Phylogeographical Pattern and Evolutionary History of an Important Peninsular Malaysian Timber Species, Neobalanocarpus heimii (Dipterocarpaceae)

LEE H. TNAH, SOON L. LEE, KEVIN K. S. NG, CHAI T. LEE, SUBHA BHASSU, AND ROFINA Y. OTHMAN

From the Forest Research Institute Malaysia, Kepong, Selangor Darul Ehsan 52109, Malaysia (Tnah, Lee, Ng, and Lee); and the Universiti Malaya, Kuala Lumpur 50603, Malaysia (Bhassu and Othman).

Address correspondence to Lee H. Tnah at the address above, or e-mail: leehong@frim.gov.my.

Abstract

Tectonic movements, climatic oscillations, and marine transgressions during the Cenozoic have had a dramatic effect on the biota of the tropical rain forest. This study aims to reveal the phylogeography and evolutionary history of a Peninsular Malaysian endemic tropical timber species, Neobalanocarpus heimii (Dipterocarpaceae). A total of 32 natural populations of N. heimii, with 8 samples from each population were investigated. Fifteen haplotypes were identified from five noncoding chloroplast DNA (cpDNA) regions. Overall, two major genealogical cpDNA lineages of N. heimii were elucidated: a widespread southern and a northern region. The species is predicted to have survived in multiple refugia during climatic oscillations: the northwestern region (R1), the northeastern region (R2), and the southern region (R3). These putative glacial refugia exhibited higher levels of genetic diversity, population differentiation, and the presence of unique haplotypes. Recolonization of refugia R1 and R2 could have first expanded into the northern region and migrated both northeastwards and northwestwards. Meanwhile, recolonization of N. heimii throughout the southern region could have commenced from refugia R3 and migrated toward the northeast and northwest, respectively. The populations of Tersang, Pasir Raja, and Rotan Tunggal exhibited remarkably high haplotype diversity, which could have been the contact zones that have received an admixture of gene pools from the northerly and also southerly regions. As a whole, the populations of N. heimii derived from glacial refugia and contact zones should be considered in the conservation strategies in order to safeguard the long-term survival of the species.

Key words: chloroplast DNA, Dipterocarpaceae, glacial refugia, Pleistocene, postglacial recolonization routes, tropical rain forest

Phylogeography was first coined in 1987 to study the principles and processes governing the geographical distribution of genealogical lineages, especially those within and among closely related species (Avise 2000). Since then, numerous phylogeographic studies have provided considerable information on the nature and locations of glacial refugia, post-glacial recolonization routes, and contemporary distribution of genetic diversity in formerly glaciated versus refugia areas (Csaki et al. 2002; Petit et al. 2003; Cheng et al. 2005; Ikeda and Setoguchi 2007; Falehi et al. 2009; Nevill et al. 2010; Saeki et al. 2011). Today, phylogeographic studies can be used to infer the evolutionary history of a species and this knowledge could have important implications for predicting current and future periods of global climate change (Provan and Bennett 2008). Also, elucidating the evolutionary history of biota is crucial to disentangle contemporary from past demographic events and identify key regions deserving priority for conservation.

Tropical rain forests are the tall, dense, evergreen forests that form the natural vegetation cover of the wet tropics, where the climate is always hot and the dry season is short or absent (Primack and Corlett 2005). Most of the land masses
that currently support tropical rain forests have a common origin in the ancient southern supercontinent of Gondwana (Morley 2000). Southeast Asia covers the third largest block of tropical rain forest on Earth and has been recognized as a major global biodiversity hotspot (Mittermeier et al. 2005). Most Southeast Asian rain forests can be characterized as “dipterocarp forests” because they are dominated by large trees of the family Dipterocarpaceae (Primack and Corlett 2005).

During the era of Cenozoic, Southeast Asia witnessed active tectonism (collisions between India and Eurasia about 50–65 Ma; Southeast Asia and Australia about 15 Ma) and dynamic climate changes (cooler and drier periods) (Hutchison 1989; Hall 1998). Everwet rain forests were rare in Southeast Asia during the Oligocene and early Miocene, but when the climate became perhumid during the middle Miocene, the rain forests spread across the region (Morley 2000). However, during the Pleistocene epoch, one of the greatest environmental influences shaping the rain forest communities has been the numerous cycles of glacial and interglacial periods. As palynological, geological, fossil and biogeographical data, plant and animal distributions all provide evidence of climate change, which considerably affected the dynamic ecosystem of Southeast Asia; the tropical rain forests are thought to have contracted into a few glacial refugia (Medway 1972; Kaas and van der Dam 1995; Verstappen 1997; Morley 2000; Thomas 2000).

During the Last Glacial Maximum (LGM) of the Pleistocene (18 kyr before the present), the temperature in Southeast Asia was 3–7 °C lower than the present (Morley 2000). The glacio-eustatic depression of sea level by approximately 120 m had fully exposed the Sunda shelf (the part of the Asian continental shelf that was exposed during the last ice age, which encompasses the Malay Peninsula and the large islands of Borneo, Sumatra, Java, and Palawan). The expansion of the Sunda shelf had prevented the winter monsoon from picking up moisture from the South China Sea, which probably led to increased drought and seasonality in the central part of the region (Gathorne-Hardy et al. 2002). The arid conditions are thought to have changed the vegetation of Southeast Asia, causing it to become a mixture of savannas and patchy deciduous forests, driving rain forest obligates to a few refugia (Brandon-Jones 1998). As the cooling periods ended, the sea rose and some of the surviving rain forests expanded and again colonized the region (Kaas and van der Dam 1995; Morley 2000). Each of the glacial episodes probably showed a similar pattern and caused the repeated expansion and contraction of the rain forests (Woodruff 2003).

Peninsular Malaysia lies in the Southeast Asian region, which existed when terrains rifted from the eastern margin of Gondwana collided with Eurasia in the Jurassic (Woodruff 2003). During most of the Tertiary, it was probably geographically the same as it is today, with a spine dominated by a north–south range of granitic mountains, 100–200 km across, over 2000 m, and surrounded by sedimentary lowlands (Wyatt-Smith and Panton 1995; Woodruff 2003). During Quaternary, changes of landscape in the region are largely attributable to eustatic sea level fluctuations (Ashton 1972). Particularly, with the advent of Pleistocene glaciations, the area was probably all vegetated by savannas or wooded savannas (Heaney 1991; Verstappen 1997; Morley 2000; Thomas 2000). Though most evidence suggested that rain forests disappeared almost entirely from Peninsular Malaysia, there are a few studies that suggested continuity of moist climates and persistence of small rain forest refugia in coastal regions of the south, east, and west of Peninsular Malaysia (Corner 1978; Geyh et al. 1979; Emmel and Curray 1982; Quack et al. 2007).

In most angiosperms, chloroplast DNA (cpDNA) is thought to evolve slowly, with low mutation rates and often maternally inherited (Wolfe et al. 1987; Corriveau and Coleman 1988; Clegg and Zurawski 1992). The growing numbers of cpDNA universal primers published in the past decades (Taberlet et al. 1991; Demesure et al. 1995; Weising and Gardner 1999; Grivet et al. 2001; Heinze 2007) have facilitated the inferences about the evolutionary history of many plant species worldwide. Particularly, these markers have been successfully applied to identify possible glacial refugia and postglacial migration routes of plant species from the temperate forest (Comes and Kadereit 1998; Taberlet et al. 1998; Cheng et al. 2005; Shepherd et al. 2007), but comparatively few studies have been dedicated to the tropical species (Chiang et al. 2001; Cannon and Manos 2003; Bänfer et al. 2006).

*Neobalanocarpus heimii*, or locally known as chengal, 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 a naturally, highly durable wood and is among the strongest timbers in the world (Thomas 1953). The species produces heavy and wingless seeds, whereby the seed dispersal usually occurs only by gravity (Symington 1943). A study of *N. heimii* in Peninsular Malaysia using nuclear microsatellites (Tnah et al. 2010) demonstrated moderate genetic differentiation among populations (*FST* = 0.127). The main flower visitors are *Trigona* spp. and *Apis* spp. (Appanah 1985, 1987). Previous studies on *N. heimii* showed that it is a diploid (*2n = 14; Jong and Lethbridge 1967) species, predominantly outcrossing, with outcrossing rates estimated at 87.5–97.9% (Konuma et al. 2000). Under the IUCN Red List of Threatened Species, it was classified as Vulnerable 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). By using *N. heimii* as a model species, this study was initiated to reveal the evolutionary history of the species in Peninsular Malaysia, including the potential glacial refugia and their postglacial recolonization routes.

**Materials and Methods**

**Sample Collection and DNA Extraction**

In the aforementioned microsatellite analysis of *N. heimii* (Tnah et al. 2010), 32 natural populations of *N. heimii*, with the sample size of 12–252 individuals were collected throughout the distribution range of *N. heimii* in Peninsular Malaysia. For
Table 1  Sampling localities chosen for the populations of Neobalanocarpus heimii in Peninsular Malaysia accompanied with the population codes

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

Eight samples per population were used in this study.

In this study, eight samples from each population with diameter at breast height more than 10 cm were investigated (Table 1). The samples were collected 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, Penzberg, Germany).

PCR Amplifications and Sequencing

In order to detect intraspecific variability, 27 universal primer pairs of cpDNA (Heinze 2007) were screened through eight individuals of N. heimii, which were selected from eight different populations. Five noncoding regions of cpDNA were found to be informative in this study: trnL intron (Taberlet et al. 1991), trnS–trnG intergenic spacer (Hamilton 1999), trnG intron (Heinze 2007), trnK intron (Demesure et al. 1995; Weising and Gardner 1999), and psbK–trnF intergenic spacer (Grivet et al. 2001; Heinze 2007). PCR amplifications were performed in 20-µL 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 MgCl₂, 0.3 µM of each primer, 0.2 mM of each dNTP, and 1 U of Taq DNA polymerase (Promega, Madison, WI). The reaction mixture was subjected to amplification using a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA), 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 GmbH, Hilden, Germany) 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 CODEON CODE Aligner version 2.0 (CODEON CODE Corporation, Dedham, MA).

Statistical Analyses

Haplotypes were determined solely based on substitution intraspecific variable sites. Insertion and deletion (indel) variable sites were excluded from the analyses, as different evolutionary rates were found between the nucleotide substitutions and indels (Hamilton et al. 2003). Haplotype diversity,
nucleotide diversity (Nei 1987), and Tajima’s D (Tajima 1989) for the departure from neutrality were estimated using DnaSP version 4.0 (Rozas et al. 2003) based on the total number of segregating sites.

Hierarchical analysis of molecular variance (AMOVA) was evaluated for two different grouping schemes: 1) the whole 32 populations of Peninsular Malaysia under this study and 2) two major genealogical cpDNA lineages (northern region: populations 1–13 and southern region: populations 14–32; Table 1) using Arlequin version 3.1 (Excoffier et al. 2005). In this case, the partition of Peninsular Malaysia into two genealogical cpDNA lineages was determined based on the haplotype distribution pattern.

Likewise, two measures of diversity and population differentiation were computed for the two different grouping schemes using the PERMUTE and CRSSR version 2.0 as described in Pons and Petit (1996) and Burban et al. (1999). The parameters included the mean within-population gene diversity (Hₒ), the total gene diversity (Hₚ), and the coefficient of genetic differentiation over all populations (Gₜ), as well as the other equivalent parameters (Vₛ, Vₚ, and Nₛ), obtained by taking into account similarities between haplotypes. Gₛ depends only on the frequencies of the haplotypes, whereas Nₛ is influenced by both haplotype frequencies and the distances between haplotypes. In order to test the null hypothesis of no phylogeographical component to the genetic structuring, the significant difference between Gₛ and Nₛ was tested via 10,000 permutation tests.

A Mantel test was also performed for the two different grouping schemes using Arlequin version 3.1. A correlation coefficient (r) and a one-tailed P value were determined for a positive relationship between pairwise population differentiation (Fₛ) and geographical distance. Significance was tested with 10,000 random permutations. Contribution of each population to the total diversity measured by the haplotype richness (CTR) was calculated according to Petit et al. (1998), using the Contrib version 1.02. The contribution was quantified in terms of contribution due to its own diversity and its differentiation from the remaining populations.

A network of haplotypes was constructed in Network 4.6.0.0 (available at http://fluxus-engineering.com) using the median-joining network algorithm (Bandelt et al. 1995) and maximum parsimony calculation (Polzin and Daneschmand 2003). The network was built based on the genetic distance between two sequences in a given data set. In order to determine an ancestral haplotype, the network was rooted by including a sequence of Hopea mengaranwan as an outgroup.

### Results

#### cpDNA Variability and Genetic Diversity

The examined sequences consisted of trnL intron (584–591 bp), trnS–trnG intergenic spacer (575–585 bp), trnG intron (660–661 bp), trnK intron (569–579 bp), and psbK–trnS intergenic spacer (679 bp). The corresponding GenBank accession numbers are EU918738–EU918763. The combined sequences of these five chloroplast noncoding regions resulted in a total of 3095 bp. Indel polymorphic sites were then removed and resulted in an aligned length of 3069 bp. In total, 10 intraspecific variable sites due to nucleotide substitutions were detected within these five noncoding regions (Table 2). Among these variable sites, three were found in each of the trnL intron and trnS–trnG intergenic spacer, two in the trnG intron, and one each in the trnK intron and psbK–trnS intergenic spacer.

The haplotype and nucleotide diversity measures for all the 32 populations are shown in Table 3. In total, 15 haplotypes were detected from the 32 populations. The total haplotype diversity was 0.749 and the nucleotide diversity per site was 0.00041. In terms of the number of haplotypes, Sungkop, Lakum, Tersang, and PRaja were the most frequent.
variable populations with four different haplotypes, whereas Bubu and Pasoh were the most homogenous populations, consisting only one haplotype. Sungkop also had the highest values of haplotype diversity ($h = 0.821$) and nucleotide diversity $(\pi = 0.00044)$. Other populations that had high levels of diversity were Lakum ($h = 0.786; \pi = 0.00034$), Tersang ($h = 0.714; \pi = 0.00040$), PRaja ($h = 0.750; \pi = 0.00034$), PantiA ($h = 0.714; \pi = 0.00033$), and GBasur ($h = 0.714; \pi = 0.00028$). The Tajima's $D$ neutrality tests showed that the observed values did not significantly $(P > 0.05)$ deviate from the expected values (Table 3).

### Genetic Differentiation

The AMOVA revealed that $56.16\%$ of observed variation was due to differences among populations and $43.84\%$ within population (Table 4, all partitions were significant at $P < 0.05$). This measure partly reflects moderate dispersal ability of the species, although long-term range fragmentation might also play a role. When the populations were grouped based on two regions, the AMOVA revealed that $65.55\%$ of the variation was apportioned between the northern and southern regions of Peninsular Malaysia, $5.04\%$ among populations within the regions, and $29.41\%$ within populations (all partitions were significant at $P < 0.05$).

Estimates of diversity and differentiation based on all $32$ populations in Peninsular Malaysia revealed that $N_{ST}$ was significantly higher than $G_{ST}$ $(P < 0.05)$, indicating that the null hypothesis of no phylogeographical component to the genetic structuring $(G_{ST} = N_{ST})$ could be rejected (Table 5). A higher value of $N_{ST}$ than $G_{ST}$ might indicate the presence of phylogeographic structure, in which closely related haplotypes were more often found in the same area than less closely related haplotypes (Pons and Petit 1996). This demonstrated either low levels of recent gene flow between populations, or common ancestry within these $32$ populations. The $G_{ST}$ and the $N_{ST}$ values of $0.390$ and $0.562$, respectively, might suggest moderate subdivision of cpDNA diversity among populations. When populations were grouped based on the southern and northern regions, southern populations $(G_{ST} = 0.151, H_{S} = 0.571, H_{I} = 0.485)$ exhibited greater differentiation and total diversity than northern populations $(G_{ST} = 0.071, H_{S} = 0.461, H_{I} = 0.429)$. Based solely on the southern region, the geographical distribution of haplotype variation in $N. heimii$ was clearly not random $(N_{ST} > G_{ST}, P < 0.05)$. For the northern region, on the contrary, the differences of $N_{ST} – G_{ST}$ were not significant. Such observation might indicate a complex history of colonization with long-term differentiated gene pools that spread over large or disjunct areas in the northern region.
In order to assess isolation by distance, correlation between pairwise $F_{ST}$ and geographical distance was checked using the Mantel test. The test indicated a significantly positive relationship between $F_{ST}$ and geographical distances for the whole 32 populations in Peninsular Malaysia ($r = 0.619, P < 0.05$) and also the southern region ($r = 0.288, P < 0.05$). However, no correlation was found for the northern region ($r = -0.027, P > 0.05$). In terms of contribution of each population to the total diversity using the haplotype richness, Sungkop, PantiA, GBasur, and PTanjung contributed most (in terms of diversity and differentiation) to the total diversity (Figure 1). The populations of Sungkop, PTanjung, GBasur, and PantiA contributed most to the differentiation component, whereas those of Sungkop, PRaja, Tersang, and Lakum contributed most to the diversity component of total diversity.

### Haplotype Distribution and Relationship Inferred from the Network of Haplotypes

The distribution of the cpDNA haplotypes throughout Peninsular Malaysia is shown in Figure 2. The most common haplotypes, h1, h2, and h3 (with 34.8%, 30.1%, and 19.9% occurrences, respectively), also had the widest distributions. Notably, eight unique haplotypes were endemic to one or two specific populations: h8 (PantiA), h9 (GBasur and Lebir), h10 (Berkelah), h11 (Lesong), h12 (Lakum and Lentang), h13 (Lakum), h14 (LenggorB), and h15 (Sungkop). Overall, haplotypes h2, h4–h7, h9, and h15 were distributed across the northern region of Peninsular Malaysia, whereas haplotypes h1, h3, h8, and h10–h14 were scattered in the southern region.

The relationships between haplotypes are shown in the maximum parsimony network in Figure 3. Overall, the statistical parsimony procedure revealed that all the cpDNA haplotypes are closely related, most differing from their closest relative by 1–2 mutational steps. The maximum number of

---

**Table 4** Hierarchical AMOVA of *N. heimii* based on two different grouping schemes

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>df</th>
<th>Sum of squares</th>
<th>Variance component</th>
<th>Percentage of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 32 populations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>31</td>
<td>97.297</td>
<td>0.35745</td>
<td>56.16</td>
</tr>
<tr>
<td>Within populations</td>
<td>224</td>
<td>62.500</td>
<td>0.27902</td>
<td>43.84</td>
</tr>
<tr>
<td>Total</td>
<td>255</td>
<td>159.797</td>
<td>0.63647</td>
<td></td>
</tr>
<tr>
<td>Partitioned into two regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>1</td>
<td>77.455</td>
<td>0.62181</td>
<td>65.55</td>
</tr>
<tr>
<td>Within populations</td>
<td>224</td>
<td>62.500</td>
<td>0.27902</td>
<td>29.41</td>
</tr>
<tr>
<td>Total</td>
<td>255</td>
<td>159.797</td>
<td>0.94862</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: df, degree of freedom.

**Table 5** Population structures of *N. heimii* for the northern region, southern region, and the whole Peninsular Malaysia

<table>
<thead>
<tr>
<th>Diversity parameters</th>
<th>Northern</th>
<th>Southern</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_{S}$</td>
<td>0.429</td>
<td>0.485</td>
<td>0.462</td>
</tr>
<tr>
<td>$H_{T}$</td>
<td>0.461</td>
<td>0.571</td>
<td>0.757</td>
</tr>
<tr>
<td>$G_{ST}$</td>
<td>0.071</td>
<td>0.151</td>
<td>0.390</td>
</tr>
<tr>
<td>$V_{S}$</td>
<td>0.425</td>
<td>0.469</td>
<td>0.334</td>
</tr>
<tr>
<td>$V_{T}$</td>
<td>0.462</td>
<td>0.572</td>
<td>0.761</td>
</tr>
<tr>
<td>$N_{ST}$</td>
<td>0.080</td>
<td>0.181</td>
<td>0.562</td>
</tr>
<tr>
<td>$N_{ST-\hat{G}_{ST}}$</td>
<td>0.009</td>
<td>0.030</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Abbreviations: NC, not calculated.

Values in parenthesis are standard deviations.

*Indicated significantly different from zero at the $P < 0.05$ level.

In order to assess isolation by distance, correlation between pairwise $F_{ST}$ and geographical distance was checked using the Mantel test. The test indicated a significantly positive relationship between $F_{ST}$ and geographical distances for the whole 32 populations in Peninsular Malaysia ($r = 0.619, P < 0.05$) and also the southern region ($r = 0.288, P < 0.05$). However, no correlation was found for the northern region ($r = -0.027, P > 0.05$). In terms of contribution of each population to the total diversity using the haplotype richness, Sungkop, PantiA, GBasur, and PTanjung contributed most (in terms of diversity and differentiation) to the total diversity (Figure 1). The populations of Sungkop, PTanjung, GBasur, and PantiA contributed most to the differentiation component, whereas those of Sungkop, PRaja, Tersang, and Lakum contributed most to the diversity component of total diversity.

---

**Figure 1.** Contribution of each population of *Neobalanocarpus heimii* to the total diversity, as described by the haplotype richness (CTR). Circles represent total diversity; solid and open bars represent contributions of differentiation and diversity to the total diversity, respectively.
mutation steps between haplotypes of *N. heimii* in the network is five. From the rooted network tree, haplotype h2 was identified as the ancestral haplotype. Also, following the coalescent theory, haplotype h2 was more likely to be the ancestral haplotype because it was located in the central part of a network, had the most linkages to other haplotypes, had higher frequencies than tip haplotypes, and showed broader geographical distributions. Overall, the haplotype tree was partitioned into two separate networks for all the cpDNA haplotypes identified. In accordance with the haplotype distribution map in Figure 2, haplotypes clustered under the northern group (h2, h4–h7, h9, and h15) were distributed across the northern region of Peninsular Malaysia, whereas haplotypes clustered under the southern group (h1, h3, h8, h10–h14) were spread across the southern region. This further supports the view that in terms of haplotype relationships, basically Peninsular Malaysia can be divided into two main regions: the northern region consisting of 13 populations (BEnggang, Sungkop, BHijau, Piah, PTanjung, Bubu, Chikus, Jeli, GBasur, Lebir, HTrengA, HTrengB, and PRaja) and the southern region comprising 19 populations (RDaun, Berkelah, Tersang, RTunggal, Lakum, BTinggi, Lentang, Lebir, HTrengA, HTrengB, Piah, and Pijiang).
Kemasul, Lesong, Gombak, Ampang, SLalang, Pelangai, Pasoh, Labis, LenggorA, LenggorB, PantiA, and PantiB. It is interesting to note also that PRaja and Tersang, which are located in the border of these two main regions, comprise a mixture of haplotypes (h1, h2, and h3) from both of the northern and southern regions.

Discussion

Hypothetical Phylogeographical History of *N. heimii* before LGM

During the middle and late Eocene, dipterocarps were first dispersed from Indian plate into Southeast Asia, as indicated by geochemical fossils, palynological and biogeography data (Ashton 1982; Ashton and Gunatileke 1987, Aarssen et al. 1990). In the earliest Miocene, the common occurrence of small tricolplate pollen in palynomorph assemblages, which is comparable to that of the dipterocarps *Shorea* and *Hopea*, might suggest extreme widespread of low-diversity dipterocarp monsoon forest (Morley 2000). However, the main radiation of dipterocarps into the everwet rain forests took place after 20 Ma (Ashton 1982). Since this period, the family Dipterocarpaceae has probably became a major component of the Southeast Asian rain forest (Morley 2000).

Molecular phylogenetic analysis of Dipterocarpaceae revealed close affinities among *Shorea*, *Hopea*, and *Neobalanocarpus* (Kajita et al. 1998; Kamiya et al. 2005; Tsumura et al. 2011). *N. heimii* is monotypic and has been postulated to be derived via hybridization between the ancestral lineages of *Shorea* and *Hopea* (Ashton 1982; Kamiya et al. 2005). Evidence from Neogene palynomorph assemblages suggests that the diversification of Southeast Asian flora continues through a major part of the Neogene (Weerd and van der Armin 1992); the species is, therefore, believed to be derived via hybridization during this period. *N. heimii* is solely endemic within Peninsular Malaysia; hence, the limited geographical distribution and relatively low number of missing/extinct haplotype appearing in the network analysis suggest that it is a rather young species. Nevertheless, the exact time for the existence of *N. heimii* remains unclear, until the discovery of new fossil evidence.

In term of Asian Tertiary migration tracks, van Steenis (1936) defined migration tracks from Asia, but more specifically, Morley (1998) suggested that the Malay Peninsula provided the main dispersal route from Asia during the Tertiary. In fact, based on the oldest geochemical fossils of dipterocarps found in Myanmar (Aarssen et al. 1990), the dipterocarps are likely to have dispersed from Myanmar to Thailand and subsequently southward to Peninsular Malaysia. Given that the ancestral haplotype, h2 is found solely in the northern region of Peninsular Malaysia, the earliest migration route of the ancestral populations of *N. heimii* is, therefore, predicted to have begun from the northern region and expanded southwards. Indeed, a star-like haplotype network, with all descendants originating from the ancestral haplotype h2 (Figure 3), indicates a scenario of range expansion across the region. Over time, specifically, both northern and southern regions could have gone through a stochastic lineage sorting and have had independent evolutionary histories for a relatively long period of time. As all haplotypes within the northern region are genetically more similar to each other than the haplotypes that are found in the southern region, it is likely that both regions have become reciprocal monophyly. Palaeoenvironmental phenomenon, for instance marine transgressions during Miocene or Pliocene, will possibly account for such divergence. Hutchison (1989) estimated the maximal Miocene and Pliocene sea levels in Southeast Asia at +220 and +140 m, respectively. Also, due to the highstand in early Pleistocene (+20 m), southeastern of Johor and Riau Archipelago have suffered prolonged isolation, partial submergence, and extensive marine erosion (Ashton 1972). By taking marine transgressions into consideration, much of the southern part of peninsula would have been submerged under shallow seas and forest habitat would have been greatly reduced compared with the more mountainous area in the north. As Miocene and Pliocene highstands in Southeast...
Asia are suggested to have played a significant role in shaping the biogeography transition on the Thai–Malay Peninsula (Woodruff 2003), we therefore believe that the marine transgressions would have caused the transitions between northern and southern populations of *N. heimii*.

Refugia and Postglacial Regions Recognized after LGM

During the advent of the Pleistocene glacial, the widespread populations were fragmented and restricted to a few refugia in Peninsular Malaysia, whereas as the perhumid conditions returned during Holocene warming, the rain forests would have again expanded out of these refugia. In this study, the cpDNA data strongly reveal glacial refugia and postglacial regions in Peninsular Malaysia. In some regions, populations were highly differentiated (e.g., Sungkop), whereas other adjacent populations were fixed for solely one haplotype (e.g., Pasoh). According to the basic expansion and contraction model (Hewitt 1996), the potential refugia and postglacial regions can be predicted based on the levels and patterns of genetic diversity. First, refugia areas should harbor high levels of genetic diversity, as a population that had persisted throughout the glacial period would have a longer demographic history than a population that had evolved during the postglacial period (Comes and Kadeireit 1998; Taberlet et al. 1998). Second, the long-term isolation of refugia sites will lead to genetic differentiation due to drift. Third, the refugia tend to harbor a high number of unique haplotypes. By considering these important genetic criteria for refugia identification, in the case of *N. heimii*, the species is postulated to have survived in the multiple refugia throughout Peninsular Malaysia (Figure 4): the northwestern region (R1: Sungkop), the northeastern region (R2: GBasur), and the southern region (R3: PantiA).

The putative glacial refugia of R1, R2, and R3 exhibited higher levels of genetic diversity, population differentiation, and the presence of unique haplotypes. These regions are located at some of the most stable areas in Malay Peninsula (in terms of tectonic stability and continuous emergence), whereby northern and eastern of Peninsular Malaysia, and also southeastern of Johor have experienced gradual uplift during the Tertiary of as much as 1000 m (Ashton 1972). Particularly, the rain forests in the northwestern (R1) and southern (R3) regions of Peninsular Malaysia could be of great antiquity, as these regions fall under the special floristic areas with high species diversity and endemism for many plant taxa (Corner 1960; Ashton 1992). Another putative glacial refugium, R2 is located at the mountainous area, with the elevation of 943 m. With the advantage of elevation, the plants would need short altitudinal shifts across the intense orography to find suitable niches following climate change (Neito Feltiner 2011). Herein, we believe that the refugium could have sufficient moisture to remain as rain forest during drier climatic phases of the Pleistocene.

Low total diversity in terms of diversity and differentiation (Figure 1) and fixation of haplotypes h1 and h3 in most of the southwestern populations suggest that these areas were probably recolonized after LGM. This finding is consistent with the palynological evidence found in the Klang Valley (western region of Peninsular Malaysia), suggesting the loss of lowland rain forest, and that the area was likely to be the grass-*Pinus* savanna during middle Pleistocene (Morley and Flenley 1987; Morley 1998). Additionally, this might be attributed to founder effects of genetic drift in these westerly populations.

Postglacial Recolonization Routes

A putative recolonization route for *N. heimii* throughout Peninsular Malaysia after LGM was sketched based on the level of genetic diversity, distribution pattern of haplotypes, and restricted dispersal of the species (Figure 4). Taken together, by looking at the northern region, Mantel test and $N_{ST}$−$G_{ST}$ indicated no correlation between $F_{ST}$ and geographical distances, and no phylogeographical structure, respectively. Hence, the northern region could be confounded by complex history of colonization from differentiated gene pools. Thus, the recolonization from refugia R1 and R2 indeed may have first expanded into the northern region of Peninsular Malaysia during the interglacial retreat and stopped at the central regions of Pahang, Terengganu, and Perak. These stocks might have migrated both northeastwards and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure4.png}
\caption{Distribution of potential refugia sites (R1, R2, and R3) inferred for *N. heimii*. Regions in black indicate mountain ranges and arrows indicate putative recolonization routes from R1, R2, and R3.}
\end{figure}
northwestwards after the climatic amelioration. In contrast, for the southern region, Mantel test and $N_{ST}-C_{ST}$ revealed positive correlation between $F_{ST}$ and geographical distances and the presence of phylogeographical structure. In line with these results, the recolonization of *N. heimii* throughout the southern region could have initially commenced from refugia R3 and migrated toward the northeast and northwest, respectively. The expansion process is clearly shown by the postglacial regions, which typically had lower genetic diversity and occupied larger geographic areas.

The topography of Peninsular Malaysia is characterized by a mountainous spine known as the Main Range or Titiwangsa mountain range running from the north (Thailand border) southwards to Negeri Sembilan, which effectively divides Peninsular Malaysia into east and west regions (Anon. 2005). The existence of this natural divider could serve as a geographical barrier between the eastern and western populations of *N. heimii*. However, the geographical distribution of haplotypes indicated the presence of haplotypes h1 and h3 on both sides of the mountain range, and it seems Peninsular Malaysia was partitioned into north–south rather than east–west orientation. In fact, this old and relatively stable mountain range was formed during the Jurassic period by the intrusion of the granite magma and consequent uplift of the Earth’s crust in the regions (Wyatt-Smith and Panton 1995). However, much of the present rain forests colonized Southeast Asia during the middle Miocene (Morley 2000). It would be unlikely for these lowland dipterocarps with limited seed dispersal to have crossed over this mountain range. Therefore, we postulate that the species could have spread over a large part of the southeast and southwest regions of Peninsular Malaysia from refugia R3.

The unraveling of postglacial recolonization routes can be further used to identify contact zones where populations from separate refugia have met. Populations of Tersang ($h = 0.750$), PRaja ($h = 0.750$), and RTunggal ($h = 0.679$) are shown to exhibit remarkably high haplotype diversity, which could be the contact zones of *N. heimii* in Peninsular Malaysia. These populations comprised both haplotypes that were endemic to the northern and also southern regions. Specifically, the contact zones with increased genetic diversity would be achieved mostly through the redistribution of the genetic information already present among populations in refugia (Petit et al. 2003). In another study, high genetic diversity in nonrefugia population has also been observed in *Castanopsis carlesii*, which showed the highest diversity due to the natural postglacial recolonization from different refugia (Cheng et al. 2005).

**Implications for Conservation**

Tectonic movement, climatic oscillation, and marine transgression during the Cenozoic have had dramatic effect on the biota of the tropical rain forest, and research addressed specifically to phylogeography would provide extremely useful information on the evolutionary history and responses of the biota to the climate change. In forest ecosystem, specifically, the genetic issues need to be considered when designing means to minimize the impacts of climate change (Frankham 2010), where it is extremely important to study how a timber species is able to adapt or migrate during the climate oscillation. However, these topics are poorly documented and little discussed. Furthermore, the recognition of the refugia sites and the recovery from prehistoric disturbance are still inconclusive. These highlight the long-term fragility of the forest ecosystem and the importance of conserving prehistoric refugia as modern forest refugia (Gathorne-Hardy et al. 2002).

In this regard, understanding the past history of extant populations is of utmost importance when developing sound conservation policies or sustainable management strategies. The putative refugia sites of *N. heimii*, including the north-western region (R1: Sungkop), the northeastern region (R2: GBasur), and the southern region (R3: PantiA), which harbor a high proportion of unique haplotypes, should be prioritized for long-term management. Besides, the total genetic diversity of a species is another key factor in conservation considerations, in which the maintenance of genetic diversity is critical for long-term survival of a species (Frankel and Soule 1981; Rauch and Bar-Yam 2005). Therefore, the contact zones of *N. heimii* (Tersang, PRaja, and RTunggal) with remarkably high haplotype diversity should also be considered in the conservation strategies in order to safeguard long-time survival of this species.

**Acknowledgments**

The Forest Departments of Kedah, Perlak, Selangor, Negeri Sembilan, Johor, Pahang, Terengganu, and Kelantan are acknowledged for granting us permission to access the forest reserves. We thank the District Forest Officers and the staffs of the Ranger Offices who provided assistance and logistic support during the field trips; and Matam Din, Ghazali Jaafar, Yahya Marbani, Ramli Ponyoh, Sharifah Talib, Suryani Che Seman, Nurul Hudaimi Mamat, and Nor Salwah Abdul Wahid for their excellent assistance in the laboratory and the field. Special thanks are also due to Dr Yoshihiko Tsumura for his helpful comments. This project was supported in part by the e-Science Research Grant (02-03-10-SF0009) entitled “Development of DNA barcode of *N. heimii* (chengal) as a tool for forensics and chain of custody certification” and the Bioversity International Agreement No. APO 05/016.

**References**


Received January 11, 2012; Revised May 21, 2012; Accepted August 3, 2012

Corresponding Editor: David B. Wagner