Comments

Integrating Phylogenetic Diversity, Complementarity, and Endemism for Conservation Assessment

DANIEL P. FAITH,*‡ C. A. M. REID,* AND JAMES HUNTER†

*The Australian Museum, 6 College Street, Sydney, New South Wales, 2010, Australia
†University of Sydney, Sydney, New South Wales, 2010, Australia

Introduction

Posadas et al.'s (2001) analysis of conservation priority setting for southern South America highlights the fact that "conservation of biodiversity requires knowledge of its history." In addressing this need, they argue that conservation of "evolutionary potential and phylogenetically rare taxa" requires integration of "phylogenetic" diversity with other factors, including complementarity and endemism. Posadas et al.'s proposed strategy for priority setting appears to achieve that integration, based on nominated definitions of *phylogenetic diversity*, *complementarity*, and *endemism*. But these terms currently have a variety of interpretations in conservation biology, and there is a danger that a mix-and-match integration may produce an unworkable framework. We argue that Posadas et al.'s approach suffers from this difficulty and that, far from providing the claimed "better use of available information," any application of their proposed methods could be detrimental to biodiversity conservation. We contrast their approach with phylogenetic diversity (PD; Faith 1992*a*), an existing, simple integration of phylogeny, complementarity and endemism. Given that PD based complementarity and endemism have had little application, we also illustrate their use in a biodiversity study of Coleoptera in New South Wales, Australia.

Phylogenetic Diversity, Complementarity, and Endemism

The link between phylogenetic pattern, complementarity and endemism can be understood as one special case of a general descriptive model for biodiversity (Faith 1994*a*; Faith & Walker 1996*a*, 1996*b*). In general, basic biodiver-

sity "units" of some kind are the items we want to count up. "Objects" (containing various units) form sets, and we evaluate the biodiversity gain from adding a specified object to an existing set. Typically, many units remain unobserved or unknown and a "pattern" of some kind defines relationships among the objects, enabling predictions of gains at the unit level. Thus, measured "pattern complementarity" predicts unit-level complementarity (Faith 1994*a*). Whereas complementarity originally (Vane-Wright et al. 1991) referred to the gain in species (or other biodiversity elements) provided by adding an area to a set of areas, complementarity more generally refers to the predicted gain in biodiversity units provided by adding an object to a set of objects (Faith 1994*a*).

For PD, the pattern is phylogeny, units are features and characters, and objects are species, and we are interested in the gain in features when adding one or more species to a set of species. The PD measure uses phylogenetic patterns of evolutionary diversification to predict feature diversity of sets of species. The total PD of a given set is the total phylogenetic branch length spanned (represented) by its member species. The PD complementarity of a species is measured by the additional branch length it represents that is not spanned by a reference set of species (Faith 1992*a*). When the reference set is all other species, the PD complementarity value is the unique PD contribution of that species; it can be thought of as "endemism" at the level of features within species (rather than species within areas).

In reserve-selection problems (e.g., Pressey et al. 1993), addition of species to sets of species occurs by adding areas to sets of areas. Phylogenetic-diversity complementarity of an area(s) is the additional branch length collectively contributed by those species in the area(s) but not in a reference set. Such calculations involve a twofold complementarity. Not only do we not consider species that are already represented, we do not count branches represented by a new species if they are already spanned by other species in the reference set.

255

A PD-based measure of endemism of areas (Faith 1992*a*; 1994*b*) results when the reference set corresponds to the set of species in all other areas. It is the amount of branch length (PD) or "evolutionary history" (Faith 1994*a*) uniquely represented by the area. For example, the PD unique to northwest Tasmania has been estimated for a phylogeny of amphipods. The PD endemism, more than conventional species-level endemism, highlighted the potential conservation importance of that area (Faith 1994*b*; for another example, see Moritz & Faith 1998).

Other recent applications and extensions of PD are reported by Smith et al. (2000), Polasky et al. (2001), Barker (2002), Rogrigues and Gaston (2002), and Sechrest et al. (2002). Discussion of PD in the context of more general phylogenetic issues in conservation can be found in Bininda-Emonds et al. (2000).

Method of Posadas et al.

Posadas et al. (2001) do not consider PD but do examine a number of alternative "phylogenetic indices," with or without modifications, to take "endemism" into account (the complexity introduced by these alternatives and variations requires a flow diagram; see their Fig. 4). Here, we focus on their preferred index, W_e. Similar problems, as noted below, would be found for the other indices they examined, based on the similar I index of Vane-Wright et al. (1991).

Posadas et al.'s preferred phylogenetic index, W_e, is a modification of the W index of Vane-Wright et al. (1991). It assigns to each species a value that is inversely related to the count of the number of groups on the cladogram for which the species is a member (in our hypothetical example tree in Fig. 1a, species d is a member of three groups). Dividing this count into the maximum count over all species (8) produces the index value for a species (species d has an index value, W, of 8/3). Species near the "base" of the tree will have a high index value because they are in few groups (W of species a is 8/2 =

4). This property is interpreted as capturing taxonomic distinctiveness (Vane-Wright et al. 1991).

The species-level index extends to provide an index for areas. Posadas et al.'s modification to the W index occurs when index values are assigned to areas, based on sums of contributions over member species. Their W_e index is regarded as incorporating "endemicity" because the W value for a species is divided equally among the areas where that species is present. An area having the sole representation of a species therefore obtains the full W score.

In selecting a set of areas to represent Posadas et al.'s version of phylogenetic diversity a form of complementarity is used. After one area is selected, based on the W_e values, the next area selected is the one with maximum complementarity. This well-established principle normally points to the number of extra components of biodiversity contributed by an area in the context of an already protected set (Margules et al. 1988; Vane-Wright et al. 1991; Pressey et al. 1993). However, Posadas et al. use what we refer to as "Jaccard-complementarity," based on the Jaccard dissimilarity between pairs of areas (the number of species in one but not both areas divided by the number of species in either area).

Example calculations (Table 1; Fig. 1) illustrate properties of their phylogenetic index and method for selecting areas. Area 1 is selected first, having the highest index value (8). Then the area having the highest Jaccard complementarity to this area is identified. Areas 3 through 8 all have maximum Jaccard-complementarity, whereas area 2 has a Jaccard-complementarity value of only 0.83. Among areas 3–8, area 3 (or 6) is selected, having the highest W_e index value. Repeating this process, areas 4, 5, 7, and 8 all have the highest possible Jaccard complementarity to the set consisting of areas 1 and 3 (Table 1). Among these, area 4 is selected, having the equal-highest W_e index value.

The first priority given to area 1 demonstrates the weakness of the W_e index of phylogenetic distinctiveness. Species a, b, and c are all closely related, but the summing

Table 1. Example calculation illustrating properties of Posadas et al. index and area-selection methods based on the hypothetical phylogenetic tree in Fig. 1a.

Area	Species present	W_e	Jaccard dissimilarity to area 1	Jaccard dissimilarity to area 1 and 3	Area's total PD*	Area's PD endemism	PD complementarity of area, given area 2
1	a,b,c	8.0	—	—	9 (3)	2	2
2	c,d,g,n	7.1	0.83	8/9	25 (10)	11	—
3	h,i	1.3	1.0	—	8 (6)	0	3
4	j,k	1.1	1.0	1.0	9 (7)	0	3
5	l,m	1.0	1.0	1.0	10 (8)	0	3
6	h,i	1.3	1.0	1.0	8 (6)	0	3
7	j,k	1.1	1.0	1.0	9 (7)	0	3
8	l,m	1.0	1.0	1.0	10 (8)	0	3

Number in parentheses is phylogenetic diversity (PD) value based on node counting.
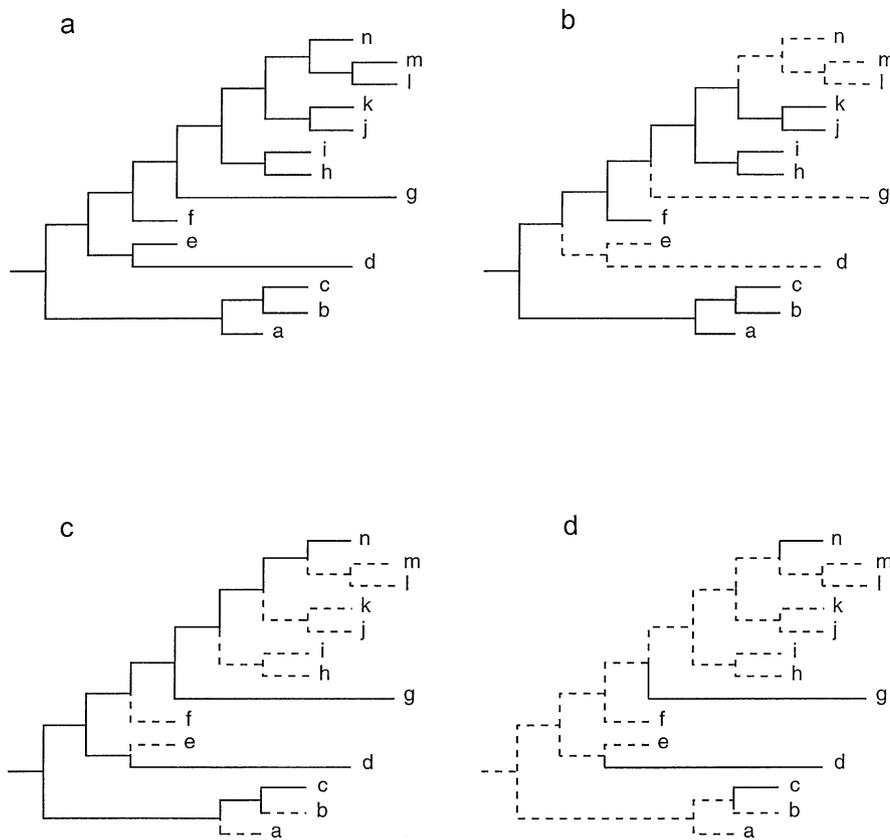
Figure 1. (a) A hypothetical phylogenetic tree, drawn in proportion to branch lengths, for species a–n. Distance from a to its most recent common ancestor with b defines one length unit. (b) Total phylogenetic diversity (PD) represented in areas 1, 3, and 4 (Table 1) (shown by solid branches relative to dashed lines for other branch segments). (c) Total PD represented in area 2 on its own (Table 1) (solid branches). (d) The PD endemicity of area 2, totaling 11 units (solid branches, excluding c).

of the individual index values by Posadas et al. to obtain the area score fails to account for the lack of complementarity among those species (this problem also would occur in the use of the *I* index). In contrast, PD recognizes area 2, with species spanning the phylogenetic tree, as representing the single best area. Area 2 has a total PD of 25 units (Table 1, last column, shows consequent PD complementarity values of other areas).

Advantages of other measures such as PD relative to the *W* index have been documented previously (see Williams et al. 1991; Faith 1992*a*, 1992*b*). Posadas et al.'s attempted integration of the *W* index with complementarity and endemism highlights further weaknesses. One problem is demonstrated by the Jaccard-complementarity values, given for area 1. Area 2 had the lowest Jaccard complementarity, yet area 2 contributes more additional species than any other area. This problem can be traced to the equation of "complementarity" with Jaccard dissimilarity. Under this complementarity measure, area 2's contribution of three additional species (not just two) is discounted because this area also shares a species (c) with area 1.

Posadas et al. appear to have misunderstood implications of the Jaccard index, claiming that a low dissimilarity value means that the areas "share many species." In fact, a low Jaccard value could arise even if the two areas share a single species, as long as there were no species found in one area and not the other.

Weaknesses of Jaccard complementarity have been noted previously in response to Colwell and Coddington's (1994) earlier interpretation of complementarity as Jaccard dissimilarity. Faith and Walker (1996*b*) presented an example in which this version of complementarity does not point to the area that would contribute the greatest number of additional species.

A second concern also relates to species-level complementarity. Area 2 has greater PD complementarity than area 1 because it contributes additional species d, g, and n, for 18 additional PD units (Fig. 1a). The total PD represented by area 1 is only 9 units (or 3 nodes, with the node-counting version of PD; Faith 1992*a*). The set of three priority areas, 1, 3, and 4, derived from Posadas et al.'s method has a total PD of only 20 units (or 10 nodes; Fig. 1b). In contrast, the PD of area 2 on its own is 25 units (or 10 nodes; Fig. 1c), reflecting the fact that its member species span the phylogeny well, including highly divergent species d and g.

Related problems arise with Posadas et al.'s attempt to incorporate "endemicity." Area 2 has three endemic species, spanning the phylogeny, and the W_e index appears reasonable in giving the full *W* score for each species to area 2. After areas 1 and 3 are selected in the example, however, neither the *W* scores for these species nor the Jaccard complementarity recognizes this additional contribution available from area 2. In contrast, for PD-based endemicity (Fig. 1d), the loss of area

2 implies a loss of 11 units of PD even if all other areas were protected.

The weaknesses of Posadas et al.'s approach arise both from poor choice of definitions of the individual terms (diversity, complementarity, endemism) and through an ad hoc integration of these key concepts. Posadas et al. concluded that the "use of a combination of cladograms and geographical information would represent a step toward including historical information in the conservation of biodiversity." We applaud this perspective but are concerned that in practice their proposed approach would be a step backward relative to the existing integration of complementarity and endemism concepts in the calculus of PD.

Phylogenetic Diversity and Beetle Genera in Northern New South Wales

We have discussed how PD-based complementarity and endemism of areas are a natural part of the calculus of PD and thus reflect marginal gains and losses in biodiversity at the level of "features," not species. Phylogenetic diversity therefore may provide information about conservation value that is not captured by conventional species-level approaches, but this possibility has been little investigated. At a large geographic scale, PD endemism may highlight different areas compared with species endemism because divergence corresponding to, say, complete genera may be restricted to one area (as for orchids in Australia; Faith 1994*b*) (for a related discussion of PD, see Polasky et al. 2001; Rogrigues & Gaston 2002; Sechrest et al. 2002). At a finer geographic scale, these two forms of endemism are more likely to correspond. Generic-level divergence, for example, is less likely to be restricted to one place.

We are exploring PD-based complementarity and endemism over several taxonomic groups at the finer scale of forest areas in northern New South Wales, Australia. Results presented here focus on the records of 10 genera in five beetle families in five key areas within northern New South Wales forests: *Pamborus* (Carabidae); *Rhyssonotus* (Lucanidae); *Aulacocyclus* and *Mastachilus* (Passalidae); *Amphistomus*, *Diorygopyx*, and the *Onthophagus capella* species group (Scarabaeidae); and *Cardiothorax egerius* and *longipes* species groups, *Coripera* and *Nototrintus* (Tenebrionidae). These groups were selected for our PD study because (1) each is monophyletic; (2) each has at least five species and therefore can theoretically provide an endemic species for each area; (3) except for Passalidae, each includes species with small ranges—eight of the groups include flightless species; (4) their taxonomy is relatively well understood, either through recent revisions or revisions currently being undertaken by C.A.M.R; (5) the insects are relatively large, easily collected, and well represented in collections, and

their ranges are therefore more accurately plotted than for most other beetle species. Figure 2 shows the estimated phylogeny for each of the 10 genera. For the PD analyses based on these phylogenetic trees, all branches (a branch is any one segment between two nodes of the tree) are assumed to have unit length (Faith 1992*a*), because confident estimates for branch lengths are not currently available.

The five geographic areas chosen for analysis all include numerous endemic species of arthropods, including Coleoptera (Adam 1987). The endemic species are generally associated with rainforest within each area. "Areas" were defined arbitrarily by a 25-km radius around a midpoint of sample-record locations. All records (published and in the Australian Museum collection) for each beetle taxon from each area were included.

The total PD of an area has a straightforward interpretation based on the phylogenetic trees in Fig. 2. Over all trees, total PD is the sum of all unit-length branches spanned by the set of species in that area. The root of each tree is included, so even when an area has only one species from a given tree, the area has a contribution to total PD, as indicated by the length of the spanning path from that species to the root of the tree. The total PD of a set of areas is equal to the total PD that would be defined by the set of species in a composite area formed by the union of the individual member areas. We can search for the best set of (e.g., 2) areas with maximum PD by using a "reserve-selection algorithm" that incorporates a PD option (Target Software privately distributed by P. A. Walker & D.P.F.).

The PD endemism of an area also extends naturally to a set of trees. It is the sum, over all trees in Fig. 2, of the branch lengths for all those branches that are *only* spanned by species found within that particular area.

Table 2 compares the five New South Wales areas based on total PD, PD endemism, species richness, and species-based endemism. First, we examined patterns among areas for total PD versus PD endemism. Area 3 has the lowest total PD and lowest PD endemism. However, total PD in general does not predict PD endemism levels. Areas 2 and 5 have the highest total PD but both have relatively low PD endemism. Area 1 has one of the lower total PD values but the highest PD endemism (Table 2; Fig. 2).

Even at the fine geographic scale of 1963 km², PD provides different information about areas relative to species counts. Areas 2 or 5 have the highest total PD and thus represent alternative choices for the best single priority area for conservation of evolutionary history, but they are not distinguishable from area 4 by species richness (Table 2). Areas 1 and 5 form the best set of two areas for PD representation (92 units). However, this is not the best set based on species counts (areas 1 + 4, or 2 + 4, with 42 species, exceed the 41 species of the set made up of areas 1 and 5), and the two best sets for species offer only 90 or 91 PD units.
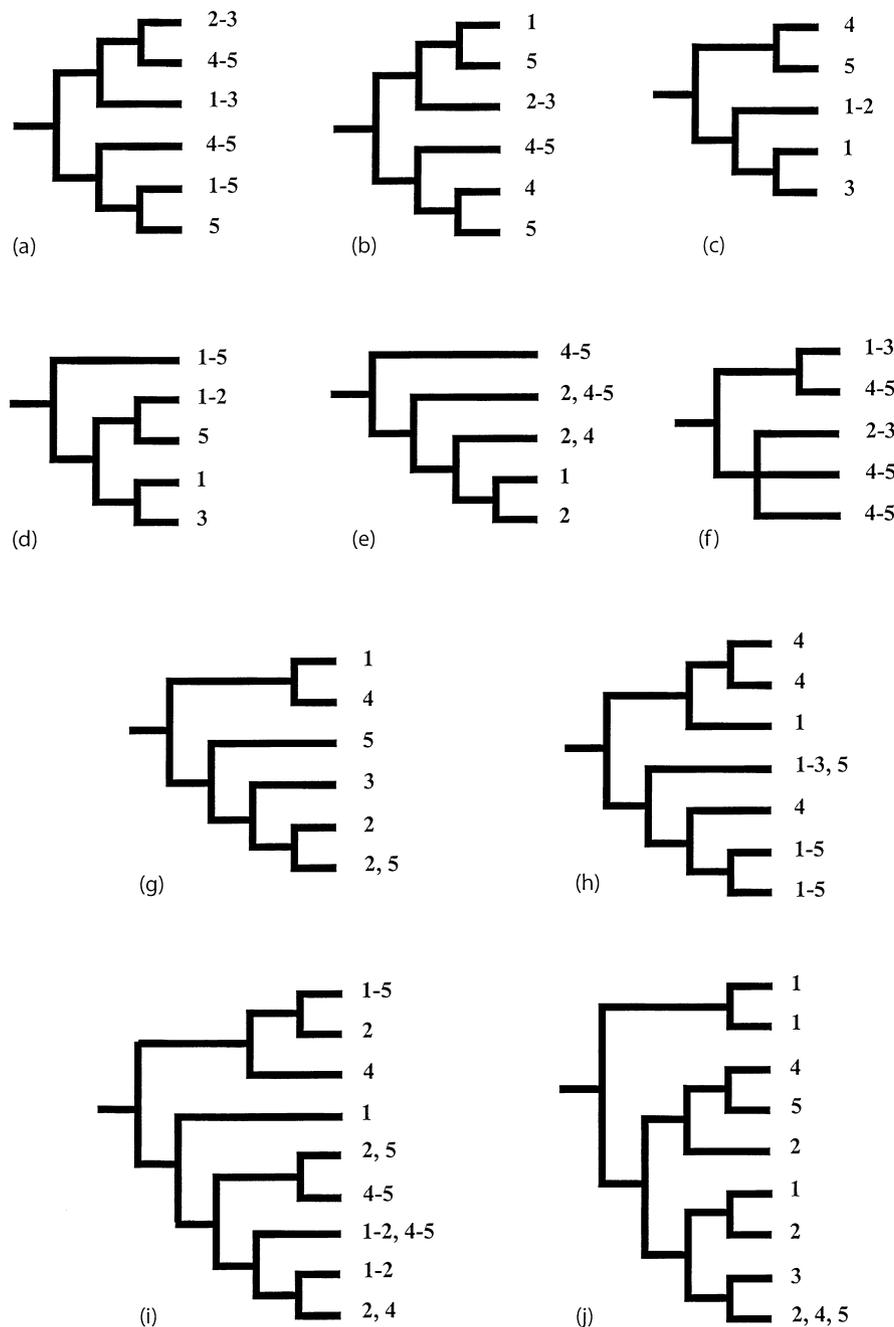
Figure 2. Estimated phylogenies for 10 taxonomic groups within Coleoptera for species in northeastern New South Wales, Australia. Branch lengths for each tree, as drawn, are arbitrary. For each phylogeny, species are labeled with numbers indicating areas of species distribution: 1, Barrington Tops massif; 2, Dorrigo Plateau; 3, eastern New England table land, Ebor Plateau; 4, Mount Warning and surrounding caldera; 5, ranges around Woodenbong. The following taxa are shown: (a) Pamborus (Carabidae); (b) Diorygopyx (Scarabaeidae); (c) Coripera (Tenebrionidae); (d) Rhyssonotus (Lucanidae); (e) Cardiothorax egerius group (Tenebrionidae); (f) Amphistomus (Scarabaeidae); (g) Nototrintus (Tenebrionidae); (h) Onthophagus capella group (Scarabaeidae); (i) Passalidae; (j) Cardiothorax longipes group (Tenebrionidae). Phylogenetic trees were estimated by cladistic parsimony applied to morphological characters (unpublished data available from C.A.M.R.). Each tree was rooted against related genera or species groups. All taxa are flightless except one Rhyssonotus, two Amphistomus, all Onthophagus species, and all Passalidae. All species are restricted to forests.

Table 2. Comparative richness and endemism values for five areas in New South Wales, based on the phylogenetic trees of Fig. 2.*

Area	Total PD	Number of species	PD endemism	Species endemism
1	61	22	11	10
2	67	26	5	5
3	46	15	4	4
4	65	26	9	8
5	67	26	7	7

Areas are numbered as follows: 1, Barrington Tops massif; 2, Dorrigo Plateau; 3, eastern New England tableland, Ebor Plateau; 4, Mount Warning and surrounding caldera; 5, ranges around Woodenbong. For details of phylogenetic diversity (PD) calculations, see text.

For these data, relative PD endemism values largely reflect the relative values of species endemism for areas (Table 2). This may in part reflect the absence of information on branch length; an area uniquely having a highly divergent species (at the end of a long branch) would have a measurable endemism contribution not detected by simple species counting. A similar consideration may help account for similarities between PD and species counting found (Polasky et al. 2001) when PD is restricted to a special case of "clock-like" evolution (in which branch lengths are ultrametric; Faith 1994a). Such considerations highlight the pitfalls involved in drawing general conclusions about PD and species congruence

based on single data sets, as in the study by Polasky et al. (2001).

Another factor affecting our New South Wales PD-species comparisons is that there are few cases in which an entire monophyletic group on the tree is found only in one area (as for the two sister species in area 1 in Fig. 2j). Such a pattern contrasts PD endemism from species endemism because species counting will not detect the shared evolutionary history of the group restricted to that area. As suggested earlier, geographic scale may be an important factor. At the courser scale of all of Australia, species endemism and PD endemism ratings of areas, based on hemipteran phylogenetic patterns, have many differences, in part reflecting whole genera restricted to certain areas (G. Cassis et al., unpublished data).

The apparent high PD (and species) endemism values at this fine geographic scale in New South Wales (further sampling may increase confidence that species absent from some areas are truly absent) are relevant to biodiversity planning. In the future, phylogenetic information may complement forest types and other coarse-scale biodiversity surrogates for New South Wales forests by revealing differences among areas in biodiversity value.

Discussion

The PD calculations in Figs. 1 and 2 highlight the fact that the PD calculus operates as if one is counting up the fundamental biodiversity units in assessing various sets of areas. We argue that any attempted integration of diversity, complementarity, and endemism that does not have this "counting" interpretation should be viewed skeptically.

Because PD implicitly counts unit features among sets of objects, it provides straightforward notions of complementarity (number of additional units gained) and endemism (number of units uniquely contained) in the context of objects and sets. These sets may be defined by the species themselves or by areas as collections of species. Indeed, PD complementarity and endemism can be applied when sets are defined in other ways. For example, we can talk about the PD endemism of an ecotype (rather than area), as for *Acidobacterium* lineages that are found globally but may represent a bacterial evolutionary history that is unique to an acidic ecotype (Wise et al. 1997; Radajewski et al. 2000).

We return to the important theme of Posadas et al.'s paper—"conservation of biodiversity requires knowledge of its history"—and argue that representation of evolutionary history, using PD, satisfies such a requirement in three ways. (1) Species that have a high PD complementarity are those that can be expected to offer novel features. Maintenance of biodiversity option value dictates that some conservation attention be focused at this level

(World Conservation Union 1980; Faith 1994*a*; Williams et al. 1994). (2) Areas that have high PD complementarity over different taxonomic groups may be expected to offer such gains for other groups as well because area-based congruence in phylogenetic patterns links to common evolutionary and biogeographic processes (Faith 1992*a*; Moritz & Faith 1998). (3) Sets of protected areas maximizing PD may represent the best single strategy for conserving not only pattern but also evolutionary processes (Faith 1992*a*; Faith 2002).

We have demonstrated that, although Posadas et al. properly argued for an integration of "phylogenetic diversity" with complementarity and endemism, only PD achieves this in a way that allows our knowledge of evolutionary history to serve biodiversity conservation.

Literature Cited

Adam, P. 1987. New South Wales rainforests: the nomination for the World Heritage List. National Parks and Wildlife Service of New South Wales, Sydney.

Barker, G. M. 2002. Phylogenetic diversity: a quantitative framework for measurement of priority and achievement in biodiversity conservation. Biological Journal of the Linnean Society **76:**165–194.

Bininda-Emonds, O. R. P., D. Vázquez, and L. L. Manne. 2000. The calculus of biodiversity: integrating phylogeny and conservation. Trends in Ecology & Evolution **15:**92–94.

Colwell, R. K., and J. A. Coddington. 1994. Estimated terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London Series B. **345:**101–118.

Faith, D. P. 1992*a*. Conservation evaluation and phylogenetic diversity. Biological Conservation **61:**1–10.

Faith, D. P. 1992*b*. Systematics and conservation: on predicting the feature diversity of subsets of taxa. Cladistics **8:**361–373.

Faith, D. P. 1994*a*. Phylogenetic pattern and the quantification of organismal biodiversity. Philosophical Transactions of the Royal Society of London Series B. **345:**45–58.

Faith, D. P. 1994*b*. Phylogenetic diversity: a general framework for the prediction of feature diversity. Pages 251–268 in P. L. Forey, C. J. Humphries, and R. I. Vane-Wright, editors. Systematics and conservation evaluation. Clarendon Press, Oxford, United Kingdom.

Faith, D. P. 2002. Quantifying biodiversity: a phylogenetic perspective. Conservation Biology **16:**248–252.

Faith, D. P., and P. A. Walker. 1996*a*. How do indicator groups provide information about the relative biodiversity of different sets of areas? On hotspots, complementarity and pattern-based approaches. Biodiversity Letters **3:**18–25.

Faith, D. P., and P. A. Walker. 1996*b*. Environmental diversity: on the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. Biodiversity and Conservation **5:**399–415.

Margules, C. R., A. O. Nicholls, and R. L. Pressey. 1988. Selecting networks of reserves to maximise biological diversity. Biological Conservation **43:**63–76.

Moritz, C., and D. P Faith. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. Molecular Ecology **7:**419–430.

Polasky, S., B. Csuti, C. Vossler, and S. M. Meyers. 2001. A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds. Biological Conservation **97:**99–105.

Posadas, P., D. R. Miranda Esquivel, and J. V. Crisci. 2001. Using phylogenetic diversity measures to set priorities in conservation: an example from southern South America. Conservation Biology **15:**1325–1334.

Pressey, R. L., C. J. Humphries, C. R. Margules, R. I. Vane-Wright, and P. H. Williams. 1993. Beyond opportunism: key principles for systematic reserve selection. Trends in Ecology & Evolution **8:**124–128.

Radajewski, S., P. Ineson, N. R. Parekh, and J. C. Murrell. 2000. Stable-isotope probing as a tool in microbial ecology. Nature **403:**646–649.

Rodrigues, A. S. L., and K. J. Gaston. 2002. Maximising phylogenetic diversity in the selection of networks of conservation areas. Biological Conservation. **105:**103–111.

Sechrest, W., T. M. Brooks, G. A. B. da Fonseca, W. R. Konstant, R. A. Mittermeier, A. Purvis, A. B. Rylands, and J. L. Gittleman. 2002. Hotspots and the conservation of evolutionary history. Proceedings of the National Academy of Science of the United States of America **99:**2067–2071.

Smith, T. B., K. Holder, D. Girman, K. O'Keefe, B. Larison, and Y. Chan, 2000. Comparative avian phylogeography of Cameroon and Equatorial Guinea mountains: implications for conservation. Molecular Ecology **10:**1505–1516.

Vane-Wright, R. I., C. J. Humphries, and P. H. Williams. 1991. What to protect? Systematics and the agony of choice. Biological Conservation **55:**235–254.

Williams, P. H., C. J. Humphries, and R. I. Vane-Wright. 1991. Measuring biodiversity: taxonomic relatedness for conservation priorities. Australian Systematic Botany **4:**665–679.

Williams, P. H., K. J. Gaston, and C. J. Humphries. 1994. Do conservationists and molecular biologists value differences between organisms in the same way? Biodiversity Letters **2:**67–78.

Wise, M. G., J. V. McArthur, and L. J. Shimkets. 1997. Bacterial diversity of a Carolina bay as determined by 16S rRNA gene analysis: confirmation of novel taxa. Applied and Environmental Microbiology **63:**1505–1514.

World Conservation Union (IUCN). 1980. World conservation strategy: living resource conservation for sustainable development. IUCN, Gland, Switzerland.