CHAPTER 3

PHYLOGEOGRAPHICAL PATTERN AND EVOLUTIONARY HISTORY OF AN IMPORTANT PENINSULAR MALAYSIA TIMBER SPECIES, *NEOBALANOCARPUS HEIMII* (DIPTEROCARPACEAE)

3.1 INTRODUCTION

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 (Csaikl *et al*., 2002; Cheng *et al*., 2005; Fjellheim *et al*., 2006; Ikeda & Setoguchi, 2007). Elucidating the evolutionary history of biota is crucial to disentangle contemporary from past demographic events and identify key regions deserving priority for conservation. Today, greater knowledge on phylogeography studies in addressing the past events would also have important implications for predicting current and future periods of global climate change (Provan & Bennett, 2008).

Southeast Asia represents one of the largest evergreen tropical rain forests on Earth. 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) (Figure 3.1a). During Pleistocene, palynological, geological, fossil and biogeographical data, plant and animal distributions all provide evidence of climate change, which considerably affected the

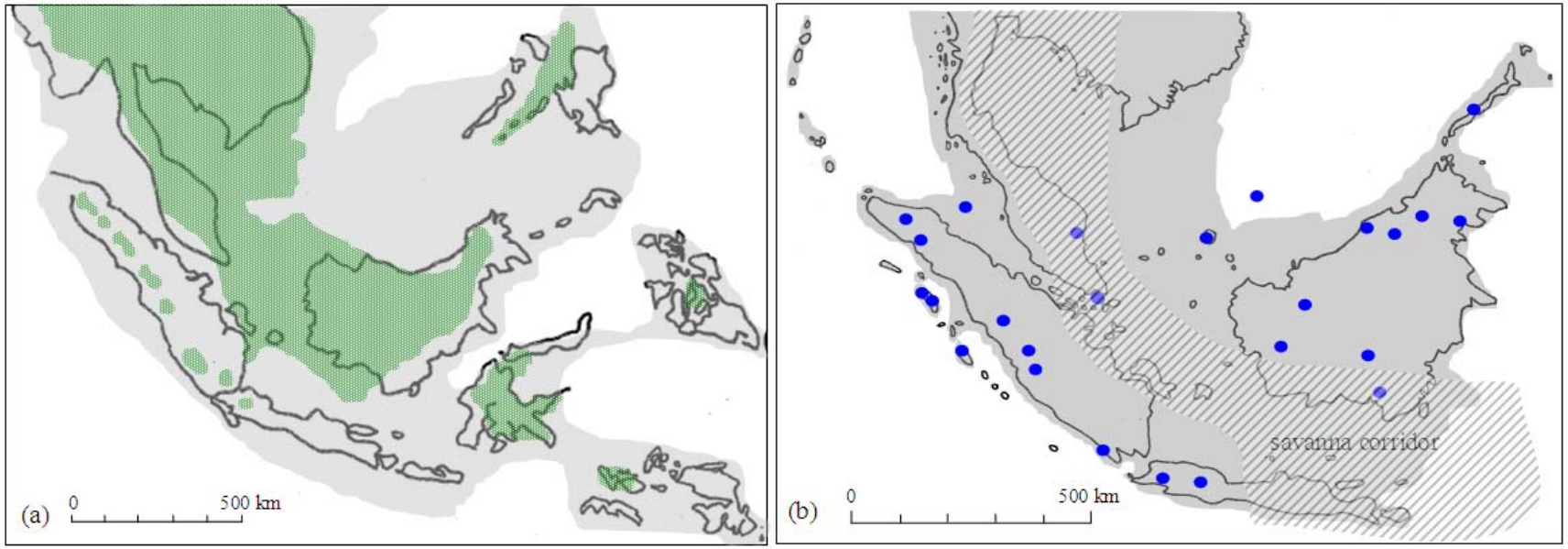

Figure 3.1: (a) Postulated distribution of land and sea in Southeast Asia at 15 Ma (Middle Miocene, adapted from Hall, 1998), everwet tropical rain forests were widespread across the region (Morley, 2000). Land areas are shaded in green colour and shallow sea in light grey; (b) Sundaland at the Last Glacial Maximum (adapted from Bird *et al*., 2005), showing "savanna corridor" proposed by Heaney (1991). Glacial refugia sites inferred from palynology, geomorphology and biogeography data are denoted by circles (Emmel & Curray, 1982; Bird *et al*., 2005; Quek *et al*., 2007).

dynamic ecosystem of Southeast Asia; the tropical rain forests are thought to have contracted into a few glacial refugia (Medway, 1972; van der Kaars & Dam, 1995; Verstappen, 1997; Cranbrook, 2000; Morley, 2000; Thomas, 2000) (Figure 3.1b).

During the Last Glacial Maximum (LGM) of the Pleistocene (20,000–18,000 years 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, which joined the mainland Southeast Asia to Sumatra, Java, Borneo and Palawan (Bird *et al*., 2005; Figure 3.1b). The expansion of the exposed 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 Sunda shelf (Gathorne-Hardy *et al*., 2002). The drought and increased seasonality 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; Figure 3.1b). As the cooling periods ended, the sea rose and some of the surviving rain forests expanded and again recolonized the region (van der Kaars & 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). This basic expansion-contraction model gives a simple paradigm for the demography of species through glacial cycles, but is rendered more complex by factors such as different species having different environmental niches and dispersal rates, the scope for altitudinal migration and interactions between communities of species (Taberlet *et al*., 1998; Hewitt, 1999).

Peninsular Malaysia, one of the Southeast Asian region, which is often credited as bearing the world's oldest flora (Morley, 2000). However, this is some what surprising where there is no phylogeography study has been reported for any plant species. Peninsular Malaysia has existed for over 100 Ma; it has probably been geographically similar to today, with a spine dominated by a north–south range of hills, which are 100–200 km across and rise to over 2,000 m (Woodruff, 2003). 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 & Curray, 1982; Quek *et al*., 2007).

As in most angiosperms, chloroplast DNA (cpDNA) is thought to evolve slowly, with low mutation rates and is known to be maternally inherited (Wolfe *et al*., 1987; Corriveau & Coleman, 1988; Clegg & Zurawski, 1992). There is no recombination occurs in cpDNA, in which the genome is in complete linkage disequilibrium and it is inherited as a single gene (Ouborg *et al*., 1999). Specifically, the transmission of chloroplast genomes through maternal would provide an exceptional opportunity for studying maternal genetic lineages of the species (Testolin & Cipriani, 1997). The growing number of cpDNA universal primers published in the past decade (Taberlet *et al*., 1991; Demesure *et al*., 1995; Weising & Gardner, 1999; Grivet *et al*., 2001; Heinze, 2007) has 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 post-glacial migration routes of plant species from the temperate forest (Comes & Kadereit, 1998; Taberlet *et al*., 1998; Huang *et al*., 2004; Cheng *et al*., 2005; Shephard *et al*., 2007), but comparatively few studies have been dedicated to the tropical species (Chiang *et al.*, 2001; Cannon & 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). The main flower visitors are *Trigona* spp. and *Apis* spp. (Appanah, 1985, 1987). A study of *N. heimi* in Peninsular Malaysia using nuclear microsatellites (Tnah *et al*., 2010) demonstrated moderate genetic differentiation among population (F_{ST} = 0.127). Thus, *N. heimii* might be expected to exhibit strong phylogeographical structure due to moderate level of gene flow. Previous studies on *N. heimii* showed that it is a diploid ($2n$ = 14; Jong & Lethbridge, 1967) and predominantly an outcrossing species, with outcrossing rates estimated at 87.5–97.9% (Konuma *et al*., 2000; Naito *et al*., 2005). In Peninsular Malaysia, little is known about the glacial

refugia and the historical event of recolonization. Therefore, the present study was aimed to unravel the past evolutionary history of *N. heimii*, including the potential glacial refugia sites and the post-glacial recolonization routes of the species in Peninsular Malaysia.

3.2 MATERIALS AND METHODS

3.2.1 Sample collection and DNA extraction

In total, 32 natural populations of *N. heimii*, with eight samples from each population, that were more than 10 cm diameter at breast height were sampled throughout the distribution range of *N. heimii* in Peninsular Malaysia (Table 3.1). A transect line method was utilized as a guide for the sampling activities. The samples were collected in the form of inner bark or leaf tissues, and immediately processed in the field, wrapped with aluminum foil and kept in liquid nitrogen. 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).

3.2.2 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 non-coding regions of cpDNA were found to be informative in this study (Table 3.2): *trn*L intron, *trn*S-*trn*G spacer, *trn*G intron, *trn*K intron and *psb*K-*trn*S spacer. 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 $MgCl_2$, 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 a 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.

Table 3.1: Sampling localities chosen for the population of the *Neobalanocarpus heimii* in Peninsular Malaysia accompanied with the population code and their haplotypes. Eight samples per population were used in this study.

Forest reserve (FR)	State	Population code	Latitude	Longitude	Haplotypes
1. Bukit Enggang	Kedah	BEnggang	05°48’	100°41’	h2, h5
2. Sungkop	Kedah	Sunkop	05°45’	100°38’	h2, h4, h5,h15
3. Bintang Hijau	Perak	BHijau	05°11’	101°00’	h2, h7
4. Piah	Perak	Piah	05°01’	101°02’	h2, h7
5. Pondok Tanjung	Perak	PTanjung	05°04’	100°47’	h2, h5, h6
6. Bubu	Perak	Bubu	04°37’	100°46’	h2
7. Chikus	Perak	Chikus	04°06’	101°12’	h2, h6
8. Jeli	Kelantan	Jeli	05°44’	101°50’	h2, h6
9. Gunung Basur	Kelantan	GBasur	05°36’	101°45’	h2, h4, h9
10. Lebir	Kelantan	Lebir	05°12’	102°20’	h2, h6, h9
11. Hulu Terengganu (compartment 31)	Terengganu	HTrengA	04°56’	102°55’	h2, h7
12. Hulu Terengganu (compartment 14A)	Terengganu	HTrengB	05°00’	102°55’	h2, h3, h6
13. Pasir Raja	Terengganu	PRaja	04°42’	102°58’	h1, h2, h3, h4
14. Rambai Daun	Terengganu	RDaun	04°36’	103°23’	h1, h3
15. Berkelah	Pahang	Berkelah	03°46’	103°08’	h1, h3, h10
16. Tersang	Pahang	Tersang	03°59’	101°49’	h1, h2, h3, h4
17. Rotan Tunggal	Pahang	RTunggal	03°47’	101°51’	h1, h3, h4
18. Lakum	Pahang	Lakum	03°37’	102°05’	h1, h3, h12, h13
19. Bukit Tinggi	Pahang	BTinggi	03°31’	101°52’	h1, h3
20. Lentang	Pahang	Lentang	03°23’	101°54’	h1, h3, h12
21. Kemasul	Pahang	Kemasul	03°25’	102°13’	h1, h3
22. Lesong	Pahang	Lesong	02°46’	103°08’	h1, h11
23. Gombak	Selangor	Gombak	03°20’	101°46’	h1, h3
24. Ampang	Selangor	Ampang	03°10’	101°47’	h1, h3
25. Sungai Lalang	Selangor	SLalang	03°05’	101°52’	h1, h3
26. Pelangai	Negeri Sembilan	Pelangai	02°48’	102°11’	h1, h3
27. Pasoh	Negeri Sembilan	Pasoh	02°59’	102°19’	h1
28. Labis	Johor	Labis	02°21’	103°10’	h1, h3
29. Lenggor (compartment 32)	Johor	LenggorA	02°11’	103°40’	h1, h3
30. Lenggor (compartment 76)	Johor	LenggorB	02°10’	103°40’	h1, h14
31. Panti (compartment 16)	Johor	PantiA	01°47’	103°57’	h1, h3, h8
32. Panti (compartment 68)	Johor	PantiB	01°49’	103°55’	h1, h3

Table 3.2: Five universal chloroplast DNA primer pairs used in this study.

Primer name	Primer sequence (5'–3')	Reference	Amplified region
ucp-c ucp-d	F: CGAAATCGGTAGACGCTACG R: GGGGATAGAGGGACTTGAAC	Taberlet *et al*., 1991 Taberlet *et al*., 1991	*trn*L intron
trnS (GCU) Hamilton trnG (UCC) Hamilton	F: GCCGCTTTAGTCCACTCAGC R: GAACGAATCACACTTTTACCAC	Hamilton, 1999 Hamilton, 1999	*trn*S-*trn*G spacer
*trn*G2-f *trn*G1-r	F: GTTTAGTGGTAAAAGTGTGATTCGTT R: CCGCATCGTTAGCTTGGAAGGC	Heinze, 2007 Heinze, 2007	*trn*G intron
ccmp1f *trn*K2r	F: CAGGTAAACTTCTCAACGGA R: CAACGGTAGAGTACTCGGCTTTTA	Demesure *et al*., 1995 Weising & Gardner, 1999	*trn*K intron
*psb*K-P1 *trn*S0r	F: GCCTTTGTTTGGCAAGCTGCTGTAAG R: GGGAGAGATGGCTGAGTGGAC	Heinze, 2007 Grivet *et al*., 2001	*psb*K-*trn*S spacer

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).

3.2.3 Statistical analyses

Haplotypes were determined solely based on substitution intraspecific variable sites. Insertion and deletion (indel) variable sites were excluded from analyses, as different evolutionary rates were found in 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 3.1) 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 distribution pattern of haplotype. Significance was tested with 10,000 permutations.

Likewise, two measures of diversity and population differentiation were computed for the two different grouping schemes using the PERMUTE and CPSSR version 2.0 as described in Pons and Petit (1996) and Burban *et al*. (1999). The parameters included the mean within-population gene diversity (H_S), the total gene diversity (H_T), and the coefficient of genetic differentiation over all populations (G_{ST}), as well as the other equivalent parameters (V_S, V_T and N_{ST}) obtained by taking into account similarities between haplotypes. G_{ST} depends only on the frequencies of the haplotypes, whereas N_{ST} 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 G_{ST} and N_{ST} 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 (F_{ST}) 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.

Nested Clade Phylogeographic Analysis (NCPA: Clement *et al*., 2000; Posada *et al*., 2000; Panchal, 2007) was performed to test for significant associations of haplotypes with geography, and to infer processes that may have led to haplotype distribution patterns. A network of haplotypes was first constructed using the package TCS version 1.18 (Clement *et al*., 2000), with the observed and inferred haplotypes were clustered hierarchically by progressively grouping vertices in the graph according to statistical parsimony (Templeton *et al*., 1992). Lower order clades were nested within higher order clades to create a nested design. The program GEODIS version 2.2 (Posada *et al*., 2000) was used to calculate various NCPA distance measures and their statistical significance. All tests of significance were made through 10,000 Monte Carlo permutation procedures. Results obtained from GEODIS were further interpreted using the revised inference key of Templeton *et al*. (2005). The statistics calculated for all clades were: (1) clade distance (D_C), (2) nested clade distance (D_N), and (3) interior-tip distances (I-TD_C and I-TD_N). Testing of significantly smaller or larger D_C and D_N distances in each nested clade was subsequently used to test the null hypothesis of no association between haplotype and geography distributions (Templeton *et al*., 1995).

3.3 RESULTS

3.3.1 Chloroplast DNA variability and genetic diversity

The examined sequences consisted of *trn*L intron (584–591bp), *trn*S-*trn*G spacer (575–585bp), *trn*G intron (660–661bp), *trn*K intron (569–579bp), and *psb*K-*trn*S spacer

(679bp). The corresponding GenBank accession numbers are EU918738–EU918763. The combined sequences of these five chloroplast non-coding regions resulted in a total of 3,095 bp. Indel polymorphic sites were then removed and resulted in an aligned length of 3,069 bp. In total, 10 intraspecific variable sites due to nucleotide substitutions were detected within these five non-coding regions (Table 3.3). Among these variable sites, three were found in each of the *trn*L intron and *trn*S-*trn*G spacer, two in the *trn*G intron, and one each in the *trn*K intron and *psb*K-*trn*S spacer.

The haplotype and nucleotide diversity measures for all the 32 populations are shown in Table 3.4. 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 haplotype, the Sungkop, Lakum, Tersang and PRaja were the most variable populations with four different haplotypes whereas the Bubu and Pasoh were the most homogenous populations that consisted of only one haplotype. The Sungkop also had the highest values of haplotype diversity (h = 0.821) and nucleotide diversity (π = 0.00044). Other populations that had high levels of diversity were Lakum (h = 0.786; π = 0.00034), Tersang (h = 0.750; π = 0.00040), PRaja (h = 0.750; π = 0.00034), PantiA (h = 0.714; π = 0.00033) and GBasur (h = 0.714; π = 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 3.4). This may indicate that the evolution was neutral and selection might not have played an important role in shaping the structure of genetic diversity within populations.

3.3.2 Genetic differentiation

The AMOVA revealed that 56.16% of observed variation was due to differences among populations and 43.84% within population (all partitions were significant at P < 0.05, Table 3.5). 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).

Table 3.3: Distribution of intraspecific variable sites in the five noncoding region of chloroplast DNA (*trn*L intron, *trn*S-*trn*G spacer, *trn*G intron, *trn*K intron and *psb*K-*trn*S spacer) of *Neobalanocarpus heimii* in Peninsular Malaysia.

Haplotype	*trn*L intron			*trn*S-*trn*G spacer			*trn*G intron		*trn*K intron	*psb*K-*trn*S spacer
	201	380	457	349	373	374	56	83	173	565
h1	C	G	G	T	A	T	A	C	A	G
h2	C	G	A	T	A	T	A	C	A	A
h3	C	G	G	T	A	T	A	C	A	A
h4	C	G	A	G	A	T	A	C	A	A
h5	C	G	A	T	A	G	A	C	A	A
h6	C	C	A	T	A	T	A	C	A	A
h7	C	G	A	T	C	T	A	C	A	A
h8	C	G	G	T	A	T	A	C	C	G
h9	C	G	A	T	A	T	T	C	A	A
h10	C	G	G	T	A	T	A	T	A	G
h11	C	G	A	T	A	T	A	C	A	G
h12	T	G	G	T	A	T	A	C	A	A
h13	C	G	G	G	A	T	A	C	A	A
h14	C	G	G	T	A	T	T	C	A	G
h15	G	G	A	T	A	T	A	G	A	A

Table 3.4: Population of *Neobalanocarpus heimii* with haplotype and nucleotide diversity associated with the results of Tajima's D neutrality test.

Population code	Haplotypes (No. of individuals)	Haplotype diversity	Nucleotide diversity x 10^3	Tajima's D ($P > 0.05$)
Total	h1-h15(256)	0.749 ± 0.014	0.41 ± 0.01	-0.528
BEnggang	h2(7), h5(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
Sunkop	h2(3), h4(2), h5(2), h15(1)	0.821 ± 0.101	0.44 ± 0.12	-0.525
BHijau	h2(7), h7(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
Piah	h2(6), h7(2)	0.429 ± 0.169	0.14 ± 0.05	0.334
PTanjung	h2(4), h5(3), h6(1)	0.679 ± 0.122	0.26 ± 0.06	0.069
Bubu	h2(8)	0.000	0.00	-
Chikus	h2(7), h6(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
Jeli	h2(7), h6(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
GBasur	h2(4), h4(2), h9(2)	0.714 ± 0.123	0.28 ± 0.07	0.414
Lebir	h2(6), h6(1), h9(1)	0.464 ± 0.200	0.16 ± 0.08	-1.310
HTrengA	h2(7), h7(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
HTrengB	h2(6), h3(1), h6(1)	0.464 ± 0.200	0.16 ± 0.08	-1.310
PRaja	h1(1), h2(4), h3(2), h4(1)	0.750 ± 0.139	0.34 ± 0.09	-0.431
RDaun	h1(4), h3(4)	0.571 ± 0.094	0.19 ± 0.03	1.444
Berkelah	h1(1), h3(5), h10(2)	0.607 ± 0.164	0.31 ± 0.09	0.932
Tersang	h1(1), h2(1), h3(4), h4(2)	0.750 ± 0.139	0.40 ± 0.09	0.204
RTunggal	h1(4), h3(3), h4(1)	0.679 ± 0.122	0.35 ± 0.12	-0.304
Lakum	h1(3), h3(3), h12(1), h13(1)	0.786 ± 0.113	0.34 ± 0.08	-0.431
BTinggi	h1(4), h3(4)	0.571 ± 0.094	0.19 ± 0.03	1.444
Lentang	h1(5), h3(1), h12(2)	0.607 ± 0.164	0.26 ± 0.08	0.069
Kemasul	h1(6), h3(2)	0.429 ± 0.169	0.14 ± 0.05	0.334
Lesong	h1(6), h11(2)	0.429 ± 0.169	0.14 ± 0.05	0.334
Gombak	h1(4), h3(4)	0.571 ± 0.094	0.19 ± 0.03	1.444
Ampang	h1(3), h3(5)	0.536 ± 0.123	0.17 ± 0.04	1.167
SLalang	h1(3), h3(5)	0.536 ± 0.123	0.17 ± 0.04	1.167
Pelangai	h1(7), h3(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
Pasoh	h1(8)	0.000	0.00	-
Labis	h1(7), h3(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
LenggorA	h1(6), h3(2)	0.429 ± 0.169	0.14 ± 0.05	0.334
LenggorB	h1(7), h14(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055
PantiA	h1(2), h3(2), h8(4)	0.714 ± 0.123	0.33 ± 0.07	1.104
PantiB	h1(7), h3(1)	0.250 ± 0.180	0.08 ± 0.06	-1.055

Table 3.5: Hierarchical analysis of molecular variance (AMOVA) of *Neobalanocarpus heimii* based on two different grouping schemes.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
All 32 populations				
Among populations	31	97.297	0.35745	56.16
Within populations	224	62.500	0.27902	43.84
Total	255	159.797	0.63647	
Partitioned into two regions				
Among regions	1	77.455	0.62181	65.55
Among population within regions	30	19.842	0.0478	5.04
Within populations	224	62.500	0.27902	29.41
Total	255	159.797	0.94862	

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 3.6). 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 & 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, southerly populations ($G_{ST} = 0.151$, $H_T = 0.571$, $H_S = 0.485$) exhibited greater differentiation and total diversity than northerly populations ($G_{ST} = 0.071$, $H_T = 0.461$, $H_S = 0.429$). 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, in contrary, the differences of $N_{ST} - G_{ST}$ were not significant. Such observation might indicate a complex history of colonization by 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 (Figure 3.2). 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 3.3). The populations of Sungkop, PTanjung, GBasur and PantiA contributed most to the differentiation component, while those of Sungkop, PRaja, Tersang and Lakum contributed most to the diversity component of total diversity.

3.3.3 Haplotype distribution and relationship inferred from NCPA

The distribution of the cpDNA haplotypes throughout Peninsular Malaysia is shown in Figure 3.4. 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

Table 3.6: Population structure of *Neobalanocarpus heimii* for the northern region, southern region and the whole Peninsular Malaysia. Values in parenthesis are standard deviation.

Diversity parameters and Mantel test	Northern	Southern	Total
H_S	0.429 (0.069)	0.485 (0.048)	0.462 (0.040)
H_T	0.461 (0.075)	0.571 (0.047)	0.757 (0.020)
G_{ST}	0.071 (NC)	0.151 (0.046)	0.390 (0.043)
V_S	0.425 (0.088)	0.469 (0.062)	0.334 (0.038)
V_T	0.462 (0.097)	0.572 (0.077)	0.761 (0.038)
N_{ST}	0.080 (NC)	0.181 (0.044)	0.562 (0.043)
N_{ST}-G_{ST}	0.009	0.030*	0.172*

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

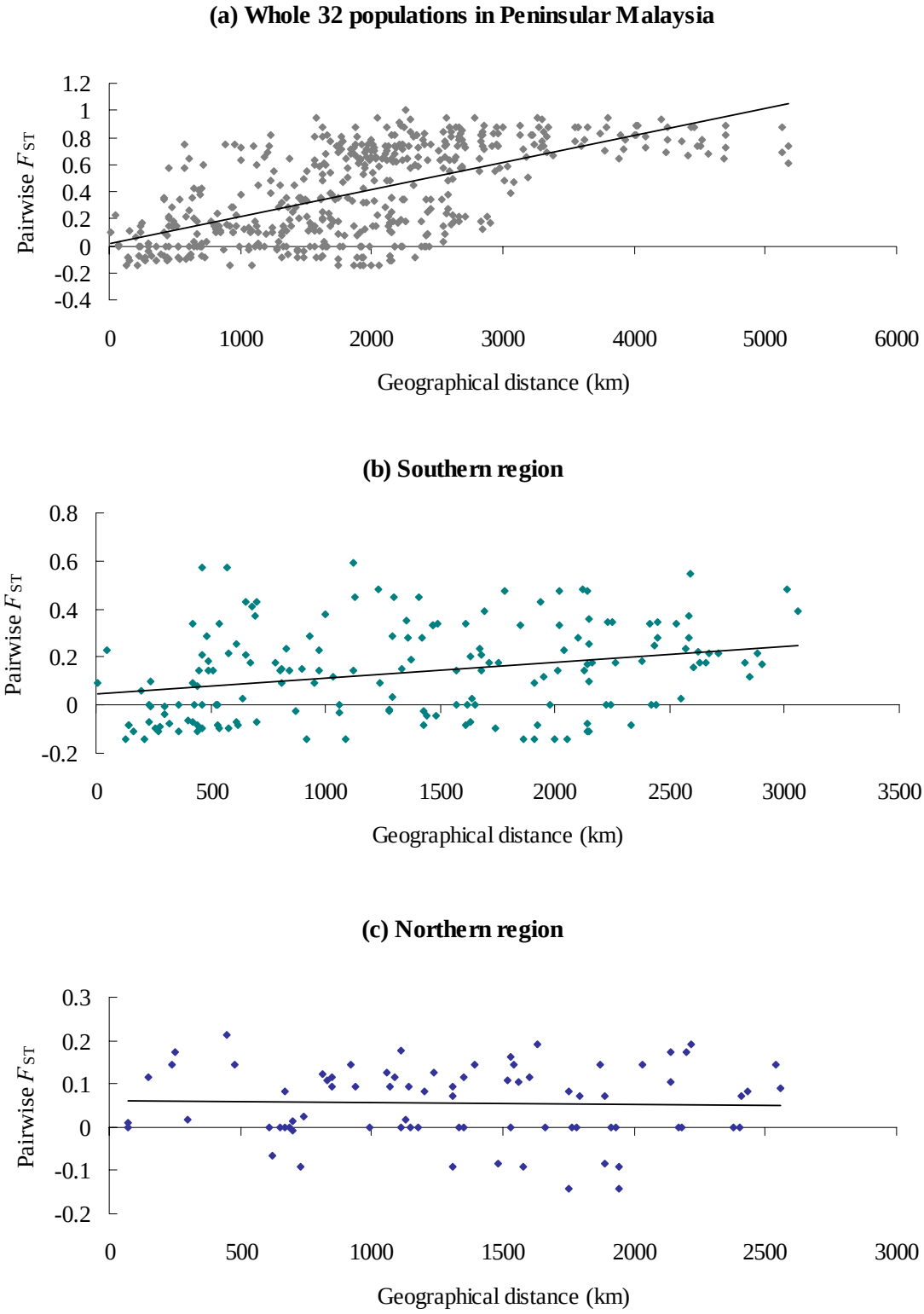

Figure 3.2: Relationship between pairwise population F_{ST} and geographical distance of *Neobalanocarpus heimii* for the (a) whole 32 populations of Peninsular Malaysia (Mantel test of correlation, r = 0.619, P = 0.0000), (b) southern region (Mantel test of correlation, r = 0.288, P = 0.0082), and (c) northern region (Mantel test of correlation, r = -0.027, P = 0.5720).

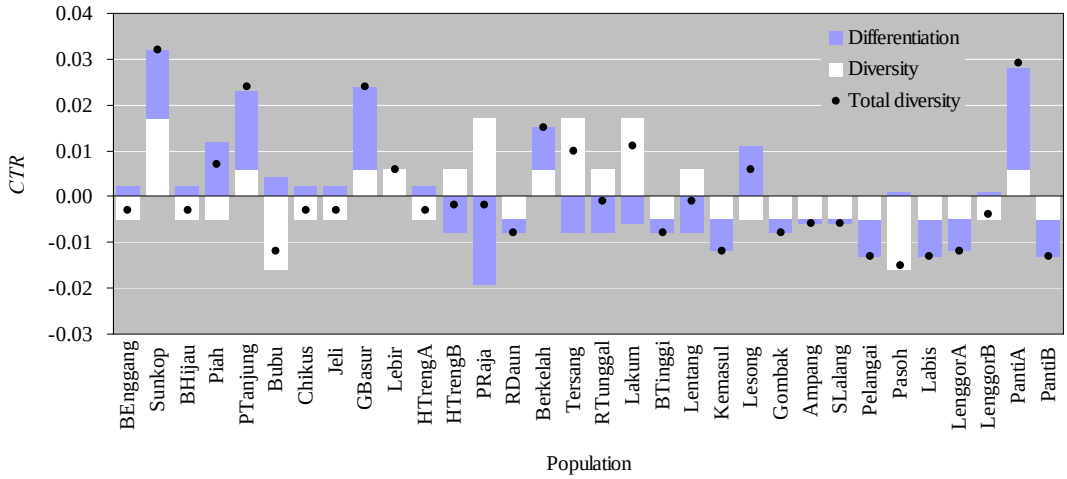

Figure 3.3: Contribution of each population of *Neobalanocarpus heimii* to the total diversity, as described by the haplotype richness (*CTR*). Circles represent total diversity; solid and open bars represent contributions of differentiation and diversity to the total diversity, respectively.

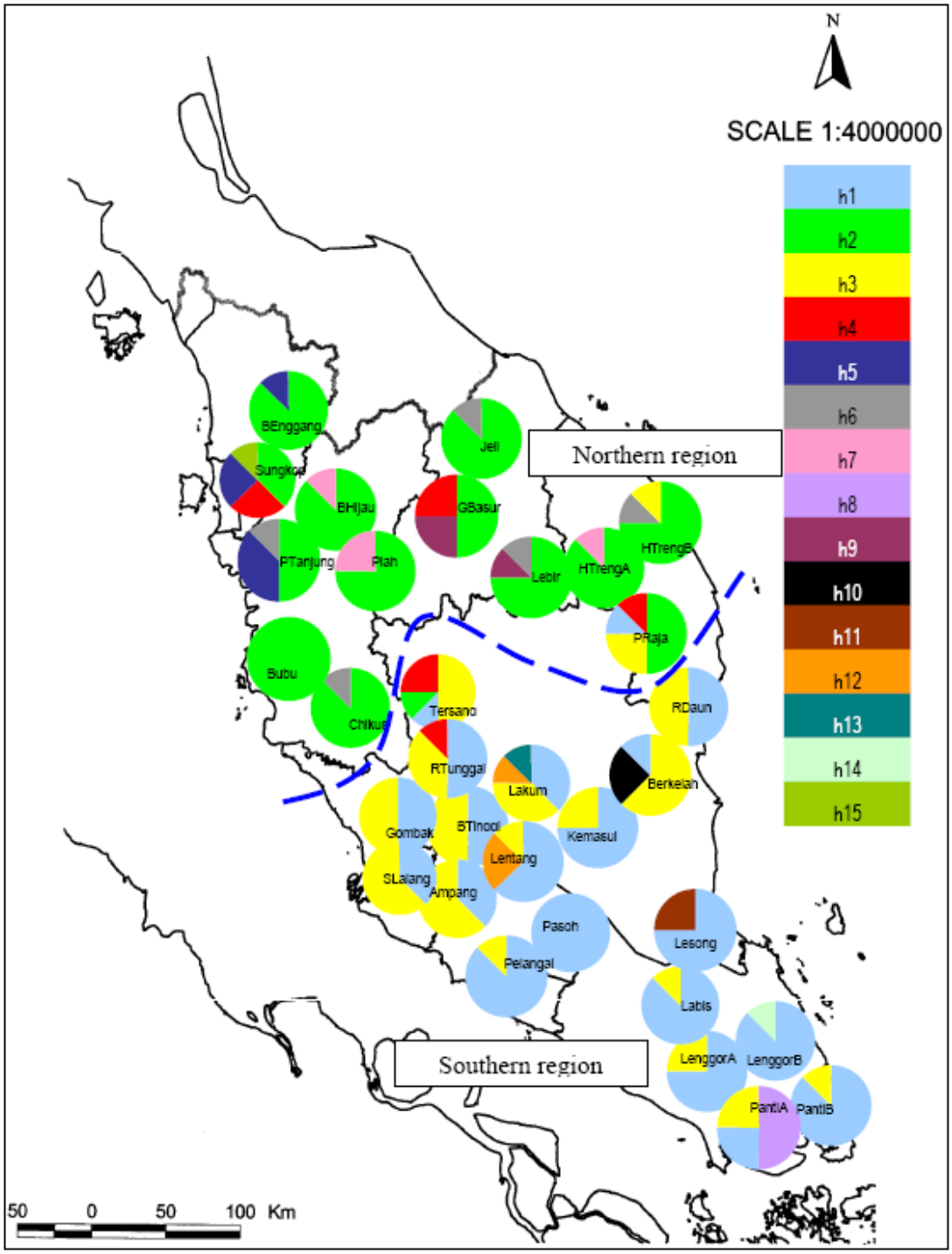

Figure 3.4: Locations of 32 populations of *Neobalanocarpus heimii* in 29 forest reserves of Peninsular Malaysia. The geographical distribution of cpDNA haplotypes (h1–h15) in each population is shown and haplotypes are colour coded and their frequencies are indicated by sectors of pies.

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.

Relationships between haplotypes are shown in the statistical parsimony network in Figure 3.5. The maximum number of mutation steps between haplotypes in the network is four. Haplotypes h1, h2 and h3 have the most linkages to other haplotypes with five, seven and four connecting haplotypes, respectively. They were in the central part of a network, had higher frequencies than tip haplotypes and showed broader geographical distributions. Subsequently, following the coalescent theory, these haplotypes are more likely to be the ancestral haplotypes. The total cladogram was partitioned into two separate networks (clades 2-1 and 2-2) for all the cpDNA haplotypes identified. In accordance with Figure 3.4, haplotypes under clade 2-2 (h2, h4–h7, h9 and h15) were distributed across the northern region of Peninsular Malaysia whereas haplotypes clustered under clade 2-1 (h1, h3, h8 and h10–h14) were spread across the southern region. This might further support that in terms of haplotype relationships, basically the 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, Kemasul, Lesong, Gombak, Ampang, SLalang, Pelangai, Pasoh, Labis, LenggorA, LenggorB, PantiA and PantiB). This pattern might also suggest that the populations of *N. heimii* in these two geographical regions have had independent evolutionary histories for a relatively long period of time. It is interesting to note also that the 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.

In general, the statistical parsimony procedure also revealed that all the cpDNA haplotypes are closely related, most differing from their closet relatively by 1–2 mutational steps. As shown in Figure 3.5, two nesting levels can be clearly defined: one-step mutation clades (1-1, 1-2, 1-3 and 1-4) and two-step mutation clades (2-1 and

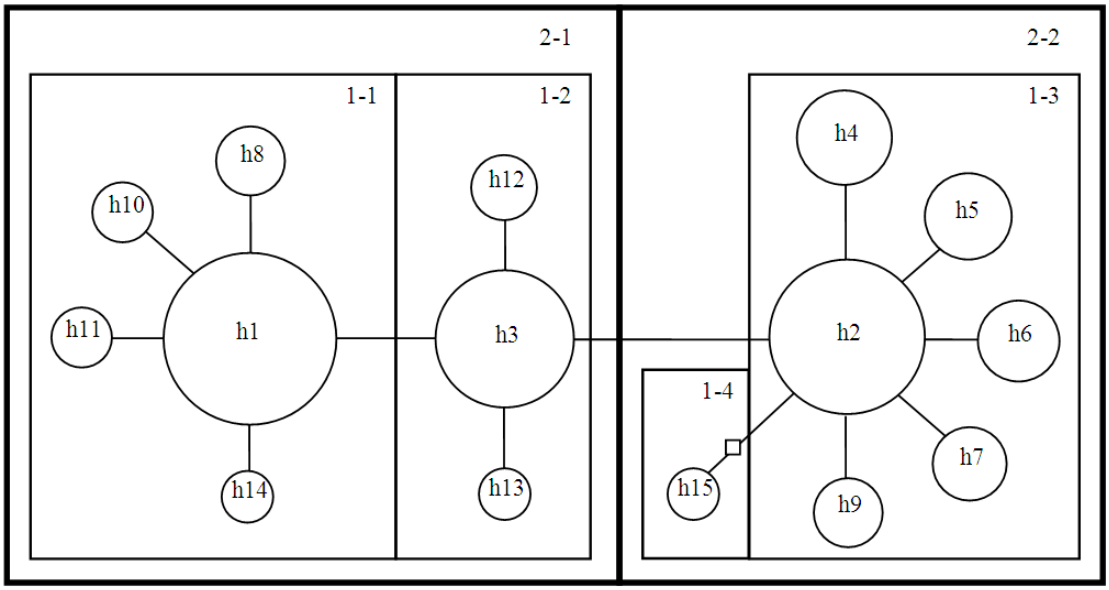

Figure 3.5: The resolved cpDNA haplotype network and resulting set of nested clades for *Neobalanocarpus heimii*. Each line connecting one haplotype to another indicates one mutational step. The relative sizes of the circles represent the frequency of these haplotypes and unsampled haplotypes are represented by square box.

2-2). The significant associations between haplotypes and geography distributions were detected on clades 1-1, 1-3, 2-1 and the total cladogram (Table 3.7). In clade 1-1, the NCPA indicated restricted gene flow with some long-distance dispersal over intermediate areas not occupied by the species; or past gene flow followed by extinction of intermediate populations. Restricted gene flow with isolation by distance was inferred for clade 1-3, whereas contiguous range expansion was inferred for clade 2-1. For the total cladogram, although geographical structuring of haplotypes was significant, non-significance of D_n or D_c resulted in inconclusive outcome of the demographic event. The absence of significant associations between haplotypes and geography distributions in clades 1-2 and 2-2 could be due to panmixia, absence of historical demographic changes, insufficient genetic variation, or inadequate sampling. In addition, the sampling populations appear to have experienced limited extinction, because there was solely one missing haplotype found in clade 1-4 (Figure 3.5).

3.4 DISCUSSION

3.4.1 Hypothetical phylogeographical history

The radiation of Dipterocarpaceae in Southeast Asia is reported during the drier Oligocene and earliest Miocene, following their dispersal from Africa via the Indian plate (Muller, 1981; Ashton, 1982; Ashton & Gunatilleke, 1987). The common occurrence of small tricolplate pollen in palynomorph assemblages, which is comparable to that of the dipterocarps *Shorea* and *Hopea*, might suggest extremely widespread of low diversity dipterocarps monsoon forest in the earliest Miocene (Morley, 2000). However, the main radiation time for Dipterocarpaceae into the everwet rain forests is taken place after 20 Ma (Ashton, 1982). Since this period, the family has probably become a major component of the Southeast Asian rain forest (Morley, 2000). *Neobalanocarous heimii*, as one of the dipterocarps, is therefore predicted to be radiated throughout the tropical rain forest of Peninsular Malaysia since Middle Miocene.

The past evolutionary history of *N. heimii* may be explained in three major phases. As for the first phase, the perhumid climates during the Middle Miocene would allow the ancestral haplotypes of *N. heimii* to expand rapidly throughout the southern and northern regions of Peninsular Malaysia. In the second phase, during the advent of the

Table 3.7: Nested contingency analysis of geographic associations and their interpretations according to the inference key of Templeton *et al*., (2005) for 32 populations of *Neobalanocarpus heimii* in Peninsular Malaysia. Clade numbers refer to numbers reported in Figure 3.5.

Clade	Chi-square	*P*	Inference chain	Inferred demographic event
1-1	170.0318*	0.0028	1-2-3-5-6-7-8 YES	Restricted gene flow but with some long-distance dispersal over intermediate areas not occupied by the species; or past gene flow followed by extinction of intermediate populations
1-2	23.2235	0.7523	Null hypothesis cannot be rejected	Panmixia, absence of historical demographic changes, insufficient genetic variation, or inadequate sampling
1-3	111.0753*	0.0043	1-2-3-4 NO	Restricted gene flow with isolation by distance
2-1	38.9110*	0.0028	1-2-11-12 NO	Contiguous range expansion
2-2	12.1165	1.0000	Null hypothesis cannot be rejected	Panmixia, absence of historical demographic changes, insufficient genetic variation, or inadequate sampling
Total	233.1984*	0.0000	1-2 IO	Inconclusive outcome

*Significant *P* value ($P < 0.05$).

Pleistocene glacial, the widespread populations were fragmented and restricted to a few refugia (Corner, 1978; Geyh *et al*., 1979; Emmel & Curray, 1982; Quek *et al*., 2007), where these refugia underwent significant genetic drift and diversified in isolation. In the third phase, as the perhumid conditions returned during Holocene warming, the rain forests expanded out of these refugia (Whitemore, 1998). Herein, the persistence of ancestral haplotypes h1 and h3 or h2 in the majority of the extant populations might suggest the contemporary populations of *N. heimii* are more likely descended from the remnant populations that survived in some relict refugia during Pleistocene.

3.4.2 Genealogical lineages

Results generated from NCPA clearly revealed two major genealogical lineages for *N. heimii*: a widespread southern region (clade 2-1) and a northern region (clade 2-2). The pronounced divergence could be attributable to the long-term isolation of populations within geographically separate refugia. During the glacial periods, populations were reserved at relatively small sizes within refugia, thus a combination of genetic drift and new mutations might have been responsible for the differentiation and apportionment of the *N. heimii* genepool in the northern and southern regions. Specifically, both diverge regions could have gone through a stochastic lineage sorting over 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 more likely that both regions have become reciprocal monophyly. Similar distribution pattern was drawn for other flora in Peninsular Malaysia, where the northern region was recognized as one of the special floristic provinces, which might be understood as a core area of seasonal Asiantic intrusion in Peninsular Malaysia (Corner, 1960; Figure 3.6a).

However, there is a paradox regarding the genealogical lineages of *N. heimii* in Peninsular Malaysia revealed from the cpDNA and nuclear microsatellite markers (Tnah *et al*., 2010). The former suggests two regions (northern and southern); the latter suggests three regions (western, northern and southern). One possible explanation for the discrepancy is that these markers reflect different mode of inheritance and mutation rate. Due to uniparental inheritance and low mutation rate, cpDNA represents genealogical lineages and provides a historical perspective on the relationships of populations, which specifically helps in discriminating between population structure and population history (Templeton *et al*., 1995; Schaal *et al*., 1998; Ennos *et al*., 1999).

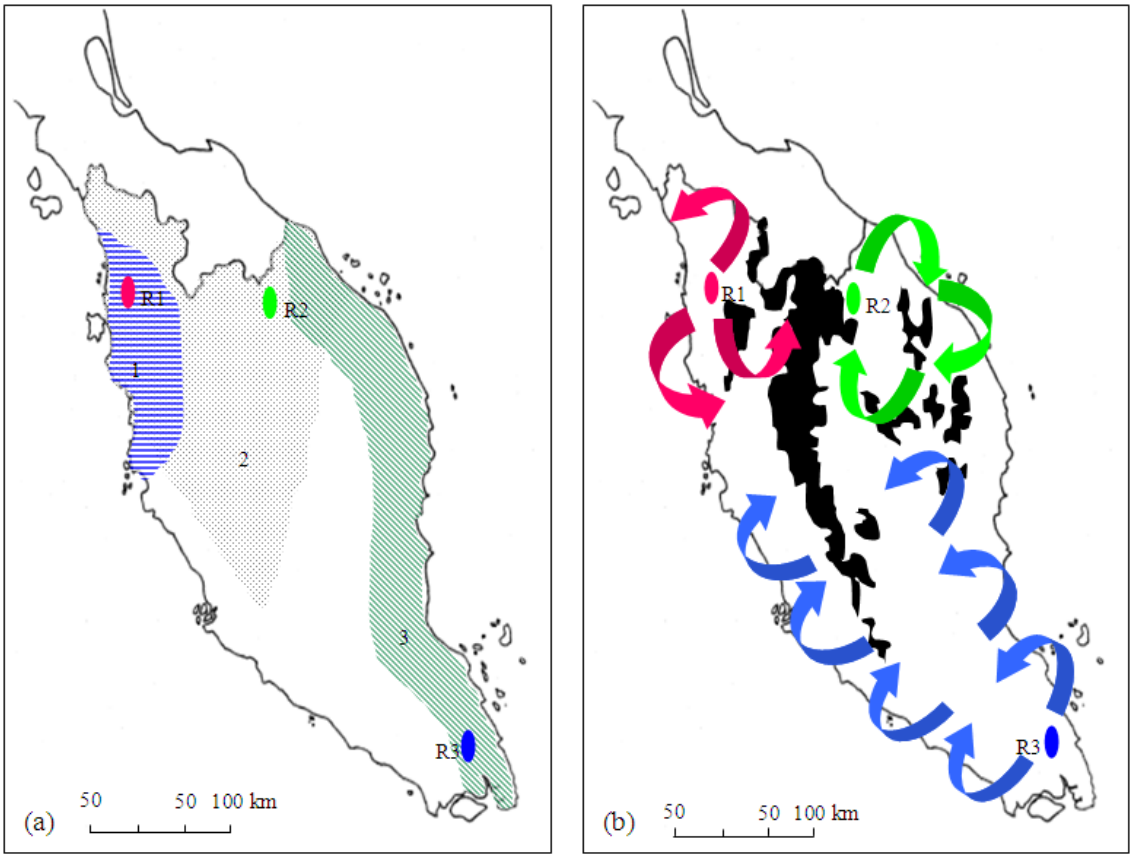

Figure 3.6: (a) Distribution of potential refugia sites (R1, R2 and R3) inferred for *Neobalanocarpus heimii* and special phytogeographical subprovinces recognized in Peninsular Malaysia include (1) Perak subprovince, (2) rough extent of core of seasonal Asiantic intrusion, and (3) the Riau Pocket (Corner, 1960; Ashton, 1992); (b) Hypothetical phylogeographical history of *N. heimii* in Peninsular Malaysia. Region in black indicates mountain range and arrows indicate putative recolonization routes from the R1, R2 and R3.

On the other hand, nuclear microsatellite marker is biparental inherited and is subject to relatively high mutation rates, as they may be in the order of 10^{-3} (Jarne & Lagoda, 1996), which is more important determinant for contemporary population structure.

3.4.3 Three potential refugia recognized in Peninsular Malaysia

The refugium of Peninsular Malaysia is still very much of a matter of conjecture as the extent of tropical rain forest refugia has never been fully investigated. In the present study, however, levels and patterns of genetic diversity could be used as a prediction key to identify the potential refugia and post-glacial regions. According to the basic expansion and contraction model (Hewitt, 1996), refugia areas should harbour 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 post-glacial period (Taberlet *et al*., 1998; Comes & Kadereit, 1998). In addition, the long-term isolation of refugia sites will lead to genetic differentiation due to drift. Also, the refugia tend to harbour a high number of unique haplotypes. In the case of *N. heimii*, based on the level and pattern of genetic diversity, population differentiation and the presence of unique haplotypes, the species is predicted to survive in the multiple refugia throughout Peninsular Malaysia (Figure 3.6a): the northwestern region (R1: Sungkop), the northeastern region (R2: GBasur), and the southern region (R3: PantiA). These putative glacial refugia exhibit higher level of genetic diversity, while the post-glacial regions were found to have a reduced level of genetic diversity with large geographic areas dominated by a single haplotype. According to Graham (1982), an area with high species diversity and endemism could be implied as a glacial refuge. In this case, this coincides at least in large part, with the three special floristic provinces recognized by Corner (1960; Figure 3.6a), which add further support to the identity of these refugia. In particular, the special floristic province recognized in Peninsular Malaysia includes (1) a special area of high diversity and endemism in the Perak region (including the adjacent part of Kedah), (2) an enclave of southward invasion by Burmese-Thai floristic elements, corresponding to a V-shaped area with its base at Kuala Lumpur and (3) the Riau Pocket, characterized by floristic affinities between northwest Borneo, the Riau archipelago, central Sumatra and along eastern and southeastern of Peninsular Malaysia (Corner, 1960; Ashton, 1992).

3.4.4 Post-glacial recolonization routes

A putative recolonization route for *N. heimii* throughout Peninsular Malaysia was sketched based on the level of genetic diversity, distribution pattern of haplotypes and restricted dispersal of the species (Figure 3.6b). Taken together, by looking at northern region, NCPA, Mantel test and $N_{ST} - G_{ST}$ indicated no statistically significant distances within the clade, 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 of 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 northeastwardly and northwestwardly after the climatic amelioration. By contrast, in the southern region, the NCPA inferred contiguous range expansion (clade 2-2) as one of the major process influencing the spatial distribution of haplotypes. In the meantime, Mantel test and $N_{ST} - G_{ST}$ 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. Along the expansion process, it is more clearly illustrated by the post-glacial regions, which typically had lower genetic diversity and occupied larger geographic areas. After the expansion event, however, the NCPA result suggests that the present-day distribution of *N. heimii* along the southeastern of Peninsular Malaysia may have achieved by past gene flow, in which some of the intermediate populations might have extinct and thereby obscuring relationships between the contemporary populations.

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 ridge, and it seems Peninsular Malaysia was partitioned into north-south rather than east-west orientation. Indeed, this

old and relatively stable mountain range was already existing before the Paleocene (Gobbett & Hutchison, 1973), while much of the present rain forests colonized Southeast Asia during the Middle Miocene (Morley, 2000). Moreover, Ashton (1988) and Filippelli (1997) suggested although the dipterocarps radiated 20 Ma, they did not become dominant and widespread elements of regional rain forests until Southeast Asia became progressively wetter during 10 Ma. Apparently, it is unlikely for these lowland dipterocarps with limited seed dispersal to cross over this mountain range. Therefore, the species could have undergone contiguous range expansion from refugia R3 in the southern flank towards the northeast and northwest regions of Peninsular Malaysia. Additionally, it is interesting to note that the homogeneity was associated for most of the populations in the western region and this would seem that these populations are derived from a recent recolonization by *N. heimii*, or this might be attributed to founder effects in these westerly populations.

The unravelling of post-glacial 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 non-refugia population has also been observed in *Castanopsis carlesii*, which shows the highest diversity due to the natural post-glacial recolonization from different refugia (Cheng *et al*., 2005).

3.4.5 Implications for conservation

Tectonic movement and climatic oscillations during the Cenozoic have had dramatic effect on the biota of the tropical rain forest, and research address specifically on 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. Further, 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 northwestern region (R1: Sungkop), the northeastern region (R2: Gunung Basur), and the southern region (R3: Panti compartment 16), which harbours 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 (Rauch & Bar-Yam, 2005; Frankel & Soulé, 1981). Therefore, the contact zones of *N. heimii* (Tersang, PRaja and RTunggal) with remarkably high haplotype diversity should also be considered in conservation strategies, in order to safeguard long-time survival of this species.