Molecular Ecology (2007) doi: 10.1111/j.1365-294X.2007.03294.x

The geography of diversification in mutualistic ants: a gene's-eye view into the Neogene history of Sundaland rain forests

S.-P. QUEK,* S. J. DAVIES,‡** P. S. ASHTON,†§ T. ITINO¶ and N. E. PIERCE*
*Museum of Comparative Zoology and †Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA, ‡Center for Tropical Forest Science — Arnold Arboretum Asia Program, 22 Divinity Avenue, Cambridge, MA 02138, USA, §Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, UK, ¶Shinshu University, Department of Biology, Faculty of Science, 3-1-1 Asahi, Matsumoto Nagano 390-8621, Japan

Abstract

We investigate the geographical and historical context of diversification in a complex of mutualistic *Crematogaster* ants living in *Macaranga* trees in the equatorial rain forests of Southeast Asia. Using mitochondrial DNA from 433 ant colonies collected from 32 locations spanning Borneo, Malaya and Sumatra, we infer branching relationships, patterns of genetic diversity and population history. We reconstruct a time frame for the ants' diversification and demographic expansions, and identify areas that might have been refugia or centres of diversification. Seventeen operational lineages are identified, most of which can be distinguished by host preference and geographical range. The ants first diversified 16–20 Ma, not long after the onset of the everwet forests in Sundaland, and achieved most of their taxonomic diversity during the Pliocene. Pleistocene demographic expansions are inferred for several of the younger lineages. Phylogenetic relationships suggest a Bornean cradle and major axis of diversification. Taxonomic diversity tends to be associated with mountain ranges; in Borneo, it is greatest in the Crocker Range of Sabah and concentrated also in other parts of the northern northwest coast. Within-lineage genetic diversity in Malaya and Sumatra tends to also coincide with mountain ranges. A series of disjunct and restricted distributions spanning northern northwest Borneo and the major mountain ranges of Malaya and Sumatra, seen in three pairs of sister lineages, further suggests that these regions were rain-forest refuges during drier climatic phases of the Pleistocene. Results are discussed in the context of the history of Sundaland's rain forests.

Keywords: biogeography, *Crematogaster*, Cytochrome Oxidase, *Macaranga*, rain forest, Southeast Asia

Received 20 August 2006; revision received 10 January 2007; accepted 16 January 2007

Introduction

The Asian tropics are home to some of the greatest concentrations of species diversity, including a group of *Macaranga* (Euphorbiaceae) trees whose hollow stems house mutualistic *Crematogaster* (Myrmicinae) ants which defend the trees against vines and herbivores in return for domatia and food bodies (Fiala *et al*. 1989). The association is obligate for both parties, and its distribution is tightly correlated with an everwet climate. Typical of ant–plant associations, it does not survive drought or seasonality (e.g. Janzen 1973) and is thus restricted to rain forests in the truest sense, as opposed to the seasonal, monsoon or savanna forests that also occur in other parts of Southeast Asia (Whitmore 1998).

The aim of this study is to investigate the geographical and historical context of diversification in this species complex of mutualistic ants that inhabit *Macaranga* trees. Because of their exclusive association with rain forests, their biogeographical and evolutionary history might illuminate some of the history of the Sunda-shelf rain forests they inhabit.

Correspondence: Swee-Peck Quek, Fax: +1617 4955667; E-mail: queksp@gmail.com
**Current Address: Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Panamá, República de Panamá.

In inferring the historical biogeography of Sundaland's rain-forest biota, the archipelagic setting and a history of fluctuating sea levels add a challenging dimension of complexity, but also provide a spatio–temporal frame of reference, as the alternating connections and disconnections among the islands correlated with dry/cool and wet/warm climate associated with glacial and interglacial phases (Haile 1971; Flenley 1984; Heaney 1991). In particular, the Pleistocene was marked by cycles of extensive cooling and drying of global climates that may have contracted rain forests worldwide to isolated refugia (Muller 1972; Haffer 1987; Colinvaux *et al*. 1996; Maley 1996; Morley 2000). Such events would have left their signatures on the geographical distribution of genetic and taxonomic diversity in rain-forest biota. At present, Malaya, Borneo and Sumatra (Fiala *et al*. 1999) constitute most of the presently emergent parts of the Sunda shelf (or Sundaland). The distribution of myrmecophytic *Macaranga* and their *Crematogaster* ants (hereon referred to by their subgeneric designation, *Decacrema*) is strictly limited to the everwet forests of the Sunda shelf (Fiala *et al*. 1999; Davies *et al*. 2001). This limited distribution may be explained by the requirement for continuous production of food bodies (on stipules or young leaves) for sustaining resident ant colonies, which likely depend in turn on the continuous production of new leaves (Davies *et al*. 2001). High-light environments, coinciding with forest gaps and edges and recently logged areas, are their preferred habitats, and thus they often constitute conspicuous members of secondary forests in the region.

Despite the publication of three molecular phylogenetic studies of the *Decacrema* from *Macaranga* (Itino *et al*. 2001; Feldhaar *et al*. 2003; Quek *et al*. 2004), a formal taxonomic treatise of the entire group is still lacking, in part because of the challenging nature of the morphological taxonomy of this group and *Crematogaster* in general (Longino 2003). Informal morphological groupings were not completely upheld by the mitochondrial DNA (mtDNA) phylogeny in both Feldhaar *et al*. (2003) and Quek *et al*. (2004). Feldhaar *et al*. (2003) recognized eight morphospecies in three groups, while Quek *et al*. (2004) identified 10 mtDNA lineages based on host specificity and geographical distribution. In this study, we extend the work of Quek *et al*. (2004), in combination with new data and the data from Feldhaar *et al*. (2003), to investigate the biogeography, phylogeography and demographic history of *Decacrema*. We aim to: (i) identify regions of taxonomic and genetic diversity and thus infer the areas that might have been refugia or centres of diversification as opposed to recently colonized areas; (ii) identify past demographic expansions and their taxonomic, geographical and temporal correlates; and (iii) reconstruct the approximate time frame for the ants' diversification, refugial occurrence and dispersals among the islands. Results are discussed in the context of the history of the Sunda shelf rain forests.

Materials and methods

Sampling and molecular data

This study incorporates data from Quek *et al*. (2004), Feldhaar *et al*. (2003) and new data totalling 433 ingroup samples from 32 locations (nine in Borneo, six in Sumatra, 14 in Malaya and six from six smaller islands neighbouring Sumatra and Malaya). The new data increase the sampling from a few locations reported in Quek *et al*. (2004) and further add nine new locations to this study — one in Borneo, six in (or near) Malaya and two in (or near) Sumatra. The distribution of sampling localities spans almost the entire range of the known distribution of *Macaranga*-associated *Decacrema*. Multiple individual trees per host species were sampled whenever possible, and most of the host species are represented in this study. A GPS unit was used to record the coordinates and altitude of each colony where possible. Host species, collection localities, elevation and GenBank accession numbers of the samples are presented in Table S1, Supplementary material.

Two overlapping datasets were used in this study. Dataset 1 comprises 565 nucleotide bp of the mitochondrial gene cytochrome oxidase subunit I (COI) for 395 ingroup samples and was used for population level analyses. Dataset 2 comprises 1879 bp (1324 bp of COI including the COI region in Dataset 1, and 555 bp of cytochrome oxidase subunit II, COII) for 21 ingroup exemplars representing the major lineages in Dataset 1, and was used for inferring relationships among lineages. For estimating divergence times, Dataset 2 minus COII was used. The polymerase chain reaction (PCR) primers and protocols for Dataset 1 are described in Quek *et al*. (2004) while for Dataset 2, the external primers reported in Feldhaar *et al*. (2003) were used as described. PCR products were directly sequenced in both directions. Sequences from Feldhaar *et al*. (2003) were obtained from GenBank (33 ingroup samples containing 483 bp of COI and 439 bp of COII) and included into Dataset 2.

Phylogenetic inference

Phylogenetic inference was optimized for maximum likelihood (ML) in the program PHYML (Guindon & Gascuel 2003). Substitution models for ML analyses were selected using the Akaike Information Criterion (AIC, Akaike 1974) and the hierarchical likelihood ratio test (hLRT, Huelsenbeck & Rannala 1997), both implemented in MODELTEST 3.06 (Posada & Crandall 1998). Where the AIC and hLRT differed in the model specified, the model with fewer parameters was used. The HKY + I + G model (Hasegawa *et al*. 1985) was used for Dataset 1 and the TrN + I + G model (Tamura & Nei 1993) for Dataset 2 with and without COII. Replicate sequences were removed and only unique haplotypes were used for phylogenetic inference. All

model parameters were optimized during tree searching in PHYML. Clade support was assessed with 500 bootstrapped data sets. To further assess the topology obtained from ML analyses, Bayesian posterior probabilities and parsimony bootstrap support were obtained using MRBAYES 3.1.1 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003) and PAUP* 4.0b10 (Swofford 2002), respectively. The HKY + I + G and GTR + I + G (in place of TrN + I + G in MRBAYES) models were used in Bayesian analyses for Datasets 1 and 2, respectively. In each Bayesian analysis, two independent runs lasting three million generations each were done, and half of the trees were discarded, well after stationarity in likelihood scores were reached. The parsimony bootstrap procedure consisted on 500 pseudoreplicates, each comprising 50 random addition sequence replicates and tree-bisection-reconnection branch swapping.

To assign the samples from Feldhaar *et al*. (2003) to the clades obtained from Datasets 1 and 2, an additional analysis was performed combining Dataset 2 with data from Feldhaar *et al*. (2003). We used phytoecious *Decacrema* inhabiting *Neonauclea*-ant plants in Sulawesi as the outgroup since an earlier analysis indicated that these were the sister taxon to the *Decacrema* from *Macaranga* (Quek *et al*. 2004). For biogeographical reconstruction, the island of origin for each lineage was traced onto the phylogeny using the most parsimonious resolving option in MACCLADE (Maddison & Maddison 1992).

Defining operational lineages

Species boundaries in the *Decacrema* species complex remain uncertain due to conflicting results from morphological taxonomy and mtDNA analyses (Feldhaar *et al*. 2003; Quek *et al*. 2004), and due to the absence thus far of multilocus genetic data for inferring relationships. In delimiting our operational lineages, we sought to approximate gene-flow boundaries based on the presence of large or obvious phylogenetic breaks giving rise to well supported monophyletic groupings, usually a cluster of shallow divergences atop a long stem. These delimitations do not necessarily reflect species boundaries, but if the samples within the defined lineages are cohesive in host association, then these delimitations may correspond to discrete species. Specificity to host species, host clade or host stem traits was determined by Chi-Square tests of the extent of departure of the observed host proportion from the expected host proportion as indicated by the frequency of the given host in all the locations where the ant clade was sampled. Five thousand Monte Carlo simulations of the sampling distribution were also performed to assess the observed sample against the null distribution. Host stem traits have been demonstrated to impose barriers to colonization by certain *Decacrema* clades (Federle *et al*. 1997; Quek *et al*. 2004). Therefore, as a conservative measure, we used this information to delimit the pool of hosts available to the ant clade in the test of host preference. The pool of available hosts were grouped into the *bancana* Clade (smooth-stemmed hosts), the waxy stemmed hosts (a paraphyletic grade), and the *Pruinosae* Clade, whose solid stems become hollow through excavation by ants. The first two groups constitute a clade with naturally hollow stems).

Population genetic and demographic history analyses

Within each lineage, genetic diversity for each location and for the entire lineage was measured using haplotype diversity, h (Nei 1987; equation 8.5) and nucleotide diversity, π (Tajima 1983), calculated in ARLEQUIN 2.000 (Schneider *et al*. 2000). Haplotype diversity is a measure of the probability that two randomly chosen samples will be different, and nucleotide diversity is the mean proportion of nucleotide sites differing in all pairwise comparisons. Grant & Bowen (1998) propose that high h and high π indicate large, stable populations with long evolutionary histories or secondary contact between differentiated lineages, high h and low π suggest a bottleneck followed by rapid growth and accumulation of mutations, and low h and low π suggest a recent bottleneck or founder event by a single or few lineages.

For lineages whose gene trees exhibit well-defined internal structures, geographical regions were mapped onto the phylogeny to infer ancestral vs. derived locations. For lineages with star-like gene trees, suggestive of sudden rapid population expansion, a minimum spanning network was constructed in TCS (Clement *et al*. 2000).

The effective population size parameter, Θ, and exponential growth rate, g, and their 95% confidence intervals, were jointly estimated using the program LAMARC (Kuhner *et al*. 2005). LAMARC implements coalescence theory and ML to estimate population parameters, summed over all possible genealogies. Each LAMARC analysis consisted of three replicates, and each replicate consisted of 15 initial chains (sampling interval 20, 2000 starting genealogies discarded and 2000 subsequent genealogies used) used as starting estimates for three final chains (sampling interval 20, 2000 starting genealogies discarded and 2000 subsequent genealogies used).

Signals of rapid demographic expansion were detected using Tajima's D (Tajima 1989a, b) and Fu's F-tests (Fu 1997). Both tests, as well as the mismatch distribution analysis (Slatkin & Hudson 1991; Rogers & Harpending 1992) were done in ARLEQUIN. The mismatch distribution (frequency distribution of observed number of pairwise differences) is expected to be unimodal in a population that has experienced sudden exponential growth, as most coalescent events are concentrated within a narrow window of time and occur relatively early, producing a star-like gene phylogeny (Slatkin & Hudson 1991). Demographic expansion tests and the mismatch analysis were conducted only for clades

showing starburst phylogenies. Significance of the deviations of the observed mismatch from the expected mismatch distribution under a model of expansion was assessed by generating 5000 simulations of the expected mismatch using parametric bootstrapping (Schneider & Excoffier 1999). The resulting distribution of the sum of square deviations (SSDs) between the observed and the simulated mismatch is used to assess the fit of the expansion model, where the *P*-value is the proportion of simulated SSDs that are larger than or equal to the observed SSD. The expansion age, *t*, was calculated using the expansion age parameter τ ($= 2\,\mu t$) and the mutation rate of the entire sequence, μ (= 1.5% divergence per million years between contemporary alleles, or 0.75% between ancestor and descendent alleles; see Age estimation methods). Ninety-five-percent confidence intervals for τ were inferred from the same 5000 simulations. The population size parameters before and after expansion, θ_0 ($= 2\,\mu N0$) and θ_1 ($= 2\,\mu N1$), were also calculated in ARLEQUIN. Confidence intervals for θ were not used because they were deemed to be too conservative (Schneider *et al*. 2000).

Age estimation

Divergence dates were estimated using COI sequences alone (Dataset 2 without COII sequences). In particular, we sought to date the divergences that resulted in sister lineages on different islands to reconstruct the time frame when *trans*-Sunda migrations might have occurred. We restricted this analysis to COI because it exhibits the least rate heterogeneity (Gaunt & Miles 2002), has been widely used for dating in arthropods (e.g. ants, Degnan *et al*. 2004), and shows a general rate of about 1.5% (mean uncorrected pairwise distance) per million years as measured in several arthropod taxa (Quek *et al*. 2004). The phylogeny was subjected to Sanderson 1997) nonparametric rate smoothing in TREEEDIT 1.0 (Rambaut & Charleston 2002). Three nodes of varying genetic divergences (4.9–10.4%) were assigned fixed ages based on 1.5% divergence per million years, and from each fixed node, the ages of the other nodes were calculated. This resulted in a spread of ages inferred for each node. Nodes with approximately equivalent levels of divergence in the descendent lineages (i.e. balanced nodes) were selected for fixing. Raw pairwise divergences at the fixed nodes were calculated with MEGA 2.1 (Kumar *et al*. 2001).

Results

Phylogeny and host specificity of lineages

The topologies resulting from ML analyses of Dataset 1 and Dataset 2 are presented in Fig. 1. Dataset 2 produced a topology with better supported nodes deeper in the phylogeny. Following Quek *et al*. (2004), operational lineages were designated by the letters A through K (except 'I'), with a new lineage, L, found mostly from the hitherto unsampled Malayan high elevation; additionally, two clades, C and G, were here divided into a further two and six lineages, respectively, resulting in 17 operational lineages defined. All sequences from Feldhaar *et al*. (2003) were unambiguously assigned to these lineages (see Table S1, Supplementary material). Because sampling was carried out without regard to a priori assumptions of species or taxonomic boundaries, the different lineage sizes approximated their natural abundances, at least in the sampled locations.

Each lineage differed from its sister lineage in geographical distribution or in the combination of geographical distribution and host preference. The Chi-Square tests of host specificity indicated that of the 17 lineages, 11 were significantly host-specific (A, B, Cb, Cms, D, E, F, G1, G5, H and K), and G2 was on the margin of statistical significance (Table 1 and Fig. 2). Clade L was not tested because it was found only in high elevations (~700–1400 m) in Malaya and Sumatra, and its host, *Macaranga hullettii*, was the only *Macaranga* myrmecophyte found at high elevations there, introducing a confounding variable. The only lineages not showing host-specificity were those with small sample sizes (G3, G4, Gs and J), possibly reflecting lack of statistical power.

Faunal relationships among islands

Borneo is the centre of diversity of the *Decacrema* associated with *Macaranga*, harbouring 11 lineages (10 of which are endemic), about twice the number found in Malaya (6) and Sumatra (5), which has one endemic each (Table 2). Parsimony-based reconstruction of ancestral areas (Fig. 2) placed their main axis of diversification in Borneo. Within Borneo, most of the diversity occurs in the northern northwest (nNW — represented in this study by the Crocker Range, Lambir and Mulu + Murud), and most of the diversity in Malaya and Sumatra occurs in mountainous areas, in Cameron Highlands and Bukittinggi, respectively (Fig. 3).

Of the six non-Bornean lineages, four are either endemic to, or have greater genetic diversity in, Sumatra compared to Malaya, despite poorer sampling in the former (Table 3): Gs is endemic there, J is predominantly Sumatran, and both K and Cms have markedly higher genetic diversities there, suggesting Sumatran origins for these four.

The lineage-level phylogeny of the ants (Fig. 2) has five terminal nodes, four of which represent divergences across the South China sea. Of particular note among these are three instances of sister lineages (Cb–Cms, L–E and G2–Gs), split between nNW Borneo and high-elevation locations in Malaya and Sumatra, corresponding to the major mountain ranges (Figs 2 and 3).

The elevational ranges of clades show differences in Borneo vs. Malaya and Sumatra. Borneo harbours lineages

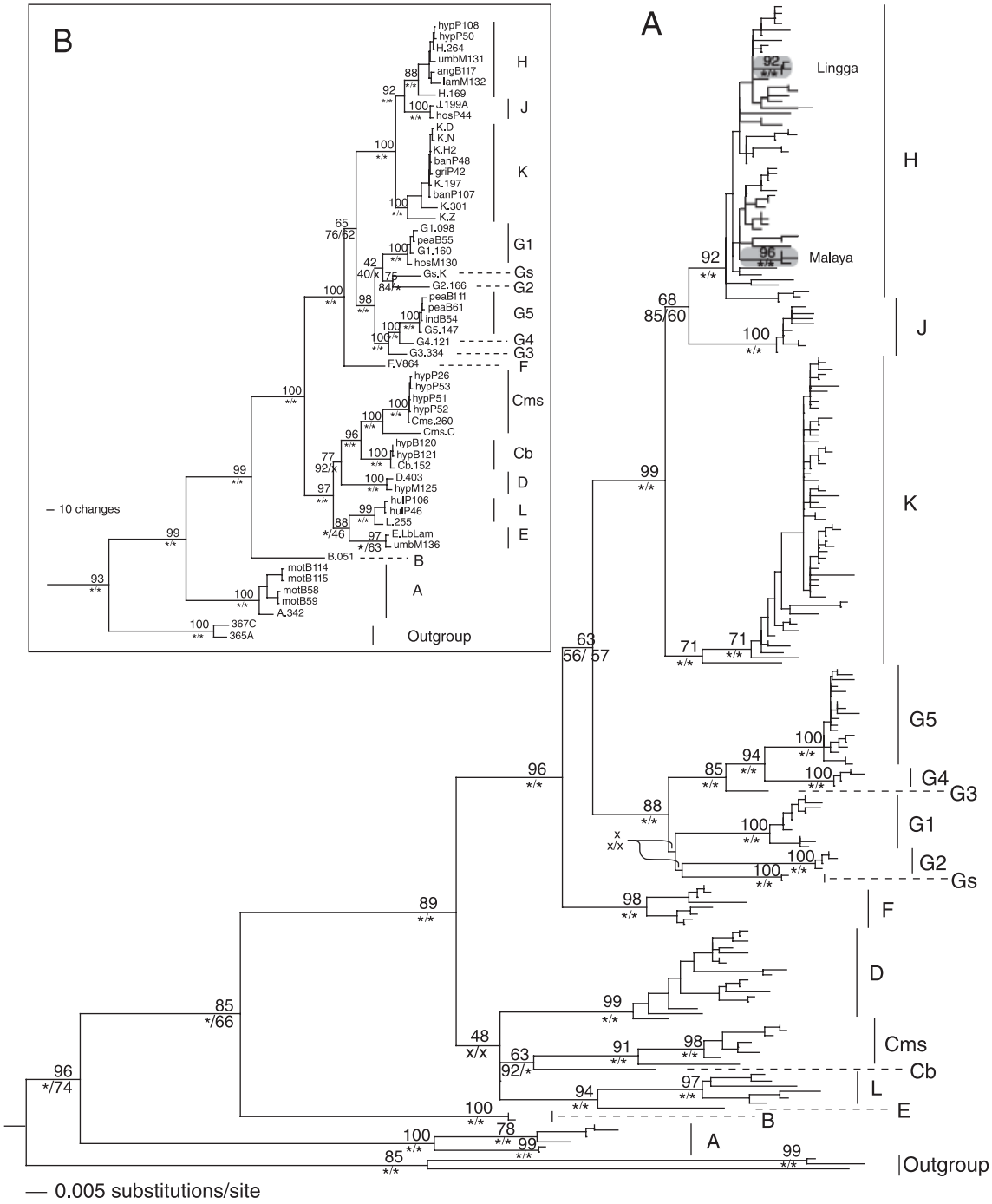

Fig. 1 (A) COI gene tree of *Decacrema* samples based on maximum-likelihood (ML) analysis of Dataset 1. (B) COI + COII phylogeny of exemplars from Dataset 1 plus data from Feldhaar *et al*. (2003) (ML analysis, Dataset 2). In both trees, only unique haplotypes were used for phylogenetic inference. ML bootstrap support is indicated above branches. Bayesian posterior probabilities, followed by parsimony bootstrap support are shown below branches. An asterisk indicates posterior probability ≥ 95% or parsimony bootstrap support ≥ 75%. An 'x' indicates the node was not present in the parsimony or ML bootstrap tree or in Bayesian analysis. *Decacrema* from ant plants in Sulawesi were used as outgroups. The exemplar tree inferred from COI + COII (shown in B) appears better resolved than the tree based on COI alone.

Table 1 Host specificity as determined by Chi-square tests for biased host association or by 5000 Monte Carlo simulations of the sampling distribution. Expected proportions are the host taxon/type tested as a proportion of the available hosts. B, *bancana* Clade; P, *Pruinosae* Clade; W, waxy hosts (see also Fig. 2). *P*-values (nondirectional) for significant host preference: **P* = 0.05; ***P* = 0.01; ****P* = 0.001; *****P* = 0.0001; ns, not significant; (—) indicates host avoidance. Data from Feldhaar *et al*. (2003) are included

Ant clade	*n*	Host taxon/ type tested	Available hosts	Expected proportion	Observed proportion	*P*
A	13	*motleyana*	W	0.54	0.85	*
		motleyana	W + B	0.22	0.85	****
		W	W + B	0.41	1.00	****
		P	W + B + P	0.13	0.00	ns
B	2	*constricta*	W	0.18	1.00	*
		constricta	W + B	0.13	1.00	**
		W	W + B	0.73	1.00	ns
		P	W + B + P	0.12	0.00	ns
C	19	*hypoleuca*	W	0.58	1.00	***
		hypoleuca	W + B	0.21	1.00	****
		W	W + B	0.36	1.00	****
		P	W + B + P	0.11	0.00	ns
Cb	4	*hypoleuca*	W	0.29	1.00	*
		hypoleuca	W + B	0.11	1.00	****
		W	W + B	0.36	1.00	*
		P	W + B + P	0.16	0.00	ns
Cms	15	*hypoleuca*	W	0.96	1.00	ns
		hypoleuca	W + B	0.33	1.00	****
		W	W + B	0.34	1.00	****
		P	W + B + P	0.06	0.00	ns
D	31	*hypoleuca* group	W	0.47	0.87	***
		hypoleuca group	W + B	0.27	0.87	****
		W	W + B	0.57	0.97	****
		P	W + B + P	0.09	0.00	* (—)
E	4	*lamellata* group	W	0.35	1.00	*
		lamellata group	W + B	0.19	1.00	***
		W	W + B	0.55	1.00	ns
		P	W + B + P	0.14	0.00	ns
F	19	B	W + B	0.50	1.00	****
		P	W + B + P	0.08	0.00	ns
G1	14	B	W + B + P	0.49	0.07	** (—)
		W	W + B + P	0.36	0.07	* (—)
		P	W + B + P	0.16	0.86	****
G2	5	B	W + B + P	0.53	1.00	0.06
		W	W + B + P	0.30	0.00	ns
		P	W + B + P	0.16	0.00	ns
G3	2	B	W + B	0.46	1.00	ns
		P	W + B + P	0.00	0.00	ns
G4	4	B	W + B + P	0.53	1.00	ns
		W	W + B + P	0.30	0.00	ns
		P	W + B + P	0.16	0.00	ns
G5	37	B	W + B + P	0.45	0.05	**** (—)
		W	W + B + P	0.42	0.57	ns
		P	W + B + P	0.13	0.38	****
Gs	2	B	W + B	0.85	1.00	ns
		P	W + B + P	0.00	0.00	ns
H	120	B	W + B + P	0.48	0.68	****
		W	W + B + P	0.43	0.26	*** (—)
		P	W + B + P	0.09	0.06	ns
J	13	B	W + B + P	0.56	0.77	ns
		W	W + B + P	0.37	0.08	ns
		P	W + B + P	0.07	0.15	ns
K	137	B	W + B + P	0.34	0.43	*
		W	W + B + P	0.53	0.48	ns
		P	W + B + P	0.13	0.09	ns

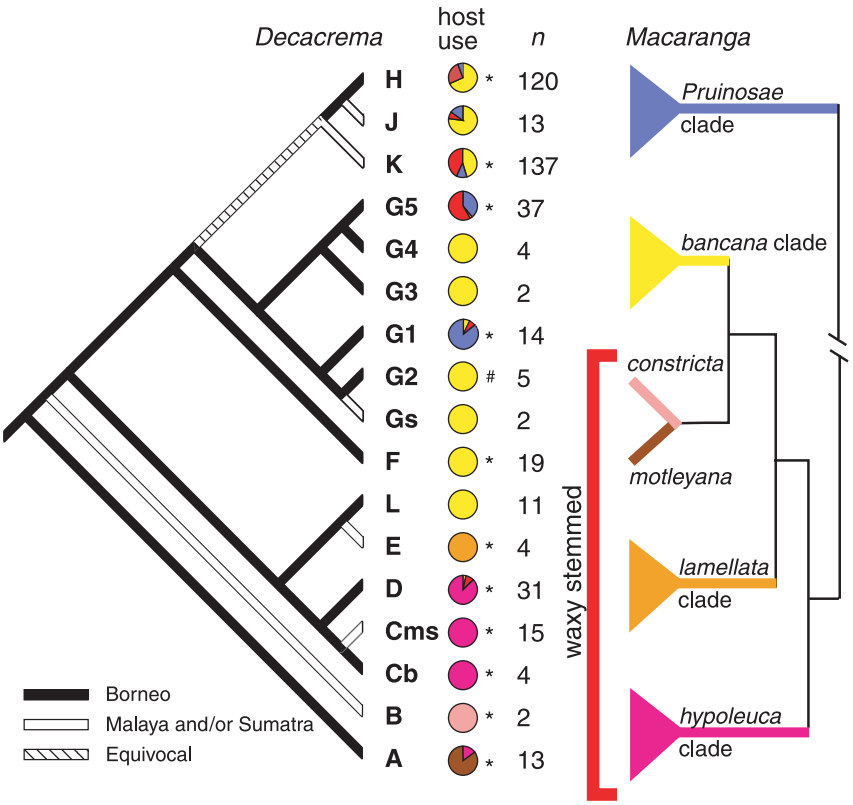

Fig. 2 Host use and specificity of *Decacrema* lineages, represented by pie charts coloured to match *Macaranga* host taxon in the phylogeny at right. Data are from Table 1. A statistically significant host preference ($P < 0.05$) is indicated by an asterisk adjacent to the sector coloured to represent the preferred host (e.g. members of the *Pruinosae* Clade, represented by blue, are the preferred host in lineage G5). # indicates $P = 0.06$. Lineage L was not tested due to the confounding factor of altitude specificity. A partial *Macaranga* phylogeny based on Davies (2001) is shown here.

Table 2 List of *Decacrema* lineages found on the three main land masses in this study. Bold font denotes endemic lineages

Borneo (11)	**A, Cb, D, E, F, H, G1, G2, G3, G4, G5**
Malaya (6)	**B**, Cms, H, J, K, L
Sumatra (5)	Cms, **Gs**, J, K, L

that are able to colonize the entire elevational range (up to 1400 m above sea level) of myrmecophytic *Macaranga*, whereas in Malaya and Sumatra each clade appears to reach its lower or upper limit in the 600–800 m zone (Fig. 4).

Local lineage and genetic diversity

The distributions of lineages and lineage compositions within sampling locations are presented in Fig. 3. The highest diversity in Malaya occurs in Cameron Highlands (four lineages), in Sumatra in the Bukittinggi region (five lineages, including one endemic), and in Borneo, in the Crocker Range, harbouring seven lineages. The great diversity in the Crocker range is contributed in large part by three apparently endemic lineages (Cb, G2 and G4). The lowest diversity locations in Borneo (Samarinda, Kuching and Siduk, with two lineages each) are in the lowlands and this is generally, but not always, true in Malaya and Sumatra (Muara Tembesi, Johor, Bauk, and the islands of Penang, Pangkor, Singapore and Bintan, with one lineage each), which have the additional confounding variable of an insular situation in many cases.

Haplotype diversities, h, and nucleotide diversities, π, are presented in Table 3. For each lineage, π-values ranged from 0.18% to 2.8% and fell into two clusters, with a break observed between 0.88% and 1.52%. This interval was used to separate high vs. low nucleotide diversities. Haplotype diversity (h) values for all lineages were high, at > 0.83 except for Clade E, measuring 0.67. High π-values ($> 1.5\%$), suggestive of demographically large, stable populations with long evolutionary histories, or geographical divergence, were found in A, Cms, D, L, E, F, H and L. Low π-values ($< 0.88\%$), suggesting bottlenecks or founding events, were found in Clades B, Cb, J, K G1, G2, G3, G4, G5 and Gs.

Malaya and Sumatra share many similarities in the patterns of genetic diversity. There, locations harbouring endemic lineages or high P-values within each lineage (bold font in Table 3) are associated with mountains or hills (e.g. Tiga Puluh Mountains, northern Bukit Barisan Range, Cameron Highlands and Beserah Hill). Lineages with long evolutionary histories and historically large populations ($\pi > 1.5\%$ in L and Cms, Table 3) are also associated with higher elevation, in addition to being patchily distributed (Figs 3 and 4). Clades there associated with founding events or bottlenecks (B, J, K and H) are found in low as well as middle elevations (Table 3 and Fig. 4). In Borneo, by

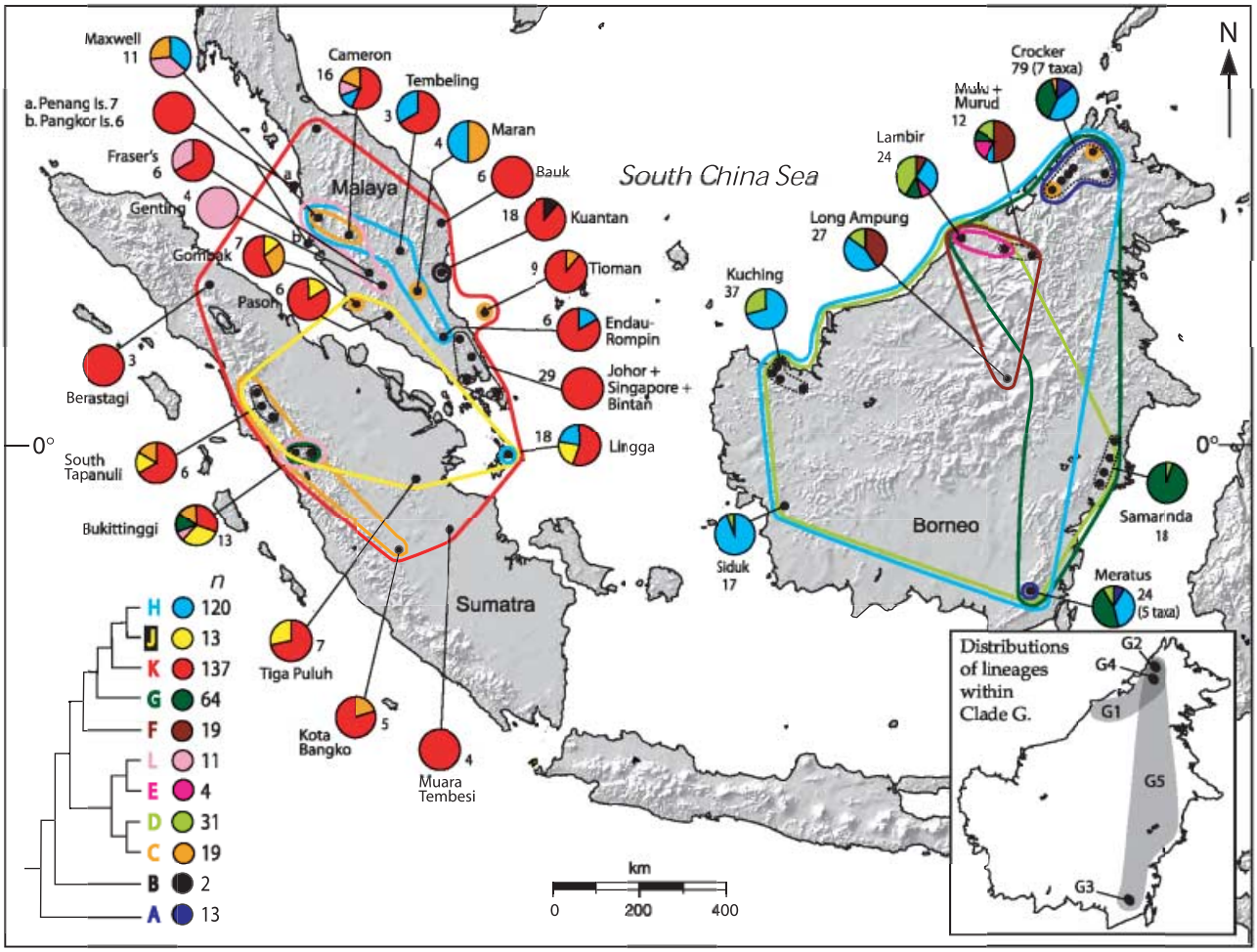

Fig. 3 Approximate distributional ranges of *Decacrema* lineages (coloured lines) as inferred by sampling locations. Pie charts show inferred lineage composition of locations (proportional to abundance as sampled), with sample sizes indicated. Pie charts and distribution lines are colour-referenced to the ant phylogeny at bottom left. Clades Cb and Cms (C in Borneo and Malaya + Sumatra, respectively), and Gs (in Sumatra), G1, G2, G3, G4 and G5 are not detailed in the phylogeny (see Fig. 1 for detail).

contrast, high genetic diversities or endemism were found in both mountainous (Crocker Range, Long Ampung, Meratus Mountains and Mulu) and lowland (Kuching and Lambir) locations, and old clades were not predominantly found in mountainous locations (as they were in Malaya and Sumatra), which also contain clades with low π-values.

In summary, locations on large mountain ranges harbour the greatest number of clades within their respective islands (Cameron Highlands in Malaya, Bukittinggi in Sumatra, Crocker Range in Borneo). In Malaya and Sumatra, mountain ranges also hold the greatest within-clade genetic diversity, even for clades predominating in the lowlands (Clades J and K in Tiga Puluh Mountains), and they are home to clades with long evolutionary histories and historically large populations. In Borneo, old and historically large clades are widespread and found in both lowland and mountainous areas, and both areas harbour high within-clade genetic diversities.

Several locations showed consistently low π-values within all their resident lineages: Lingga island ($\pi < 0.25\%$), Meratus Mountains ($\pi < 0.2\%$) and, despite having high lineage diversity, Lambir ($\pi < 0.41\%$).

Demographic history and divergence dates

Results for tests of population expansion are presented in Table 4. Only clades showing star-like topologies were tested (G5, H, J and K in Fig. 1a). Clades G5, H and K had significantly negative test statistics for one or both of Tajima's D and Fu's F, while Clade J was on the margin of significance for both tests.

Mismatch distributions for G5, H, J and K are shown in Fig. 5 and are not significantly different from the expected expansion model under the SSD test (Table 4). All demographic expansions, including their 95% confidence intervals, were estimated to have occurred in the Pleistocene.

Table 3 Haplotype diversity (*h*) and nucleotide diversity (π) of *Decacrema* clades within sampling locations, based on COI. For each species, locations with the highest π-values are indicated in bold. The COI fragment used in all analyses is from the 565-bp Dataset 1, except for Clade Cb and Clade E (marked with *), where the COI region from Feldhaar *et al*. (2003) was used

Ant clade	Location	*h*	π (%)	*n*	No. of haplotypes	No. of changes	Seq. length
A	pooled (B)	0.92 ± 0.09	2.78 ± 1.56	9	7	37	541
	Meratus	1.00 ± 0.50	0.19 ± 0.26	2	2	1	541
	Crocker	0.86 ± 0.14	**1.06 ± 0.66**	7	5	12	541
B	Kuantan (M)	1.00 ± 0.50	0.18 ± 0.26	2	2	1	546
Cb	Crocker (B)	0.83 ± 0.22	0.76 ± 0.58	4	3	7	463*
Cms	pooled	0.93 ± 0.07	1.89 ± 1.05	11	8	36	555
	Sumatra pooled	1.00 ± 0.18	2.76 ± 1.88	4	4	30	555
	S Tapanuli & Bukittinggi (S)	1.00 ± 0.27	**3.36 ± 2.59**	3	3	28	555
	Kota Bangko (S)	—	—	1	1	—	555
	Malaya pooled	0.81 ± 0.13	0.77 ± 0.50	7	4	12	555
	Cameron (M)	0.67 ± 0.31	0.24 ± 0.25	3	2	2	555
	Tioman Island (M)	—	—	1	1	—	555
	Maxwell (M)	—	—	3	1	—	555
D	pooled (B)	0.94 ± 0.03	2.09 ± 1.08	30	17	56	558
	Lambir & Mulu	0.62 ± 0.10	0.40 ± 0.27	11	3	5	558
	Lambir	0.64 ± 0.10	0.41 ± 0.27	10	3	5	558
	Kuching	0.93 ± 0.07	**2.30 ± 1.27**	11	8	35	558
	Meratus	—	—	2	1	—	558
	Siduk	—	—	1	1	—	558
	Long Ampung	0.83 ± 0.22	0.45 ± 0.36	4	3	4	558
	Samarinda		—	—	1	1	—
L	pooled	0.83 ± 0.13	2.48 ± 1.40	9	6	38	553
	Malaya pooled	0.79 ± 0.15	2.14 ± 1.24	8	5	28	553
	Cameron (M)	1.00 ± 0.50	**3.07 ± 3.16**	2	2	17	553
	Fraser's (M)	1.00 ± 0.50	1.09 ± 1.17	2	2	6	553
	Maxwell (M)	—	—	4	1	—	553
	Bukittinggi (S)	—	—	1	1	—	553
E	Lambir & Mulu (B)	0.67 ± 0.20	1.52 ± 1.08	4	2	11	483*
	Lambir	—	—	2	1	—	483*
	Mulu	—	—	2	1	—	483*
F	pooled (B)	0.87 ± 0.06	1.72 ± 0.93	18	8	25	565
	Long Ampung	0.69 ± 0.13	0.61 ± 0.38	11	4	8	565
	Murud	—	—	2	1	—	565
	Mulu	0.67 ± 0.31	0.12 ± 0.15	3	2	1	565
	Lambir	—	—	2	1	—	565
G1	pooled (B)	0.91 ± 0.07	0.81 ± 0.48	12	8	12	548
	Crocker	0.83 ± 0.10	0.38 ± 0.26	9	5	4	548
	Lambir	1.00 ± 0.27	0.37 ± 0.34	3	3	3	548
G2	Crocker (B)	0.90 ± 0.16	0.33 ± 0.27	5	4	4	548
Gs	Bukittinggi (S)	1.00 ± 0.50	0.18 ± 0.26	2	2	1	548
G3	Meratus (B)	—	—	2	1	—	548
G4	Crocker (Borneo)	1.00 ± 0.18	0.18 ± 0.18	4	4	2	548
G5	pooled (B)	0.87 ± 0.05	0.51 ± 0.31	34	18	22	548
	Crocker	0.96 ± 0.08	0.73 ± 0.47	8	7	10	548
	Meratus	0.22 ± 0.17	0.04 ± 0.06	9	2	1	548
	Samarinda	0.90 ± 0.05	0.40 ± 0.26	17	10	14	548
H	pooled	0.98 ± 0.01	1.81 ± 0.93	110	50	83	537
	Long Ampung (B)	0.94 ± 0.06	**1.37 ± 0.78**	12	9	19	537
	Crocker (B)	0.89 ± 0.03	**1.39 ± 0.74**	31	11	25	537
	Kuching (B)	0.95 ± 0.02	**1.45 ± 0.78**	23	14	40	537
	Lambir & Murud (B)	0.86 ± 0.14	0.67 ± 0.45	7	5	12	537
	Lambir (B)	0.88 ± 0.13	0.32 ± 0.25	6	4	4	537
	Siduk (B)	0.83 ± 0.08	**1.34 ± 0.74**	16	8	29	537
	Crocker coastal flank (B)	0.56 ± 0.17	0.40 ± 0.29	9	3	5	537

Table 3 *Continued*

Ant clade	Location	*h*	π (%)	*n*	No. of haplotypes	No. of changes	Seq. length
	Meratus (B)	—	—	9	1	—	537
	Lingga Island (near S)	0.50 ± 0.27	0.09 ± 0.12	4	2	1	537
	Malaya	0.46 ± 0.20	0.21 ± 0.17	8	3	3	537
J	pooled (M + S)	0.91 ± 0.08	0.69 ± 0.42	12	9	17	545
	Tiga Puluh Mountains (S)	1.00 ± 0.50	**1.10 ± 1.19**	2	2	6	545
	Bukittinggi (S)	1.00 ± 0.18	0.43 ± 0.35	4	4	4	545
	S Tapanuli & Bukittinggi (S)	1.00 ± 0.13	0.59 ± 0.43	5	5	7	545
	Lingga island (near S)	—	—	4	1	—	545
	Pasoh (Malaya)	—	—	1	1	—	545
K	pooled (M + S)	0.95 ± 0.01	0.88 ± 0.48	128	57	74	510
	Sumatra pooled	0.94 ± 0.04	1.33 ± 0.73	21	15	41	510
	Tiga Puluh Mountains (S)	1.00 ± 0.13	**2.59 ± 1.65**	5	5	26	510
	Bukittinggi (S)	1.00 ± 0.27	0.39 ± 0.37	3	3	3	510
	Muara Tembesi (S)	0.50 ± 0.27	0.59 ± 0.46	4	2	6	510
	Kota Bangko (S)	0.67 ± 0.31	0.26 ± 0.27	3	2	2	510
	Berastagi (S)	0.83 ± 0.22	0.30 ± 0.26	4	3	3	510
	S Tapanuli (S)	1.00 ± 0.50	**2.37 ± 2.46**	2	2	12	510
	Lingga island (near S)	0.67 ± 0.16	0.24 ± 0.19	10	5	6	510
	Bintan island (near S)	1.00 ± 0.50	0.20 ± 0.28	2	2	1	510
	Malaya pooled	0.95 ± 0.01	0.85 ± 0.47	81	34	47	510
	SE Johor (M)	0.82 ± 0.08	0.49 ± 0.31	21	12	13	510
	SE Johor & Endau-Rompin (M)	0.86 ± 0.06	0.46 ± 0.29	26	13	14	510
	Endau-Rompin (M)	0.70 ± 0.22	0.31 ± 0.26	5	3	4	510
	Bauk (M)	0.60 ± 0.13	0.24 ± 0.20	6	2	2	510
	Kuantan (M)	0.80 ± 0.08	**1.27 ± 0.71**	15	6	17	510
	Cameron (M)	0.94 ± 0.07	0.45 ± 0.31	9	7	6	510
	Fraser's (M)	1.00 ± 0.18	**1.97 ± 1.37**	4	4	17	510
	Pasoh (M)	—	—	5	1	—	510
	Southern Thailand (M)	—	—	1	1	—	510
	Kuala Tembeling (M)	—	—	2	1	—	510
	Singapore Island (near M)	0.53 ± 0.17	0.11 ± 0.12	6	2	1	510
	Tioman Island (near M)	0.79 ± 0.15	0.16 ± 0.14	8	5	2	510
	Pangkor Island (near M)	0.73 ± 0.16	0.90 ± 0.60	6	3	9	510
	Penang Island (near M)	—	—	7	1	—	510

Table 4 Population expansion parameters and test statistics. Significant *P*-values are shown in bold text. Note that the critical level equivalent to 0.05 for Fu's *F* is 0.02

Ant		Tajima's D		Fu's *F*		Mismatch Distribution					Expansion
Lineage	*n*	D	*P*	F	*P*	SSD	*P* (SSD)	τ	θ_0	θ_1	Age in Ka (95% CI)
G5	34	−1.64	0.042	−10.91	0.000	0.0127	0.554	4.20	0.006	5.741	511 (162–1010)
J	12	−1.45	0.070	−3.18	0.029	0.0254	0.235	4.89	0.001	24.004	598 (254–938)
H	110	−1.24	0.106	−19.25	0.001	0.0015	0.547	10.91	0.000	61.768	1355 (890–1666)
K	128	−2.13	0.006	−25.60	0.000	0.0030	0.532	1.48	2.598	1985.625	194 (48–641)

The effective population size parameter Θ (= 2*N*μ for haploid data; *N* = effective female population size; μ = mutation rate per nucleotide site per generation) and exponential growth rate (g), estimated by Lamarc for each clade, are presented in Table 5. Clades with sample sizes under nine were excluded from this analysis. The Lamarc analyses were roughly in agreement with tests for demographic expansion — the largest values of Θ were seen in the three clades, G5, H and K, that showed significant demographic expansions, and all the four clades tested were in the upper half of the exponential growth rate values. Clades A and F had large negative values of g, indicating population

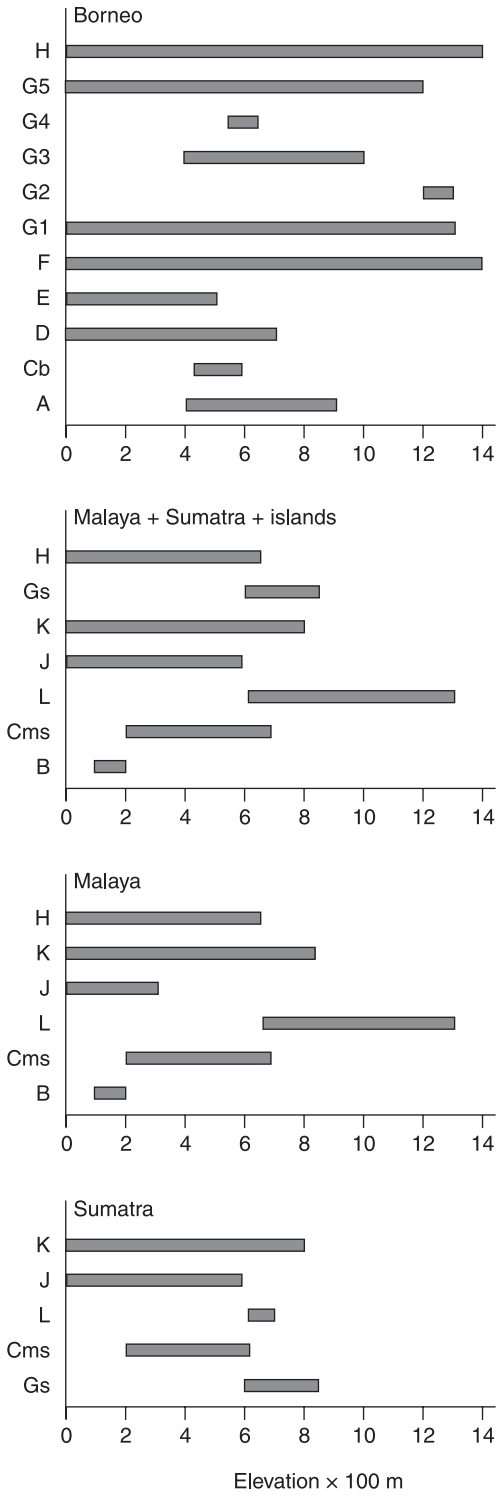

Fig. 4 Approximate elevational ranges of *Decacrema* lineages (details in Table S1, Supplementary material). The graphs reveal that only Bornean lineages show elevational ranges spanning the entire range of myrmecophytic *Macaranga* (0 to 1400 m).

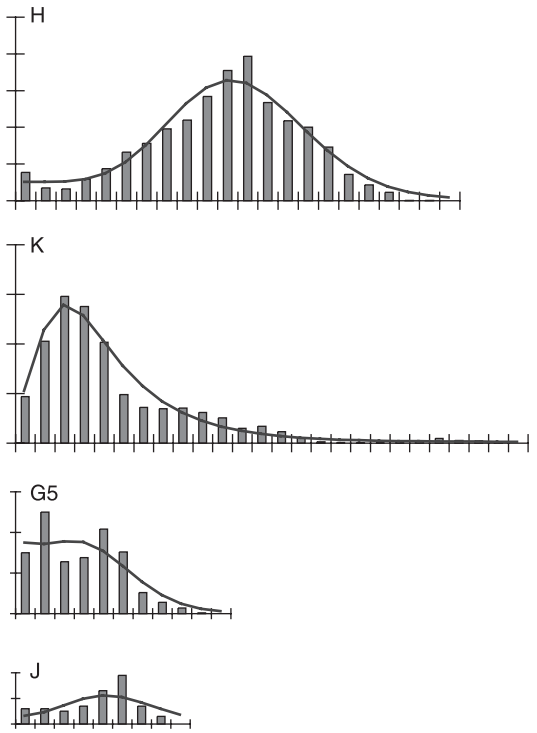

Fig. 5 Mismatch distributions of the *Decacrema* lineages H, K, J and G5 examined for demographic expansion. Bars indicate observed mismatch distribution and lines represent the mismatch distribution expected under a model of sudden expansion.

Table 5 Estimates of effective population size parameter (Θ) and exponential growth rate (g)

Clade	Θ	95% CI, Θ	g	95% CI, g
A	0.016	0.0059–0.0511	–25.7669	–163.298–72.549
Cms	0.029	0.0116–0.0843	17.4306	–104.942–125.776
D	0.062	0.0333–0.1231	61.8967	–40.157–168.648
L	0.027	0.0091–0.1015	22.8746	–141.829–182.317
F	0.012	0.0054–0.0287	–19.1408	–234.267–158.148
G1	0.015	0.0045–0.0642	189.5148	–233.650–917.234
G5	0.086	0.0170–1.0300	1247.756	415.686–3060.960
H	0.103	0.0729–0.1657	117.0698	25.964–223.240
J	0.069	0.0134–5.0840	719.7938	78.431–2916.01
K	0.083	0.0594–0.1233	130.0749	40.872–249.250

decline. However, it should be noted that of all the clades tested, only the four clades inferred to be significant or almost significant for Fu's *F*-test (J, H, K, G5) had phylogenies with no internal subdivisions (indicative of unhindered gene flow, an assumption underlying the Lamarc analyses).

Inferred ages of nodes using nonparametric rate smoothing are shown in Table 6 and Fig. 6. The tree in Fig. 6, based on COI alone, does not resolve the relationship between the three lineages Cb + Cms, D and L + E, in contrast to the tree based on COI plus COII (Fig. 1b). This does not affect

Node	% divergence	Age (Myr, Node 3 fixed)	Age (Myr, Node 5 fixed)	Age (Myr, Node 6 fixed)
3 (CDEL–FGHJK)	10.419	**6.95**	8.52	9.87
5 (G–HJK)	7.539	4.10	**5.03**	5.82
6 (K–JH)*	4.935	2.32	2.84	**3.29**
10 (Cb–Cms)*	6.899	4.10	5.03	5.82
9 (E–L)*	5.287	3.47	4.26	4.94
7 (H–J)*	4.513	1.87	2.29	2.66
8 (Gs–G2)*	5.060	2.49	3.06	3.54
2 (B–sister lineage)*	13.067	10.69	13.11	15.18
4 (F–GHJK)	7.297	4.54	5.57	6.45
1 (A–sister lineage)	15.860	15.94	19.56	22.65

Table 6 Percent COI divergences and ages of nodes on the *Decacrema* phylogeny obtained by nonparametric rate smoothing, with various nodes fixed and an assumed 1.5% uncorrected divergence per million years between sister lineages (fixed ages indicated in bold). An asterisk (*) indicates Bornean vs. non-Bornean sister lineages. The phylogeny for this analysis was inferred using 1324 bp of COI for 21 ingroup exemplars

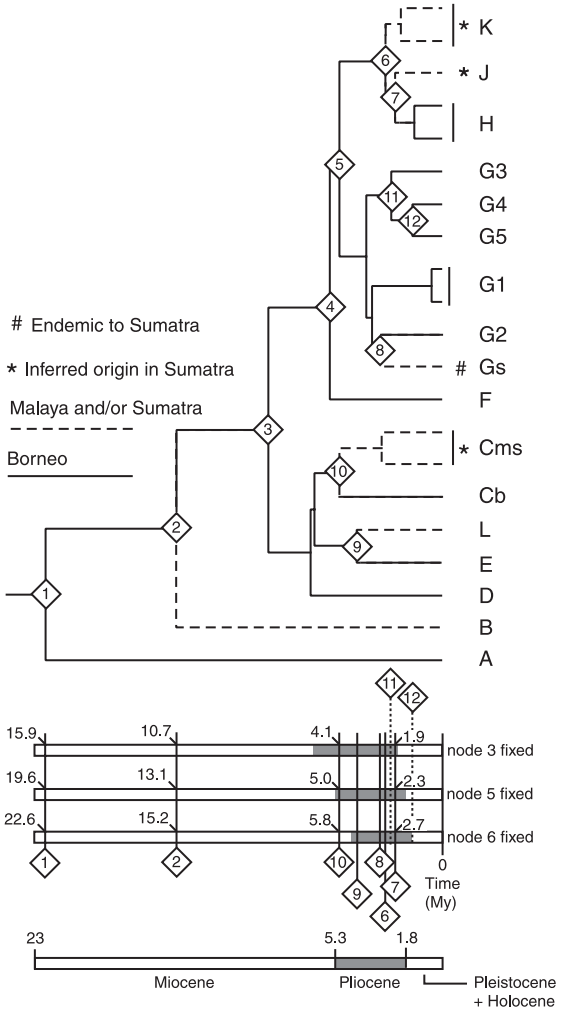

Fig. 6 Phylogeny of *Decacrema* inhabiting *Macaranga* with branch lengths proportional to time as inferred by nonparametric rate smoothing (1324 bp COI, 21 ingroup exemplars). Three timelines are inferred by fixing the ages of three nodes (3, 5, 6) using 1.5% divergence per million years in COI. Data from Table 5. Nodes mapped on the timelines: Node 1 represents the minimum age of the entire species complex and Nodes 2, 6, 7, 8, 9 and 10 represent inter-island splits in the phylogeny (Bornean vs. non-Bornean).

our results since our conclusions are not based on divergence dates between these lineages. The minimum age of the *Decacrema* inhabitants of *Macaranga* was estimated at 15.9 to 22.6 Myr, placing their first diversification in the early to middle Miocene. Perhumid climate, a prerequisite for the *Decacrema–Macaranga* association, became well established in the region from about 20 Ma onwards (Morley 2000), thus the ants are unlikely to predate this time. The time frame estimated from fixing node 6, giving 22.6 Myr as the minimum age for the ants (Fig. 6), is therefore less probable than the other two time frames. All lineages were already in place before the Pleistocene, with the bulk of cladogenetic events occurring in the Pliocene. Consequently, the divergences of pairs of sister lineages on either side of the South-China Sea (Borneo vs. Malaya + Sumatra) also occurred in the Pliocene (1.8–5.3 Myr).

The minimum spanning network for Clade H (not shown), a predominantly Bornean lineage, contains two well-supported emigrant clades found in Malaya and Lingga, respectively (see Fig. 1a, shaded ovals). The Malayan haplotypes were closest to two haplotypes from the Crocker Range, both of which differed from the closest Malayan sequence by nine changes (out of 537; or 1.68%). The minimum distance between Lingga and Bornean sequences, occurring in Kuching and Siduk in the west, was four changes (0.74%). The maximum distance among the Malayan haplotypes was three changes, while that of Lingga was one change (Table 3). If we assume a rate of 1.5% per Myr, the Malayan branch of Clade H could have been there for at least 350 000 years, but not more than 1.1 Myr, and Lingga would have been colonized after 0.5 Ma.

Discussion

The phylogeographical study presented here is the most comprehensive to date among such studies focusing on tropical Asia. We have drawn from analyses at macro-evolutionary and population levels to investigate the history of a species complex of rain forest ants. We emphasize

that this is a study of genes in space and time and not necessarily a study of speciation history, although it is likely that many of the lineages, in particular L and A through F, will turn out to be good approximations for species. Additionally, some of the patterns we observed in the data are likely to have been shaped by past extinctions, the extents of which remain elusive. Our inferences were drawn from the data taken at face value, and could be affected by artefacts such as patchy sampling and unequal sampling effort among locations. Correcting for these is not feasible since it is not possible to isolate the confounding factor(s) in most cases. Sumatra is poorly sampled overall, and sampling in Malaya is characterized by many locations with smaller sample sizes per location, as compared to Borneo which has better sampling per location but fewer locations. The distributional ranges in Fig. 3 are therefore approximations, and increased sampling could expand the distributions of some clades, especially within Sumatra and Borneo, which were sparsely sampled spatially. Further sampling is also likely to reveal more haplotypes, possibly resulting in changes in the distribution of genetic diversity. However, even though Sumatra was much more poorly sampled than Malaya, Sumatran locations (e.g. Tiga Puluh Mountains) often harboured greater genetic diversity within a given clade than better sampled locations in Malaya. Thus, the gross patterns of genetic diversity observed are likely to be robust.

The time frames inferred herein are also only approximations. While a general conservatism has been observed in COI rates across distant arthropod groups (Gaunt & Miles 2002; Quek *et al*. 2004), variations around this rate are to be expected. However, our proposed dates are not likely to be excessively speculative as they are consistent with (i) the age of Sundaland rain forests (Morley 2000); (ii) a mammalian molecular clock applied to rain forest rodents there (Gorog *et al*. 2004); and (iii) a reconstruction of Miocene and Pliocene land distribution in Sundaland (Hall 2001), which is discussed below).

A biogeographical scenario

Where did the association between *Macaranga* and *Decacrema* originate and subsequently diversify? Both parties attain their greatest diversity in Borneo, particularly for *Decacrema*, in the northernmost reaches of the island (the Malaysian state of Sabah), implying an origin and subsequent diversification there. The following three observations also support this idea: (i) the sister taxon to the Sundaland *Decacrema* occurs in Sulawesi, to the east of Borneo, where they inhabit stem domatia of *Neonauclea* (Quek *et al*. 2004); (ii) *Neonauclea*-inhabiting *Decacrema* are also known in the southern Philippines (Ridsdale 1989; Bolton 1995), to the northeast of Borneo; and (iii) the earliest branching lineage (Clade A) was restricted to the eastern part of Borneo (Fig. 3). The ancestral area reconstruction shows a Bornean backbone in the topology, indicating that the major diversification occurred there. Based on a chloroplast genealogy of myrmecophytic *Macaranga* and their allies, Bänfer *et al*. (2006) also infer a Bornean origin for section *Pachystemon*, the clade containing the vast majority of myrmecophytes. However, given that extinctions and range changes must have accompanied climatic and land area fluctuations of the Pleistocene (and Pliocene, to a smaller extent) in Malesia, the precise area of origin and major diversification cannot be unequivocally established.

We can, however, sketch the history of each island's *Decacrema* assemblage. All of Borneo's clades are likely to have originated *in situ*, or in areas close to the present-day Borneo. Of Sumatra's five lineages, all but L likely have Sumatran origins, with three (Cms, J and K) subsequently colonizing Malaya. All but K are sister to Bornean lineages, sharing last common ancestors in the Pliocene (the ancestral region for the node at which K, the fifth lineage, diverged is equivocal). Of Malaya's six lineages, one represents a Pleistocene migration from Borneo (H), three probably colonized the peninsula from Sumatra (K, J, Cms), and one (L) is sister to a Bornean lineage, also sharing a last common ancestor in the Pliocene. The presence of the relictual lineage, B, on Malaya's east coast suggests massive extinction or (less likely) suppressed diversification in its history.

At the end of the Miocene, four lineages were present (assuming no extinctions) but by the end of the Pliocene, almost all of the present lineage diversity (all 17 recognized lineages, excepting either G4 or G5) was already in existence (Fig. 6). Thus, in contrast to the Pleistocene species pumps postulated for Amazonia (Haffer 1987), most of the present lineage diversity (14 of the 17 lineages) can be traced to cladogenetic events in the Pliocene (A and B have Miocene origins and F is at the borderline, depending on which timeline is used in Fig. 6).

From the presumed Bornean stem, the divergence of the four originally Sumatran lineages (Gs, K, L, Cms) occurred during the Pliocene, while the divergence of the two originally Malayan lineages (B and L) occurred earlier on the whole, during the Miocene and earlier Pliocene (Fig. 6). An earlier Malayan connection to Borneo is consistent with land area reconstructions showing a Miocene Sundaland comprising predominantly much of Borneo, Malaya and the intervening sea bed, while Sumatra during that time consisted mostly of a chain of volcanic islands to the west of this land mass (Hall 2001). Only later towards the Pliocene did a substantial land bridge to this island chain (the emerging Sumatra) begin to form.

The three pairs of sister lineages (L–E, Cms–Cb, and G2–Gs) that represent divergences between nNW Borneo and mountain ranges in Malaya and Sumatra appear to have constricted ranges in Borneo and constricted or patchy distributions in Malaya and Sumatra (Fig. 3). That these

sister pairs are so widely dispersed, yet geographically restricted as individual lineages, indicate dramatic range reductions in the past, and their present distributions thus surely represent refugia. Biotic connections between NW Borneo and Malaya + Sumatra have also been observed in floristic studies of western Malesia (van Steenis 1964). Indeed, palynologist Muller (1972) contended that 'In NW Borneo, these [pollen] have provided evidence of uniformity of climate through the Tertiary period, of the existence of past cordilleras and land connections allowing floristic immigration pathways no longer extant, …' Our study also suggests that these immigration pathways predate the Pleistocene. These *Decacrema* distributions look suspiciously like the Riau Pocket of Corner (1940; see also Ashton 1992), characterized by floristic affinities between NW Borneo, the Riau archipelago and parts of Malaya and Sumatra, and their antiquity may have some bearing for the age of this enigmatic floristic distribution.

The Bornean vs. non-Bornean distributions seen in *Decacrema* clades have also been observed in other rain-forest animals such as rodents (Gorog *et al*. 2004) and frogs (Inger & Voris 2001) and indicate the biotic isolation of Borneo relative to other islands of the Sunda shelf. There is also growing evidence that the repeated and extensive exposures of the Pleistocene Sunda shelf, which experienced generally lower temperatures and drier conditions than the present (Morley 2000), rarely facilitated the spread of aseasonal rain-forest elements throughout the region. The *trans*-Sunda distributions of rain-forest rodents and some frogs are attributed to Pliocene rather than Pleistocene migrations (Inger & Voris 2001; Gorog *et al*. 2004). Within Sundaland *Decacrema*, all six nodes that straddle the South-China Sea predate the Pleistocene and five are of Pliocene age (Fig. 6). However, the two instances of dispersals from Borneo to Malaya and Lingga (within Clade H) within the past one Myr indicate the occurrence of successful Pleistocene migrations, possibly via forested riparian corridors (ideal habitat for *Macaranga* myrmecophytes) that may have dissected the drier vegetation of that time — indeed, extensive fossilized riverbeds are known from the South-China-Sea floor (Heaney 1991; Voris 2000).

The Pleistocene witnessed demographic expansions in a few *Decacrema* lineages. Clades H and K (as well as J, which was marginally significant for Fu's *F*-test) are predominantly specialists on *bancana* Clade hosts (Table 1, Fig. 2) which currently comprises 13 species. This might indicate the time frame for the proliferation of species and/or abundance of these hosts. If the expansion in Clade H reflects speciation activity in *bancana* Clade hosts, then Pleistocene origins may be likely for some *Macaranga* species, and thus also for some of the rain-forest biodiversity as suggested by Ashton (1972). The expansion of Malaya's and Sumatra's dominant clade, the primarily lowland-distributed K, occurred in the late Pleistocene and possibly reflects a later expansion of lowland rain forests in the region relative to Borneo (Fig. 5). It is worth noting, however, that although myrmecophytic *Macaranga* are strictly associated with everwet climate, and hence with evergreen forest, most are also restricted to gaps or edges where light and nutrients are abundant (Davies *et al*. 1998). Therefore, population expansions in *Decacrema*, presumed to reflect proliferation of their hosts, may not necessarily coincide with the maximal distribution of perhumid forests but may instead reflect transitory periods between rain-forest minima and maxima, or periods of disturbance, during which the habitat of *Macaranga* myrmecophytes would have been maximized at the edges of expanding rain-forest fragments.

Historical land-area and climate reconstructions by C.H. Cannon (unpublished) suggest that the present configuration and warm, wet conditions of Sundaland are just one of at least a dozen other similar intrusions, driven by Milankovitch cycles, into an otherwise cooler, drier and more extensive Sundaland in the past one Myr. It is thus conceivable that some of the present-day patterns we are observing could be recreated anew, or with varying degrees of alteration with each iteration of these intrusions. The inferred demographic expansions and distributions of genetic diversity, for example, could be attributed to any one or several of these cycles, and not necessarily to the onset of the Holocene.

Altitude generalists in Borneo and specialists in Malaya

What explains the presence of altitude generalists in Borneo but not in Malaya? Analyses of seabed sediments by Liechti *et al*. (1960) near Lambir suggest orogenic activity, followed by rapid erosion of the cordillera towards the end of the Miocene. Since the early Pliocene, the Kinabalu mountain massive (connected to the Crocker Range) has been undergoing uplift and subsequent erosion, and the geomorphology of Borneo remains dynamic to the present (Liechti *et al*. 1960). In Malaya, the bulk of orogenic activity had ceased before the Paleocene, producing the main mountain range of west-central Malaya (Gobbett & Hutchison 1973). The continuing dynamism in the geomorphology of Borneo may have fostered biota that were able to adapt to rapidly changing elevations, whereas the older and relatively stable mountains of Malaya may have been conducive to altitude specialization.

The disparity in altitudinal ranges is also apparent in the flora of Borneo, Malaya and Sumatra (Ashton 2003). Many lowland rain-forest tree species in Borneo also occur in the lower montane forest types, extending above 1200 m, whereas in Malaya and Sumatra, lowland rain-forest elements are rarely found above 1200 m. However, the suggested explanation is ecological rather than historical and related to the replacement of soil types from the lowlands to the lower montane zones in Malaya and

Sumatra but not in Borneo (Ashton 2003). On the other hand, myrmecophytic *Macaranga* species in Borneo do not show the same elevational spread as their ants — the species found in the lowlands extend at most to 1000 m, and species occurring above 1200 m extend at most to 800 m (Davies 2001 and S.-P. Quek, personal observation). Their ants, however, are specific to host clades, rather than to individual host species in most cases. The reason for the differences in altitudinal ranges in the ants is thus unlikely to have been driven by habitat or soil specificity of their hosts, and may have been shaped by the differing geomorphological histories of Malaya and Borneo.

Refugia versus recently colonized areas

Decacrema clades inferred to have stable populations and long histories are relatively widespread in Borneo, whereas in Malaya + Sumatra, they are restricted to or associated with mountains. Clades inferred to have undergone bottlenecks or founding effects are in the lowlands in Malaya and Sumatra. Thus the rain forests of Borneo (at least in the lowlands) might have been less perturbed than that in Sumatra and Malaya, where much of the present lowland rain forests may represent a relatively young Pleistocene community.

A curious finding is that Lambir and Meratus, both locations with high lineage diversity (five lineages each), show low population genetic diversities for all their lineages ($\pi = 0.41$ in Lambir, $\pi = 0.19$ in Meratus; Table 3). These locations probably represent recently colonized areas in close proximity to a diverse *Decacrema* source, or to a refuge(s) where rain forests have persisted through the Pleistocene. Although a high correlation between sample size and π-value was found for the Lambir data ($R^2 = 0.55$), a scenario of recent colonization is consistent with studies showing that it was part of a shallow sea in the early Holocene (Liechti *et al*. 1960). The Meratus data indicates more definitively a recent colonization. It shows a negative correlation between sample size and π-value ($R^2 = -0.27$), attributed largely to Clades G5 and H, each having nine samples representing only two and one haplotype(s), respectively.

In contrast, the western Borneo locations (Kuching and Siduk) show low lineage diversity (two lineages each) but generally high population genetic diversity ($\pi = 1.45$ and 2.3 in Kuching; $\pi = 1.34$ in Siduk; Table 3). The low diversity in western Borneo is also reflected in its flora and has been attributed to its relative youth as a terrestrial habitat (Slik *et al*. 2003).

In Malaya and Sumatra, mountains were certainly important for the survival of rain forests through the Pleistocene, when the lowlands may have been invaded by drier vegetation during the north temperate glacial intervals. Bukit (= 'hill' in Malay) Beserah on the east coast of Malaya is a small refuge (or a tiny sample of a more widespread east-coast refuge), harbouring at least two extremely rare taxa that we now know of, Lineage B within *Decacrema* and its host, *Macaranga constricta*. The Tiga Puluh mountains in Sumatra, centre of genetic diversity for the younger Clades J and K, may represent younger refugia, and/or a centre of origin for these lineages.

The rain forests in parts of NW Borneo, however, must be of great antiquity, certainly predating the Pleistocene. This region is regarded as the diversity hub for many plant taxa (Ashton 1972; Davis *et al*. 1995), including *Macaranga*, and is thought to have been a refuge also for rain-forest primates (Brandon-Jones 1996). Of all the sampling locations, the Crocker range of Sabah harbours the greatest number of lineages — of the seven found there, three (Cb, G2 and G4) appear to be endemic. The high intensity of sampling at this site is at best a minor explanation for this diversity and endemism — biotic/genetic isolation of locations in Sabah from the rest of Borneo has also been observed in several other rain forest taxa, including myrmecophytic *Macaranga* (Bänfer *et al*. 2006 and references therein). This high diversity is likely due in large to topographic relief — this region harbours the highest peak in Southeast Asia — Mount Kinabalu.

The World Wide Fund For Nature and The World Conservation Union (Davis *et al*. 1995) list numerous locations in the northern part of northwestern Borneo, the montane regions of Malaya, locations along the main mountain range of Sumatra (Bukit Barisan Range) and the Tiga Puluh Mountains as centres of plant diversity and endemism. In South America, the most diverse rain-forest communities and centres of endemism are thought to have experienced the longest continuity of wet climate, and have thus sustained rain forests for the longest period of time; likewise, a similar climatic history has been inferred for NW Borneo (Muller 1972). The persistence of rain forest communities on the mountains of Malaya and Sumatra may have been permitted by uninterrupted orographic rain through the Pleistocene, as is thought to be the case for neotropical rain forests (Haffer 1987). More importantly, these findings will be useful in forming priorities for the protection not only of species-rich regions, but also of areas that might harbour genetically diverse populations of endangered rain-forest species.

Despite their importance as one of the few centres of distribution of rain forests worldwide, the Asian tropics have been subject to relatively few studies on phylogeography and biogeography, in contrast to the numerous studies on the Neotropical biota. When more such studies are undertaken, a consilience of the region's rain forest history will hopefully come to light.

Acknowledgements

We are grateful to C. Cannon, J. Seger, C. Marshall, S. Russo, R. Eastwood, R. Mallarino and two anonymous reviewers for critical

comments, and to T. Rock and the late M. Knapp for assistance in the laboratory. This research was funded by National Science Foundation grants (DEB 9629601 to P. S. Ashton and DEB 0106866 to NEP & SPQ), and the following Harvard University grants: Putnam Expedition Grant (Museum of Comparative Zoology) and Department of Organismic and Evolutionary Biology Student Dissertation Grant. SJD thanks Sarawak Forest Department for permission to work in Sarawak, SPQ thanks Sabah Parks (Sabah, Malaysia) and National Parks Board (Singapore) for collecting permission. We thank S. K. Rambe for assistance in Indonesia.

Supplementary material

the supplementary material is available from
http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3294/MEC33294sm.htm

Table S1 List of samples. Ref. 1 refers to Feldhaar *et al*. 2003 (GenBank accession nos AF499935–AF499946, AF499948–AF499967, AF499970–AF599981 and AF499983–AF500002); ref. 2 refers to Itino *et al*. 2001. In the 'Location' column, B refers to Borneo, M to Malaya and S to Sumatra. In the 'Host' column, '*Neo*' refers to *Neonauclea* sp., and *Macaranga* host species are abbreviated as follows: *aët: aëtheadenia; ang: angulata; ban: bancana; bec: beccariana; con: constricta; cal: calcicola; gla: glandibracteolata; gri: griffithiana; hav: havilandii; hos: hosei; hul: hullettii; hyp: hypoleuca; ind: indistincta; kin: kingii; lam: lamellata; mot: motleyana; pea: pearsonii; pet: petanostyla; pru: pruinosa; pub: puberula; tra: trachyphylla; umb: umbrosa; vel: velutina*. Ogp: outgroup.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Ashton PS (1972) The Quaternary geomorphological history of Western Malesia and lowland forest phytogeography. In: *The Quaternary Era in Malesia. Transactions of the 2nd Aberdeen-Hull Symposium on Malesian Ecology* (eds Ashton PS, Ashton M), pp. 63–83. University of. Hull Department of Geography Miscellaneous Series no. 13, Hull.

Ashton PS (1992) Plant conservation in the Malaysian Region. In: *Harmony with Nature* (eds Yap SK, Lee SW), pp. 86–93. Malayan Nature Society, Kuala Lumpur.

Ashton PS (2003) Floristic zonation of tree communities on wet tropical mountains revisited. *Perceptions in Plant Ecology, Evolution and Systematics*, **6/1** (2), 87–104.

Bänfer G, Moog U, Fiala B, Mohamed M, Weising K, Blattner FR (2006) A chloroplast genealogy of myrmecophytic *Macaranga* species (Euphorbiaceae) in Southeast Asia reveals hybridization, vicariance and long-distance disersals. *Molecular Ecology*, **15**, 4409–4424.

Bolton B (1995) *A New General Catalogue of the Ants of the World*. Harvard University Press, Cambridge, Massachusetts.

Brandon-Jones D (1996) The Asian Colobinae (Mammalia: Cercopithecidae) as indicators of Quaternary climatic change. *Biological Journal of the Linnean Society*, **59**, 327–350.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Colinvaux PA, DeOliveira PE, Moreno JE, Miller MC, Bush MB (1996) A long pollen record from lowland Amazonia: forest and cooling in glacial times. *Science*, **274**, 85–88.

Corner EJH (1940) *Wayside Trees of Malaya*. Government Printer, Singapore.

Davies SJ (2001) Systematics of *Macaranga* sections *Pachystemon* and *Pruinosae* (Euphorbiaceae). *Harvard Papers in Botany*, **6**, 371–448.

Davies SJ, Palmiotto PA, Ashton PS, Lee HS, Lafrankie JV (1998) Comparative ecology of 11 sympatric species of *Macaranga* in Borneo: tree distribution in relation to horizontal and vertical resource heterogeneity. *Journal of Ecology*, **86**, 662–673.

Davies SJ, Lum SKY, Chan R, Wang LK (2001) Evolution of myrmecophytism in western Malesian *Macaranga* (Euphorbiaceae). *Evolution*, **55**, 1542–1559.

Davis SD, Heywood VH, Hamilton AC (1995) *Centres of Plant Diversity: a Guide and Strategy for Their Conservation*, Vol. 2. Asia, Australasia and The Pacific. The World Wide Fund for Nature (WWF) and IUCN — The World Conservation Union.

Degnan PH, Lazarus AB, Brock CD, Wernegreen JJ (2004) Host-symbiont stability and fast evolutionary rates in an ant–bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *Systematic Biology*, **53**, 95–110.

Federle W, Maschwitz U, Fiala B, Riederer M, Hölldobler B (1997) Slippery ant-plants and skilful climbers: selection and protection of specific ant partners by epicuticular wax blooms in *Macaranga* (Euphorbiaceae). *Oecologia*, **112**, 217–224.

Feldhaar H, Fiala B, Gadau J, Mohamed M, Maschwitz U (2003) Molecular phylogeny of *Crematogaster* subgenus *Decacrema* ants (Hymenoptera: Formicidae) and the colonization of *Macaranga* (Euphorbiaceae) trees. *Molecular Phylogenetics and Evolution*, **27**, 441–452.

Fiala B, Maschwitz U, Pong TY, Helbig AJ (1989) Studies of a Southeast Asian ant–plant association: protection of *Macaranga* trees by *Crematogaster borneensis*. *Oecologia*, **79**, 463–470.

Fiala B, Jakob A, Maschwitz U (1999) Diversity, evolutionary specialization and geographic distribution of a mutualistic ant-plant complex: *Macaranga* and *Crematogaster* in South East Asia. *Biological Journal of the Linnean Society*, **66**, 305–331.

Flenley JR (1984) Late Quaternary changes of vegetation and climate in the Malesian mountains. *Erdwissenschaftliche Forschung*, **18**, 261–267.

Fu Y-X (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Molecular Biology and Evolution*, **19**, 748–761.

Gobbett DJ, Hutchison CS (1973) Geology of the Malay Peninsula (West Malaysia and Singapore). In: *Regional Geology Series* (ed. De Sitter LU), p. 438. Wiley-Interscience, New York.

Gorog AJ, Sinaga MH, Engstrom MD (2004) Vicariance or dispersal? Historical biogeography of three Sunda shelf murine rodents (*Maxomys surifer*, *Leopoldamys sabanus* and *Maxomys whiteheadi*). *Biological Journal of the Linnean Society*, **81**, 91–109.

Grant WS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, **89**, 415–426.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Haffer J (1987) Biogeography and Quaternary history in tropical America. In: *Biogeography and Quaternary History in Tropical America* (eds Whitmore TC, Prance GT), pp. 1–18. Clarendon Press, Oxford.

Haile NS (1971) Quaternary shorelines in West Malaysia and adjacent parts of the Sunda shelf. *Quaternaria*, **15**, 333–343.

Hall R (2001) Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and Sea. In: *Faunal and Floral Migrations and Evolution in SE Asia-Australia* (eds Metcalfe I, Smith JMB, Morwood M, Davidson I), pp. 35–56. Swets and Zeitlinger Publishers, Lisse.

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Heaney LR (1991) A synopsis of climatic and vegetational change in Southeast Asia. *Climatic Change*, **19**, 53–61.

Huelsenbeck JP, Rannala B (1997) Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science*, **276**, 227–232.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

Inger RF, Voris HK (2001) The biogeographical relations of the frogs and snakes of Sundaland. *Journal of Biogeography*, **28**, 863–891.

Itino T, Davies SJ, Tada H *et al.* (2001) Cospeciation of ants and plants. *Ecological Research*, **16**, 787–793.

Janzen DH (1973) Dissolution of mutualism between Cecropia and its *Azteca* ants. *Biotropica*, **5**, 15–28.

Kuhner MK, Yamato J, Beerli P *et al.* (2005) *LAMARC, Version 2.0*. University of Washington, Washington.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA 2.1: Molecular Evolutionary Genetics Analysis software. *Bioinformatics*, **17**, 1244–1245.

Liechti P, Roe FW, Haile NS (1960) *The Geology of Sarawak, Brunei and the Western Part of North Borneo Bulletin 3*. Geological Survey Department, British Territories in Borneo, Kuching, Sarawak.

Longino JT (2003) The *Crematogaster* (Hymenoptera, Formicidae, Myrmicinae) of Costa Rica. *Zootaxa*, **151**, 1–150.

Maddison WP, Maddison DR (1992) *Macclade: Analysis of Phylogeny and Character Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Maley J (1996) The African rain forest — main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. In: *Essays on the Ecology of the Guinea — Congo Rain Forest*, vol. 104 (eds Alexander IJ, Swaine MD, Watling R), pp. 31–73. Royal Society of Edinburgh, Edinburgh.

Morley RJ (2000) *Origin and Evolution of Tropical Rain Forests*. John Wiley & Sons, Chichester.

Muller J (1972) Palynological evidence for change in geomorphology, climate and vegetation in the Mio-Pliocene of Malesia. In: *The Quaternary Era in Malesia. Transactions of the 2nd Aberdeen-Hull Symposium on Malesian Ecology* (eds Ashton PS, Ashton M), pp. 63–83. University of Hull Department of Geography Miscellaneous Series no. 13, Hull.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitutions. *Bioinformatics*, **14**, 817–818.

Quek S-P, Davies SJ, Itino T, Pierce NE (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*, **58**, 554–570.

Rambaut A, Charleston M (2002) *TreeEdit 1.0 Alpha 10: Phylogenetic Tree Editor*. Department of Zoology, University of Oxford, Oxford.

Ridsdale CE (1989) A revision of *Neonauclea* (Rubiaceae). *Blumea*, **34**, 177–275.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, **14**, 1218–1231.

Schneider S, Excoffier L (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates very among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin, Version 2000: a Software for Population Genetics Data Analysis Genetics and Biometry Laboratory*. University of Geneva, Geneva.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Slik JWF, Poulsen AD, Ashton PS *et al.* (2003) A floristic analysis of the lowland dipterocarp forests of Borneo. *Journal of Biogeography*, **30**, 1517–1531.

van Steenis CGGJ (1964) Plant geography of the mountain flora of Mt. Kinabalu. *Proceedings of the Royal Society of London, Series B*, **161**, 7–38.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony*. Ver. 4. 0b10. Sinauer Associates, Sunderland, Massachusetts.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

Tajima F (1989a) The effect of change in population size on DNA polymorphism. *Genetics*, **123**, 597–601.

Tajima F (1989b) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in Humans and Chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153–1167.

Whitmore TC (1998) *An Introduction to Tropical Rain Forests*. Oxford University Press, New York.

This study was part of Swee-Peck Quek's PhD dissertation, co-supervised by Naomi Pierce and Stuart Davies in the Department of Organismic and Evolutionary Biology at Harvard. Swee-Peck's interests are in the natural history and human history of Southeast Asia. Naomi Pierce is Hessel Professor of Biology at Harvard University. Her research focuses on behavioral ecology and the evolution of species interactions, particularly symbioses involving ants. Stuart Davies is currently Director of the Center for Tropical Forest Science, coordinating a global network of large-scale tropical forest research plots. Stuart investigates ecological and evolutionary influences on variation in rain forest communities across the tropics, and also specialises in plant taxonomy and ecosystems of Southeast Asia. Peter Ashton is Charles Bullard Professor of Forestry, emeritus, at Harvard, and recipient of the 2007 Japan Prize in the category of Science and Technology of Harmonious Co-Existence. His life-long interest is in the systematics and ecology of trees in Asia's tropics, and their forests. He is currently writing a synthesis of his experience: Essays on the forests of tropical Asia. Takao Itino is professor of biology at Shinshu University. He researches the ecology and evolution of species interactions, focusing on mutualistic, parasitic, and/or manipulative associations between ants and plants, sap-feeding insects, and their parasites.
