View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Malaya Students Repository

CHAPTER 4

GEOGRAPHICAL TRACEABILITY OF AN IMPORTANT TROPICAL TIMBER (*NEOBALANOCARPUS HEIMII*) INFERRED FROM CHLOROPLAST DNA

4.1 INTRODUCTION

New methods to match a timber log with its population of origin would signify an important forensic component in the context of stolen log traceability for the control of illegal logging and also the approach in chain of custody developed for the certification of timber from sustainably managed forests (Lyke, 1996; Chihambakwe *et al*., 1997). Indeed, illegal logging is a problem that not only destroys forest ecosystems in its own right but also threatens the viability of forest certification by depressing the price of timber and creating extremely low-priced competitor products (Cashore *et al.*, 2004). Although only timber products with legality licenses are allowed to enter European markets (Commission, 2003), it is estimated that 50% of the tropical timbers traded in the European markets are illegal (Richert, 2003).

In response to the increasing concern over illegal logging, nearly 10% of commercial forests worldwide have been certified as being “well-managed” by the Forest Stewardship Council in 2005 (FSC, 2005a). Driven by the rapid growth of the forest certification, in some countries, forest certification has become a regular aspect in the logging industries, and the timber trades have become more transparent, since the origins of certified timbers are known (Visseren-Hamakers & Glasbergen, 2007). Similarly, chain of custody approach and ecolabelling schemes have also proliferated in recent years and have attracted widespread participation of forest landowners (FSC, 2005b; Global Ecolabelling Network, 2008).

The implementation of forest certification and ecolabelling contributes to halting deforestation and ensuring a sustainable use of forest resources globally. However, there

may be products bearing ecolabels that do not actually meet the label's environmental standards (Global Ecolabelling Network, 2008). Most timber-auditing systems rely on tagging or certificates of origin issued in the source country to verify the legality of the timbers, but these methods are susceptible to falsification (Carr, 2007). Therefore, there is a need for a harmonized timber tracking system to trace and verify the origin of a suspected timber in order to reduce illegal timber trade and hence strengthen the cooperation between producers and consumer countries (Asia Forest Partnership, 2005).

In the past, spectrometry and isotopic methods have been applied and proposed to differentiate wood samples from different geographical origins that will permit their geographical origins to be determined with varying degrees of certainty (Perez-Coello *et al.*, 1997; Durand *et al.*, 1999; English *et al.*, 2001). However, these approaches are influenced by the local environment, variability of chemical composition and are limited by the fact that such markers can show a discrepancy between individuals from the same population or even between different tissues from the same individual (Hoffman *et al.*, 1994; Towey & Waterhouse, 1996). Hence, this has led to major advances in the use of inbuilt unique properties of DNA within the timber to support the determination of identity and provenance (Asia Forest Partnership, 2005). The use of DNA track-back system, once thought to be impossible for wood, is now feasible, though in its infancy. A good example is shown in the European white oaks. The strong geographical structure and differentiation of western vs. eastern populations were used for the oak wood traceability (Deguilloux *et al.*, 2003).

There are two very different ways in which DNA could be applied in timber tracking and forensic forestry investigations. First, cpDNA markers showing enough geographical structure could be used to differentiate the origin of one source of timber from another. Second, a highly polymorphic nuclear short tandem repeat (nSTR) marker could be used to generate DNA profiling databases for individual identification, in which an illegal timber log could be matched with its original stump (Tnah *et al.*, 2010). In combating illegal logging, both of these tools require rapid development of large

comprehensive databases, detailing the distribution of genetic markers and incorporating these DNA-based techniques into the traceability systems. It is important that these databases can be established as soon as possible, in order to capture the "natural" conditions, before the important patterns have been completely erased by human activity (Asia Forest Partnership, 2005).

In plants, molecular techniques using chloroplast DNA (cpDNA) markers have provided tools for studying the phylogeography or migration footprints of a species (Avise, 2000). Chloroplast DNA is thought to evolve slowly, with low mutation rates, and is known to be maternally inherited in most angiosperms (Wolfe *et al.*, 1987; Corriveau & Coleman, 1988; Clegg & Zurawski, 1992). Maternally inherited DNA markers generally reveal much greater genetic structure in comparison with biparentally inherited nuclear markers (Petit *et al.*, 1993a, b) and these markers have been successfully applied to identify possible glacial refugia and species migration routes of many plant species (Huang *et al.*, 2004; Cheng *et al.*, 2005; Fjellheim *et al.*, 2006; Shephard *et al.*, 2007). In principle, the geographical origin of wood samples can be checked with the cpDNA markers that show enough geographical structure. For example, a study on *Cedrela odorata* throughout Mesoamerica using cpDNA markers indicated a strong geographical pattern which can be used to determine the geographical origin of Mesoamerican *C. odorata* timbers (Cavers *et al.*, 2003).

Neobalanocarpus heimii or locally known as chengal is endemic but widely distributed in Peninsular Malaysia. It is found in diverse localities, on low-lying flat land as well as on hills of up to 900 m (Symington, 1943). *N. heimii* produces a naturally, highly durable wood and is among the strongest timbers in the world. It is used for heavy constructions, bridges, boats, buildings, and wherever strength is considered essential (Thomas, 1953). Under the IUCN Red List of Threaten Species, 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 might become endangered in the near future. Therefore, this study was aimed to (1) provide a detailed picture of the distribution of chloroplast haplotypes of *N. heimii* throughout Peninsular Malaysia and (2) identify specific haplotypes that could be used to generate population identification database and serve as a tracking and monitoring tools in the context of illegal logging, forest certification and chain of custody certification.

4.2 MATERIALS AND METHODS

4.2.1 Sample collection and DNA extraction

In order to generate a comprehensive database of *N. heimii* for population identification, sample collection was conducted throughout the distribution range of *N. heimii* in Peninsular Malaysia. Thirty-two natural populations of *N. heimii* from 29 forest reserves, with a total of 256 individuals of more than 10 cm diameter at breast height (eight samples per population) were investigated in this study (Table 4.1). The samples were collected either in the form of inner bark or leaf tissues. Total DNA was extracted using the procedure described by Murray and Thompson (1980) with modification, and further purified using the High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH).

4.2.2 PCR amplifications and sequencing

Twenty-seven cpDNA universal primer pairs of higher plants (Heinze, 2007) were screened to identify intraspecific variability; five cpDNA non-coding regions were proven to be informative in the characterization of haplotypes in *N. heimii*. However, one of the regions, *trn*S-*trn*G spacer was unsuccessfully amplified from the dry wood of *N. heimii*. Therefore, only four cpDNA regions: *trn*L intron (Taberlet *et al*., 1991), *trn*G intron (Heinze, 2007), *trn*K intron (Demesure *et al*., 1995; Weising & Gardner, 1999) and *psb*K-*trn*S spacer (Grivet *et al*., 2001; Heinze, 2007) were chosen to sequence all the 256 individuals of *N. heimii*. The PCR amplifications were performed in 20 µL

Table 4.1: Sampling localities chosen for the population of the *Neobalanocarpus heimii* in Peninsular Malaysia accompanied by the population codes. Eight samples per population were used in this study.

Forest reserve (FR)	State	Pop. code	Latitude	Longitude
1. Bukit Enggang	Kedah	BEnggang	05°48’	100°41’
2. Sungkop	Kedah	Sunkop	05°45’	100°38’
3. Bintang Hijau	Perak	BHijau	05°11’	101°00’
4. Piah	Perak	Piah	05°01’	101°02’
5. Pondok Tanjung	Perak	PTanjung	05°04’	100°47’
6. Bubu	Perak	Bubu	04°37’	100°46’
7. Chikus	Perak	Chikus	04°06’	101°12’
8. Jeli	Kelantan	Jeli	05°44’	101°50’
9. Gunung Basur	Kelantan	GBasur	05°36’	101°45’
10. Lebir	Kelantan	Lebir	05°12’	102°20’
11. Hulu Terengganu (compartment 31)	Terengganu	HTrengA	04°56’	102°55’
12. Hulu Terengganu (compartment 14A)	Terengganu	HTrengB	05°00’	102°55’
13. Pasir Raja	Terengganu	PRaja	04°42’	102°58’
14. Rambai Daun	Terengganu	RDaun	04°36’	103°23’
15. Berkelah	Pahang	Berkelah	03°46’	103°08’
16. Tersang	Pahang	Tersang	03°59’	101°49’
17. Rotan Tunggal	Pahang	RTunggal	03°47’	101°51’
18. Lakum	Pahang	Lakum	03°37’	102°05’
19. Bukit Tinggi	Pahang	BTinggi	03°31’	101°52’
20. Lentang	Pahang	Lentang	03°23’	101°54’
21. Kemasul	Pahang	Kemasul	03°25’	102°13’
22. Lesong	Pahang	Lesong	02°46’	103°08’
23. Gombak	Selangor	Gombak	03°20’	101°46’
24. Ampang	Selangor	Ampang	03°10’	101°47’
25. Sungai Lalang	Selangor	SLalang	03°05’	101°52’
26. Pelangai	Negeri Sembilan	Pelangai	02°48’	102°11’
27. Pasoh	Negeri Sembilan	Pasoh	02°59’	102°19’
28. Labis	Johor	Labis	02°21’	103°10’
29. Lenggor (compartment 32)	Johor	LenggorA	02°11’	103°40’
30. Lenggor (compartment 76)	Johor	LenggorB	02°10’	103°40’
31. Panti (compartment 16)	Johor	PantiA	01°47’	103°57’
32. Panti (compartment 68)	Johor	PantiB	01°49’	103°55’

reaction mixture, consisting of approximately 10 ng of template DNA, 50 mM of KCl, 20 mM of Tris–HCl (pH 8.0), 1.5 mM of $MgCl_2$, 0.4 μM of each primer, 0.2 mM of each dNTP, and 1 U of *Taq* DNA polymerase (Promega). The reaction mixture was subjected to amplification using GeneAmp PCR System 9700 (Applied Biosystems), for an initial denaturing step of 94 °C for 5 min, 30 cycles of 94 °C for 1 min, 50–55 °C annealing temperature for 1 min, and 72°C for 1 min. This was followed by further primer extension at 72°C for 8 min.

The PCR products were purified using the MinElute PCR Purification Kit (Qiagen) and sequenced in both directions using the BigDye Terminator Sequencing Kit (Applied Biosystems) based on the standard dideoxy-mediated chain termination method. The sequencing thermal 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 3130xl Genetic Analyzer (Applied Biosystems). Sequencing data were edited and assembled using CODONCODE ALIGNER version 2.0 (CodonCode Corporation). Haplotypes were determined from nucleotide substitutions and indels (insertions and deletions).

4.2.3 Population identification database and test of conformity of origin

As a result of their uniparental inheritance, cpDNA were expected to show pronounced levels of population differentiation and large proportion of population specific haplotypes. These haplotypes were used to generate a haplotype distribution map throughout Peninsular Malaysia. The population identification database was constructed based on the entire 256 samples dataset, which was fitted to classify the samples into their population of origin based on the significant intraspecific variable sites.

To test the conformity of origin, a statistical procedure based on either presence or absence of haplotype on wood sample was used to test whether a suspected wood lot conforms to a given geographical origin (Deguilloux *et al.*, 2003). The null hypothesis H_O: 'the wood lot is of the presumed origin' is defined. For each haplotype i identified

in the samples, $p(i)$ indicates its frequency in the region of presumed origin, while N denotes the number of the wood samples. The probability of $P(i)$ of observing a rare haplotype at least once is an indication of non-conformity, but the value would depend on the sample size. By using the classical threshold of $p(i) = 0.05$ as the criterion for rarity, the following are obtained:

$$\text{if } p(i) < 0.05, \text{ then } P(i) = 1 - [1 - p(i)]^N,$$
$$\text{if } p(i) \geq 0.05, \text{ then } P(i) = 1.$$

The probability P to observe the particular configuration of haplotypes in the samples is given by the $P = \Pi P(i)$. If $P < \mathrm{S}$, the hypothesis H_O is rejected and the wood lot is declared as non-conformity, S being the maximum risk accepted where S is set as 0.05, with a risk $\alpha \leq S$ (α is the probability of type I error).

The type I (false positive) and type II (false negative) errors of the southern and northern regions of population identification database were estimated by simulating 10,000 artificial wood lot origins from either the southern or northern provenance. The artificial wood lots were designed to have random haplotypes that were proportional to the haplotype frequencies determined in the 32 populations of *N. heimii* in Peninsular Malaysia. The conformity of each artificial wood lot origin from the northern or southern provenance was tested against the database with southern and northern regions. The proportion of wood lot that was considered as non-conformity was measured.

4.3 RESULTS

4.3.1 Population identification database and haplotype distribution

The examined sequences consisted of *trn*L intron (584-591bp), *trn*G intron (660-661bp), *trn*K intron (569-579bp) and *psb*K-*trn*S spacer (679bp). The corresponding GenBank accession numbers are EU918738–EU918743 and EU918751–EU918763. In total, 10 intraspecific variable sites were detected within these four noncoding regions (Table

Table 4.2: Distribution of significant intraspecific variable sites in the non-coding region of chloroplast DNA of *Neobalanocarpus heimii* in the population identification database. The positions of the variable sites in the *trn*L intron, *trn*G intron, *trn*K intron and *psb*K-*trn*S spacer, and (-) deletion are indicated.

Haplotype	*trn*L intron				*trn*G intron			*trn*K intron		*psb*K-*trn*S spacer
	201	306-312	380	457	56	83	574	173	440-449	565
P1	C	-	G	G	A	C	G	A	-	G
P2	C	-	G	A	A	C	-	A	$(AT)_3$	A
P3	C	-	G	G	A	C	G	A	-	A
P4	C	-	G	A	A	C	-	A	-	A
P5	C	-	G	A	A	C	G	A	$(AT)_3$	A
P6	C	-	C	A	A	C	-	A	$(AT)_3$	A
P7	C	-	G	G	A	C	G	C	-	G
P8	C	CCTTTTT	G	A	A	C	G	A	$(AT)_3$	A
P9	C	-	G	A	T	C	G	A	$(AT)_3$	A
P10	C	-	G	A	A	C	G	A	$(AT)_5$	A
P11	T	-	G	G	A	C	G	A	-	A
P12	C	-	G	G	A	T	G	A	-	G
P13	C	-	G	G	A	C	-	A	-	A
P14	C	-	G	G	A	C	G	A	AT	A
P15	C	-	G	A	A	C	G	A	-	G
P16	C	CCTTTTT	G	A	A	C	G	A	-	G
P17	C	-	G	G	A	C	-	A	-	G
P18	C	-	G	G	T	C	G	A	-	G
P19	C	-	C	A	A	C	G	A	$(AT)_3$	A
P20	G	-	G	A	A	G	G	A	$(AT)_2$	A
P21	C	-	G	A	T	C	G	A	-	A

4.2). Among these variable sites, four sites were found in the *trn*L intron, three in the *trn*G intron, two in the *trn*K intron and one in the *psb*K-*trn*S spacer. In particular, seven variable sites were caused by base substitutions and three were the result of indels (Table 4.2).

The distributions of cpDNA haplotypes in the population identification database are shown in Figure 4.1. A total of 21 haplotypes were defined from the 10 variable sites. The most common haplotype, P1, was present in 34.4% of the samples, whereas the second most common, P2, was found in 24.7% of the samples. Both haplotypes displayed nearly disjunct distributions. Haplotype P1 was found widespread from the central to the southern region of Peninsular Malaysia, where it was either found alone or along with haplotype P3. In contrast, haplotypes P2 and P4 were solely found in the northern region of Peninsular Malaysia. In part, haplotypes P1, P2 and P3 were found jointly only in PRaja and RTunggal. Of the 21 haplotypes, 12 haplotypes (P7, P8, P10, P12 and P14–P21) were endemic to specific populations, as shown in Figure 4.1. Based on the cpDNA haplotypes, the results clearly showed that only the northern and southern regions of Peninsular Malaysia were distinguishable. Thirteen populations were grouped under the northern region (BEnggang, Sungkop, BHijau, PTanjung, Piah, Bubu, Chikus, Jeli, GBasur, Lebir, HTrengA, HTrengB and PRaja), while 19 populations were grouped under the southern region (Gombak, SLalang, Ampang, Pelangai, Pasoh, Tersang, RTunggal, BTinggi, Lentang, Lakum, Kemasul, Berkelah, RDaun, Lesong, Labis, LenggorA, LenggorB, PantiA and PantiB).

4.3.2 Test of conformity of origin

The type I error for the southern region of population identification database was estimated by simulating 10,000 artificial wood lot origins from the southern of Peninsular Malaysia against the database of the southern region (Figure 4.2). With 30 samples per lot, 0.3% of wood lots showed non-conformity at 5.0% threshold. In order to estimate the type II error, similar simulations were performed with wood lot origin from the northern provenance tested against the database of the southern region.

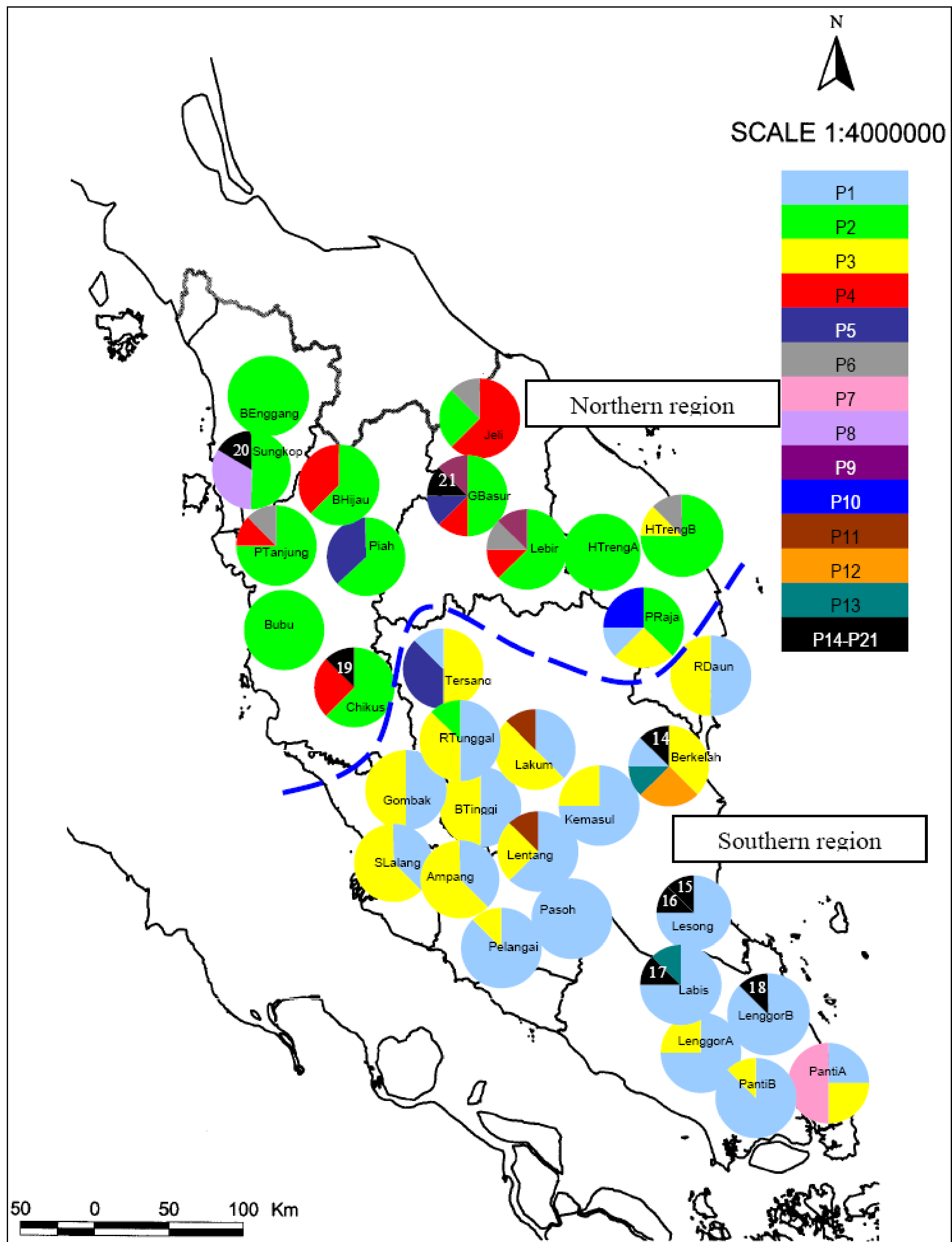

Figure 4.1: Locations of 32 populations of *Neobalanocarpus heimii* from 29 forest reserves of Peninsular Malaysia. The distributions of 13 haplotypes in the population identification database that occurred in more than two individuals are shown in different colours. The remaining eight unique haplotypes that occurred in only one individual are each shown in black.

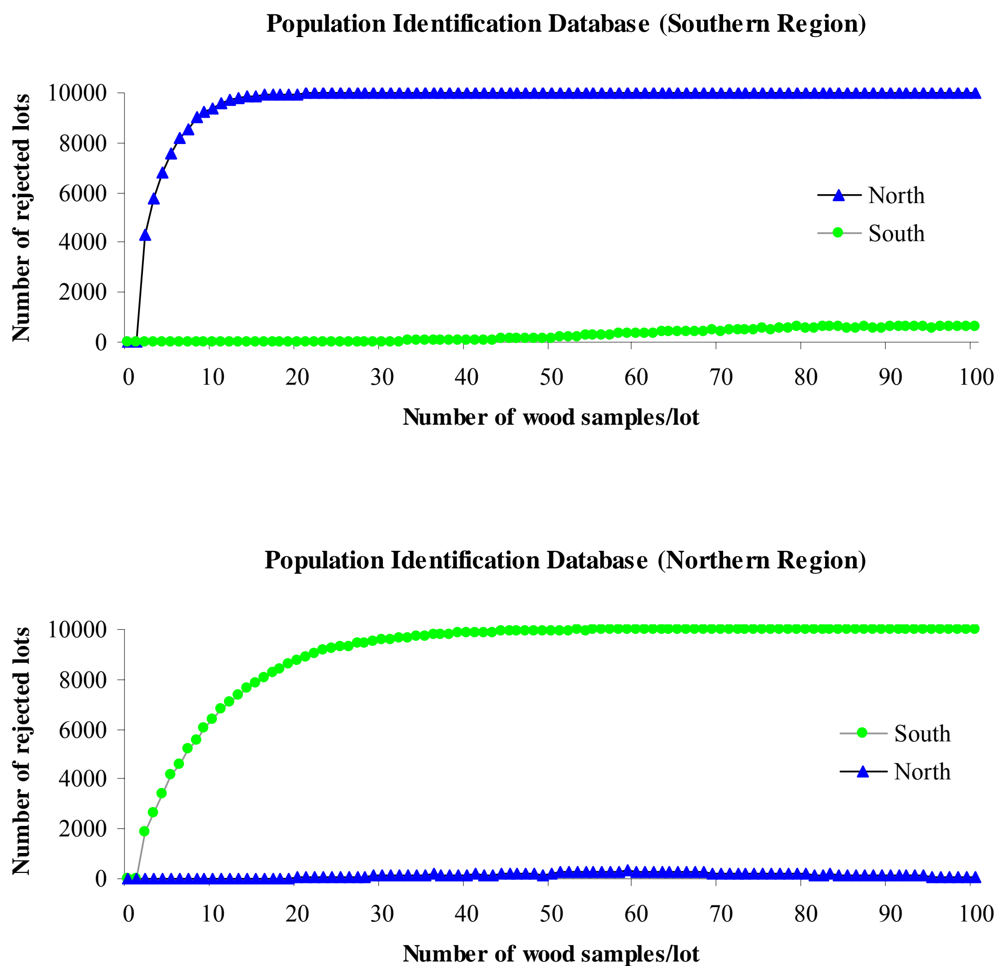

Figure 4.2: Simulation of the number of wood samples/lots considered as non-conformity with the size of lot ranging from 1–100, originating from the southern and northern of Peninsular Malaysia were tested against the southern and northern regions of population identification database.

Likewise, with 30 samples per lot, 100.0% of those from the northern provenance resulted in non-conformity at the 5.0% threshold. Similarly, for the northern region of population identification database, the type I error was estimated by simulating 10,000 wood lots origin from the northern region of Peninsular Malaysia against the database of the northern region (Figure 4.2). With 30 samples per lot, 1.3% of wood lots showed non-conformity at 5.0% threshold. For the type II error, 96.0% of those from the southern provenance resulted in non-conformity at the 5.0% threshold. Overall, the observed types I and II errors resulting from the simulation in population identification database showed good concordance with the predicted 5.0% threshold.

4.4 DISCUSSION

In the context of forest certification and forensic forestry, two important criteria were considered when evaluating appropriate cpDNA region for population identification: (i) significant intraspecific variability and (ii) an appropriately short sequence length so as to facilitate DNA amplification from dry wood. A number of universal primers compiled in cpDNA primer database (Heinze, 2007) were screened in order to search for the most variable sequences; *trn*L intron, *trn*G intron, *trn*K intron and *psb*K-*trn*S spacer, with fragment length 569–679bp have proven to meet the criteria needed for maximum utility (variability and short length). Indeed, the paucity of variable sites in cpDNA usually limits its discriminating power to differentiate between populations; therefore more cpDNA regions were combined to increase the power of discrimination. With regard to sequence length, Deguilloux *et al.* (2002) demonstrated the potentialities to recover cpDNA sequence (ranging from 566 to 1483bp) from dry wood through PCR amplification. An important rationale for using short sequence length resides in the need to amplify highly degraded DNA from dry wood, which is a major technical prerequisite for timber certification and forensic forestry.

A common theme of authenticity testing is the requirement for a comprehensive database, in which a suspect sample can be compared to establish its conformity or

reveal its provenance. To hold promise as a powerful identification tool, the population identification database should be able to specifically partition the whole Peninsular Malaysia until the level of single population or forest reserve. Conversely in this study, at face value, though the databases covered most of the distribution range of *N. heimii* throughout Peninsular Malaysia, notably, only the northern and southern regions of Peninsular Malaysia are clearly distinguishable. As most populations shared the common haplotypes, it is impossible to partition each forest reserve individually. In fact, the boundary of a forest reserve is defined by human, in which the natural populations might not cluster according to human's terminology. The findings clearly demonstrated the database could only be used to determine the unknown origin at the regional level. In addition, it is important to highlight a common limitation that the effort to identify all possible haplotypes is constrained by the limitation to sample every single forest reserve in Peninsular Malaysia. More haplotypes could be observed if more sampling sites were incorporated. Hence, the database needs to be continuously updated with novel cpDNA sequence in order to trace the availability of the rare and unsampled haplotypes.

In terms of application, this database could be applied for traceability in two different circumstances: (1) to verify the provenance of a wood lot in the context of forest certification and chain of custody certification and (2) to identify the potential population of origin of the suspected illegal harvested wood lot. The first circumstance would provide an answer as to whether a wood lot is conformed to a presumed population, whereas the second circumstance would postulate a potential population that the questioned wood lot is belonging to. As a whole, for both circumstances, the database would be primarily used to confirm or reveal either the wood lot originated from the northern or southern provenance. Once the potential source region is determined, by using the individual identification profiling database that was previously developed (Tnah *et al.*, 2010), an assignment test could be conducted based on multilocus genotypes of STR to determine the source of suspected timber until the level of specific population or forest reserve. The assignment test would provide the probability of the questioned wood lot belonging to a population, thus permitting a link

to allocate the wood lot to its potential population of origin. In general, the results from many analytical measurements rely on the strengths of the mathematical methods applied to the data for evaluation. It must be able to withstand legal challenge and the best authenticity testing would not rely on complicated statistical tests that do not provide a definite answer. Therefore, the simple statistical approach used in population identification database and assignment test that focuses only on the composition (presence/absence of haplotypes and multilocus genotypes of STR) of the wood lot would arguably be the most appropriate analytical tool for accurately revealing and verifying the provenance of a wood lot.

In short, this study demonstrates how the DNA authenticity testing can serve as an important technical tool to monitor and verify the legality of the suspected timber. Although the population identification database has limitation to determine the provenance of the suspected timber until the level of specific population or forest reserve, the drawback could be complemented by the application of assignment test using individual identification profiling database. For the moment though, there is a need for population and individual identification databases to be established for premium timber and for these to be used as benchmarks in ongoing scientific developments in the near future. To this end, a foolproof tracking system to verify timber legality is something very much in demand from the market place. However, overcoming the need for these databases and tracking system in all commonly traded tropical timber will be an extremely challenging task.