HAL open science

Phylogenetic conservation prioritization with uncertainty

Alain Billionnet

To cite this version:

Alain Billionnet. Phylogenetic conservation prioritization with uncertainty. Biodiversity and Conservation, 2018, 27 (12), pp.3137-3153. 10.1007/s10531-018-1593-z . hal-02465032

HAL Id: hal-02465032

https://hal.science/hal-02465032v1

Submitted on 7 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Phylogenetic Conservation Prioritization with Uncertainty

Alain Billionnet

Laboratoire CEDRIC, École Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise, 1, square de la Résistance, 91025 Évry cedex, France; E-mail: Alain.Billionnet@ensiie.fr

Abstract. We consider a set of species *S* and are interested in the assessment of the subsets of *S* from a phylogenetic diversity viewpoint. Several measures can be used for this assessment. Here we have retained phylogenetic diversity (PD) in the sense of Faith, a measure widely used to reflect the evolutionary history accumulated by a group of species. The PD of a group of species *X* included in *S* is easy to calculate when the phylogenetic tree associated with *S* is perfectly known but this situation is rarely verified. We are interested here in cases where uncertainty regarding the length of branches and the topology of the tree is reflected in the fact that several phylogenetic trees are considered to be plausible for the set *S*. We propose several measures of the phylogenetic diversity to take account of the uncertainty arising from this situation. A natural problem in the field of biological conservation is to select the best subset of species to protect from a group of threatened species. Here, the best subset is the one that optimizes the proposed measures. We show how to solve these optimal selection problems by integer linear programming. The approach is illustrated by several examples.

Keywords: Biodiversity conservation, phylogenetic diversity, uncertainty, protected species, optimization.

1. Introduction

The risk of extinction affecting a large number of species threatens biodiversity (Stork 2010). Many causes can be identified for this alarming situation (Collen et al. 2013). The international community agrees that there is an urgent need to try to halt the decline of biodiversity, but the resources available are of course limited (Convention on Biological Diversity 2011). It is therefore essential to have criteria for selecting the species to be protected as a priority. Many measures have been proposed to assess the value of protecting a given set of species (see, for example, (Moreno et al. 2017) and also the very extensive book by MacLaurin and Sterelny (2008) on this subject). In this article we have retained phylogenetic diversity, a measure widely used to reflect the evolutionary history accumulated

by a community of species. The reader can refer to (Maclaurin & Sterelny 2008, Chap.7 and Pellens & Grandcolas 2016) for a very comprehensive presentation of the role of phylogenetic diversity, in large sense, in the field of biological conservation. Many measures of phylogenetic diversity have been proposed (Vellend et al. 2010; Lean and Maclaurin 2016). We have retained here the measure proposed by Faith (1992a, 1992b). This measure that we note PD is easy to use when the phylogenetic tree (branch lengths and their arrangement in the tree) related to the set of species considered is perfectly known but this is an abstraction. Although many studies involving the notion of phylogeny rely on a single tree, uncertainty is inevitably present. It has many causes (Donoghue and Ackerly1996; Swenson 2009; Rangel et al. 2015; Nipperess 2016). Thus, uncertainty regarding the length of branches and the topology (branching pattern) of the tree can be reflected in the fact that several phylogenetic trees are plausible for a given set of species (Jetz et al. 2012; Arponen and Zupan 2016; Nipperess 2016). As Collen (2015) notes, establishing how best to make robust decisions with limited and uncertain information is an important avenue of research. The crucial question discussed here is how to estimate the phylogenetic diversity of a group of species by taking into account the multiple trees that are associated with this group. In an attempt to answer this question, we propose several measures derived directly from Faith's phylogenetic diversity and discuss the pros and cons of these measures in considering uncertainty. For a group of species *X*, these measures are based on the aggregation of the PD of *X*, calculated on the different trees envisaged. Indeed, if there is no strong argument in favor of one of the trees, all must be taken into account (Mimouni et al. 2016). Techniques for aggregating criteria to ultimately achieve a single criterion are widely used in many disciplines. For a given set of species, the values of the suggested measures are easy to calculate. This is not the case if one tries to determine, among a given set of species, the subset that optimizes these values while respecting specific constraints such as a budgetary constraint. We show that this optimal selection problem can be easily solved for these different measures by integer linear programming. Several examples are used to illustrate the approach.

2. Phylogenetic diversity

A rooted phylogenetic tree associated with a set of species *S* can be considered as a directed tree, $T = (V, A, S, \lambda)$, where *V* is the set of nodes (or vertices), *A*, the set of arcs – or directed edges or branches –, *S*, the set of leaves representing the species, and λ, the vector of branch lengths. The branches of *A* connect two nodes and are directed from the root of *T* to its

leaves. A vertex i is said to be a successor of a vertex j if an arc of A goes from j to i and, conversely, i is said to be a predecessor of j if an arc of A goes from i to j. The root of T is the only vertex that has no predecessor and the leaves of T have no successor. The vertices of T that are neither root nor leaf have a single predecessor and at least two successors. For all $a \in A$, we denote by L_a the set of species for which there is a path from the terminal node of a to these species. In other words, the survival of any species of L_a permits to preserve all the evolutionary history represented by the path of T going from the root to the leaf representing this species. The phylogenetic diversity created by Faith (1992a) and used here is by definition equal to $\sum_{a \in A, L_a \cap X \neq \varnothing} \lambda_a$, for any set X included in S. Intuitively, it represents the total amount of evolutionary history embodied in the set of species X, since the time of the most recent common ancestor of the set (Faith 1992a; Chao et al. 2010). For example, the PD of the set of species $\{s_2, s_4, s_6\}$ calculated from phylogenetic tree 1 in Figure 1 is equal to 3.11. The reader is referred to (Maclaurin and Sterelny 2008; Faith 2013, 2016; Collen 2015; Pellens and Grandcolas 2016) for a comprehensive presentation of the concepts of PD and especially of its relevance in the field of biological conservation. The phylogenetic diversity requires the precise knowledge of the phylogenetic tree associated with the species concerned but this tree is generally not easy to establish. As mentioned in the introduction, we are interested in defining the phylogenetic diversity of a subset of species X included in S when there are uncertainties on the phylogenetic tree associated with S. These uncertainties are reflected in the existence of several phylogenetic trees that can be envisaged for the set of species S. In the following we propose several ways to evaluate the phylogenetic diversity of a subset X of S taking into account this multiplicity of trees. The proposed approaches apply as well in the case where a probability cannot be associated with each tree as in the opposite case (for example, equiprobable trees). However, it is assumed that there are no strong enough arguments to eliminate some trees. The final objective is to propose a method for selecting, under constraints and in a context of uncertainty, the 'best' set of species to be protected on the basis of the phylogenetic diversity criterion.

3. Phylogenetic diversity measures with uncertainty on the phylogenetic tree

We consider a set of species $S = (s_1, s_2, ..., s_m)$ and a set of n possible phylogenetic trees for these species, $T = (T_1, T_2, ..., T_n)$. We note M the set of indices $\{1,...,m\}$ and N, the set of indices $\{1,...,n\}$. Each tree T_t of T is represented by the quadruplet (V^t, A^t, S, λ^t) where V^t is

the set of nodes, A^t, the set of arcs , *S*, the set of leaves, and λ^t, the vector of branch lengths. For any subset *X* of *S*, we note $PD_t(X)$ the PD of *X* in the tree T_t. We are faced with the question of being able to quantify in a certain way the phylogenetic diversity of a group of species by taking into account the uncertainties on the phylogenetic tree associated with these species. Below we propose different ways of expressing this phylogenetic diversity.

3.1. Average and weighted average phylogenetic diversity (avPD and wavPD)

One way to take into account the multiplicity of trees to assess the phylogenetic diversity of a group of species is to study the mean of its PDs on all trees (Nipperess 2016, Section Future directions). This is a very classic measure. We denote by *avPD*(*X*) the average phylogenetic diversity of a group of species *X* included in *S*. It can be expressed as follows:

$$avPD(X) = \frac{1}{n}\sum_{t\in N} PD_t(X).$$

Considering the weighted mean of the PD on all possible trees is a way of making one tree more important than another. We denote it by *wavPD*(*X*) and it can be expressed as follows:

$$wavPD(X) = \frac{\sum_{t\in N} w_t PD_t(X)}{\sum_{t\in N} w_t},$$

where w_t denotes the weight assigned to the phylogenetic tree T_t. Note that if w_t=1 for any *t* then *avPD*(*X*)=*wavPD*(*X*). The advantage of avPD and wavPD lies in their simplicity, but they have a lot of drawbacks. The measure avPD is in fact the mathematical expectation of phylogenetic diversity that would be obtained by attributing an equal probability to all trees. Similarly wavPD is the mathematical expectation corresponding to the probability $w_t / \sum_{t\in N} w_t$ attributed to the tree T_t ($t \in N$). As is well known, an important disadvantage of these measures is that they are strongly influenced by extreme values of the PD. Moreover they allow compensation between a bad score and a good score. Thus, a set of species that optimizes an average phylogenetic diversity value can be very bad from the point of view of its PD calculated on some trees. Note that if the uncertainty regarding the phylogenetic tree concerned only the length of the branches constituting this tree, we could be interested, for any set of species *X* included in *S*, in the mathematical expectation of the PD of *X* provided that we could associate a probability to each possible length of the different branches (Nipperess 2016).

3.2. Robust phylogenetic diversity (rPD)

When uncertainty affects the phylogenetic tree associated with a set of species, a robust measure of phylogenetic diversity can be investigated. The term *robust solutions* encompasses a set of solutions designed to protect the decision-maker against uncertainty. Several robustness measures can be used. For a very complete presentation of the concept of robustness in the presence of several scenarios, the reader can refer to (Kouvelis and Yu 1997). In this section, we focus on a common robust measure that we note rPD. It is a very prudent measure which ensures that the PD of any subset of species X included in S is at least equal to the rPD of X regardless of the 'true' phylogenetic tree associated with S (among the trees considered). This measure is therefore very conservative since it takes into account the worst-case scenario. In practice, it consists of calculating the phylogenetic diversity of X in each tree and then retaining the lowest of the values obtained. We note it $rPD(X)$. It can be expressed as follows:

$$rPD(X) = \min_{t \in N} PD_t(X) .$$

This measure is pessimistic because, for a given group of species, it retains the worst situation (in terms of phylogenetic diversity). We will see in the following that searching for a set of species $X \subseteq S$ that maximizes $rPD(X)$ is like searching for a set of species $X \subseteq S$ that performs well regardless of the 'true' phylogenetic tree associated with S. This is the way the robust solution protects against uncertainty. This measure, which is interesting whatever the probabilities associated with each tree, is particularly useful in the case where all the considered trees are equiprobable. Note that the rPD measure is in the same vein as the "phylogenetic risk" approach advocated by Faith (2008). The difference lies in the fact that in (Faith 2008) the uncertainty associated with the conservation of a set of species comes from the extinction probabilities of these species whereas here the uncertainty comes from the fact that several phylogenetic trees are plausible for this set of species.

We could also define another measure that is a compromise between the pessimistic measure, rPD, which we have just seen, and an optimistic measure that would consist in retaining, for a set of species X in S, its maximum PD on all trees (Hurwicz's criterion). This measure is therefore a weighted average of extreme consequences. Given a pessimistic coefficient $\alpha \in [0,1]$ it is written:

$$\alpha \min_{t \in N} PD_t(X) + (1-\alpha) \max_{t \in N} PD_t(X) .$$

3.3. Ordered weighted average phylogenetic diversity (owaPD)

We saw in the previous section a measure associated with a set of species X included in S that took into account the worst situation (associated with the phylogenetic tree providing the lowest value of $PD_t(X)$) and also another measure that took into account both the worst situation and the best situation. We now propose to use another measure that takes into account all situations in some way. Ordered weighted averaging (OWA) operators were introduced by Yager (1998) as a tool for information aggregation. In the case we are interested in, this operator provides a measure of X that takes into account first the lowest value of $PD_t(X)$ then the value that immediately follows and so on until the best value is taken into account. We denote this measure by $owaPD(X)$. Let $w_1,w_2,\ldots,w_n$ be a decreasing set of weights lying in the unit interval and summing to one. The calculation of $owaPD(X)$ can be performed as follows. First, calculate $PD_t(X)$ for any $t \in N$, then multiply the weight w_t with the lowest t'th value obtained and finally add all of it. This measure can be expressed as follows:

$$owaPD(X) = \sum_{t \in N} w_t PD_{\sigma(t)}(X)$$

where $\sigma(t)$ denotes the index of the tree corresponding to the lowest t'th value of the set $\{PD_t(X) : t \in N\}$. The value of $owaPD(X)$ is between the minimum and the maximum values of $PD_t(X)$. Compared to rPD, owaPD is less conservative since it enables trade-offs between several scenarios. Note however that owaPD includes rPD as a special case (w_1=1 and $w_2=w_3=\ldots=w_n=0$). One of the difficulties that arise in using owaPD is the definition of weights $w_1,w_2,\ldots,w_n$. The meaning of these weights is indeed very different from that of the weights used in wavPD. This last measure violates the requirement of impartiality as it assigns the weights to the specific trees in order to reflect the importance of each tree and the trees are known (Ogryczak and Sliwinski 2003). On the contrary, the weight involved in owaPD reflect the importance given to the worst value, the one that comes immediately after and so on, but neither these values nor the tree from which they come are known. The reader can refer to the following references for a presentation of OWA operators, their interest and their use in different applications: (Yager and Kacprzyk 1997; Calvo et al. 2002; Llamazares 2007; Olender 2016).

3.4. Highest guaranteed phylogenetic diversity for δ trees (h_δPD)

A value *v* for this measure means that there are δ trees of *T* for which $PD_t(X)$ takes at least the value *v*. Since we are interested in the highest possible value, *v* is therefore equal to the (*n*-δ+1)'th lowest value of the set $\{PD_t(X) : t \in N\}$. This measure therefore does not take into account the worst (*n*-δ) values. It is denoted by $h_\delta PD(X)$ and can be expressed as follows:

$$h_\delta PD(X) = \max_{\substack{E \subseteq \{1,2,\ldots,n\} \\ |E| = \delta}} \min_{t \in E} PD_t(X) = \min_{t \in E_\delta} PD_t(X)$$

where E_δ denotes the δ indices belonging to *N* and corresponding to the highest δ values of $PD_t(X)$. In other words, for δ trees the phylogenetic diversity of *X* is at least equal to $h_\delta PD(X)$. This measure can be considered as a special case of owaPD provided that the weights $w_1, w_2, \ldots, w_n$ are allowed not to be necessarily decreasing. If $w_1 = 0, \ldots, w_\delta = 0, w_{\delta+1} = 1, w_{\delta+2} = 0, \ldots, w_n = 0$ then $h_\delta PD(X)$=*owaPD*(*X*). In the particular case where δ=*n*, $h_\delta PD(X)$=*rPD*(*X*). Note that in the case where a probability can be associated with each tree T_t of *T*, the probability that the phylogenetic diversity of a set $X \subseteq S$ is greater than or equal to $h_\delta PD(X)$ is greater than or equal to the sum of the probabilities associated with the trees whose indices belong to E_δ (for example, δ/n when all the trees are equiprobable).

3.5. Largest gap from maximum phylogenetic diversity (lgapPD)

This measure which is part of the robust measures is by nature slightly different from previous ones. Consider a set of species, *X*, included in *S* and verifying a set of constraints *C*. This measure involves the highest PD that can be associated with a set of species included in *S* and satisfying *C*, and this for each tree considered. To evaluate a set of species *X* included in *S* (and verifying *C*) with this measure, the difference between the PD of *X* and the highest PD value that could be obtained on the same tree for a set of species included in *S* and satisfying *C* is calculated for each tree. The largest of these differences is then retained. This measure is denoted by *lgapPD*(*X*) and can be expressed as follows:

$$lgapPD(X) = \max_{t \in T} \left\{ PD_C^{*t} - PD_t(X) \right\}$$

where PD_C^{*t} is equal to the maximum PD of a set of species included in *S* and satisfying *C*, in tree T_t. Thus, for any *X* included in *S* and verifying *C*, the distance between the PD of *X* and

the maximum PD of a set of species included in S and satisfying C is guaranteed in each tree insofar as this distance is less than or equal to $lgapPD(X)$. This measure is therefore based on a regret concept.

We will denote by $\alpha PD(X)$ these 6 measures (α represents *av*, *wav*, *r*, *owa*, h_δ, *lgap*)

4. Example

Consider the 4 hypothetical phylogenetic trees in Figure 1. These trees are associated with the set of 6 species $\{s_1,s_2,s_3,s_4,s_5,s_6\}$. They were generated with the sole purpose of illustrating, as best as possible, the PD measures presented in the previous section. Table 1 shows the $PD_t(X)$ values in each tree as well as the 6 $\alpha PD(X)$ values when $X=\{s_1,s_2,s_3\}$. In this example $lgapPD(X)$ is calculated assuming that the constraints C express only the fact that 3 species out of 6 must be selected.

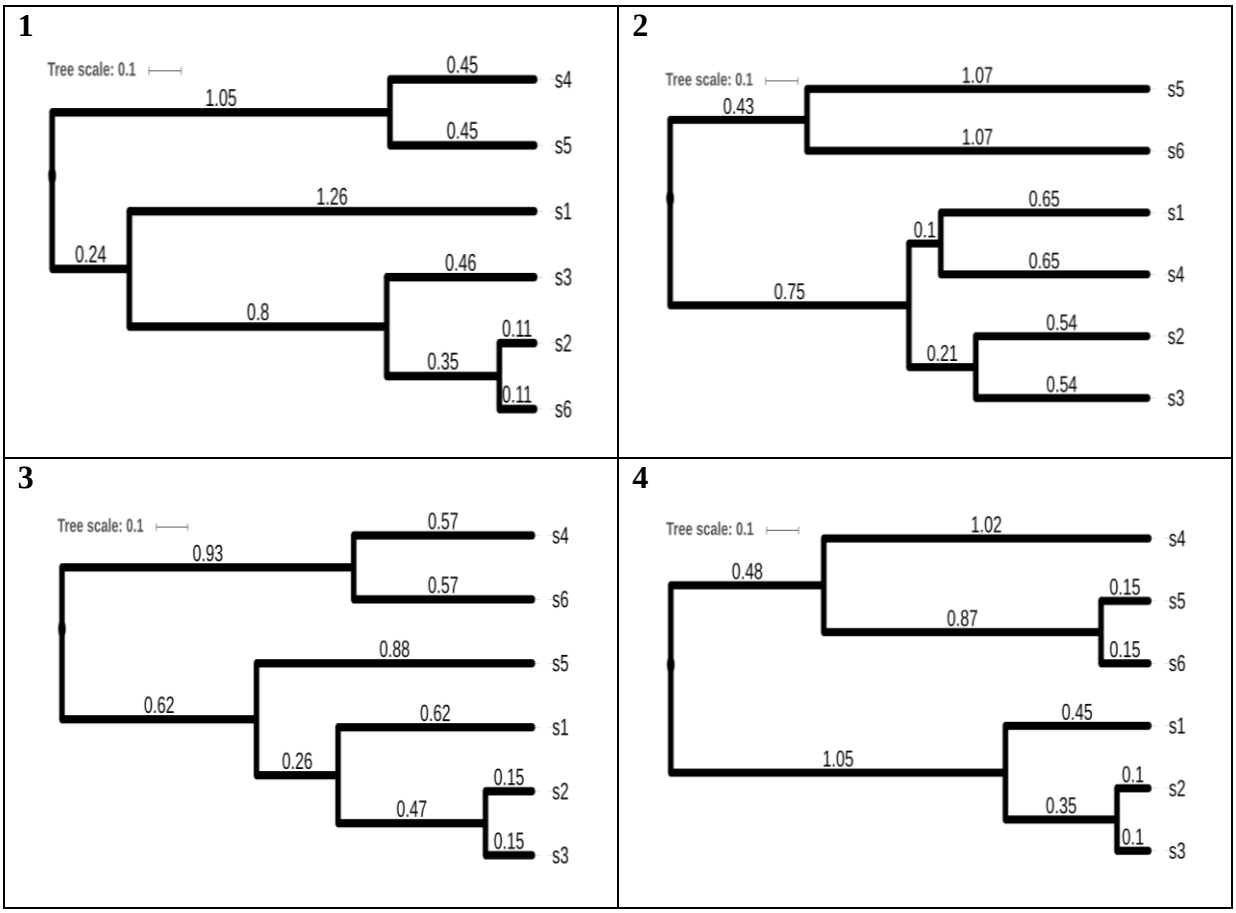

Figure 1. Four hypothetical ultrametric phylogenetic trees associated with 6 species (drawn with iTOL).

Table 1. PDs of the set $\{s_1,s_2,s_3\}$ in the four trees of Figure 1 and αPDs.

PD Tree 1	PD Tree 2	PD Tree 3	PD Tree 4	**avPD**	**wavPD** w=(0.5,0.2,0.2,0.1)	**owaPD** w=(0.4,0.3,0.2,0.1)	**rPD**	**h_2PD**	**lgapPD**
3.22	2.79	2.27	2.05	2.5825	2.827	2.381	2.05	2.79	1.97

5. Application to optimization problems

As noted in the introduction, $\alpha PD(X)$ is easy to calculate for a given set of species X included in S. On contrast, determining a set X included in S that optimizes $\alpha PD(X)$ under some constraints is much more difficult. We show in the following that integer linear programming is an effective tool to address this issue. We will focus on the following 4 classic selection problems summarized by Chernomor et al. (2016) in the context of split diversity:

Problem I (Species Selection): Given a set of n phylogenetic trees for m species, find a subset X of k species that optimizes $\alpha PD(X)$ over all subsets of k species.

Problem II (Budgeted Species Selection): Given a set of n phylogenetic trees for m species and a conservation cost for each species, find a subset of species X whose total conservation costs do not exceed a predefined budget while optimizing $\alpha PD(X)$ over all subsets of species whose total conservation costs do not exceed this predefined budget.

Problem III (Reserve Selection): Given a set of n phylogenetic trees for m species distributed in p areas, find a subset of k areas that optimizes $\alpha PD(U)$ over all subsets of k areas where U is the set consisting of all species present in at least one of the k selected areas.

Problem IV (Budgeted Reserve Selection): Given a set of n phylogenetic trees for m species distributed in p areas and conservation costs for each area, find a subset of areas whose total conservation costs do not exceed a predefined budget while optimizing $\alpha PD(U)$ over all subsets of areas whose total conservation costs do not exceed this predefined budget where U is the set consisting of all species present in at least one of the selected areas.

Optimization means maximization for α=av, wav, owa, r, h_δ and minimization for α=lgap.

Given a set of possible phylogenetic trees associated with a set of species S, it is very easy to calculate $\alpha PD(X)$ for α=*av*, *wav*, *owa*, *r*, h_δ, for any subset X of S. For α=*lgap* the calculation is a little more complicated since it is first necessary to determine, for each tree, the maximum

PD of a set of species verifying some constraints. For example, these constraints may express that the set of species must correspond to the species present in k areas. This is an optimization problem in itself, which can be more or less difficult to solve. If constraints only express the fact that k species must be selected then the problem can be solved easily by a greedy algorithm (Hartmann & Steel 2006, Pardi 2009). The algorithm begins with an empty set, X, and sequentially adds to X the species, s_i, which maximizes the PD of the set $X \cup \{s_i\}$ until the cardinal of X is k.

Many conservation-relevant issues include determining, under specific constraints, a subset of species X included in S that is optimal with respect to phylogenetic diversity. We are interested here in the problems of determining sets of species that, under some constraints, optimize αPD (Problems I-IV). These problems are generally difficult. Indeed, the method which would consist in calculating $\alpha PD(X)$ for all the subsets X of S respecting the constraints is not practicable as soon as the number of species in X and/or S is large. We propose here a method based on integer linear programming to determine, under some constraints, the set X included in S which optimizes $\alpha PD(X)$.

6. Formulation of Problems I-IV by integer linear programs

6.1. Optimizing avPD, wavPD, rPD, h$_\delta$PD and lgapPD

Let us recall that $T = (T_1, T_2, ..., T_n)$ is the set of phylogenetic trees that has to be considered for the set of species concerned, that A^t is the set of arcs of T_t and that λ^t is the vector of the branch lengths of T_t. More precisely, $\lambda^t = \{\lambda_a^t : a \in A_t\}$ where λ_a^t denotes the length of the branch a in the tree T_t. L_a^t refers to the set of species (leaves) under the arc a in the tree T_t. In other words, the survival of at least one species of L_a^t preserves the evolutionary history of the arc a if, however, T_t is the 'true' phylogenetic tree. The cost associated with the selection of species s_i will be designated by c_i. We will note $R=\{R_1,R_2,...,R_p\}$ the set of areas considered, C_j the cost associated with the selection of the area R_j and B the available budget. We will note $P=\{1,\ldots,p\}$. Finally, b_{ij} is a parameter that is equal to 1 if species s_i is present in the area R_j and 0 if not. In order to formulate Problems I-IV by integer linear programs, we will use the following variables:

$x_i \in \{0,1\}$ $(i \in M)$: $x_i = 1$ if and only if species s_i is selected;

$z_a^t \in \{0,1\}$ $(t \in N, a \in A^t)$: $z_a^t = 1$ if and only if the set of selected species allows to preserve the evolutionary history of the arc a in the tree T_t;

$y_j \in \{0,1\}$ $(j \in P)$: $y_j = 1$ if and only if area R_j is selected;

$\alpha \geq 0$: α is a working variable for calculating the minimum of several quantities;

$\beta_t \in \{0,1\}$ $(t \in N)$: β_t is a working variable allowing to take into account or not the tree T_t in calculating the optimal value of $h_\delta PD$;

$PD_t \geq 0$ $(t \in N)$: phylogenetic diversity, calculated in the tree T_t, of the set of selected species, i.e. of the set $\{s_i : i \in M, x_i = 1\}$.

Let us now examine the different constraints that will be used to formulate Problems I-IV with the 6 different measures proposed to assess phylogenetic diversity.

(C_0): $z_a^t = \max_{i \in L_a^t} x_i$ $(t \in N, a \in A_t)$. The Boolean variable z_a^t must take the value 1 if and only if at least one of the species of L_a^t is selected. This constraint is not linear, we will replace it with the linear constraint C_1 below;

(C_1): $z_a^t \leq \sum_{i \in L_a^t} x_i$ $(t \in N, a \in A_t)$. In the case of maximizing z_a^t, this constraint is equivalent to C_0;

(C_2): $PD_t = \sum_{a \in A_t} \lambda_a^t z_a^t$ $(t \in N)$;

(C_3): $\sum_{i \in M} x_i = k$. The number of species selected must be equal to k;

(C_4): $\sum_{i \in M} c_i x_i \leq B$. The total cost of the selected species should not exceed the available budget, B;

(C_5): $x_i \leq \sum_{j \in P} b_{ij} y_j$ $(i \in M)$. Species x_i can only be selected (x_i=1) if at least one of the areas where it occurs is itself selected;

(C_6): $\sum_{j \in P} y_j = k$ .The number of areas selected must be equal to k;

(C_7): $\sum_{j \in P} C_j y_j \leq B$. The total cost associated with the selected areas should not exceed the available budget, B.

Once these variables and constraints are defined most of the mathematical programs that solve Problems I-IV, with $\alpha PD(X)$ as a measure of the phylogenetic diversity, are easy to establish. These programs when α is equal to *av*, *wav*, *r*, h_δ and *lgap* are given in Table 2.

Table 2. Integer linear programs associated with Problems I-IV when phylogenetic diversity is measured by αPD with α = av, wav, r, h_δ and lgap.

avPD and wavPD	rPD	h_δPD	lgapPD
$\max \frac{1}{n}\sum_{t\in N} w_t PD_t$ subject to:	$\max \alpha$ subject to: $\alpha \le PD_t \quad (t \in N)$	$\max \alpha$ subject to: $\alpha \le PD_t + H\beta_t \quad (t \in N)$ $\sum_{t\in N} \beta_t = n-\delta$	$\min \alpha$ subject to: $\alpha \ge PD_C^{*t} - PD_t \quad (t \in N)$
	C_1,C_2,C_3 (Problem I) C_1,C_2,C_4 (Problem II) C_1,C_2,C_5,C_6 (Problem III) C_1,C_2,C_5,C_7 (Problem IV)		

For avPD and wavPD the objective function, $(1/n)\sum_{t\in N} w_t PD_t$, expresses wavPD for the set of selected species and avPD in the particular case where w_t=1 for all $t \in N$. It should be noted that in all these programs, the set of species selected corresponds either to a set of directly selected species or to a set of species whose selection results from the selection of a set of areas. For rPD, according to constraint $\alpha \le PD_t \ (t \in N)$ and considering that α is the objective function to be maximized, α will take, at the optimum, the smallest of the *n* values $PD_t \ (t \in N)$, i.e. the smallest of the *n* PD values (in the *n* trees considered) associated with the set of selected species. In the program associated with h_δPD, *H* is a sufficiently large constant that allows to make inactive the constraints in which β_t=1. The objective is to maximize the variable α. Because of constraints $\alpha \le PD_t + H\beta_t \ (t \in N)$, the value of α will be equal, at the optimum, to the smallest of the quantities $PD_t + H\beta_t \ (t \in N)$ but, according to constraint $\sum_{t\in N} \beta_t = n-\delta$, n-δ variables β_t can take the value 1. In order to maximize α, the values of β_t which will take the value 1 are those whose index corresponds to the smallest n-δ values of $PD_t \ (t \in N)$. At the optimum, α will therefore be equal to the value that comes immediately after. Recall that in the program associated with lgapPD, PD_C^{*t} denotes the maximum PD of a feasible set of species (directly selected or resulting from the selection of

areas), calculated in the tree T_t. According to constraint $\alpha \geq PD_C^{*t} - PD_t \ (t \in N)$ and considering that α is the objective function to be minimized, α will take, at the optimum, the largest of the n values $PD_C^{*t} - PD_t \ (t \in N)$, i.e. the largest deviation (in the n trees considered) we are interested in.

6.2. Maximizing owaPD

Problems I-IV are somewhat more difficult to formulate by a mathematical program when the phylogenetic diversity of a set of species X is measured by *owaPD*(X). Here we show how to formulate the optimization of *owaPD*(X) by an integer linear program using the results established by Ogryczak and Sliwinski (2003) for linear programs with OWA objective functions. Let us use the Boolean variable $e_{ij} \ (i \in N, j \in N)$ which takes the value 1 if and only if the weight w_i is assigned to the value of the PD of X in the tree T_j. For a set of species X included in S and fixed, *owaPD*(X) is equal to the optimal value of the linear program in Boolean variables $\mathrm{P}_1(X)$.

$$\mathrm{P}_1(X): \begin{cases} \min \sum_{j \in N} w_j \sum_{j \in N} PD_i(X) e_{ij} \\ \text{s.t.} \left| \begin{array}{lll} \sum_{j \in N} e_{ij} = 1 & i \in N & (a) \\ \sum_{i \in N} e_{ij} = 1 & j \in N & (b) \\ e_{ij} \in \{0,1\} & i \in N, j \in N & (c) \end{array} \right. \end{cases}$$

Indeed, given the objective function to minimize, the decreasing weights $w_1, w_2, \ldots, w_n$ will be assigned to the PDs of X sorted in increasing order, respectively. The objective function thus expresses well, at the optimum of $\mathrm{P}_1(X)$, owaPD for the selected species. It is known that in this type of program the integrality constraints (c) can be relaxed, i.e. replaced by the constraints $0 \leq e_{ij} \leq 1 \ (i \in N, j \in N)$ and finally by the constraints $0 \leq e_{ij}$ since constraints (a) and (b) prevent the e_{ij} variables from exceeding value 1. Consider the dual program of the program thus obtained by associating to constraints (a) and (b) the real variables μ_i and ν_i, respectively. The mathematical program $\mathrm{P}_2(X)$ is obtained, whose optimal value is equal to that of $\mathrm{P}_1(X)$.

$$\mathrm{P}_2(X):\begin{cases}\max \sum_{i\in N}(\mu_i+\nu_i) \\ \text{s.t.} \left|\begin{array}{ll}\mu_i+\nu_j \le w_j PD_i(X) & i\in N. j\in N \\ \mu_i\in R, \nu_i\in R & i\in N\end{array}\right.\end{cases}$$

To solve the 4 problems under consideration, it is sufficient to express, in $\mathrm{P}_2(X)$, $PD_i(X)$ as a function of variables x_i, for problems I and II, and also as a function of variables y_i for problems III and IV. The resulting formulations are presented below by program P_3.

$$\mathrm{P}_3:\begin{cases}\max \sum_{i\in N}(\mu_i+\nu_i) \\ \text{s.t.} \left|\begin{array}{ll}\mu_i+\nu_j \le w_j PD_i & i\in N, j\in N \\ \mu_i\in R, \nu_i\in R & i\in N \\ \mathrm{C}_1,\mathrm{C}_2 & \\ \mathrm{C}_3 \text{ (if Problem I), } \mathrm{C}_4 \text{ (if Problem II),} & \\ \mathrm{C}_5,\mathrm{C}_6 \text{ (if Problem III), } \mathrm{C}_5,\mathrm{C}_7 \text{ (if Problem IV)} & \end{array}\right.\end{cases}$$

7. Example

Let us return to the 4 phylogenetic trees from Figure 1 and consider Problems I (resp. III) in which we seek to select a set of 3 species (resp. 2 areas) that optimizes αPD for α=r, owa, h_2 and lgap. For Problem III, we assume that 4 areas R_1, R_2, R_3 and R_4 are concerned and that the 6 species s_1, s_2, s_3, s_4, s_5 and s_6 are spread over these 4 regions as follows: s_1 and s_2 in R_1, s_3 and s_4 in R_2, s_4 and s_5 in R_3 and s_5 and s_6 in R_4. The optimal solutions and their values are presented in Table 3. This table also presents, for comparison purposes, the solutions to Problems I and III when each tree is considered separately. In other words, we give for problem I (resp. III) and for each of the 4 trees, the set of 3 species (resp. 2 areas) that maximizes the PD in the tree considered. This table also shows the worst selection of species or areas that can be made based on one of the 4 phylogenetic trees while the 'true' phylogenetic tree is another of these 4 trees. For example, let us look at problem III with the rPD criterion. With this criterion, the optimal solution is to select areas R_1 and R_3. This selection ensures a PD of 4.40 regardless of the 'true' tree among the 4 trees considered. This solution is the best, i.e. there are no other selections of 2 areas that ensure a PD strictly greater than 4.40 on the 4 trees. Row 2 of Table 3 indicates that selecting 2 optimal areas based on only one of the trees could lead to a rather poor solution if this tree is not the right one. In the

worst case, the PD obtained would be equal to 3.44. In this case, the optimal solution for tree 1 ($\{R_1, R_2\}$) was chosen while tree 2 is the 'true' tree. Choosing the solution of rPD max is therefore a strong protection against uncertainty since it ensures a PD of 4.40 in all cases while, as we have just seen, a bad choice of the 'true' tree results in a PD of 3.44 (about 22% less). Note that, in this small example, the solution that maximizes rPD is also the solution that minimizes the greatest deviation, for each tree, from the solution of PD max. For this solution, the largest of these deviations is equal to 0.42 whereas if the solution of maximum PD for tree 1 ({R1, R2}) is retained, the largest of these deviations is equal to 1.38.

Table 3. Problems I and III on the 6-species instance described by Figure 1: Solutions optimizing PD and αPD (α = r, owa, h_2 and lgap) and the worst solutions that can be retained due to a bad choice of the 'true' tree.

	Problem I 6 species, 4 trees, 3 species to be selected			Problem III 6 species, 4 trees, 4 areas, 2 areas to be selected		
1. Solutions of PD max for each tree	Sets of species of PD max for each tree, and their PD:	$\{s_1,s_3,s_5\}$ $\{s_1,s_5,s_6\}$ $\{s_1,s_4,s_5\}$ $\{s_1,s_4,s_5\}$	4.26 4.07 3.88 4.02	Sets of areas of PD max for each tree, and their PD:	$\{R_1,R_2\}$ $\{R_1,R_4\}$ $\{R_1,R_3\}$ $\{R_1,R_3\}$	4.72 4.82 4.50 4.47
2. Worst PD value that could be obtained	Select the set of species of PD max for tree 2,$\{s_4,s_5,s_6\}$, when tree 4 is the 'true' tree. PD obtained for tree 4:		2.67	Select the set of areas of PD max for tree 1, $\{R_1,R_2\}$, when tree 2 is the 'true' tree. PD obtained for tree 2:		3.44
3. Largest error that could be committed	Select the set of species of PD max for tree 2, $\{s_4,s_5,s_6\}$, when tree 4 is the 'true' tree. Error committed for tree 4:		1.35	Select the set of areas of PD max for tree 1, $\{R_1,R_2\}$, when tree 2 is the 'true' tree. Error committed for tree 2:		1.38
4. Solutions of rPD max	Set of species of rPD max and its value: $\{s_1,s_4,s_6\}$ PD of the set of rPD max, for each tree: 4.26, 3.65, 3.57, 4.02		3.57	Set of areas of rPD max and its value: $\{R_1,R_3\}$ PD of the set of rPDmax, for each tree: 4.71, 4.40, 4.50, 4.47		4.40
5. Solutions of owaPD max with w=(0.4,0.3,0.2,0.1)	Set of species of owaPD max and its value: $\{s_1,s_4,s_6\}$ PD of the set of owaPD max, for each tree: 4.26, 3.65, 3.57, 4.02		3.753	Set of areas of owaPD max, and its value: $\{R_1,R_3\}$ PD of the set of owaPD max, for each tree: 4.71, 4.40, 4.50, 4.47		4.472
6. Solutions of h_δPD max with δ=2	Set of species of h_δPD max and its value: $\{s_1,s_5,s_6\}$ PD of the set of h_δPD max, for each tree 4.26, 4.07, 3.88, 3.15		4.07	Set of areas of h_δPD max and its value: $\{R_1,R_4\}$ PD of the set of h_δPD max, for each tree 4.37, 4.82, 4.50, 3.60		4.50
7. Solutions of lgapPD min	Set of species of lgapPD min, and the gap: $\{s_1,s_4,s_6\}$ PD of the set of species of lgapPD min, for each tree: 4.26, 3.65, 3.57, 4.02		0.42	Set of areas of lgapPD min, and the gap: $\{R_1,R_3\}$ PD of the set of areas of lgapPD min, for each tree: 4.71, 4.40, 4.50 4.47		0.42

In order to test the applicability of the proposed approaches, we tested them on a large sized instance with 10 phylogenetic trees covering 100 species. We have generated many

hypothetical rooted ultrametric binary trees with a pure birth model based on Yule's process (Yule 1924; Steel and McKenzie 2002; Steel and Mooers 2010; Mooers et al. 2012). In this model the time that a given lineage persists until it speciates has an exponential distribution with a given rate. This random time provides the length of the corresponding branch. For this experiment, the birth rate was set at 1 and the time elapsed from the root to the leaves (species) was set at $\log(|S|/2) = 3.91$ units. We start with a rooted tree composed of a root and its two successor nodes and repeat the following procedure until the time elapsed from the root to each pendant node (node without successor) is at least 3.91: select the pendant node for which the time separating it from the root is minimum and make it the direct ancestor of two new descendent nodes. The length of the two corresponding branches is drawn at random following the exponential distribution of rate 1. When the procedure has been completed, the pendant nodes (leaves) represent the species to be considered. The length of all the pendant branches (connected to a leaf) is then adjusted so that the time elapsed from the root to the present moment is exactly equal to 3.91. We thus obtain an ultrametric tree whose number of species can be different from 100. We then selected 10 trees, out of all the trees produced, covering exactly 100 species. These 100 species are numbered from 1 to 100 in the order in which they are generated by the process described above. These 10 trees therefore reflect the phylogenetic uncertainty regarding the 100 species considered. These 100 species are considered to be distributed over 10 areas as follows: area R_j is home to species s_i if and only if the double inequality, $1+10(j-1.3) \le i \le 1+10j$, is verified. Thus, area R_1 is home to species $s_1,...,s_{11}$, area R_2 is home to species $s_8,...,s_{21}$, area R_3 is home to species $s_{18},...,s_{31}$, etc. We solved problems I (optimal selection of 25 species) and III (optimal selection of 4 areas) with the αPD criterion when α=r, owa, h_δ and lgap. Results are presented in Table 4. Let us look at some of the results of Table 4 for Problem I. First, we see that a wrong tree choice can lead to a very poor solution. Indeed, if we retain tree 4 to determine the best set of species while tree 10 is the 'true' tree, we obtain a PD equal to only 30.05 (line 2 of table 4). On the other hand, the robust solution (the one that maximizes rPD, line 4 of Table 4) ensures a PD of 51.73 (about 72% more) whatever the 'true' tree (among the 10 considered). If one looks at the gap on each tree between the maximum PD and the PD of the robust solution, the largest of these gaps is 61.68-52.56=9.12; it corresponds to tree 3. In addition, the solution that minimizes the largest of these gaps (line 7 of Table 4) results in a gap of only 3.77. However, the corresponding solution can lead to a PD of 49.29 (tree 10). For problem III and in these experimental conditions, the error that can be committed by choosing the worst tree is

relatively limited since it is equal to slightly less than 10% (line 3 of Table 4). The solutions of maximum rPD or minimum lgapPD in this case bring only a slight improvement. Arponen and Zupan (2016) had already observed, in a specific case concerning 275 European terrestrial mammals, that uncertainty about phylogenies may not have much impact with regard to the prioritization of evolutionary hotspots. But, as these authors point out, this result is not general. It may also depend on the conservation objective.

Table 4. Problems I and III on a 100-species instance: Solutions optimizing PD and αPD (α = r, owa, h_6 and lgap) and worst solutions that can be retained due to a bad choice of the 'true' tree.

	Problem I 100 species, 10 trees, 25 species to be selected	Problem III 100 species, 10 trees, 10 areas, 4 areas to be selected
1. Solutions of PD max, for each tree	PD max for each tree: 58.37, 60.38, 61.68, 56.68, 54.97, 53.79, 55.28, 61.94, 53.83, 52.54	PD max for each tree: 72.59, 73.78, 82.42, 71.17, 69.88, 66.94, 71.88, 80.89, 76.16, 65.99
2. Worst PD value that could be obtained	Select the set of species of PD max for tree 4 when tree 10 is the 'true' tree. PD obtained: 30.05	Select the set of areas of PD max for tree 8 when tree 10 is the 'true' tree. PD obtained: 62.93
3. Largest error that could be committed	Select the set of species of PD max for tree 8 when tree 1 is the 'true' tree. Error committed: 26.67	Select the set of areas of PD max for tree 1 when tree 8 is the 'true' tree. Error committed: 7.37
4. Solutions of rPD max	rPD max: 51.73 PD of the set of species of rPD max, for each tree: 53.72, 55.29, 52.56, 51.81, 51.81, 51.76, 51.73, 53.93, 51.74, 51.83	rPD max: 65.99 PD of the set of areas of rPD max, for each tree: 71.83, 71.66, 79.80, 70.06, 67.89, 66.94, 71.88, 75.21, 71.62, 65.99
5. Solutions of owaPD max with w=(0.35,0.2,0.15,0.1, 0.06,0.05,0.03,0.025, 0.02,0.015)	owaPD max: 52.1012 PD of the set of species of owaPD max, for each tree: 53.13, 53.86, 54.60, 52.71, 53.07, 52.60, 52.41, 56.91, 52.26, 50.93	owaPD max: 68.2073 PD of the set of areas of owaPD max, for each tree: 71.83, 71.66, 79.80, 70.06, 67.89, 66.94, 71.88, 75.21, 71.62, 65.99
6. Solutions of h_δPD max with δ=6	h_δPD max: 55.28 PD of the set of species of h_δPD max, for each tree 55.42, 55.60, 56.62, 55.44, 50.38, 48.55, 55.28, 55.71, 43.68, 44.65	h_δPD max: 71.62 PD of the set of areas of h_δPD max, for each tree 71.83, 71.66, 79.80, 70.06, 67.89, 66.94, 71.88, 75.21, 71.62, 65.99
7. Solutions of lgapPD min	gap: 3.77 PD of the set of species of lgapPD min, for each tree 54.77, 56.92, 58.18, 53.05, 52.41, 50.07, 51.51, 58.17, 51.20, 49.29	gap : 3.03 PD of the set of areas of lgapPD min, for each tree: 70.69, 72.12, 82.42, 70.83, 69.88, 64.67, 68.85, 79.54, 74.43, 64.06

We solved another instance of Problem III by increasing the number of areas from 10 to 20 (Table 5). The 100 species are distributed over the 20 areas as follows: area R_j is home to species s_i if and only if the double inequality, $1+5(j-1.3) \le i \le 1+5j$, is verified. In this case,

the error that can be committed by choosing the worst tree is much larger (about 17%) as shown by the results presented in Table 5.

Table 5. Problem III on a 100-species instance: worst solutions that can be retained due to a bad choice of the 'true' tree.

	Problem III 100 species, 10 trees, 20 areas, 4 areas to be selected
1. Solutions of PD max for each tree	PD max for each tree: 54.76, 55.13, 58.55, 49.20, 50.11, 49.29, 53.81, 55.98, 51.12, 48.05
2. Worst PD value that could be obtained	Select the solution of PD max for tree 4 when tree 10 is the 'true' tree. PD obtained: 40.74
3. Largest error that could be committed	Select the solution of PD max for tree 4 when tree 1 is the 'true' tree. Error committed: 9.10

8. Concluding remarks

In this article, we are interested in measuring the phylogenetic diversity of a group of species in the presence of uncertainties on the phylogenetic tree associated with these species. Many assessments of a group of species based on phylogenetic diversity, regardless of the precise meaning given to this concept, assume that the phylogenetic tree of these species is perfectly known. This is of course a simplifying hypothesis. Here, uncertainty is reflected in the fact that several trees are possible. Assessing the phylogenetic diversity of a group of species in the presence of uncertainty about the associated phylogenetic tree is a complex issue. We show on examples that this assessment can be quite distant from the true assessment if we are mistaken about the correct tree. We propose several measures taking into account the fact that several trees are plausible. These measures are easy to calculate. They aggregate the phylogenetic diversities, in the sense of Faith, calculated on each tree. We then focus on the natural and classical problem of selecting species (or areas) to maximize the associated phylogenetic diversity. For this purpose, decision makers may use any of the proposed measures according to their objectives. They can, for example, adopt a very conservative approach by selecting the group of species that minimizes the risk (on the PD value) due to an error on the tree. Instead, they can adopt a more optimistic approach. For the suggested measures, solving the selection problem is a difficult problem. We show that it can be easily formulated by integer linear programming. The problem is therefore easy to solve if an integer

linear programming software is available. Several examples based on hypothetical phylogenetic trees are presented to illustrate the proposed approaches. A natural extension of this study would be to consider extinction probabilities for each species considered (while maintaining the assumption that several trees are plausible) and to replace the notion of phylogenetic diversity with the well-known notion of expected phylogenetic diversity.

References

Arponen, A., Zupan, L. 2016. Representing hotspots of evolutionary history in systematic conservation planning for European mammals. In: Pellens R., Grandcolas P. (Eds), Biodiversity Conservation and Phylogenetic Systematics, Topics in Biodiversity and Conservation 14, Springer.

Calvo, T., Mayor, G., Mesiar, R. (Eds.). 2002. Aggregation operators: New trends and applications. Physica-Verlag, Heidelberg.

Chao, A., Chiu, C.-H., Jost, L. 2010. Phylogenetic diversity measures based on Hill numbers. Phil. Trans. R. Soc. B 365, 3599-3609.

Chernomor, O., Klaere, S., von Haeseler, A., Minh, B.Q. 2016. Split Diversity: Measuring and Optimizing Biodiversity Using Phylogenetic Split Networks. In: Pellens R., Grandcolas P. (Eds), Biodiversity Conservation and Phylogenetic Systematics, Topics in Biodiversity and Conservation 14, Springer.

Collen, B., Pettorelli, N., Baillie, J.E.M., Durant, S.M. (Eds). 2013. Biodiversity monitoring and conservation: Bridging the gap between global commitment and local action. Wiley-Blackwell, Hoboken.

Collen, B. 2015. Conservation prioritization in the context of uncertainty. Animal Conservation, 18, 315-317.

Convention on Biological Diversity. 2011. Strategic plan for biodiversity 2011-2020, including Aichi biodiversity targets. Available at: http://www.cbd.int/sp/.

Donoghue, M.J., Ackerly, D.D. 1996. Phylogenetic Uncertainties and Sensitivity Analyses in Comparative Biology. Phil. Trans. R. Soc. Lond. B , 351, 1241-1249.

Faith, D.P. 1992a. Conservation evaluation and phylogenetic diversity. Biol. Cons., 61, 1-10.

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

Faith, D.P. 2008. Threatened species and the potential loss of phylogenetic diversity: Conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. Conservation Biology, 22, 1461-1470.

Faith, D.P. 2013. Biodiversity and evolutionary history: useful extensions of the PD phylogenetic diversity assessment framework. Annals of the New-York Academy of Sciences, 1289, 69-89.

Faith, D.P. 2016. The PD phylogenetic diversity framework: linking evolutionary history to feature diversity for biodiversity conservation. In: Pellens R., Grandcolas P. (Eds), Biodiversity Conservation and Phylogenetic Systematics, Topics in Biodiversity and Conservation 14, Springer.

Hartmann, K., Steel, M. 2006. Maximizing phylogenetic diversity in biodiversity conservation: greedy solutions to the Noah's Ark Problem. Syst. Biol., 55, 644-651.

Isaac, N.J.B., Turvey, S.T., Collen, B., Waterman, C., Baillie, J.E.M. 2007. Mammals on the EDGE : Conservation priorities based on threat and phylogeny. PloS ONE 2 (3): e296.

iTOL, version 4.0.3. Interactive Tree Of Life. An online tool for the display, annotation and management of phylogenetic trees. Available at https://itol.embl.de/.
Jetz, W., Thomas, G.H., Joy, J.B, Hartmann, K., Mooers, A.O. 2012. The global diversity of birds in space and time. Nature, 491, 444-448.

Kouvelis, P., Yu, G. 1997. Robust discrete optimization and its applications. The Netherlands: Kluwer Academic Publishers.

Lean, C., Maclaurin, J. 2016. The Value of Phylogenetic Diversity. In: Pellens R., Grandcolas P. (Eds), Biodiversity Conservation and Phylogenetic Systematics, Topics in Biodiversity and Conservation 14, Springer.

Llamazares, B. 2007. Choosing OWA operator weights in the field of Social Choice. Information Sciences, 177, 4745-4756.

MacLaurin, J., Sterelny, K. 2008. What Is Biodiversity? The University of Chicago Press, Chicago, 217 p.

Mimouni, E.-A., Beisner, B.E., Pinel-Alloul, B. 2016. Phylogenetic diversity and its conservation in the presence of phylogenetic uncertainty: a case study of cladoceran communities in urban waterbodies. Biodiversity and Conservation, 25, 2113-2136.

Mooers, A., Gascuel, O., Stadler, T., Li, H., Steel, M. 2012. Branch lengths on birth-death trees and the expected loss of phylogenetic diversity. Syst. Biol. 61, 195-203.

Moreno, C.E. et al. 2017. Measuring biodiversity in the Anthropocene: a simple guide to helpful methods. Biodiversity and Conservation, 26, 2993-2998.

Nipperess, D.A. 2016. The rarefaction of phylogenetic diversity: Formulation, extension and application. In: Pellens R., Grandcolas P. (Eds), Biodiversity Conservation and Phylogenetic Systematics, Topics in Biodiversity and Conservation 14, Springer.

Ogryczak, W., Sliwinski, T. 2003. On solving linear programs with the ordered weighted averaging objective. Eur. J. of Operat. Res.,148, 80-91.

Olender, P. 2016. Hybrid Models for the OWA Optimization. Journal of Telecommunications and Information Technology, 4, 22-30.

Pardi, F. 2009. Algorithms on phylogenetic trees. PhD thesis, University of Cambridge.

Pellens, R., Grandcolas, P. (Eds). 2016. Biodiversity conservation and phylogenetic systematics: preserving our evolutionary heritage in an extinction crisis. Springer International Publishing, Springer Open, 390 p.

Rangel, T.F., Colwell, R.K., Graves, G.R., Fučíková, K., Rahbek, C., Diniz-Filho, J.A.F. 2015. Phylogenetic uncertainty revisited: Implications for ecological analyses. Evolution, 69, 1558-5646.

Steel, M., McKenzie, A. 2002. The "shape" of phylogenies under simple random speciation models. Proceedings of Biological Evolution and Statistical Physics Conference, Lässig, M., Valleriani, A. (Eds), LNP 585, Springer-Verlag, 162-180.

Steel, M., Mooers, A. 2010. The expected length of pendant and interior edges of a Yule tree. Applied Mathematics Letters, 23, 1315-1319.

Stork, N.E. 2010. Re-assessing current extinction rates. Biodivers. Conserv., 19, 357-371.

Swenson, N.G. 2009. Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities. PLoS One 4:e4390.

Vellend, M., Cornwell, W.K., Magnuson-Ford, K., Mooers, A.Ø. 2010. Measuring phylogenetic biodiversity. In: Magurran, A.E., McGill, B.J. (Eds.), Biological diversity: frontiers in measurement and assessment. Oxford University Press, Oxford.

Yager, R.R. 1988. On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics, 18, 183-190.

Yager, R.R., Kacprzyk, J. (Eds.). 1997. The Ordered Weighted Averaging Operators: Theory and Applications. Kluwer Academic Publishers, Dordrecht.

Yule, G.U. 1924. A mathematical theory of evolution: based on the conclusions of Dr. J.C. Willis. Philos. Trans. R. Soc. Lond. Ser. B 213, 21-87.