



Historical biogeography inference in Malesia

CAMPBELL O. WEBB AND RICHARD REE

8.1 Introduction

The Malesian region (the floristic region incorporating the nation states of Indonesia, Malaysia, Singapore, Brunei, the Philippines and Papua New Guinea) has long been recognised as a place of great biogeographic interest (Wallace 1869), and has stimulated a large literature (e.g. Croizat 1958, MacArthur and Wilson 1967, Michaux 1991). The ancestors of the Malesian biota have arrived by three major routes: (1) Laurasian (including Boreotropical) clades arriving from the west (e.g. for plants, *Trigonobalanus* and *Lithocarpus*, Morley 2000); (2) Gondwanan clades arriving from the west, via Africa or the Indian raft and Sundaland (e.g. Dipterocarpaceae; many Annonaceae, Richardson et al. 2004, Crypteroniaceae, Moyle 2004); and (3) Gondwanan clades arriving from the east, via the Australian raft (e.g. Proteaceae, Barker et al. 2007; Cunoniaceae; Monimiaceae; *Phyllocladus*, *Nothofagus*, Morley 2000; *Eucalyptus*, Ladiges et al. 2003). This biotic interchange has occurred on a complex, ever-changing landscape of continents and large and small islands, with barriers to species' movements not only of great over-ocean distances, but also of fluctuating climates, which sometimes caused large areas of unsuitable habitat. Given these multiple interacting factors, it is not surprising that we still poorly understand the detailed historical movements of most clades of organisms. Yet this very complexity makes Malesia a fascinating area to study, and tempts us with the promise that detecting shared patterns of movement and

interaction in this region will help us better understand the historical assembly of local biota in general. Typical biogeographic questions in the region include:

- When and how did taxa cross the narrowing divide between Asia and Melanesia?
- Why does Wallace's line clearly demarcate faunas but not floras?
- What was the nature and effect of interaction when ecologically similar taxa from West Malesia and from East Malesia encountered each other?
- Why have some lineages diversified only in the east or the west of Malesia despite having been present throughout the region for long periods?

Answering such questions will require the application of methods of historical biogeographic analysis across many clades, with a synthetic comparison of results (Sanmartín et al. 2008). Methods of biogeographic inference employed in the region include comparison of taxon lists from different areas ('Q-mode analysis', e.g. Simberloff and Connor 1979, van Balgooy, 1987, van Welzen et al. 2005, and 'R-mode analysis', e.g. Holloway, 1998), super-imposition of range maps (e.g. Baker et al. 1998), panbiogeography (Hedges, 2003) and studies of fossils (e.g. Truswell et al. 1984, Morley 2000). Phylogenies, both morphology-based (Turner 1996, Ridder-Numan 1998) and molecular (Scharaschkin and Doyle 2005, Brown et al. 2006b, Pfeil and Crisp 2008), have added hypotheses about the relationships among taxa, and more recently, estimates of the ages of ancestral taxa. Phylogenies have enabled the search for generalised, hierarchical patterns of connected areas (e.g. area cladograms inferred by Brooks Parsimony Analysis; Ruedi 1996, van Welzen et al. 2003), and the reconstruction of ancestral ranges by treating areas as pseudo-characters (e.g. Crisp et al. 2010). 'Event-based' analyses have added explicit consideration of the probability of the processes of dispersal, speciation and extinction. In particular, dispersal–vicariance analysis (DIVA; Ronquist 1997) is an accessible method and has been widely applied (e.g. Kreier and Schneider 2006, Jönsson et al. 2010). However, the Malesian region presents challenges to the assumptions of most existing analytical biogeography methods:

1. The terranes of the region are generally still approaching each other and accreting rather than fragmenting. Hence the area relationships themselves are better represented as a coalescing network rather than a branching tree (Fig 8.1). Approaches that attempt to reconcile trees (using cladistic or other methods) to find a common area tree may thus be imposing an incorrect underlying area hypothesis (Holloway 1998).
2. Because of the relatively small scale of the region, and possibly the existence of many 'island' species, with good dispersal ability, many taxa exhibit wide-spread (multi-terrane) distributions (e.g. van Welzen et al. 2005). This creates extensive polymorphism when treating areas as pseudo-characters, and most phylogenetic ancestral state reconstruction programs (the most widely

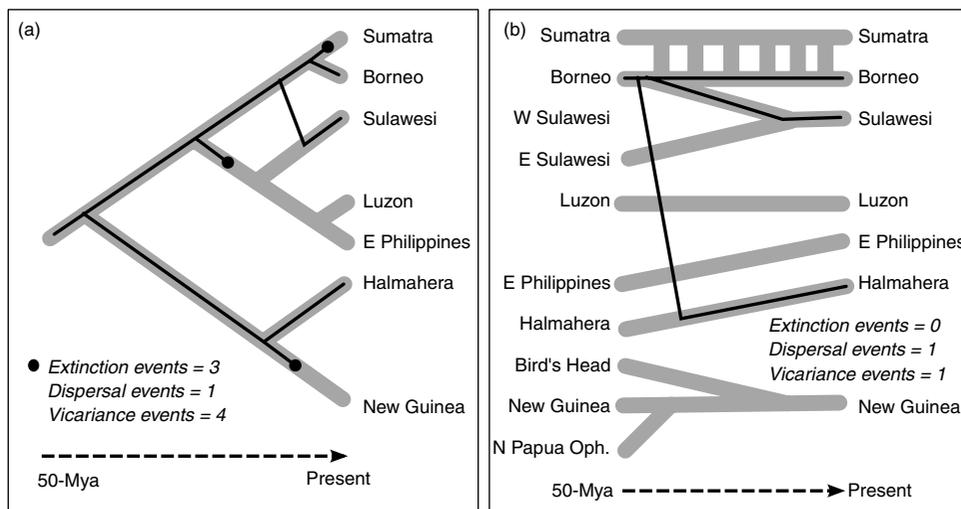


Figure 8.1 Schematic of hypotheses of land connectivity (grey) through time in Malesia, with associated organismal lineages. (a) A tree-like hypothesis of land vicariance that underlies tree-based analyses. (b) A connectivity-through-time network, based on terrane movement in Hall (2001). Clustering of the ends of the grey lines on the right side of (b) indicates spatial proximity; note similarity to Fig 9.26 in Morley (2000). The fine black lines represent the time–space trajectories of a currently widespread lineage (in Borneo, Sulawesi, Halmahera); solid circles indicate extinction events. For example, in (a), an initial vicariance event leaves descendants in the Halmahera–New Guinea area, and after the subsequent splitting of Halmahera and New Guinea, with associated lineage vicariance, the New Guinea lineage goes extinct. Note the extensive explanatory role for vicariance in (a), and its limited role in (b).

applied historical biogeography techniques; Sanmartín et al. 2008) are unable to accept polymorphisms (although this can be circumvented by recoding polymorphisms as combinations of independent present/absent characters, e.g. Hardy and Linder 2005).

3. The insular nature of Malesia means that dispersal, extinction and intra-island speciation will be the dominant processes structuring distributions (Keppel et al. 2009). Methods that assume vicariance to have a lower cost than dispersal (which include dispersal–vicariance analysis; Sanmartín et al. 2008, Kodandaramaiah, 2010) are likely to misinterpret the data.
4. Similarly, the changing distances between islands, and changing island areas, are not well represented by models with fixed costs (e.g. DIVA; Kodandaramaiah 2010).

Overall, reconstructing lineage movements in this complex area is unlikely to be successful without explicitly incorporating external information about temporal

changes in area connectivity into our models. Recently, several stochastic modelling approaches have emerged that do incorporate and even test hypotheses about contemporary and historical land conformations. Sanmartín et al. (2008) applied Bayesian Markov chain Monte Carlo (MCMC) transition matrix methods to models of dispersal between island groups in the Canary Islands. Ree and Smith (2008) applied a maximum likelihood inference method (Ree et al. 2005) to reconstruct ancestral ranges for *Psychotria* in Hawaii. These methods have great promise in the Malesian region, for which there are well-developed hypotheses about terrane history (e.g., Hall 2001, Hall this volume, Chapter 3).

In this chapter, we apply the maximum likelihood method of Ree and Smith (2008), implemented in the program LAGRANGE, to an example case of reconstructing the ancestral distributions of *Rhododendron* section *Vireya*. We also introduce a new method for ancestral area reconstruction based on simulation of ranges on a given phylogeny (SHIBA). We compare both methods with the results of dispersal–vicariance analysis (DIVA) and discuss the benefits and drawbacks of the three methods. Finally, we highlight the critical need for more data for solving some of the most fascinating problems of Malesian historical biogeography.

8.1.1 Incorporating landscape history into historical biogeography inference

Some properties of phylogenetic methods that should improve our ability to correctly infer historical biogeography in complex regions like Southeast Asia are:

1. Incorporating empirical constraints on dispersal implied by the spatial configuration of areas, guided by the general principle that the dispersal rate between two areas should be inversely proportional to the distance or strength of the barrier between them, e.g. open water, mountain range or other inhospitable tract (e.g. MacArthur and Wilson 1967). This becomes increasingly important as the number of areas becomes large, because more areas means more complexity in spatial relationships (Ree and Sanmartín 2009).
2. Restricting inferences to ancestral ranges that are biologically plausible, for example those that form connected networks of possible gene flow. The rationale is that a species is unlikely to remain cohesive if its range is fragmented or disjunct, with sub-ranges that are disrupted by intervening areas in which the species is absent (Ree and Smith 2008). Ranges should therefore be ‘contiguous’, for example so that each area of a species’ range should be connected to at least one other by a possible migration route. This spatial configuration is important when enumerating scenarios of range subdivision and inheritance at speciation events (see below), especially in the face of increasing geographic complexity as the number of areas becomes large. Species with disjunct ranges do occur in Malesia (van Welzen et al. 2005), but are rare.

3. Incorporating temporal information relevant to probabilities of range evolution through time. For example, the dynamics of area movement and accretion modelled by geologists could be linked to estimates of clade ages and lineage divergence times obtained from molecular clock analysis. To the extent that clades can be independently dated to absolute time, it is desirable to use this information in order to improve inferences of geographic ancestry (Ree et al. 2005), while acknowledging that geological models are also hypotheses, with varying degrees of error.
4. Along similar lines, including temporal information relevant to constraints on ancestral ranges. Examples of this include fossils constituting positive evidence that a lineage was present in a certain area at a particular time (Ree et al. 2005), or geological evidence of areas being uninhabitable (or underwater) at particular times (Ree and Smith 2008). The objective is to reject, or down-weight, historical inferences that are inconsistent with the evidence, for example, an ancestral range that includes an area before it exists, or does not include an area at a time when fossils place the lineage there.
5. Incorporating our understanding of how island area influences lineage survival. The positive relationship between island size and local species survival, part of the Theory of Island Biogeography (MacArthur and Wilson 1967), is a well-supported cornerstone of contemporary biogeography, and by applying it to historical hypotheses of lineage spread we are likely to greatly increase the accuracy of our reconstructions.

Thus, methods should ideally incorporate external hypotheses about land connectivity and size through time. At a minimum, they should include contemporary probabilities of dispersal events between different areas. Models can then be developed that allow the evolution of ranges at ancestral nodes of a branching phylogeny. Such models require an explicit statement of the spatial conformation of speciation events ('how do daughter species inherit spatial ranges?'). Finally an analytical engine can then optimise the fit of the model to the input data: phylogeny, current distribution, historical fossil locations, changing spatial conformations and even variation in the physical area of different spatial units. Uncertainty in phylogenetic, geological or event/process hypotheses can be assessed either through sensitivity analysis (i.e. re-running analyses with different combinations of parameters) or through incorporating informative prior probabilities in a Bayesian framework (e.g. Sanmartín et al. 2008). In the following examples we illustrate how desirable properties 1 and 2, above, can be addressed using LAGRANGE, and how 1 and 3–5 can be addressed using SHIBA.

LAGRANGE

This biogeographic inference method was inspired by perceived limitations of DIVA largely relating to its reliance on parsimony (Ree et al. 2005, Ree and Smith

2008). LAGRANGE retains a common conceptual framework with DIVA, in which geographic areas are discrete, species' ranges are coded as binary presence-absence data, and ranges evolve along phylogenetic branches by lineage movement (dispersal from one area to another, causing range expansion) and local extinction (extirpation within an area, causing range contraction). The primary objective also remains the same, namely inference of biogeographic history by fitting a model of range evolution, including geographic scenarios of speciation, to an observed phylogenetic tree with species ranges arrayed at its tips. With DIVA, this is achieved by finding the most parsimonious set of ancestral geographic ranges that minimises the number of dispersal and extinction events required to explain the data on the tree. With LAGRANGE, it involves optimising the likelihood of stochastic dispersal and local extinction along branches, and scenarios of range inheritance at nodes where lineages diverge. The primary differences of LAGRANGE from DIVA are: (1) probabilities of dispersal and local extinction are functions of time, and calculated analytically from a rate matrix; (2) geographic modes of speciation include a wider range of scenarios, including persistence of widespread ancestral ranges through cladogenesis events; and (3) flexible constraints reflecting spatial relationships, habitability, etc. of areas may be imposed on geographic ranges and rate parameters.

SHIBA

Another approach to estimating ancestral ranges is to simulate lineage movement in a discrete spatial and temporal model. During each time slice a species' range can expand via dispersal or contract via extinction, in an externally specified landscape. Where a phylogenetic branching event occurs within a time slice, the daughter taxa can be distributed to represent the products of either sympatric or allopatric speciation. This approach mirrors the 'experimental biogeography' (Posadas et al. 2006) of Colwell and Winkler (1984) and Haydon et al. (1994), but rather than generating simulated phylogenies, it takes a phylogeny as input and determines the probabilities of ancestral ranges and lineage movement. A necessary condition is that estimates for the ages of branching events in the phylogeny exist so that they can be anchored in the corresponding dynamic land-area model. To capture additional aspects of biogeographic reality, the extinction probability in an area can be modelled as a function of the physical area of that unit, and the dispersal probability a negative function of the distance to be crossed (Fig 8.2). This incorporates the fundamental components of the Island Biogeography Theory of MacArthur and Wilson (1967), and has led us to dub the software application that implements these methods SHIBA: 'simulated historical island biogeography analysis'. The software is open source, and available for download at <http://phylodiversity.net/shiba/>; it is a command-line C program, and can be compiled on any operating system. The input data are: (1) A distance matrix for spatial

distance between x areas, for each of t time slices (i.e. of size $x \times x \times t$). A distance of zero between two areas implies they are directly connected. (2) A matrix of physical areas for each unit area, of size $x \times t$. (3) A list of n taxa, and a chronogram; for example a hypothesis for the phylogenetic relationships among the n taxa (with L edges), with node ages set to represent actual times of divergence. This chronogram is 'sliced' into t slices, in each of which a particular edge either branches (speciates into two or more new taxa) or continues as the same taxon. (4) An $L \times x$ presence/absence matrix of contemporary distribution for n taxa, and historical distribution of any fossils on the L edges. See Appendix B for an example of input data. The range evolution algorithm for each run is: (1) Choose a starting area or areas for the stem of the phylogeny, either at random or pre-determined. (2) Determine whether the lineage disperses to each of the other areas during the time slice, based on the distance between the existing location and the new one (Fig 8.2). A vagility parameter is 'tuned' during the simulation to match the total number of occurrences on islands with that observed (e.g. too low a dispersal rate will cause many islands to end up unoccupied). (3) Determine if the population in the original area goes extinct, based on the physical area of the site (Fig 8.2). While there has been some discussion about the actual form of the extinction rate versus area curve (Gilpin and Diamond 1976), we chose a model of $survival = \log area$ as the simplest to parameterise.

The speciation process model is similar to that employed in LAGRANGE: if a lineage is present in only one area, the speciation is sympatric (i.e. 'duplication'). If

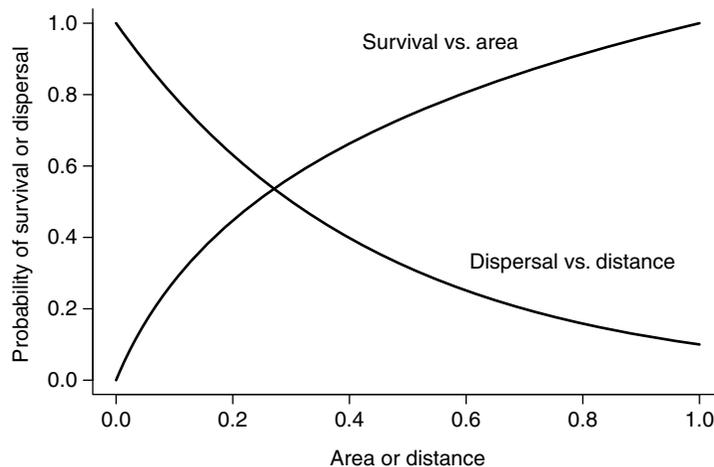


Figure 8.2 General form of the survival and dispersal curves in SHIBA: $survival = S \times \log area$ and $dispersal = e^{-1 \times D \times distance}$, where S and D are constants (here set at 1.0). Areas and distances are expressed as proportions of maximum unit area and maximum inter-terrene distance.

the parent lineage is present in two or more areas, one is chosen randomly as the location of one daughter lineage, while populations in the other areas become the other daughter species (i.e. allopatric speciation). This is also sometimes described as a daughter species branching off of a continuing parent species.

The simulation is re-run many times. A run is terminated prematurely if (1) any lineage dies out in all areas before reaching the present, or (2) if a particular lineage does not pass through a specific fossil location in the specified time slice. If all lineages reach the present, the distribution of taxa is compared to the observed distribution. Only if the simulated distribution is the same as the observed distribution is a run kept as a 'success', and the ranges of its ancestral taxa recorded. After a pre-specified number of successes (r), the simulation terminates and histograms of locations of ancestral taxon locations are reported. These histograms can be described thus: 'A lineage L_p at time t_q occurred h_0 times out of r successful runs in area x_0 , h_1 times in area x_1 , h_3 times in area x_3 , etc.' In this way, the probabilities of ancestral ranges and movements that can lead to the observed distribution are determined, given the model structure and parameters. Indeed, these simulations provide a probabilistic means to reject hypotheses; for example, if less than 925 runs placed the root taxon in an area A , we could with confidence say that the model does not support the historical scenario that the origin of the clade was in A .

8.2 An example: *Rhododendron* section *Vireya* in Malesia

8.2.1 The taxa

To demonstrate the application of these biogeographic methods in the Malesian region, we searched published studies for a clade of plants having species in both West and East Malesia, and a well-sampled molecular phylogeny with well-documented range data. The most suitable we could find was a study of *Vireya* rhododendrons by Brown et al. (2006b), which had detailed range maps, but included only 65 out of an estimated 300 species. *Vireya* is a primarily Malesian clade, with species easily recognised by their leaf and twig scales, and its species are much prized by gardeners. Brown et al. (2006b) mapped the 63 species (plus two outgroup species) onto 20 contemporary areas of endemism and constructed a molecular phylogeny from chloroplast markers (*trnT-trnL* and *psbA-trnH* intergenic spacer regions). They found that *Vireya* was divided into two major clades: one comprising species of Western and Central Malesia and one comprising species of New Guinea and further east. Using paralogy-free subtree analysis (Nelson and Ladiges 1996), Brown et al. explored the cladistic connections between areas, based on the *Vireya* phylogeny, and equivocally concluded that either (1) the clade is old, with *Vireya* having rafted north from Gondwanaland on the Australian

continent, entering Malesia from the east, while sect. *Rhododendron* arrived from the west via India, or (2) the clade is much younger and only arrived from the west, via India, with the other *Rhododendron* clades (cf. similar hypothesis of Irving and Hebda 1993). Hence the outstanding questions in this example of *Vireya* in Malesia are: (1) How old is the clade? and (2) Did the clade diversify from west to east or from east to west?

8.2.2 The landscape

Increasingly detailed models of the complex tectonic history of the region (Hall 1998, 2001, 2009, this volume, Chapter 3) contribute much to our biogeographic understanding. Western Indonesia, on the Sunda shelf, has been part of the Eurasian mainland, and at equatorial latitudes, at least since the mid-Cretaceous (Morley 2000). The Australian plate broke away from Antarctica at *c.* 60 Ma and has approached the Asian plates from the South, only taking up its current conformation at *c.* 10 Ma. The islands of Wallacea are terranes that have variously been split from Sundaland (West Sulawesi), rafted in from the West Pacific (East Sulawesi, East Halmahera), arrived as outliers of the Australian plate (Seram, Buton), been extruded (West Halmahera, Sumbawa), or been uplifted (many small islands; Hall 2001). New Guinea is a complex assembly of Australian plate, recently accreted West Pacific terranes and a young, active centre.

For the SHIBA analysis presented here, which incorporates changing landscape conformations over time, we used the maps of Hall (2001, Figs 2–5), which present a hypothesis for terrane position and extent at 55, 45, 35, 25, 15 and 5 Ma. We digitised each image and computed (1) terrane area and (2) geographical distance among each pair of terranes, for each time period. While many separate terranes might be recognised from Hall's maps (the SHIBA software distribution includes data for 12), in this example we modelled four areas: Sundaland, Western Sulawesi, Luzon (North Philippines) and Central New Guinea, which are the current day locations of the taxa used in the SHIBA analysis. There is no theoretical limit to including more terranes, which might have acted as historical stepping-stones, but increasing the number of areas decreases the rate of finding solutions and slows the analyses. Values for distance and area were interpolated linearly for one time slice between each age of the landscape in the seven maps to give 12 time periods. Because we do not know the above-water land area for the terranes in the past, we made the assumption that terrane area is an appropriate substitute for land area.

8.2.3 LAGRANGE analysis

Rationale and methods

Our objective in revisiting *Vireya* with LAGRANGE was to explore the utility of its dispersal–extinction–cladogenesis (DEC) model in making historical biogeographic inferences in a region of substantial spatial complexity. A major challenge posed by

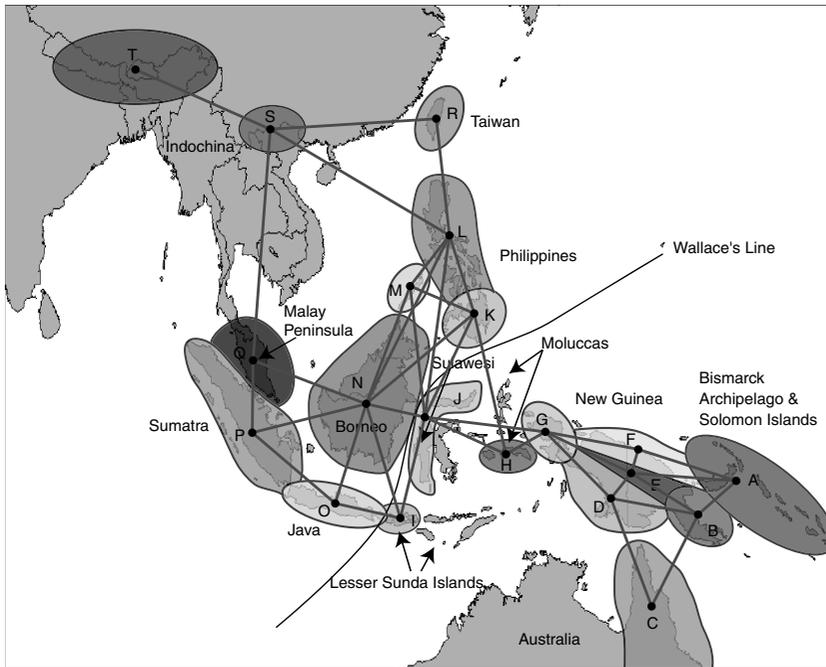


Figure 8.3 Areas of endemism for *Vireya* defined by Brown et al. (2006b), showing adjacency connections used to constrain dispersal routes in reanalysis using LAGRANGE. See plate section for colour version.

the $n = 20$ areas of endemism defined by Brown et al. (2006b) stems from the number of geographic ranges theoretically possible, $2^n = 1\,048\,576$, if all subsets of areas are considered valid ranges. Computing probabilities of ancestor-descendant range evolution for a phylogenetic branch requires integration of a $2^n \times 2^n$ rate matrix, meaning that for *Vireya*, the size of this matrix makes computing this integral impractical using standard software tools. Fortunately, however, this is neither necessary nor desirable if we use the spatial structure of areas to simplify the model.

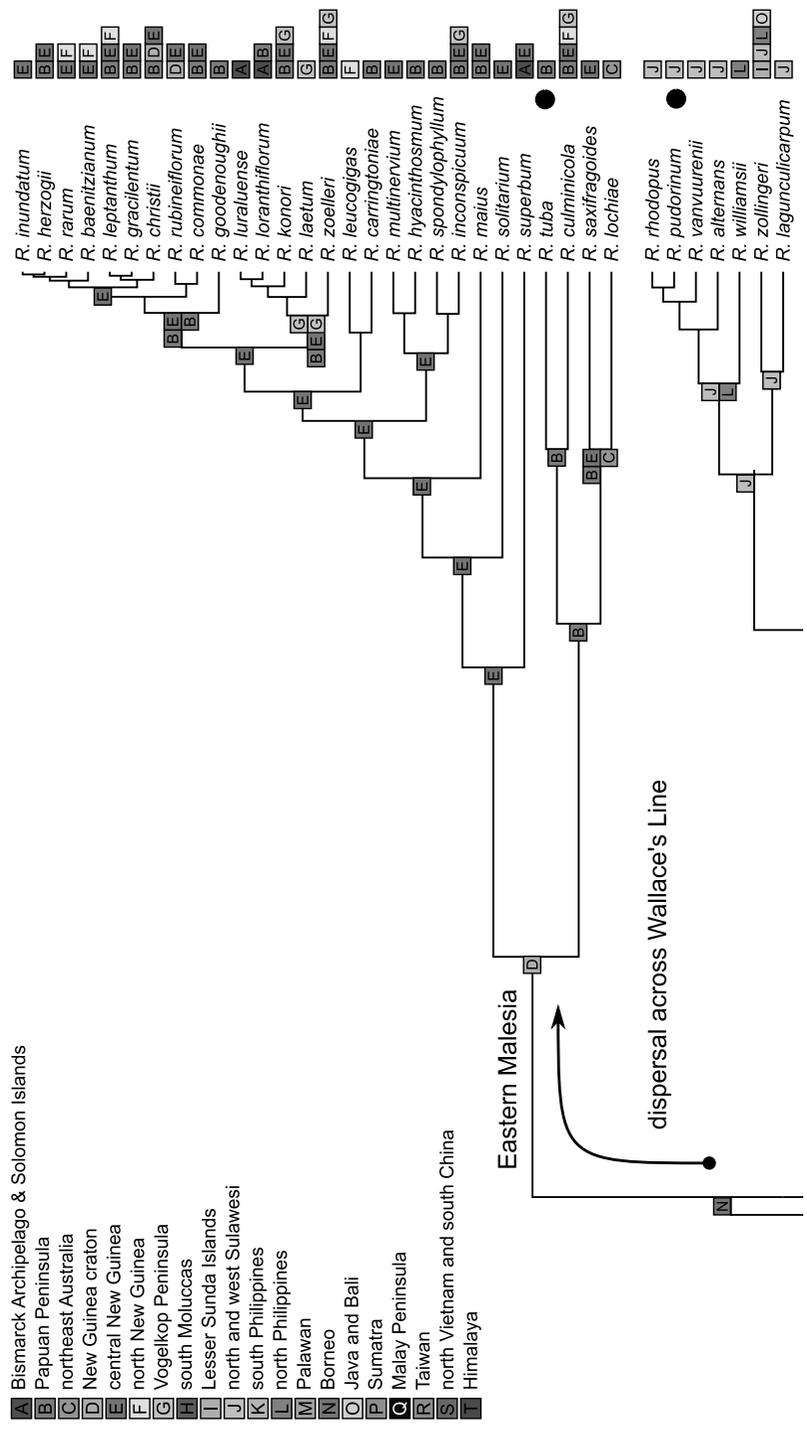
Reasoning that dispersal and gene flow are most likely between spatially proximate areas, we created an ‘adjacency graph’ of nearest-neighbour relationships for the 20 areas of endemism based on their present-day positions (Fig 8.3). This graph allowed us to identify subsets of areas forming disconnected components, which we excluded from consideration on the grounds that they represented disjoint and therefore biologically unrealistic geographic ranges of species. In addition, we excluded ranges greater in extent than observed in extant species, setting the maximum number of areas to three (see below). This reduced the dimensions of the rate matrix considerably, to a manageable $n = 405$. We also used the adjacency graph to constrain dispersal, such that its rate parameter d was constant and symmetrical between connected areas and zero elsewhere.

Some caveats of this analysis must be noted. One is that we assumed static area positions, meaning that the constraints described above were constant across all branches of the *Vireya* phylogeny. This is, of course, a suboptimal approach. Ideally, we would have incorporated knowledge of terrane dynamics during the Cenozoic (Hall 2001, this volume, Chapter 3) by constructing a time series of adjacency graphs reflecting change in dispersal opportunities (as with the SHIBA analysis). Other caveats relate to the geographic range data from Brown et al. (2006b). In general we took these at face value, without considering issues such as incomplete sampling of extant species. However, in this particular case, we also reduced the size of the largest ranges, i.e. those exceeding three areas, in order for LAGRANGE to achieve convergence when optimizing dispersal and extinction rates. This meant modifying the given ranges of *Rhododendron malayanum*, *R. javanicum*, *R. zollingeri*, *R. culminicola* and *R. zoelleri*, by reducing each to a single representative area shared by its closest relative. The necessity of this step is perhaps clarified by considering the size of the phylogeny (63 leaf nodes) relative to the size of the transition matrix (405 elements square). We suspect that a larger phylogeny would be required for convergence in parameter estimation with a larger transition matrix. We acknowledge that these caveats reduce the empirical value of this analysis, and therefore emphasise it here as more of an illustration of how adjacency graphs can be used in spatially complex regions.

We inferred the phylogeny of the 65 *Vireya* species treated in Brown et al. (2006b) using sequences of *psbA-trnH* and *trnT-trnL* from Brown et al. (2006a). We aligned the sequences using MUSCLE (Edgar 2004) and estimated a maximum-likelihood tree given the GTRCAT nucleotide model implemented in RAXML 7.2.0 (Stamatakis 2006, Stamatakis and Ott 2008), with substitution parameters partitioned by marker. An identical topology and similar branch lengths were obtained from PHYML with the TVM models selected by Brown et al. (2006a) using AIC. We converted the tree to a relative-time chronogram by nonparametric rate smoothing (use of penalised likelihood was precluded by the presence of several terminal branches with effectively zero length), as implemented in APE (Paradis 2004). We then estimated maximum-likelihood values for rates of dispersal (d) and local extinction (e), using LAGRANGE with a DEC model constrained by the adjacency graph as described above. Finally, we used these parameter values to reconstruct maximum-likelihood ancestral ranges at each internal node on the tree.

Results

The phylogeny estimated using RAXML was congruent with the results of Brown et al. (2006a), in recovering a ‘Euvireya’ clade, within which was nested a pair of sister clades corresponding to those labelled ‘Eastern Malésia’ and ‘Western and Middle Malésia,’ respectively (Brown et al. 2006a, Fig 1). The relative-time chronogram of this tree, annotated with ancestral ranges inferred using LAGRANGE, is shown in Fig 8.4. At the root node of *Vireya*, a northern ancestral range was inferred



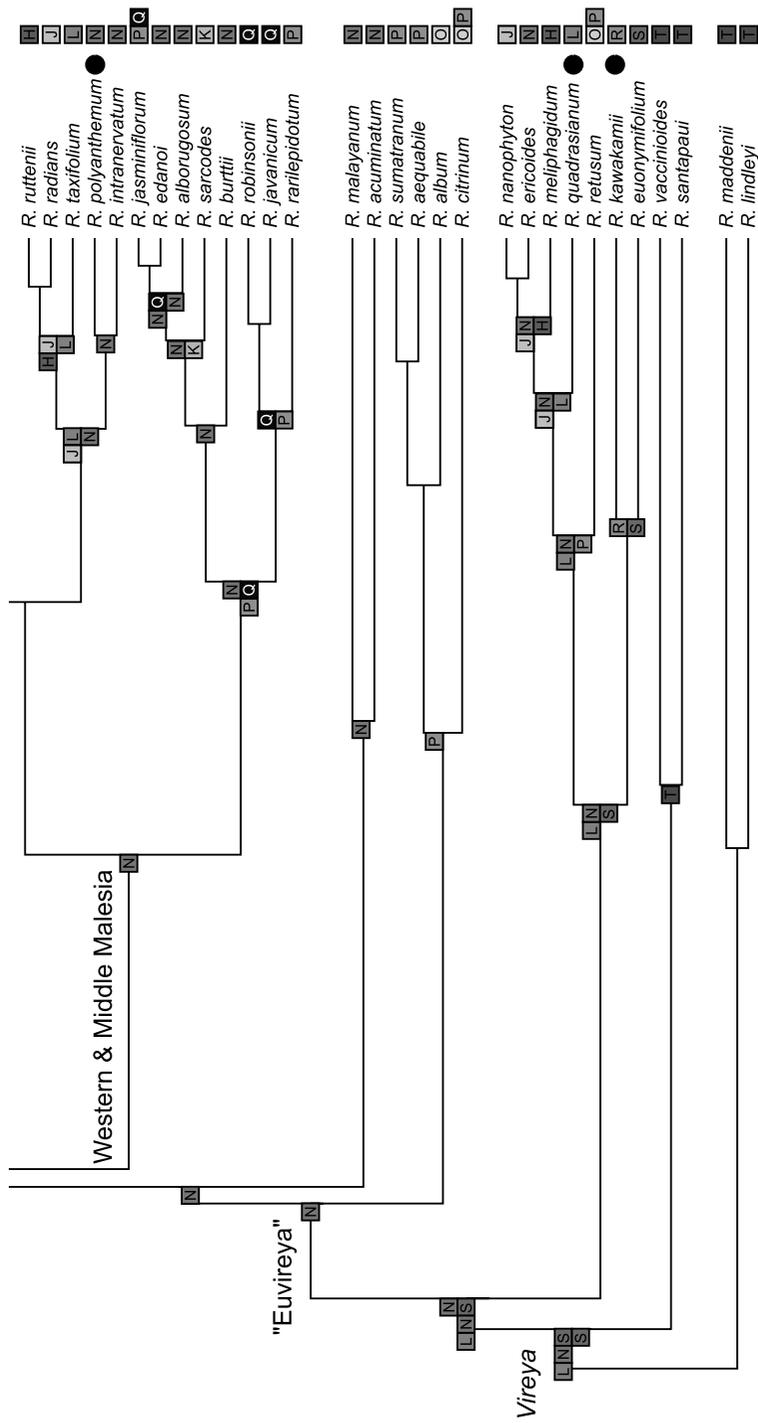


Figure 8.4 Results of LAGRANGE re-analysis of *Vireya* biogeography given 20 areas of endemism in southeast Asia, following Brown et al. (2006b). The relative-time chronogram was produced by applying nonparametric rate smoothing to a maximum likelihood (ML) molecular phylogeny. Nodes are annotated with ML estimates of ancestral geographic ranges. Rates of dispersal and local extinction were estimated subject to constraints implied by the adjacency graph shown in Figure 8.3 (see text for details). Colonisation of eastern Malasia is inferred as dispersal across Wallace's Line. Species used in the SHIBA runs are indicated with a dot. See plate section for colour version.

including southern China (S), Philippines (L), and Borneo (N). From this origin, lineage movement was generally to the south and east, with Borneo serving as a hub. Migration across Wallace's Line was reconstructed along the stem branch of the 'Eastern Malesia' clade.

8.2.4 SHIBA analysis

Simulation details

Because SHIBA searches through the complete possibility space defined by a set of areas, times and lineage divisions, the rate of finding successful runs (i.e. giving observed taxa distributions) falls rapidly with increasing taxa and areas. Hence we chose a subset of *Vireya* taxa and Malesian areas to represent the biogeographic behaviour of the whole clade over the full set of terranes. We chose *R. kawakamii* and *R. quadrasianum* from the basal 'Pseudovireya' clade, *R. polyanthemum* and *R. pudorinum* from the 'western' clade and *R. tuba* from the 'eastern' clade (Fig 8.4). We chose areas: (1) Sundaland and mainland Asia, (2) West Sulawesi, (3) Luzon (North Philippines), and (4) Central New Guinea (Fig 8.5).

The utility of the SHIBA approach lies in its capacity to reconstruct lineage movements over shifting historical areas (also theoretically possible with LAGRANGE) and the intuitive interpretation of its output. We ran two analyses, one setting the root of the clade to 55 Ma, the other to 11 Ma, thus bracketing the likely range of ages for *Vireya* diversification in the region. We divided the chronogram of the five taxa (pruned from the chronogram used in the LAGRANGE run) into 12 'slices' (of 5 Ma and 1 Ma respectively) for which a land area and terrane distance model was prepared (see above). When they are known, fossils can be included in the model to constrain lineages to pass through particular places at certain times, but we were not aware of any *Vireya* fossils to use in this case. Survival parameter *S* was set to 1.0. Dispersal parameter *D* was initially set to 0.2, but is subsequently and repeatedly adjusted by the program to give simulated occurrence values (proportion of total possible areas \times taxa occupied) that match the observed value (4 out of 20 = 0.2). A single area of the root taxon was chosen at random from the ancestral four areas. Random selections of multiple areas could also have been used as the root starting distribution, but in order to compare SHIBA results with other methods, and also to determine the most likely origin direction (east or west), we set the initial distribution to be a single area in these simulations. Simulations were run until 1000 successes were found.

Results

Both 'old root' and 'young root' simulations indicated that Sundaland and mainland Asia was most probably the ancestral area of root of *Vireya* (see bar chart at root node of Fig 8.5). In only 307 and 310 out of 1000 runs (in the 55 Ma and 11 Ma root-age analyses, respectively) was the ancestor of *R. tuba* already in New Guinea

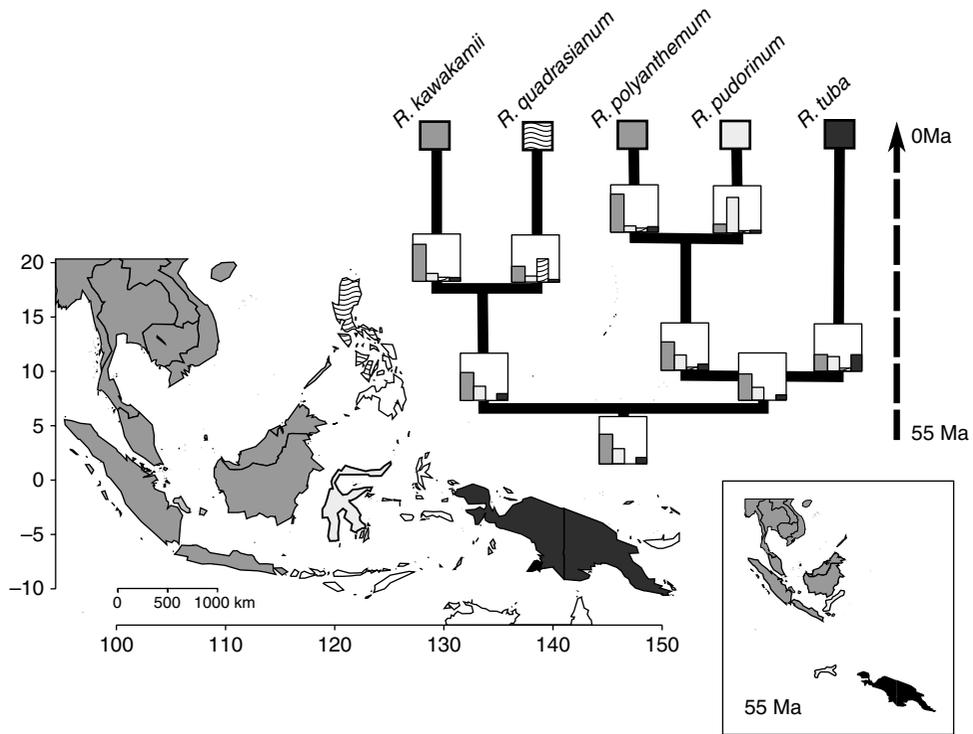


Figure 8.5 Results of SHIBA analysis of range evolution of five representative *Rhododendron* sect. *Vireya* species in Malesia, using four regions: mainland Southeast Asia and Sundaland (mid-grey), western Sulawesi (light grey), Northern Philippines (wavy lines), and mainland New Guinea (dark grey). Boxes on the branches of the phylogeny (pruned from Fig 8.4) indicate the proportion of runs (out of 1000) that the basal populations of a lineage were in a particular area. As an example, the direct ancestor of *R. quadrasianum* was present in the Philippines in 478/1000 runs, and in Asia/Sundaland in 297/1000 runs. Inset figure shows approximate locations of areas at 55 Ma; the Luzon terrane was probably still submerged (Hall 2001).

at the time of its separation from the western lineage, indicating that it most likely dispersed into New Guinea soon after the origin of the eastern clade, followed by radiation within New Guinea (leading to *c.* 170 *Vireya* species). A different speciation scenario is indicated in the case of *R. pudorinum* and *R. polyanthemum*: the majority of runs show the ancestor of *R. pudorinum* to already be in Sulawesi and the ancestor of *R. polyanthemum* to be in Borneo (Sundaland) at the time of origination of these two lineages, indicating origination by allopatric speciation of their common ancestor.

One clear difference between the ‘old root’ and ‘young root’ simulations is the timing of arrival of the ancestors of *R. quadrasianum* in Luzon. Because the

terrane was probably not above water at 55 Ma (Hall 2001), the Luzon physical area in our model was set to zero at 55 Ma, preventing this area being chosen as a starting area for the *Vireya* root. In the 11 Ma simulation, Luzon was near its present position and dry land, and we found that the *Vireya* root taxon occurred in Luzon in as many as 139 out of 1000 successes.

The success rate (number of runs giving the observed distribution divided by the total number of runs) is an indication of the likelihood of the solution, given the parameters chosen and the structure of the model. The success rate for a 55 Ma root was 6.71×10^{-7} , while that for a 11 Ma root was 6.49×10^{-7} . It is tempting to conclude that the older root age is slightly more probable, but these values are close and a more comprehensive sensitivity analysis would be needed to assess the confidence of any such conclusion.

8.2.5 DIVA analysis

Because DIVA is the most widely used phylogenetic biogeography tool, we conducted DIVA analyses that can be compared with the LAGRANGE and SHIBA analyses. For the LAGRANGE comparison, we used the distribution of the 63 taxa in 15 areas; collapsing areas N-T in Brown et al. (2006b) to a single 'Sundaland and mainland Asia' (N) area, because DIVA will only accept a maximum of 15 areas. We used the maximum likelihood phylogeny re-generated from the original sequence data, as described above. When the number of ancestral areas was not constrained, DIVA reconstructed a widespread *Vireya* root ancestor (most recent common ancestor of *R. santapau* to *R. inundatum*; Fig 8.4), occurring from Mainland Asia to New Guinea (BCEFGHJLN or ABCEFGHJLN); these widespread solutions always have the lowest cost in DIVA in the absence of other constraints (Kodandaramaiah 2010). When we constrained the distribution to occur only in a maximum of three areas, the root taxon occurred only in 'Sundaland and mainland Asia' (area N). In this constrained model we also found that the distribution of the ancestor of the 'eastern' clade was only in Central New Guinea (area E), in agreement with both the LAGRANGE and SHIBA analyses. For the SHIBA comparison, with four areas and five taxa in DIVA, the *Vireya* root ancestor was again widespread in the unconstrained DIVA case, and limited to the 'Sundaland and Mainland Asia' area when the analysis was constrained to only two ancestral areas.

8.3 Discussion

8.3.1 *Vireya* example

All analyses, using the three different methods, inferred the ancestor of the *Vireya* clade to have occupied Western Malesia, and that the dispersal into New Guinea

and Australia and subsequent radiation occurred relatively recently, thus supporting Hypothesis 2 ('Laurasian origin') of Brown et al. (2006b). Visual inspection of the phylogeny, showing the outgroup and the basally positioned taxa occurring in the west (Fig 8.4), plus the presence of fossil *Rhododendron* primarily in Laurasia (references in Brown et al. 2006b), intuitively suggest the same interpretation, so these results are not surprising. However, the extent to which this result is empirically robust depends on the ranges and phylogenetic positions of unsampled species. Because these represent a substantial proportion of the clade (only 63 out of *c.* 300 species were included), the ancestral ranges estimated here should be regarded as very preliminary in nature. If the unsampled taxa in a clade are missing at random, with respect to geographic area, we do not expect any systematic biases in our biogeographic reconstruction (i.e. the general issue of 'taxon sampling'; Ackerly 2000). It is, however, actually quite likely that there will be spatial structure in the unsampled taxa, caused by difficult access to parts of the total spatial range of the clade (e.g. poor sampling of Wallacea and West New Guinea).

Another question of interest relates to the age of *Vireya* and the eastern Malesian sub-clade. If the crown age of *Vireya* is old (e.g. Paleocene), and the relative branch lengths of our chronogram are correct (i.e. the eastern Malesian sub-clade is at least 50% of the age of the *Vireya* root), the most likely interpretation is a single long-distance dispersal event onto the New Guinea terranes at *c.* 35 Ma followed by a period of radiation in isolation. If the *Vireya* root age is young, dispersal into New Guinea would have involved shorter distances, and lineage diversification within New Guinea must then have occurred much faster. In both scenarios, it is notable that there has been no back-dispersal of New Guinea clade members into Western Malesia.

How might historical biogeographic analysis help resolve questions relating to clade age? Fossil-calibrated molecular clocks suggest that *Vireya* is relatively young: Milne (2004) estimated that subgenus *Rhododendron* is 46–32 Myr old, suggesting that the origin of *Vireya* is more recent. A more recent chronogram by Alex Twyford and James Richardson (pers. comm.), using the nuclear marker *RPB2*, and building on the work of Goetsch et al. (2005), dates the crown group age of *Vireya* at only *c.* 13 Ma. In the face of uncertainty about the statistical confidence and precision of these estimates, we suggest that biogeographic likelihoods are a potential source of relevant information. By this we mean comparing likelihoods of extant species ranges having evolved, given a history of area connections through time, and a set of alternative chronograms reflecting uncertainty in node ages. The rationale is that if the availability of dispersal routes has strongly influenced the movements of lineages over phylogenetic timescales, the maximum likelihood score of a LAGRANGE analysis and the success rate of a SHIBA run should be highest when the 'true' chronogram of a clade is overlaid on the land-connection history. For example, one could set the root age of a clade to be

a free parameter (in LAGRANGE), or an axis along which to conduct sensitivity analyses (in SHIBA). This approach would lead to estimates of the root age (and ancestral range) that relate directly to biogeographic likelihoods. In this study, the difference between the success rate of the 55 Ma and 11 Ma root-age SHIBA runs, though small, demonstrates how, of the two levels, the former provided a slightly more likely solution; however, in this case we are not inclined to take this as strong evidence in favour of a Paleocene age for *Vireya*.

This idea in effect turns traditional biogeographic analysis on its head, and places weight on area history as a line of evidence relevant to making inferences about node ages and ancestral ranges. Naturally, this assumes that knowledge of area history can be obtained with greater confidence and precision than, say, divergence times. It also assumes that the ‘true’ area history has had a significant effect on the proliferation and movements of species in the clades of interest. The validity of both assumptions will be region- and clade-dependent. For Malesia, the generally accepted description of terrane movement has been relatively stable for more than a decade, indicating an increasing confidence in this hypothesis of area history. Similarly, while we still do not understand the details, we can be sure that the insular nature of Malesia has had a strong influence on species proliferation and movement.

A further possibility in the interaction of geological models with biogeographic models is that the latter could even be used to test the likelihood of the former. If confidence in a chronogram is high, and especially if there are fossils that can anchor particular ancestral lineages to particular areas, both LAGRANGE and SHIBA could be run over a range of geological models and the likelihood (or success rate) for each geological model used to indicate its validity.

8.3.2 Methodological approaches and future directions

While DIVA is an accessible and widely used tool, the uses of LAGRANGE and SHIBA described here demonstrate some important advantages, especially for the Malesian region. For example, neither method assumes vicariance to be the null model for widespread range inheritance, and both can model range evolution according to spatial and temporal constraints. They each in turn have particular strengths and weaknesses. LAGRANGE is more computationally efficient than SHIBA, and for that reason can handle larger numbers of areas and taxa. Neither method accounts for the phylogeny itself being the stochastic outcome of (possibly range-dependent) speciation and extinction processes (Ree and Sanmartín 2009). As a result, LAGRANGE analyses tend to infer artificially wide ancestral ranges and low local extinction rates, to condition on observed data (Lamm and Redelings 2009). SHIBA is slow relative to LAGRANGE because it attempts to search across all possible histories of range evolution. For example, a simulation of 1000 successes took *c.* 2.5 hours on a 3 Ghz 8-core Intel Xeon CPU with 16.2 GB

RAM; increasing simulation efficiency is a high priority for further development of SHIBA. This means that applying SHIBA may require the user to subsample taxa in the clade of interest, while being mindful of the issue of spatial bias in sampled taxa, as discussed above. On the other hand, SHIBA currently offers a more comprehensive framework for incorporating dynamic histories of land connection. Additionally, only SHIBA can currently incorporate variation in area size and its effect on extinction rates. These differences make LAGRANGE a better choice for reconstructing the biogeographic history of large clades in Malesia over the fairly recent past (when the islands' conformation was not greatly different than the present one), and SHIBA a better choice for smaller clades over a longer history of land movement.

Both methods use a similar algorithm for determining how daughter species are distributed spatially during vicariant speciation in a widespread species, invoking allopatric speciation with single areas for one daughter and remainder areas for the other (see Ree et al. 2005). These algorithms, while defensible as parsimonious solutions, do represent particular choices and as such may misrepresent the biology of speciation for some groups of organisms, and are open to criticism (e.g. Lamm and Redelings 2009). As is always the case with modelling complex systems, we would urge all users to be aware of the assumptions and simplifications involved. We intend in future to carry out a thorough assessment of the ability of SHIBA to infer biogeographic history in simulated test cases for which the history of lineage movement is fully known.

Finally, obtaining meaningful inferences from phylogenetic comparative methods will always depend on obtaining adequate numbers of representative taxa. To answer the most fundamental questions about biogeographic dynamics in Malesia, perhaps the greatest need is for more intensive sampling of both taxa and areas in molecular phylogenetic studies (Crisp et al. 1995, 2004). Despite the long history of botanical research in the Malesian region (Webb et al. 2010), few plant clades have well-sampled molecular phylogenies; the general requirement of fresh material to successfully extract DNA from plant species places a premium on new collections, especially from undersampled and remote regions. Several animal clades have been relatively well sampled (e.g. Evans et al. 2003, Honda et al. 2006), but we remain far from being able to recognise general patterns. More complete phylogenetic sampling will also allow analyses of historical biogeography to address key questions about disparities in species richness across the region. One kind of inference in particular need of methodological attention relates to links between range dynamics (the timing and direction of movement) and variation in regional rates of speciation and extinction (Ree and Sanmartín, 2009). In the interim, we suggest that process-based models for reconstructing historical biogeography in Malesia, tailored to the unique geological history of the region, will enable the most effective use of the data at hand.

Acknowledgements

We are grateful to A. Antonelli, L. Rüber and three anonymous reviewers for valuable comments, and to M. Donoghue and B. Moore for early discussions on modelling approaches. Webb was partially supported by a US NSF grant (DEB-1020868) during preparation of this chapter.

References

- Ackerly, D. D. (2000). Taxon sampling, correlated evolution, and independent contrasts. *Evolution*, **54**, 1480–92.
- Baker, W. J., Coode, M. J. E., Dransfield, J. et al. (1998). Patterns of distribution of Malesian vascular plants. In *Biogeography and Geological Evolution of SE Asia*, ed. R. Hall and J. D. Holloway. Leiden, The Netherlands: Backhuys Publishers, pp. 243–58.
- Barker, N. P., Weston, P. H., Rutschmann, F. and Sauquet, H. (2007). Molecular dating of the ‘Gondwanan’ plant family Proteaceae is only partially congruent with the timing of the break-up of Gondwana. *Journal of Biogeography*, **34**, 2012–27.
- Brown, G., Craven, L., Udovicic, F. and Ladiges, P. (2006a). Phylogeny of *Rhododendron* section *Vireya* (Ericaceae) based on two non-coding regions of cpDNA. *Plant Systematics and Evolution*, **257**, 57–93.
- Brown, G. K., Nelson, G. and Ladiges, P. Y. (2006b). Historical biogeography of *Rhododendron* section *Vireya* and the Malesian Archipelago. *Journal of Biogeography*, **33**, 1929–44.
- Colwell, R. K. and Winkler, D. W. (1984). A null model for null models in biogeography. In *Ecological Communities: Conceptual Issues and the Evidence*, ed. D. R. Strong, D. Simberloff, L. G. Abele and A. B. Thistle. Princeton, NJ: Princeton University Press, pp. 344–59.
- Crisp, M. D., Linder, H. P. and Weston, P. H. (1995). Cladistic biogeography of plants in Australia and New Guinea: congruent pattern reveals two endemic tropical tracks. *Systematic Biology*, **44**, 457–73.
- Crisp, M. D., Cook, L. and Steane, D. (2004). Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **359**, 1551–71.
- Crisp, M. D., Isagi, Y., Kato, Y., Cook, L. G. and Bowman, D. M. J. S. (2010). *Livistona* palms in Australia: ancient relics or opportunistic immigrants? *Molecular Phylogenetics and Evolution*, **54**, 512–523.
- Croizat, L. (1958). *Panbiogeography*. Caracas: Published by the author.
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **5**, 113.
- Evans, B. J., Brown, R. M., McGuire, J. A. et al. (2003). Phylogenetics of fanged frogs: testing biogeographical hypotheses at the interface of the Asian and Australian faunal zones. *Systematic Biology*, **52**, 794–819.

- Gilpin, M. E. and Diamond, J. M. (1976). Calculation of immigration and extinction curves from the species-area-distance relation. *Proceedings of the National Academy of Sciences of the USA*, **73**, 4130–4.
- Goetsch, L., Eckert, A. J. and Hall, B. D. (2005). The molecular systematics of *Rhododendron* (Ericaceae): a phylogeny based upon RPB2 gene sequences. *Systematic Botany*, **30**, 616–26.
- Hall, R. (1998). The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. In *Biogeography and Geological Evolution of SE Asia*, ed. R. Hall, R. and J. D. Holloway. Leiden, The Netherlands: Backhuys Publishers, pp. 99–131.
- 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- Australasia*, ed. I. Metcalfe, J. M. B. Smith, M. Morwood and I. D. Davidson. Lisse, The Netherlands: A. A. Balkema (Swets and Zeitlinger Publishers), pp. 35–56.
- Hall, R. (2009). Southeast Asia's changing palaeogeography. *Blumea*, **54**, 148–61.
- Hardy, C. R. and Linder, H. P. (2005). Intraspecific variability and timing in ancestral ecology reconstruction: a test case from the Cape Flora. *Systematic Biology*, **54**, 299–316.
- Haydon, D., Radtkey, R. R. and Pianka, E. R. (1994). Experimental biogeography: interactions between stochastic, historical, and ecological processes in a model archipelago. In *Species Diversity in Ecological Communities: Historical and Geographical Perspectives I*, ed. R. E. Ricklefs and D. Schluter. Chicago: University of Chicago Press, pp. 117–30.
- Heads, M. (2003). Ericaceae in Malésia: vicariance biogeography, terrane tectonics and ecology. *Telopea*, **10**, 311–449.
- Holloway, J. D. (1998). Geological signal and dispersal noise in two contrasting insect groups in the Indo-Australian tropics: R-mode analysis of pattern in Lepidoptera and cicadas. In *Biogeography and Geological Evolution of SE Asia*, ed. R. Hall and J. D. Holloway. Leiden, The Netherlands: Backhuys Publishers, pp. 291–314.
- Honda, M., Ota, H., Murphy, R. W. and Hikida, T. (2006). Phylogeny and biogeography of water skinks of the genus *Tropidophorus* (Reptilia: Scincidae): a molecular approach. *Zoologica Scripta*, **35**, 85–95.
- Irving, E. and Hebda, R. (1993). Concerning the origin and distribution of rhododendrons. *Journal of the American Rhododendron Society*, **47**, 139–62.
- Jønsson, K. A., Bowie, R. C. K., Moyle, R. G. et al. (2010). Historical biogeography of an Indo-Pacific passerine bird family (Pachycephalidae): different colonization patterns in the Indonesian and Melanesian archipelagos. *Journal of Biogeography*, **37**, 245–57.
- Keppel, G., Lowe, A. J. and Possingham, H. P. (2009). Changing perspectives on the biogeography of the tropical South Pacific: influences of dispersal, vicariance and extinction. *Journal of Biogeography*, **36**, 1035–1054.
- Kodandaramaiah, U. (2010). Use of dispersal-vicariance analysis in biogeography: a critique. *Journal of Biogeography*, **37**, 3–11.
- Kreier, H.-P. and Schneider, H. (2006). Phylogeny and biogeography of the staghorn fern genus *Platyserium* (Polypodiaceae, Polypodiidae).

- American Journal of Botany*, **93**, 217–25.
- Ladiges, P. Y., Udovicic, F. and Nelson, G. (2003). Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography*, **30**, 989–98.
- Lamm, K. S. and Redelings, B. D. (2009). Reconstructing ancestral ranges in historical biogeography: properties and prospects. *Journal of Systematics and Evolution*, **47**, 369–82.
- MacArthur, R. H., and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.
- Michaux, B. (1991). Distributional patterns and tectonic development in Indonesia: Wallace reinterpreted. *Australian Systematic Botany*, **4**, 25–36.
- Milne, R. I. (2004). Phylogeny and biogeography of *Rhododendron* subsection *Pontica*, a group with a tertiary relict distribution. *Molecular Phylogenetics and Evolution*, **33**, 389–401.
- Morley, R. J. (2000). *Origin and Evolution of Tropical Rainforests*. New York: Wiley.
- Moyle, R. G. (2004). Calibration of molecular clocks and the biogeographic history of Crypteroniaceae. *Evolution*, **58**, 1871–3.
- Nelson, G. and Ladiges, P. Y. (1996). Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates*, **3167**, 1–58.
- Paradis, E. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–90.
- Pfeil, B. E. and Crisp, M. D. (2008). The age and biogeography of citrus and the orange subfamily (Rutaceae: Aurantioideae) in Australasia and New Caledonia. *American Journal of Botany*, **95**, 1621–31.
- Posadas, P., Crisci, J. V. and Katinas, L. (2006). Historical biogeography: a review of its basic concepts and critical issues. *Journal of Arid Environments*, **66**, 389–403.
- Ree, R. H. and Sanmartín, I. (2009). Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography*, **36**, 1211–20.
- Ree, R. H. and Smith, S. A. (2008). Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.
- Ree, R. H., Moore, B. R., Webb, C. O. and Donoghue, M. J. (2005). A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, **59**, 2299–311.
- Richardson, J. E., Chatrou, L. W., Mols, J. B., Erkens, R. H. J. and Pirie, M. D. (2004). Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, **359**, 1495–508.
- Ridder-Numan, J. W. A. (1998). Historical biogeography of *Spatholobus* (Leguminosae- Papilionoideae) and allies in SE Asia. In *Biogeography and Geological Evolution of SE Asia*, ed. R. Hall and J. D. Holloway. Leiden, The Netherlands: Backhuys Publishers, pp. 259–77.
- Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.
- Ruedi, M. (1996). Phylogenetic evolution and biogeography of southeast Asian shrews (genus *Crocidura*: Soricidae). *Biological Journal of the Linnean Society*, **58**, 197– 219.

- Sanmartín, I., van der Mark, P. and Ronquist, F. (2008). Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography*, **35**, 428–49.
- Scharaschkin, T. and Doyle, J. A. (2005). Phylogeny and historical biogeography of *Anaxagorea* (Annonaceae) using morphology and non-coding chloroplast sequence data. *Systematic Botany*, **30**, 712–35.
- Simberloff, D. and Connor, E. F. (1979). Q-mode and R-mode analyses of biogeographic distributions: null hypotheses based on random colonization. In *Contemporary Quantitative Ecology and Related Ecometrics*, ed. G. P. Patil and M. L. Rosenzweig. Fairland, MD: International Co-operative Publishing House, pp. 123–38.
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–90.
- Stamatakis, A. and Ott, M. (2008). Efficient computation of the phylogenetic likelihood function on multi-gene alignments and multi-core architectures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3977–84.
- Truswell, E. M., Kershaw, A. P. and Sluiter, I. R. (1984). The Australian-South-east Asian connection: evidence from the paleobotanical record. In *Biogeographical Evolution of the Malay Archipelago*, ed. T. C. Whitmore. Oxford: Clarendon Press, pp. 32–49.
- Turner, H. (1996). Sapindaceae and the biogeography of Eastern Australia. *Australian Systematic Botany*, **8**, 133–67.
- van Balgooy, M. M. J. (1987). A plant geographical analysis of Sulawesi. In *Biogeographical Evolution of the Malay Archipelago*, ed. T. C. Whitmore. Oxford: Clarendon Press, pp. 94–102.
- van Welzen, P. C., Turner, H. and Hovenkamp, P. (2003). Historical biogeography of Southeast Asia and the West Pacific, or the generality of unrooted area networks as historical biogeographic hypotheses. *Journal of Biogeography*, **30**, 181–92.
- van Welzen, P. C., Slik, J. W. F. and Alahuhta, J. (2005). Plant distribution patterns and plate tectonics in Malésia. *Biologiske skrifter*, **55**, 199–217.
- Wallace, A. R. (1869). *The Malay Archipelago*. London: Macmillan and Co.
- Webb, C. O., Slik, J. W. F. and Triono, T. (2010). Biodiversity inventory and informatics in Southeast Asia. *Biodiversity and Conservation*, **19**, 955–972.

Appendix A

Input data for Lagrange

The following indicates the data taken as input for the LAGRANGE analysis. Full data can be found via the Lagrange page at <http://www.reelab.net>.

PHYLOGENY

```
((Rhododendron_santapau:0.4837224249,Rhododendron_vaccinioides:0.4837224249):0.481376557,(((Rhododendron_euonymifolium:0.2488067527,Rhododendron_kawakamii:0.2488067527):0.2524890665,(Rhododendron_retusum:0.2629986131,(Rhododendron_quadrasianum:0.1370131223,(Rhododendron_meliphagidum:0.07023346485, ... (63 taxa total)
```

DISTRIBUTION

```
63 20
Rhododendron_santapau 00000000000000000001
Rhododendron_vaccinioides 00000000000000000001
Rhododendron_euonymifolium 00000000000000000010
Rhododendron_kawakamii 000000000000000000100
Rhododendron_retusum 0000000000000110000
Rhododendron_quadrasianum 00000000000100000000
... (63 rows total)
```

AREA NAMES

```
A B C D E F G H I J K L M N O P Q R S T
```

RATE MATRIX

```
1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 1 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
1 1 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
1 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
... (20 rows total)
```

INCLUDED DISTRIBUTIONS

```
10000000000000000000,11000000000000000000,11100000000000000000,
11010000000000000000,11001000000000000000,11000100000000000000,
10011000000000000000,10001000000000000000, ... (159 items total)
```

Appendix B

Input data for SHIBA

The following indicates the data taken as input for a set of SHIBA runs, though is not the input file itself. A '#' indicates a comment that is not involved in the computation. The data are: no. time slices, no. areas, matrix of area \times time, distances matrices for each time, no. taxa, phylogeny with branch lengths, extant distribution and optional fossil constraints (e.g. fossil indicates that edge n_2 passed through area s_2), various parameters. A few other operational parameters (number of runs, stopping rules, etc.) are not shown.

TIME: 12 # t1: 60-55; t2: -50; t3: -45; t4: -40; t5: -35; t6: -30;
 # t7: -25; t8: -20; t9: -15; t10: -10; t11: -5; t12: 5-0 Ma

SPACE: 4 # s1: Borneo; s2: W. Sulawesi; s3: C. New Guinea; s4: Luzon

LAND AREA (arbitrary units):

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12
	==	==	==	==	==	==	==	==	==	==	==	==
s1 :	10	10	10	10	10	10	10	10	10	10	10	10
s2 :	3	3	3	3	3	3	3	3	3	3	3	3
s3 :	12	12	12	12	12	12	12	12	12	12	12	12
s4 :	0	0	3	3	3	3	3	3	3	3	3	3

DISTANCE BETWEEN LAND UNITS:

TIME = t1:					TIME = t2:				
	s1	s2	s3	s4		s1	s2	s3	s4
	====	====	====	====		====	====	====	====
s1 :	0	773	2690	na	s1 :	0	623	2546	na
s2 :		0	1965	na	s2 :		0	1941	na
s3 :			0	na	s3 :			0	na
s4 :				na	s4 :				na

TIME = t3...t11					TIME = t12:				
	s1	s2	s3	s4		s1	s2	s3	s4
	====	====	====	====		====	====	====	====
...					s1 :	0	591	1965	1173
					s2 :		0	1387	1245
					s3 :			0	1948
					s4 :				0

TAXA: 5 # x1 = tuba, x2 = pudor., x3 = polya., x4 = quadr., x5 = kawak.

PHYLOGENY (branch lengths in units of time):

((x1:9, (x2:5, x3:5)n3:4)n2:2, (x4:6, x4:6)n4:5)n1:1 ;

EXTANT DISTRIBUTION AND FOSSIL CONSTRAINTS:

	n1	n2	x1	n3	x2	x3	n4	x4	x5
	==	==	==	==	==	==	==	==	==
s1 :	.	.	0	.	0	1	.	0	1
s2 :	.	.	0	.	1	0	.	0	0
s3 :	.	.	1	.	0	0	.	0	0
s4 :	.	.	0	.	0	0	.	1	0

NUMBER OF START AREAS: 1

POSSIBLE START AREAS: s1 OR s2 OR s3 OR s4

PROBABILITY OF SURVIVAL 1.0

PROBABILITY OF DISPERSAL 0.2