# Phylogenetic tests of distribution patterns in South Asia: towards an integrative approach

SAYANTAN BISWAS[1,*] and SAMRAAT S PAWAR[2]

[1]*Deptartment of Biological Sciences, George Washington University, 2023 G St. NW., Washington DC 20052, USA*
[2]*Section of Integrative Biology, University of Texas, Austin, TX 78712, USA*

**Corresponding author (Fax, 1-202-994-6100 Email, sayantan@gwu.edu)*

The last four decades have seen an increasing integration of phylogenetics and biogeography. However, a dearth of phylogenetic studies has precluded such biogeographic analyses in South Asia until recently. Noting the increase in phylogenetic research and interest in phylogenetic biogeography in the region, we outline an integrative framework for studying taxon distribution patterns. While doing so, we pay particular attention to challenges posed by the complex geological and ecological history of the region, and the differences in distribution across taxonomic groups. We outline and compare three widely used phylogenetic biogeographic approaches: topology-based methods (TBMs), pattern-based methods (PBMs) and event-based methods (EBMs). TBMs lack a quantitative framework and utilize only part of the available phylogenetic information. Hence, they are mainly suited for preliminary enquiries. Both PBMs and EBMs have a quantitative framework, but we consider the latter to be particularly suited to the South Asian context since they consider multiple biogeographic processes explicitly, and can accommodate a reticulated history of areas. As an illustration, we present a biogeographic analysis of endemic Sri Lankan agamid lizards. The results provide insights into the relative importance of multiple processes and specific zones in the radiation of two speciose lizard clades.

[Biswas S and Pawar S S 2006 Phylogenetic tests of distribution patterns in South Asia: towards an integrative approach; *J. Biosci*. **31** 95–113]

## 1. Introduction

For over two centuries now, features in the distribution of organisms, such as endemism (restriction to a particular geographical area) and disjunction (geographical discontinuity in distribution), have attracted the attention of biologists. The quest to understand historical processes underlying these aspects of geographical distributions lies at the heart of biogeography. In the last few decades, significant progress has been made in this field due to the infusion of **phylogenetic systematics**[*] (Crisci *et al* 2003; Lomolino and Heaney 2004). Phylogenetic analyses allow inferences about the pattern and time scale of divergence between lineages and postulation of historical processes at levels of detail that were hitherto impossible without direct fossil evidence. However, in South Asia, such biogeographic analyses have begun to appear only recently, following an upsurge of phylogenetic studies (e.g. Bossuyt and Milinkovitch 2001; Conti *et al* 2002; Gower *et al* 2002; Bell and Donoghue 2003; Bossuyt *et al* 2004; Roelants *et al* 2004). In view of these developments, we discuss methodological frameworks that have been, and furthermore, can be used to study biogeographic patterns in the region with phylogenetic data. We stress on two aspects that deserve particular attention in the South Asian context: the challenges posed by the unique geological and ecological history of the region, and the need to consider differences in distribution patterns between taxonomic groups. Having described the three methodological frameworks, we suggest an integrative approach for biogeographic analyses in the region, and as an illustration, perform a preliminary analysis of endemic agamid lizards from Sri Lanka.

**Keywords.** Agamidae; disjunct distribution; dispersal; phylogenetic biogeography; Sri Lanka; vicariance

---

[*] Refer to glossary for text in bold.

### 1.1 *The new face of biogeography*

From the mid 19th to the mid 20th century, biogeography developed as a discipline founded on detailed descriptions of distribution patterns and explanatory theories supported by fossil and geological data (Brown and Lomolino 1998; Crisci *et al* 2003; Lomolino and Heaney 2004). However, the fossil record is often incomplete, and this fundamentally retrospective discipline lacked a framework within which the relative roles of the processes of speciation, dispersal, and extinction could be evaluated. For example, when closely related taxa showed disjunction, evaluations of these different historical processes were essentially descriptive, with a strong reliance on dispersal to explain disjunct distributions (Brown and Lomolino 1998).

In the 1960s, two developments changed the face of biogeography from a largely descriptive science to a hypothesis-driven one. These were the revival of Plate Tectonic Theory (see Kearey and Vine 1996), and the introduction of phylogenetics (Hennig 1966). Phylogenetics made it possible to reconstruct the historical relationships of species without direct fossil evidence. The potential of juxtaposing **cladograms** of species with their areas of distribution was soon recognised, and work in the 1970s (Platnick and Nelson 1978; Rosen 1978; Nelson and Platnick 1981; Crisci *et al* 2003), furthered Croizat's (1964) and Hennig's (1966) seminal contributions that led to the emergence of the field of phylogenetic biogeography.

At the same time, the widespread acceptance of plate tectonics shifted the focus of biogeography towards relationship of areas through explanations involving geographical **vicariance** and the associated origin of species due to isolation (termed "allopatric speciation"; see Mayr 1963; Coyne and Orr 2004). During this phase, vicariance became a biogeographic 'paradigm' at the cost of alternative processes, and dispersalist arguments in particular were criticized as ad-hoc and untestable (Rosen 1978; Crisci *et al* 2003; Donoghue and Moore 2003). Both philosophical and methodological reasons prevented a more pluralistic biogeography during this period. Little need was felt to include competing non-vicariant biogeographic processes to be as primary factors explaining distribution patterns, or to understand the biogeographic history of individual taxonomic groups. At the same time, there was an absence of analytical methods that could include multiple processes and molecular techniques to accurately infer the timing of events (Page 2003; Crisci *et al* 2003; Donoghue and Moore 2003; McDowall 2004).

However, a new face of biogeography has emerged over the last few decades. As data from the biogeography of a greater diversity of taxonomic groups have accumulated, the need to consider non-vicariant processes and events in order to arrive at realistic biogeographic inferences has become apparent (Page 1990; Ronquist and Nylin 1990; Ronquist 1997a, b). The plausibility of other non-vicariant processes has increasingly been appreciated, especially for studies examining geographical scales where various taxonomic groups are likely to have a different response to historical changes, generating incongruent patterns. This pluralistic shift in biogeography is particularly relevant in the context of South Asia, because of its complex ecological, climatic and geological history.

### 1.2 *The challenges of realistic biogeographic analyses*

Although considering multiple biogeographic processes is necessary, analytical complexity increases substantially with the inclusion of multiple processes or events. Figure 1 depicts a few possible biogeographic events, indicating this potential complexity. The classical dispersal and vicariance scenarios are represented in figure 1a–c. Figure 1d-g illustrate sympatric speciation, failure to speciate in response to a vicariant event, extinction, and loss of a pre-existing barrier respectively. Biogeographic events illustrated in figure 1a-d and f have been either modelled explicitly, or included as ad-hoc explanations for deviant patterns in the current methods discussed below. However, to appreciate the challenges of adopting such a pluralistic approach, consider the alternative **area cladograms** in figure 2. In figure 2a, each area is inhabited by only a single taxon unique to that area, a distribution pattern explainable by a single process (vicariance or dispersal). In reality however, such (relatively) unambiguous patterns are rare. Figure 2b-d represents alternative scenarios: widespread distribution (e.g. due to dispersal of Sp3 to SL; *cf.* figure 1b-c); presence of two taxa in the same area (e.g. due to a past event of sympatric speciation in NEI; *cf.* figure 1d) and absence of a taxon from one of the areas (e.g. no species in SEA due to extinction; *cf.* figure 1f). Such cases are more common than that in figure 2a, and highlight the need for considering multiple processes. However, in such cases, many alternative combinations of processes are possible, and the challenge is to arrive at the most tenable one(s). For instance, in figure 2a, considering only vicariant speciation requires invoking just three vicariance events corresponding to the three nodes. However, just considering vicariance + dispersal changes the situation dramatically. If dispersal is considered as likely as vicariance and independent of it, $2^3= 8$ different combinations of vicariance and dispersal events are possible, representing eight different area cladograms. This is a considerable addition of complexity to the pure single process scenario. Now, if one considers widespread taxa, where the same species occur in more than one area of endemism (*cf.* figure 2b), the number of possible area

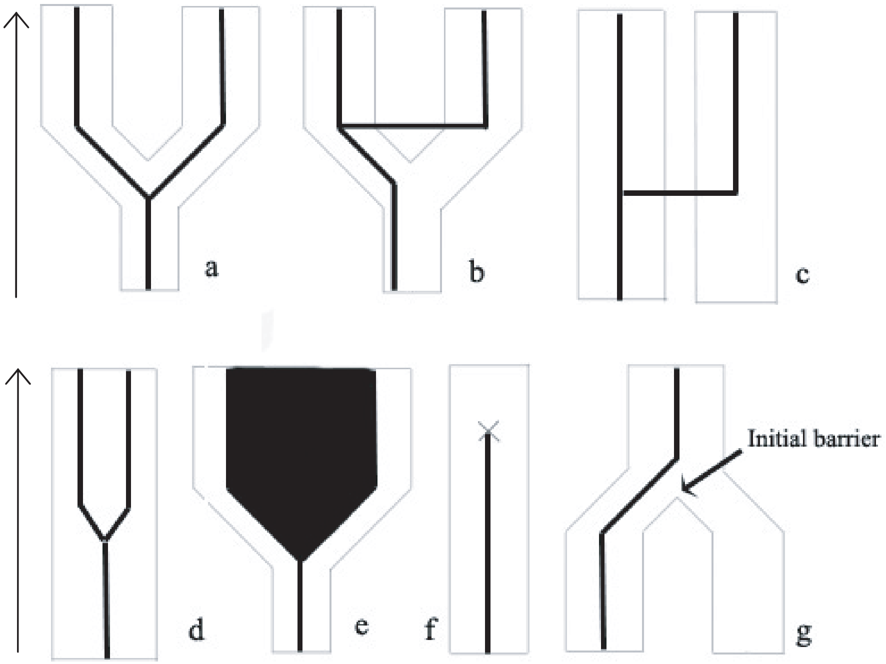


**Figure 1.** A few possible biogeographic events. The lighter and bolder lines represent areas and lineages respectively. Arrows along the left margin represent increasing evolutionary time (adapted from Page 2003; Ronquist 2003). (**a**) Vicariant event fragmenting a widely distributed taxon into two areas, leading to speciation of a daughter lineage in each. (**b**) Dispersal across the barrier to a neigbouring closely associated area after a vicariance event leading to subsequent speciation. (**c**) Dispersal across areas never closely associated followed by speciation. (**d**) Speciation independent of a vicariance event. (**e**) Failure to speciate in response to a vicariant event, resulting in a widespread taxon. (**f**) Extinction of a lineage within an area. (**g**) Dispersal due to loss of a barrier and subsequent expansion of range.

cladograms are

$$2^j + \sum_i n_i i,$$

where $j$ is the number of nodes in the taxon cladogram, $i$ the number of widespread taxa, and $n$ the number of ways in which each widespread taxon (in this case, Sp3) could have come to occupy more than one area, which depends upon the size of the set of all biologically and geologically feasible dispersal possibilities. Adding on sympatric or redundant distributions (e.g. due to sympatric speciation) and missing areas (e.g. from extinction), means that in effect, the possible alternative hypotheses for processes underlying an area cladogram will increase exponentially with the number of taxa and areas considered.

In the following sections, we introduce different frameworks of phylogenetic biogeographic analyses that deal with multiple processes. We use representative techniques and examples for illustration, and provide a comparative discussion in the context of South Asian biogeography.

## 2. Contemporary phylogenetic biogeographic methods: a brief overview

Since the 1980's, many phylogenetic biogeographic techniques have been formalised, differing mainly in their assumptions about the relative importance of vicariance, dispersal, and extinction, and their treatment of sympatric taxa, widespread taxa, and missing areas (refer to Crisci *et al* 2003 for a comprehensive review). Most of these current techniques can be divided into **topology-based** methods (TBMs), **pattern-based methods** or **cladistic biogeographic** methods (PBMs), and **event-based** methods (EBMs). The latter two are commonly identified as such in the literature (Ronquist 1997a; Crisci *et al* 2003), whereas TBMs are a predominantly descriptive class of analyses that we identify here, and have developed concurrently with PBMs and EBMs. We follow Sanmartín and Ronquist (2002) in referring to the cladistic biogeographic framework as pattern-based because its proponents treat processes implicitly in the search of patterns of **area relationships** (see PBMs in § 2.4 for further details). In this paper, we will

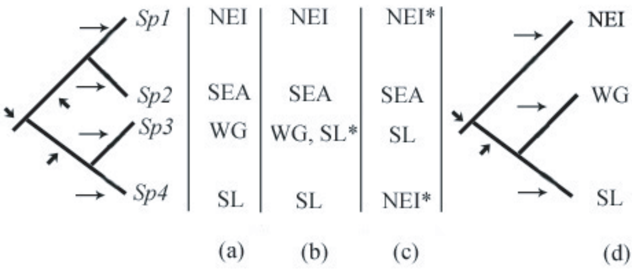


**Figure 2.** Relationship between taxa and areas in a phylogenetic biogeographic framework: Sp1 to Sp4 represent the taxa, and their areas of distribution (NEI, Northeast India; SEA, Southeast Asia; SL, Sri Lanka; WG, Western Ghats). (**a**) Exact match between taxa and areas with one area-one unique species associations. (**b**) Case of a widespread taxon due to Sp3 being distributed in WG and SL (marked by asterisk), resulting in a taxon area cladogram (TAC) with more than one area at one of its terminals. (**c**) Case of redundant distribution as Sp1 and Sp4 are both distributed in NEI (marked by asterisk), resulting in a TAC with more than one terminal with the same area. (**d**) Case of missing area, as no taxon occurs in SEA. Note that (**c**) is also a case of missing area as no taxon occurs in WG. See text for further discussion. Longer and shorter arrows respectively denote terms and components in the area cladogram.

henceforth invoke "cladistic biogeography" only for the sake of historical relevance.

A few authors have also categorized current methods as "*à priori*" and "*à posteriori*" (Van Veller and Brooks 2001; Van Veller *et al* 2002, 2003), depending on the stage at which non-vicariant processes are invoked to explain deviant patterns from a starting assumption of vicariance-induced area relationships. *A priori* methods (e.g. Component Analysis; see below) consider vicariance and other processes in individual monophyletic taxonomic groups, and then summarize common patterns across multiple groups. In contrast, *à posteriori* methods (e.g. Brooks Parsimony Analysis), combine data from all monophyletic taxonomic groups to produce 'general' vicariance-based area relationships, and then invoke non-vicariant processes to explain deviations. Proponents of this classification have interpreted the *à posteriori* approach as "deductive: or "evidence-based" because it involves rejecting vicariance-based null hypothesis, and the à priori approach as "inductive" or "model-based" (Van Veller and Brooks 2001; Van Veller *et al* 2002, 2003). Given the lack of comparative examination of biogeographic methods (e.g. Van Veller *et al* 2000; Ronquist 2003, and references therein), an appropriate classification covering relevant methods based on clearer theoretical and empirical understanding remains to be proposed. Hence, considering the state of the field, we prefer to use the distinction of PBM and EBM in this article as it clearly reflects the *modus operandi* of current quantitative phylogenetic biogeographic methods.

### 2.1 *Topology-based phylogenetic biogeographic methods*

TBMs are perhaps the most widely used technique currently adopted for phylogenetic biogeographic analyses. We classify TBM as any approach that infers vicariance and dispersal events in **monophyletic groups** using data from geology, phylogenetic tree **topology** and estimates of divergence times of lineages. As far as we know, there has been no formal exposition of this approach. TBMs are characterized by the following features:

(i) They start by developing alternative hypotheses about expected area cladograms for testing with data. These hypotheses are arrived at by comparing the geographical distributions of taxa, which are surmised to form monophyletic and **paraphyletic groups** and/or branches in a particular order in the **phylogenetic tree** in response to the geological, climatic and ecological history of the region (e.g. Macey *et al* 1998, 2000). This allows qualitative comparisons between vicariance and dispersal, and/or testing of different orders of occurrences of dispersal and vicariance events (see figure 3a).

(ii) Along with predictions about the nature of groupings and order of branching, the occurrence of specific biogeographic events may be inferred by comparing splitting time of **sister taxa** with the timing of past geological and/or ecological events (e.g. Bossuyt and Milinkovitch 2001; Gower *et al* 2002).

(iii) The tree topology derived from the data may be tested statistically for fits with the topologies predicted by alternative hypotheses (Macey *et al* 2000).

(iv) TBMs do not involve any quantitative analysis to evaluate the relative effects of multiple biogeographical events.

Here, using examples from studies on South Asian biota, we outline the TBM approach and point out a recent misrepresentation of the same. Consider a recent paper, that dealt with biotic relationships between the Western Ghats (Southern India) and Sri Lanka using multiple vertebrate and invertebrate groups (Bossuyt *et al* 2004). The authors concluded that Sri Lankan fauna was derived from mainland India, and there had been fewer instances of exchanges of taxa between the two regions than had been postulated earlier. Both of these inferences were drawn from examining the distribution of monophyletic and paraphyletic taxa. The inference of Sri Lankan fauna being derived from India was based on a nested position of the former within the latter (forming monophyletic groups), rendering the

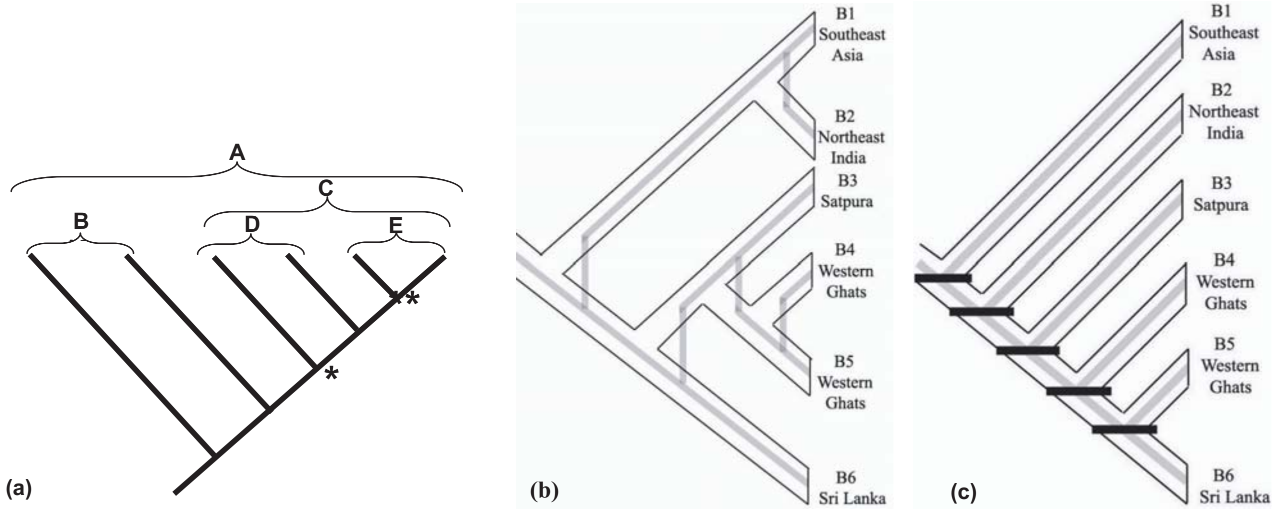


**Figure 3.** (**a**) A cartoon cladogram to illustrate how TBMs work. Here, the most inclusive monophyletic group A is composed of two sets of taxa B and C. C in turn is composed of two sets of taxa D and E. If A is considered to comprise of only B or C it will render A paraphyletic, because not all the descendants of the common ancestor at the base of the cladogram A are included. Similarly, considering just D or E with respect to the clade C will render C paraphyletic. Now imagine that distributions of these sets of taxa are as follows: B, E = India; D = Sri Lanka. This suggests that the common ancestor of the monophyletic group C was derived from India, suggesting an exchange between the two landmasses sometime during diversification (marked a single asterisk). The same argument holds for the monophyletic group E when only the cladogram C is considered (the exchange marked by a double asterisk). In totality, this suggests two exchanges from India to Sri Lanka and back to India, an argument used by Bossuyt *et al* (2004). However, to postulate specific biogeographic events (e.g. dispersal or vicariance) one needs to compare the timing of divergences of these taxa at hypothesized instances of exchanges with timing of separation of these two landmasses. Though this is the basic framework used in TBMs, more complicated situations with repeated dispersal or vicariance, or a combination of both can be inferred using a topology with branching order as in (**b**) or (**c**), or a combination of both. (**b, c**). The phylogenetic tree topology-based dispersal and vicariance models of Karanth (2003). These figures are meant to illustrate that distinguishing between vicariance and dispersal scenarios is not possible based on the phylogenetic tree topology alone, as presented by Karanth. (**b**) Uses modification of the "vicarance model" topology of Karanth to illustrate that a dispersal scenario fits it equally well. For all the divergences, the vicariant events (hollow lines) occurs first, followed by dispersal of ancestral populations across barriers and subsequent speciation into two descendent lineages on either side of the barrier. (**c**) The 'step-like' tree topology in the "dispersal model" of Karanth can equally accommodate successive vicariance events fragmenting the distribution of ancestral taxa and subsequent speciation into descendent lineages. Thus a distinction between these two scenarios can only be made when timing information for lineage divergences and fragmentation of areas is compared to assess their relative occurrences (*cf.* figure 1a-c). See text for further discussion.

Indian taxa paraphyletic with respect to the Sri Lankan ones (see figure 2 of Bossuyt *et al* 2004; figure 3a). The conclusion of limited exchange was also drawn similarly, with the direction of exchange inferred as being from India to Sri Lanka (because Sri Lankan species nested within an otherwise Indian **clade**) or from Sri Lanka to India (a few Indian species in turn nested within an otherwise Sri Lankan clade). It should be noted that this indicates exchange between the two regions, not the roles of specific biogeographic events (e.g. dispersal or vicariance). This would be possible when reconstruction of timing of divergences at postulated instances of faunal exchange (see table 4 in Bossuyt *et al* 2004) are correlated with the timing of past geological events. Thus, if divergences coincide with the formation of a barrier, or substantially post-date it, they may be categorized to be due to vicariance or dispersal, respectively. See the above references in this section for more detailed application of TBMs (see figure 3a).

One important limitation of the TBM approach is that it lacks an analytical framework to infer multiple events on the same area cladogram. As an illustration, consider the case of shield-tail snakes (Family: Uropeltidae). Shield-tails are endemic to southern India and Sri Lanka, and have undergone prolific radiation into the subterranean habitat (Gans 1976), with *ca*. 50 species currently recognized across 8 genera (a number that is likely increase significantly with further taxonomic work). However, diversification of these snakes remains largely unexplained and any comprehensive phylogenetic biogeographic analysis should investigate distributions at two geographical scales: those between the

Sri Lanka and Western Ghats lineages and those within the two regions. The phylogenetic analyses currently available for this group address only the former (Cadle *et al* 1990; Bossuyt *et al* 2004). Shield-tails are likely to exhibit all the features that complicate biogeographic analyses, (e.g. sympatry and widespread taxa), indicating that a strict vicariance and/or dispersal-based TBM approach would be inadequate because of the need to consider multiple processes. Nevertheless, when robust phylogenies and detailed geological history are available, TBMs within their limitations will allow testing vicariance and dispersal scenarios for simpler distribution patterns (*cf.* figure 2).

Appropriate use of TBMs hinges foremost on correctly setting up the framework for analysis, and also on recognizing the limitations of this framework. A recent approach suggested by Karanth (2003) for testing disjunct distributions in South Asian wet-zone taxa fails on both counts, and illustrates these points. Karanth (2003) proposes three phylogenetic tree topology based models: 'pure' vicariance, 'pure' dispersal, and convergence, and concludes that (i) the convergence model can be used to differentiate between true and false disjunct patterns and (ii) once "true disjuncts" are identified phylogenetic tree topology is sufficient to test whether the disjunctions are driven by dispersal or vicariance. As far as relationships among wet zone species are concerned, Karanth's (2003) dispersal and vicariance models make similar predictions: wet zone species from India and Sri Lanka should be more closely related to those from Northeast India and Southeast Asia than to nearby dry zone species (*cf.* Karanth 2003 for definition of zones). From Karanth's viewpoint, dispersal and vicariance models are mutually exclusive and can be differentiated based on the topology of the phylogenetic tree and position of **basal** vs. recent taxa. Under this criterion, the phylogenetic tree branching pattern appears step-like for dispersal, and reflects the sequence of fragmentation events for vicariance (figures 4 and 5 of Karanth 2003). However, in our opinion, none of the topologies *per se* are unique to a dispersal or vicariance scenario, and can be explained by competing scenarios, as depicted in figure 3b,c. Phylogenetic tree topology represents pattern of common ancestry, but not whether the ancestral taxa were fragmented and then evolved to two new species (vicariance scenario), or whether the ancestral taxa crossed an existing barrier and then evolved to two species on either sides of the barrier (dispersal scenario). The justification he provides for the above two models suggests that they offer an alternative way to distinguish vicariance from dispersal, since "from current faunal distribution and geological data, it is difficult to determine the relative age of populations versus their barriers" (Karanth 2003, p. 1278, column 1). In practice, Karanth implicitly assumes that timing of fragmentation and concurrent divergence of taxa are known. Hence, it might seem that tree topology alone is enough to arrive at such inferences about processes. But as we discuss above and illustrate in figure 3b,c, absence of timing data allows the same tree topology to be explained by either of the competing scenarios.

Thus on the whole, provided that the TBM is properly conceptualized, it can be used effectively, either as an exploratory tool, or for analyses of relatively broad-scale patterns.

### 2.2 *Pattern-based phylogenetic biogeographic methods*

The PBM framework seeks to maximize **congruence** of area cladograms from multiple monophyletic groups, with or without assuming a single **general area cladogram** (GAC). Depending on the method, this is achieved with or without *à priori* modification of the taxon area cladograms (TACs). Most methods involve *à posteriori* explanations to account for discrepancies from their assumptions (Van Veller and Brooks 2001; Van Veller *et al* 2003). One of the earliest PBM was Component Analysis (CA), developed by Nelson and Platnick (1981), and later implemented algorithmically and within a formal statistical framework by Page (1988, 1989). Many other pattern-based methods have been developed, often adapted from analogous techniques in host-parasite and species-gene tree research (Page 2003; see Crisci *et al* 2003). Here, we discuss CA alone as it serves to illustrate the PBM framework. CA involves the following steps:

(i) Identification of areas of endemism: This involves identifying three or more areas, each with one or more endemic species (Harold and Mooi 1994; Morrone 1994; Linder 2001). For South Asia, the Western Ghats, Northeast India, and Sri Lanka are a few such examples. Replacement of species in the phylogenetic tree with their areas yields a TAC (figure 2).

(ii) Extraction of components: The next step is to extract components i.e., the internal branches (the non-singleton clusters) of the area cladogram (see figure 2), which essentially summarizes the hypothesis of relationships among areas.

(iii) Derivation of parsimonious or **resolved area cladogram(s)** (RAC) for each monophyletic taxonomic group: To derive one or more RACs, CA deals with deviations from a straightforward vicariance model by making a one area-one unique species assumption. This is because a TAC, which is directly derived from taxon relationships, might shroud the underlying pattern of area relationships due to widespread taxa, redundant distributions (or sympatric taxa), and missing areas (Zandee and Roos 1987; Page 1988; Van Veller *et al* 2000, 2001; figure 2). A combination of biogeographic processes, termed assumptions 0 (A0), 1 (A1) and 2 (A2) are used to deal with deviations from a single area occupied by single unique species assumption (Nelson and Platnick 1981; Zandee and Roos 1987; Page 1988, 1989, 1990). Although the choice of assumptions to apply seems to depend upon the biogeographer's viewpoint, Page (1989) and Platnick (1981) have discussed how relative geographical distributions of the areas may help in such decisions (also see Van Veller *et al* 1999, who have suggested alternative criteria). The set of RACs suggested by application of the three assumptions may exhibit different degrees of overlap, but A0 and A2 may be considered the most and least restrictive of the rules respectively, allowing the lowest and highest number of alternative relationships among areas. Various combinations of these rules may also be applied for scenarios in the same area cladogram (e.g. A1 for widespread taxa and A2 for redundant distributions (but see Van Veller *et al* 1999). RACs are derived by applying these rules algorithmically, or by recoding the TACs and analysing the resulting component matrix using **parsimony** or other phylogenetic methods (Page 1988), either of which yield one or more RACs, depending on the assumption used. For each set of RACs for a taxonomic group, the ones that are least different from the original TAC are chosen. This is done by counting the differences in number of components and **terms** (the singleton clusters, or **terminals** of a cladogram) between the original TACs and the RAC(s) (termed "items of error" by Nelson and Platnick 1981).

(iv) Assessment of congruence between RACs of different co-distributed taxonomic groups; construction of a GAC: The next step is to assess the degree of congruence among most economical RACs for each taxonomic group, and arrive at a GAC. Occurrence of single, multiple or no GACs common to different RACs compared suggest one, more than one or no common biogeographic history among the areas of endemism, respectively (Crisci *et al* 2003). When there are multiple or no RACs at the intersection, similarity among the RACs is assessed using an appropriate measure of tree comparison and statistical significance of the observed similarity is tested against a random association of taxa and areas. Significant congruence is subsequently summarized (Page 1988, 1989, 1990). The hypothesis of area relationships indicated by the GACs is further compared with geological history to assess their plausibility. Congruence among area cladograms from different taxonomic groups is considered to be due to their common response to vicariant events. Dispersal is considered a random process, incapable of producing congruence (Rosen 1978). While they laid the foundation of quantitative phylogenetic biogeography, PBMs focussed primarily on general patterns of area relationships, and helped sustain the vicariance paradigm. Non-vicariant processes are considered in the PBM framework only to explain 'discrepancies' and their implicit treatment confounds biogeographic patterns with phylogenetic relationships (Page 1988, 1989, 1990).

### 2.3 *Event-based phylogenetic biogeographic methods*

Compared to the PBM framework, EBMs are relatively new, having appeared over the last decade (Ronquist 1997a). PBMs, which dominated the early phase of quantitative phylogenetic biogeography, have come under increasing criticism for their ad-hoc and restrictive treatment of alternative biogeographic processes, and the failure to explicitly reconstruct ancestral area for taxa (Sanmartín and Ronquist 2002; Ronquist 2003). The EBM framework provides a more balanced and explicit treatment of multiple processes within a single framework by assigning relative costs to different biogeographical events, and then deriving minimum-cost taxon area cladograms. Here, we discuss Dispersal-Vicariance Analysis (DIVA), an event-based approach in historical biogeography that has also been implemented as software (Ronquist 1996a).

The methodology of DIVA was originally developed in the context of host-parasite association studies, and was later adapted for phylogenetic biogeographic analyses (Ronquist and Nylin 1990; Ronquist 1996b, 1997a). In contrast to PBMs, DIVA does not assume the existence of a GAC or a **hierarchical history** of areas (Ronquist 1997a, b, 2003). The method considers individual taxonomic groups separately, and explicitly reconstructs ancestral areas of distribution for each pair of sister taxa**.** The methodology of DIVA can be summarized as follows:

(i) Consider a phylogeny, where terminals represent the taxa, and **internal nodes** the **hypothetical ancestors**. If there has been a shift between the distribution of the ancestral and terminal taxa, it has occurred somewhere along the branch connecting them.

(ii) Costs are assigned to changes between the distributional states in the descendants with respect to their immediate ancestors. The internal nodes are then assigned the distribution state through a series of **optimizations** that result in the lowest cost of biogeographic events over the whole area cladogram.

(iii) The reconstruction of ancestral nodes is carried out by a three-dimensional cost matrix, which is similar to a two dimensional cost matrix of ancestral-descendant character state transition. But the former has an additional dimension because each cost value consists of a combination of transitions from one ancestor to two descendent taxon distributions in a bifurcating tree (Ronquist 1997b, 1998).

(iv) Four events (or processes) are considered: vicariant speciation, dispersal, vicariance-independent speciation (i.e. 'duplication' of a lineage within an area), and extinction (*cf.* figure 1a–d, f).

(v) Being a parsimony approach, DIVA attempts to choose a set of events that result in the least overall cost for the entire area cladogram. Therefore, relative costs are assigned to the events based on a null model of associations between taxa and areas over **evolutionary time**, which assumes no change of association between taxa and areas over successive events of speciation. Of the four events, vicariance-independent speciation (duplication) and vicariant speciation satisfy this condition because descendent taxa will either remain restricted to the same ancestral area, or split into descendents occupying fragments of the ancestral area. These two are therefore assigned a value of zero. On the other hand, extinction and dispersal cause deviations from the null assumption because they change the association between taxa and areas over evolutionary time, and are therefore assigned a cost of one for each area lost (extinction) and/or added (dispersal).
(vi) The assignment of ancestral areas may be any combination of terminal areas, but these can result in computationally intractable optimizations. Hence, two rules are used to restrict the number of combinations in defining the ancestral area: (a) the optimal distribution at any ancestral node cannot include any area not occupied by its descendents, and (b) the **optimal ancestral area** should include at least one area from each descendent node (Ronquist 1997a).

The logic of DIVA has been extended to include multiple taxonomic groups and inference of GACs within an event-based framework (Sanmartín and Ronquist 2002; Ronquist 2003), and implemented in the software TreeFitter (also see Page 1994 for an earlier approach; Ronquist 2002). The GACs are reconstructed either from paleogeographic reconstructions of area relationships (Sanmartín *et al* 2001; Sanmartín and Ronquist 2004), or by comparing area cladograms from individual taxonomic groups with all possible topologies of area cladograms for all the areas considered, and choosing the one that postulates least costly biogeographic events (Sanmartín *et al* 2001; Ronquist 2002; Sanmartín and Ronquist 2004). TreeFitter also allows assigning a range of different costs for biogeographic events and exploration of the resultant parameter space through sensitivity analysis (see Sanmartín and Ronquist 2004).

## 3. The promises and the challenges of South Asian biogeography

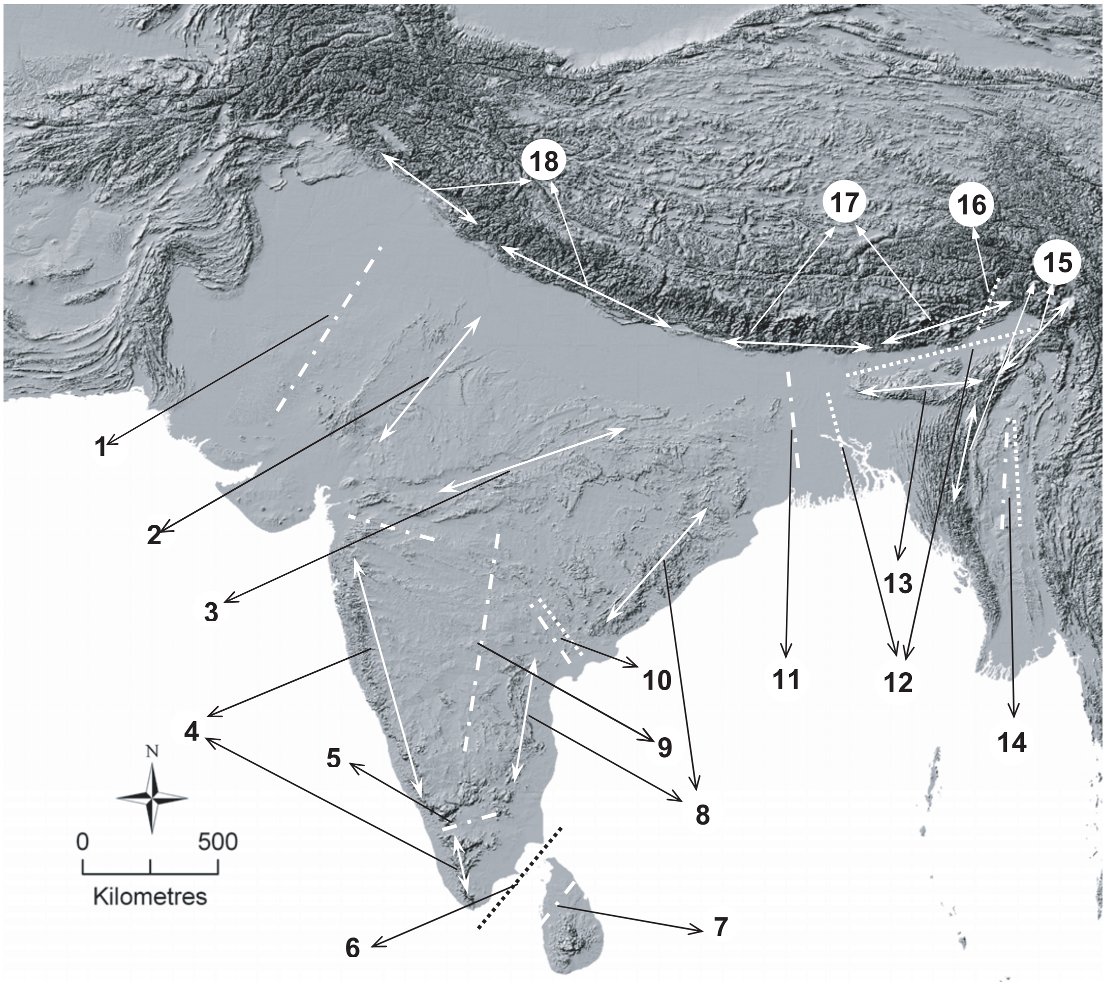
Given the three methodological frameworks described above, the moot question is, “Which method is appropriate?” There is of course no straightforward answer to this, but considering the specific context of South Asia may provide some directions. South Asia is an exceptionally interesting and challenging region for biogeographers due to its antiquity, unique plate tectonic and palaeoclimatic history, location at the confluence of biogeographical realms, and astounding array of physiognomic and habitat diversity (Mani 1974). These factors have contributed to a considerable diversification of biota in the region (Mani 1974; Fernando 1984; Morley 2000). Additionally, changes in the extents of habitats, especially of tropical moist forests in space and time, have resulted in striking disjunctions in ranges of almost all taxonomic groups, inviting much debate about processes underlying them (Randhawa 1945; Hora 1949; Dilger 1952; Mani 1974; Fernando 1984; Swan 1993). As an illustration, consider figure 4, which shows important physiognomic and ecological features that have been considered in previous biogeographical analyses of terrestrial wet forest biota. Clearly, the South Asian region contains a complex assortment of barriers and dispersal routes varying in their nature as well as spatial scale to which multiple co-distributed taxonomic groups are unlikely to show a congruent response. To compound the challenges posed by this setting, the poor fossil record from this region has hitherto limited biogeographical analyses to a great extent.

### 3.1 *The role of phylogenetic biogeography in the South Asian context*

With the burgeoning of phylogenetic work on regional taxa, the time is, however, ripe for testing and refining biogeographical hypotheses within a quantitative phylogenetic biogeographic framework for different taxonomic groups. These include (but are not restricted to), understanding relationships of regionally restricted biota, and exchange between areas [e.g. Did the Western Ghats separate first from the central Indian moist forests and northern Orissa or did Northeast India? How often did “out of India” and “into India” dispersals occur (Bossuyt and Milinkovitch 2001; Gower *et al* 2002)? How many times have the Western Ghats and Sri Lanka actually exchanged biota (Bossuyt *et al* 2004)?]. One should expect different taxonomic groups to have responded differently to geological and paleoecological events, and even within the same clade, different effects could be seen across lineages. Moreover, these analyses will also help establish the relative roles of biogeographic processes such as vicariance, dispersal, extinction and sympatric speciation in generating current patterns.

### 3.2 *Taxonomic problems*

Given the lack of previous phylogenetic work in almost all invertebrates and most major groups of vertebrates and plants in South Asia, employing current taxonomy for drawing biogeographic inferences may lead to spurious results. This critical point has also been raised by Karanth (2003), who presents a “convergence model” (referring to the possible convergent evolution of morphological and



1. Thar desert and Northwestern arid zone of Indian Subcontinent
2. Aravalli range and the Brij region
3. Satpura-Vindhya ranges
4. Western Ghats
5. Palghat Gap
6. Gulf of Mannar
7. Northwestern arid zone of Sri Lanka
8. Eastern Ghats
9. Dry zone of Peninsular India
10. Godavari river and basin
11. Garo-Rajmahal Gap
12. Brahmaputra river and Ganga-Brahmaputra delta
13. Garo, Khasi, and Jaintia Hills
14. The Ayerawaddy river and basin
15. Rakhine-Yoma ranges
16. Siang River
17. Outer-Eastern Himalayas
18. Outer-Western Himalayas and Siwaliks

**Figure 4.** Digital elevation model of South Asia highlighting some of the features that have been considered important in previous biogeographical literature on terrestrial wet forest biota. Dotted lines represent potential aquatic barriers, dash-dotted lines terrestrial ones (dry zones), and bidirectional arrows potentially contiguous dispersal routes. Note that the impact of barriers and dispersal routes need to be evaluated with respect to the life history characteristics of the specific organisms or taxonomic groups in question, and both, timing and persistence of the processes involved. Also, though not depicted in this map, exchange of biota between south Asia and neighbouring regions is also an important consideration for understanding of the biogeography of the region.

ecological traits). Here, "false disjuncts" appear due to incorrect taxonomy which needs to be first eliminated before proceeding with biogeographic analyses. For example, Karanth cites the ostensible disjunction in distribution of the wet-zone langur monkey species in South Asia (currently, the genus *Trachypithecus*). Because recent phylogenetic data indicate that phenotypically similar species hitherto considered disjunct are actually not sister taxa (Karanth 2003), this is probably a false pattern (convergence in Karanth's definition).

This is a valid point, but Karanth's TBM approach of identifying convergence is inappropriate in a biogeographic context unless it is reframed as a scenario underpinned by a process such as parapatric speciation or divergence across dry-wet zones or other ecological gradients (Moritz *et al* 2000). Moreover, Karanth suggests that if there is disagreement between taxonomy and the TBM analysis, taxonomic revision would be in order (Karanth 2003, p. 1279, column 2). This might be true in cases when wet zone species are wrongly classified together to the exclusion of dry zone species (as in the langur example above). However, we would like to point out that taxonomic revision (or lack thereof) does not always affect phylogenetic biogeographic analyses *per se*, because given a robust phylogenetic tree that includes relevant taxa (irrespective of current taxonomy), comparisons are automatically made between appropriate sister taxa.

### 3.3 *Choosing a phylogenetic biogeographic framework for analysis*

Having outlined some of the challenges facing phylogenetic biogeographic analyses in South Asia, we now return to the issue of the choice of methods. Figure 5 presents a framework for such analyses in the region. We do not claim that this is the best possible decision tree for the South Asian region, as rapid developments in phylogenetic biogeography are bound to require frequent revisions of this framework. As discussed above, the EBM framework presently appears to be the most flexible and most capable of a pluralistic treatment, compared to TBMs and PBMs (also see conclusions below). In addition, in the South Asian context, apart from a few taxonomic groups, the choice of a method will currently be limited by availability of phylogenetic, distribution and geological data. The nature of the question being asked is, of course, also an important factor. For example, in analyses at larger spatial scales, as between Peninsular India and Northeast India, a TBM approach might suffice to answer questions about faunal exchange between the subregions. But at finer-scales where localized exchanges (such as across the Palghat Gap or the Brahmaputra river) result in redundant distributions and widespread taxa, an EBM would be more appropriate.

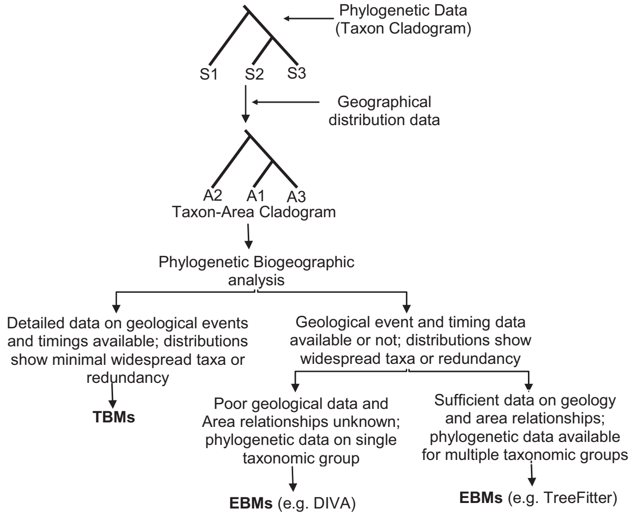


**Figure 5.** An integrated methodological framework for phylogenetic biogeographic analyses. Replacing the terminals of a taxon cladogram with their respective areas of distribution yields a Taxon-Area Cladogram, which can then be subjected to biogeographic analyses. Note that to start with, TACs are rarely well-resolved, and can be quite complicated, as hinted in figure 2. See text for further discussion.

## 4. An empirical example: the endemic agamid lizards of Sri Lanka

We now present a phylogenetic biogeographic analysis of endemic agamid lizards from Sri Lanka. We have two motivations for performing this analysis. Firstly, this will illustrate a quantitative phylogenetic biogeographic approach. Secondly, being the first such analysis, it will also provide insights for future biogeographical work on these lizards and the island's other endemic biota. We use the EBM, DIVA for this analysis because (i) it accommodates multiple processes, an important factor considering the redundancy and widespread taxa evident from the distribution of these agamids, (ii) it can be applied without *à priori* understanding of area relationships, which remain ambiguous due to the complex geological and ecological history of Sri Lanka (Cooray 1967; Ashton and Gunatilleke 1987; Erdelen and Preu 1990), and (iii) it provides an analytical framework within which individual taxonomic groups can be studied without the restrictions imposed by the GAC-oriented approach of PBMs.

The agamid lizards of Sri Lanka represent a unique radiation exhibiting high degree of endemism (> 82%). Among other things, these poorly known lizards are characterized by morphological features that find few parallels elsewhere (Manamendra-Arachchi and Liyanage 1994; Pethiyagoda and Manamendra-Arachchi 1998; Bahir and Maduwage 2005; Bahir and Silva 2005).

### 4.1 *Methods*

There are 17 species of agamid lizards distributed in Sri Lanka, represented by two or more clades (Pethiyagoda and Manamendra-Arachchi 1998; Schulte *et al* 2004; Bahir and Maduwage 2005; Bahir and Silva 2005). Phylogenetic information was gathered from recent analyses of Sri Lankan agamid lizards inferred from a large molecular data set and broad taxon sampling (Macey *et al* 2000; Schulte II *et al* 2002; Schulte *et al* 2004). We have only included the *Ceratophora-Lyriocephalus-Cophotis* and the *Calotes* clades. The *Sitana ponticeriana-Otocryptis weigmanni* clade, though included in the original phylogeny, was excluded because its position is poorly supported in the analyses (*cf.* Macey *et al* 2000; Schulte *et al* 2004). Since Sri Lankan agamids do not form a single monophyletic group, the two clades included were analysed separately. Based upon distribution maps from Manamendra-Arachchi and Liyanage (1994) and Pethiyagoda and Manamendra-Arachchi (1998), species were assigned to one or more biogeographic zones based on a slightly modified version of Senanayake *et al.*'s (1977) classification (figure 6), which is supported by multiple taxonomic groups. The resulting distributions are summarized in table 2.

One of the current drawbacks of DIVA is that it tends to produce widespread ancestral taxa towards the **root** of the tree because each tree is part of a larger tree of life, and the algorithm needs further information from this larger tree to reliably reconstruct the ancestral areas towards the root (Ronquist 1996a). This problem may be reduced by including **outgroups** and parts of the phylogenetic tree other than the **ingroup**. Another option is to constrain the number of ancestral areas to two (Ronquist 1996a; Voelker 1999; Austin *et al* 2003), though constraining the number of ancestral areas is likely to increase number of postulated dispersal events. We adopted both these approaches in our analysis. As described above, DIVA reconstructs ancestral distributions by considering the costs of four events (*cf.* Ronquist 1997a): vicariant speciation and vicariance-independent speciation within an area (cost = 0), dispersal between areas (cost = 1 unit per distribution added), and extinction (cost = 1 unit per area of distribution deleted).

### 4.2 *Results and discussion*

For both clades, the analyses indicate a combination of dispersal, vicariant speciation, and vicariance-independent speciation.

*Ceratophora-Cophotis-Lyriocephalus* group: When the number of ancestral areas was left unconstrained, the analysis suggested six dispersals, six vicariance-independent speciations and three vicariant events for the Sri Lankan endemics (*Ceratophora*, *Cophotis*, *Lyriocephalus*), with three possible ancestral areas for the common ancestor of *Cophotis ceylanica* and *Lyriocephalus scutatus* (see figure 7i-iii). The genus *Ceratophora* seems to have evolved within the wet zone of peneplain 2 (wzp2), following a vicariant event separating wzp2 from other zones (abde) in Sri Lanka. However, subsequent to their divergence from the **most recent common ancestor** (MRCA), three of the species, *C. stoddartii*, *C. tennentii* and *C. aspera*, also dispersed to three other zones; *C. stoddartii* to peneplain3 (e) and dry zone of peneplain2 (d) and *C. tennentii* and *C. aspera* to peneplain 3(e) and wet zone of peneplain1 (b), respectively.

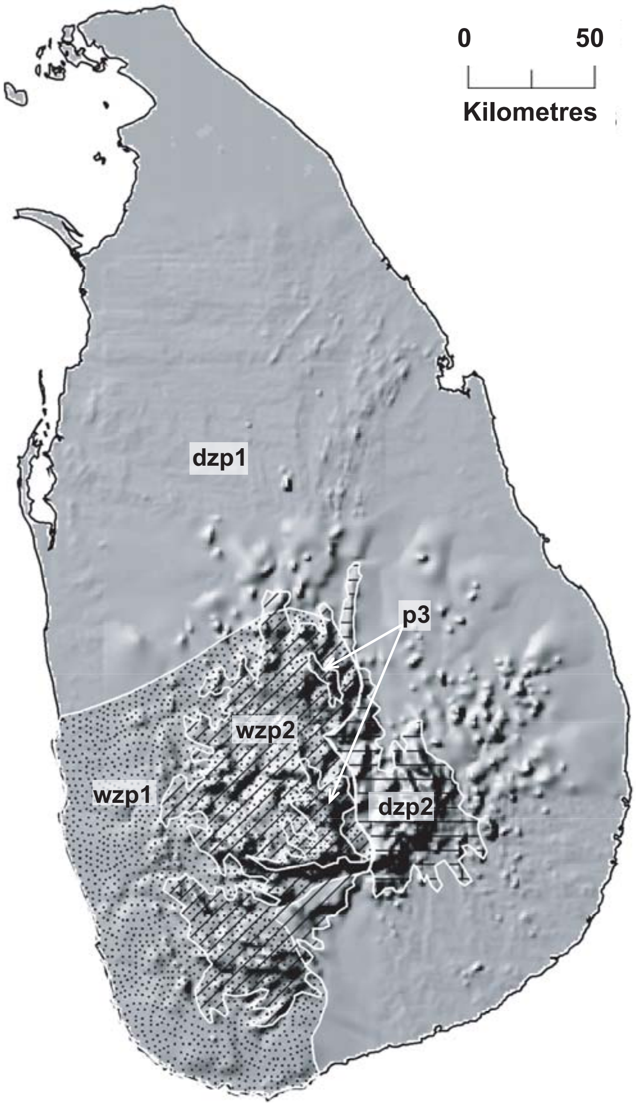


**Figure 6.** Digital elevation map of Sri Lanka showing areas of endemism used in the biogeographic analysis of Sri Lankan agamid lizards (*cf.* table 2 and figure 7). The zones are, dzp1, Dry zone of peneplain 1; wzp1**,** wet zone of peneplain 1; wzp2, wet zone of peneplain 2; dzp2, dry zone of peneplain 2; p3, peneplain 3. Classification of zones is based upon Senanayake *et al.* (1977), with their peneplain 2 further divided into wet and dry zones (wzp2 and dzp2, respectively). The three peneplains are located at successively higher elevations. Location of the three zones outside Sri Lanka used in the analysis can be located from the following features marked in figure 4: Western Ghats – 4; Peninsular India – 8–10; Northeast India – the region marked by 11–17; Southeast Asia – the region southeast of 14–15.

The alternative events hypothesized to have affected the ancestral distribution of *Cophotis ceylanica* and

**Table 1.** Three ‘rules’ used in cladistic biogeography. These rules are used to resolve deviations from the assumption of vicariance-driven one area-one unique species associations. Note that missing areas are treated identically under Assumption 1 and Assumption 2. See text for further details.

| | Rules | | |
|---|---|---|---|
| Scenarios | Assumption 0 | Assumption 1 | Assumption 2 |
| Missing Areas (= absence of taxa from one or more areas) | Absence due to 'failure' of the ancestral taxon to disperse into that area or extinction | Absence due to 'failure' to disperse into that area, extinction or lack of sampling | Absence due to 'failure' to disperse into that area, extinction or lack of sampling |
| Widespread taxa (= two or more areas with same taxon) | Areas suggested by widespread taxon are sister areas, caused by ancestral taxon's 'failure' to vicariate | Failure of ancestral taxon to vicariate, possibly in combination with subsequent extinction; the way one area is treated is also applicable for other areas | Failure of ancestral taxon to disperse, or any combination of these; areas may be sister to each other, or occurring at any position on the tree; each area treated separately |
| Redundant distributions (= two or more taxa in same area) | Duplicated areas due to vicariance independent speciation or undersampling | Duplicated areas due to vicariance independent speciation and extinction; the way one redundant area is treated is also applicable for other areas | Duplicated areas due to vicariance independent speciation, extinction and dispersal; each redundant area treated separately |

**Table 2.** The distribution of Sri Lankan agamid lizards in areas of endemism used in the biogeographic analysis shown in figure 7.

| | Areas of endemism | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Taxa | dzp1 | wzp1 | wzp2 | dzp2 | p3 | wg | pi | nei | sea |
| *Calotes liolepis* | X | X | X | | | | | | |
| *C. liocephalus* | | | X | | X | | | | |
| *C. ceylonensis* | X | | X | | X | | | | |
| *C. nigrilabris* | | | X | X | X | | | | |
| *C. calotes* | X | X | X | X | X | X | X | | |
| *C. mystaceus* | | | | | | | | X | X |
| *C. emma* | | | | | | | | X | X |
| *Lyriocephalus scutatus* | | X | X | | | | | | |
| *Cophotis ceylanica* | X | | X | X | X | | | | |
| *Ceratophora karu* | | | X | | | | | | |
| *C. erdeleni* | | | X | | | | | | |
| *C. stoddarti* | | | X | X | X | | | | |
| *C. tennenti* | | | X | | X | | | | |
| *C. aspera* | | X | X | | | | | | |
| *Aphaniotis fusca* | | | | | | | | | X |
| *Bronchocela cristatella* | | | | | | | | | X |
| *Gonocephalus grandis* | | | | | | | | | X |

Presence of lizards are marked with ‘X’. Area codes used here and in the text are as follows: dzp1, Dry zone of peneplain 1; wzp1**,** wet zone of peneplain 1; wzp2**,** wet zone of peneplain 2; dzp2**,** dry zone of peneplain 2; p3, peneplain 3; wg, Western Ghats; pi, peninsular India, excluding WG; nei, northeast India; sea, Southeast Asia including Andaman and Nicobar islands and south China. Also see figure 6. Distribution data are from Erdelen (1996), Manamendra-Arachchi and Liyanage (1994), Pethiyagoda and Manamendra-Arachchi (1998), and Uetz (2004).

*Lyriocephalus scutatus* are as follows. The first scenario, represented in figure 7i, suggests a vicariant event between the wzp1 (b) and dzp1, dzp2 and p3 (ade), followed by independent dispersal of each of the species to the wzp2 (c). The second scenario (figure 7ii) suggests a vicariant event separating the ancestral distribution of these two taxa into dzp1, dzp2, wzp2 and p3 (acde) and wzp1 (b) followed by dispersal of *Lyriocephalus scutatus* into dzp2 (c). The third (figure 7iii) suggests a vicariant event separating the ancestral distribution of these two taxa into dzp1, dzp2 and p3 (ade) and wzp1 and wzp2 (bc) followed by dispersal of *Cophotis ceylanica* into dzp2 (c). Successive constraining of ancestral areas to two increased the proposed number of dispersal events from six to eight in these endemic genera (results not shown).

*Calotes* group: When the number of ancestral areas was left unconstrained, the analysis suggested ten dispersals, four vicariance-independent speciation events and two vicariant events for the Sri Lankan *Calotes* spp. (see figure 7iv). For the two interior nodes (the root of the tree and MRCA of *C. mystaceus* and *C. emma*), three alternative areas were inferred. However, these are not discussed further as they do not alter the optimization of ancestral areas for the focal Sri Lankan *Calotes* spp. (*cf.* Thornton *et al* 2001). All endemic *Calotes* spp. from Sri Lanka are proposed to have speciated within the wzp2 (C), followed by dispersal to all the other zones of the peneplains. *Calotes calotes*, which is currently distributed in Sri Lanka, Western Ghats and Peninsular India, is hypothesized to have been widely distributed in the most of the above areas before dispersing into wzp2 of Sri Lanka. The endemic *Calotes* species dispersed three times to the p3 (e) (*C. ceylonsis*, *C. liocephalus*, and *C. nigrilabris*): twice to dzp1 (a) (*C. ceylonensis* and *C. liolepis*) and once to wzp1 (b) (*C. liolepis*) and dzp2 (d) (*C. nigrilabris*) (Erdelen 1988). Constraining the ancestral areas up to two for the *Calotes* species resulted in an increase in dispersal events from 10 in unrestricted analysis to 15 (results not shown)

To the best of our knowledge, this preliminary analysis is one of the first applications of quantitative phylogenetic biogeography in South Asia (also see Bell and Donoghue 2003), and the first to address within-Sri Lanka diversification of endemic reptiles. The first noteworthy detail of our results is that all the endemic Sri Lankan agamid species, except *Cophotis ceylanica* and *Lyriocephalus scutatus* probably evolved within the wzp2 (c). This is broad agreement with the information that the wet zone harbours the majority of endemic Sri Lankan vertebrates (Senanayake *et al* 1977; Senanayake and Moyle 1982; Das 1996; Pethiyagoda and Manamendra-Arachchi 1998). The analyses also suggest that taxa have colonized wzp2 multiple times (see figure 7). Another striking result is the dispersal of five of the eleven endemic species from wzp2 to p3. Also, given the present location of the different peneplain zones, certain postulations of ancestral distributions are probably unlikely. This refers to the two dispersal events into the wzp2 (c) associated with the ancestors of *Cophotis ceylanica* (figure 7i and 7ii), and *Calotes calotes* (figure 7iv). In both cases, the ancestors occur in p3 (e) along with other zones before dispersing into wzp2. However, p3 is situated completely within the wzp2c, which raises the question of how the ancestral taxon occurred in p3 without occurring in the intervening wzp2. Such events are only possible if jump dispersal (of the ancestral taxon) into p3 has occurred, surpassing wzp2, or if p3 and wzp2 had a different relative configuration in the past. Also, identification of areas of diversification is likely to be more informative in cases of datasets where multiple, competing species-rich candidate areas are present. In the case of Sri Lankan agamids this is not the case, and it comes as no surprise that the analyses identify the most species-rich area wzp2 (see table 2). However, it must be noted that the value of the EBM approach in this case lies not so much in the identification of wzp2 specifically as the area of diversification, but more in the fact that multiple biogeographic events have been implicated, thus generating fine-scale biogeographic hypotheses that can be tested with further data.

The plausibility of dispersal events hypothesized above may be further tested using phylogeographic analyses (Nielsen and Wakeley 2001; Knowles 2004; Templeton 2004) of conspecific populations distributed across multiple peneplain zones. For example, populations from which dispersal into other peneplains have been proposed may exhibit significantly higher genetic divergence and deeper population structure than the latter. Apart from phylogeographic analyses, correlation of the biogeographic events postulated above with past geologic and climatic changes will also test the validity of these results. However, lack of fossils and reliable molecular dating estimates along with poor data on past geology and climate of Cenozoic (<65 mybp) Sri Lanka (Cooray 1967) renders it difficult to assess the influence of wzp2 on diversification of endemic agamid taxa at the moment. In this context of diversification, it is interesting to note also that DIVA suggests non-vicariant speciation in these lizards (through parapatric, peripatric, or sympatric speciation; see Coyne and Orr 2004). Despite this, and the fact that profuse diversification in both agamid clades has occurred in a relatively small area, we exercise caution in invoking non-allopatric speciation because without more detailed palaeoecological, and phylogeographic data, fine-scale allopatry cannot be overruled.

Such situations also raise the question how information from fossils, past geology and climate, and phylogeny-based analyses may contribute to the understanding of regional biogeography. South Asia is unlikely to offer the ideal situation where phylogeny-based hypotheses of past biogeographic events can be tested with fossils, past geology,

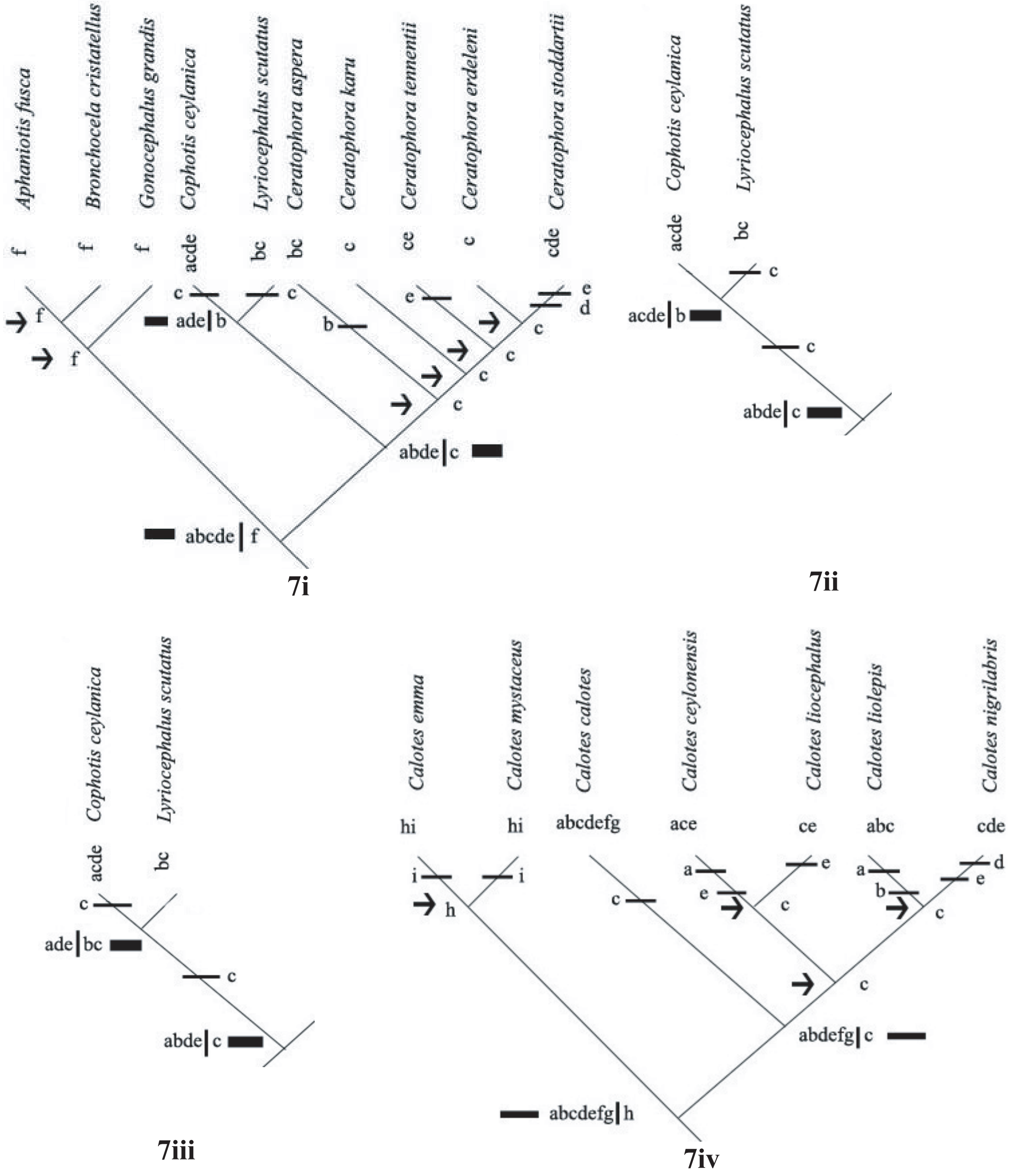


**Figure 7.** Reconstruction of ancestral areas and hypotheses of biogeographic events for two Sri Lankan agamid lizard clades using DIVA. These area cladograms show species and their known areas of distribution at terminals of the cladograms. Refer to table 2 for area codes and a summary of the distributions, and figure 6 for location of the areas. Phylogenies were extracted and adapted from Schulte *et al.*(2004). Hypothesized biogeographic events are marked as follows: vicariance (black rectangle) suggested between areas separated by the vertical bar, dispersal (horizontal bar) shown across the branch, vicariance-independent speciation (arrow). 7i refers to the *Ceratophora*-*Cophotis*-*Lyriocephalus* clade, with 7ii and 7iii depicting the two other alternatives for the ancestral distribution of *Cophotis ceylanica* and *Lyriocephalus scutatus*. 7iv includes the endemic *Calotes* spp. of Sri Lanka. Area codes are (a) Dry zone of peneplain 1; (b) wet zone of peneplain 1; (c) wet zone of peneplain 2; (d) dry zone of peneplain 2; (e) peneplain 3; (f) Western Ghats; (g) peninsular India, excluding WG; (h) northeast India; (i) Southeast Asia including Andaman and Nicobar islands and south China.

and paleoclimatic data. Instead, various sources are likely to play complementary roles in providing support to such hypotheses. For example, preliminary hypotheses of biogeographic events may be gathered from phylogenetic analysis and partitioned across different time periods using molecular dating (Donoghue and Moore 2003). These may lead to predictions that may be further tested by phylogeographic analyses and comparative data from other taxonomic groups. Finally, whenever available, data on past geologic events, modelling of past climatic conditions and fossils may provide further support.

The geographical contiguity of Sri Lanka and India following fragmentation of the Gondwana (Cooray 1967) allowed the possibility of biotic exchange between the two regions until the insularisation of Sri Lanka during the Miocene (Cooray 1967; Erdelen 1996). This scenario, contrasting with the high endemicity of Sri Lankan herpetofauna, has generated two apparently opposing hypotheses of repeated faunal exchange (Erdelen and Preu 1990) and *in-situ* diversification (Bossuyt *et al* 2004) as determinant of the island's endemic fauna. Our results, while supporting the latter, provides insights for further examination. Hypotheses suggesting faunal exchange (Erdelen and Preu 1990) and independent evolution (Bossuyt *et al* 2004) are not necessarily mutually exclusive, but are likely to have differential impact depending upon the taxonomic group. Traditionally, hypotheses of tropical diversification have stressed the importance of isolation and distance, though recent studies increasingly indicate the importance of diversification across ecological gradients without any apparent barriers (Moritz *et al* 2000). Hence a more realistic approach would be to consider both the scenarios, and assess the importance of each. Detailed phylogenetic and phylogeographic studies and estimates of dates of divergences for endemic and non-endemic herpetofauna from Sri Lanka and adjoining mainland India and Southeast Asia will significantly contribute to such reassessments. More definitive understanding will require sampling of taxa from south India and other parts of Asia, and more robust phylogenetic hypotheses for deeper divergences, along with investigation of other taxonomic groups.

Finally, the limitations of EBMs such as DIVA also need to be considered. For example DIVA requires a strictly bifurcating tree, tends to postulate a greater number of ancestral areas towards the root, and rarely implicates extinction. As far as extinctions are concerned, only when relationships of areas are available in the form of a GAC as in a framework where multiple taxonomic groups are compared (Sanmartín *et al* 2001; Sanmartín and Ronquist 2004), absence data can be inferred, and enforcing geographical constraints then leads to postulation of extinction events (Ronquist 1996a). Otherwise, for analyses of single taxonomic groups, absence data are explained by lack of dispersal to an area that precludes inference of extinction events.

## 5. Conclusions

The results of this analysis clearly illustrate the utility of using a phylogenetic biogeographic framework that can explicitly accommodate multiple events indicated for both the lizard groups. There will of course be future improvements in current methods, probably including but not limited to, consideration of additional biogeographic events, development of more rigorous statistical frameworks, consideration of phylogenetic reconstruction uncertainty, inference of ancestral areas not shared by descendent taxa, incorporation of biological attributes of taxa (e.g. relative dispersal ability), and assessment of congruence among distribution patterns of different taxa partitioned over comparable evolutionary time (Donoghue and Moore 2003). Improved resolution of phylogenetic biogeographic inferences will also be achieved through greater integration with **phylogeography** (Rannala and Michalakis 2003; Wakeley 2004; Lomolino and Heaney 2004), and incorporation of ecological niche modelling (Peterson *et al* 1999; Graham *et al* 2004a, b).

No matter what direction phylogenetic biogeography takes, analyses in South Asia will first have to overcome the deficiency of phylogenetic studies in South Asia. As more phylogenetic data accumulates, and the renewed interest in South Asian biogeography continues to surge, the next challenge will be to adopt a methodologically and conceptually integrative approach. There is no simple path towards a better understanding of the subcontinent's complex biogeography. Here, we have tried to highlight both the challenges involved in analysing the biogeography of the region, and the methods that might yield insights and provide direction for further enquiry. All available information has been attenuated by loss over evolutionary time, and there will always be cases where past events influencing present distributions just cannot be inferred. Therefore, comparative biogeography of multiple taxonomic groups spanning spatial scales and evolutionary times is particularly important. Only then will generalities as well as important specifics of the region's biogeographical history begin to emerge.

## Acknowledgements

We thank Michelle Koo for generously providing us with digital elevation base maps and the Smithsonian Institution Libraries for rare literature. SB thanks the Ashoka Trust for Research in Ecology and the Environment for providing computer access during the last quarter of this article's preparation. We thank Juan Carlos Santos and two anonymous reviewers for comments on this article.

## Glossary (for terminology highlighted in the text)

| | |
|---|---|
| Area relationships | Historical geological relationships between areas |
| Area cladogram | A tree depicting historical relationships between geographical areas, either hypothesized by replacing each taxon in a phylogenetic tree or cladogram with the area(s) it occurs in, or reconstructed from direct geological evidence |
| Basal taxa | Taxa diverging closer to the root (see below) of a phylogenetic tree |
| Clade (or Monophyletic group) | A group of taxa that includes a common ancestor and only all of its descendents |
| Cladistic biogeography (or pattern-based methods) | An approach in historical biogeography focussed mainly on understanding patterns of area relationships resulting from vicariance events |
| Cladogram (or Phylogenetic Tree | A branching or bifurcating diagram representing a hypothesis of hierarchical relationships between taxa based upon shared derived character states (i.e., two or more alternative expressions of a character) |
| Congruence | Degree of concordance between two or more patterns such as the branching of cladograms, phylogenetic trees, or area cladograms |
| Event-based methods | An approach in historical biogeography where possible events or processes are explicitly hypothesized and assigned relative costs before performing analysis |
| Evolutionary time | The time scale over which significant shifts in mean genetic or phenotypic trait values takes place in populations |
| General area cladogram | The cladogram at the intersection of Resolved Area Cladogram(s) (RAC) (see below) from two or more co-distributed monophyletic taxonomic groups. This is arrived at using (such as a consensus technique; *cf.* Kitching *et al* 1998) to summarize the topology information shared between RACs. |
| Hierarchical history | Successive splitting and divergence of descendant lineages or separation of areas over time from a common ancestor or area |
| Hypothetical ancestor | Interpreting the node at which two or more taxa diverge (or arise) as being the ancestor of all the descendent taxa arising from that node |
| Ingroup | An assumed monophyletic group whose phylogenetic relationships are the focus of the study (*cf.* Outgroup) |
| Internal node | A point in a phylogenetic tree where two or more branches join (i.e., arise from; *cf.* Hypothetical ancestor) |
| Monophyletic group | See Clade. |
| Most recent common ancestor | The hypothesized ancestor at the node joining two or more descendent taxa on a phylogenetic tree |
| Optimal ancestral area(s) | The best assignment of ancestral areas under a particular optimality criterion derived by maximizing a function that defines how well data fit a particular hypothesis (e.g. a phylogenetic tree or area cladogram). |
| Optimization | Set of iterative passes from the root to the tip of the tree and vice versa, during which the values of the optimality criterion function are updated (*cf.* Optimal Ancestral Area). |
| Outgroup | One or more taxa phylogenetically outside the focal study group (*cf.* ingroup) used for comparative purposes |
| Paraphyletic | A group that includes a most recent common ancestor plus only some of its descendents |
| Parsimony | A general scientific criterion for choosing between competing hypotheses, based on the principle of minimizing the number of events or steps needed to explain the data (e.g. character states or biogeographic events in phylogenetic and biogeographic analysis respectively) |
| Pattern-based methods | See Cladistic biogeography |
| Phylogenetic systematics | A set of methods used to infer evolutionary relationships among |

| | |
|---|---|
| | taxa by grouping them hierarchically into nested sets, which are then interpreted through a phylogenetic tree |
| Phylogenetic tree | See cladogram |
| Phylogeography | The study of the geographical context of evolutionary or genealogical relationships within a species' population based upon principles of population genetic theory and phylogenetic systematics |
| Resolved area cladogram | One or more area cladogram(s) obtained for a monophyletic taxonomic group under certain assumptions; see table 1 |
| Root | Internal node that is the starting point or base of a cladogram or phylogenetic tree |
| Sister taxa | Two or more descendent lineages branching from a most recent common ancestor, and are thus nearest relatives of each other |
| Term | See Terminal branch |
| Terminal | See Terminal branch |
| Terminal branch (or just terminal or term) | The branch in a phylogenetic tree connected to one node and leading to a taxon or a tip on which a taxon is placed |
| Topology-based method | An approach in biogeography that uses phylogenetic tree topology, timing of lineage splitting and geologic events to qualitatively infer past dispersal and/or vicariance events |
| Topology | The pattern of branching in a phylogenetic tree or area cladogram |
| Vicariance | Fragmentation of a geographical area and therefore, the continuous distribution of a taxon by the formation of geographical barrier(s) |

## References

Ashton P S and Gunatilleke C V S 1987 New light on the plant geography of Ceylon. I. Historical plant geography; *J. Biogeog.* **14** 249–285

Austin J D, Lougheed S C, Moler P E and Boag P T 2003 Phylogenetics, zoogeography, and the role of dispersal and vicariance in the evolution of the *Rana catesbeiana* (Anura: Ranidae) species group; *Biol. J. Linn. Soc.* **80** 601–624

Bahir M M and Maduwage K P 2005 *Calotes desilvai,* a new agamid lizard from Morningside Forest, Sri Lanka; *Raffles Bull. Zool. (Suppl. No. 12)* 381–392

Bahir M M and Silva A 2005 *Otocryptis nigristigma*, a new species of agamid lizard from Sri Lanka; *Raffles Bull. Zool. (Suppl. No. 12)* 393–406

Bell C D and Donoghue M J 2003 Phylogeny and biogeography of Morinaceae (Dipsacales) based on nuclear and chloroplast DNA sequences; *Org. Divers. Evol.* **3** 227–237

Bossuyt F, Meegaskumbura M, Beenaerts N, Gower D J, Pethiyagoda R, Roelants K, Mannaert A, Wilkinson M, Bahir M M, Manamendra-Arachchi K, Ng P K L, Schneider C J, Oommen O V, and Milinkovitch M C 2004 Local endemism within the Western Ghats-Sri Lanka biodiversity hotspot; *Science* **306** 479–481

Bossuyt F and Milinkovitch M C 2001 Amphibians as indicators of early Tertiary "Out-of-India" dispersal of vertebrates; *Science* **292** 93–95

Brown J H and Lomolino M B 1998 *Biogeography* (Massachusetts: Sinauer Associates)

Cadle J E, Dessauer H C, Gans C and Gartside D F 1990 Phylogenetic relationships and molecular evolution in uropeltid snakes (Serpentes: Uropeltidae): allozymes and albumin immunology; *Biol. J. Linn. Soc.* **40** 293–320

Conti E, Eriksson T, Schönenberger J, Systema K J and Baum D A 2002 Early Tertiary Out-of-India dispersal of Crypteroniaceae: evidence from phylogeny and molecular dating; *Evolution* **56** 1931–1942

Cooray P G 1967 An introduction to the geology of Ceylon; *Spolia Zeylan.* **31** 1–324

Coyne J A and Orr H A 2004 *Speciation* (Massachusetts: Sinauer Associates)

Crisci J V, Katinas L and Posadas P 2003 *Historical biogeography: an introduction* (Cambridge: Harvard University Press)

Croizat L 1964 *Space, time, form: The biological synthesis* (Caracas)

Das I 1996 *Biogeography of the reptiles of South Asia* (Malabar: Krieger)

Dilger W C 1952 The Brij Hypothesis as an explanation for the tropical faunal similarities between the Western Ghats and the eastern Himalayas, Assam, Burma, and Malaya; *Evolution* **40** 977–996

Donoghue M J and Moore B R 2003 Toward an integrative historical biogeography; *Integr. Comp. Biol.* **43** 261–270

Erdelen W 1988 Population dynamics and dispersal in three species of agamid lizards in Sri Lanka: *Calotes calotes*, *C. versicolor* and *C. nigrilabris*; *J. Herpetol.* **22** 42–52

Erdelen W 1996 Biogegraphy and conservation of south Asian herpetofauna; in *Biology and conservation of the amphibians, reptiles and their habitats in South Asia* (ed.) A deSilva (Colombo: Amphibian and Research Organisation of Sri Lanka) pp 76–96

Erdelen W and Preu C 1990 Quarternary coastal and vegetation dynamics in the Palk strait region, south Asia – the evidence and hypotheses; in *Vegetation and erosion* (ed.) J B Thornes (London: John Wiley) pp 491–504

Fernando C H 1984 *Ecology and biogeography in Sri Lanka* (The Hague: Dr W Junk Publishers)

Gans C 1976 Aspects of the biology of uropeltid snakes; in *Morphology and biology of reptiles* (ed.) D Bellairs (London: Academic Press) pp 191–204

Gower D J, Kupfer A, Oommen O V, Himstedt W, Nussbaum R A, Loader S P, Presswell B, Müller H, Krishna S B, Boistel R and Wilkinson M 2002 A molecular phylogeny of ichthyophiid caecilians (Amphibia: Gymnophiona: Ichthyophiidae): out of India or out of South East Asia; *Proc. R. Soc. London B* **269** 1563–1569

Graham C H, Ferrier S, Huettman F, Moritz C and Peterson A T 2004a New developments in museum-based informatics and applications in biodiversity analysis; *Trends Ecol. Evol.* **19** 497–503

Graham C H, Ron S R, Santos J C, Schneider C J and Moritz C 2004b Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs; *Evolution* **58** 1781–1793

Harold A S and Mooi R D 1994 Areas of endemism: Definition and recognition criteria; *Syst. Biol.* **43** 261–266

Hennig W 1966 *Phylogenetic systematics* (Urbana: University of Illinois Press)

Hora S L 1949 Satpura hypothesis of the distribution of the malayan fauna and flora in peninsular india; *Proc. Natl. Inst. Sci. India* **15** 309–314

Karanth K P 2003 Evolution of disjunct distributions among wet-zone species of the Indian subcontinent: Testing various hypotheses using a phylogenetic approach; *Curr. Sci.* **85** 1276–1283

Kearey P and Vine F J 1996 *Global tectonics* (Boston: Blackwell Science)

Kitching I J, Forey P L, Humphries C J and Williams D M 1998 *Cladistics: the theory and practice of parsimony analysis* (New York: Oxford University Press)

Knowles L L 2004 The burgeoning field of statistical phylogeography; *J. Evol. Biol.* **17** 1–10

Linder H P 2001 On areas of endemism, with an example from the African Restionaceae; *Syst. Biol.* **50** 892–912

Lomolino M V and Heaney L R 2004 *Frontiers of biogeography: new directions in the geography of nature* (Massachusetts: Sinauer Associates)

Macey J R, Schulte II J A, Ananjeva N B, Larson A, Rastegar-Pouyani N, Shammakov S M and Papenfuss T J 1998 Phylogenetic relationships among agamid lizards of the *Laudakia caucasia* species group: testing hypotheses of biogeographic fragmentation and an area cladogram for the Iranian Plateau; *Mol. Phyl. Evol.* **10** 118–131

Macey J R, Schulte II J A, Larson A, Anajeva N B, Wang Y, Pethiyagoda R, Rastegar-Pouyani N and Papenfuss T J 2000 Evaluating Trans-Tethys migration: an example using acrodont lizard phylogenetics; *Mol. Phyl. Evol.* **49** 233–256

Manamendra-Arachchi K and Liyanage S 1994 Conservation and distribution of the agamid lizards of Sri Lanka with illustrations of the extant species; *J. South Asian Nat. Hist.* **1** 77–96

Mani M S 1974 *Ecology and biogeography in India* (The Hague: Dr W Junk b v Publishers)

Mayr E 1963 *Animal species and evolution* (Cambridge: Belknap Press)

McDowall R M 2004 What biogeography is: a place for process; *J. Biogeogr.* **31** 345–351

Moritz C, Patton J L, Schneider C J and Smith T B 2000 Diversification of rainforest faunas: an integrated molecular approach; *Ann. Rev. Ecol. Syst.* **31** 533–563

Morley R J 2000 *Origin and evolution of tropical rain forests* (Chichester: John Wiley)

Morrone J J 1994 On the identification of areas of endemism; *Syst. Biol.* **43** 438–441

Nelson G J and Platnick N I 1981 *Systematics and biogeography: cladistics and vicariance* (New York : Columbia University Press)

Nielsen R and Wakeley J 2001 Distinguishing migration from isolation: a Markov chain Monte Carlo approach; *Genetics* **158** 885–896

Page R D M 1988 Quantitative cladistic biogeography: constructing and comparing area cladograms; *Syst. Zool.* **37** 254–270

Page R D M 1989 Comments on component-compatibility in historical biogeography; *Cladistics* **5** 167–182

Page R D M 1990 Component Analysis: A valiant failure?; *Cladistics* **6** 119–136

Page R D M 1994 Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas; *Syst. Biol.* **43** 58–77

Page R D M 2003 Introduction; in *Tangled trees: phylogeny, cospeciation, and coevolution* (ed.) R D M Page (Chicago: The University of Chicago Press) pp 1–21

Peterson A T, Soberón J and Sanchez-Cordero V 1999 Conservation of ecological niches in evolutionary time; *Science* **285** 1265–1267

Pethiyagoda R and Manamendra-Arachchi K 1998 A revision of the endemic Sri Lankan agamid lizard genus *Ceratophora* Gray, 1835, with description of two new species; *J. South Asian Nat.Hist.* **3** 1–50

Platnick N I 1981 Widespread taxa and biogeographic congrunce; in *Advances in Cladistics* (eds) V A Funk and D R Brooks (New York: New York Botanical Gardens) pp 223–227

Platnick N I and Nelson G 1978 A method of analysis for historical biogeography; *Syst. Zool.* **27** 1–16

Randhawa M S 1945 Progressive dessication of northern India; *J. Bom. Nat. Hist. Soc.* **45** 558–565

Rannala B and Michalakis Y 2003 Population genetics and cospeciation: from process to pattern; in *Tangled trees: phylogeny, cospeciation, and coevolution* (ed.) R D M Page (Chicago: The University of Chicago Press) pp 120–143

Roelants K, Jiang J and Bossuyt F 2004 Endemic ranid (Amphibia: Anura) genera in southern mountain ranges of the Indian subcontinent represent ancient frog lineages: evidence from molecular data; *Mol. Phyl. Evol.* **31** 730–740

Ronquist F 1996a *DIVA* [Version: 1.1 ] Uppsala University, Uppsala. http://www.ebc.uu.se/systzoo/research/diva/diva.html

Ronquist F 1996b Reconstructing the history of host-parasite associations using generalised parsimony; *Cladistics* **11** 73–89

Ronquist F 1997a Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography; *Syst. Biol.* **46** 195–203

Ronquist F 1997b Phylogenetic approaches in coevolution and biogeography; *Zool. Scripta* **26** 313–322

Ronquist F 1998 Three-dimensional cost-matrix optimization and maximum cospeciation; *Cladistics* **14** 167–172

Ronquist F 2002 *TreeFitter* [Version: 1.0] Uppsala University, Uppsala. http://www.ebc.uu.se/systzoo/research/treefitter/treefitter.html

Ronquist F 2003 Parsimony analysis of coevolving species associations; in *Tangled trees: phylogeny, cospeciation, and coevolution* (ed.) R D M Page (Chicago: The University of Chicago Press) pp 22–64

Ronquist F and Nylin S 1990 Process and pattern in the evolution of species associations; *Syst. Zool.* **39** 323–344

Rosen D E 1978 Vicariant patterns and historical explanation in biogeography; *Syst. Zool.* **27** 159–188

Sanmartín I, Enghoff H and Ronquist F 2001 Patterns of animal dispersal, vicariance and diversification in the Holartic; *Biol. J. Linn. Soc.* **73** 345–390

Sanmartín I and Ronquist F 2002 New solutions to old problems: widespread taxa, redundant distributions and missing areas in event-based biogeography; *Anim. Biodiv. Conserv.* **25** 75–93

Sanmartín I and Ronquist F 2004 Southern Hemisphere biogeography inferred by event-based models: Plant versus animal patterns; *Syst. Biol.* **53** 216–243

Schulte II J A, Macey J R, Pethiyagoda R and Larson A 2002 Rostral horn evolution among agamid lizards of the genus *Ceratophora* endemic to Sri Lanka; *Mol. Phyl. Evol.* **22** 111–117

Schulte J A, Vindum J V, Win H, Thin T, Lwin K S and Shein A K 2004 Phylogenetic relationships of the genus *Ptyctolaemus* (Squamata: Agamidae), with a description of a new species from the Chin Hills of western Myanmar; *Proc. Cal. Acad. Sci.* **55** 222–245

Senanayake F R and Moyle P B 1982 Conservation of freshwater fishes of Sri Lanka; *Biol. Conserv.* **22** 181–195

Senanayake F R, Soulé M E, and Senner J W 1977 Habitat values and endemicity in the vanishing rain forests of Sri Lanka; *Nature (London)* **265** 351–354

Swan L W 1993 The Satpura Hypotheses: a biogeographical challenge to geology; *J. Bom. Nat. Hist. Soc.* **90** 141–157

Templeton A R 2004 Statistical phylogeography: methods of evaluating and minimizing inference errors; *Mol. Ecol.* **13** 789–809

Thornton I W B, Cook S, Edwards J S, Harrison R D, Schipper C, Shanahan M, Singadan R and Yamuna R 2001 Colonization of an island volcano, Long Island, Papua New Guinea, and an emergent island, Motmot, in its caldera lake. VII. Overview and discussion; *J. Biogeogr.* **28** 1389–1408

Uetz P 2004 *The EMBL reptile database* (http://www.embl-heidelberg.de/~uetz/)

Van Veller M G P and Brooks D R 2001 When simplicity is not parsimonious: a priori and a posteriori methods in historical biogeography; *J. Biogeogr.* **28** 1–11

Van Veller M G P, Brooks D R and Zandee M 2003 Cladistic and phylogenetic biogeography: the art and the science of discovery; *J. Biogeogr.* **30** 319–329

Van Veller M G P, Kornet D J and Zandee M 2000 Methods in vicariance biogeography: Assessment of the implementations of assumptions 0, 1, and 2; *Cladistics* **16** 319–345

Van Veller M G P, Kornet D J and Zandee M 2002 A posteriori and a priori methodologies for testing hypotheses of causal processes in vicariance biogeography; *Cladistics* **18** 207–217

Van Veller M G P, Zandee M and Kornet D J 2001 Measures for obtaining inclusive sets of area cladograms under assumptions zero, 1, and 2 with different methods for vicariance biogeography; *Cladistics* **17** 248–259

Van Veller M G P, Zandee M and Kornet D J 1999 Two requirements for obtaining valid common patterns under different assumptions in vicariance biogeography; *Cladistics* **15** 393–406

Voelker G 1999 Dispersal, vicariance, and clocks: historical biogeography and speciation in a cosmopolitan passerine Genus (*Anthus*: Motacillidae); *Evolution* **53** 1536–1552

Wakeley J 2004 Inferences about the structure and history of populations: coalescents and intraspecific phylogeography; in *The evolution of population biology* (eds) R S Singh and M K Uyenoyama (Cambridge, UK: Cambridge University Press) pp 193–215

Zandee M and Roos M C 1987 Component-compatibility in historical biogeography; *Cladistics* **3** 305–332

*MS received 9 May 2005; accepted 4 January 2006*

*e*Publication: 18 February 2006

Corresponding editor: RENEE M BORGES