ORIGINAL ARTICLE

Parsimony analysis of endemicity describes but does not explain: an illustrated critique

Ivonne J. Garzón-Orduña[1,2], Daniel Rafael Miranda-Esquivel[2]* and Mariano Donato[3]

[1]*Department of Biological Sciences, University of New Orleans, New Orleans, LA, USA,* [2]*Laboratorio de Sistemática & Biogeografía (LSB), Escuela de Biología, Universidad Industrial de Santander, Bucaramanga, Colombia and* [3]*Laboratorio de Sistemática y Biología Evolutiva (LASBE), Museo de La Plata, La Plata, Argentina*

ABSTRACT

Aim To demonstrate that parsimony analysis of endemicity (PAE) is not analogous to a cladistic biogeographical analysis.

Location We used six data sets from previously published studies from around the world.

Methods In order to test the efficiency of PAE in recovering historical relationships among areas, we performed an empirical comparison of nodes recovered with PAE, primary Brooks parsimony analysis (BPA), and an event-based method using three models (maximum codivergence, reconciled trees, and the default model of the TREEFITTER program) for six data sets. We measured the performance of PAE in recovering historical area relationships by counting the number and examining the content of nodes recovered by PAE and by historical methods. The dispersal/vicariance ratio was calculated to assess the prevalence of dispersal or vicariance in each reconstruction and its relationship to the performance of PAE.

Results Our results show that PAE recovers an average of 17.25% of historical nodes. PAE and BPA tend to provide similar results; however, in relation to the event-based models, PAE performance was poor under all the tested scenarios. Although in some cases PAE reconstructions are more resolved than historical reconstructions, this does not necessarily mean that PAE produces more informative answers. These additional nodes correspond to unsupported statements that are based solely on the distributional data of taxa and not on their phylogenetic history. In other words, these nodes were not found by the historical methods, which take phylogenetics into account. The number of historical nodes recovered using PAE was in general negatively correlated with the dispersal/vicariance ratio.

Main conclusions Our results show that PAE is unable to recover historical patterns and therefore does not fit into the current paradigm of historical biogeography. These findings raise doubts regarding conclusions derived from biogeographical studies that interpret PAE trees as area cladograms. We acknowledge that PAE aims to describe but does not explain the current distribution of organisms. It is therefore a useful tool in other biogeographical or ecological analyses for exploring the distribution of taxa or for establishing hypotheses of primary homology between areas.

Keywords
Cladistic biogeography, dispersal, event-based methods, historical biogeography, PAE, primary homology, vicariance.

*Correspondence: Daniel Rafael Miranda-Esquivel, Laboratorio de Sistemática & Biogeografía (LSB), Escuela de Biología, Universidad Industrial de Santander, A.A. 678. Bucaramanga, Colombia.
E-mail: dmiranda@uis.edu.co*

INTRODUCTION

The main goal of historical biogeography is to identify common distribution patterns between taxa from which a historical sequence of area relationships can be derived. Although all approaches in historical biogeography have the same ultimate goal, several methods have been used (Crisci, 2001), and various taxonomies have been proposed to classify these historical biogeographical methods. Under Ronquist's taxonomy, methods are divided as pattern-based or process-based (Ronquist & Nylin, 1990; Ronquist, 1997, 1998). Other schemes discriminate between a priori and a posteriori analyses (see van Veller et al., 2001), or between methods used to study the biogeographical history of a particular taxon, and those used to study the history of areas (Hovenkamp, 1997). In general, a method can be regarded as 'historical' if it uses phylogenetic information on the taxa involved or if it produces historical answers concerning the ancient configuration of the area. In the latter case, however, the answers are considered historical regardless of the source of data (for example topologies based on distributional data and interpreted as area cladograms). For the purposes of this paper, we consider historical biogeography to be a discipline that rests on the theoretical principles and application of phylogenetic systematics, and thus relies on the use of phylogenies to produce historical reconstructions of area relationships.

Morrone & Crisci (1995) described cladistic biogeography as a historical biogeographical approach that derives a general area cladogram based on the assumption that there is a correspondence between phylogenetic and area relationships. Therefore, historical biogeography relies on cladogenetic events represented by the nodes of a tree and on the accuracy of the distributional data. A general biogeographical pattern must be inferred from the congruence between patterns obtained for each monophyletic group (a node). Methods such as reconciled trees (Page, 1994), primary Brooks parsimony analysis (Wiley, 1987, 1988), component analysis (Nelson & Platnick, 1981), three-area statements (Nelson & Ladiges, 1991), and paralogy-free subtree (Nelson & Ladiges, 1996) fit this approach.

Parsimony analysis of endemicity (PAE; Rosen, 1988; Rosen & Smith, 1988) aims to search for similarities between areas based on the taxa that occur there. PAE is based on an 'area by taxa' matrix in which the presence of a given taxon in an area is coded as 1 and its absence as 0. An additional row with all zero observations is included to 'root' the final topology. The most parsimonious solution for this matrix is optimized with any of the regular software packages used in a cladistic analysis. In this way, PAE classifies localities, areas, or quadrats according to their shared taxa by means of the most parsimonious topology, resulting in a hierarchical classification of geographical units (Rosen & Smith, 1988; Crisci et al., 2000, 2003). The matrix used in PAE is similar to those used by other methods based on Assumption 0 (e.g. Brooks parsimony analysis), and therefore shares a similar modus operandi: areas inhabited by one widespread taxon are considered to be sister areas, and the missing areas are treated as 'absent' (Zandee & Ross, 1987). Rosen (1988) noted that, as in cladistic biogeography, PAE generates historical (and hence geological) hypotheses from present-day distributions. However, he considered two additional issues. First, PAE only uses distributional data, whereas cladistic biogeography uses 'taxonomic characters to fingerprint areas' (Rosen, 1988, p. 457); and second, as PAE only uses contemporary distributions, the meaning and scope of the results are not clear when there are taxa from different strata at the same locality. He also pointed out that the analogy between distribution data and characters has to be considered, and therefore the theoretical basis for the historical inferences obtained from PAE has yet to be developed satisfactorily.

Craw (1988) proposed a modification of Rosen's (1988) method using areas as study units and adding taxonomic information from monophyletic clades. Cracraft (1991) also proposed a version of PAE based on areas of endemism, similar to Craw's, and recently revamped it in the form of cladistic analysis of distributions and endemism (CADE) in Porzecanski & Cracraft (2005). The similarity between these methods rests on the fact that both cluster taxa at a supra-specific rank, and therefore a recoding of Craw's multi-state characters as binary makes the matrices equivalent (Crisci et al., 2000, 2003). This addition of taxonomic information on supra-specific taxa is considered a misuse of primary Brooks parsimony analysis (Crisci et al., 2000, 2003). We remark that Cracraft (1991) stressed that PAE 'does not constitute a general method of biogeographic analysis; it cannot be informative about area relationships if dispersal and plesiomorphic distributions contribute significant biogeographic noise' (Cracraft, 1991, p. 212).

Morrone (1994a) proposed another application of PAE. The goal of his method was to delimit areas of endemism. To achieve this, the study area is divided into quadrants and a data matrix is constructed and analysed following the original rules of PAE (Rosen, 1988). Based on the concept presented by Platnick (1991), an area of endemism is defined as those quadrants that form a monophyletic group supported by at least two taxa (namely distributional congruence). PAE has been one of the most frequently used methods to define areas of endemism. However, its use has now been extended to reconstruct 'area cladograms' and infer vicariance events (see Table 1). This view has been so widely accepted and shared that PAE is considered a 'cladistic parsimony method of biogeography' (Ron, 2000; contra Crisci, 2001). Indeed, da Silva & Oren (1996) and De Grave (2001) both consider PAE analogous to a cladistic analysis.

Phylogenetic information has played a central role in historical biogeography (Nelson & Platnick, 1981). As PAE does not use phylogenetic information, the historical content of its results may be far from the pattern revealed by a cladistic method. A number of authors have contested the application of PAE as a historical method (Morrone & Crisci, 1995; Humphries & Parenti, 1999; Posadas & Miranda-Esquivel, 1999; García-Barros et al., 2002; Santos, 2005). The criticisms have also highlighted the fact that PAE has phenetic characteristics

Table 1 Major uses of PAE in biogeography published in recent years

PAE as	Author(s)
Cladogram of area relationships	Aguilar-Aguilar *et al.* (2003), Bates *et al.* (1998), Bellan & Bellan-Santini (1997), Bisconti *et al.* (2001), Costa *et al.* (2000), Crisci *et al.* (2001), da Silva & Oren (1996), De Grave (2001), Emerson *et al.* (1997), Fernandes *et al.* (1995), García-Barros *et al.* (2002), Geraads (1998), Glasby & Álvarez (1999), Goldani & Gervasio (2003), Katinas *et al.* (2004), Luna-Vega *et al.* (2001), Morrone & Coscarón (1996), Morrone & Escalante (2002), Morrone & Lopretto (1995), Morrone (1994b, 1998), Morrone *et al.* (1997), Myers (1991), Posadas (1996), Posadas *et al.* (1997), Racheli & Racheli (2003, 2004), Ron (2000), Sfenthourakis & Giokas (1998), Trejo-Torres & Ackerman (2001), Watanabe (1998)
Tool to discover primary homology	Luna-Vega & Alcántara Ayala (2001), Morrone & Márquez (2001), Mota *et al.* (2002), Posadas *et al.* (1997)
Tool to identify areas of endemism	Caviares *et al.* (2001), Espinosa-Organista *et al.* (2000), García-Barros *et al.* (2002), Ippi & Flores (2001), Linder (2001), Morrone (1994a, 1998), Morrone *et al.* (1999), Posadas (1996), Posadas *et al.* (1997)

analogous to those in distance methods. Other critical problems are the lack of quantitative and qualitative phylogenetic information and the presumption that vicariance is the only type of event to affect the pattern. Based on these criticisms, the results of PAE should be considered as a raw approximation to the historical relationships between areas (e.g. a hypothesis of primary homology), or these relationships would have to be interpreted on grounds other than historical ones.

The recent application of PAE to reconstructing historical area relationships raises two questions: is phylogenetic information necessary to reconstruct historical relationships, or is the congruence of distributional records sufficient evidence to establish a claim of historical connections between areas? Given the implications of these questions, it is first necessary to verify empirically if PAE can produce the same results as historical methods. Only if this were confirmed would it be acceptable to use PAE and therefore ignore phylogenies, dispersal and extinctions. Therefore, the aim of this paper is to assess the ability of PAE to reconstruct historical patterns among areas.

METHODS

Comparison of methods

To compare the performance of PAE with that of historical methods, we evaluated the number and content of nodes recovered by PAE vs. the number and content of those recovered by historical methods. Since we consider biogeographical methods using a phylogenetic hypothesis to be more likely to recover historical statements about area relationships, we used the topology obtained from the historical methods as the reference topology. Historical biogeographical reconstructions were made using primary Brooks parsimony analysis (hereafter BPA) (Wiley, 1987, 1988), and an event-based analysis using three distinct models of cost (Ronquist, 2002a). We wanted to verify whether historical statements (in the form of nodes) could be recovered using PAE. The use of six distinct data sets allowed us to set up a variety of biogeographical scenarios produced under different assumptions, and we were

therefore able to test PAE in a broad context. Fig. 1 compares the methodological paths of PAE, BPA, and the event-based method.

Data sets and analyses

We used six data sets – five from previous published studies and one from an unpublished study. The requirements used by each author to choose the phylogenies can be found in the respective source references of the data sets. All included phylogenies corresponded to monophyletic groups, and the areas used corresponded to previously defined areas of endemism. Two data sets include 73 phylogenies corresponding to the Holarctic region based on non-marine animals (hereafter Family and Genera data sets) (Enghoff, 1996); two others correspond to the world-wide distribution of animals and plants and include 73 phylogenies (hereafter Plants and Animals data sets) (Sanmartín & Ronquist, 2004); another data set includes nine phylogenies corresponding to Neotropical *Simulium* subgenera (hereafter *Simulium* data set) (Miranda-Esquivel, 2001); and the final data set corresponds to phylogenetic and distributional information of 58 Amazonian animal taxa, including birds, insects, mammals, squamates, and anurans (hereafter Amazonia data set). The general area cladogram for the Amazonian data set was reconstructed using Haffer's areas of endemism (Haffer, 1969).

The data sets as a whole comprise 213 phylogenies plus the corresponding distributional information for each taxon. This sampling allowed us to achieve a scenario with different levels of biogeographical noise, beyond Rosen's (1978) classical *Xiphophorus* and *Heterandria* data set. This is in agreement with Morrone & Carpenter's (1994) view, as reconstructions of clean data sets will produce almost the same answer regardless of the method used.

Using the dispersal–vicariance analysis (DIVA; Ronquist, 1996, 1997), we calculated, for each data set, the frequency of dispersal and vicariance events suggested by the reconstructions. In the results section we show and refer to these values as a ratio (dispersal/vicariance). DIVA is an event-based method that optimizes a multi-dimensional matrix of costs and allows
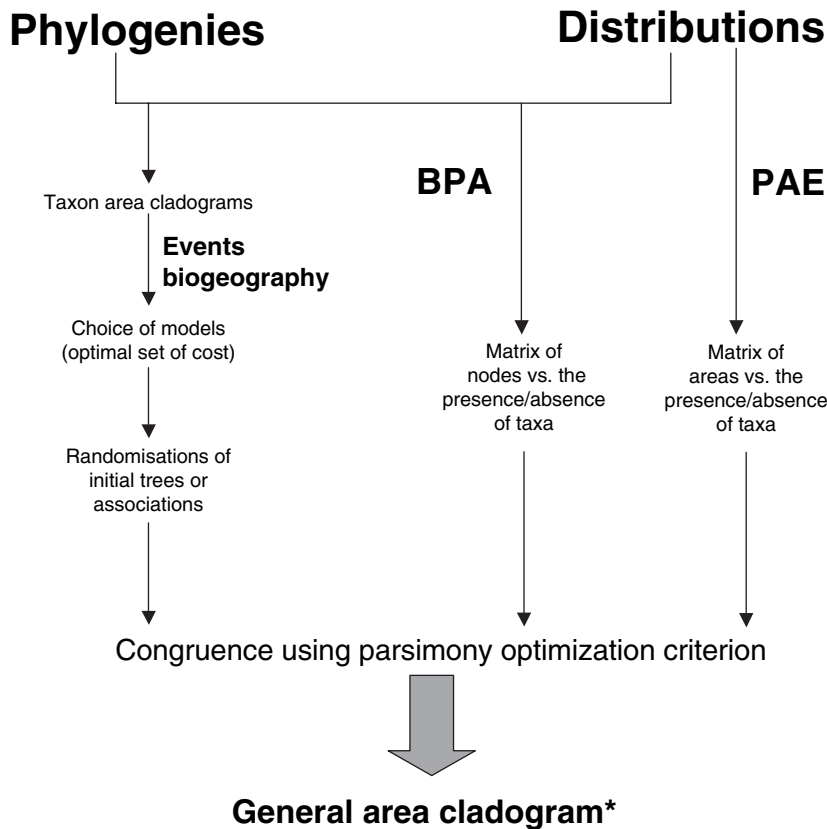
Phylogenies

Distributions

Taxon area cladograms

BPA

PAE

Events biogeography

Choice of models
(optimal set of cost)

Matrix of
nodes vs. the
presence/absence
of taxa

Matrix of
areas vs. the
presence/absence
of taxa

Randomisations of
initial trees or
associations

Congruence using parsimony optimization criterion

General area cladogram*

Figure 1 Flow chart showing the methodology followed by PAE, BPA, and event biogeography methods. *The topology obtained with PAE is not considered an area cladogram, as it is based only on the congruence of distributional data.

the description of non-hierarchical patterns (for further details see Ronquist, 1997). We then performed five analyses for each data set:

1. PAE;

2. BPA;

3. an event-based analysis using three distinct cost models to reconstruct a general area cladogram:

the maximum codivergence vicariance (MC) model;

the reconciled trees (REC) model;

the default model of the TREEFITTER software (RONQ).

The use of an explicit cost model in this empirical comparison allows assessments of PAE performance under different biogeographical scenarios. Whereas the MC model produces a reconstruction that maximizes the number of vicariance events, the REC model prohibits dispersal, and the RONQ model considers differential costs for all events. In the case of the RONQ model, the cost attributable to each event was assigned by Ronquist (2002a), who searched for a combination capable of producing a reconstruction constrained by the phylogenetic hypotheses of the taxa.

PAE and BPA were performed using NONA (Goloboff, 1998). TREEFITTER (Ronquist, 2002b) was used to produce the reconstructions of the event models (MC, REC, and RONQ). Comparisons were performed for each data set by counting the number of nodes shared between PAE trees and each of the historical methods. This value was then scaled with respect to the number of nodes recovered by the various historical

methods (BPA, MC, REC, RONQ) and by the number of nodes in PAE. We used strict consensus trees when three or more reconstructions were produced by any of the methods; otherwise, we used the single tree or the two reconstructions obtained.

The indexes used were as follows (see Figs 2 and 3).

1. The historical match index (Hmatch), defined as the ratio ESN/NCM, where ESN is the number of exact nodes shared between PAE and each of the historical methods, and NCM is the number of nodes present in the area cladogram obtained using the historical method (BPA; the event models MC, REC, or RONQ). The ratio ESN/NCM can be regarded as a historical match index, as it accounts for the number of 'historical' nodes recovered by PAE. A historical match index equal to 1.0 (or 100%) implies that PAE recovered all the nodes present in a given historical method (but not necessarily that all nodes in PAE are historical; see below), whereas a value of 0 means that PAE and the historical method do not share any nodes.

2. The PAE match index (Pmatch), defined as the ratio ESN/NPAE, where ESN is the number of exact nodes shared between PAE and each of the historical methods and NPAE is the number of nodes present in the PAE topology. This ratio can be considered a PAE match index, as it accounts for the number of PAE nodes that are historical. A Pmatch index equal to 1 (or 100%) implies that all nodes present in the PAE topology are historical, whereas a Pmatch index of 0 implies that none of the PAE nodes is historical.
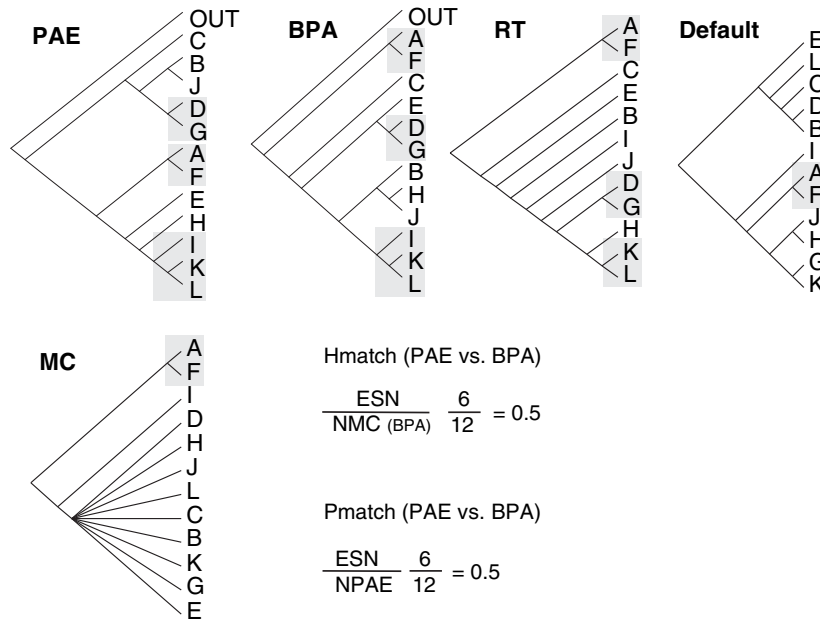
Figure 2 Area cladograms obtained for the Animals data set. Note that the PAE topology shares some nodes with the historical methods (shaded). This data set had the lowest dispersal/vicariance ratio (0.27). Labels used: PAE, PAE topology; BPA, BPA cladogram; RT, cladogram from reconciled trees; DEFAULT, cladogram from the default model; MC, cladogram from the maximal codivergence model. The historical match index (Hmatch) is defined as the ratio between the number of exact nodes shared between PAE and each of the historical methods (ESN), and the number of nodes present in the area cladogram (NMC) obtained using the historical method (BPA; the event models MC, REC, or RONQ). The PAE match index (Pmatch) is defined as the ratio between the number of exact nodes shared between PAE and each of the historical methods (ESN), and the number of nodes present in the PAE topology (NPAE).

We evaluated whether the Pmatch, Hmatch and the dispersal/vicariance ratio were correlated with the recovery of historical nodes by PAE, using Pearson's correlation coefficient. Finally, we calculated the resolution of the area cladograms obtained by each method as the number of nodes present in the cladogram scaled by the number of nodes present in the PAE topology. This value was calculated for each data set.

RESULTS

Resolution and dispersal–vicariance analysis

For almost all data sets, the resolution of PAE was equal to or lower than the resolution of the historical methods, and only in two instances did the topology of PAE have one or two additional nodes (in the Amazonia and Family data sets)
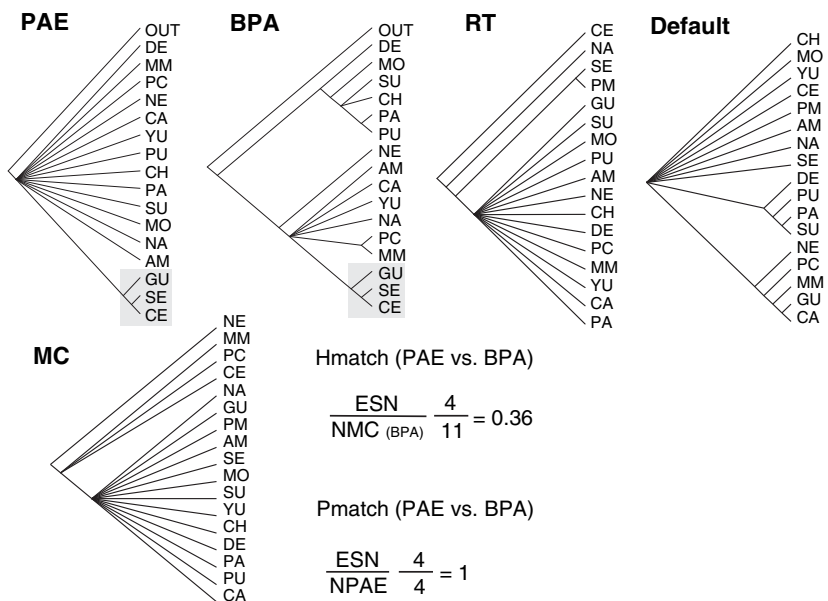

Figure 3 Area cladograms obtained for the *Simulium* data set. Using this data set, which had the highest dispersal/vicariance ratio (0.96), the agreement between PAE and historical methods decreased as dispersal increased. Labels, Hmatch and Pmatch as in Fig. 2.

	Data Set					
	Animals	Plants	Family	Genera	*Simulium*	Amazonia
Number of terminals	12	11	5	8	16	19
Number of nodes						
PAE topology	12*	4	5	4	4	1
BPA strict consensus tree	12*	11*	5*	4*	11*	1
REC strict consensus tree	11*	7	1	1	5	8
REC majority rule consensus	–	10	1	6	10	12
RONQ strict consensus tree	11*	9	4*	5	8	17
RONQ majority rule consensus	–	9	–	–	8	17
MC strict consensus tree	4	5	4	1	3	17
MC majority rule consensus	10	9	4	4	10	17
Relative resolution						
PAE	1	1	1	1	1	1
BPA	1	3	1	1	3	1
MC	0	1	1	0	1	17
REC	1	2	0	0	1	8
RONQ	1	2	1	0	2	17

Table 2 General information about the results of PAE and historical methods. The number of terminals, nodes and relative resolution are shown for each data set analysed under each method. Relative resolution was calculated using the strict consensus tree for each method scaled by the resolution of PAE (number of nodes). PAE, parsimony analysis of endemicity; BPA, Brooks parsimony analysis; REC, reconciled trees model; RONQ, default model; MC, maximum codivergence model

*A single tree was generated.

(Table 2). In general, the results showed that PAE and BPA tend to provide similar results, and for most of the data sets PAE and BPA produced trees with the same number of nodes (Animals, Family, Genera, and Amazonia). The dispersal/vicariance ratios recovered with DIVA for each data set are presented in Table 3. The Neotropical *Simulium* data set has the highest dispersal/vicariance ratio, which means that there was almost as much dispersal as vicariance. The Family and Genera data sets follow the *Simulium* data set, with similarly high dispersal/vicariance ratios. The Plants and Animals data sets have the lowest ratios, indicating a higher proportion of vicariance in relation to dispersal.

PAE vs. historical methods

Given that each data set provides a particular scenario because each represents the biogeographical history of a different area, below we present the results of the tree comparisons separately

Table 3 Results of all the analyses shown as the values obtained for historical match (Hmatch) and PAE match (Pmatch) indexes (shown as percentages). The number of phylogenies and the dispersal/vicariance ratio are also presented as general descriptors of each data set. The correlation between the dispersal/vicariance ratio and the number of nodes recovered was calculated using a Pearson correlation

	Data Set						Pearson's correlation to dispersal/vicariance ratio	Average performance
	Animals	Plants	Family	Genera	*Simulium*	Amazonia		
Number of phylogenies	54	19	35	38	9	58		
Dispersal/vicariance ratio	0.27	0.37	0.80	0.73	0.96	0.61		
Historical match (Hmatch)								
BPA.strict	50.00	27.27	100.00	100.00	36.36	100.00	0.31	68.94
REC.strict	27.27	0.00	0.00	0.00	0.00	0.00	−0.66	4.55
REC.majority	–	0.00	0.00	16.67	0.00	0.00	0.09	3.33
RONQ.strict	9.09	0.00	50.00	0.00	0.00	0.00	0.21	9.85
RONQ.majority	–	0.00	–	–	0.00	0.00	0.00	
MC.strict	25.00	20.00	0.00	0.00	0.00	0.00	−0.90	7.50
MC.majority	20.00	11.11	0.00	25.00	0.00	0.00	−0.49	9.35
Mean for historical recuperation								17.25
PAE match (Pmatch)								
BPA.strict	50.00	75.00	100.00	100.00	100.00	100.00	0.87	87.50
REC.strict	25.00	0.00	0.00	0.00	0.00	0.00	−0.66	4.17
REC.majority	–	0.00	0.00	25.00	0.00	0.00	0.09	5.00
RONQ.strict	8.33	0.00	40.00	0.00	0.00	0.00	0.19	8.06
RONQ.majority	–	0.00	0.00	–	0.00	0.00	0.00	
MC.strict	8.33	25.00	0.00	0.00	0.00	0.00	−0.70	5.56
MC.majority	16.67	25.00	0.00	25.00	0.00	0.00	−0.58	11.11
Mean for PAE recuperation								20.23

and then assess the congruence of PAE with respect to each of them.

Animals data set

Figure 2 shows the area cladograms obtained for the Animals data set using all methods, and Table 3 shows the Hmatch and Pmatch values as percentages. According to the Hmatch index, PAE performs best in relation to BPA. The Hmatch index shows that PAE recovered 50% of the historical nodes. However, the Pmatch index value shows that these historical nodes represented only 50% of the total PAE nodes. Both examples indicate that PAE may present an over-resolution of non-historical nodes, and that these nodes account for the majority of the nodes found by PAE. The worst performance of PAE was with the RONQ model, where PAE was able to recover only 9% (Hmatch) of the historical nodes (a single node); furthermore, the Pmatch index indicates that this node represented only 8% of PAE nodes. Once again, in this case, the topology from PAE is a well-resolved tree, but 91% of its nodes are non-historical.

Plants data set

The best performance of PAE for the Plants data set was obtained in relation to BPA and to the MC model (Table 3). PAE recovered 27% and 20% of the historical nodes reconstructed by each of these methods, respectively. The Pmatch index shows that the historical nodes represent 75% and 25% of the PAE nodes, respectively. However, the PAE topology had only four nodes and was, in this case, the least-resolved tree. The worst performance of PAE was with the RONQ and REC models, in which case PAE did not recover any of the nodes present in the historical reconstructions.

Family data set

The best performance of PAE for the family data set was with BPA; PAE recovered all (Hmatch 100%) nodes present in BPA, and, likewise, all the nodes present in the PAE topology were present in the BPA tree (Pmatch 100%) (Table 3). With respect to the RONQ model, PAE recovered 50% of the historical nodes; however, these nodes accounted for only 40% of the nodes in the PAE topology. Finally, PAE did not recover any of the nodes in the MC or the REC reconstruction.

Genera data set

As in the previous cases, for the Genera data set PAE performed well when compared with BPA, recovering all of the BPA nodes. PAE performed equally poorly with respect to the other methods (Table 3). PAE obtained relatively high values for the Hmatch and Pmatch indexes with the MC (Hmatch, 25%; Pmatch, 25%) and the REC (Hmatch 16.7%; Pmatch 25%) models. However, these values were obtained using majority consensus trees for the comparison; majority consensus trees were used instead of strict consensus because of the poor resolution of the historical trees when compared with the well-resolved PAE tree (Table 2). For this data set, PAE recovered only one node from each of these two historical methods, and none from the RONQ model.

Simulium data set

Table 3 shows that PAE recovered only 36% of the BPA historical nodes (Hmatch) for the *Simulium* data set and did not recover any nodes from the other historical methods. The area cladograms obtained with the various methods are shown in Fig. 3. The Pmatch index with respect to BPA showed that 100% of the nodes in the area cladogram reconstructed with PAE were historical; however, this value represents the only node present in the PAE topology.

Amazonia data set

Using the Amazonia data set, the PAE and BPA trees represented topologies with no resolution and had only one node (the root). The index values (Hmatch and Pmatch) of 100% are the result of the fact that PAE and BPA have one node and therefore share it. As a consequence of the lack of resolution for PAE, it performed very poorly with respect to the other historical methods, as it did not recover any of their nodes (Table 3).

Figures 2 and 3 show that the agreement between PAE and the historical methods decreased as dispersal increased (Table 3). The area cladograms obtained for the data set with the lowest dispersal/vicariance ratio (Animals) and for the data set with the largest dispersal/vicariance ratio (*Simulium*) are shown, with congruent nodes shaded.

DISCUSSION

The values of the Hmatch index show that PAE usually performs best in comparison with BPA, but not in all cases. The average performance (historical nodes recovered) of PAE in relation to BPA was not high (68.94%, Table 3) and was as low as 27.3% in the Plants data set. In general, the recovery by PAE of historical nodes was poor in the three scenario types we tested: those driven mainly by vicariance (Animals data set), those driven by dispersal (*Simulium* data set), and those with an almost equal frequency of vicariance and dispersal events (Family data set). This result shows that PAE will perform poorly in all possible scenarios for any data set. PAE performance decreases as the dispersal/vicariance ratio increases (Table 3) (Brooks & van Veller, 2003) or when the model involves more events. The mean Hmatch–Pmatch values are highest when PAE is compared with BPA: note that both methods are based on Assumption 0. In contrast, the values for Hmatch–Pmatch are low when the PAE results are compared with reconstructions based on other analytical models. Our results highlight the importance of evaluating the role of the dispersal/vicariance ratio as a measure of noise

(Miranda-Esquivel, 2001; Posadas & Morrone, 2001). This ratio is a characteristic of the data that is seldom explored in current biogeographical studies, in spite of previous research regarding the effects of dispersal in biogeographical reconstructions (Zink *et al.*, 2000; Miranda-Esquivel, 2001; Givnish *et al.*, 2004; Cook & Crisp, 2005; Heads, 2005).

PAE performance in relation to the models tested was poor, particularly so with respect to the REC model. Inasmuch as the REC model does not use dispersal events to produce a reconstruction, the disagreement between the two methods may indicate that the presence of dispersal has a critical effect on the biogeographical reconstructions. The poor performance of PAE compared with MC suggests that PAE is not able to recover all the vicariance events present in the various scenarios evaluated. The agreement between PAE and some cladistic methods depends on the properties of the data set used. In data sets with a strong vicariance structure and little dispersal, it is more plausible that PAE performs well and recovers many of the same nodes as a cladistic method. But, again, the performance of PAE worsens as dispersal increases (Fig. 2).

The main aim of historical biogeographical studies is to reveal and explain the origin of the biota and the relationships between areas (Andersson, 1996; Enghoff, 1996; Palmer & Cambefort, 2000; Donato *et al.*, 2003; Sanmartín & Ronquist, 2004). PAE is, however, unable to meet this goal using the methodological background of the early stages in the development of biogeography (e.g. vicariance-based). Moreover, since it uses only the distributional information of the organisms without considering their phylogenies, PAE can only be used to describe the distributional pattern of the taxa and the biotic similarities between areas. We believe that this conclusion is reasonably well supported by the results of our study. We consider PAE a helpful tool for exploring distributional data, and in this sense PAE could be seen as a heuristic tool to link biogeography and biodiversity analysis (Posadas & Miranda-Esquivel, 1999). For instance, parsimony analysis of locality data could be used to analyse the temporal and spatial variation of taxa in specific areas, or to identify patterns of ecological replacement or succession events (Pellens *et al.*, 2005; Wenzel & Luque, 2008). PAE assumes that all the areas sharing the presence of an organism are related (Assumption 0), and hence the treatment of widespread taxa becomes a critical issue in the performance of PAE. Indeed, this is the most problematic issue when PAE is interpreted as a historical method. On the basis that dispersal has been recorded many times as having a significant effect on reconstructions, the use of PAE to reconstruct historical relationships should be completely discarded. At this point in the development of biogeography, we consider Assumption 0 to be excessively unrealistic if it is based solely on distributional information. An obvious agreement between PAE trees and BPA area cladograms was evident throughout our analyses. This result was not unexpected, because PAE and BPA treat widespread taxa in the same way. However, the agreement was not complete, indicating that the information

from the phylogeny, which is expressed as a character in BPA, does have a strong bearing on the final result.

PAE might give a misleading sense of a 'good' solution because sometimes it will produce better resolved trees than other methods. Although the resolution achieved with PAE could be viewed as an ideal attribute for a historical method in biogeography, the analyses based on the historical match index (Hmatch) and the PAE match index (Pmatch) show that few of the PAE nodes could be considered historical. Given our results regarding the quality of the 'historical statements' produced by PAE, the extra resolution of PAE represents an undesirable feature and provides a false sense of accuracy. To conclude, we stress that PAE can be used in the same way as a panbiogeographic approach, to define a primary biogeographical homology (Morrone, 2001), but its use as a test or assessment of biogeographical history is misleading.

ACKNOWLEDGEMENTS

REFERENCES

Aguilar-Aguilar, R., Contreras-Medina, R. & Salgado-Maldonado, G. (2003) Parsimony Analysis of Endemicity (PAE) of Mexican hydrological basins based on helminth parasites of freshwater fishes. *Journal of Biogeography*, **30,** 1861–1872.

Andersson, L. (1996) An ontological dilemma: epistemology and methodology of historical biogeography. *Journal of Biogeography*, **23,** 269–277.

Bates, J.M., Hackett, S.J. & Cracraft, J. (1998) Area-relationships in the Neotropical lowlands: an hypothesis based on raw distributions of Passerine birds. *Journal of Biogeography*, **25,** 783–793.

Bellan, G. & Bellan-Santini, D. (1997) Utilizzazione delle analisi di parsimonia (cladistica) in sinecologia bentonica: Esempi in una zona inquinata. *Societá Italiana di Ecologia Atti*, **18,** 247–250.

Bisconti, M., Landini, W., Bianucci, G., Cantalamessa, G., Carnevale, G., Ragaini, L. & Valleri, G. (2001) Biogeographic relationships of the Galápagos terrestrial biota: parsimony analyses of endemicity based on reptiles, land birds and Scalesia land plants. *Journal of Biogeography*, **28,** 495–510.

Brooks, D.R. & van Veller, M.G.P. (2003) Critique of parsimony analysis of endemicity as a method of historical biogeography. *Journal of Biogeography*, **30,** 819–825.

Caviares, L., Mihoc, M., Marticorena, A., Marticorena, L., Matthei, O. & Squeo, F. (2001) Determinación de áreas

prioritarias para la conservación: Análisis de Parsimonia de Endemismos (PAE) en la flora de la IV Región de Coquimbo. *Libro Rojo de la Flora Nativa y de los Sitios Prioritarios para su Conservación: Región de Coquimbo* (ed. by F. Squeo, A.G. Arancio and J.R. Gutiérrez), pp. 159–170. Ediciones Universidad de La Serena, La Serena, Chile.

Cook, L.G. & Crisp, M.D. (2005) Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. *Journal of Biogeography*, **32**, 741–754.

Costa, L.P., Leite, Y.L.R., da Fonseca, G.A.B. & da Fonseca, M.T. (2000) Biogeography of South American forest mammals: endemism and diversity in the Atlantic forest. *Biotropica*, **32**, 872–881.

Cracraft, J. (1991) Patterns of diversification within continental biotas: hierarchical congruence among the areas of endemism of Australian vertebrates. *Australian Systematic Botany*, **4**, 211–227.

Craw, R.C. (1988) Continuing the synthesis between panbiogeography, phylogenetic systematics and geology as illustrated by the Chatham Islands. *Systematic Zoology*, **37**, 291–310.

Crisci, J.V. (2001) The voice of historical biogeography. *Journal of Biogeography*, **28**, 157–168.

Crisci, J.V., Katinas, L. & Posadas, P. (2000) *Introducción a la teoría y practica de la biogeografía histórica*. Sociedad Argentina de Botánica, Buenos Aries.

Crisci, J.V., Freire, S.E., Sancho, G. & Katinas, L. (2001) Historical biogeography of the Asteraceae from Tandilia and Ventania mountain ranges (Buenos Aires, Argentina). *Caldasia*, **23**, 21–41.

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

De Grave, S. (2001) Biogeography of Indo-Pacific Pontoniinae (Crustacea, Decapoda): a PAE Analysis. *Journal of Biogeography*, **28**, 1239–1253.

Donato, M., Posadas, P., Miranda-Esquivel, D.R., Ortiz Jaureguizar, E. & Cladera, G. (2003) Historical biogeography of the Andean region: evidence from *Listroderina* (Coleoptera: Curculionidae: Rhytirrhinini) in the context of the South American geobiotic scenario. *Biological Journal of the Linnean society*, **80**, 339–352.

Emerson, B.C., Wallis, G.P. & Patrick, B.H. (1997) Biogeographic area relationships in southern New Zealand: a cladistic analysis of Lepidoptera distributions. *Journal of Biogeography*, **24**, 89–99.

Enghoff, H. (1996) Historical biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. *Cladistics*, **11**, 223–263.

Espinosa-Organista, D., Morrone, J.J., Aguilar, C. & Llorente Bousquets, J. (2000) Regionalización biogeográfica de México: Provincias bióticas. *Biodiversidad, taxonomía y biogeografía de artrópodos de México: Hacia una síntesis de su conocimiento*, Vol. II (ed. by J. Llorente Bousquets, E. González and N. Papavero), pp. 61–94. Conabio, México.

Fernandes, M.E.B., da Silva, J.M. & Silva, J. Jr (1995) The monkeys of the islands of the Amazon estuary, Brazil: a biogeographic analysis. *Mammalia*, **59**, 213–221.

García-Barros, E., Gurrea, P., Luciáñez, M.J., Cano, J.M., Munguira, M.L., Moreno, J.C., Sainz, H., Sanz, M.J. & Simón, J.C. (2002) Parsimony analysis of endemicity and its application to animal and plant geographical distributions in the Ibero-Balearic region (western Mediterranean). *Journal of Biogeography*, **29**, 109–124.

Geraads, D. (1998) Biogeography of circum-Mediterranean Miocene-Pliocene rodents; a revision using factor analysis and parsimonious analysis of endemicity. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **137**, 273–288.

Givnish, T.J., Millam, K.C., Evans, T.M., Hall, J.C., Pires, J.C., Berry, P. & Sytsma, K.J. (2004) Ancient vicariance or recent long-distance dispersal? Inferences about phylogeny and South American–African disjunctions in Rapateaceae and Bromeliaceae based on ndhF sequence data. *International Journal of Plant Science*, **165**, S35–S54.

Glasby, C.J. & Álvarez, B. (1999) Distribution patterns and biogeographic analysis of Austral Polychaeta (Annelida). *Journal of Biogeography*, **26**, 507–533.

Goldani, A. & Carvalho, G.S. (2003) Parsimony analysis of endemicity of Neotropical spittlebugs (Hemiptera, Cercopidae). *Revista Brasileira de Entomologia*, **47**, 437–442.

Goloboff, P. (1998) *NONA, Version 2.0. Software and user manual*. Available at http://www.zmuc.dk/public/phylogeny/Nona-PeeWee.

Haffer, J. (1969) Speciation in Amazonian forest birds. *Science*, **165**, 131–137.

Heads, M.J. (2005) Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society*, **84**, 675–723.

Hovenkamp, P. (1997) Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics*, **13**, 67–79.

Humphries, C.J. & Parenti, L.R. (1999) *Cladistic biogeography*, 2nd edn. Oxford University Press, Oxford.

Ippi, S. & Flores, V. (2001) Las tortugas neotropicales y sus áreas de endemismo. *Acta Zoológica Mexicana*, **84**, 49–63.

Katinas, L., Crisci, J.V., Wagner, W. & Hoch, P.C. (2004) Geographic diversification of tribes Epilobieae, Gongylocarpeae, and Onagreae (Onagraceae) in North America, based on parsimony analysis of endemicity and track compatibility analysis. *Annals of the Missouri Botanical Garden*, **91**, 159–185.

Linder, H.P. (2001) On areas of endemism, with an example from the African Restionaceae. *Systematic Biology*, **50**, 892–912.

Luna-Vega, I. & Alcántara Ayala, O. (2001) Análisis de simplicidad de endemismos (PAE) para establecer un modelo de vicarianza preliminar del Bosque Mesófilo de Montaña Mexicano. *Introducción a la biogeografía en Latinoamerica: Teorías, conceptos y aplicaciones* (ed. by J.L. Bousquets and J.J. Morrone), pp. 273–277. Facultad de Ciencias, UNAM, Mexico.

Luna-Vega, I., Morrone, J.J., Alcántara Ayala, O. & Espinosa Organista, D. (2001) Biogeographical affinities among

Neotropical cloud forest. *Plant Systematics and Evolution*, **228**, 229–239.

Miranda-Esquivel, D.R. (2001) Efectos de la dispersión sobre la reconstrucción por árboles reconciliados y el patrón de distribución de los subgeneros Neotropicales de *Simulium* (Diptera: Simuliidae). *Caldasia*, **23**, 3–20.

Morrone, J.J. (1994a) On the identification of areas of endemism. *Systematic Biology*, **43**, 373–401.

Morrone, J.J. (1994b) Distributional patterns of species of Rhytirrhinini (Coleoptera: Curculionidae) and the historical relationships of the Andean provinces. *Global Ecology and Biogeography*, **4**, 188–194.

Morrone, J.J. (1998) On Udvardy's Insulantarctica province: a test from the weevils (Coleoptera: Curculionoidea). *Journal of Biogeography*, **25**, 947–955.

Morrone, J.J. (2001) Homology, biogeography and areas of endemism. *Diversity and Distributions*, **7**, 1–3.

Morrone, J.J. & Carpenter, J.M. (1994) In search of a method for cladistic biogeography: an empirical comparison of component analysis, Brooks parsimony analysis, and three-area-statements. *Cladistics*, **10**, 99–153.

Morrone, J.J. & Coscarón, M. del C. (1996) Distributional patterns of the American Peiratinae (Heteroptera: Reduviidae). *Zoologische Mededelingen (Leiden)*, **70**, 1–15.

Morrone, J.J. & Crisci, J.V. (1995) Historical biogeography: introduction to methods. *Annual Review of Ecology and Systematics*, **26**, 373–401.

Morrone, J.J. & Escalante, T. (2002) Parsimony analysis of endemicity (PAE) of Mexican terrestrial mammals at different area units: when size matters. *Journal of Biogeography*, **29**, 1095–1104.

Morrone, J.J. & Lopretto, E.C. (1995) Parsimony analysis of endemicity of freshwater Decapoda (Crustacea: Malacostraca) from southern South America. *Neotropica*, **41**, 3–8.

Morrone, J.J. & Márquez, J. (2001) Halffter's Mexican Transition Zone, beetle generalized tracks, and geographical homology. *Journal of Biogeography*, **28**, 635–650.

Morrone, J.J., Katinas, L. & Crisci, J.V. (1997) A cladistic biogeographic analysis of Central Chile. *Journal of Comparative Biology*, **2**, 25–42.

Morrone, J.J., Espinosa Organista, D., Aguilar, C. & Llorente Busquets, J. (1999) Preliminary classification of the Mexican biogeographic provinces: a parsimony analysis of endemicity based on plant, insect, and bird taxa. *Southwestern Naturalist*, **44**, 507–514.

Mota, J.F., Pérez-García, F.J., Jiménez, M.L., Amate, J.J. & Peñas, J. (2002) Phytogeographical relationships among high mountain areas in the Baetic Ranges (South Spain). *Global Ecology and Biogeography*, **11**, 497–504.

Myers, A.A. (1991) How did Hawaii accumulate its biota? A test from the Amphipoda. *Global Ecology and Biogeography Letters*, **1**, 24–29.

Nelson, G. & Ladiges, P.Y. (1991) Three-area statements: standard assumptions for biogeographic analysis. *Systematic Zoology*, **40**, 470–485.

Nelson, G. & Ladiges, P.Y. (1996) Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *American Museum Novitates*, **3167**, 1–58.

Nelson, G. & Platnick, N. (1981) *Systematics and biogeography: cladistics and vicariance.* Columbia University Press, New York.

Page, R.D.M. (1994) Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Systematic Biology*, **43**, 58–77.

Palmer, M. & Cambefort, Y. (2000) Evidence for reticulate palaeogeography: beetle diversity linked to connection-disjunction cycles of the Gibraltar strait. *Journal of Biogeography*, **27**, 403–416.

Pellens, R., Grandcolas, P. & Guilbert, E. (2005) Phylogenetic algorithms and the evolution of species communities in fragmented forests. *Cladistics*, **21**, 8–14.

Platnick, N.I. (1991) On areas of endemism. *Australian Systematic Botany*, **4**, 9–10.

Porzecanski, A.L. & Cracraft, J. (2005) Cladistic analysis of distributions and endemism (CADE): using raw distributions of birds to unravel the biogeography of the South American aridlands. *Journal of Biogeography*, **32**, 261–275.

Posadas, P. (1996) Distributional patterns of vascular plants in Tierra de Fuego: a study applying parsimony analysis of endemicity (PAE). *Biogeographica*, **72**, 161–177.

Posadas, P. & Miranda-Esquivel, D.R. (1999) El PAE (Parsimony Analysis of Endemicity) como una herramienta en la evaluación de la biodiversidad. *Revista Chilena de Historia Natural*, **72**, 539–546.

Posadas, P. & Morrone, J.J. (2001) Biogeografía cladística de la subregión subantártica: un análisis basado en taxones de la familia Curculionidae (Insecta: Coleoptera). *Introducción a la biogeografía en Latinoamérica: Teorías, conceptos y aplicaciones* (ed. by J.L. Bousquets and J.J. Morrone), pp. 267–272. Facultad de Ciencias, UNAM, Mexico.

Posadas, P.E., Estevez, J.M. & Morrone, J.J. (1997) Distributional patterns and endemism areas of vascular plants in the Andean subregion. *Fontqueria*, **48**, 1–10.

Racheli, L. & Racheli, T. (2003) Historical relationship of Amazonian areas of endemism based on raw distributions of parrots (Psittacidae). *Tropical Zoology*, **16**, 33–46.

Racheli, L. & Racheli, T. (2004) Patterns of Amazonian area relationships based on raw distributions of papilionid butterflies (Lepidoptera: Papilionidae). *Biological Journal of the Linnean Society*, **82**, 345–357.

Ron, S. (2000) Biogeographic area relationships of lowland Neotropical rainforest based on raw distributions of vertebrate groups. *Biological Journal of the Linnean Society*, **71**, 379–402.

Ronquist, F. (1996) *DIVA. Version 1.2. Software and user manual.* Available at http://www.ebc.uu.se/systzoo/research/diva/diva.html

Ronquist, F. (1997) Dispersal–vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**, 195–203.

Ronquist, F. (1998) Three-dimensional cost matrix optimization and maximum cospeciation. *Cladistics*, **14,** 167–172.

Ronquist, F. (2002a) Parsimony analysis of coevolving species associations. *Cospeciation* (ed. by R.D.M. Page), pp. 22–64. Chicago University Press, Chicago, IL.

Ronquist, F. (2002b) *TREEFITTER version 1.3b. Software and User Manual.* Available via http://www.ebc.uu.se/systzoo/research/treefitter/treefitter.html.

Ronquist, F. & Nylin, S. (1990) Process and pattern in the evolution of species associations. *Systematic Zoology*, **39,** 323–344.

Rosen, D.E. (1978) Vicariant patterns and historical explanation in biogeography. *Systematic Zoology*, **27,** 159–188.

Rosen, B.R. (1988) From fossils to Earth history: applied historical biogeography. *Analytical biogeography: an integrated approach to the study of animal and plant distributions* (ed. by A.A. Myers and P.S. Giller), pp. 437–481. Chapman & Hall, London.

Rosen, B.R. & Smith, A.B. (1988) Tectonics from fossils? Analysis of reef-coral and sea-urchin distributions from late Cretaceous to Recent using a new method *Geological Society, London, Special Publications*, **37,** 275–306.

Sanmartín, I. & Ronquist, F. (2004) Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology*, **53,** 216–243.

Santos, C.M.D. (2005) Parsimony analysis of endemicity: time for an epitaph? *Journal of Biogeography*, **32,** 1284–1286.

Sfenthourakis, S. & Giokas, S. (1998) A biogeographical analysis of Greek Oniscidean endemism. *Israel Journal of Zoology*, **44,** 273–282.

da Silva, J.M.C. & Oren, D.C. (1996) Application of parsimony analysis of endemicity in Amazonian biogeography: an example with primates. *Biological Journal of the Linnean Society*, **39,** 427–437.

Trejo-Torres, J.C. & Ackerman, J.D. (2001) Biogeography of the Antilles based on a parsimony analysis of orchid distributions. *Journal of Biogeography*, **28,** 775–794.

van Veller, M.G.P., Kornet, D.J. & Zandee, M. (2001) A posteriori and a priori methodologies for testing hypotheses of causal processes in vicariance biogeography. *Cladistics*, **18,** 207–217.

Watanabe, K. (1998) Parsimony analysis of the distribution patterns of Japanese primary freshwater fishes, and its application to the distribution of the bagrid catfishes. *Ichthyology Research*, **45,** 259–270.

Wenzel, J.W. & Luque, G.M. (2008) Parsimony analysis of ecological successions, a powerful tool for interpreting changes in community structure. *Cladistics*, in press.

Wiley, E.O. (1987) Methods in vicariance biogeography. *Systematics and evolution: a matter of diversity* (ed. by P. Hovenkamp), pp. 283–306. Utrecht University, Utrecht.

Wiley, E.O. (1988) Parsimony analysis and vicariance biogeography. *Systematic Zoology*, **37,** 271–290.

Zandee, M. & Ross, M.C. (1987) Component-compatibility in historical biogeography. *Cladistics*, **3,** 305–332.

Zink, R.M., Blackwell, R.C. & Ronquist, F. (2000) The shifting roles of dispersal and vicariance in biogeography. *Proceedings of the Royal Society B: Biological Sciences*, **267,** 497–503.

BIOSKETCHES

Ivonne J. Garzón-Orduña is a second year PhD student in the program of Conservation Biology at the University of New Orleans; she is currently interested in cladistic biogeography and the phylogenetic history of *Hamadryas* butterflies.

Daniel Rafael Miranda-Esquivel is a professor at the Biology School of Universidad Industrial de Santander in Colombia. His research interests include the phylogenetic history within Neotropical black flies (Diptera: Simuliidae), the biogeographical origin of the Neotropics, quantitative approaches to identifying areas of endemism, and weighting schemes in cladistic analysis.

Mariano Donato is a CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina) researcher at Museo de La Plata, Argentina. His research interests include the systematics and biogeography of the Chironomidae (Diptera), the biogeography of southern South America, and cladistic methods.

Editor: Malte Ebach