



Phylogeography of the Southeast Asian stone oaks (*Lithocarpus*)

Charles H. Cannon* and Paul S. Manos *Department of Biology, Duke University, Durham, NC, USA*

Abstract

Aim To describe current geographical patterns of genetic diversity and infer the historical population dynamics of the stone oaks (*Lithocarpus*) in Southeast Asia.

Location We sampled three populations in Indochina: (1) Yunnan province, China; (2) Pyin Oo Lwin area, Myanmar and (3) north-western Vietnam; two in western Borneo: (1) South-western Sarawak and (2) West Kalimantan, Indonesia; two in central Borneo: (1) north coastal Sarawak and (2) north-eastern Sarawak, Malaysia; and two in northern Borneo: (1) Central Sabah and (2) Northern Sabah, Malaysia.

Methods A phylogenetic reconstruction of chloroplast DNA sequence variation from numerous individuals of multiple species was used to determine geographical distribution of genetic diversity. A resampling scheme was used to determine the significance of these patterns at different hierarchical levels of the phylogeny. Results were compared with a previously published set of nuclear DNA sequence data.

Results A high level of chloroplast sequence variation was found, which was divided equally between two major clades separated by four non-homoplasious changes. One clade was confined to the island of Borneo, while the other was widespread. Strong geographical structure was observed in the chloroplast sequence variation. The Indochinese populations were much more closely related than expected, comparable with the highly endemic and isolated population on the western coast of Borneo. Conversely, individuals from the Kelabit Highlands were found to be more distantly related than expected. The highest levels of genetic endemism were observed in western Borneo. More geographical structure was observed in the Bornean clade than in the Widespread clade, because of limited genetic diversity in the Widespread clade. Relatively weak geographical structure was found in the nuclear sequence variation: only populations in southern China and central Sabah were significantly related.

Conclusions The high levels of chloroplast genetic diversity and the persistence of an ancestral haplotype that is a single step away from a haplotype found in *Castanopsis* indicates the continuous presence of tropical rain forest in Southeast Asia throughout the evolutionary history of the genus (*c.* 40 Myr). This conclusion is supported by the high frequency of numerous endemic types observed in every population and the relatively few number of 'missing' haplotypes. This situation suggests both limited migration and limited extinction. In contrast, the nuclear genetic diversity contained less geographical structure, indicating that our taxonomic sampling among populations was unbiased and that gene flow mediated through pollen is less geographically restricted and contains less geographical structure than purely seed-mediated (chloroplast) gene flow. The most likely scenario suggested by the evidence involves four major patterns: (1) the widespread presence of an ancestral haplotype; (2) the large degree of separation (four non-homoplasious base pairs) between the types found in the two major clades; (3) the concentration of derived types from both major clades found in central and northern Borneo; and (4) the molecular endemism found in each location. These patterns suggest four primary things about the population dynamics of *Lithocarpus* since the late Eocene: (1) populations have either spanned the entire region throughout much of the

*Correspondence: Department of Biology, Texas Tech University, Box 43131, Lubbock, TX 79409, USA. E-mail: chuck.cannon@ttu.edu

evolutionary history of the genus or substantial populations have persisted in both Indochina and Borneo with limited migration between them; (2) significant fragmentation has occurred subsequently between the Asian mainland and the Melasian island archipelago, leading to independent genetic diversification in both regions; (3) several locations possessing significant independent histories, have experienced little migration and have never gone completely extinct; and (4) that the central highlands of Borneo have been re-invaded from the north and the west. The timing of these events is difficult to ascertain but probably predate the Quaternary Period, suggesting that although the recent ice ages might have affected the overall distribution of rain forest in Southeast Asia, it managed to persist in most regions even through the most dramatic drying events.

Keywords

Fagaceae, chloroplast DNA sequence, *ITS* sequence, Sundaland, Indochina, biogeography, resampling statistics.

INTRODUCTION

Signs of past change in the distribution and connectivity of tree populations on a regional scale have been detected in current patterns of genetic diversity, particularly among seed-mediated molecular markers (Ferris *et al.*, 1995; Dumolin-Lapegue *et al.*, 1997; Petit *et al.*, 1997; Fineschi *et al.*, 2000). While most previous studies of plant phylogeography have been concerned with detecting the effects of climatic change during the Quaternary Period (Comes & Kadereit, 1998), patterns dating back to the Pliocene or Oligocene may still be present in equatorial rain forests, because climatic fluctuations during the more recent period are expected to have been milder than in temperate forests (Flenley, 1998; Morley, 2000). This is particularly true for old-growth forest taxa, with long generation times, possessing fruit with little long distance dispersal potential and intolerance of climatic extremes. For trees with generation times of several centuries, fewer than 100 generations are necessary to reach beyond the last glacial maximum (LGM) and less than 10,000 to reach the upper boundary of the Tertiary Period. Life span is the relevant measure of generation time, not time to first reproduction, because the vast majority of reproductive effort in trees results in complete failure (a tree may be very ancient before it ever successfully reproduces). Adequate sampling of appropriate taxa may provide evidence across several episodes of climate change, not just the most recent.

The vast majority of molecular phylogeographical studies of tree species have focused on Old World temperate forests (Comes & Kadereit, 1998; Ferris *et al.*, 1998; King & Ferris, 1998; Schaal *et al.*, 1998; Taberlet *et al.*, 1998; Tomaru *et al.*, 1998; Ennos *et al.*, 1999; Newton *et al.*, 1999; Oddou-Muratorio *et al.*, 2001; Fuji *et al.*, 2002; Petit *et al.*, 2002) and few studies of tropical species have been performed on a regional or landscape scale (Aide & Rivera, 1998; Lowe *et al.*, 2000; Olsen, 2002); none in Southeast Asia. A previous study of allozyme diversity in *Shorea leprosula* Miq. found little

geographical structure but these markers are prone to balancing selection pressure and no attempt at phylogenetic reconstruction was attempted (Lee *et al.*, 2000).

Southeast Asia, including the island archipelago reaching the island of Papua (Wallace, 1869), is a particularly interesting region for the application of phylogenetic analysis in a biogeographical context. Not only has the region been subject to climate change, but also a correlated change in the distribution of land area has occurred as well (see Fig. 1) through changes in sea level during glacial periods (Inger & Voris, 2001). This additional layer of complexity will not make interpreting patterns any easier but will provide insight across a spectrum of historical situations and varying degrees of connectivity between populations.

The effects of past climatic fluctuations on forest tree communities in Southeast Asia remain ambiguous (Morley, 2000), although several new studies are providing a clearer picture for the Quaternary Period (van der Kaars & Dam, 1997; Dam *et al.*, 2001b). Overall, climate change was less extreme than in temperate zones and forest areas expanded and contracted in size but never vanished entirely (Flenley, 1998). Most records indicate cooling and drying trends at the time of the LGM across a number of sites in the region (Morley, 1998; Anshari *et al.*, 2001; Dam *et al.*, 2001a; Hope, 2001; Penny, 2001; van der Kaars *et al.*, 2001). The further expansion of our palynological knowledge, particularly into the more distant past (Metcalfe *et al.*, 2001), is crucial to a complete understanding of past vegetation dynamics.

The stone oaks (*Lithocarpus*) are a good model group for phylogeographical study in palaeotropical rain forests. They have been an important element of old growth forests in the area, richest in the uplands, since at least the Miocene (Morley, 2001) and presently exist in diverse mixed species communities throughout their range (Young & Herwitz, 1995; Aiba & Kitayama, 1999; Cannon, 2001). Their presence in the pollen record is often used to infer cooler temperatures in lowland sites (Dam *et al.*,

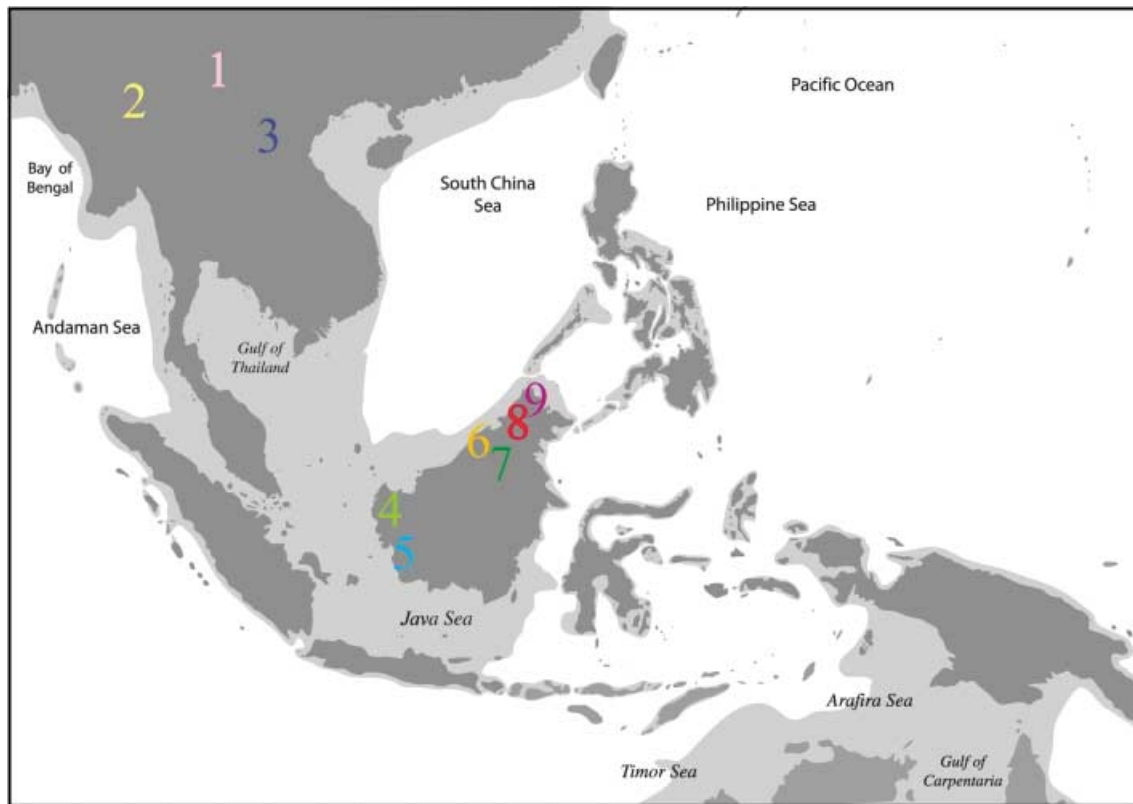


Figure 1 Distribution of the stone oaks and collection sites for this study. Dark landmasses illustrate the majority of the distribution of *Lithocarpus*, with extensions into Japan and Assam not shown. Light grey area represents the distribution of land during the last glacial maximum, when sea levels were lowered by 120 m. Sample collection sites are numbered 1–9, labelled in Table 1, and colour coded as in Figure 2. Adapted from Inger & Voris (2001).

2001a) and current limits to their geographical distribution, e.g. not in the lesser Sunda Islands, restriction to southern tropical and semi-tropical parts of Indochina and coastal east Asia (Soepadmo, 1972; Huang *et al.*, 2000), indicate an intolerance of seasonal climates. They produce single nuts seated in a fused cupule of various forms (Langdon, 1939; Barnett, 1944; Soepadmo, 1970; Cannon & Manos, 2000), which are probably scatter-hoarded by spiny rats and terrestrial squirrels, among other mammals (Leighton & Leighton, 1983; Payne *et al.*, 1985; Corlett, 1998; Vander Wall, 2001). They are closely related to other well-studied taxa, including *Quercus* and *Castanea*, but the relationships among these groups are difficult to resolve (Manos & Steele, 1997; Manos *et al.*, 2001). Finally, they share a wide range of ecological characteristics with other major dominants of palaeotropical forests (Ashton, 1988), including obligate ectomycorrhizal symbioses and a generalized insect pollination system (Sakai *et al.*, 1999; Cannon, 2001).

Although the species biology of *Lithocarpus* is poorly known, interspecific hybridization is potentially an important element in stone oak evolutionary history. In the Fagales, reproductive isolation among closely related species is gen-

erally weak, more so among *Quercus* species (Hardin, 1975). Species-independent and population-dependent distribution of haplotypes has been shown in mixed communities of *Quercus* species in the white oak group (Whittemore & Schaal, 1991; Petit *et al.*, 1993a) and a similar pattern was found in an extensive survey of haplotype distribution in eight species of European white oaks: significant genetic overlap among the most common species related more to geographical location than taxonomic composition of populations (Petit *et al.*, 2002). Because of their close relationship with *Quercus* and the likelihood of poorly isolated species, an infrageneric sampling approach was adopted in this study. While this interspecific phylogeographical approach weakens the statistical power for any one taxonomic unit in the study, this sampling approach avoids possible misinterpretation of biogeographical patterns, given little knowledge of how taxonomic boundaries relate to gene flow in tropical trees.

The following questions are addressed in the current study: (1) What are the levels of variation in chloroplast types across both the geographical and taxonomic range of the genus? (2) Is the variation geographically structured? (3) Is the variation structured in relation to a nuclear marker?

(4) Does the structure relate to recent Quaternary or more ancient events or both?

MATERIALS AND METHODS

Study species and sampling strategy

All stone oak species are medium- to large-sized trees, predominantly found scattered at low densities in old growth forest above 1000 m elevation and also range from 0 to over 2000 m elevation, on most soil types and drainage conditions. All species are monoecious (Kaul, 1986), the male flowers are produced in copious quantities and emit a heavy musky scent which attracts a broad range of insects while visitation of receptive female flowers appears to be infrequent, because of lack of attractive structures and delayed anthesis (Cannon, 2001). Anecdotal evidence about reproductive phenology is limited (Kaul, 1989) but field experience indicates frequent and aseasonal flowering with suprannual production of fruit crops in most species, except the smallest fruited species. The nuts possess no obvious features to facilitate dispersal, although delayed hypogeal germination may occur among the larger fruited species (Ng, 1991). The trees are often poorly formed and not usually considered valuable timber species (Burgess, 1966), particularly because the wood tends to split upon drying because of the multi-seriate transverse rays in the wood (pers. obs.). Historically, their distribution may date back to 40 Myr, including parts of Europe (Kvacek & Walther, 1989) and western North America (Manchester, 1994) but the megafossil record is limited (Herendeen *et al.*, 1995; Sims *et al.*, 1998). *Lithocarpus densiflorus*, often considered a relictual species in the Pacific Northwest, apparently diverged prior to the monophyletic Southeast Asian radiation of the genus (Manos *et al.*, 2001).

The infrageneric structure of the genus has been described, based almost entirely on fruit morphology (Camus, 1952–54). Fruit morphology is the most reliable way to identify these trees (Soepadmo, 1972) and the groups identified by

Camus generally agree with a molecular study based on nuclear sequence data, although a certain amount of convergent evolution has apparently occurred (Cannon & Manos, 2001). The most obvious and perhaps most meaningful division is between species that produce nuts which are free from the cupule and those which are entirely fused to it (Barnett, 1944; Sang & Soepadmo, 2000; Cannon, 2001). The species that produce 'free' nuts are grouped into a few large subgenera or sections, marked by the presence or absence of cupular appendages. The importance of these groups is questionable, as they are occasionally polyphyletic or paraphyletic (Manos *et al.*, 2001) – the presence (including type) or absence of cupular appendages is probably labile, in an evolutionary sense. The species that produce 'fused' nuts are split among several small groups, based on the presence and absence of cupular appendages and fruit and leaf morphology. Most of these groups appear to be monophyletic (Cannon & Manos, 2001; Manos *et al.*, 2001), although the evidence supporting them is not overwhelming. Our sample includes all the major infrageneric groups proposed by Camus (1952–54) and almost all the smaller ones. Subsequent analyses will focus directly on the complex relationship of taxonomy to molecular data from both genomes.

Current sampling locations for the genus (Fig. 1) focused on representing the biogeographical range and the centres of species diversity in Indochina (Huang *et al.*, 2000) and Borneo (Soepadmo, 1970). General geological and biogeographical evidence across much of the distributional range of the genus is summarized elsewhere (Whitmore, 1987; Hall & Holloway, 1998; Ridder-Numan, 1998; Metcalfe *et al.*, 2001). The approximate geographical location and taxonomic composition of the nine collection locations are shown in Table 1. Samples were collected from old-growth trees in areas that have experienced limited human disturbance. Individuals in each location were not necessarily collected from a contiguous population but geographical distance within locations was trivial in comparison with distances among locations.

Table 1 Collection locations, their taxonomic composition, and chloroplast genetic variation. Sectional definitions follow Camus (1952–54). Approximate latitude–longitude values are shown. 'Types' refers to absolute number of types observed in each location and 'Max' to the maximum percentage divergence between types found in that location

Place	Location	Sample size			Genetic diversity	
		Inds	Spp.	Section	Types	Max
(1) C – Yunnan, China	25° N, 102° E	18	10	7	7	1.1
(2) M – Myanmar	22° N, 95°30' E	5	5	3	4	1.8
(3) VN – Vietnam	22° N, 103°30' E	28	11	5	10	1.0
(4) SEM – Semitau, MY	1°30' N, 110° E	33	13	4	15	2.2
(5) GP – Gunung Palung, IND	1°30' S, 110°15' E	19	11	4	10	1.5
(6) LH – Lambir Hills, MY	4°15' N, 114° E	9	5	3	7	2.0
(7) KH – Kelabit Highlands, MY	3°45' N, 115°15' E	20	13	6	13	2.3
(8) CR – Crocker Range, MY	5°30' N, 116°15' E	21	12	5	8	2.0
(9) KP – Kinabalu, MY	6° N, 116°45' E	39	15	6	9	1.5

Molecular data

Mature fresh leaves were collected from each individual, along with herbarium voucher specimens. As previous extractions from silica gel dried samples produced uncooperative DNA, isolation directly from fresh leaves was necessary. Whenever possible, leaf samples were chilled until isolations could be performed. Leaf material was ground in a 1.5-mL Eppendorf tubes (Fisher Scientific, Pittsburgh, PA, USA) using plastic pestles on sterile sand, occasionally with the aid of liquid nitrogen. DNA isolations were performed using DNeasy Plant Kits (Qiagen Inc., Santa Clarita, CA, USA). DNA was precipitated with ammonium acetate, dried, and returned to the laboratories at Duke University.

A pilot study was performed to identify the informative portion of the chloroplast genome at the level of this study. Previous results revealed little variation in the intergenic spacer regions among the various *trn* genes (Taberlet *et al.*, 1991; Demesure *et al.*, 1995). Around thirty individuals, representing the taxonomic and geographical breadth of the sample, were sequenced from both directions for several other candidate regions. The intergenic spacer region between the *rbcL* and *atpB* genes (Hodges & Arnold, 1994) was found to vary significantly, marking several groups by significant and non-homoplasious sequence changes or insertion/deletion events.

All 197 individuals included in the study were partially sequenced for the chloroplast spacer region (GenBank accession numbers: AY182835–AY182952). Initial alignment was performed in Sequencher (GeneCodes, Ann Arbor, MI, USA) and all base pair changes were verified by visually examining the chromatograms. As sequences were incorporated into the overall alignment, duplicate types were discarded. Once all sequences were incorporated, the types were exported to GeneDoc (Nicholas & Nicholas, 1997) and Mathematica 4 (Wolfram, 1998). Visual alignment was performed, resulting in a length of 588 base pairs, and informative indels were coded in a binary fashion.

To assess difference between cytoplasmic and nuclear markers, we analysed sequence data from the *ITS* regions of the nuclear rDNAs from a subset of fifty-seven individuals (Manos *et al.*, 2001), all included in the chloroplast survey. Preliminary results indicated that this pollen-mediated, nuclear molecular marker was more species-dependent, even among widespread *Lithocarpus* species (Manos *et al.*, 2001; Manos and Cannon, unpubl. data). The subset of fifty-seven individuals represents considerable taxonomic and geographical range in the genus, including most of the sections and geographical regions. Only a subset of individuals was used because of the difficulty involved in obtaining homologous copies of the internal transcribed spacer regions as most individuals are heterogeneous for this universally primed locus. Several different families of *ITS* loci persist in *Quercus* (Muir *et al.*, 2001) and we have observed the same in *Lithocarpus*. Population samples for this locus are currently being expanded and will be presented elsewhere.

Phylogenetic reconstruction

As initial heuristic searches, using PAUP* (Swofford, 2000), produced thousands of equally parsimonious trees, an effort was made to objectively choose a 'mean' phylogenetic hypothesis, representing the central tendency of the equally parsimonious hypotheses. A 'mean' phylogenetic hypothesis was chosen to test the range of phylogenetic uncertainty through the following steps. Ten separate heuristic searches were performed, using a random-addition sequence of taxa each time with the maximum number of trees for each analysis set at 500. Duplicate trees of the resulting 5000 trees were discarded and the remaining trees were filtered to retain only those meeting the minimum tree length of the most parsimonious tree (all characters equally weighted, including indels – see the score in the Results).

Distances among these equally parsimonious trees were measured using the symmetric difference metric (Penny & Hendy, 1985) and the two most different trees, lying at opposite ends of the distribution of the metric, were chosen to represent the breadth of phylogenetic uncertainty. The tree found halfway between these two extreme trees was chosen as the 'mean' tree. Character changes were mapped on the 'mean' tree, to create a phylogram for the construction of a haplotype network.

Statistical analysis of structure

A matrix of genetic distance, using the 'F84' model for sequence evolution (Felsenstein, 1984), was produced using PAUP* (Swofford, 2000) for both the nuclear and chloroplast data sets (unpublished tests of different models of sequence evolution did not affect the results). Indels were included in this genetic distance by coding them as transitions ($A > G = 0 > 1$) and combining them with the sequence data. These matrices were read into Mathematica 4 (Wolfram, 1998) and a distribution of pairwise genetic distances was produced for the pooled data. A continuous normal distribution was fit to the pooled data using the observed mean and standard deviation. This model was used as the 'global' model to which all populations were compared. 'Fit' was determined by the cumulative probability density functions of each observation given the 'global' model (Strait *et al.*, 1996). The expected distribution of genetic distances for each population in the analysis was generated from 1000 resamplings of the pooled data. The fit and summary statistics of each resampled data set were calculated, given the 'global' model. This technique was applied to both the chloroplast and nuclear molecular data. To allow direct comparison of the two data sets, all genetic distances were standardized using the mean genetic distance from the pooled data set (Petit *et al.*, 2002). This analysis was also performed on each of the major clades found in the chloroplast data set separately, similar to the nested-clade approach (Templeton, 1998) but working from bottom to top, instead of in an unrooted network.

Resampling schemes, such as the one used in this study, can test basic hypotheses about various elements of structure in the data and do not make any assumptions about the

underlying distribution (Fox, 1987; Manly, 1994; Simon, 1995; Webb & Peart, 2000). In this study, this technique was used to test for the presence of geographical structure in the distribution of genetic diversity. In the 'global' test, each resampling of the data assumed no structure and matched sampled individuals in each population randomly with haplotypes taken from the pool of all haplotypes. If no geographical structure existed in the data, the distribution of genetic diversity among the resampled populations should resemble the observed distribution. This technique also has the advantage of allowing significance values for observed patterns to be generated from the resampled distributions. Significance tests were performed by estimating the cumulative density function of the observed values on the distribution of fits, mean values and variances from each resampled data set. Observed values that lay beyond the 5% threshold on either tail of the distribution were considered significant. Using this method, there are three ways in which the observed values for a population can vary from the expected values, given the global model: the shape of its distribution, its mean and its variance.

RESULTS

Genetic variation and phylogenetic reconstruction

The final alignment was 588 base pairs long. Eight insertion/deletion events were coded as presence/absence data, varying in length from 3 to 13 base pairs, and combined with the sequence data. The final data set contained seventy-six haplotypes from the 197 individuals sampled. Two additional types were observed in the outgroup *Castanopsis* individuals. Sampling effort in a location did not correlate well with genetic diversity (Table 1). The sample from Kinabalu was the largest in terms of individuals (thirty-nine) and species (fifteen) and yet only nine haplotypes were observed with a maximum sequence divergence of 1.5%, while the sample from the Kelabit Highlands of twenty individuals and thirteen species contained thirteen haplotypes separated by a maximum of 2.3%. Two of the Indo-chinese populations (China and Vietnam) contained the lowest sequence divergence (1.1 and 1.0%, respectively) despite fairly large sample sizes. Additionally, genetic diversity did not correlate well with geographical distance between locations: the Semitau populations contained the largest number of haplotypes (fifteen) found in any population with the second highest maximum sequence divergence (2.2%), while the nearby Gunung Palung population in west Borneo contained ten types separated by a maximum 1.5% sequence divergence.

The sample was nearly evenly divided between two major clades, separated by four non-homoplasious changes (Fig. 2). These changes each represent a set of presumably independent mutations in DNA sequence and individuals in each clade possess one of the two sets of four base pair combinations. The clades were labelled 'Widespread' and 'Bornean' because of their distribution patterns, containing 111 individuals/thirty-nine types and eighty-six individuals/thirty-

five types, respectively. The group of most parsimonious trees was 115 steps long (excluding outgroups) and the strict consensus tree is shown (Fig. 2a). The breadth of phylogenetic uncertainty among the most parsimonious trees was twenty-eight symmetric differences (Penny & Hendy, 1985) and the 'mean' tree was sixteen symmetric differences from each of these two extremes. The haplotype network (Fig. 2b) represents the relationships and inferences on character change obtained from the 'mean' tree, which had a CI of 0.80. All equally parsimonious trees were rooted by the outgroup haplotypes at type A1 (the most common type; Fig. 2b). Most of the uncertainty about reconstruction occurred within the Bornean clade, particularly in how the two major shared types (S17 and SB7) are connected to the third missing type, counting from A1. One of the extreme trees suggests an almost complete split between western and northern Borneo.

A large proportion of observed haplotypes was endemic to a location (Fig. 3), particularly in Gunung Palung in western Borneo (Loc. 5 – GP) where only one observed type was shared with another location (Loc. 4 – SEM) and no types from the Widespread clade were observed. The populations in northern Borneo possessed the lowest frequency of endemic types, particularly in the Crocker Range (Loc. 8 – CR) population. The SEM population in northwest Borneo contained the largest number of shared types from both clades.

Assuming no geographical structure within the data, the distribution of standardized pairwise genetic distances could be accurately described by a normal distribution, given a mean of 0.76 and a standard deviation of 0.44 (Table 2; Fig. 4A). The two largest samples from Indochina (C and VN) both strongly deviated from this assumption, in terms of mean, standard deviation and distribution of relatedness (Table 2; Fig. 4i, iii), indicated that individuals in these populations are much more closely related than expected and that strong geographical structure exists in the distribution of genetic diversity in this location. The GP population was the only one on Borneo with similar levels of deviation (Table 2; Fig. 4v), while the nearby SEM population did not deviate from the null model for any of the three measures, suggesting that it either lies near the centre of the origin of the group or is at an equilibrium in terms of migration. Kinabalu (KP) was the only other Bornean population that was more closely related than expected (Table 2; Fig. 4ix). Conversely, individuals in the central Bornean population (KH) were more distantly related than expected (Table 2; Fig. 4vii), suggesting invasion of the region and *in-situ* derivation of subtypes from both major lineages.

Much less geographical structure was found in the sample, according to nuclear *ITS* sequence data taken from the subset of individuals (Table 2; Fig. 5). This result indicates little phylogenetic bias existed in our sample, despite the large number of species in the study. Three of the four populations on Borneo were found to be near random expectations, assuming no geographical structure (Table 2; Fig. 5iv, v, ix). Surprisingly, the only population (CR) showing geographical structure according to the nuclear

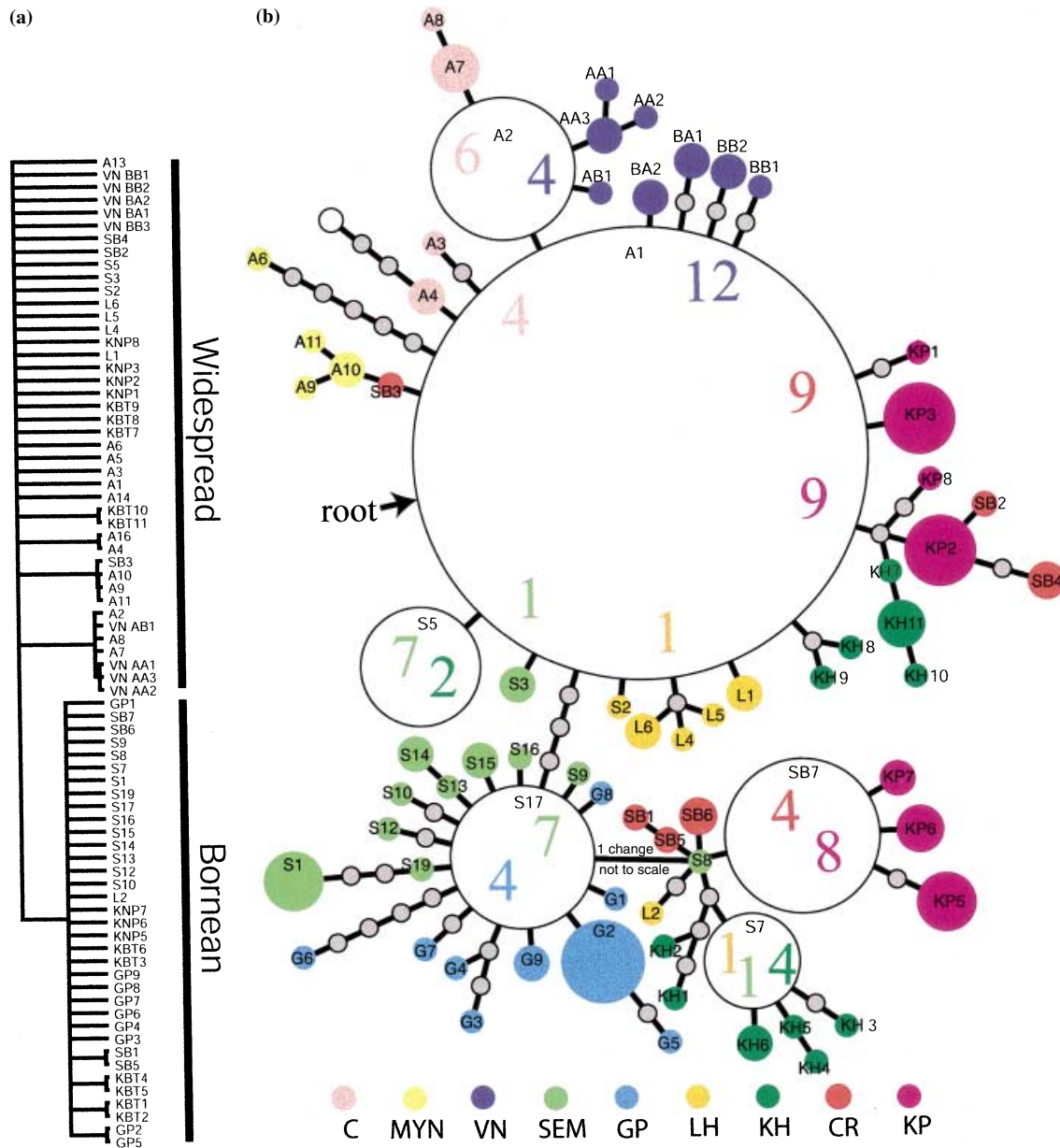


Figure 2 Phylogenetic relationship of chloroplast types. The cladogram on the left is the strict consensus of all equally parsimonious trees. The two major clades are labelled 'Widespread' and 'Bornean' according to their geographical distributions. The genetic network on the right was constructed from the 'mean' most parsimonious tree. The clear and coloured circles represent observed types while the grey circles represent 'missing' types inferred from the reconstruction. All types are labelled (GenBank accession numbers: AY182835–AY182952). The diameter of the circle indicates the number of individuals found possessing that type, the smallest being a single individual. Clear circles were found in more than one location (shared types), the number of individuals in each location shown by the coloured numbers within the circle. Solid coloured circles illustrate types found in only one location (endemic types). Locations are colour coded in the following way: China – pink, Myanmar – yellow, Vietnam – dark blue, Semitau – light green, Gunung Palung – light blue, Lambir Hills – orange, Kelabit Highlands – dark green, Crocker Range – dark red, Kinabalu – purple.

data was near random expectations according to the chloroplast data (Table 2; Fig. 5viii). A slight bias was apparent between the Chinese and Bornean populations (Table 2; Fig. 5i), although the strength of the geographical

structure among nuclear types is much weaker than among chloroplast haplotypes.

When the two major clades are considered separately, the individuals in the SEM population were geographically

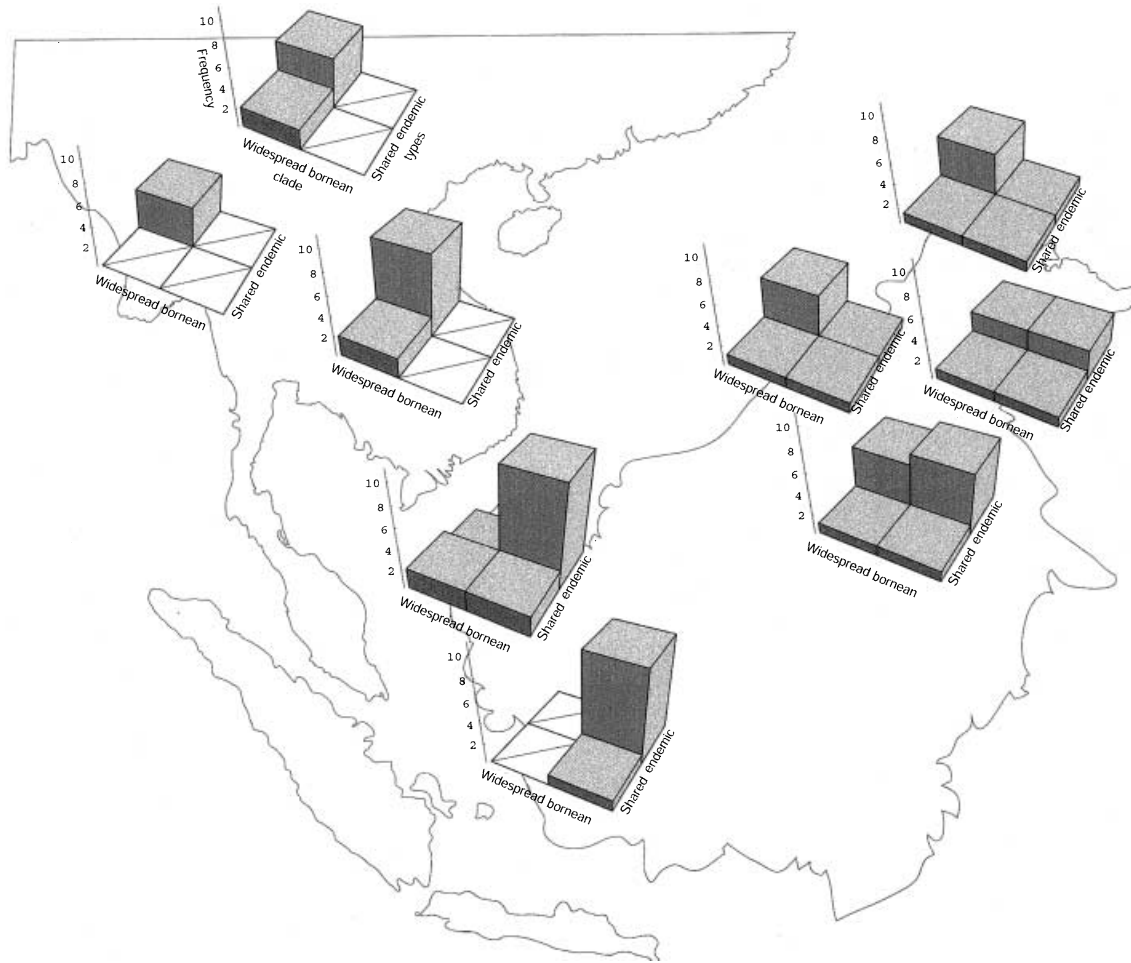


Figure 3 The geographical distribution of shared vs. endemic types from each of the major clades. Charts are shown for each location, the vertical axis represents the frequency of types for each category. Types from the Widespread clade are shown in the left two columns and types from the Bornean clade are shown in the right columns. The front row represents shared types while the back row represents endemic types.

structured in both clades (Table 3). This result indicates that the population lies near the centre of origin of all haplotype diversity but is not at equilibrium for migration in the two clades, being heavily biased towards the Bornean clade. Individuals in the KH population were slightly more closely related than expected in each of the two clades (Table 3 – significance tests for the separate clades was greatly hampered by small sample sizes in each population). This result was in sharp contrast to the more distant relationship observed when all haplotypes were considered (Table 2) and clearly demonstrates the mixture of the two clades in the KH population.

CONCLUSIONS

Ancient persistence

Four aspects of the data suggest that most of the locations sampled in this study represent ancient populations, which have experienced limited migration and limited extinction:

- (1) the overall high genetic diversity found in the study;
- (2) the high levels of endemism observed in most locations;
- (3) the presence of an ancestral type (A1) throughout most of the group's range and
- (4) the relatively few 'missing' ancestral types throughout the network. How ancient these populations may be is hard to determine but given that (1) the ancestral type, A1, is only a single genetic change away from a type observed in the sister genus *Castanopsis* with no observed overlap in chloroplast types between the two genera, and (2) the limited evidence for long distance migration, these populations, excluding perhaps GP in west Borneo and KH in central Borneo, may have been persistent throughout the evolutionary history of the genus.

Overall, the level of genetic diversity in the stone oaks greatly exceeds that observed in European trees (Petit *et al.*, 1997; Ferris *et al.*, 1998; King & Ferris, 1998), although this may not be surprising. This study examined a large number of individuals from tropical rain forests, which have not been affected by climate change as profoundly as north temperate populations during the Quaternary Period. But, in

Table 2 Standardized genetic distance among individuals for molecular markers from two genomes. Mean and standard deviation are shown for all independent pairwise distances. Numbers in boldface are significantly different than random expectations (see Materials and methods, Fig. 4)

	CpDNA	nDNA
Pooled	0.76 (0.44)	1.0 (0.44)
1. C (18,12)	0.25***† (0.19***)	0.71** (0.34)
2. M (5,0)	0.61† (0.60*)	N/A
3. VN (28,0)	0.27***† (0.20***)	N/A
4. SEM (33,6)	0.64 (0.41)	0.93 (0.37)
5. GP (19,10)	0.35***† (0.27***)	0.9 (0.44)
6. LH (9,3)	0.76 (0.50)	N/A
7. KH (20,2)	0.90***† (0.56***)	N/A
8. CR (21,4)	0.60 (0.44)	0.53* (0.37)
9. KP (39,25)	0.54*** (0.37*)	1.1† (0.48)

Significance levels are: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Population sample sizes are shown in parentheses for the chloroplast and nuclear data sets, respectively.

†The distribution of pairwise differences is significantly different from expectations at $P < 0.05$ level.

a study of eight species of *Quercus* in over 2600 populations across the European continent, Petit *et al.* (2002) observed only thirty-two chloroplast types after sampling four separate fragments of cpDNA, while only one fragment was examined in this study. Diversity of chloroplast types in neotropical trees does appear to be relatively high: seven types from five fragments were observed in 163 individuals from a single species in several largely contiguous populations (Hamilton, 1999); six types in 100 individuals (Dutech *et al.*, 2000) and twelve types in 250 individuals (Caron *et al.*, 2000) from single species across French Guiana.

The difference between this study and many of the other studies may also reflect the differences between polymerase chain reaction-restricted fragment length polymorphism (PCR-RFLP) based techniques (Petit *et al.*, 1993b; Dumolin *et al.*, 1995; Muloko-NToutoume *et al.*, 2000) and DNA sequencing. In a study of chloroplast sequence variation from a single species of cycle-cup oak (*Cyclobalanopsis glauca*) in northern East Asia, thirteen haplotypes were detected in 140 individuals from thirty-two populations (Huang *et al.*, 2002), further supporting the idea that sequence data more efficiently detects haplotype variation. The detection of sequence mutations using PCR-RFLP depends on screening of fragments with the appropriate restriction enzyme, whereas all changes are observed using sequence data. Additionally, most changes detected by PCR-RFLP are length differences (Dutech *et al.*, 2000; Petit *et al.*, 2002), commonly inferred as insertion–deletion events (McCauley, 1995) but may also refer to extension of highly repetitive motifs that appear to be prone to higher levels of homoplasy (van Oppen *et al.*, 2000). One such region (a string of eight to twelve adenine bases) was excluded in this analysis because it did not appear to amplify reliably, but could have been interpreted as a length difference by the PCR-RFLP

technique. Additionally, phylogenetic reconstruction of DNA sequence is more rigorous and model-driven than the presence–absence data of PCR-RFLP data. More importantly, sequence data allows unambiguous rooting of the network and definition of subclades, allowing objective determination of evolutionary direction in the model.

The high frequency of endemics in most populations suggests prolonged *in-situ* sequence evolution (Fig. 3). Some locations, like GP in western Borneo, possess a majority of endemic types and contain a high relative fraction of the genetic differentiation found in the entire network (eight changes from tip to tip in GP alone vs. fifteen for the entire sample). Given a clock-like evolution of the chloroplast sequence, this result suggests that the GP population is almost half as old as the entire genus (possibly 40 Myr). Chloroplast DNA sequence evolution in Fagaceae was calibrated using fossils (Crepet, 1989; Franscaria *et al.*, 1993), suggesting that roughly 1.0% non-coding sequence divergence equals 17 Myr (Manos & Stanford, 2001). Calibrating the rate of chloroplast evolution of the Southeast Asian stone oaks will require some external corroborating evidence (Baldwin & Sanderson, 1998), given the potential for different historical population dynamics in tropical forests, in comparison with temperate regions. Fluctuations in generation time and population levels as a result of climatic and ecosystem change would be expected to cause rate heterogeneity among the nodes of the genetic network.

The high frequency of an ancestral haplotype (A1) in Indochina and north Borneo suggests limited genetic drift in either population (Fig. 2B). The persistence of this ancestral haplotype is particularly striking because nuclear markers indicate the presence of geographical structure between Indochina and Borneo (Table 2), and deep phylogenetic structure among the species found in the two regions (Manos *et al.*, 2001). Ancient persistence of a neutral polymorphism in cichlid fishes has been estimated to last more than a million years (Nagl *et al.*, 1998). Given the radical difference in generation time between the two organisms, the persistence of neutral polymorphisms could certainly persist for more than 10 Myr. The separation of these two populations probably coincides with the narrowing of Sundaland and the opening of the Bay of Thailand during the middle of the Miocene (Hall, 1998). While this type is concentrated in northern Borneo, it and numerous derived types are present in almost all Bornean populations. The maintenance of this type in both Indochina and north Borneo through migration is very unlikely because no other types are shared between the two landmasses, suggesting that only individuals possessing the ancestral type or types derived from it migrate. The migration of derived types would not result in the high levels of endemism for these types in each population.

Finally, most of the populations also appear to have experienced limited extinction, because other than the three ‘missing’ types separating the two major clades, few other major gaps exist in the genetic network. The addition of more samples may provide the other ‘missing’ haplotypes or they may potentially have become extinct during a severe bottleneck event. If genetic drift and frequent extinction of

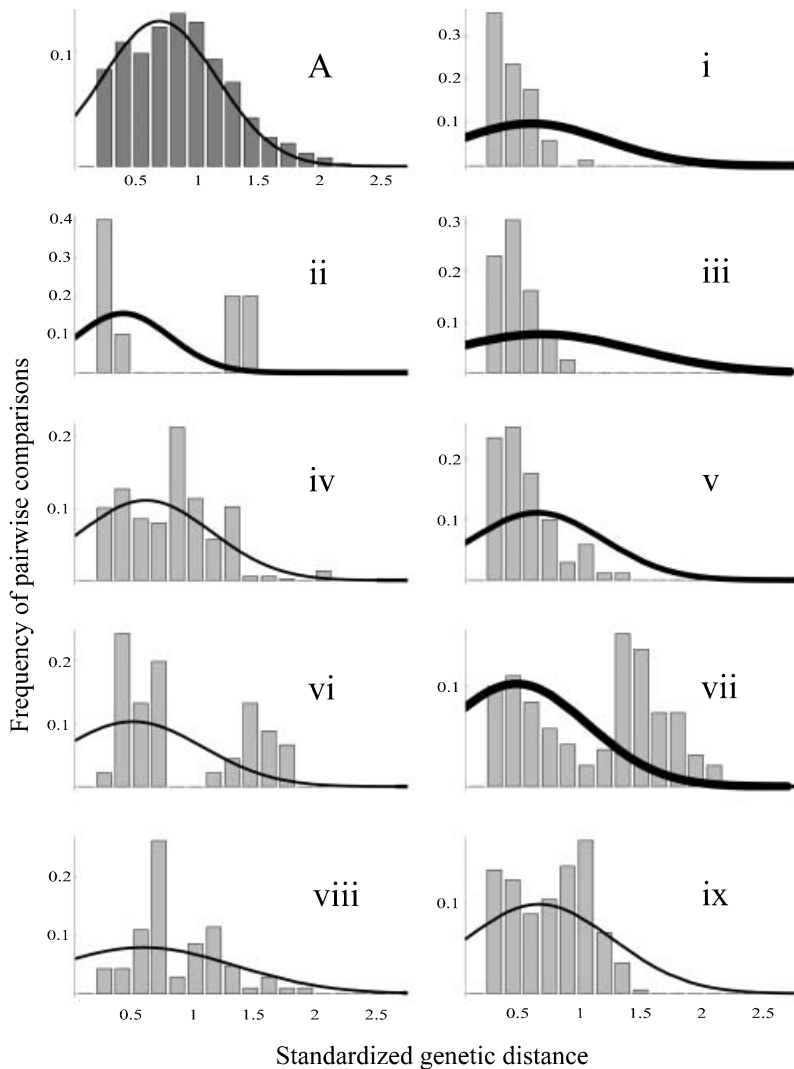


Figure 4 Distribution of pairwise genetic distances among individuals for chloroplast DNA sequence data. 'A' represents the individuals from all locations pooled together, the other locations are labelled according to Table 1. All genetic distances were standardized by the mean pairwise distance from the 'A' sample. Bars represent observed distributions. Curves represent the expectation for the best fitting normal distribution for each location, assuming no geographical structure in the data (see Materials and methods). Thick curves show observed distributions that deviated significantly from expectations, with the level of significance shown by the thickness of the curve: 4 pt. ($P < 0.01$), 6 pt. ($P < 0.001$).

haplotypes were occurring, the genetic network would have numerous spans of 'missing' types separating the observed types.

Biogeographical implications

The results indicate three major phases in the evolution of the genus in Southeast Asia. Initially, the ancestral group became quite widespread, either in continuous populations throughout the Indochina–Sundaland area or by a rapid expansion of their range, allowing the ancestral haplotype to reach either end of the subcontinent without undergoing any change. This widespread population was then fragmented and underwent a severe bottleneck in central Sundaland. This bottleneck and subsequent fragmentation caused the Bornean lineage (which was not necessarily in the geographical region we now call 'Borneo' at the time) to undergo significant genetic drift and to diversify in isolation. Subsequently, the populations containing the Bornean

lineage migrated into central and northern Borneo from north-western Sundaland.

This scenario would explain several aspects of the results, including (1) the large genetic distance between the two major clades, (2) the great breadth of diversity of haplotypes found in north-western Borneo, including the relatively low frequency of the ancestral A1 haplotype, and (3) north and central Borneo as a mixing ground for the two clades. Alternatively, it raises a few conflicting questions: if Indochina and North Borneo have maintained stable populations over the last few tens of millions of years of evolution in the genus, why has so little genetic diversity accumulated in the Widespread clade in comparison with the Bornean clade and why have so few endemic haplotypes been produced? This difference may ultimately be the result of simple rate heterogeneity in sequence evolution between the two clades.

The low genetic diversity of the ancestral type in combination with the broad phylogenetic breadth of types in the Semtau population and the complete absence of the

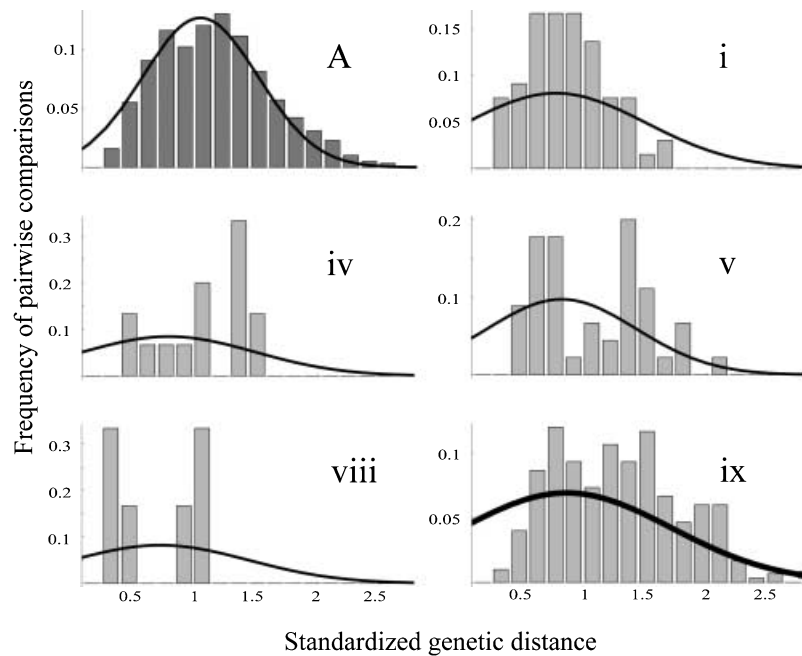


Figure 5 Distribution of pairwise genetic distances among individuals for nuclear DNA sequence data. 'All' represents the individuals from all locations pooled together, the other locations are labelled according to Table 1. All genetic distances were standardized by the mean pairwise distance from the 'All' sample. Bars represent observed distributions. Curves represent the expectation for the best fitting normal distribution for each location, assuming no geographical structure in the data (see Materials and methods). Thick curves show observed distributions that deviated significantly from expectations: the level of deviation shown by each location label: * $P < 0.05$.

ancestral type in the GP population agree with the idea of an extreme bottleneck in central Sundaland, after which these populations re-established themselves. Given recent reconstructions for land area during these periods for greater Sundaland (Moss & Wilson, 1998), this area probably represents the oldest upland tropical habitat in the region since the early Cenozoic. Pollen records indicate substantial climatic fluctuations during the Pliocene, causing rain forest habitat to significantly contract (Morley, 2000). These fluctuations may have caused the diversification of the Bornean lineage.

Table 3 Standardized genetic distance among individuals each of the two major clades of chloroplast sequence data. Mean and standard deviation are shown for all independent pairwise distances. Numbers in boldface are significantly different than random expectations (see Materials and methods, Fig. 4)

	Widespread	Bornean
Pooled	1.00* (0.69)	1.0 (0.57)
1. C (18,0)	0.69 (0.52)	N/A
2. M (5,0)	1.60* (1.60*)	N/A
3. VN (28,0)	0.73* (0.50)	N/A
4. SEM (10,23)	0.34**† (0.36)	0.72** (0.49)
5. GP (0,19)	N/A	0.62*** (0.47)
6. LH (8,1)	1.1 (0.49)	N/A
7. KH (9,11)	0.86 (0.47)	0.74 (0.54)
8. CR (13,8)	0.67† (0.74)	0.51* (0.32)
9. KP (20,18)	0.65* (0.49)	0.32*** (0.27*)

Significance levels are * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Population sample sizes are shown in parentheses for each major clade. †The distribution of pairwise differences is significantly different from expectations at $P < 0.05$ level.

The bimodal nature of the distribution of genetic relatedness among individuals in the central Bornean populations and their tendency to be more distantly related than expected strongly supports their position as a mixing ground between the two major lineages on the island. They also contain substantial fractions of unique genetic diversity, which suggests that they have also maintained populations, although they may have been severely fragmented at various times in the past. The population in the Kelabit Highlands is the only population in Borneo containing both clades that did not also contain the ancestral type (A1), suggesting a severe bottleneck and the loss of this ancestral type. Subsequent re-invasion from the north (containing the 'Widespread' clade) and from the west (containing the 'Bornean' clade) was of roughly equal proportion and these populations have persisted since that period.

The question of why Indochina and Kinabalu contain such low levels of genetic diversity remains. If they have acted as refugia for the genus throughout its evolutionary history and predate the populations in western Borneo, one would expect equal if not much greater genetic diversity and uniqueness, which are not observed in the data. Two main hypotheses can be put forward. (1) These populations remained both large and contiguous, with few barriers to gene flow, either geographically or taxonomically. This would prevent genetic substructuring within the population and prevent genetic drift, although conversely one would expect a higher frequency of endemic types. These mutants may simply be extremely rare and therefore not sampled in this study. (2) Alternatively, these areas may have been affected by the recent climatic changes during the Quaternary Period, experiencing significant population bottlenecks and habitat reductions in the recent past, leaving only a fraction of the original genetic diversity. This

hypothesis requires that population bottlenecks were both repeated and recent, otherwise the gaps in the genetic network would be more pronounced, i.e. more 'missing' types would have been inferred.

One of the most intriguing aspects of this data is the level of species independence observed in haplotype distribution. Although most individuals of a taxon are contiguous to one another on the network, several taxa possess a wide range of types and several types are found in a wide range of species. This situation may either be the result of incomplete lineage sorting of ancestral polymorphisms or widespread introgression of these markers across traditional species boundaries, similar to patterns suggested by chloroplast DNA variation observed in a single species of Asteraceae (Mason-Gamer *et al.*, 1995). Although introgression between species with radically different fruit morphologies seems unlikely, e.g. between sections *Synaedrys* and *Pasania* (Cannon & Manos, 2001; Manos *et al.*, 2001), a continuum of 'leakage' between closely related species pairs, similar to a 'ring species' (Irwin *et al.*, 2001), may allow homogenization of types within diverse populations (Whittemore & Schaal, 1991). Sample sizes for individual taxonomic units in this study are currently too small to perform tests for taxonomic structure in the distribution of haplotypes but initial patterns do not suggest high levels of introgression. Species of *Lithocarpus* appear to behave like 'cohesive' units (Van Valen, 1982; Templeton, 1989), which have inherited a diversity of ancestral polymorphisms. These 'cohesive' units retain a diversity of chloroplast types, unlike the different sections in *Quercus* (Manos *et al.*, 1999) and the types present in most species are not randomly distributed across the network but involve subclades (Cannon, 2000). These different tendencies towards fixation within sections or closely related species may relate to relative differences in age or basic species biology in the two genera.

The species independence of cpDNA markers in this study and others (Golden & Bain, 2000; Petit *et al.*, 2002) should provide a cautionary tale for phylogeographical studies of cytoplasmic markers, particularly when species boundaries are poorly understood. This situation is probably true for most tropical tree species, where high infrageneric species diversity and mixed-species communities are common, e.g. *Shorea* (Dipterocarpaceae), *Syzygium* (Myrtaceae), *Garcinia* (Clusiaceae) (Kochummen & LaFrankie, 1990; LaFrankie *et al.*, 1995; Curran & Leighton, 2000). An initial broad sampling of taxonomic units at the generic and infrageneric levels should be conducted, until clear divisions for gene flow and evolutionary history are apparent. Otherwise, possible misinterpretation of patterns could be possible using a single species.

Conversely, the species dependence of these markers means that the geographical signal may not be as strong as it appears. Geographical patterns of species composition in section *Synaedrys* have been described based on fruit morphology (Cannon & Manos, 2000) and a deep split between Chinese and Bornean taxa based upon *ITS* sequence data (Manos *et al.*, 2001) obviously suggest that some of the geographical pattern in chloroplast sequence variation is because of irrevocable reproductive isolation (speciation)

between taxonomic groups. Unfortunately, most species are not represented by enough individuals to perform rigorous tests and to properly estimate the amount of bias imposed upon the analysis by this confounding factor. Infrageneric groups, such as sections, are still too poorly defined to provide adequate guidance as well. Only the ancestral haplotype 'A1' was observed in nine individuals of *L. turbinatus*, the only example of a 'fixed' species represented by numerous samples. This species is restricted to perhumid, exposed mountain ridges, normally above 2000 m elevation, in northern Borneo and probably suffers from chronic population bottlenecks because of the rarity and specificity of its habitat. This kind of population dynamic would lead to fixation quickly. More data from independent sources is necessary to understand why it fixed on the ancestral type.

More extensive population sampling, particularly on the landmasses remaining from central Sundaland, should be pursued. This region may harbour the three missing types between the two major clades (Fig. 2B). This evidence will also provide a much clearer understanding of migration routes and refugia in Southeast Asian rain forests and allow the construction of a rigorous framework to examine the effects of past climate change on the distribution of rain forest, perhaps reaching back to the Pliocene. Comparative studies with lowland dominants, like the Dipterocarpaceae, or trees with different fruit dispersal syndromes, like the Myrtaceae, would also provide a complementary view to patterns emerging about the upland-dominant Fagaceae.

Generally, each geographical location appears to have had a relatively distinct and rich history. Little evidence for widespread mixing of populations exists, as most of the patterns can best be explained through long-term persistence, limited migration, and *in-situ* diversification. Similar patterns were suggested by the low level of overlap in community species composition in Malaysian-Borneo riverine frogs, despite a range of stream conditions in each location sampled (Inger & Voris, 1993). Only GP in the west has become fixed on one of the major lineages, thus indicating a severe bottleneck in its evolutionary past, while all other locations have maintained both ancestral haplotypes. At the moment, timing these migrations and diversifications is difficult.

ACKNOWLEDGMENTS

The authors would like to thank the Institute for Biodiversity and Environmental Conservation at the University of Malaysia, Sarawak; the Forest Research Center of the Sarawak Forestry Department; the Parks Office in Sabah; the Indonesian Institute of Sciences (LIPI), Department of Forestry, the University of Indonesia, and the Biological Research Center in Bogor; the Kunming Institute of Botany, China; and the Institute of Ecology and Biological Resources of Vietnam for their assistance and support during the collection of samples for this study. Particular thanks goes to J.H. Beaman, G. ('Charlie') Ismail, F. Liew, J. Tarawe, Z.-K. Zhou, Dan Harder, Phan Ke Loc and Nguyen Tien Hip. Laboratory assistance was provided by M.E. Jones, G. Chen, K. Hancock, and N. Hamilton. J.S. McLachlan and an

anonymous reviewer provided helpful criticism. Funding was provided by the U.S. Academy for Educational Development, the Mellon Foundation, and NSF DEB-0108690.

REFERENCES

- Aiba, S.I. & Kitayama, K. (1999) Structure, composition and species diversity in an altitude-substrate matrix of rain forest tree communities on Mount Kinabalu, Borneo. *Plant Ecology*, **140**, 139–151.
- Aide, T.M. & Rivera, E. (1998) Geographic patterns of genetic diversity in *Poulsenia armata* (Moraceae): implications for the theory of Pleistocene refugia and the importance of riparian forest. *Journal of Biogeography*, **25**, 695–705.
- Anshari, G., Kershaw, P. & van der Kaars, S. (2001) A Late Pleistocene and Holocene pollen and charcoal record from peat swamp forest, Lake Sentarum Wildlife Reserve, West Kalimantan, Indonesia. *Palaeogeography, Palaeoclimatology, and Palaeoecology*, **171**, 213–228.
- Ashton, P.S. (1988) Dipterocarp biology as a window to the understanding of tropical forest structure. *Annual Review of Ecology and Systematics*, **19**, 347–370.
- Baldwin, B.G. & Sanderson, M.J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences of the USA*, **95**, 9402–9406.
- Barnett, E.C. (1944) Keys to the species groups of *Quercus*, *Lithocarpus*, and *Castanopsis* of eastern Asia, with notes on their distributions. *Transactions of the Botanical Society of Edinburgh*, **34**, 159–204.
- Burgess, P.F. (1966) *Timbers of Sabah*. Forest Department, Sabah Malaysia.
- Camus, E. (1952–54) Les chenes. *Encyclopedie Economique de Sylviculture*, **8**, 511–1196.
- Cannon, C.H. (2000) Macroevolutionary and phylogeographic studies of *Lithocarpus* (Fagaceae): gene flow and fruit evolution in a tropical tree. PhD Dissertation, Department of Botany, Duke University, Durham, NC, 158 pp.
- Cannon, C.H. (2001) Morphological and molecular diversity in *Lithocarpus* (Fagaceae) of Mount Kinabalu. *Sabah Parks Nature Journal*, **4**, 45–69.
- Cannon, C.H. & Manos, P.S. (2000) The Bornean *Lithocarpus* Bl. section *Synaedrya* (Lindl.) Barnett (Fagaceae): its circumscription and description of a new species. *Botanical Journal of Linnean Society*, **133**, 343–357.
- Cannon, C.H. & Manos, P.S. (2001) Combining and comparing continuous morphometric descriptors with a molecular phylogeny: the case of fruit evolution in the Bornean *Lithocarpus* (Fagaceae). *Systematic Biology*, **50**, 1–21.
- Caron, H., Dumas, S., Marque, G., Messier, C., Bandou, E., Petit, R.J. & Kremer, A. (2000) Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology*, **9**, 1089–1098.
- Comes, H.P. & Kadereit, J.W. (1998) The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.
- Corlett, R.T. (1998) Frugivory and seed dispersal by vertebrates in the Oriental (Indomalayan) Region. *Biology Review of Cambridge Philosophical Society*, **73**, 413–448.
- Crepet, W.L. (1989) History and implications of the early North American fossil record of Fagaceae. *Evolution, systematics, and fossil history of the Hamamelidae* (eds P.R. Crane and S. Blackmore), pp. 45–66. Clarendon, Oxford.
- Curran, L.M. & Leighton, M. (2000) Vertebrate responses to spatiotemporal variation in seed production of mast-fruiting dipterocarpaceae. *Ecological Monograph*, **70**, 101–128.
- Dam, R.A.C., Fluin, J., Suparan, P. & van der Kaars, S. (2001a) Palaeoenvironmental developments in the Lake Tondano area (N. Sulawesi, Indonesia) since 33,000 yr B.P. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 147–183.
- Dam, R.C.A., van der Kaars, S. & Kershaw, P. (2001b) Quaternary environmental change in the Indonesian region. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 91–95.
- Demesure, B., Sodzi, N. & Petit, R.J. (1995) A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129–131.
- Dumolin, S., Demesure, B. & Petit, R.J. (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theoretical and Applied Genetics*, **91**, 1253–1256.
- Dumolin-Lapegue, S., Demesure, B., Fineschi, S., LeCorre, V. & Petit, R.J. (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.
- Dutech, C., Maggia, L. & Joly, H.I. (2000) Chloroplast diversity in *Vouacapoua americana* (Caesalpinaceae), a neotropical forest tree. *Molecular Ecology*, **9**, 1427–1432.
- Ennos, R.A., Sinclair, W.T., Hu, X.S. & Langdon, A. (1999) Using organelle markers to elucidate the history, ecology and evolution of plant populations. *Molecular systematics and plant evolution* (eds P.M. Hollingsworth, R.M. Bateman and R.J. Gornall), pp. 1–19. Taylor & Francis, London.
- Felsenstein, J. (1984) Distance measures for inferring phylogenies: a justification. *Evolution*, **38**, 368–376.
- Ferris, C., Oliver, R.P., Davy, A.J. & Hewitt, G.M. (1995) Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Ecology*, **4**, 731–738.
- Ferris, C., King, R.A., Vainola, R. & Hewitt, G.M. (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity*, **80**, 584–593.
- Fineschi, S., Turchini, D., Villani, F. & Vendramin, G.G. (2000) Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill. (Fagaceae) throughout southern European countries. *Molecular Ecology*, **9**, 1495–1503.
- Flenley, J.R. (1998) Tropical forests under the climates of the last 30,000 years. *Climatic Change*, **39**, 177–197.
- Fox, B.J. (1987) Species assembly and the evolution of community structure. *Evolutionary Ecology*, **1**, 201–213.
- Franscaria, N., Maggia, L., Michaud, M. & Bousquet, J. (1993) The *rbcL* gene sequence from chestnut indicates a slow rate of evolution in the Fagaceae. *Genome*, **36**, 668–671.
- Fuji, N., Tomaru, N., Okuyama, K., Koike, T., Mikami, T. & Ueda, K. (2002) Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. *Plant Systematics and Evolution*, **232**, 21–33.

- Golden, J.L. & Bain, J.F. (2000) Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packeria* (Asteraceae) species in southwestern Alberta. *Evolution*, **54**, 1566–1579.
- Hall, R. (1998) The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 99–131. Backhuys, Leiden.
- Hall, R. & Holloway, J.D. (eds) (1998) *Biogeography and geological evolution of Southeast Asia*. Backhuys, Leiden.
- Hamilton, M.B. (1999) Tropical tree gene flow and seed dispersal. *Nature*, **401**, 129.
- Hardin, J.W. (1975) Hybridization and introgression in *Quercus alba*. *Journal of Arnold Arboretum*, **56**, 336–363.
- Herendeen, P.S., Crane, P.R. & Drinnan, A.N. (1995) Fagaceous flowers, fruits, and cupules from the Campanian (Late Cretaceous) of central Georgia, USA. *International Journal of Plant Science*, **156**, 93–116.
- Hodges, S.A. & Arnold, M.L. (1994) Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences of the USA*, **91**, 5129–5132.
- Hope, G. (2001) Environmental change in the Late Pleistocene and later Holocene at Wanda site, Soroako, South Sulawesi, Indonesia. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 129–145.
- Huang, C.J., Zhang, Y.T. & Bartholomew, B. (2000) Fagaceae. *Flora of China: Cycadaceae through Fagaceae* (eds Z.Y. Wu and P.H. Raven), pp. 314–400. Science Press, Missouri Botanical Garden, Beijing, St Louis, MO.
- Huang, S.S.F., Hwang, S.-Y. & Lin, T.-P. (2002) Spatial pattern of chloroplast DNA variation of *Cyclobalanopsis glauca* in Taiwan and East Asia. *Molecular Ecology*, **11**, 2349–2358.
- Inger, R.F. & Voris, H.K. (1993) A comparison of amphibian communities through time and from place to place in Bornean forests. *Journal of Tropical Ecology*, **9**, 409–433.
- Inger, R.F. & Voris, H.K. (2001) The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography*, **28**, 863–891.
- Irwin, D.E., Irwin, J.H. & Price, T.D. (2001) Ring species as bridges between microevolution and speciation. *Genetica*, **112**, 223–243.
- van der Kaars, S. & Dam, R. (1997) Vegetation and climate change in west Java, Indonesia during the last 135,000 years. *Quaternary International*, **37**, 67–71.
- van der Kaars, S., Penny, D., Tibby, J., Fluin, J., Dam, R.A.C. & Suparan, P. (2001) Late Quaternary palaeoecology, palynology and palaeolimnology of a tropical lowland swamp: Rawa Danau, West Java, Indonesia. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 185–212.
- Kaul, R.B. (1986) Evolution and reproductive biology of inflorescences in *Lithocarpus*, *Castanopsis*, *Castanea*, and *Quercus* (Fagaceae). *Annals of the Missouri Botanical Garden*, **73**, 284–296.
- Kaul, R.B. (1989) Fruit structure and ecology in paleotropical *Lithocarpus* (Fagaceae). *Evolution, systematics, and fossil history of the Hamamelidae* (eds P.R. Crane and S. Blackmore), pp. 67–86. Clarendon Press, Oxford.
- King, R.A. & Ferris, C. (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151–1161.
- Kochummen, K.M. & LaFrankie, J.V. (1990) Floristic composition of Pasoh Forest reserve, a lowland rain forest in peninsular Malaysia. *Journal of Tropical Forest Science*, **3**, 1–13.
- Kvacek, Z. & Walther, H. (1989) Paleobotanical studies in Fagaceae of the European Tertiary. *Plant Systematics and Evolution*, **162**, 213–229.
- LaFrankie, J.V., Tan, S. & Ashton, P.S. (1995) *Species list for the 52-ha forest dynamics research plot: Lambir Hills National Park, Sarawak, Malaysia*. Center for Tropical Forest Science, Smithsonian Tropical Research Institute.
- Langdon, D.M. (1939) Ontogenetic and anatomical studies of the Fagaceae and Juglandaceae. *Botanical Gazette*, **101**, 301.
- Lee, S.L., Wickneswari, R., Mahani, M.C. & Zakri, A.H. (2000) Genetic diversity of a tropical tree species, *Shorea leprosula* Miq. (Dipterocarpaceae), in Malaysia: implications for conservation of genetic resources and tree improvement. *Biotropica*, **32**, 213–224.
- Leighton, M. & Leighton (1983) Vertebrate responses to fruiting seasonality within a Bornean rain forest. *Tropical rain forests: ecology and management* (eds S.L. Sutton, T.C. Whitmore and A.C. Chadwick). Blackwell Science, Oxford.
- Lowe, A.J., Gillies, A.C.M., Wilson, J. & Dawson, I.K. (2000) Conservation genetics of bush mango from central/west Africa: implications from random amplified polymorphic DNA analysis. *Molecular Ecology*, **9**, 831–841.
- Manchester, S.R. (1994) *Fruits and seeds of the middle Eocene Nut Beds flora, Clarno Formation, Oregon*. Paleontological Research Institution, Ithaca, NY.
- Manly, B.F.J. (1994) *Randomization and Monte Carlo methods in biology*. Chapman & Hall, New York.
- Manos, P.S., Doyle, J.J. & Nixon, K.C. (1999) Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12**, 333–349.
- Manos, P.S. & Stanford, A.M. (2001) The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the Northern hemisphere. *International Journal of Plant Science*, **162**, S77–S93.
- Manos, P.S. & Steele, K.P. (1997) Phylogenetic analyses of 'higher' Hamamelididae based on plastid sequence data. *American Journal of Botany*, **84**, 1407–1419.
- Manos, P.S., Zhou, Z.K. & Cannon, C.H. (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *International Journal of Plant Science*, **162**, 1361–1379.
- Mason-Gamer, R.J., Holsinger, K.E. & Jansen, R.K. (1995) Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). *Molecular Biology and Evolution*, **12**, 371–381.
- McCauley, D.E. (1995) The use of chloroplast DNA polymorphisms in studies of gene flow in plants. *Trends in Ecology and Evolution*, **10**, 198–202.
- Metcalf, I., Smith, J.M.B., Morwood, M. & Davidson, I. (2001) *Faunal and floral migrations and evolution in SE Asia-Australasia*. A.A. Balkema Publishers, Lisse.
- Morley, R.J. (1998) Palynological evidence for Tertiary plant dispersals in the SE Asian region in relation to plate tectonics and climate. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 211–234. Backhuys, Leiden.

- Morley, R.J. (2000) *Origin and evolution of tropical rain forests*. John Wiley & Sons, New York.
- Morley, R.J. (2001) Why are there so many primitive angiosperms in the rain forests of Asia-Australasia? *Faunal and floral migrations and evolution in SE Asia-Australasia* (eds I. Metcalfe, J.M.B. Smith, M. Morwood and I. Davidson). A.A. Balkema Publishers, Lisse.
- Moss, S.J. & Wilson (1998) Biogeographic implications of the Tertiary palaeogeographic evolution of Sulawesi and Borneo. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 133–163. Backhuys, Leiden.
- Muir, G., Fleming, C.C. & Schlotterer, C. (2001) Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Molecular Biology and Evolution*, **18**, 112–119.
- Muloko-Ntoutoume, N., Petit, R.J., White, L. & Abernethy, K. (2000) Chloroplast DNA variation in a rainforest tree (*Aucoumea klaineana*, Burseraceae) in Gabon. *Molecular Ecology*, **9**, 359–363.
- Nagl, S., Tichy, H., Mayer, W.E., Takahata, N. & Klein, J. (1998) Persistence of neutral polymorphisms in Lake Victoria cichlid fish. *Proceedings of the National Academy of Sciences of the USA*, **95**, 14238–14243.
- Newton, A.C., Allnutt, T.R., Gillies, A.C.M., Lowe, A.J. & Ennos, R.A. (1999) Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, **14**, 140–145.
- Ng, F.S.P. (1991) *Manual of forest fruits, seeds, and seedlings*. Forest Research Institute Malaysia, Kuala Lumpur.
- Nicholas, K.B. & Nicholas H.B. (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments, Version 2.6.
- Oddou-Muratorio, S., Petit, R.J., Le Guerroue, B., Guesnet, D. & Demesure, B. (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123–1135.
- Olsen, K.M. (2002) Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. *Molecular Ecology*, **11**, 901–911.
- van Oppen, M.J.H., Rico, C., Turner, G.F. & Hewitt, G.M. (2000) Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Molecular Biology and Evolution*, **17**, 489–498.
- Payne, J., Francis, C.M. & Philipps, K. (1985) *A field guide to the mammals of Borneo*. The Sabah Society, Kuala Lumpur.
- Penny, D. (2001) A 40,000 year palynological record from north-east Thailand; implications for biogeography and palaeo-environmental reconstruction. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **171**, 97–128.
- Penny, D. & Hendy, M.D. (1985) The use of tree comparison metrics. *Systematic Zoology*, **34**, 75–82.
- Petit, R.J., Kremer, A. & Wagner, D.B. (1993a) Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, **87**, 122–128.
- Petit, R.J., Csaikl, U.M., Bordacs, S., Burg, K., Coart, E., Cottrell, J., van Dam, B., Deans, J.D., Dumolin-Lapegue, S., Fineschi, S., Finkeldey, R., Gillies, A., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Matyas, G., Munro, R.C., Olalde, M., Pemonge, M.H., Popescu, F., Slade, D., Tabbener, H., Turchini, D., de Vries, S.G.M., Ziegenhagen, B. & Kremer, A. (2002) Chloroplast DNA variation in European white oaks – phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.
- Petit, R.J., Wagner, D.B. & Kremer, A. (1993b) Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. *Annales des Sciences Forestières*, **50**, 41s–47s.
- Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducouso, A. & Kremer, A. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Science (USA)*, **94**, 9996–10001.
- Ridder-Numan, J.W.A. (1998) Historical biogeography of *Spatholobus* (Leguminosae-Papilionoideae) and allies in SE Asia. *Biogeography and geological evolution of SE Asia* (eds R. Hall and J.D. Holloway), pp. 133–163. Backhuys, Leiden.
- Sakai, S., Momose, K., Yumoto, T., Nagamitsu, T., Nagamasu, H., Hamid, A.A. & Nakashizuka, T. (1999) Plant reproductive phenology over four years including an episode of general flowering in a lowland dipterocarp forest, Sarawak, Malaysia. *American Journal of Botany*, **86**, 1414–1436.
- Sang, J. & Soepadmo, E. (2000) Fagaceae. *Tree flora of Sabah and Sarawak* (eds E. Soepadmo, K.M. Wong and L.G. Saw). Sabah Forestry, Forest Research Institute of Malaysia, Sarawak Forestry, Kuala Lumpur.
- Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T. & Smith, W.A. (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.
- Simon, J.L. (1995) *Resampling Stats user guide*. Resampling Stats, Arlington, VA.
- Sims, H.J., Herendeen, P.S. & Crane, P.R. (1998) New genus of fossil Fagaceae from the Santonian (Late Cretaceous) of central Georgia, USA. *International Journal of Plant Science*, **159**, 391–404.
- Soepadmo, E. (1970) Florae Malesianae precursors XLIX. Malesian species of *Lithocarpus* BL. (Fagaceae). *Reinwardtia*, **8**, 197–308.
- Soepadmo, E. (1972) Fagaceae. *Flora Malesiana* (ed. C.G.G.J. van Steenis). Noordhoff International Publishing, Leyden, The Netherlands.
- Strait, D.S., Moniz, M.A. & Strait, P.T. (1996) Finite mixture coding: a new approach to coding continuous characters. *Systematic Biology*, **45**, 67–78.
- Swofford, D.L. (2000) PAUP* phylogenetic analysis using parsimony (* and other methods), Version 4.0. Sinauer Associates, Sunderland, MA.
- Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.
- Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Templeton, A.R. (1989) The meaning of species and speciation: A genetic perspective. *Speciation and its consequences* (ed. D.O.J.A. Endler), pp. 3–27. Sinauer Associates, Sunderland, MA.

- Templeton, A.R. (1998) Nested clade analyses of phylogenetic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.
- Tomaru, N., Takahashi, M., Tsumura, Y., Takahashi, M. & Ohba, K. (1998) Intraspecific variation and phylogeographic patterns of *Fagus crenata* (Fagaceae) mitochondrial DNA. *American Journal of Botany*, **85**, 629.
- Van Valen, L.M. (1982) Integration of species: stasis and biogeography. *Evolutionary Theory*, **6**, 99–112.
- Vander Wall, S.B. (2001) The evolutionary ecology of nut dispersal. *Botanical Review*, **67**, 74–117.
- Wallace, A.R. (1869) *The Malay Archipelago*. Harper and Brothers, New York.
- Webb, C.O. & Peart, D.R. (2000) Habitat associations of trees and seedlings in a Bornean rain forest. *Journal of Ecology*, **88**, 464–478.
- Whitmore, T.C. (1987) *Biogeographical evolution of the Malay Archipelago*. Oxford University Press, Oxford.
- Whittemore, A.T. & Schaal, B.A. (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Science (USA)*, **88**, 2540–2544.
- Wolfram, S. (1998) *The mathematica book*, Version 4.0.1. Wolfram Media/Cambridge University Press, Cambridge.
- Young, S.S. & Herwitz, S.R. (1995) Floristic diversity and co-occurrences in a subtropical broad-leaved forest and two contrasting regrowth stands in central-west Yunnan Province, China. *Vegetatio*, **119**, 1–14.

BIOSKETCHES

Chuck Cannon is Assistant Professor of Biology at Texas Tech University. He has been studying various aspects of tropical rain forest in Southeast Asia since 1987, primarily on the evolution and maintenance of tree diversity.

Paul Manos is Associate Professor of Biology at Duke University. Over the past 15 years he has worked on the phylogenetic biology of tree species in the angiosperm order Fagales (Fagaceae, Juglandaceae, Nothofagaceae).