Conserving Phylogenetic Diversity
can be a Poor Strategy for Conserving Functional Diversity

Florent Mazel\textsuperscript{1,*}, Arne Mooers\textsuperscript{1}, Giulio Valentino Dalla Riva\textsuperscript{2}, and Matthew W. Pennell\textsuperscript{2}

\textsuperscript{1}Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
\textsuperscript{2}Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;

* Corresponding Author; flo.mazel@gmail.com
Abstract

For decades, academic biologists have advocated for making conservation decisions in light of evolutionary history. Specifically, they suggest that policymakers should prioritize conserving phylogenetically diverse assemblages. The most prominent argument is that conserving phylogenetic diversity (PD) will also conserve diversity in traits and features (functional diversity; FD), which may be valuable for a number of reasons. The claim that PD-maximized (‘maxPD’) sets of taxa will also have high FD is often taken at face value and in cases where researchers have actually tested it, they have done so by measuring the phylogenetic signal in ecologically important functional traits. The rationale is that if traits closely mirror phylogeny, then saving the maxPD set of taxa will tend to maximize FD and if traits do not have phylogenetic structure, then saving the maxPD set of taxa will be no better at capturing FD than criteria that ignore PD. Here, we suggest that measuring the phylogenetic signal in traits is uninformative for evaluating the effectiveness of using PD in conservation. We evolve traits under several different models and, for the first time, directly compare the FD of a set of taxa that maximize PD to the FD of a random set of the same size. Under many common models of trait evolution and tree shapes, conserving the maxPD set of taxa will conserve more FD than conserving a random set of the same size. However, this result cannot be generalized to other classes of models. We find that under biologically plausible scenarios, using PD to select species can actually lead to less FD compared to a random set. Critically, this can occur even when there is phylogenetic signal in the traits. Predicting exactly when we expect using PD to be a good strategy for conserving FD is challenging, as it depends on complex interactions between tree shape and the assumptions of the evolutionary model. Nonetheless, if our goal is to maintain trait diversity, the fact that conserving taxa based on PD will not reliably conserve at least as much FD as choosing randomly raises serious concerns about the general utility of PD in conservation.

Keywords. Conservation, Functional diversity, Evolutionary diversity, Trait evolution, Species prioritization
In the face of the current biodiversity crisis, society needs to decide how to distribute limited funds and effort to conservation. Conservation biologists and policy makers have presented many proposals for making rational and scientific decisions about which species warrant the most protection (Bottrill et al. 2008).

One prominent prioritization scheme uses evolutionary history to place a quantitative value on species and sets of species. The idea is that when making conservation policy, we should try to conserve the set of species or habitats that harbour the greatest amount of evolutionary history (Vane-Wright et al. 1991). While there are many, overlapping metrics for measuring the evolutionary history encompassed by a set of species (Winter et al. 2013; Tucker et al. 2016), the most common is the sum of all branch lengths connecting a set of species to a common root (Faith 1992), called Phylogenetic Diversity (PD). This measure is vague insofar as the units of “branch length” are unspecified, but it is the metric whose maximization has been proposed as a conservation prioritization strategy.

While PD has been only sparingly used in actual policy decisions (for one example, see the EDGE program of the Zoological Society of London; www.edgeofexistence.org), it has caught the attention of researchers; according to Google Scholar, the original Faith (1992) paper on the topic has been cited more than 1900 times as of April 2017. Indeed, Faith’s paper has spawned an entire subfield in which biologists and mathematicians have worked out complex solutions to measuring and maximizing PD (e.g. Rodrigues and Gaston 2002; Forest et al. 2007; Bordewich et al. 2008; Bennett et al. 2014; Chao et al. 2015; Pollock et al. 2015; Thuiller et al. 2015). Faith and other researchers have proposed several key reasons why conserving PD is worthwhile. Prioritizing species’ conservation to maximize PD may help ensure that: (i) remarkable species that occur as evolutionary isolated lineages (e.g. tailed frogs, tuataras, Welwitschia) are prioritized (Rosauer and Mooers 2013); (ii) essential ecosystem functions and services are maintained (Cadotte et al. 2008, but see Srivastava and Vellend, 2005 for a discussion on the link with applied conservation); and that (iii) we maximize ‘evolutionary potential’ (Faith 1992; Forest et al. 2007). All of these ideas have are underpinned by the claim that phylogenetically diverse sets of taxa contain a disproportionately large amount of trait/feature/functional diversity. Hereafter, we will
not make a distinction between trait, feature and functional diversity and we will refer to
them as functional diversity, or FD.

Like evolutionary history, functional diversity is an ambiguous concept with
many potential measures. Villéger et al. (2008) suggest that FD has three components:
richness, divergence, and evenness. Functional richness generally measures ‘how much
trait space is filled, while functional divergence and evenness indices describe how this
space is filled’ (Schleuter et al. 2010). Functional richness represents the amount of
functional trait space that is encapsulated by a set of species, is usually correlated with
species richness, and can be related to the functioning of ecosystems (see, e.g. Cadotte et
al. 2011). The second component, functional divergence, is largely independent of
species richness and describes how species are clustered in trait space, which may be
valuable to conservation biologists interested in, e.g., ecosystem services (e.g. Diaz et al.
2007). These two classes of measure are often used in trait-ecology and conservation (e.g.
Devictor et al. 2010; Mouillot et al. 2014) and we assume here that conserving functional
richness and/or divergence is a valuable conservation objective. While functional
divergence relates to some measure of mean trait distances between species, functional
evenness relates to the variance of these trait distances. This last FD dimension describes
the extent to which species are clustered with their (direct) neighbours versus being
regularly spaced in trait space. We did not consider any measure of functional evenness
(such as the Functional Evenness Index, Villéger et al. 2008) in what follows because
could not identify any potential causal link between evenness in trait space and
ecosystem function or services, and trait evenness has generally not been a concern of
conservation biologists.

In this paper, we ask whether maximizing PD help to conserve functional
diversity. The common rationale for using PD as a proxy for FD is that many ecologically
relevant traits harbour some degree of phylogenetic signal (see, e.g., Winter et al. 2013).
At a glance, this seems logical: if the data shows strong phylogenetic signal, then picking
distantly related taxa seems a sensible way to ensure that you have captured species from
across trait space. And indeed, if we assume that traits have evolved according to a
Brownian motion (BM, Felsenstein 1985) process, then this will be true (see below). The
converse is also true: if traits do not show phylogenetic signal, other methods for
Capturing FD are needed (see, e.g., Faith 2015). A number of studies from across evolutionary biology, ecology, and conservation biology have evaluated the amount of phylogenetic signal (measured in a variety of ways, see Münkemüller et al. 2012) in ecologically important traits (see, e.g., Freckleton et al. 2002; Blomberg et al. 2003; Chamberlain et al. 2012). Recently, Kelly et al. (2014) specifically focused on the implications of phylogenetic signal for the use of PD in conservation. They constructed trees using a wide variety of morphological traits and found that while closely related species often shared many trait combinations, these traits were not informative for deeper splits in the tree. They argued that this was evidence that maximizing PD would not reliably maximize feature diversity.

The results of these studies (along with, likely many more) have been widely variable: some traits in some taxa in some regions contain a lot of phylogenetic signal while others do not. This led Winter et al. (2013, p 201) to conclude: “If the conservation goal is to conserve functional diversity, considering phylogenetic diversity might be either well suited or totally misleading”. We argue that there is an important and underappreciated assumption in this line of reasoning: that the degree of phylogenetic signal in some key trait(s) is indicative of the effectiveness of using PD to conserve FD.

There are two reasons to be suspicious of this assumption. First, our thinking about phylogenetic signal has been informed by considering a few simple models of trait evolution; other, completely different, classes of models may generate variation in phylogenetic signal that are far less intuitive. Second, the motivating idea is that policymakers should use PD to pick sets of taxa to conserve. These sets are, by definition, non-random and therefore may have different statistical properties from the clade as a whole.

In this paper, we simulate data under different models of evolution and, for the first time to our knowledge, directly test how much FD the set of taxa that maximize PD (‘the maxPD set’ hereafter) contains compared to alternative possible sets.

Specifically, we contrast the outcome for FD conservation of conserving the maxPD set of taxa -- and letting everything else go extinct -- with conserving a random set of taxa of the same size. Here, random simply means conservation decisions that ignore phylogenetic position and the functional traits we are considering. As such, random sets provide a natural point of comparison to understand the properties of
maxPD. We note that we are not testing whether conserving maxPD will maximize the amount of FD it is possible to conserve. While this claim is likely what some advocates of PD have in mind, and is what Kelly et al. (2014) actually aimed to test, it is a rather high bar to meet. Indeed, it is easy to concoct scenarios in which this will not hold; if, for example, traits were so labile that there was no phylogenetic signal (i.e., the “white noise” model), then we would expect that maxPD sets would contain no more or less FD on average than any other set. It therefore seems too high a bar to expect for PD to always maximize FD in order to declare it useful for conserving FD. Instead, we believe we must first clear a much lower bar -- does prioritizing species based on maximizing PD do better at capturing FD than prioritizing a random set?

Below, we demonstrate that both the model of trait evolution and the tree shape are relevant for deciding whether or not PD is a good strategy for conserving FD. And, more surprisingly, we show that even when there is phylogenetic signal in the sampled traits, using PD to guide conservation decisions can lead to choice outcomes for conserving FD that are worse than if we were choosing randomly. This counter-intuitive result suggests that we need to re-assess both the way in which we intuitively consider phylogenetic signal in conservation biology, and the justification for phylogenetically-based prioritization.

**METHODS**

We wanted to test the following conjecture under a variety of evolutionary scenarios:

If we select a set $S$ of $m$ taxa from a clade of size $n$ such that the sum of the branch lengths connecting $S$ is at least as large as that stemming from any other possible subset (i.e. PD is maximized), then $S$ will contain at least as much FD on average as a randomly chosen subset of size $m$.

Four things are notable about this test. First, as stated above, we are not trying to determine whether the maxPD set will actually maximize FD (i.e., that $S$ would contain at least as much FD as any other set of the same size). Second, we are interested in the


expectation, or average. Evolution certainly can take interesting turns such that some sub-
clades span the functional diversity of the entire group (e.g., different clades of African
rift cichlids have independently evolved the same breadth of functional diversity in
different lakes; Muschick et al. 2012). Or, a trait important for ecosystem functioning
may also evolve only once and we would like to make sure we capture this lineage
(Davies et al. 2016). Average properties are critical, however, because PD’s utility in
conservation comes precisely when we don’t know the traits or functions that matter; the
best we can hope for is that, on average, we expect it to perform well. Third, we do not
require $S$ to uniquely maximize PD. We use the greedy algorithm proposed by Bordewich
et al. (2008) to find our maxPD set of species $S$. For a given tree there are likely multiple,
and possibly very many, sets of with the same PD as $S$. As this number will vary across
simulations and could, in some case, be very large, we have chosen to select only one set
per simulation. This allows us to carry out more simulations, increasing the generality of
our results. And last, we are assuming that all of the taxa we select will survive and that
every other taxa in the clade will go extinct with certainty. This is, of course,
unreasonable and unrealistic but is useful for the purposes of illustration (see Discussion).

Simulations

To explore a broad range of tree shapes, we simulated trees under three different
diversification models. First, we simulated trees under a Yule process (no extinction).
Second, to obtain trees that were more ‘tippy’ (i.e., having more speciation events close
to the present), we used a coalescent model. In both cases, we simulated trees with 32 and
with 64 taxa. To obtain trees that were more unbalanced than those typically produced by
the Yule or coalescent processes, we simulated trees where the speciation rate evolved as
a continuous trait along the tree (Rabosky 2010; Beaulieu and O’Meara 2015). This
allowed some groups within a tree to diversify faster than others, with this heterogeneity
being phylogenetically clustered.

To do the latter, we used R scripts from Beaulieu (2015, modified from Rabosky
2010) and set the initial speciation rate to .06. Each tree was subsequently pruned to $n =$
64 and to $n = 32$. We then kept the 100 first trees that encompassed a wide range of
imbalance values: we kept 10 trees by bins of 0.4 imbalance value (as measured by $\beta$, Blum et al. 2006) from -1.6 to 2. For a point of comparison, we also used fully imbalanced ($\beta = -2$) and balanced trees ($\beta = 10$) of 32 and 64 species.

To explore a range of continuous trait evolution models, we used 1) the BM model setting the drift parameter $\sigma^2 = 1$ (we did not explore multiple values of $\sigma^2$ because it does not influence the phylogenetic signal of the data and thus will not impact our results); 2) the Ornstein-Uhlenbeck (OU, Hansen 1997), with $\sigma^2 = 1$ and $\alpha = \{1.4, 7\}$ corresponding to half life of .1 and .5 for a tree with total height rescaled to 1; and 3) the early burst (EB, Harmon et al. 2010, $r = -5$ and -1). For discrete traits, we used the Markov model of evolution (Pagel 1994). We used a simple Markov model with 4 character states and all transitions rates equal to .1 or 1. Speciation models, in which trait evolution occurs (at least in part) when lineages split, were also used for both continuous and discrete traits by applying a Pagel $\kappa$ transformation to the original tree (Pagel 1999). We simulated datasets with $N = \{1, 2, 4\}$ independently evolving traits. As we wanted to keep the simulations simple, we did not include variations such as multi-rate BM (O’Meara et al. 2006; Eastman et al. 2011) or multi-optima OU models (Butler and Kings 2004; Ingram and Mahler 2013; Uyeda and Harmon 2014). For each set of parameters and number of trait values, we simulated 1000 datasets. In each case, we also computed the phylogenetic signal contained in the data by calculating the spearman correlation between phylogenetic and trait distance matrices (following Kelly et al., 2014) in addition to more commonly used measures such as Blomberg’s K (Blomberg et al. 2003); Blomberg’ K does not allow for the possibility of “anti-signal” in traits, wherein close relatives are more dissimilar than distantly related taxa.

Analysis

For each dataset, we selected two sets of $m$ species ($m = \{8, 16\}$) out of the total number of $n$ species in the tree ($n = \{32, 64\}$) for a total of four parameter combinations. One set was chosen at random and the other was a set that maximized PD (i.e., maxPD), using the algorithm of Bordewich et al. 2008 that we implemented in R. We then computed FD for both the random and the maxPD sets. Functional richness was
estimated using the convex hull volume (Cornwell et al. 2006), which measures the total volume encapsulated by all species in trait space. In a single dimension, this simply equals the range of values. This broadly used metric in ecology is set monotonic with species richness, a property generally assumed desirable in conservation whereby the addition of a new species can never decrease the metric's value (Ricotta 2005). Functional divergence was estimated using Rao’s quadratic entropy (Rao 1982; Botta-Dukát 2005), which represents the mean trait distance between pairs of species (including the null distance of a species with itself) and is highly correlated to the trait variance across tips (de Bello et al. 2016). While this index is not set monotonic with species richness, we feel that it might be of interest to test the robustness of our results. By using functional richness and functional divergence, we are able to capture both the spread of the data in trait space as well as how clustered it is, since it is not immediately clear what quantity is most relevant for the use of PD in conservation. For discrete traits, the convex hull volume is less meaningful than for continuous traits. Therefore, we used the number of unique trait states in the set as a measure of FD for discrete traits (Petchey and Gaston 2006; Mouillot et al. 2014).

For each simulation, we then computed the relative amount of FD in the two sets using the following metric:

\[
\text{rFD} = \frac{\text{FD}_{\text{maxPD}}}{(\text{FD}_{\text{Random}} + \text{FD}_{\text{maxPD}})}
\]  

(eq. 1)

A rFD value of greater than or equal to 0.5 means that the PD set contains at least as much FD as the random set and a rFD value less than 0.5 means that it contains less.

All analyses were run in R, making special use of the ape (Popescu et al. 2012), ade4 (Dray et al. 2007), phytools (Revell 2012), geiger (Pennell et al. 2014a), geometry (Habel et al. 2015), apTreeshape (Bortolussi et al. 2006), and mvMORPH (Clavel et al. 2015) packages. All code to run the analyses is available at https://github.com/FloMazel/PD_FD.

**RESULTS**
We found that, under many common models of trait evolution, conserving the maxPD set of taxa will on average conserve more FD than conserving a random set of the same size (i.e. rFD is always > .5, see Table 1, note that rFD is an average over all simulations but individual simulation may have rFD < .5). This is because related species tend to be on average closer in trait space than distantly ones (Figure 2a-d), so that selecting distantly related species increases FD. This result is more pronounced for very early evolution (as modelled by an early burst model of evolution) because in this case distantly related species are always well separated in the functional space. On the contrary, very late evolution, or very strong stabilizing selection (as modelled by the OU process) tends to erase the differences between set of species, but never leads (on average) to the maxPD set of species to harbour less FD than the random set. Overall, an increase of phylogenetic signal tends to increase the difference between FD of the two sets (Table 1 and Supp. Tables). Our results also hold for alternative tree sizes (Supp. Table 1) and Functional Divergence (measured as Rao's Quadratic entropy, see Supp. Table 2). Also, the difference between FD of the two sets of species is largest when a small proportion of tree size is selected and tends to decrease when more species are selected (Supp. Table 2). This is expected: if 100% of the species are selected, the FD of the random and maxPD sets will be equal and equal to the FD of the entire clade (rFD = .5).

However, this result cannot be generalized to all classes of models. When traits evolve on an imbalanced tree under a speciational model (Figure 1), early diverging species are always selected to maximize PD (species No 1, 2 and 3 in Figure 2d-g) but are functionally relatively similar since their traits have not diverged much. Here, a random choice of species will, on average, select species that are much less functionally similar, yielding higher FD and thus an rFD < .5 (Figure 2d-h). As with other models, the difference between FD of the two sets of species is strongest when a small proportion of tree size is selected and tends to decrease when more species are selected (Fig. S1). This result also holds using Rao's measure of Functional Divergence (Fig. S2).

Above we described the results for n=2 traits. Multiple traits are likely important for maintaining ecosystem functions and services and for potentially promoting diversification. However, our results do not qualitatively depend on how many traits we consider. If we use convex hull volume as a measure of FD, then the patterns we see in
one or two dimensions are only exacerbated in higher trait dimensions (Fig. S1): in cases where maxPD does poorly, adding more traits makes it do worse, and in cases where it does well, more traits accentuate its success. When we measured FD using Rao’s quadratic entropy, there was no difference between results at two or higher dimensions (Fig. S2). This is because Rao’s quadratic entropy represents the mean functional distance between species (including comparing a species to itself) and we know that, for a BM model, increasing the number of traits simply decreases the variance of functional distances between species (see e.g. Letten and Cornwell 2015) and thus will not impact the average of the rFD metric. Importantly all our results are also robust to variation in tree size and number of selected species (Fig. S1-2) and also hold when a speciational model of evolution for discrete traits is applied instead (i.e. a Markov model, see Fig. S3).

After seeing our results, we naively thought that if there was a non-negative correlation between the traits and the phylogeny (i.e., “phylogenetic signal” broadly construed), this would mean that PD should on average do at least as well as random. Our intuition here was wrong. Indeed, even in our “worst case” scenario, when the tree is perfectly imbalanced and trait evolution only occurs at speciation, the correlation between the trait covariance matrix and the phylogenetic covariance matrix is still positive -- close relatives resemble one another but selecting the maxPD set of taxa captures less FD than a randomly chosen set on average (Fig S4-5)! The key to resolving this apparent paradox is recognizing that the phylogenetic signal of the entire dataset is not expected to equal the phylogenetic signal of non-random subsets of the data. In particular, the set of species that maximized PD is expected to occupy a very particular position in the phylogenetic and functional distances space.

To intuitively understand this point, we present in figure 2 (panels d-h) a simplified toy example with a fully imbalanced phylogeny of 16 species from which four species are selected, either at random (squares in the figure) or in order to maximize PD (maxPD set, represented by triangles). In this case, species 1, 2 and 3 will always be selected to maximize PD, while the fourth one will be chosen at random among the remaining species (Fig 2d). In the case of a speciational model of evolution, three out of four species from the maxPD set (species 1,2 and 3) will be, on average, relatively clumped in the trait space (Fig. 2e, triangles) and thus harbour small trait distances, while
being distantly related in the phylogeny (Fig. 2g). On the contrary, the random subset (squares) will be more spread in the trait space (Fig. 2f) and thus harbour relatively higher trait distances, while being relatively less distant in the phylogeny (Fig. 2g). So, the random set will harbour more FD than the maxPD set (Fig 2h). While the overall (i.e. for all species) relationship between trait and phylogenetic distances is slightly positive (and not negative), the same relationship restricted to random and maxPD subsets becomes negative (imagine a line between squares and triangles on figure 2g). It thus appears that the overall trend between all species are not representative of the trend between members of the maxPD and random sets; the measure of phylogenetic signal on the whole phylogeny may not be a good proxy for the representativeness of FD by the maxPD set of species.

**DISCUSSION**

Most of the arguments for using PD in conservation decisions reason that conserving phylogenetically diverse sets of taxa is valuable because it conserves some sort of trait diversity; for other rationales for conserving PD see e.g., Vane-Wright et al. (1991) or Rosauer and Mooers (2013). Trait diversity may be valuable if it helps maintain ecosystem functioning and services (e.g. Best et al. 2013; Winter et al. 2013; Gross et al. 2017), if it captures ‘evolutionary potential’ (Faith 1992; Forest et al. 2007), or if trait diversity increases the probability of encompassing rare traits that are deemed valuable for their rareness *per se* (e.g. egg-laying in mammals, Rosauer and Mooers, 2013). Here, we are agnostic as to why traits are valuable to conserve; we only assume that they are.

Our main results speak to at least on other recent paper that also purported to test whether PD was a good proxy of feature diversity. Using a wide variety of morphological traits previously used to infer phylogenies, Kelly et al. (2014) showed that, while closely related species often share many trait combinations, these traits are not informative for deeper splits in the tree – i.e. that phylogenetic signal decays rapidly in the tested character matrices. A second key finding of the Kelly et al. study was that the trait distances between the two most distant species in the tree (i.e. considering FD of the maxPD sets of two species) is lower than the maximal trait distance in the dataset. Our test is both more stringent and more general than that of Kelly et al. First, we did not test
whether preserving the maxPD species will maximize the amount of FD it is possible to preserve, but rather if the maxPD set capture more FD than a random set, a much lower bar to meet. For example, even in the situation where we found the maxPD set to harbour more FD than random (e.g. in the case of a simple BM model), it is likely that this set does not maximize FD. Second, while Kelly et al. focused on the FD of the maxPD set that comprises only two species, we consider here sets of taxa with a broader range of sizes (8 and 16 species). This allowed us to show that the measure of phylogenetic signal on the whole phylogeny may not be a good proxy for the representativeness of FD by the maxPD set of species.

Our analysis is, of course, rather oversimplified in some ways. In the real world, we do not have full control over which species survive and which are lost. Conservation prioritization itself is a result of a complex interplay of social, economic, political, and scientific priorities and is not always species-centred. And even if we did have the power to decide, we would neither conserve everything we chose, nor would everything we didn’t choose go extinct. Furthermore, the extinction proportions used in our simulations (e.g. 75%) are beyond dystopic. But the simplicity of our simulations allows us to evaluate the logic underlying the (seemingly obvious, but not actually obvious at all) claim that conserving phylogenetic diversity will result in conserving trait diversity. We realize also that some of the situations which produce rFD values of less than 0.5 may not be biologically realistic. It is unlikely that most trait evolution is speciational (Pennell et al. 2014b) and, while empirical trees are more unbalanced than those produced by Yule models (Mooers and Heard 1997), totally unbalanced trees are rare. While, such extreme scenarios are not necessary to reliably get rFD values of less than 0.5, we think that these cases are useful for critically evaluating the underlying logic behind the use of PD and will perhaps stimulate the production of more direct tests of the usefulness of PD to represent FD.

While there have been several meta-analyses comparing the fit of various trait models across clades (Harmon et al. 2010; Pennell et al. 2015), these have been limited to a few simple models, all of which are in the part of parameter space where PD performs well as a proxy for FD. More comprehensive meta-analyses of the fit of models to comparative data are required to allow us to assess where in model space traits of interest
generally fall. Furthermore, recent innovations using simulation-based approaches (e.g. Slater et al. 2012; Sukumaran et al. 2016; Clarke et al. 2017) may allow us to expand beyond our limited set of process models. A simpler empirical test of the utility of PD is to gather empirical datasets and to repeat our analytical procedure on these. We would then be able to ask for these empirical datasets whether the maxPD set of taxa will contain more FD than a randomly chosen set. To our knowledge, no such test has been performed. While this test would not provide a definitive answer to the utility of PD, it would at least provide some indication of how concerned we should be given our results.

That said, if we had some approximate idea as to how likely it is the maxPD fails to capture FD, policy recommendations might still be difficult. If maxPD does better than random in, say, 80% of clades/traits, should this be interpreted as an endorsement of the use of PD in conservation or a denouncement? What level of increase in FD is important? A formal decision-theoretic framework (Robert 2007) might be needed for navigating these thorny problems.

CONCLUSION

Given the interest in using PD in conservation decisions and the amount of work that has gone into the problem of how to measure and prioritize PD, it is surprising that there has not been direct theoretical or broad empirical evaluations of what exactly PD captures. Here, we find that under many common models of trait evolution and tree shapes, conserving the maxPD set of taxa will indeed conserve more FD than conserving a random set of the same size. However, under other biologically plausible scenarios, using PD to select species can actually lead to less FD compared to a random set. Importantly, this can occur even when there is phylogenetic signal in the traits. The fact that conserving taxa based on PD will not always reliably conserve at least as much FD as choosing randomly may raise serious concerns about the utility of PD in conservation if our goal is to save a diverse set of traits.

Acknowledgements
This paper is a joint effort of the sCAP working group funded by the Synthesis Centre of the German Centre for Integrative Biodiversity Research Halle-Jena-Leipzig and the Canadian Institute for Ecology and Evolution. FM was supported by an NSERC Accelerator Grant to AOM. GDR thanks Mike Steel. We thank Risa Sergeant at the University of Ottawa for hospitality, the Crawford Lab for Evolutionary Studies at SFU for discussion, Luke Harmon and Will Cornwell for statistical advice, Caroline Tucker, William Pearse, Marc Cadotte, Wilfried Thuiller, Laura Pollock, Thomas Buckley, and two anonymous reviewers for comments on a previous version of this manuscript.

References


Bottrill M.C., Joseph L.N., Carwardine J., Bode M., Cook C., Game E.T., Grantham H.,
Kark S., Linke S., McDonald-Madden E., Pressey R.L., Walker S., Wilson K.A.,


Devictor V., Mouillot D., Meynard C., Jiguet F., Thuiller W., Mouquet N. 2010. Spatial


Jennings W., Kozak K.H., McPeek M.A., Moreno-Roark F., Near T.J., Purvis A.,
Ricklefs R.E., Schluter D., Schulte II J.A., Seehausen O., Sidlauskas B.L., Torres-
Carvajal O., Weir J.T., Mooers A.O. 2010. Early bursts of body size and shape
Ingram T., Mahler D. 2013. SURFACE: detecting convergent evolution from
comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike
Letten A.D., Cornwell W.K. 2015. Trees, branches and (square) roots: why evolutionary
relatedness is not linearly related to functional distance. Methods Ecol. Evol. 6:439–
444.
Mouillot D., Villeger S., Parravicini V., Kulbicki M., Arias-Gonzalez J.E., Bender M.,
Functional over-redundancy and high functional vulnerability in global fish faunas
Münkemüller T., Lavergne S., Bzeznik B., Dray S., Jombart T., Schiffers K., Thuiller W.
O’Meara B.C., Ané C., Sanderson M.J., Wainwright P.C. 2006. Testing for different rates
Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the
45.
884.


Table 1. For common trait macroevolution models, sets of species that maximize PD always harbour, on average, at least as much FD as random sets of species of the same size. The table presents, for each combination of macroevolutionary models (column 1), specific set of parameters (column 2-3, the transition rate for the Markov model is 1, see also methods) and number of independent traits (column 4), a measure of the relative amount of FD (rFD) between maxPD and random sets of species for pure birth Yule trees (column 5-6) and coalescent trees (column 9-10). These results correspond to a tree of 64 species from which 8 are selected either at random or to maximize PD (other combinations of these parameters are presented in Supp. Tables). The comparison of FD (as captured by the convex hull measure) between the two sets of species is quantified with the following metric: rFD = FD_{maxPD}/(FD_{Random} + FD_{maxPD}). A value <.5 means PD is doing worse than random, a value >.5 means PD is doing better than random and a value of .5 means PD is doing the same as random. The phylogenetic signal for Yule trees (column 7-8) and coalescent trees (column 11-12) is measured with

<table>
<thead>
<tr>
<th>Evolutionary Model</th>
<th>Type of Trees</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yule Tree</td>
<td>Coalescent Tree</td>
<td></td>
<td>Yule Tree</td>
<td>Coalescent Tree</td>
<td></td>
<td>Yule Tree</td>
<td>Coalescent Tree</td>
<td></td>
<td>Yule Tree</td>
<td>Coalescent Tree</td>
<td></td>
<td>Yule Tree</td>
<td>Coalescent Tree</td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.53</td>
<td>0.09</td>
<td>0.98</td>
<td>0.43</td>
<td>0.54</td>
<td>0.09</td>
<td>1.04</td>
<td>0.93</td>
<td>0.54</td>
<td>0.09</td>
<td>1.04</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.55</td>
<td>0.13</td>
<td>1</td>
<td>0.31</td>
<td>0.6</td>
<td>0.12</td>
<td>1.01</td>
<td>0.64</td>
<td>0.54</td>
<td>0.09</td>
<td>1.01</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.61</td>
<td>0.19</td>
<td>1</td>
<td>0.21</td>
<td>0.74</td>
<td>0.17</td>
<td>1.01</td>
<td>0.45</td>
<td>0.54</td>
<td>0.09</td>
<td>1.01</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td>0.08</td>
<td></td>
<td></td>
<td>0.54</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>1</td>
<td></td>
<td></td>
<td>0.53</td>
<td>0.07</td>
<td></td>
<td></td>
<td>0.56</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>2</td>
<td></td>
<td></td>
<td>0.53</td>
<td>0.05</td>
<td></td>
<td></td>
<td>0.57</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markov</td>
<td>4</td>
<td></td>
<td></td>
<td>0.53</td>
<td>0.05</td>
<td></td>
<td></td>
<td>0.57</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>-5</td>
<td>1</td>
<td>0.55</td>
<td>0.09</td>
<td>4.32</td>
<td>2.51</td>
<td>0.56</td>
<td>0.11</td>
<td>9.18</td>
<td>7.36</td>
<td>0.54</td>
<td>0.09</td>
<td>1.52</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0.53</td>
<td>0.09</td>
<td>1.34</td>
<td>0.62</td>
<td>0.55</td>
<td>0.09</td>
<td>1.52</td>
<td>1.33</td>
<td>0.54</td>
<td>0.09</td>
<td>1.52</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>-5</td>
<td>2</td>
<td>0.62</td>
<td>0.13</td>
<td>4.33</td>
<td>1.83</td>
<td>0.66</td>
<td>0.15</td>
<td>9.32</td>
<td>5.55</td>
<td>0.54</td>
<td>0.09</td>
<td>1.52</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>-1</td>
<td>4</td>
<td>0.77</td>
<td>0.16</td>
<td>4.23</td>
<td>1.36</td>
<td>0.82</td>
<td>0.16</td>
<td>9.23</td>
<td>4.17</td>
<td>0.54</td>
<td>0.09</td>
<td>1.52</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>-1</td>
<td>4</td>
<td>0.65</td>
<td>0.17</td>
<td>1.34</td>
<td>0.35</td>
<td>0.74</td>
<td>0.17</td>
<td>1.49</td>
<td>0.67</td>
<td>0.54</td>
<td>0.09</td>
<td>1.52</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>OU</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
<td>0.51</td>
<td>0.09</td>
<td>0.54</td>
<td>0.16</td>
<td>0.54</td>
<td>0.09</td>
<td>0.44</td>
<td>0.3</td>
<td>0.54</td>
<td>0.16</td>
<td>0.54</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>OU</td>
<td>0.5</td>
<td>0</td>
<td>2</td>
<td>0.51</td>
<td>0.13</td>
<td>0.53</td>
<td>0.12</td>
<td>0.59</td>
<td>0.13</td>
<td>0.45</td>
<td>0.22</td>
<td>0.54</td>
<td>0.13</td>
<td>0.59</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>OU</td>
<td>0.5</td>
<td>0</td>
<td>4</td>
<td>0.51</td>
<td>0.13</td>
<td>0.54</td>
<td>0.12</td>
<td>0.59</td>
<td>0.13</td>
<td>0.45</td>
<td>0.22</td>
<td>0.54</td>
<td>0.13</td>
<td>0.59</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>OU</td>
<td>0.5</td>
<td>0</td>
<td>4</td>
<td>0.51</td>
<td>0.23</td>
<td>0.03</td>
<td>0.59</td>
<td>0.19</td>
<td>0.14</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the Blomberg K (for multiple traits, the mean across traits is given). All statistics are based on 1000 simulations in each case.

Figure 1. Under a speciational model trait evolution on imbalanced trees, sets of species that maximize PD harbour less FD than random sets of species of the same size. The figure presents rFD (the relative amount of FD captured by the convex hull measure between the maxPD set and random sets of species) as a function of tree imbalance (as measured by $\beta$, Y-axis) and the degree of speciational vs. gradual evolution (as measured by Pagel $\kappa$, X-axis). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10 different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree. Results are based on sets of 8 species out of 64 (tree size) and two traits. The two specific positions ‘i’ and ‘ii’ drawn on the figure refers to the parameter space position of the examples presented in figure 2, panels a-d and e-h, respectively. The tree presented in Figure 2a (corresponding to the position marked by ‘i’ in the present figure) has an imbalance of $\beta=3.5$. 

$$rFD = \frac{FD_{PD_{max}}}{FD_{PD_{max}} + FD_{rd}}$$

$FD_{PD_{max}}$ means the maximal PD value of the convex hull between maxPD and random sets, $FD_{rd}$ means the PD value of random sets.
Figure 2. Examples of cases where the set of species that maximize PD harbours more (a-d) or less (e-h) FD than a random set of species. For each example, the original phylogenetic tree (panels a and d), the position of species in trait space and their corresponding convex hull (panels b and f), the relationship between phylogenetic and trait distances (panels c and g) and the corresponding relative amount of FD between PD_{max} set and random sets are given (panels d and h). Example (a-d) corresponds to a BM model on a relatively balanced tree while example (e-h) corresponds to a speciational model (Pagel κ = 0) on a fully imbalanced tree. Both examples are also reported in figure 1, but note that here, for the purpose of simplicity, we used a tree with only 16 species from which four species were selected.
SUPPLEMENTAL MATERIAL

Conserving phylogenetic diversity can be a poor strategy for conserving functional diversity

Florent Mazel*, Arne Mooers¹, Giulio Valentino Dalla Riva¹, and Matthew W. Pennell²

¹ Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada
² Department of Zoology, Biodiversity Research Centre, University of British Columbia, Vancouver, BC V6T 1Z4, Canada;

*flo.mazel@gmail.com

List of supplemental Materials

Supplemental Tables.

Supplemental Table 1. For common trait macroevolution models, sets of species that maximize PD always harbour, on average, at least as much FD (Convex hull measure) as random sets of species of the same size.

Supplemental Table 2. For common trait macroevolution models, sets of species that maximize PD always harbour, on average, at least as much FD (Rao quadratic entropy measure) as random sets of species of the same size.

Supplemental Figures.

Figure S1. Extension of our results to multiple traits and varying tree sizes and selected species number using Convex Hull as a measure of FD.

Figure S2. Extension of our results to multiple traits and varying tree sizes and selected species number using Rao quadratic entropy as a measure of FD.

Figure S3. Extension of our results to discrete trait evolution.

Figure S4. Variability of the phylogenetic signal as measured by the Bloomberg K.
Figure S5. Variability of the phylogenetic signal as measured by the Spearman correlation between trait and phylogenetic distances.

Supplemental Table 1. For common trait macroevolution models, sets of species that maximize PD always harbour, on average, at least as much FD (Convex hull measure) as random sets of species of the same size. The table presents, for each combination of macroevolutionary models (column 1), specific set of parameters (column 2-3), number of independent traits (column 4), tree size (column 5) and number of selected species (column 6) a measure of the relative amount of FD between PD... and random sets of species for pure birth Yule trees (column 7-8) and coalescent trees (column 11-12). The comparison of FD (as measured
by the convex hull measure) between the two sets of species is quantified with the following metric: \( \frac{FD_{\text{conv}}}{(FD_{\text{conv}} + FD_{\text{rand}})} \). A value <.5 means PD is doing worse than random, a value >.5 means PD is doing better than random and a value of .5 means PD is doing the same as random. The phylogenetic signal for Yule trees (column 9-10) and coalescent trees (column 13-14) is measured with the Bloomberg K (for multiple traits, the mean across traits is given). All statistics are based on 1000 simulations in each case.

<table>
<thead>
<tr>
<th>Evolutionary Model</th>
<th>Traits and Species</th>
<th>Yule Tree</th>
<th>Type of Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>( \alpha )</td>
<td>( r )</td>
<td>Measured Traits</td>
</tr>
<tr>
<td></td>
<td>( \beta_{\text{conv}} )</td>
<td>( \beta_{\text{rand}} )</td>
<td>( \beta_{\text{conv}} )</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>3</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>4</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>5</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>6</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>6</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>7</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>8</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>9</td>
<td>64</td>
</tr>
<tr>
<td>BM</td>
<td>0</td>
<td>10</td>
<td>52</td>
</tr>
<tr>
<td>EB</td>
<td>0</td>
<td>10</td>
<td>64</td>
</tr>
</tbody>
</table>

Supplemental Table 2. For common trait macroevolution models, sets of species that maximize PD always harbour, on average, at least as much FD (Rao quadratic entropy measure) as random sets of species of the same size. The table presents, for each combination of macroevolutionary models (column 1), specific set of parameters (column 2-3), number of independent traits (column 4), tree size (column 5) and number of selected species (column 6) a measure of the relative amount of FD between PD_\text{conv} and random sets of species for pure birth
Yule trees (column 7-8) and coalescent trees (column 11-12). The comparison of FD (as measured by the Rao quadratic entropy) between the two sets of species is quantified with the following metric: $FD_{wcoor}/(FD_{wcoor} + FD_{maxPD})$. A value <.5 means PD is doing worse than random, a value >.5 means PD is doing better than random and a value of .5 means PD is doing the same as random. The phylogenetic signal for Yule trees (column 9-10) and coalescent trees (column 13-14) is measured with the Bloomberg K (for multiple traits, the mean across traits is given). All statistics are based on 1000 simulations in each case.
Figure S1. Extension of our results to multiple traits and varying tree sizes and selected species number using Convex Hull as a measure of FD. The figure presents the variation of a measure of the relative amount of FD (as measured by the convex Hull measure) between PD-maximized and random set of species (see legend) in function of tree imbalance (as measured by $\beta$, Y-axis, “fullbal” refers to fully balanced tree and “fullImbal” refers to fully imbalanced tree) and the degree of speciational vs. gradual evolution (as measured by Pagel $\kappa$, X-axis). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10
different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree. Each panel corresponds to a different set of parameters (tree size (n), selected number of species (m) and number of traits).

Figure S2. Extension of our results to multiple traits and varying tree sizes and selected species number using Rao quadratic entropy as a measure of FD. The figure presents the variation of a measure of the relative amount of FD (as measured by Rao quadratic entropy).
between PD-maximized and random set of species (see legend) in function of tree imbalance (as measured by $\beta$, Y-axis, “fullbal” refers to fully balanced tree and “fullImbal” refers to fully imbalanced tree) and the degree of speciational vs. gradual evolution (as measured by Pagel $\kappa$, X-axis). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10 different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree. Each panel corresponds to a different set of parameters (tree size (n), selected number of species (m) and number of traits).

2 Traits

Sigma=0.1  m=8  n=64

Sigma=1   m=8  n=64

Pagel Kappa

$0.44$ $0.48$ $0.52$ $0.56$

FD[pd]/(FD[pd]+FD[rd])
**Figure S3. Extension of our results to discrete trait evolution.** The figure presents the variation of a measure of the relative amount of FD (as measured by the number of character state combinations) between PD-maximized and random set of species (see legend) in function of tree imbalance (as measured by β, Y-axis, “fullbal” refers to fully balanced tree and “fullImbal” refers to fully imbalanced tree) and the degree of speciational vs. gradual evolution (as measured by Pagel κ, X-axis). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10 different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree. Results are based on sets of 8 species out of 64 (tree size). The different panels correspond to different values of the transition rate parameter of the Markov model.
Figure S4. Variability of the phylogenetic signal as measured by the Bloomberg K. The figure presents the variation of phylogenetic signal across the parameter space presented on figure 1 of the main text. Mean Bloomberg K statistic (see legend) are presented in function of tree imbalance (as measured by $\beta$, Y-axis, “fullbal” refers to fully balanced tree and “fullImbal” refers to fully imbalanced tree) and the degree of speciation vs. gradual evolution (as measured
by Pagel κ, X-axis). Each panel corresponds to a different set of parameters (tree size (n), selected number of species (m) and number of traits). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10 different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree.

Figure S5. Variability of the phylogenetic signal as measured by the Spearman correlation between trait and phylogenetic distances. The figure presents the variation of phylogenetic
signal across the parameter space presented on figure 1 of the main text. Mean mantel (spearman) statistic (see legend) are presented in function of tree imbalance (as measured by β, Y-axis, “fullbal” refers to fully balanced tree and “fullImbal” refers to fully imbalanced tree) and the degree of speciational vs. gradual evolution (as measured by Pagel κ, X-axis). Each panel corresponds to a different set of parameters (tree size (n), selected number of species (m) and number of traits). The color of each grid cell reflects the mean value of the metric over 100 trait simulations on 10 different trees (for a total of 1000 simulations) or, in the case of fully balanced and fully imbalanced trees, 1000 simulations on one single tree.