genotype information for a sample, allele dropout can potentially result in a false negative or incorrect exclusion of two samples that come from a common source (Butler 2005). There are several possible solutions to overcome the allele dropout (Urquhart *et al.* 1994, Holt *et al.* 2002, Nelson *et al.* 2002) but the easiest is to simply drop the STR locus.

Forensic STR analysis is based on comparing DNA fragment (allele) sizes for known samples and also the query samples over a sufficient number of predetermined loci. Typing for length polymorphism will not resolve differences in allele sequence. Virtually many STR loci have alleles that exhibit complex repeat sequences (Kobilinsky *et al.* 2005). Unless sequencing is done, the difference between alleles goes unnoticed. In the present study, although a relatively few number of STR alleles were sequenced, it was the only way to examine the variation in the repeat motifs, size homoplasy and the conservativeness of non-native STR in *N. heimii*. Subsequently, the sequence information is particularly important to understand the origin, mutational processes and structure of STR.

The sequence data showed that the two non-native loci (*Sle111a* and *Shc09*) were changed from dinucleotide perfect repeats to mononucleotide perfect repeats. These two loci should be excluded from forensic analysis since mononucleotide repeats often cause complexity in allele sizing and instability during PCR amplifications (Hancock 1999). The sequence data also clearly showed the tendency of non-native loci changing from perfect to compound (*Hbi161* and *Sle392*) or imperfect repeats (*Hbi016*, *Sle605*, *Slu044a* and *Shc04*) after cross-amplification in *N. heimii* (Figure 2). For example, at locus *Sle392*, a single unit of TA was reported to flank GA repeats in *S. leprosula* (Lee *et al.* 2004b). However, after cross-amplifying in *N. heimii*, four repeat units of TA were found to flank the GA repeats. Similarly, in locus *Hbi161*, a single unit of GGGAGA was reported to flank GGA repeats in *H. bilitonensis* (Lee *et al.* 2004a), but three repeat units of GGGAGA were found in *N. heimii* after cross-amplification. This also meant that the repeat units of TA and GGGAGA had undergone expansion in *N. heimii*. Short tandemly repeated sequences, with as few as two to four repeat units, are the starting points for subsequent STR expansion (Primmer & Ellegren 1998). Therefore, the increase in the numbers of TA and GGGAGA repeats might have taken place independently in *N. heimii* after the divergence of *Neobalanocarpus* from *Shorea* and *Hopea* genera. A similar expansion of STR in *Pinus* had been reported but the authors suggested that base

substitution could be one of the factors that provided material for replication slippage or some other inserting mechanism which enabled the expansion of the STR (Karhu *et al.* 2000).

The locus *Sle605* contained numerous interruptions after cross-amplification in *N. heimii*, i.e. the middle region had degenerated due to several indels and base substitutions. Such interruption of STR repeats was also observed in *Arabis* (Van Treuren *et al.* 1997) and *Glycine* (Peakall *et al.* 1998). Interruptions of perfect repeats may be a mechanism to stabilise STRs. For instance, an interruption of a STR in yeast decreases the mutation rates fivefold (Petes *et al.* 1997). Also, many studies concluded that indels and base substitutions could influence subsequent evolutionary rates and should occur frequently between close relatives (Estoup *et al.* 1995, Angers & Bernatchez 1997). Interruption tends to break up perfect repeats and reduce slippage, so the accumulation may eventually lead to degradation and loss of STR loci (Zhu *et al.* 2000). STRs comprising interrupted repeats correspond to the death stage of STR, whereas compound STR can be represented as the late stage in STR evolution (Kutil & Williams 2001).

The sequence data also indicated that variation in fragment length might not be solely due to variations in the number of tandem repeats but could also be due to indel mutations in the flanking regions. This possibility is supported by the occasional presence of odd-sized alleles in otherwise even-sized alleles or vice versa within a locus with di-, tetra- or hexanucleotide repeat motifs. For example, the *Nhe011* and *Sle605* loci with dinucleotide repeat motifs exhibited a mix of odd- and even-sized alleles. Several studies postulated that the indel mutations might be responsible for the observation of odd- and even-sized alleles within a locus (Grimaldi & Crouau-Roy 1997). In contrast, the presence of several odd-sized alleles among even-sized alleles is presumably caused by the formation of secondary structures that altered the gel mobility (Bull *et al.* (1999).

Size homoplasy was observed in 4 out of 16 loci, which was either due to single-base substitutions or variations in the number of tandem repeats. This also meant that different alleles could be identical in length but not identical by descent or they have the same length but yet they are entirely different alleles. The single-nucleotide substitution in the flanking regions appeared to be the major factor contributing to size homoplasy. Given that relatively few STR alleles were being sequenced in the present study, size homoplasy might be common at other loci as well. In order to detect size homoplasy, analysing a sufficiently large number of individuals and loci within a population is crucial to draw a clear relationship for size homoplasy in a population genetics context (Estoup *et al.* 2002). Empirical studies in population genetics showed that there is a substantial influence of size homoplasy within and between populations, and size homoplasy is normally rare within populations and common between populations (Estoup *et al.* 1995, Primmer & Ellegren 1998, Viard *et al.* 1998, Jones *et al.* 1999). The occurrence of size homoplasy is well known in human (Möller *et al.* 1995, Schneider *et al.* 1998, Walsh *et al.* 2003) and animal STRs (Eichmann *et al.* 2004). According to the European DNA Profiling Group (EDNAP), size homoplasy might not correspond to a significant problem for DNA profiling and subsequently size homoplasy would not affect evidence interpretation (Gill *et al.* 1997). The high degree of polymorphism at STR loci is often largely compensated by their homoplasious evolution, hence the gain associated with the detection of size homoplasy data appear to be minimum in most cases (Estoup *et al.* 2002). In addition, a match in forensic analysis is usually made against many loci, not just one (Butler 2005).

In this study, the STR loci comprised mainly dinucleotide repeat motifs. In human and animal, STRs with tetra- or pentanucleotide repeats are generally preferable than STRs with di- or trinucleotide repeats (Bacher & Schumm 1998, Eichmann *et al.* 2004, Butler 2005). Tetra- or pentanucleotide repeats have been claimed to generate few stutter peaks or slippage artifacts. The percentage of stutter peak formation in tetranucleotide repeats is generally less than 15%, whereas for di- and trinucleotides repeats the percentage can reach as high as 30% (Butler 2005). In addition, stutter peaks have been reported to be more common in longer alleles than in shorter ones (Butler 2005). Nevertheless, despite the existence of stutter peaks, they generally do not cause confusion in the interpretation of the locus genotype except when mixtures exist (Kobilinsky *et al.* 2005). Moreover, the amount of stutter peak formation may be reduced when using DNA polymerase with faster processivity (Butler 2005).

STR polymorphisms have been reported to correlate with a number of tandem repeats (Weber 1990, Bryan *et al.* 1997, Smulders *et al.* 1997) but this relationship is not always apparent (Szewc-McFadden *et al.* 1996). For example, in this study, *Nhe004* with 6–15 repeat units consisted of a large number of alleles. Similarly, in *Lycopersicon* (Smulders *et al.* 1997) and *Arabis* (Van Treuren *et al.* 1997), STRs with four repeat units also showed high degrees of polymorphisms. This indicates that STRs with shorter repeats may well be polymorphic and these loci should not be ignored as potentially informative markers (Peakall *et al.* 1998).

In summary, after an extensive evaluation of 51 STRs developed for Dipterocarpaceae, finally, five native (*Nhe004*, *Nhe005*, *Nhe011*, *Nhe015* and *Nhe018*) and seven non-native (*Hbi161*, *Sle392*, *Sle605*, *Slu044a*, *Shc03*, *Shc04* and *Shc07*) STR loci, which showed specific amplification, high polymorphism, single-locus mode of inheritance, absence of null alleles and absence of mononucleotide repeat motifs, were selected for DNA profiling of *N. heimii*. It is anticipated that the number of STR loci used in the current study is sufficient in accordance with others forensic study. For example, 10 STR loci were used in *Cannabis sativa* to screen against a worldwide population of 255 individuals representing 33 countries (Zinnamon & Keim 2003). Gilmore *et al.* (2003) used five STR loci to discriminate between individuals and accessions of 93 samples of *C. sativa* worldwide, whilst Howard *et al.* (2009) used 10 STR loci to discriminate 510 *C. sativa* samples representing both fibre and drug varieties. In humans, North American Laboratories use 13 core CODIS loci, while European laboratories largely use the SGM Plus kit that amplifies only 10 STR loci (Butler 2005).

ACKNOWLEDGEMENTS

The Forest Departments of Negeri Sembilan and Johore are acknowledged for granting us permission to access the forest reserves. We thank the District Forest Officers and the staff of the Ranger Offices who provided assistance and logistic support during the sample collections; and D Mariam, J Ghazali, M Yahya, P Ramli, T Sharifah, CS Suryani and AW Nor Salwah for their excellent assistance in the laboratory and in the field. This project was supported in part by the e-Science Research Grant (02-03-10-SF0009) entitled 'Development of DNA barcode of *Neobalanocarpus heimii* (chengal) as a tool for forensics and chain of custody certification' and the Biodiversity International Agreement No. APO 05/016.

REFERENCES

Angers B & Bernatchez L. 1997. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Molecular Biology and Evolution* 14: 230–238.

Ashton PS. 1982. Dipterocarpaceae. Flora Malesiana. Series I. *Spermatophyta* 9: 237–552.

Bacher J & Schumm JW. 1998. Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. *Profiles in DNA* 2: 3–6.

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB & Gale MD. 1997. Isolation and characterization of microsatellites from hexaploid bread wheat. *Theoretical and Applied Genetics* 94: 557–563.

Bull LN, Pabón-Peña CR & Freimer NB. 1999. Compound microsatellite repeats: practical and theoretical features. *Genome Research* 9: 830–838.

Butler JM. 2005. *Forensic DNA Typing: Biology, Technology, and Genetics of STR Marker*. Second edition. Elsevier Academic Press, Burlington.

Chua LSL. 1998. *Neobalanocarpus heimii*. In *IUCN 2008, 2008 IUCN Red List of Threatened Species*. Available at http://www.iucnredlist.org.

Coyle HM, Ladd C, Palmbach T & Lee HC. 2001. The green revolution: botanical contributions to forensics and drug enforcement. *Croatian Medical Journal* 42: 340–345.

Coyle HM, Palmbach T, Juliano N, Ladd C & Lee HC. 2003. An overview of DNA methods for the identification and individualization of marijuana. *Croatian Medical Journal* 44: 315–323.

Craft KJ, Owens JD & Ashley MV. 2007. Application of plant DNA markers in forensic botany: genetic comparison of *Quercus* evidence leaves to crime scene trees using microsatellites. *Forensic Science International* 165: 64–70.

Dayanandan S, Bawa KS & Kesseli R. 1997. Conservation of microsatellites among tropical trees (Leguminosae). *American Journal of Botany* 84: 1658–1663.

Dayanandan S, Dole J, Bawa K & Kesseli R. 1999. Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianesis* (Meliaceae). *Molecular Ecology* 8: 1585–1592.

De la Rosa R, James CM & Tobutt KR. 2002. Isolation and characterization of polymorphic microsatellites in olive (*Olea europeae* L.) and their transferability to other genera in the Oleaceae. *Molecular Ecology Notes* 2: 265–267.

Eichmann C, Berger B & Parson W. 2004. A proposed nomenclature for 15 canine-specific polymorphic STR loci for forensic purposes. *International Journal of Legal Medicine* 118: 249–266.

Estoup A, Jarne P & Cornuet JM. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591–1604.

Estoup A, Tailliez C, Cornuet JM & Solignac M. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Molecular Biology and Evolution* 12: 1074–1084.

Gill P, Brinkmann B, d'Aloja E, Andersen J, Bar W, Carracedo A, Dupuy B, Eriksen B, Jangblad A, Johnsson V, Kloosterman AD, Lincoln P, Morling N, Rand S, Sabatier M, Scheithauer R, Schneider PM & Vide MC. 1997. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. *Forensic Science International* 87: 185–192.

Gilmore S, Peakall R & Robertson J. 2003. Short tandem repeat (STR) DNA markers are hypervariable and informative in *Cannabis sativa*: implications for forensic investigations. *Forensic Science International* 131: 65–74.

Grimaldi MC & Crouau-Roy B. 1997. Microsatellite allelic homoplasy due to variable flanking sequences. *Journal of Molecular Evolution* 44: 336–340.

Guicking D, Rana TS, Blattner FR & Weising K. 2006. Microsatellite markers for the palaeotropic pioneer tree genus *Macaranga* (Euphorbiaceae) and their cross-species transferability. *Molecular Ecology Notes* 6: 245–248.

Hancock JM. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. Pp. 1–9 in Goldstein DB & Schlötterer C (Eds.) *Microsatellites: Evolution and Applications*. Oxford University Press, Oxford.

Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD & Walsh PS. 2002. TWGDAM validation of AmpFISTRTM PCR amplification kits for forensic DNA casework. *Journal of Forensic Sciences* 47: 66–96.

Howard C, Gilmore S, Robertson J & Peakall R. 2009. A *Cannabis sativa* STR genotype database for Australian seizures: forensic applications and limitations. *Journal of Forensic Sciences* 54: 556–563.

Iwata H, Konuma A & Tsumura Y. 2000. Development of microsatellite markers in the tropical tree *Neobalanocarpus heimii* (Dipterocarpaceae). *Molecular Ecology* 9: 1661–1686.

Jones AG, Rosenqvist G, Berglund A & Avise JC. 1999. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* 152: 1057–1063.

Kajita T, Kamiya K, Nakamura K, Tachida H, Wickneswari R, Tsumura Y, Yoshimaru H & Yamazaki T. 1998. Molecular phylogeny of Dipterocarpaceae in Southeast Asia based on nucleotide sequences of *mat*K, *trn*L intron, and *trn*L-*trn*F intergenic spacer region in chloroplast DNA. *Molecular Phylogenetics and Evolution* 10: 202–209.

Kamiya K, Harada K, Tachida H & Ashton PS. 2005. Phylogeny of *PgiC* gene in *Shorea* and its closely related genera (Dipterocarpaceae), the dominant trees in Southeast Asian tropical rain forest. *American Journal of Botany* 92: 775–788.

Karhu A, Dietrich JH & Savolainen O. 2000. Rapid expansion of microsatellite sequences in pines. *Molecular Biology and Evolution* 17: 259–265.

Kobilinsky L, Liotti T & Oeser-Sweat J. 2005. *DNA: Forensic and Legal Applications*. John Wiley & Sons, Hoboken.

Korpelainen H & Virtanen V. 2003. DNA fingerprinting of mosses. *Journal of Forensic Sciences* 48: 804–807.

Kutil BL & Williams CG. 2001. Triplet-repeat microsatellites shared among hard and soft pines. *Journal of Heredity* 92: 327–332.

Lee CT, Lee SL, Faridah QZ, Siraj SS, Ng KKS, Norlia B & Mat-Isa MN. 2006a. Isolation and characterization of microsatellite markers in *Koompassia malaccensis* (Leguminosae), an important tropical timber species. *Molecular Ecology Notes* 6: 1198–1201.

Lee SL, Ng KKS, Saw LG, Lee CT, Norwati M, Tani N, Tsumura Y. & Koskela J. 2006b. Linking the gaps between conservation research and conservation management of rare dipterocarps: a case study of *Shorea lumutensis*. *Biological Conservation* 131: 72–92.

Lee SL, Tani N, Ng KKS & Tsumura Y. 2004a. Characterization of 15 polymorphic microsatellite loci in an endangered tropical tree *Hopea bilitonensis* (Dipterocarpaceae) in Peninsular Malaysia. *Molecular Ecology Notes* 4: 147–149.

Lee SL, Tani N, Ng KKS & Tsumura Y. 2004b. Isolation and characterization of 20 microsatellite loci for an important tropical tree *Shorea leprosula* (Dipterocarpaceae) and their applicability to *S. parvifolia*. *Molecular Ecology Notes* 4: 222–225.

Levinson G & Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203–221.

Mestel R. 1993. Murder trial features trees genetic fingerprint. *New Science* 138: 6.

Möller A, Schürenkamp M & Brinkmann B. 1995. Evaluation of an ACTBP2 ladder composed of 26 sequenced alleles. *International Journal of Legal Medicine* 108: 75–78.

Murray M & Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4325.

Nelson MS, Levedakou EN, Matthews JR, Early BE, Freeman DA, Kuhn CA, Sprecher CJ, Amin AS, McElfresh KC & Schumm JW. 2002. Detection of a primer-binding site polymorphism for the STR locus D16S539 using the PowerPlex(R) 1.1. system and validation of a degenerate primer to correct for the polymorphism. *Journal of Forensic Sciences* 47: 345–349.

Oosterhout V, Hutchinson WF, Wills DM & Shipley P 2004. MICRO-CHECKER: a software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

Peakall R, Gilmore S, Keys W, Morgante M & Rafalski A. 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants. *Molecular Biology and Evolution* 15: 1275–1287.

Petes TD, Greenwell PW & Dominska M. 1997. Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics* 146: 491–498.

Primmer CR & Ellegren H. 1998. Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution* 15: 997–1008.

Schneider HR, Rand S, Schmitter H & Weichhold G. 1998. ACTBP2-nomenclature recommendations of GEDNAP. *International Journal of Legal Medicine* 111: 97–100.

Selkoe KA & Toonen RJ. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629.

Siver PA, Lord WD & McCarthy DJ. 1994. Forensic limnology: the use of freshwater algal community ecology to link suspects to an aquatic crime scene in southern New England. *Journal of Forensic Sciences* 39: 847–853.

Smulders MJM, Bredemeijer G, Ruskortekaas W, Arens P & Vosman B. 1997. Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. *Theoretical and Applied Genetics* 94: 264–272.

Stacy EA, Dayanandan S, Dancik BP & Khasa PD. 2001. Microsatellite DNA markers for the Sri Lanka rainforest tree species, *Shorea cordifolia* (Dipterocarpaceae), and cross-species amplification in *S. megistophylla*. *Molecular Ecology Notes* 1: 53–54.

Steinkellner H, Lexer C, Turetschek E & Glössl J. 1997. Conservation of $(GA)_n$ microsatellite loci between *Quercus* species. *Molecular Ecology* 6: 1189–1194.

Szewc-McFadden AK, Kresovich S, Bliek SM, Mitchell SE & McFerson JR. 1996. Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. *Theoretical and Applied Genetics* 93: 534–538.

Thompson JD, Higgins DG & Gibson TJ. 1994. clustalw: improving the sensitivity of progressive multiple alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

Tsumura Y, Kawahara T, Wickneswari R & Yoshimura K. 1996. Molecular phylogeny of Dipterocarpaceae in Southeast Asia using RFLP of PCR-amplified chloroplast genes. *Theoretical and Applied Genetics* 93: 22–29.

Ujino T, Kawahara T, Tsumura Y, Nagamitsu T, Yoshimaru H & Wickneswari R. 1998. Development and polymorphism of simple sequence repeat DNA markers for *Shorea curtisii* and other Dipterocarpaceae species. *Heredity* 81: 422–428.

Urquhart A, Kimpton CP, Downes TJ & Gill P. 1994. Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. *International Journal of Legal Medicine* 107: 13–20.

Van Treuren R, Kuittinen H, Kärkkäinen K, Baena-Gonzalez E & Savolainen O. 1997. Evolution of microsatellites in *Arabis petraea* and *Arabis Iyrata*, out-crossing relatives of *Arabidopsis thaliana*. *Molecular Biology and Evolution* 14: 220–229.

Viard F, Franck P, Dubois MP, Estoup A & Jarne P. 1998. Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution* 47: 42–51.

Walsh SJ, Robinson SL, Turbett GR, Davies NP & Wilton AN. 2003. Characterization of variant alleles at the HumD21S11 locus implies unique Australasian genotypes and re-classification of nomenclature guidelines. *Forensic Science International* 135: 35–41.

Weber JL. 1990. Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. *Genomics* 7: 524–530.

Weising K, Nybom H, Wolff K & Kahl G. 2005. *DNA Fingerprinting in Plants: Principles, Methods, and Applications*. Second edition. Taylor & Francis Group, Boca Raton.

White E, Hunter J, Dubetz C, Brost R, Bratton A, Eded S & Sahota R. 2000. Microsatelllite markers for individual tree genotyping: application in forest crime prosecutions. *Journal of Chemical Technology and Biotechnology* 75: 923–926.

White GW & Powell W. 1997. Isolation and characterization of microsatellite loci in *Swietenia humillis* (Meliaceae): an endangered tropical hardwood species. *Molecular Ecology* 6: 851–860.

Whitton J, Rieseberg LH & Ungerer MC. 1997. Microsatellite loci are not conserved across the Asteraceae. *Molecular Biology and Evolution* 14: 204–209.

Yoon CK. 1993. Botanical witness for the prosecution. *Science* 260: 894–895.

Zhu Y, Queller DC & Strassmann JE. 2000. Phylogenetic perspective on sequence evolution in microsatellite loci. *Journal of Molecular Evolution* 50: 324–338.

Zinnamon KN & Keim P. 2003. Identification of individual marijuana (*Cannabis sativa*) plants using short tandem repeat (STR) markers: forensic applications. *Proceedings of American Academy of Forensic Sciences Annual Meeting*. 17–22 February 2003, Chicago.