Forest Ecology and Management 259 (2010) 1436–1446

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Forensic DNA profiling of tropical timber species in Peninsular Malaysia

Lee Hong Tnah [a], Soon Leong Lee [a,*], Kevin Kit Siong Ng [a], Qamaruz-Zaman Faridah [b], Ibrahim Faridah-Hanum [b]

[a] Forest Research Institute Malaysia, 52109 Kepong, Selangor Darul Ehsan, Malaysia
[b] Universiti Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia

ARTICLE INFO

Article history:
Received 7 September 2009
Received in revised form 12 January 2010
Accepted 12 January 2010

Keywords:
Dipterocarpaceae
Forestry forensic
Random match probability
Neobalanocarpus heimii
Short tandem repeats (STRs)

ABSTRACT

Illegal logging poses a significant threat to the sustainability of tropical forest ecosystems. By using *Neobalanocarpus heimii* (Dipterocarpaceae) as an example, the study assessed the feasibility of using short tandem repeats (STRs) as a tool to identify the source of illegally logged timber. Thirty natural populations of *N. heimii* were profiled using 12 STRs to develop the DNA profiling databases. As the cluster analysis divided the 30 populations into three genetic clusters, corresponding to three subregions within Peninsular Malaysia. The DNA databases were characterised at the levels of population, subregion and Peninsular Malaysia. Independence tests within and among loci were violated in all the databases due to significant levels of population differentiation and inbreeding. Thus, the effects of population substructure and inbreeding should be incorporated into the calculation of random match probability. The random match probabilities estimated using subpopulation and subpopulation-cum-inbreeding models were biased in favour of the defendant, whereas the random match probabilities estimated using product rule were biased in favour of the prosecutor. The conservativeness tests showed that the subregion and Peninsular Malaysia databases were conservative, and these databases should be able to provide legal evidence for court proceedings against illegal loggers in Peninsular Malaysia.

1. Introduction

Tropical forest coverage has been declining at an exponential rate and large areas of forest are either being lost to conversion for agriculture or degraded through poor logging practices or illegal logging without regard to sustainability and biodiversity (Asia Forest Partnership, 2005). The illegal logging activities can span the entire commercial timber production chain, from the actual felling of trees without authorisation or without meeting the required operating standards, to the transport, processing and export of products (Smith, 2004). The vast extent of the illegal logging could contribute to increased poverty and land/social conflicts, causes significant losses of tax revenues, and more importantly, poses a significant threat to the sustainability of forest ecosystems (Asia Forest Partnership, 2005).

DNA profiling has long been used in humans for legal proceedings to prove guilt or innocence, resolve unestablished paternity, identify remains of missing persons or victims of mass disasters and establish citizenship by proving blood relationships in immigration laws (Butler, 2005; Kobilinsky et al., 2005). Recently DNA profiling has also been applied for the identification of animals where the issues of endangered species and breeding are significant (Manel et al., 2002; Withler et al., 2004). In forensic botany, samples of plant materials are used to solve criminal and civil cases, and plant DNA profiling have been used as evidence to link the individual on whom the plant material was found to a crime scene (Yoon, 1993; Siver et al., 1994; Craft et al., 2007). In Malaysia, in the case of illegal logging, foresters at the moment have to depend on wood anatomy evidence to link the suspected timber thefts to the source trees. But this is inadequate as identification could only be done on the group of trees and not to the species and individual levels. Thus, in order to establish a linkage between the evidentiary sample and the source, utilisation of DNA profiling evidence to match logs and the stumps from which the timber is believed to have originated would provide valuable evidence in legal cases.

Some progress has already been made towards the development of plant DNA profiling methods for discriminating among varieties and individuals, such as strawberry (Congiu et al., 2000), *Thuja plicata* (White et al., 2000), *Cannabis sativa* (Gilmore et al., 2003; Howard et al., 2009), *Albies alba* (Ziegenhagen et al., 2003) and *Acer rubrum* (Bless et al., 2006). These DNA-based methods included randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and short tandem repeat

* Corresponding author at: Genetic Unit, Forest Research Institute Malaysia (FRIM), Kepong, 52109 Selangor Darul Ehsan, Malaysia. Tel.: +60 3 62797145; fax: +60 3 62804614.
E-mail address: leesl@frim.gov.my (S.L. Lee).

0378-1127/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.foreco.2010.01.017

(STR), which offer the potential to aid the forensic community by enabling the identification of botanical samples and determination of the provenance of seized samples (Yoon, 1993; Becker et al., 1995; Gilmore et al., 2003). The short tandem repeats are superior to RAPD and AFLP for individual identification and individualisation, because in principle alleles and genotypes can be unambiguously assigned, and primer sequences can be easily distributed and shared among different laboratories (Weising et al., 2005).

In DNA testimony, it is necessary to provide an estimate of the weight of the evidence. However, since it is not possible to generate a DNA profile of every individual in a particular population, this weight needs to be assigned by applying population genetic principles and models (Buckleton et al., 2006). There are three possible outcomes of a DNA test: no match, inconclusive, or match between samples examined. However, only the third outcome requires statistics to answer the following question: are they from the same individual or is there something else out there which might just happen to match the evidence in question by chance? Therefore, before a marker system can be introduced into forensic casework, a population database must be established for statistical evaluation of the evidence to extrapolate the probability of a random match, since it would not be scientifically justifiable to speak of a match as proof of identity in the absence of underlying data that permit some reasonable estimate of how rare the matching characteristics actually are (NRC II, 1996).

Typically, the computation of the weight of the evidence is based on the product rule, which relies for its validity on an assumption of Hardy–Weinberg equilibrium (HWE) and linkage equilibrium (LE) at the population level (Buckleton et al., 2006). In a real population, however, genetic mixing of alleles is not completely random because parents often share some common ancestry. The consequence of this non-random mating usually causes a decrease in heterozygotes and an increase in homozygotes (Ayres and Overall, 1999). In such circumstances, the product rule is violated and the estimator will normally give a lower random match probability than the true match probability (Balding and Nichols, 1994; Donnelly, 1995). In addition, objections have been raised on the basis that population substructuring might exist in a real population and might induce both between- and within-locus dependencies (Lander, 1989, 1991). Various statistical models have been proposed to overcome these problems. These include the use of subpopulation model which assumes HWE and LE at the subpopulation level (Balding and Nichols, 1994) and the subpopulation-cum-inbreeding model which incorporates both the inbreeding and coancestry coefficients into the calculation of random match probability (Ayres and Overall, 1999).

By using *Neobalanocarpus heimii* (Dipterocarpaceae) as an example, this study was aimed at utilising STRs for individual identification of important timber species in forensic investigations in forestry. Locally known as chengal, *N. heimii* is endemic and widely distributed in Peninsular Malaysia. It is found in diverse localities, on low-lying flat land as well as on hills up to 900 m (Symington, 1943). It produces heavy and wingless seeds, in which the seed dispersal usually occurs only by gravity (Symington, 1943). Previous studies on *N. heimii* showed that it is a diploid ($2n = 14$; Jong and Lethbridge, 1967), and predominantly an outcrossing species, with outcrossing rates estimated at 87.5–97.9% (Konuma et al., 2000). *N. heimii* produces a naturally, highly durable wood and is among the strongest timbers in the world. It is used for heavy construction, bridges, boats, buildings, and wherever strength is considered essential (Thomas, 1953). Under the International Union for Conservation of Nature (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. Herein, the specific objectives of this study were: (i) to evaluate the genetic structure of *N. heimii* within Peninsular Malaysia and within a population; (ii) to generate a comprehensive DNA profiling database of *N. heimii* throughout Peninsular Malaysia; and (iii) to examine the statistical models for the calculation of random match probability.

2. Methods

2.1. Sample collection and DNA extraction

In order to generate a comprehensive database of *N. heimii* for individual identification, sample collections were conducted throughout the distribution range of *N. heimii* in Peninsular Malaysia. Thirty natural populations of *N. heimii* were collected from 27 forest reserves, and a total of 1081 individuals, that were more than 10 cm diameter at breast height (dbh), were sampled (Table 1). A transect line method was utilised as a guide for the sampling activities for all the populations except Pasoh which was based on a 50-ha ecological plot for intra-population statistical analysis. 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 modification, and further purified using High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH).

2.2. PCR amplifications and electrophoresis condition

Twelve STR loci: *Nhe004*, *Nhe005*, *Nhe011*, *Nhe015* and *Nhe018* (Iwata et al., 2000), *Hbi161* (Lee et al., 2004a), *Shc03*, *Shc04* and *Shc07* (Ujino et al., 1998), *Sle392* and *Sle605* (Lee et al., 2004b) and *Slu044a* (Lee et al., 2006), which show specific amplification, single-locus mode of inheritance, absence of mononucleotide repeat motifs and null alleles were selected for DNA profiling of *N. heimii* (Tnah et al., 2010). PCR amplifications were performed in 10 μl reactions 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).

2.3. Intra-population STR diversity

The Pasoh population, with 252 individuals was subjected to intra-population statistical analysis. Spatial genetic structure within population was analysed based on the number of alleles in common using the program Spatial Genetic Software (SGS) (Degen et al., 2001). The spatial distribution of alleles was tested for 20 continuous distance classes, each of 20 m, from 0–20 to 380–400 m. Significant deviation from random spatial distribution at 95% confidence interval was tested using Monte Carlo simulations (1000 permutations). Genetic relatedness among individuals was quantified using the PowerMarker (Liu and Muse, 2005) based on D_{SA} shared allele distance (Chakraborty and Jin, 1993), with cluster