CHAPTER 2

LITERATURE REVIEW

2.1 Phylogeography

Since the term "phylogeography" was first introduced in 1987 (Avise, 2000), the field has rapidly developed and has had great impact on scientific literature. As part of historical biogeography, phylogeography is defined as a field of study that aims to study the principles and historical processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species (Avise, 2000). In particular, phylogeography seeks to understand how the historical events have helped to shape the current dispersion of genes, populations and species; therefore it embraces both aspects of time and space (Freeland, 2005). The field of phylogeography has gained much of its current popularity by providing a linkage between those disciplines concentrating on population and species responses to relatively recent evolutionary processes or paleoclimatic events in Earth's history and those that have traditionally been more concerned with the association between biological diversification and geological events in Earth's history (Riddle & Hafler, 2007).

During the 1960s, the developments of plate tectonics theory (Hess, 1962) and cladistic systematics (Hennig, 1966) have played an important role in shaping modern phylogeography. At the present day, phylogeography incorporates a variety of phylogenetic as well as population genetic approaches and methods to address questions that span the traditional boundaries between the macroevolutionary and microevolutionary arenas (Lomolino *et al*., 2006). The field has strong association with molecular evolutionary genetics, natural history, population biology, paleontology, historical geography and speciation analysis (Avise, 2000). Also, some recent studies have promoted the use of fossil evidence and geological, palynological, phytogeographical, paleoecological and paleoclimatological models in conjunction with phylogeography to improve the inference of historical distributions (Cruzan &

Templeton, 2000; Gathorne-Hardy *et al*., 2002; Hugall *et al*., 2002; Seppä, 2003). Additionally, coalescent theory is integrated in newly developed approaches, which can more accurately address the relative roles of different historical forces in shaping current distribution of gene genealogies (Cruzan & Templeton, 2000).

2.1.1 Geologic time scales

The geologic time scale is a framework which relates stratigraphy to time used by geologists, paleontologists and other earth scientists to describe the history of the Earth (Gradstein *et al*., 2004). Kardong (2005) gave a detailed description of the framework (Figure 2.1). The Earth's history spans from 4.6 billion years ago until the present, and is partitioned into four eons: Hadean, Archaen, Proterozoic and Phanerozoic. In the early Archaen, the Earth received heavy meteorite bombardment, which caused the crust to melt and puncture, leaving it cratered and scarred. After several billion years, crustal and plate tectonic movements have then churned and reworked the Earth's pockmarked surface. During the Proterozoic eon, the world's large continental landmass was broken into small continents and microfossils of prokaryotic, eukaryotic and stromatolites are recorded. As for the Phanerozoic, this eon is marked by the appearance of multicellular organisms, which is partitioned into three eras, Paleozoic (also termed as Age of Fishes), Mesozoic (Age of Reptiles) and Cenozoic (Age of Mammals). By zooming into Cenozoic, the era is divided into the periods of Paleogene and Neogene. The period of Paleogene is then further divided into three epochs (Paleoence, Eocene and Oligocene) while the Neogene is divided into four epochs (Miocene, Pliocene, Pleistocene and Holocene). The boundaries between epochs are marked by changes in the characteristics of geographic site or fauna.

2.1.2 Geological events and climatic changes in Southeast Asia

During the era of Cenozoic, Southeast Asia was formed by active tectonism since the Eocene and was governed by dynamic climate changes during the Pleistocene (Hall, 1998). Particularly, the geological history of Southeast Asia is primarily associated with two major plate collisions: (1) between India and Eurasia about 50–65 Ma (Courtillot *et*

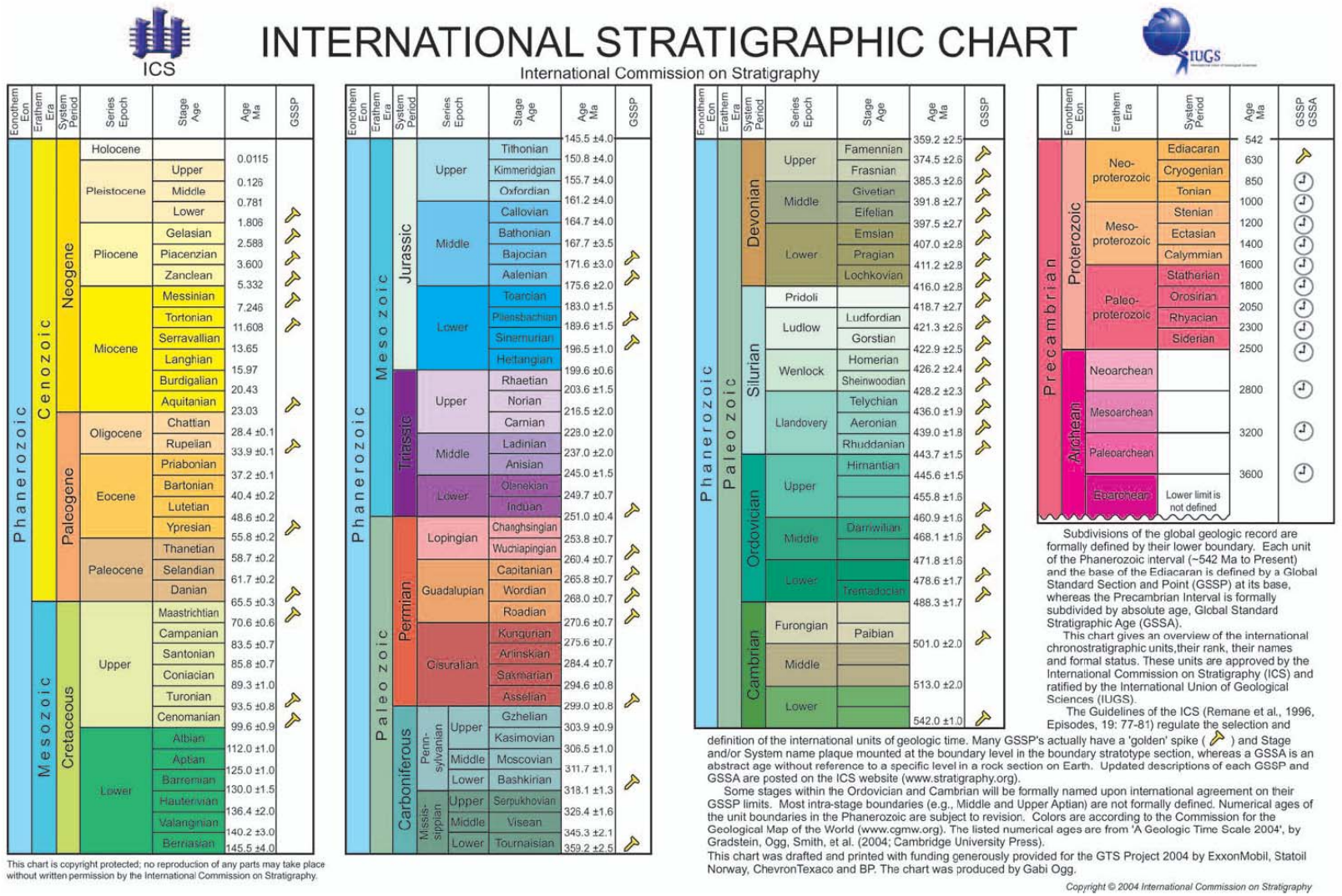

Figure 2.1: The International Stratigraphic Chart summarizing the set of chronostratigraphic units (geologic stages, periods) and their computed ages, which are the main framework for Geologic Time Scale 2004 (Grastein *et al*., 2004).

al., 1986; Beck *et al*., 1995; Hall, 1998); and (2) between Southeast Asia and Australia about 15 Ma (Tapponnier *et al*., 1982; Hutchison, 1989). In addition, the Miocene (23.0–5.3 Ma) was also a significant period of tectonic movement (Hall, 1998), with the Philippine Sea plate rotating clockwise, and the Borneo plate and the northern Sumatra/southern Malay peninsular plate rotating counter-clockwise. The northern Thai–Malay peninsular plate, however, rotated clockwise and remained attached both to Indochina and to the southern Malay plate (Hall, 1998). It is speculated that the Southeast Asia had assumed its current geography by 10 Ma (Hall, 1998).

It is believed that the collision between India and Eurasia was the main cause of the uplift of the Himalaya-Tibetan Plateau and subsequently led to the formation of the modern pattern of summer- and winter-monsoon circulation (Kutzbach *et al*., 1993; Ruddiman, 1998; Zhisheng *et al*., 2001). The intensity of these monsoonal winds had widely been influenced by the emergence of Sunda shelf, due to the growth of continental glaciers during Pleistocene glacial periods (Hutchison, 1989; Heaney, 1991). The winter monsoon is caused by air from cold high pressure areas in Siberia and the Tibetan plateau blowing to the hot low pressure zone over Australia. The winter monsoon picks up water from the West Pacific Warm Pool and the South China Sea and is the cause of most of the rainfall in the Sunda shelf (Whitten *et al*., 1984, 1996; MacKinnon *et al*., 1997; An, 2000). In contrast, the summer monsoon picks up water from the Indian Ocean, causing rain in the summer months (An, 2000).

During the Last Glacial Maximum (LGM) of the Pleistocence (20,000–18,000 years before present), the temperature in Southeast Asia was 3–7 °C lower than at present (Morley, 2000). The sea level was approximately 120 m below the present level, which linked the major land masses of insular Southeast Asia, exposing approximately 1,800,000 km^2 of the Sunda shelf (Shackleton, 1987; Whitemore, 1987; Pirazzoli, *et al*., 1991; Voris, 2000). This exposed continent is known as “Sundaland” (Mollengraaff, 1921). The Sundaland has 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.

The "Pleistocene Refuge Theory" has been widely used to explain the distribution of organisms, which reduced to a fraction of their present distribution as dry climate vegetation expanded during glacial advances (Haffer, 1969). However, the diversification of tropical rain forest biota indeed continue through a major part of the Cenozoic and was not exclusively a Pleistocene phenomenon. Therefore, the application of the refuge theory in this ecosystem needs to be modelled over the whole Cenozoic, with changing distributions of rain forest being determined not only by climate, but also the successive drowning and re-exposure of the Sunda shelf (Morley, 2000; Figure 2.2).

2.1.3 Origin of tropical rain forest

Since the Eocene, the collision of India with Eurasia and the movement of the Australian plate towards the equator have brought floras and faunas with very different histories into juxtaposition (Hall, 1998, 2002). Also, evidence from megafossil plants indicated the existence of "monsoon forests" in Asia as far back as the Eocene (Guo, 1993). Yet, everwet rain forests were rare in Southeast Asia during the Oligocene and Early Miocene, as the climate was predominantly seasonal and dry. By the Middle Miocene, the climate had changed to a perhumid one; reef building coral occurred at much higher latitudes (Wilson & Rosen, 1998) and the rain forest spread across the region (Morley, 2000).

In Southeast Asia rain forest, the tree family Dipterocarpaceae dominates forests in Borneo, Sumatra, Java, and the Malay Peninsula, as well as the wetter parts of the Philippines, with the majority of the large trees being the members of this family (Primack & Corlett, 2005). The dipterocarps appear to have reached Southeast Asia from Africa via the Indian plate, as evidenced from their pollen (Muller, 1981), wood record and biogeography (Ashton, 1982; Ashton & Gunatilleke, 1987). But they did not become dominant and widespread in rain forests until Southeast Asia became

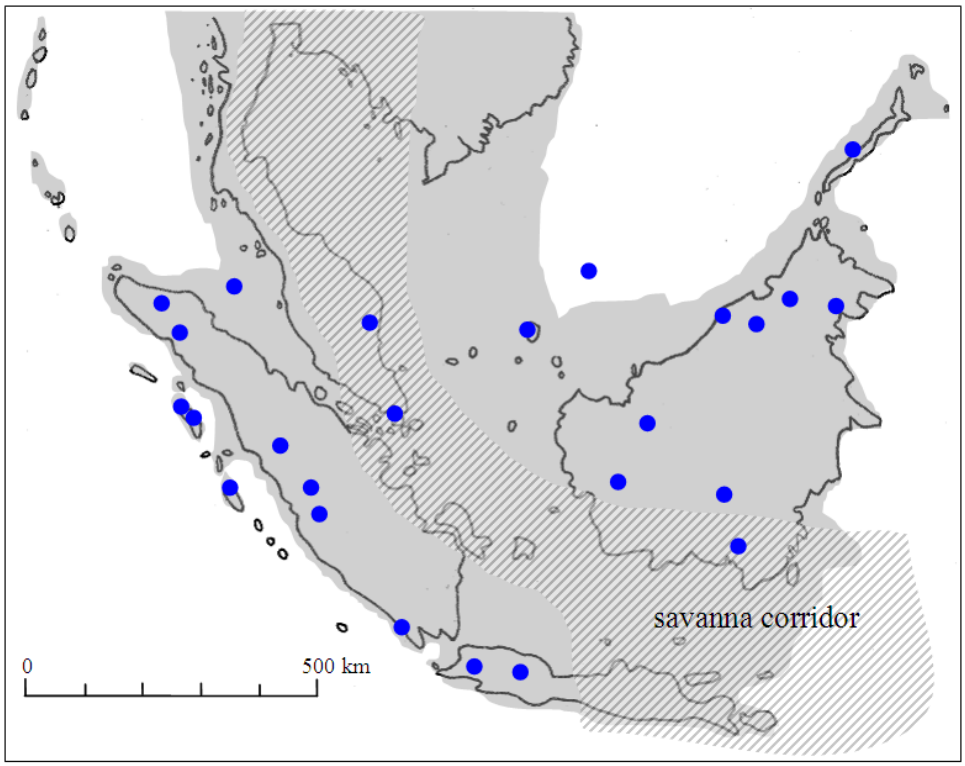

Figure 2.2: Map of Sundaland at the Last Glacial Maximum (adapted from Bird *et al.*, 2005), showing "savanna corridor" proposed by Heaney (1991). Glacial refugia sites inferred from palynology, geomorphology and biogeography data are denoted by circles (Emmel & Curray, 1982, Bird *et al.*, 2005; Quek *et al.*, 2007).

progressively wetter in 10 Ma, following the intensification of the monsoons associated with the uplift of the Himalayas and the Tibetan Plateau (Ashton, 1988; Filippelli, 1997). However, the climate oscillations have caused the repeated expansion and contraction of the rain forest. During the hypothermals period of Pliocene and Pleistocene, it is thought that pine woodlands and savanna grasslands expanded and replaced rain forests over large areas (Morley, 2000).

2.1.4 Evidence of climate change and areas of continuing rain forest

The Pleistocene was marked by cycles of extensive cooling and drying of global climates that may have constricted rain forests worldwide to isolated refugia (Haffer, 1987; Colinvaux *et al*., 1996; Morley, 2000). With the advent of Pleistocene glaciations, dramatic decreases of a planktonic foraminifera, *Globigerinoides ruber*, indicate that the winter monsoon became episodically stronger about every 400,000 year (Jian *et al*., 2003), which probably led to increased drought in the centre of the Sunda shelf. Most geological (Verstappen, 1992, 1997; Thomas, 2000), palynological (van der Kaars & Dam, 1995; Morley, 2000), fossil and modern animal distributions (Medway, 1972) suggested that seasonal and savanna-like system would perhaps be the main type of vegetation in Southeast Asia, including Peninsular Malaysia, West Java, Southern and Northern Sumatra, Sarawak and West Kalimantan. Particularly, Heaney (1991) postulated a wide "savanna corridor" extending down Peninsular Malaysia and across the flooded region between Borneo and Java, which was flanked east and west by tropical rain forest (Figure 2.2).

Although the whole of Southeast Asia was likely to have felt the effects of the increased drought and seasonality, many areas of everwet rain forest still remained as refugia, including Mentawai islands, Northern Sumatra, Brunei, Sabah, Northern Sarawak, East Peninsular Malaysia, East and West Java (Figure 2.2; Ashton, 1972, 1995; Stuijts, 1993; Brandon-Jones, 1998; Newsome & Flenley, 1988; van den Bergh, 1999; Cranbrook, 2000; Morley, 2000; Quek *et al*., 2007). In principle, it is likely that these areas had sufficient moisture to remain as rain forest for two reasons (Gathorne-Hardy

et al., 2002): (1) proximity to the continental shelf and therefore to the sea, even at low sea levels, allowing convergent rainfall with moisture from the sea; and (2) altitude, allowing the interception of the weak summer monsoon and fog, and as they were cold, forcing precipitation. Herein, particularly in Malaysia and Sumatra, mountains were certainly important for the survival of rain forests through the Pleistocene, when the lowlands might have been invaded by drier vegetation during the north temperate glacial intervals (Quek *et al*., 2007).

2.1.5 The persistence of rain forest refugia in Peninsular Malaysia

Most palynological and geological evidence postulated that rain forests disappeared almost entirely from Peninsular Malaysia which was then vegetated with open pine woodlands and savanna grasslands (Verstappen, 1997; Morley, 2000; Thomas, 2000; Brandon-Jones, 2001). However, Kershaw *et al*. (2001) believe that the Sunda shelf was still covered with rain forests during the LGM. Instead, they found evidence of rain forests between Peninsular Malaysia and Borneo. Although these forests might have been restricted to riverbanks in a savanna–grassland matrix, it is clear that the rain forests did not disappear completely during the LGM. Moreover, Quek *et al*. (2007) contended that the distribution of mutualistic *Crematogaster* ants and *Macaranga* trees in the equatorial rain forests of Southeast Asia has provided evidence of small rain forest refuge in the east coast of Peninsular Malaysia. In addition, Geyh *et al*. (1979) provided evidence of perhumid climatic conditions in the southern Malacca Strait before the sea level arose following the LGM. Also, Emmel and Curray (1982) suggested abundant vegetation in the emergent Malacca Strait, probably resembling the lowland vegetation of tropical regions, with mangroves in the low-lying areas and Nipah palms along the banks of muddy creeks. Tropical rain forests would cover the higher drier parts of this area. It is therefore possible that wet conditions and tropical, closed forests prevailed in the west of Peninsular Malaysia.

2.2. Tracing the geographic origin

Tracing the geographic origin would require a priory sampling and profiling of reference populations throughout the distribution area of the target species (Nielsen & Kjaer, 2008). Particularly, the feasibility to trace an unknown sample to its geographic origin strongly depends on the sample size, number of variable markers and spatial genetic structure of the reference populations (Degen *et al*., 2001; Cavers *et al*., 2005; Nielsen & Kjaer, 2008). A clear example is shown in the case of European oak (Deguilloux *et al*., 2003). According to Deguilloux *et al*. (2003), the traceability test is applied to the context of the French barrel industry as French coopers have developed a certification scheme that guarantees the barrels are made only with French oak wood. Although it seems possible to trace oak wood to its presumed provenance regions by the use of chloroplast DNA markers, a drawback is that the resolution of these markers is not high enough to partition between smaller regions or single forest. In most cases, it will probably be difficult to reveal the correct origin down to population level, but the system may prove valuable for controlling purposes as long as some between-population variation exists (Nielsen & Kjaer, 2008). In practice, the analysis should actually not identify the correct origin, but rather test if the announced origin may be false, where the methods will be valuable to verify the geographical origin (Nielsen & Kjaer, 2008). Besides, an assignment test based on nuclear short tandem repeat could be applied to check the conformity of genetic composition of suspected samples with the announced geographic site. It is possible to obtain significant power in a test of excluding a given origin if sufficient reference populations are sampled (Nielsen & Kjaer, 2008).

2.2.1 Issues of illegal logging, forest certification, eco-labelling, and chain of custody certification

In recent years, the declining of forest coverage is a major global environmental concern. During the past three decades, almost three billion hectares, half of the original forest covers worldwide, have been destroyed (Visseren-Hamakers & Glasbergen, 2007). Large areas of forest are either being lost to conversion for agriculture or degraded

through poor logging practices without regard to sustainability and biodiversity (Asia Forest Partnership, 2005). In particular, one of the major threats is uncontrolled illegal logging and the timber trade. The impacts of illegal logging are numerous and worrisome, which include billions in lost revenue to a nation and forest-dependent communities. In some producer countries, the amount of illegally produced timber exceeds the legally produced amount. For instance, it is estimated 80% and 70% of the logging in Brazil and Indonesia respectively are illegal (Richert, 2003). Also, it is estimated that 50–80% of the timber traded in the market from Amazonian, Central Africa and Southeast Asia is of illegal origin (Toyne *et al*., 2002; Greenpeace, 2003). Numerous initiatives have therefore been taken to tackle the issues, including the implementation of the Forest Law Enforcement, Governance and Trade (FLEGT) action plan, where only those timber products with a proof of legality will be permitted in the European market (Commission, 2003).

With regard to forest sustainability, various forest certification schemes have been developed for sustainable industrial logging. The best-known schemes are the Forest Stewardship Council (FSC), the Programme for the Endorsement of Forest Certification Schemes (PEFC), the Canadian Standards Association (CSA), the Sustainable Forestry Initiative (SFI) and the Malaysian Timber Certification Council (MTCC). Recently, a total of 73 million hectares of forests were reported to be FSC certified (FSC, 2006), while over 187 million hectares of forests are PEFC certified (including forests certified under the CSA and SFI) (PEFC, 2005). The FSC is the most stringent, inclusive and performance- based scheme, while the PEFC is the world's largest certification scheme (Visseren-Hamakers & Glasbergen, 2007). The effectiveness of the certification scheme has played a major role in the implementation of sustainable forest management and even replaced the role of governments in forest governance. In most countries, notably forest certification has become an obligation in the logging forestry, for example, 87% of the logging companies researched in USA were certified (Dyke *et al*., 2005). However, these trustworthy certification schemes are not easy to implement in large parts of the world.

Ecolabels have proliferated in recent years to help consumers make informed choices by offering information on the environmental impact of particular products or services (Gulbrandsen, 2005). Particularly, the ecolabelling practice relies mainly on the more professional customers who have expressed a preference for environmentally friendly products, and have urged producers to adhere to sustainable management practices (Gulbrandsen, 2005). Ecolabelling in forestry is initiated by the FSC to promote sustainable forest practices and encourage consumers to support such practices by buying labelled forest products. One important feature of the forest product labelling system is the opportunity to track the origin of the products through every stage of the supply chain, which is referred to the as "chain of custody". The FSC has issued more than 4,200 chain of custody certificates to manufacturers, traders and retailers according to more rigorous standards (FSC, 2005a). In part, the FSC has made sure the solid wood products bearing FSC label should contain 100 percent of certified wood, while chip, fibre, and component products should have at least 70 percent FSC-certified content (Cashore *et al*., 2004). However, many forest companies have encountered problems in meeting these strict labelling requirements as their timbers are sourced from both certified and non-certified forest owners. As a result, more lenient labelling schemes, such as chain of custody certificate based on a "percent in, percent out" approach was launched by the PEFC (Cashore *et al*., 2004). The more lenient industry-dominated schemes may generally increase the availability of labelled timber products in the market place, but they seem to confuse consumers and may even reduce the trust in well-established labels in the long run (Jordan *et al*., 2003).

2.2.2 Authenticity testing and databases for traceability

At the present day, the concerns about food safety, animal health, international trade, consumer awareness, and improving supply chain management have made authenticity testing and databases for traceability essential. In the food industry, there is growing enthusiasm among consumers for high-quality food with clear regional identity (Kelly *et al*., 2005). For instance, multi-element and multi-isotopic measurements have been widely applied to determine the geographical origin of beef (Hegerding *et al*., 2002),

lamb (Piasentier *et al*., 2003), milk (Renou *et al*., 2004), coffee (Weckerle *et al*., 2002) and rice (Yasui & Shindoh, 2000). In Australia, the most inclusive "National Livestock Identification System" has been created to allow an individual animal to be traced from its property of birth to its slaughter destination (Tonsor & Schroeder, 2004). Similarly, a species database of fish, molluscs and crustaceans has been established with the aim to identify species of origin of seafood products (Maldini *et al*., 2006).

In forestry, the implementation of chain of custody certification for legal timber traceability has greatly improved the sustainability of forest management. Low technology tools such as paint and ultraviolet paint have been widely used to mark and track the flow of timbers. However, these relatively inexpensive tools are not particularly useful in broader-based or national-level efforts to combat illegal logging or trade. More high technology tools involve a combination of databases; the physical tagging of logs and some form of spot checking are required in national tracking systems (Smith, 2004). Still, most of these timber tracking systems rely on tagging or certificates of origin issued in the source countries to verify the legality of the timbers which are susceptible to falsification in many developing countries (Carr, 2007). Spectrometry and isotopic analyses have been used to reveal provenance (Perez-Coello *et al*., 1997; Durand *et al.*, 1999; English *et al*., 2001; Kagawa *et al*., 2007), yet they are limited by inconsistent results (Hoffman *et al*., 1994; Towey & Waterhouse, 1996). With the advent of molecular techniques, the use of DNA markers to identify geographic origin of timber offers many advantages over methods currently used to verify whether a timber product has been harvested from legitimate source.

The DNA track-back system can be applied at the point of consumer country importation, and can be used to overrule questionable certification documentation that may have been introduced along the supply chain (Lowe, 2007). There are several good examples of timber tracking studies that have been undertaken, for instance Simmonds Lumber, one of Australia's largest timber importing firms has implemented DNA testing to stem the import of illegally logged or endangered timber (Carr, 2007). A

timber verification company, Certisource, located in Singapore has developed a DNA-based technique to verify the legality of *Intsia palembanica* (Nielsen & Kjaer, 2008). In Canada, the British Columbian Ministry of Forests has developed DNA markers and DNA bank for *Thuja plicata* for the purpose of tracking illegally harvested timber (an article published in The Vancouver Province, August 2006; Nielsen & Kjaer, 2008). As a whole, the accessibility of such DNA databases and tracking systems is extremely important in the timber trade and industry, yet a significant challenge is faced with the aim to genotype the large number of timbers in the market place (Carr, 2007).

2.3 DNA extraction from dry wood

Wood can be a suitable material to be used in molecular genetic analyses if appropriate DNA quality for PCR amplification is obtainable. In fact, the feasibility to perform molecular genetic analyses on dry wood may have great prospects in forest certification, wood species and taxonomic identification, forest crime prosecution, archaeology, paleobotany or phylogeography (Deguilloux *et al*., 2002). Notably, the possibility of applying these molecular analyses to dry wood would be of considerable interest in many fields. Detailed studies on the potential of dry wood as source of DNA have been reported for oak (Dumolin-Lapègue *et al*., 1999; Deguilloux *et al*., 2002), *Gonystylus bancanus* (Asif & Cannon, 2005), and dipterocarps (Rachmayanti *et al*., 2006, 2009; Yoshida *et al*., 2007). However, a simple evaluation of basic information on wood anatomy, structure and biochemistry makes it obvious that the adaptation of usual DNA extraction methods used for fresh tissues to dry wood will not be straightforward (Deguilloux *et al*., 2002).

2.3.1 General consideration in plant DNA extraction

The quality and quantity of the extracted plant DNA are greatly influenced by the condition of the starting material (Weising *et al*., 2005). Fresh and young tissues are generally preferable. Difficulties may arise when the plant species has to be sourced from remote areas, or the plant material has to be stored for longer periods, e.g. weeks, months or even several years. Nevertheless, empirical studies have been proposed to

optimize field collection and preservation of plant material, for instance rapid drying of plant material using silica gel (Chase & Hills, 1991), chemical preservation of plant material in ethanol, CTAB–NaCl solution (Rogstad, 1992) or CTAB–NaCl–azide solution (Bhattacharjee *et al*., 2004).

Due to the extreme sensitivity of PCR, extra precaution is necessary to avoid contamination of the plant material with other organisms (Taylor & Swann, 1994). Fungi, insects, dirt, lichens and epiphytes could be the potential contaminants (Weising *et al*., 2005). For example, Staub *et al*. (1996) discovered different DNA profiles when extracting DNA from plants infected by fungal parasites, as compared with uninfected plants. These external contaminants could be removed through surface sterilization. However, contamination by endophytic fungi which are widely present in most plants, yet without any obvious symptoms, is almost unavoidable (Saar *et al*., 2001; Schultz *et al*., 1993). Additionally, another potential source of the contaminant is from carry-over of alien PCR product (Weising *et al*., 2005). This could create a severe problem especially when dealing with a small amount of poor quality DNA. Nonetheless, the set-up of fully dedicated rooms and equipment for pre- and post-PCR, laminar flow hood and UV light irradiation, and bleach cleaning of every surface could be taken to avoid such carry-over events (Deguilloux *et al*., 2002).

Different biochemical compositions found in plant tissues and species have necessitated the development of extraction methods tailored specifically for each plant species (Weising *et al*., 2005). Large numbers of plant DNA extraction protocols have therefore been published. Most of the protocols are designed to extract total cellular DNA, which are applicable to all PCR-based marker methods. However, there are some protocols that are tailored to extract solely nuclear DNA (Watson & Thompson, 1986), chloroplast DNA (Dally & Second, 1989), and mitochondrial DNA (Wilson & Choury, 1984).

In most DNA extraction protocols, after tissue disruption and homogenization, liberation of DNA from nuclei or organelles is performed using extraction buffer, which also functions as lysis buffer (Weising *et al*., 2005). Typically, it comprises (1) detergents to destroy membranes, denature proteins and dissociate proteins from DNA, e.g. CTAB and sodium dodecyl sulphate (SDS), (2) buffer system, e.g. Tris–HCl to maintain pH and prevent the optimal activity of degrading enzymes, (3) high salt concentrations, e.g. more than 1 M of NaCl to separate nuclear proteins from DNA (Aljanabi & Martinez, 1997), retain polysaccharides in solution during ethanol precipitation and salt out PCR inhibitors (Cheung *et al*., 1993), (4) reducing agents, e.g. β-mercaptoethanol, to protect DNA against oxidation processes and (5) chelating agents, e.g. ethylenediaminetetraacetic acid (EDTA), to detain bivalent metal ions that stimulate metal-dependent DNases that are released from the cells. The dissolved proteins and polysaccharides are subsequently removed by one or more rounds of extraction with either phenol-chloroform or chloroform-isoamyl alcohol mixtures (Weising *et al*., 2005). To precipitate DNA, isopropanol is preferably used rather than ethanol, because isopropanol has been reported to dissociate high molecular DNA from polysaccharides and only small volumes (0.6 volumes) are needed (Dellaporta *et al*., 1983). RNA is often co-extracted along with DNA; however, it is usually removed by an RNase treatment in either early or late step in the DNA extraction protocol (Weising *et al*., 2005).

Extraction of DNA is difficult when the plant species and tissues contain higher amounts of polysaccharides and secondary metabolites such as phenols, alkanoids, terpenes and flavonoids. Pectin-like polysaccharides are often water soluble and usually co-extracted with DNA (Weising *et al*., 2005). They often cause highly viscous DNA solutions that inhibit the activity of restriction enzymes (Do & Adams, 1991) and *Taq* DNA polymerase (Demeke & Adams, 1992). Nevertheless, Barker *et al*. (1999) reported that limiting the heat incubation during CTAB extraction to a maximum of 15 min and precipitation of the DNA at room temperature would leave most polysaccharides in the supernatant. In terms of secondary metabolites, phenols and

polyphenols are easily oxidized by intrinsic enzymatic activities and resulting quinonic compounds (oxidizing agents) which cause damage to DNA and proteins, browning of DNA solutions and render the DNA inaccessible for some enzymes (Weising *et al*., 2005). However, the inclusion of polyphenol absorbents such as bovine serum albumin (BSA), soluble polyvinylpyrrolidone (PVP), and phenoloxidase inhibitors, e.g. diethyldithiocarbamic acid (DIECA), in the extraction buffer should reduce the detrimental effects of polyphenols and their oxidation products (Weising *et al*., 2005).

2.3.2 Structure of wood

Wood is composed mainly of hollow, elongated, spindle-shaped cells that are arranged parallel to each other along the trunk of a tree (Forest Products Laboratory, 1999). A cross- section of a tree illustrates the following well-defined features (from the outside to the centre): outer bark, inner bark, cambium, sapwood, heartwood, and the pith as shown in Figure 2.3. A detailed review of the structure and function of each wood tissue has been given by the Forest Products Laboratory (1999) and Wiedenhoeft and Miller (2005). Beginning from the outside of a tree, the outer bark helps to limit evaporative water loss and provides mechanical protection to the softer inner bark. The inner bark, also known as phloem, is the tissue through which sugars produced by photosynthesis are translocated from the leaves to the roots or growing portions of the tree. The cambium, the layer between the inner bark and the wood (xylem), is responsible for producing both phloem and xylem tissues. Sapwood containing both living and dead tissues is responsible for conduction of sap from the roots to the leaves and also heartwood formation. Heartwood is made up of inactive cells and is found as a core of lighter or darker-coloured wood in the middle of most trees, which is formed by a gradual change in the sapwood. The pith located at the center of trunk is the remnants of the early growth of the trunk. In particular, wood or secondary xylem is formed by a succession of five major steps, including cell division, cell expansion (elongation and radial enlargement), cell-wall thickening (involving cellulose, hemicellulose, cell-wall proteins, and lignin biosynthesis and deposition), programmed cell death, and lastly heartwood formation (Plomion *et al*., 2001).

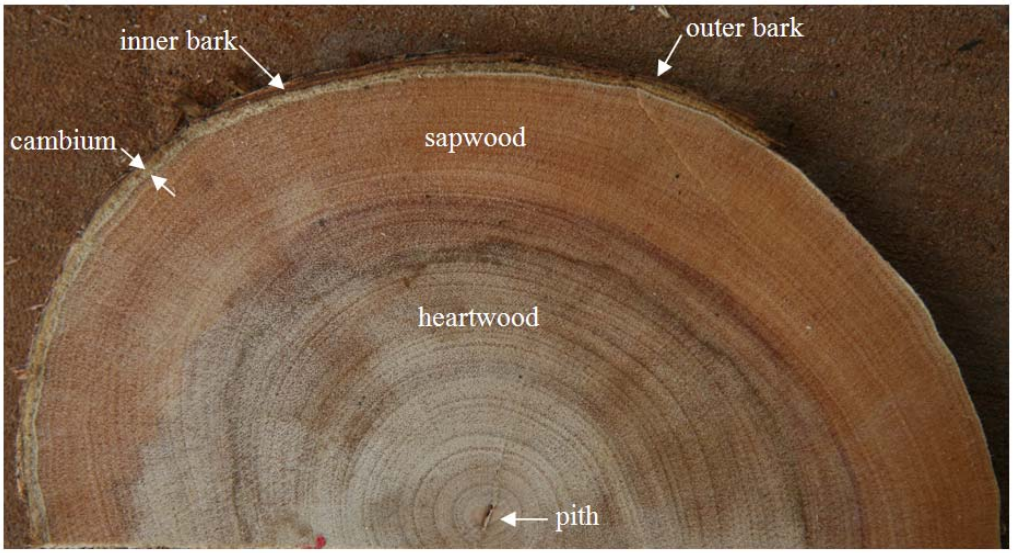

Figure 2.3: Cross-section of a *Neobalanocarpus heimii* trunk: outer bark, inner bark, cambium, sapwood, heartwood and pith.

2.4 Chloroplast DNA marker

In the last few decades, many studies have been initiated to determine the complete DNA sequences for nuclear, mitochondrial and chloroplast genomes. For instance, all three genomes of the plant model organism, *Arabidopsis thaliana* (Unseld *et al*., 1997; Sato *et al*., 1999; The *Arabidopsis* Genome Initiative, 2000) and rice (Hiratsuka *et al*., 1989; Goff *et al*., 2002; Notsu *et al*., 2002) have been fully sequenced. In plants, the nuclear genome comprises a high proportion of non-functional DNA, which consists mainly of tandem repeats and may be considered as junk DNA (Graur & Li, 2000). The mitochondrial genome is characterized by a large size, slow nucleotide substitution rates and extensive levels of intramolecular recombination; thus it has limited use in the ecology and evolutionary history of plant species (Provan *et al*., 2001).

In contrast, the chloroplast genome has been used extensively to address plant evolutionary issues. The chloroplast is postulated to arise from endosymbiosis between a photosynthetic bacterium and a non-photosynthetic host (Howe *et al*., 1992). In general, chloroplast genome codes for key functions of photosynthesis in most plants. It is highly abundance in the plant cell, for instance a cell of a plant leaf usually contains 400 to 1,600 copies of chloroplast genome (Pyke, 1999). Most chloroplast genomes are circular DNA molecules, typically range from 120 to 200 kbp in length and have a quadripartite organization of two copies of inverted repeats, which divides the rest of the genome into a large-single-copy and a small-single-copy-region (Freeland, 2005; Yang, *et al*., 2010). It is maternally inherited in most angiosperms, and paternally inherited in gymnosperms, and notably, the clonal inheritance of chloroplast lineages can be traced over time and space with relative ease (Freeland, 2005). It is coupled with a highly conserved gene order, a general lack of heteroplasmy and recombination (Olmstead & Palmer, 1994). Furthermore, the chloroplast genome evolves very slowly, which is four times slower than nuclear DNA (Wolfe *et al*., 1987). Hence, primers can be designed with the purpose of working across species to study molecular variation among closely related species, or among separate sets of populations within species, by analysing intron and spacer (Heinze, 2007). Additionally, since chloroplast genome is

haploid, its effective population size is half that of diploid nuclear DNA; thus cpDNA markers could be a good indicator of historical bottleneck, founder effects and genetic drift (McCauley, 1995; Ennos *et al*., 1999).

The development of molecular techniques using chloroplast genome has had a great impact on many aspects of ecology, history and evolution of plant populations. In 1991, Taberlet *et al*. (1991) recommended the use of universal primers in chloroplast genome across plant genera, and for the intention to study species intra-specific variation. Since then, the approach has been widely adopted and has accelerated molecular research tremendously. *Nicotiana tabacum* is reported as the first fully sequenced chloroplast genome (Shinozaki *et al*., 1986). To date, there are numerous fully sequenced chloroplast genomes deposited in GenBank, including *Atropa belladonna, Spinacia oleracea, Arabidopsis thaliana, Oryza sativa, Pinus thunbergii, Marchantia polymorpha, Zea mays, Oenothera elata, Acorus calamus, Eucalyptus globulus* and *Medicago trunculata* (Heinze, 2007). By using *Nicotiana tabacum* as an example (Figure 2.4), the chloroplast genome is made up of approximately 113 genes, containing 21 ribosomal proteins, 4 ribosomal RNAs, 30 transfer RNAs and 29 genes that are necessary for functions associated with photosynthesis and 11 genes that are involved with chlororespiration (Sugiura, 1992).

In recent years, the choice of appropriate chloroplast markers has been facilitated by the growing availability of conserved oligonucleotide primers and complete chloroplast genome sequence for both coding and non-coding regions (Heinze, 2007). The degree of 'universality' of these chloroplast primers would virtually amplify all terrain plants and many algae species (Taberlet *et al*., 1991). Particularly, non-coding regions of chloroplast genome such as *trn*L intron, *trn*L-F, *trn*T-L, *atp*B-*rbc*L, *pet*G-*trn*P and *pet*A-*psb*J intergenic spacers have been widely sequenced to survey the plant population structure and phylogeography (Huang *et al*., 2004; Chung *et al*., 2007; Ikeda & Setoguchi, 2007). Although low levels of genetic variation were reported in chloroplast genome, virtually all recently published plant phylogeography studies have

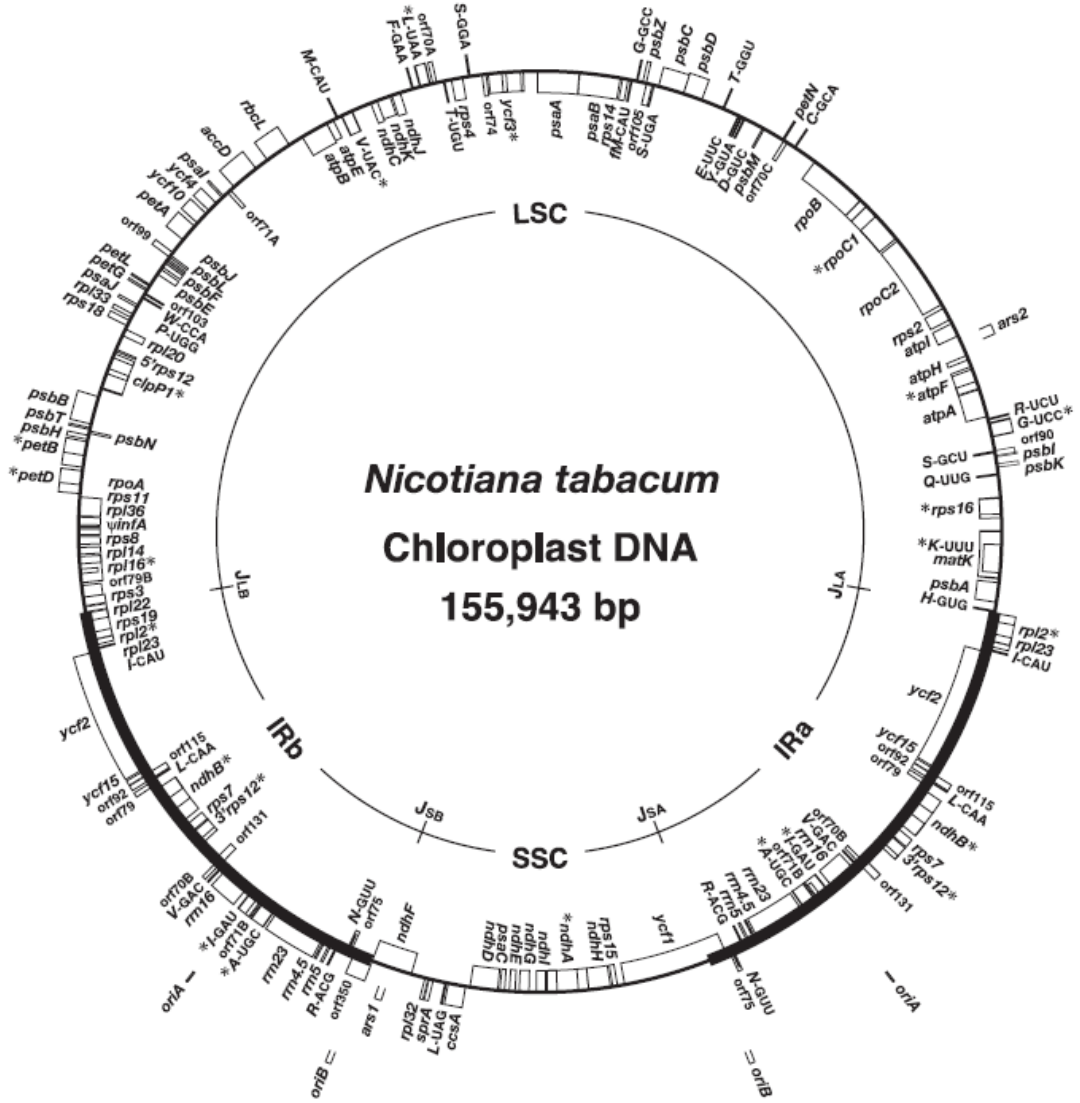

Figure 2.4: Gene map of the chloroplast genome from *Nicotiana tabacum*. Genes shown on the inside of the circle are transcribed clockwise, and those on the outside are transcribed counter-clockwise. Genes for transfer RNAs are represented by 1-letter code of amino acids with anticodons. Asterisks denote split genes. Open-reading frames are shown by orf plus codon number (≥70 codons). When two genes overlap, the one that is located downstream or inside the other gene is displayed with a lower-height box (by Yukawa *et al*., 2005).

benefited by data from more than one chloroplast region, which allow for finer phylogeography resolution. Small insertions/ deletions (indels) are relatively frequent when compared with base substitutions (Heinze, 1998). For highly conserved exon sequences, e.g. *rbc*L, *mat*K, *ndh*F and *rpl*16, they have been widely used to establish and verify phylogenies (Heinze, 2007). Also, these chloroplast regions have been proposed as a short cut that would provide species identification, namely "DNA barcoding" (Rubinoff *et al*., 2006).

Chloroplast DNA (cpDNA) markers are generally preferred in the majority of plant phylogeographical studies. Based on intraspecific polymorphisms in cpDNA, three different marker types were used (Weising *et al*., 2005): (1) PCR restriction fragment length polymorphisms (PCR-RFLP) or cleaved amplified polymorphic sequences (CAPS), (2) chloroplast microsatellites (cpSSR) and (3) direct DNA sequencing of one or several cpDNA regions.

PCR-RFLP or CAPS are generated through PCR amplification of a defined cpDNA region, followed by digestion with a restriction enzyme (Konieczny & Ausubel, 1993). The digested amplification products may or may not reveal polymorphisms after separation on agarose gel. PCR-RFLP has been studied extensively in plants and has proven to be valuable for molecular systematic and phylogeographic studies (Clegg, 1993; Jansen *et al*., 1998; King & Ferris, 1998). However, because only a subset of base substitutions is targeted, and small indels events may escape detection, PCR-RFLP markers are therefore less informative than direct sequence analysis of PCR product (Weising *et al*., 2005).

For cpSSR, it has been found in all completely sequenced plant genomes and also in a huge number of partial chloroplast sequences (Provan *et al*., 2001). In the chloroplast genome of higher plants, cpSSR is made up of short poly (A) or poly (T) tracts, with maximum sizes of about 20 bp (Weising & Gardner, 1999). The amplified cpSSR products exhibit length variation, which corresponds to the expansion or

contraction of the repeat region (Weising *et al*., 2005). Extensive levels of length polymorphism in cpSSR loci have been observed in species of *Glycine* (Powell *et al*., 1995a), *Pinus* (Powell *et al*., 1995b) and *Abies* (Parducci *et al*., 2001). An investigation based on rice species (*Oryza*), which had previously been examined using PCR-RFLPs, revealed a fourfold increase in levels of diversity when the species were studied using cpSSR (Provan *et al*., 1997). However, regardless of the hypervariability, cpSSR is confounded by the problem of length homoplasy. Technically, only direct sequencing of cpDNA would circumvent the problem of length homoplasies which usually occur when using cpSSR (Parker *et al*., 1998). Generally, direct DNA sequencing provides highly robust, reproducible and informative datasets, though it can be prohibitively tedious and expensive when large amounts of samples have to be assayed (Weising *et al*., 2005). In recent years, the availability of complete chloroplast genomes has allowed for the design of universal PCR primers which have greatly facilitated DNA sequencing and is widely used in phylogeography study (Cannon & Manos, 2003; Huang *et al*., 2004; Cheng *et al*., 2005; Shephard *et al*., 2007).

2.5 Description of *Neobalanocarpus heimii* (King) Ashton

Neobalanocarpus heimii (King) Ashton belongs to the family Dipterocarpaceae, the main timber family in Southeast Asia (Figure 2.5). It is locally known as chengal and is widely distributed throughout Peninsular Malaysia (Symington, 1943) except in the states of Perlis and Melaka. It is distributed in diverse localities, on low-lying flat land as well as on hills up to 900 m (Symington, 1943). 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).

Symington (1943) gave a detailed description of *N. heimii*, where the chief diagnostic characters of *N. heimii* are the dark-brown-coloured bole, sometimes with purplish tinge, longitudinally and more or less regularly-fissured. The outer bark is hard and rather brittle, frequently showing small exudations of almost colourless dammar the

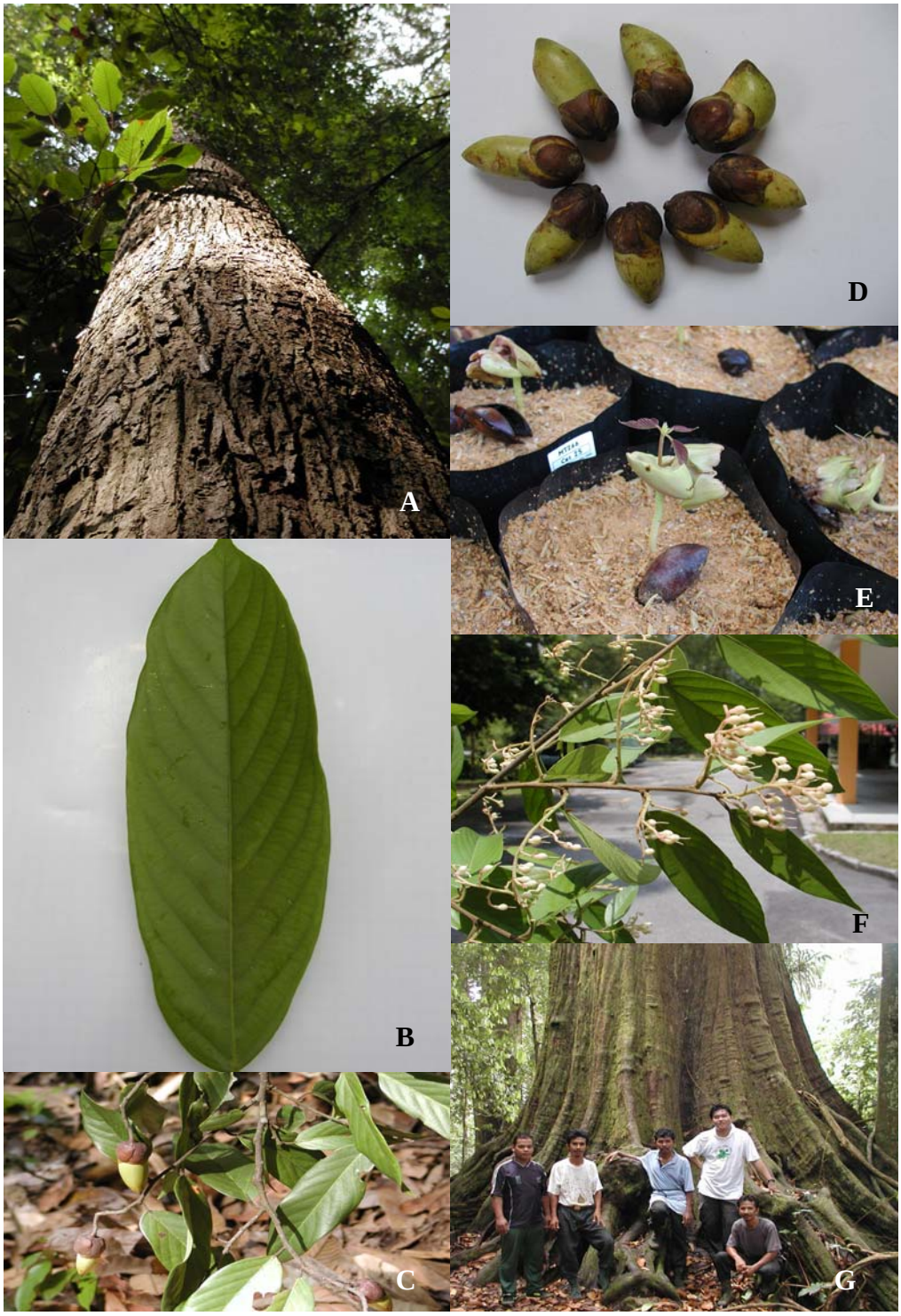

Figure 2.5: Morphological characteristics of *Neobalanocarpus heimii*. (A) the mature tree; (B) leaf; (C) fruits of the mature tree; (D) intact seeds; (E) seedlings; (F) stipules and (G) the largest tree of *N. heimii* in Pasir Raja Forest Reserve, Peninsular Malaysia.

edge of fissures or where there has been an injury (Foxworthy, 1927). The tree's large buttresses are usually found prominent at only a few feet above the ground (Ashton, 1982). The leaves are simple, alternate, thinly coriaceous and oblong acuminate with rounded base. The flowers are hermaphrodite, in pale yellowish colour and arranged insmall axillary clusters. The fruits are large, wingless, ovoid shaped, apiculate and distinctly bent to one side (Foxworthy, 1927; Symington, 1943; Soerianegara & Lemmens, 1994).

Neobalanocarpus heimii flowers are in synchrony for about two weeks annually (Appanah & Weinland, 1993), and the flowers open in the early hours of the morning, with bees (*Apis* sp. and *Trigona* sp.) being the main flower visitors and likely pollinators (Symington, 1943). Fruiting is staggered over a long period after flowering; fruit-fall can start about four months after flowering, and the process can go on for another year after that, with a few fruits falling every day (Élouard *et al*., 1994). The heavy, wingless seeds have no means of dispersal beyond the area covered by the crown of the parent tree, except, perhaps, by rolling down hill slopes or being carried by animals (Symington, 1943). Direct estimation of gene flow using STR markers showed that the pollen dispersal of *N. heimii* in a lowland dipterocarp forest was moderately extensive (mean pollen flow distance = 524 m), which might be responsible for the observed significant spatial genetic structure among the 30 potentially flowering trees (Konuma *et al*., 2000).

In economic terms, *N. heimii* produces the naturally durable, heavy construction or "primary hardwood" timbers with a density of 915–980 kg m^{-3} air dry (Wong, 2002). In Peninsular Malaysia, it is the best-known timber having a deservedly high reputation. This timber is also used as a standard by which other timbers are judged. The timber is fully discussed by Desch (1941), who groups it on the basis of its anatomical similarities with some species of *Hopea*. *Neobalanocarpus heimii* although dense is easy to work with and no preservation treatment is necessary. It is among the strongest timbers in the world, being 50% stronger than teak. Furthermore, it is resistant to

termites and fungi. The timber is suitable for all forms of heavy construction, bridges, railway sleepers, boat, building, wharves, power-line poles and wherever strength is considered essential (Thomas, 1953; Wong, 2002). In 2007, the export of sawn *N. heimii* timber was 925 m^3 with a value of RM1,961,739 (Malaysia Timber Council, 2008).