Modelling Phylogenetic Diversity[*]

KLAUS NEHRING

Department of Economics, University of California at Davis

Davis, CA 95616, U.S.A.

and

CLEMENS PUPPE

Department of Economics, University of Bonn

Adenauerallee 24-42, D – 53113 Bonn, Germany

Revised, September 2002

Abstract *The paper proposes a model for measuring and valuing biodiversity based on evolutionary information, called the phylogenetic tree model. While avoiding the strong restrictions of Weitzman's (1992) "cladistic" approach, the phylogenetic tree model retains much of the mathematical simplicity of the cladistic model. In particular, in the phylogenetic tree model the diversity of any set of species can be recursively determined from the pairwise dissimilarities between them. The restrictions imposed by the phylogenetic tree model on the underlying dissimilarity metric are characterized and shown to be weaker than those entailed by the cladistic model. An especially parsimonous version of the phylogenetic tree model is obtained by invoking an appropriate notion of translation invariance.*

1   Introduction

This paper addresses the problem of the modelling of biodiversity. From an economic perspective, the central task is to provide ways of constructing society's preferences over different conservation policies. From a biological perspective, the issue is to develop applicable measures of biodiversity. In "A Theory of Diversity" (Nehring and Puppe (2002), henceforth TD (2002)), we have developed a *multi-attribute approach* to valuing and measuring diversity. Its basic idea is to think of the diversity of a set of entities as derived from the number and weight of the different *attributes* possessed by them. Measures of diversity that are based (explicitly or implicitly) on the general idea of counting attributes ("features," "characteristics") have been proposed frequently in the literature (see, among others, Vane-Wright, Humphries and Williams (1991), Faith (1992), Weitzman (1992, 1998), Solow, Polasky and Broadus (1993), Faith and Walker (1994), Williams and Humphries (1996), Tilman (1997), Weikard (2002), and the volumes edited by Gaston (1996) and Polasky (2001)). One goal of TD (2002) is to formalize this idea in a general and unified framework. The multi-attribute approach allows for various different interpretations of the "entities" whose diversity is measured. These may be genes, individual organisms, species, or ecosystems. Depending on the chosen level of description, the relevant attributes will be different. In the case of species, examples of relevant attributes are "suckling their young," "living for more than 1000 years," "feeding on molluscs," or "descending from archeopteryx." Examples of attributes of ecosystems are "containing a particular species," or "containing a particular set of complementary species" such as particular predator-prey relationships. In this paper, we will be concerned with *species* diversity of a particular kind.

One needs to clearly distinguish the level of measurement from the level of conservation action. As to the latter, there is a broad consensus that effective biodiversity conservation policies need to be directed at reserve sites or even entire ecosystems; frequently, it is impractical, or even outright infeasible, to target single species in isolation. Nonetheless, it will often be the conservation of species (not reserve sites) that one ultimately cares about. This two-pronged approach is taken, for instance, by Ando, Camm, Polasky and Solow (1998) (see also Armsworth, Kendall and Davis (2002) for an overview). Note that on this view, the ecological interrelations among species that determine the outcomes of conservation policies are analytically separate from the assessment of the resulting biodiversity. In the language of economics, the former are part of the "technology" of biodiversity conservation, while the latter is a matter of societal preferences (see Section 2.6 below for an elaboration).

In a biological context, one can distinguish three main types of species diversity: genetic diversity, functional diversity, and phylogenetic diversity. Here, our focus is on the latter. By phylogenetic diversity we mean the morphological diversity of species based on their evolutionary descent. Our empirical starting point is thus the evolutionary tree describing the genealogy of species. The relevant attributes are "homologies," i.e. shared characteristics inherited from a common ancestor, in contrast to "homoplasies," i.e. shared characteristics due to the adaptation to a common way of life. For instance, the similarity between a birds' wing and a human arm represents a homology (the inherited but differently adapted reptilian forelimb), while the similarity between the wing of a bird and the wing of a bat corresponds to a homoplasy.

One of the first workable models of diversity based on evolutionary information was developed by Weitzman (1992). Effectively, his model assumes that all relevant

1

attributes correspond to the characteristics shared by *all* species in the same "clade," i.e. by all species that descend from a common ancestor. Thus, we will refer to Weitzman's model as the *cladistic model*. The major restriction of this model derives from the fact that all cladistic attributes "live forever:" once a species has founded a clade, all descendants share its cladistic characteristics, by definition. This entails, for instance, that later species are superior to earlier ones in an extreme form: a species has *no* marginal diversity value as long as one of its successors survives. A related criticism has been put forward by Faith (1994) who argued that, implicitly, the cladistic model assumes a constant speed of evolution.

To overcome these criticisms, we propose here the "phylogenetic tree model" as a natural generalization of Weitzman's cladistic model, retaining much of its mathematical simplicity. The main new feature of the phylogenetic tree model is its ability to account for non-cladistic attributes such as those derived from the "Linnean" taxonomy of species. Indeed, the present paper can be viewed as an argument in favor of a neo-linnean broadening of taxonomy (for a plea for "Neolinnean Impressionism," see Tudge (2000)).

The paper is organized as follows. In the following Section 2, we briefly review the basic concepts and tools of the multi-attribute approach introduced in TD (2002). In particular, we define the notion of a diversity function as an appropriate aggregator of attribute weights. A diversity function naturally induces pairwise dissimilarities between species. We also introduce the concept of "monotonicity in dissimilarity," which requires the diversity of an arbitrary set of species to be determined by the pairwise dissimilarities between its elements in a monotone fashion. The section concludes with a brief discussion on the economic interpretation of diversity theory.

In Section 3, we turn to evolutionary trees and discuss two polar ways of modelling diversity based on evolutionary information. As a model of minimal specifity, we introduce the "tree model" and show that it is too unstructured to serve as a useful model of biodiversity. We then review Weitzman's cladistic model which represents the other extreme of maximal specifity. The cladistic model is charaterized by a simple condition on the family of relevant attributes, the "Nesting Property," which states that any two attributes are either completely unrelated or one is more special than the other. As a consequence of this property, diversity in the cladistic model can be determined using a simple recursion formula. However, the cladistic model offers too little flexibility since the Nesting Property imposes very strong restrictions on the family of admissible attributes. We then discuss Faith's (1992) frequently used model of phylogenetic diversity. While that model overcomes the weaknessnes of the cladistic model, it entails other problematic restrictions, as shown in Subsection 3.4.

In Section 4, we present our proposal for modelling biodiversity, the phylogenetic tree model which allows one to combine evolutionary with taxonomic similarity information. As the main result, we characterize the phylogenetic tree model in terms of a qualitative compatibility restriction on the family of relevant attributes, called the "Weak Nesting Property." It can be used as an empirical criterion to verify the applicability of the model. A preliminary discussion suggests that the Weak Nesting Property may stand up well empirically. We also show that phylogenetic tree models admit a simple representation via an (ordinal) index of "remoteness from the origin."

A corollary of the characterization in terms of the Weak Nesting Property is the recursive character and monotonicity in dissimilarity of the phylogenetic tree model, as shown in Section 5. The possibility of determining the diversity of an arbitrary

set of species from their pairwise dissimilarities has obvious advantages in practical applications, since the number of pairwise dissimilarities grows quadratically with the total number of species, while in the unconstrained multi-attribute framework the number of potential attributes grows exponentially.

Given the property of monotonicity in dissimilarity, one would like to know which dissimilarity metrics are consistent with a phylogenetic tree model. This question is addressed in Section 6, where we characterize the restrictions that the phylogenetic tree model imposes on the underlying dissimilarity metric. The key restriction is a condition of "weak ultrametricity" which relaxes the well-known ultrametricity condition of the cladistic model. Section 6 is mathematically more demanding than the rest of the paper and may be skipped without loss of continuity.

The phylogenetic tree model allows one to incorporate information about differential speed of evolution. In the final Section 7, we offer an interpretation along these lines under the name of "evolutionary clock model." We also propose an especially parsimonous version invoking a notion of translation invariance. Translation invariance amounts to identical durability patterns of attributes through the entire evolutionary tree. In particular, this allows one to reduce comparisons between species that are far apart to "local" comparisons of neighboring species. All proofs of formal results are collected in an appendix.

2   Background: The Multi-Attribute Model of Diversity

This section briefly reviews the basic concepts and tools of the general multi-attribute framework developed in TD (2002).

2.1   An Introductory Example

As a simple example, consider a universe X consisting of three distinct species: whales (wh), rhinoceroses (rh) and sharks (sh). Intuitively, judgements on the diversity of different subsets of these species will be based on their possessing different *features*. For instance, whales and rhinos possess the feature "being a mammal," while sharks possess the feature "being a fish." Let F be the totality of all features deemed relevant in the specific context, and denote by $R \subseteq X \times F$ the "incidence" relation that describes the features possessed by each object, i.e. $(x, f) \in R$ whenever object $x \in X$ possesses feature $f \in F$. A sample of elements of R in our example is thus (wh, f_{mam}), (rh, f_{mam}), and (sh, f_{fish}), where f_{mam} and f_{fish} denote the features "being a mammal" and "being a fish," respectively. The basic idea behind the multi-attribute model is to view the diversity of a set S of species as being determined by the number and the value of the different features possessed by the species in S. Specifically, for each relevant feature $f \in F$, let $\lambda_f \geq 0$ quantify the value of the realization of f. Upon normalization, λ_f can thus be thought of as the relative importance, or *weight* of feature f. The *diversity value* of a set S of species is defined as

$$v(S) := \sum_{f \in F : (x,f) \in R \text{ for some } x \in S} \lambda_f. \qquad (2.1)$$

3

Hence, the diversity value of a set of species is given by the total weight of all different features possessed by some species in S. Note especially that each feature occurs at most once in the sum. In particular, each single species contributes to diversity the value of all those features that are not possessed by any already existing species.

The relevant features can be classified according to which sets of objects possess them, as follows. First, there are all idiosyncratic features of the above species, the sets of which we denote by $F_{\{wh\}}$, $F_{\{rh\}}$ and $F_{\{sh\}}$, respectively. Hence, $F_{\{wh\}}$ is the set of all features that are possessed exclusively by whales, and analogously for $F_{\{rh\}}$ and $F_{\{sh\}}$. For instance, sharks being the only fish in this example, $F_{\{sh\}}$ contains the feature "being a fish." On the other hand, there will typically exist features jointly possessed by several objects. For any subset $A \subseteq X$ of species denote by F_A the set of features that are possessed by *exactly* the objects in A; thus, each feature in F_A is possessed by all elements of A and not possessed by any element of $X \setminus A$. For instance, whales and rhinos being the only mammals in the example, the feature "being a mammal" belongs to the set $F_{\{wh,rh\}}$. With this notation, (2.1) can be rewritten as

$$v(S) := \sum_{A \cap S \neq \emptyset} \sum_{f \in F_A} \lambda_f. \tag{2.1$'$}$$

Intuitively, any feature shared by several objects corresponds to a similarity between these objects. For instance, the joint feature "mammal" renders whales and rhinos similar with respect to their taxonomic classification. Suppose, for the moment, that the feature of "being a mammal" is in fact the only non-idiosyncratic feature deemed relevant in our example, and let λ_{mam} denote its weight. In this case, (2.1) or (2.1$'$) yield $v(\{wh, sh\}) = v(\{wh\}) + v(\{sh\})$, i.e. the diversity value of whale and shark species together equals the sum of the value of each species taken separately. On the other hand, since $v(\{wh, rh\}) = v(\{wh\}) + v(\{rh\}) - \lambda_{mam}$, the diversity value of whale and rhino species together is *less* than the sum of the corresponding individual values by the weight of the common feature "mammal." This captures the central intuition that the diversity of a set is reduced by similarities between its elements.

As in TD (2002), we will suppress explicit reference to the underlying description F of relevant features by identifying features *extensionally*. Specifically, for each subset $A \subseteq X$ denote by $\lambda_A := \sum_{f \in F_A} \lambda_f$ the total weight of all features with extension A, with the convention that $\lambda_A = 0$ whenever $F_A = \emptyset$. With this notation, (2.1$'$) can be further rewritten as

$$v(S) = \sum_{A \cap S \neq \emptyset} \lambda_A. \tag{2.1$''$}$$

The totality of all features $f \in F_A$ will be identified with their extension A, and we will refer to the subset A as a particular *attribute*. Hence, a set A viewed as an attribute corresponds to the family of all features possessed by exactly the elements of A. For instance the attribute $\{wh\}$ corresponds to the conjunction of all idiosyncratic features of whales ("being a whale"), whereas the attribute $\{wh, rh\}$ corresponds to "being a mammal."[1] The function λ that assigns to each attribute A its weight λ_A, i.e. the total weight of all features with extension A, is referred to as the *attribute weighting*

---

[1] Subsets of X thus take on a double role as sets to be evaluated in terms of diversity on the one hand, and as weighted attributes on the other. In order to notationally distinguish these roles we will always denote generic subsets by the symbol "A" whenever they are viewed as attributes, and by the symbol "S" otherwise.

function. The set of *relevant* attributes is given by the set

$$\Lambda := \{A : \lambda_A \neq 0\}.$$

An element $x \in X$ *possesses* the attribute A if $x \in A$, i.e. if x possesses one, and therefore all, features in F_A. Furthermore, say that an attribute A is *realized* by the set S if it is possessed by at least one element of S, i.e. if $A \cap S \neq \emptyset$. According to $(2.1'')$, the diversity value $v(S)$ is thus the total weight of all attributes realized by S.

2.2 Diversity Functions

A function v of the form $(2.1'')$ with $\lambda_A \geq 0$ for all A is called a *diversity function*, and we will always assume the normalization $v(\emptyset) = 0$. Clearly, any given attribute weighting function $\lambda \geq 0$ determines a particular diversity function via formula $(2.1'')$. Conversely, any given diversity function v *uniquely* determines the corresponding collection λ_A of attribute weights (see TD (2002, Fact 2.1)). In particular, any given diversity function v unambiguously determines the corresponding family Λ of relevant attributes. This basic fact allows one to describe properties of a diversity function in terms of corresponding properties of the associated attribute weighting function.

An essential property of a diversity function is that the marginal value of a species x *decreases* in the size of existing species; formally, for all S, T and x

$$S \subseteq T \Rightarrow v(S \cup \{x\}) - v(S) \geq v(T \cup \{x\}) - v(T). \tag{2.2}$$

Indeed, using $(2.1'')$, one easily verifies that

$$v(S \cup \{x\}) - v(S) = \sum_{A \ni x, A \cap S = \emptyset} \lambda_A,$$

which is decreasing in S due to the non-negativity of λ. Property (2.2), known as *submodularity*, is a very natural property of diversity; indeed, it captures the fundamental intuition that it becomes the harder for an object to add to the diversity of a set the larger that set already is.[2]

2.3 Dissimilarity

Any diversity function v naturally induces a notion of pairwise dissimilarity between species. Specifically, define the *dissimilarity from x to y* by

$$d(x, y) := v(\{x, y\}) - v(\{y\}). \tag{2.3}$$

The dissimilarity $d(x, y)$ from x to y is thus simply the marginal diversity of x in a situation in which y is the only other existing species. Using $(2.1'')$ one easily verifies that

$$d(x, y) = \sum_{A \ni x, A \not\ni y} \lambda_A,$$

that is, $d(x, y)$ equals the weight of all attributes possessed by x but not by y. Note that, in general, d need not be symmetric, and thus fails to be a proper metric; it

---

[2] A somewhat stronger property, called *total* submodularity, in fact characterizes diversity functions, see TD (2002, Fact 2.2).

does, however, always satisfy the triangle inequality. The function d is symmetric if and only if $v(\{x\}) = v(\{y\})$ for all $x, y \in X$, i.e. if and only if single species have identical diversity value. We will refer to a diversity function that gives equal value to all singletons as a *uniform* diversity function.

2.4 Models of Diversity

A non-empty family of attributes $\mathcal{A} \subseteq 2^X \setminus \{\emptyset\}$ is referred to as a *model (of diversity)*. A diversity function v is *compatible* with the model $\mathcal{A}$ if the corresponding set Λ of relevant attributes is contained in $\mathcal{A}$, i.e. if $\Lambda \subseteq \mathcal{A}$. A model thus represents a *qualitative* a priori restriction, namely that no attributes outside $\mathcal{A}$ can have strictly positive weight. For instance, in a biodiversity context, an example of such an a priori restriction would be the requirement that all relevant attributes are biological taxa, such as "being a vertebrate," "being a mammal," etc. This requirement leads to an especially simple functional form of any compatible diversity function, as we shall see below.

2.5 Diversity as Aggregate Dissimilarity

In practical applications, one will have to construct the diversity function from primitive data. One possibility is, of course, to first determine appropriate attribute weights and to compute the diversity function according to $(2.1'')$. Determining attribute weights is a complex task, however, since there are as many potential attributes as there are non-empty *subsets* of species, i.e. $2^n - 1$ when there are n species. An appealing alternative is to try to derive the diversity of a set from the pairwise dissimilarities between its elements. This is a much simpler task since, with n species, there are at most $n \cdot (n-1)$ non-zero dissimilarities. Specifically, say that a model $\mathcal{A}$ is *monotone in dissimilarity* if, for any compatible diversity function v and any S, the diversity $v(S)$ is uniquely determined by the value of all single species in S and the pairwise dissimilarities within S, and if, moreover, the diversity $v(S)$ is a monotone function of these dissimilarities.

The problem of determining the conditions under which a model is monotone in dissimilarity is one the main tasks of TD (2002). Say that a model $\mathcal{A}$ is *acyclic* if for no $m \geq 3$ there exist species $x_1, ..., x_m$ and attributes $A_1, ..., A_m \in \mathcal{A}$ such that, for all $i = 1, ..., m-1$ $A_i \cap \{x_1, ..., x_m\} = \{x_i, x_{i+1}\}$, and $A_m \cap \{x_1, ..., x_m\} = \{x_m, x_1\}$. Thus, for instance in the case $m = 3$, acyclicity requires that there be no triple of species such that each pair of them possesses an attribute that is not possessed by the third species. A main result of TD (2002) establishes that a model of diversity is monotone in dissimilarity if and only if it is acyclic.[3]

2.6 On the Economic Interpretation of Diversity Theory

From an economic perspective, the problem posed by biodiversity consists in the choice of an appropriate conservation policy, such as investment in conservation sites, restrictions of land development, anti-poaching measures, or the reduction of carbondioxid emission. This can be modelled along the following lines. A policy determines at each point of time a probability distribution over sets of existing species and consumption. Formally, a policy p can be thought of a sequence $p = (p^t)_{t \geq 0}$, where each p^t is a probability distribution on $2^X \times \mathbf{R}_+^N$ with $p^t(S^t, c^t)$ as the probability that at time t

---

[3] The necessity of acyclicity hinges on a weak regularity requirement, see TD (2002, Section 6).

the set S^t is the set of existing species and c^t is the consumption vector. Denoting by P the set of feasible policies, society's problem can thus be written as

$$\max_{p \in P} \int_0^\infty e^{-\delta t} \cdot E_{p^t}[v(S^t) + u(c^t)]dt, \qquad (2.4)$$

where δ denotes the discount rate and E_p the expectation with respect to p. The objective function in (2.4) is composed of utility from aggregate consumption $u(c^t)$ and the existence value services $v(S^t)$ from the set S^t of surviving species.

Diversity theory tries to help us determine the intrinsic value we put on the survival of different species which is represented by the function v. The probabilities p^t reflect societies expectations about the consequences of its actions; these, in turn, reflect our knowledge of economic and ecological processes. For instance, the role of keystone species that are crucial for the survival of an entire ecosystem will be captured in the relevant probability distribution. Thus, the value derived from the presence of such species *qua* keystone species enters as an indirect rather than instrinsic utility.[4]

As a simple example, consider two species y and z each of which can be saved forever (at the same cost); moreover, suppose that it is not possible to save both of them. Which one should society choose to save? Assuming constant consumption *ceteris paribus*, the utility gain at t from saving species x, given that otherwise the set S^t of species survives, is

$$v(S^t \cup \{x\}) - v(S^t) = \sum_{A \ni x, A \cap S^t = \emptyset} \lambda_A.$$

Denote by $Q^t(x) := \sum_{A \ni x} \lambda_A \cdot prob(A \cap S^t = \emptyset)$ the expected marginal value at t of saving x, which is given by the sum of the weights of all attributes possessed by x multiplied by the probability that x is the unique species possessing them. The expected present value of the utility gain from saving x is given by

$$\int_0^\infty e^{-\delta t} \cdot Q^t(x)dt.$$

For concreteness, let y be one of the few species of rhinoceroses, and z a unique endemic species which currently has a sizeable number of fairly distant relatives. In view of the fact that all rhino species are currently endangered, this leads to the following trade-off between maximizing diversity in the short-run and in the long-run. Saving the endemic species z yields a significant short-run benefit, while the expected benefit from safeguarding the last rhino species would be very high. This suggests the following qualitative behavior of the streams of intertemporal benefits accruing from the two policies:

---

[4]Alternatively, the multi-attribute framework can also be interpreted in terms of *option value*, as explained in TD (2002, p.1168). As a result, measures of biodiversity based on that notion, such as the one proposed in Polasky, Solow and Broadus (1993), also fit into our framework.
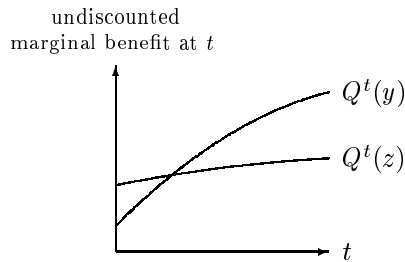
Figure 1: Streams of expected marginal benefits

The strong increase in the expected marginal value of saving y stems from the fact that, due to the limited current number of rhinos, the extinction probability of their unique attributes becomes high as t grows. Clearly, the rhino species y should be saved if the discount rate is low enough; otherwise, z should be saved. The decision thus depends on three factors: the discount rate, the value of the relevant attributes at stake, and the probability of the survival of close relatives over time.

3 Diversity Based on Evolutionary Information

We now turn to our main issue, the modelling of diversity based evolutionary information. Let X be the set of all species that ever existed, and suppose that the genealogy of species is described by a partial order $\geq_{ev}$ on X, with the interpretation that $y \geq_{ev} x$ if x is an ancestor of y. The partial order $\geq_{ev}$ is assumed to be a *tree order*, i.e. X has a minimal element (the root, denoted by x_0) and, for all $y \in X$, the set $\{x : y \geq_{ev} x\}$ of all ancestors of y is totally ordered by $\geq_{ev}$. In this case, $\geq_{ev}$ gives rise to a graphic tree τ_{ev}, in which each species is connected by edges to its immediate ancestor and to all of its immediate descendants (see Figure 2, in which $x \leq_{ev} y$).
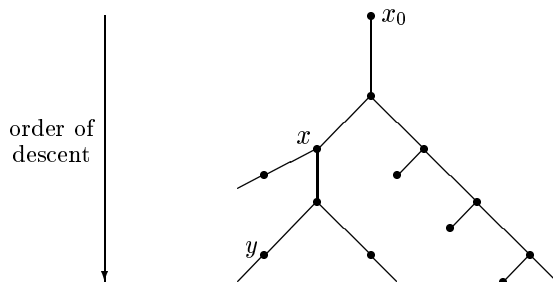

Figure 2: An evolutionary tree

3.1 The General Tree Model

A natural minimal requirement on a model of diversity based on evolutionary information is that all relevant attributes be connected in the evolutionary tree. An attribute

8

A is τ_{ev}-*connected* whenever A contains with any two species the entire shortest path in τ_{ev} connecting them. Denote by $\mathcal{T}_{ev}$ the family of all τ_{ev}-connected subsets of X; by convention, let $\emptyset \in \mathcal{T}_{ev}$. We will refer to $\mathcal{T}_{ev}$ as the *tree model* associated with $\geq_{ev}$. As is easily verified, the requirement $\Lambda \subseteq \mathcal{T}_{ev}$ amounts to the following two conditions in terms of the order $\geq_{ev}$. For all $A \in \Lambda$,

TR1 (Common Ancestor) $x, y \in A$ implies $w \leq_{ev} x$ and $w \leq_{ev} y$ for some $w \in A$.

TR2 (Unbroken Lineage) $x, y \in A$ and $x \leq_{ev} z \leq_{ev} y$ implies $z \in A$.

Condition TR1 says that any two species sharing an attribute must descend from a common ancestor with the same attribute. Condition TR2 says that an attribute once left behind by an evolutionary lineage is never recovered. Both seem to be natural requirements under a genealogical interpretation of attributes.

While plausible as a *minimal* requirement, the assumption of τ_{ev}-connectedness is arguably not enough. Indeed, the general tree model $\mathcal{T}_{ev}$ is insufficiently restrictive as a model of diversity based on evolutionary information. The reason is that the tree model $\mathcal{T}_{ev}$ yields no similarity restrictions on the set $E(X) \subseteq X$ of *extant species*. Without loss of generality, we identify the set of currently existing species with the terminal nodes of the evolutionary tree, i.e. $x \in E(X)$ if and only if x has no successor.[5] The key observation is the fact that any subset $B \subseteq E(X)$ of terminal nodes can be obtained as the intersection of some attribute $A \in \mathcal{T}_{ev}$ with B. In other words, for *any* subset B of currently existing species there is a potential attribute that is possessed by all species in B but by no other currently existing species.[6]

3.2   The Cladistic Model

As we have just argued, the general tree model $\mathcal{T}_{ev}$ yields no restrictions on the similarity relations between extant species. Since one would expect such similarity relations to hold, one has to find more specific restrictions on the family of admissible attributes reflecting these similarities. The proposal made in Weitzman (1992, 1998) can be viewed as an example of this strategy. Given the evolutionary tree τ_{ev}, Weitzman selects the family of all "clades" as the family of admissible attributes. Specifically, he proposes the model $\mathcal{H}_{ev} := \{C_x : x \in X\}$, where $C_x := \{y : y \geq_{ev} x\}$ is the *clade* founded by species x, i.e. the set of all species that descend from x. In the following, we refer to $\mathcal{H}_{ev}$ as the *cladistic model*. Clearly, any clade is τ_{ev}-connected, hence $\mathcal{H}_{ev} \subseteq \mathcal{T}_{ev}$. Indeed, the cladistic model is characterized by condition TR1 together with the following strengthening of condition TR2. For all A,

TR2* $x \in A$ and $y \geq_{ev} x$ implies $y \in A$.

Moreover, the cladistic model $\mathcal{H}_{ev}$ satisfies the following property. For all $A \in \mathcal{H}_{ev}$,

NP (Nesting Property) $A \cap B \neq \emptyset \Rightarrow [A \subseteq B \text{ or } B \subseteq A]$.

Thus, in the cladistic model any two attributes are either disjoint or one is contained in ("more specialized than") the other. Following TD (2002), we refer to attribute families satisfying this property as *hierarchical* models. Diversity in a hierarchical model admits

---

[5]Of course, there are already extinct species which also had no descendants. For our purposes such species are irrelevant and we neglect them here for simplicity.

[6]For verification, consider for any given subset B of terminal nodes the set A of all ancestors of the species in B. Clearly, $A \cap E(X) = B$ and A is τ_{ev}-connected, hence A represents an admissible attribute $A \in \mathcal{T}_{ev}$ that is possessed exactly by the set B of currently existing species.

the following simple recursion formula, originally proposed by Weitzman (1992). For all S and all v compatible with some hierarchical model,

$$v(S \cup \{x\}) - v(S) = \min_{y \in S} d(x, y). \tag{3.1}$$

Thus, the marginal diversity of a species x given the set S of already existing species is simply the minimal dissimilarity of x from any member of S. It is immediate from (3.1) that the cladistic model $\mathcal{H}_{ev}$, and more generally any hierarchical model, is monotone in dissimilarity. Indeed, (3.1) allows one to recursively determine the diversity of any set of species from their value as singletons and their pairwise dissimilarities; moreover, the diversity is monotonically increasing in the dissimilarities. In TD (2002), we show that the recursion formula (3.1) in fact characterizes the class of all hierarchical models. As a result, the recursion cannot hold for general dissimilarity metrics. The dissimilarity metrics that are consistent with a hierarchical model are the so-called "ultrametrics." In the uniform case (equal valuation of singletons), ultrametricity is the requirement that the two greatest pairwise dissimilarities between three elements must be equal.

3.3 Critique of the Cladistic Model

We now want to argue that the cladistic model is too restrictive to appropriately reflect basic features of phylogenetic diversity. As a starting point, consider the following evolutionary tree.
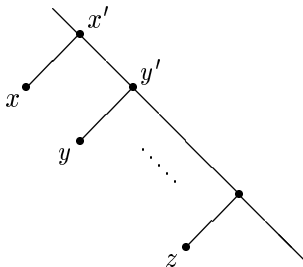

Figure 3: A restriction of the cladistic model

In the situation depicted in Figure 3, the cladistic model entails that

$$d(y, x) \geq d(y, z), \tag{3.2}$$

no matter how far z has evolved from y', its common ancestor with y. To verify this, observe that $d(y, x) = \lambda_{\{y\}} + \lambda_{C_{y'}}$, i.e. the dissimilarity of y from x equals the weight of y's idiosyncratic characteristics plus the weight of $C_{y'}$, the only proper clade of y that it does not share with x. On the other hand, $d(y, z) = \lambda_{\{y\}}$ since z belongs to any (proper) clade to which y belongs.

Empirically, (3.2) is a problematic and undesirable restriction. To see this, consider the following concrete example.
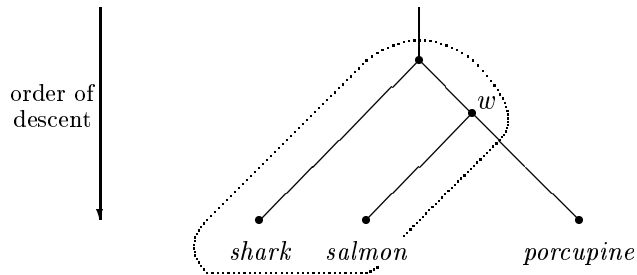
Figure 4: A segment of the evolutionary tree

In Figure 4, w denotes a common ancestor of *salmon* and *porcupine* that is not an ancestor of *shark*. In analogy to the above restriction (3.2), the cladistic model forces $d(salmon, porcupine) \leq d(salmon, shark)$. In particular, for any uniform diversity function with $\Lambda \subseteq \mathcal{H}_{ev}$ the set $\{salmon, porcupine\}$ is (weakly) less diverse than the set $\{salmon, shark\}$, which seems counterintuitive. The reason is that the cladistic model neglects the important commonality of *salmon* and *shark* that derives from their belonging to the same (non-cladistic) taxon "fish" (corresponding to the dotted attribute in Figure 4).

As another example illustrating the problematic character of (3.2), consider the following remarkably frequent pattern of speciation, the "caterpillar tree" (for the many examples of this pattern, see Tudge (2000)).
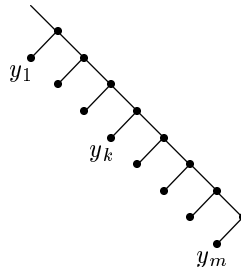

Figure 5: A caterpillar tree

In the situation depicted in Figure 5 a model should not put a priori restrictions on the relative magnitude of the dissimilarity $d(y_k, y_1)$ of the middle species y_k from y_1 as compared to the dissimilarity $d(y_k, y_m)$ of y_k to y_m. Intuitively, these dissimilarities should only depend on the relative "degree of evolvedness" of y_k. For instance, if y_k has branched off close to y_1, while y_m is much more evolved, one would expect that $d(y_k, y_1) < d(y_k, y_m)$. However, by (3.2), one again obtains an unambiguous answer by the cladistic model: $d(y_k, y_1) \geq d(y_k, y_m)$.

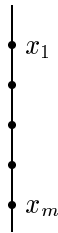As a final example, consider a simple lineage of descent, as shown in Figure 6.

11

Figure 6: A lineage of descent

Along a lineage of descent, the cladistic model implies that $d(x_k, x_l) = 0$ whenever $k \leq l$. Indeed, the descendant x_l belongs to any clade to which its ancestor x_k belongs. Since the cladistic model rules out that ancestors have any attributes not shared with all their descendants, it forces one to treat later species as superior to earlier ones in an extreme form: an earlier species has *no* marginal diversity value as long as one successor survives. This, indeed, seems to be the fundamental drawback of the cladistic model: all attributes (i.e. clades) are *immortal*. A related criticism against Weitzman's cladistic model has been put forward by Faith (1994), who argued that it implicitly assumes a constant speed of evolution. For instance, on this account the evolutionary clock has run faster on the branch from w to *salmon* in Figure 4 above than on the branch from w to *porcupine*, making *salmon* closer than *porcupine* to *shark*.[7]

Faith (1992) has proposed a measure which overcomes these difficulties and which has been widely used in practice. We will discuss the corresponding model in the following subsection and argue that it has difficulties of its own. Later, in Section 4, we will show that one can accommodate all of these critiques while maintaining the desirable property of monotonicity in dissimilarity.

3.4   The "Quasi-Cladistic" Model

Faith (1992) has proposed the following measure of biodiversity based on evolutionary information. For any set of species S, denote by $Sp(S)$ the set of all species that are on some shortest τ_{ev}-path between two species in S (i.e. the species on the subtree spanned by S). Furthermore, for any $y \neq x_0$, denote by y^- the immediate predecessor of y. The measure v_δ is defined as follows. For all S,

$$v_\delta(S) = c + \sum_{y:\{y,y^-\}\subseteq Sp(S)} \delta(y^-, y), \tag{3.3}$$

where $c \geq 0$ is a constant, and $\delta(\cdot, \cdot) \geq 0$ is an exogeneously given symmetric distance function that is *additive* in the sense that

$$\delta(x, z) = \delta(x, y) + \delta(y, z),$$

whenever y is on the path between x and z. Note that the summation in (3.3) is taken over all distances such that both y and y^- are in $Sp(S)$; in particular, $v_\delta(\{x\}) = c$ for all single species x. For any two species x and z, let $(y_1, y_2, ..., y_l)$ with $y_1 = x$ and

---

[7]A formal representation of the notion of a constant or variable "speed of evolution" is given in Section 7 below.

$y_l = z$ be a path connecting x with z in τ_{ev}. Note that this implies either $y_{i+1} = y_i^-$, or $y_i = y_{i+1}^-$. By (3.3) and symmetry of δ one obtains,

$$v_\delta(\{x, z\}) = c + \sum_{i=2}^{l} \delta(y_{i-1}, y_i),$$

and hence, $d_\delta(x, z) := v_\delta(\{x, z\}) - v_\delta(\{z\}) = \sum_i \delta(y_{i-1}, y_i)$. By additivity of δ, this implies

$$d_\delta(x, z) = \delta(x, z),$$

i.e. the dissimilarity metric associated with v_δ is simply the given distance function δ.

We will now show that the functional form (3.3) corresponds to a specific structure of the underlying attributes. Specifically, each relevant attribute A is either a clade ($A = C_x$ for some x), or an "anti-clade," i.e. the complement of a clade ($A = X \setminus C_x$ for some x). Formally, let

$$\mathcal{T}^0 := \mathcal{H}_{ev} \cup \{X \setminus A : A \in \mathcal{H}_{ev}\},$$

which we refer to as the "quasi-cladistic" model.

Proposition 3.1 *A function v_δ of the form (3.3) has a unique representation in terms of an attribute weighting function λ, given by*

$$\lambda_A = \begin{cases} 0 & \text{if} \quad A \notin \mathcal{T}^0 \\ \delta(x^-, x) & \text{if} \quad A = C_x \text{ and } x \neq x_0 \\ \delta(x^-, x) & \text{if} \quad A = X \setminus C_x \text{ and } x \neq x_0 \\ c - \Delta & \text{if} \quad A = X \end{cases}, \tag{3.4}$$

where $\Delta := \sum_{y \neq x_0} \delta(y^-, y)$. Conversely, for any attribute weighting function satisfying (3.4) the corresponding function v_δ defined by $v_\delta(S) = \sum_{A : A \cap S \neq \emptyset} \lambda_A$ is of the form (3.3). In particular, v_δ is a diversity function if and only if $c \geq \Delta$.

While the quasi-cladistic model overcomes the criticisms of the cladistic model put forward in the previous subsection, it has problems of its own. These problems appear in their starkest form in a lineage of descent as in Figure 6 above. While now an ancestor has positive marginal diversity value, intermediate species do not. Concretely, one has $v_\delta(\{x_1, x_2, ..., x_m\}) = v_\delta(\{x_1, x_m\})$ whenever $x_2, ..., x_{m-1}$ are intermediate between x_1 and x_m. More generally, the quasi-cladistic model implies that, for all S,

$$v(S) = v(Sp(S)). \tag{3.5}$$

In fact, among all uniform diversity functions compatible with the tree τ_{ev}, property (3.5) characterizes the quasi-cladistic ones.[8] Thus, the entire additional content of the quasi-cladistic model stems from the problematic restriction (3.5).

In defense of the quasi-cladistic model, one could argue that (3.5) entails no direct restriction on the diversity of extant species $E(X)$ (terminal nodes), since $Sp(S)$ can differ from S only in non-extant species. Consider, however, again the "caterpillar" pattern.

---

[8]Indeed, if $v(S) \neq v(Sp(S))$, then there must exist an attribute $A \notin \mathcal{T}^0$ with strictly positive weight.

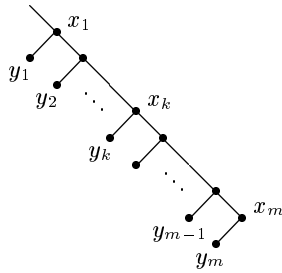*Figure 7: The caterpillar again*

Assume that $\delta(x_i, y_i) = \bar{\delta}$ for all $i = 1, ..., m$, and that $\delta(x_{m-1}, x_m) > 0$. In that case, the quasi-cladistic model implies that

$$v_\delta(\{y_1, y_{m-1}, y_k\}) < v_\delta(\{y_1, y_{m-1}, y_m\}),$$

no matter how large $\delta(y_1, y_k)$ and $\delta(y_k, y_m)$ are, and no matter how small $\delta(y_{m-1}, y_m)$ is. In other words, given the survival of species y_1 and y_{m-1}, the marginal diversity of y_m is always greater than that of *any* species $y_k \in \{y_2, ..., y_{m-2}\}$, which seems implausible.

The diagnosis of the restrictiveness of the quasi-cladistic model is that it rules out *combinations* of attributes. For instance, in Figure 7 all potentially relevant attributes of the form $C_{x_i} \cap (X \backslash C_{x_j})$ for $i < j$ ("descending from x_i but not from x_j") are excluded. But the distinctiveness of y_k from y_1 and y_m is naturally described precisely by such attributes. In general, it seems that combinations of relevant attributes (formally: their intersections) are often relevant themselves, a point that has frequently been observed in the literature (see, e.g., Williams and Humphries (1996)). Note, however, that if one was to include all intersections of clades and anticlades, one would be thrown back to the general tree model $\mathcal{T}_{ev}$. In the next section, we will make a proposal that ensures that combinations of relevant attributes are relevant as well. More specifically, we are looking for a model $\mathcal{A}$ with $\mathcal{H}_{ev} \subset \mathcal{A} \subset \mathcal{T}_{ev}$ that is stable under taking intersections (combinations) of attributes. By the preceding argument, we are forced to exclude *some* anticlades. Arguably, while some anticlades certainly correspond to relevant attributes, not all do. For instance, the anticlade corresponding to the first reptile (i.e. the attribute "pre-reptile") is probably relevant.[9] By contrast, the anticlade "pre-porcupine" seems to be biologically hardly significant.

The so-called "p-median model" proposed by Faith and Walker (1994) is also closed under intersections and does, therefore, not suffer from the just stated weakness of the quasi-cladistic model. In contrast to the phylogenetic tree model developed below, however, it typically excludes some clades as relevant attributes; moreover, it fixes attribute weights in a specific way.

---

[9]Indeed, the attribute "fish" critical to our argument surrounding Figure 4 above can be construed as the intersection of the clade "vertebrates" and the anticlade "pre-reptile."

14

4 The Phylogenetic Tree Model

4.1 Definition

The basic idea of the phylogenetic model to be described in this section is to enrich the family of cladistic attributes in a controlled way. This is done in two steps: First, an appropriate set of further attributes $\mathcal{H} \subseteq \mathcal{T}_{ev}$ with a hierarchical structure is added. In a second step, all intersections of the attributes in $\mathcal{H}_{ev}$ and $\mathcal{H}$, respectively, are included.

Definition A model $\mathcal{PH}$ is called a *phylogenetic tree model* if there exists a hierarchical family $\mathcal{H} \subseteq \mathcal{T}_{ev}$ such that

$$\mathcal{PH} \subseteq (\mathcal{H} \cup \mathcal{H}_{ev})^* := \{A \cap B : A \in \mathcal{H} \text{ and } B \in \mathcal{H}_{ev}\}.^{10}$$

To motivate this definition, consider again the segment of the evolutionary tree depicted in Figure 4. We have argued, that one problem with the cladistic model is its neglect of the non-cladistic taxon "fish" (dotted in Figure 4) as a potential attribute. One would thus want to combine the underlying evolutionary tree with the similarity information in the taxonomic classification of species. A prime candidate for the family $\mathcal{H}$ in the above definition is thus a non-cladistic "Linnean" taxonomic hierachy $\mathcal{H}_{tax}$ of species. Including intersections of cladistic and taxonomic attributes is natural since intersections correspond to conjunctions of the attribute-defining features.

As a second example, consider again the caterpillar tree in Figure 5. Here, the inappropriate restrictions on dissimilarities entailed by the cladistic model are naturally overcome by including the following selection of anticlades as further relevant attributes: $A_k = \{y_1, ..., y_k\} \cup \{x : x \leq_{ev} y_k\}$ (see Figure 8). Note that, in contrast to the quasi-cladistic model, not *all* anticlades are included. In particular, the family $\mathcal{H} = \{A_k : k = 1, ..., m\}$ has a hierarchical structure, as required in the above definition.
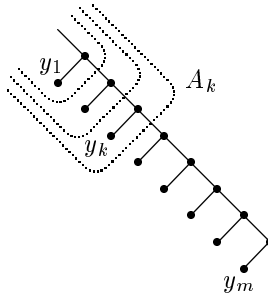

Figure 8: Augmenting the cladistic model in the caterpillar

Note that if one relativizes the corresponding phylogenetic tree model to the set of extant species $Y = \{y_1, ..., y_m\}$, the resulting model $\mathcal{PH}_{cat}|_Y$ consists of all sets of the form $B = \{y_j, ..., y_k, ..., y_l\}$ with $j \leq k \leq l$. This model on Y is characterized by the

---

[10]For a model $\mathcal{A}$, we denote by $\mathcal{A}^*$ the "intersection-closure," i.e. the family of all intersections of the elements of $\mathcal{A}$.

property that whenever B contains two species, then it also contains all species that are intermediate in terms of their "degree of evolvedness." Observe that, in contrast to the general tree model, $\mathcal{PH}_{cat}|_Y$ does impose the plausible restrictions of the form $d(y_k, y_j) \leq d(y_l, y_j)$ whenever $j \leq k \leq l$.

4.2   The Structure of Phylogenetic Tree Models

What restrictions are entailed by assuming that a model can be described as a phylogenetic tree model, for *some* appropriate family $\mathcal{H}$? This question is answered by the characterization of the entire *class* of phylogenetic tree models in terms of the following consistency property on the family of relevant attributes.

WNP (Weak Nesting Property) For all attributes A, B, and all x, y, z with $z \leq_{ev} x$ and $z \leq_{ev} y$, if $z \in A \cap B$ and $x \in A \setminus B$, then not $y \in B \setminus A$.

Note that if one were to drop the requirement in WNP that z be a common ancestor of x and y, one would obtain the Nesting Property. The Weak Nesting Property may be interpreted as a requirement of consistent ordering of attributes in terms of "entrenchment," as illustrated by the following figure. In the r.h.s. of Figure 9, the attribute A is "more entrenched" than B in the sense that in the evolution of species the attribute A cannot be lost without also loosing B, while the converse does not hold (see, e.g. species x). In the l.h.s. of Figure 9, when WNP is violated, no such relation between A and B exists.
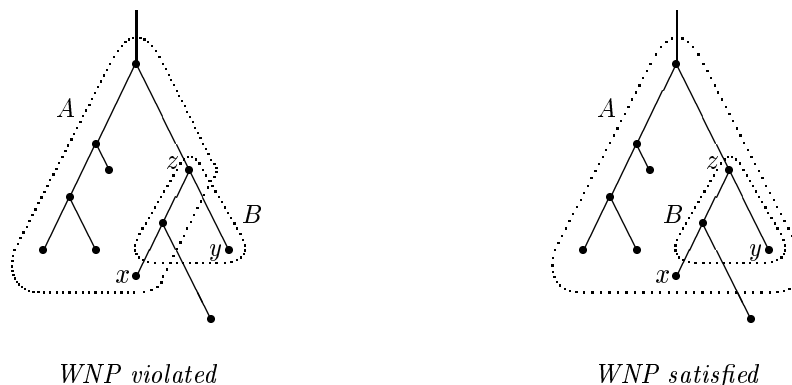

WNP violated   *WNP satisfied*

Figure 9: The Weak Nesting Property

Theorem 1 *A model is a phylogenetic tree model if and only if it is contained in $\mathcal{T}_{ev}$ and satisfies the Weak Nesting Property.*

By Theorem 1, WNP exhaustively describes the qualitative structure of phylogenetic tree models, and thus provides a simple empirical test of the applicability of that model. Are there empirical counterexamples to the Weak Nesting Property? The closest we could get to such an example is as follows.[11] Let z denote the "Ur-reptile" and consider the clade C_z founded by z, containing e.g. all birds and all mammals. Let x denote some bird, and y some mammal. Furthermore, denote by A the set of all species in C_z

---

[11] That we could not come up with a more compelling example may be due to our limited knowledge of biology; we hope that some biologists will find it worthwhile to examine the empirical validity or at least plausibility of the WNP in greater detail.

16

that are not mammals, and by B the set of all spieces in C_z that are not birds. If one deems both attributes A ("non-mammal") and B ("non-bird") as relevant, one obtains a violation of WNP as shown in the following figure.
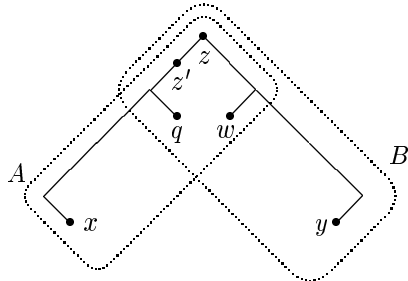

Figure 10: Reptiles, birds and mammals

A couple of remarks are in order. Note first that, in order to obtain a violation of WNP that can be motivated in terms of terminal nodes of the tree, it is necessary that the attributes A and B include species "on the other side" of z, indicated here by w for A and q for B. As a matter of empirical fact, according to Tudge (2000, Table 17), all *extant* non-mammalian members of the reptile clade are in the position of q, none in that of w. As a result, for the purpose of describing (dis)similarities among *extant* species, the attribute A is equivalent to the cladistic attribute $C_{z'}$; replacing A by $C_{z'}$ evidently removes the violation of WNP. In any case, it seems plausible that the relatively few extinct reptile species in the position of w can be viewed as "pre-mammals" in which case the biological appeal of A as an attribute would seem weak.

On the other hand, the example clearly illustrates that there is nothing in the notion of phylogenetic diversity *per se* which would preclude violations of the WNP due to the existence of relevant attributes of the form $C_z \setminus C_{y_0}$ and $C_z \setminus C_{x_0}$, where C_{y_0} and C_{x_0} are disjoint subclades of C_z (here the clades of mammals and birds, respectively). It simply appears to be a (rather remarkable) empirical fact that such non-nested pairs of attributes occur only exceptionally as "natural kinds," if at all. If one accepts the "empirical validity" of the WNP, one can read Theorem 1 as explaining mathematically *why* it is possible to arrive at a satisfactory model of phylogenetic diversity by augmenting the family of cladistic attributes by a family of supplementary attributes that is organized *hierarchically* in the manner of Linnean taxa. In confirmation, it is interesting to note that Tudge's (2000) more than 20 non-cladistic attributes recorded in his global evolutionary tree satisfy the Nesting Property.

4.3   Remoteness-Representation of Phylogenetic Tree Models

Phylogenetic tree models admit a simple representation in terms of an index of "remoteness" from the root, as follows. For any set S of species, denote by $lcaS$ the latest common ancestor of the species in S. Given a model $\mathcal{A} \subseteq \mathcal{T}_{ev}$, define a binary relation on X as follows. For all x, y,

$$y \succ_\mathcal{A} x :\Leftrightarrow [\text{ there exists } A \in \mathcal{A} \text{ with } lca\{x, y\} \in A, x \in A \text{ and } y \notin A] \qquad (4.1)$$

Intuitively, $y \succ_\mathcal{A} x$ means that species y is more remote from the root than species x since y has lost some attribute that x still shares with their latest common ancestor.

17

Note that WNP is equivalent to the asymmetry of $\succ_{\mathcal{A}}$. Under WNP, the interpretation of $\succ_{\mathcal{A}}$ as describing "remoteness from the root" is confirmed by the following observation.

Fact 4.1 *If $\mathcal{A} \subseteq \mathcal{T}_{ev}$ satisfies WNP, then the relation $\succ_{\mathcal{A}}$ defined in (4.1) is irreflexive, asymmetric and transitive.*

From Fact 4.1 it follows that we can choose an index $\rho : X \to \mathbf{R}$ such that

$$y \succ_{\mathcal{A}} x \Rightarrow \rho(y) > \rho(x). \tag{4.2}$$

Suppose an index ρ has been derived via (4.1) and (4.2) for a model $\mathcal{A} \subseteq \mathcal{T}_{ev}$ satisfying WNP. Any attribute $A \in \mathcal{A}$ can be described as follows: If A originates from x, then A consists exactly of all successors of x that are not more remote from the root than some given value $r \geq \rho(x)$. Formally, for any $x \in X$ and any $r \geq \rho(x)$, denote

$$A_{x,r} := \{y : y \geq_{ev} x \text{ and } \rho(y) \leq r\}.$$

Fact 4.2 *Let $\mathcal{A} \subseteq \mathcal{T}_{ev}$ satisfy WNP and fix ρ according to (4.2). Then, any attribute $A \in \mathcal{A}$ is of the form $A = A_{x,r}$ for some x and some $r \geq \rho(x)$.*

Now suppose that with the tree τ_{ev} on X there is independently given an index $\rho : X \to \mathbf{R}$ satisfying

$$y \geq_{ev} x \Rightarrow \rho(y) \geq \rho(x), \tag{4.3}$$

and consider the family

$$\mathcal{PH}_{\rho} := \{A_{x,r} : x \in X, r \geq \rho(x)\}.$$

Theorem 2 *A model is a phylogenetic tree model if and only if it is contained in $\mathcal{PH}_{\rho}$ for some index ρ satisfying (4.3).*

By this result, we can assume without loss of generality that a generic phylogenetic tree model takes the form $\mathcal{PH}_{\rho}$. This simplifies the following analysis considerably, as suggested by the following simple graphic representation of $\mathcal{PH}_{\rho}$. The crucial feature in Figure 11 is that each attribute ends on any branch at the same ρ-level.
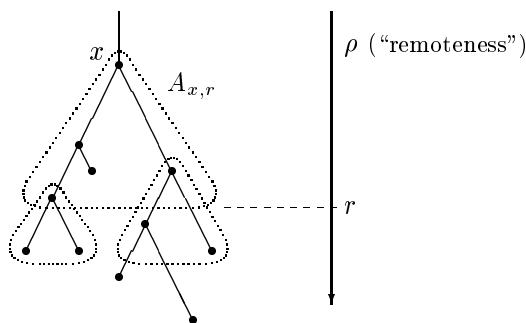

Figure 11: Graphic Representation of $\mathcal{PH}_{\rho}$

18

5 Diversity in Phylogenetic Tree Models

In this section, we generalize the recursion formula characterizing the cladistic model. Consider any diversity function v with $\Lambda \subseteq \mathcal{PH}_\rho$ and an arbitrary set $S = \{x_1, ..., x_k\}$ of species. Without loss of generality, assume that the species x_i are enumerated such that $\rho(x_1) \leq \rho(x_2) \leq ... \leq \rho(x_k)$. Denote by $\hat{x}_k$ any species in $\{x_1, ..., x_{k-1}\}$ that has the most recent common ancestor with x_k. Formally, $\hat{x}_k$ is defined by

$$lca\{\hat{x}_k, x_k\} \geq_{ev} lca\{x_i, x_k\} \text{ for all } i < k. \tag{5.1}$$

Note that $\hat{x}_k$ is not uniquely determined by (5.1). Figure 12 illustrates the choice of $\hat{x}_k$; in the depicted situation one has $\hat{x}_k = x_j$ (although $\rho(x_j) < \rho(x_{k-1})$).
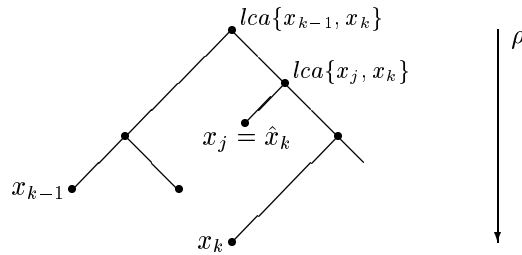

Figure 12: Determining $\hat{x}_k$

Fact 5.1 *If $\hat{x}_k$ satisfies (5.1),*

$$v(\{x_1, ..., x_k\}) = v(\{x_1, ..., x_{k-1}\}) + d(x_k, \hat{x}_k).$$

Applying the same argument to the set $\{x_1, ..., x_{k-1}\}$, one obtains for some appropriately chosen species $\hat{x}_{k-1} \in \{x_1, ..., x_{k-2}\}$,

$$v(\{x_1, ..., x_k\}) = v(\{x_1, ..., x_{k-2}\}) + d(x_{k-1}, \hat{x}_{k-1}) + d(x_k, \hat{x}_k).$$

By induction, we thus have the following result which implies that any phylogenetic tree model is monotone in dissimilarity.

Proposition 5.1 *Let v be compatible with some phylogenetic tree model $\mathcal{PH}_\rho$, and let $S = \{x_1, ..., x_k\}$ be any set of species such that $\rho(x_1) \leq ... \leq \rho(x_k)$. Then,*

$$v(\{x_1, ..., x_k\}) = v(\{x_1\}) + \sum_{i=2}^{k} d(x_i, \hat{x}_i). \tag{5.2}$$

6 Metric Restrictions of Phylogenetic Tree Models

As we have just seen, a generalized version of the recursion formula (3.1) remains valid for the phylogenetic tree model. In particular, the underlying dissimilarity metric (together with the value of singletons) uniquely determines the diversity of any set of

species. In order to further simplify the model it may seem natural to assume that dissimilarity is additive along lines of descent, i.e. that

$$d(x,z) = d(x,y) + d(y,z)$$

whenever z descends from y, and y from x.[12] This, however, turns out to be remarkably restrictive. In the general tree model, it forces all relevant attributes to be clades or anticlades, i.e. it reduces the general tree model to the quasi-cladistic one. By consequence, a phylogenetic tree that admits a non-degenerate additive dissimilarity metric must take the form of a single line of descent. Formally, one has the following results.

Proposition 6.1 *A τ_{ev}-compatible diversity function v induces a dissimilarity metric d that is additive along lines of descent if and only if $\Lambda \subseteq \mathcal{T}^0$.*

Corollary 6.1 *Let v be compatible with a phylogenetic tree model $\mathcal{PH}_\rho$. If the associated dissimilarity metric is additive along lines of descent and strictly positive, then τ_{ev} must be a line. Conversely, if τ_{ev} is a line, diversity functions with the stated properties exist.*

The latter result implies that in a proper phylogenetic tree, dissimilarity must be *sub*additive. What other properties do dissimilarity metrics associated with phylogenetic tree models possess? In the following, we want to determine them by answering the following question. Consider any phylogenetic tree model $\mathcal{PH}_\rho$ on the evolutionary tree τ_{ev}; furthermore, suppose we are given a particular valuation $v^0(x)$ of all single species, and, for any pair of species x, y, the dissimilarity $d(x, y) \geq 0$. Under what conditions on these data does there exist a diversity function v that is compatible with the given model $\mathcal{PH}_\rho$ and that induces precisely the given valuation of singletons and the given pairwise dissimilarities? It turns out that the metric restrictions can be classified into those *along* lines of descent and those *across* lines of descent. We consider the former first.

Say that a given dissimilarity metric is *bounded* if, for all x, y, $d(x, y) \leq v^0(x)$. Next, consider four species x', x, y and y' such that $x' \leq_{ev} x$, $x \leq_{ev} y$, $x \leq_{ev} y'$ and $\rho(y) \leq \rho(y')$ (see Figure 13).
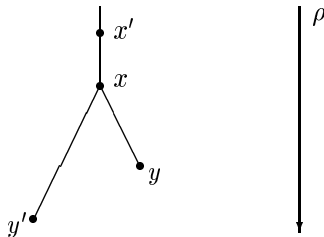

Figure 13: Monotonicity and Submodularity of d

Say that d is *monotone* if in this situation (i) $d(y, x') \geq d(y, x)$, and (ii) $d(x, y') \geq d(x, y)$. Furthermore, say that d is *submodular* if

$$d(x', y) - d(x, y) \geq d(x', y') - d(x, y').$$

---

[12]Note that this is weaker than the additivity condition considered in Section 3.4 above, since additivity is required here to hold only along lines of descent.

Intuitively, submodularity says that the increase in dissimilarity from a species y when replacing x by some of its ancestors is the smaller the more remote y is. Observe that submodularity implies subadditivity along lines of descent by taking $x = y$. As shown by Theorem 3 below, the conditions of boundedness, monotonicity and submodularity exhaustively describe the metric restrictions along lines of descent.

We now turn to the restrictions across lines of descent. Consider three species x, y_1 and y_2, where $\rho(x) \geq \max\{\rho(y_1), \rho(y_2)\}$ and $lca\{x, y_1\} = lca\{x, y_2\}$ (see Figure 14). In this case, compatibility with a phylogenetic tree model implies

$$d(x, y_1) = d(x, y_2). \tag{6.1}$$

To see this, simply observe that any attribute A possessed by x contains y_1 if and only if it contains y_2. Indeed, if $A \in \mathcal{PH}_\rho$ contains x and one element of $\{y_1, y_2\}$, it must contain $lca\{x, y_1\} = lca\{x, y_2\}$; but then $A \supseteq \{y_1, y_2\}$ since $\rho(x) \geq \max\{\rho(y_1), \rho(y_2)\}$.
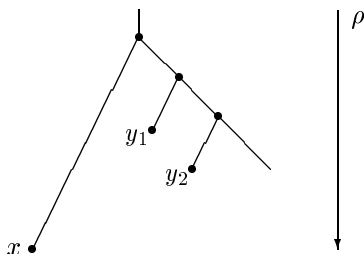

Figure 14: A situation in which necessarily $d(x, y_1) = d(x, y_2)$

Theorem 3 *Let $\mathcal{PH}_\rho$ be a phylogenetic tree model, and consider a given valuation $v^0 : X \to \mathbf{R}_+$ of all single species and a dissimilarity metric $d : X \times X \to \mathbf{R}_+$. There exists a diversity function v with $\Lambda \subseteq \mathcal{PH}_\rho$ such that $v(\{x\}) = v^0(x)$ and v induces the given dissimilarities if and only if d is bounded, monotone, submodular and satisfies (6.1). Moreover, v is uniquely determined by v^0 and d.*

The significance of Theorem 3 is that it *exhaustively* describes the restrictions imposed on the dissimilarities by the phylogenetic tree model. It thus provides a simple empirical test whether a given dissimilarity metric is consistent with a phylogenetic tree model.

Of the necessary properties of the dissimilarity metric underlying a phylogenetic tree model, the restriction (6.1) is arguably the most problematic. Note that (6.1) describes the restriction entailed by a *particular* phylogenetic tree model (i.e. for a given index ρ). More generally, consider three species x, y, z such that $z = lca\{x, y\}$ as in the following figure.
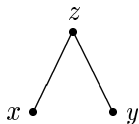

Figure 15: $z = lca\{x, y\}$

By (6.1), compatibility of the dissimilarity metric with *some* phylogenetic tree model requires in this situation,

$$d(x, y) = d(x, z) \text{ or } d(y, x) = d(y, z). \tag{6.2}$$

In the uniform case of constant valuation of single species, (6.2) is equivalent to requiring equality of the two greatest dissimilarities among any two species of the triple. This is the well-known ultrametricity condition of the cladistic model, but now restricted to the special constellation where one of the three species is the latest common ancestor of the two others. This reflects the relaxation of the Nesting Property to the Weak Nesting Property.

The "weak ultrametricity" condition (6.2) may appear counterintuitive. For instance, the very picture of the tree in Figure 15 might suggest that z is strictly between x and y, hence that $d(x, z) < d(x, y)$ and $d(y, z) < d(y, x)$. This is certainly true from a genetic point of view, but not necessarily from our present phylogenetic perspective. Indeed, rejecting the restriction (6.2) simply amounts to rejecting the Weak Nesting Property. To see more concretely what is involved, consider the following figure.
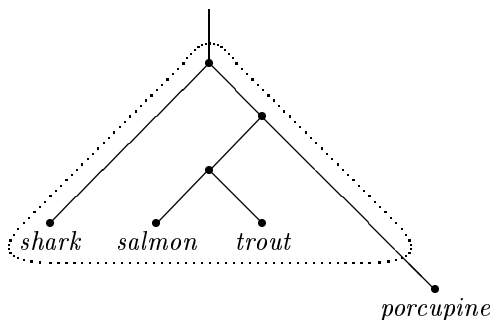

Figure 16: Miscellaneous Fish and Porcupines

The presence of the dotted attribute "fish" in Figure 16 entails that *porcupine* is more remote from the origin than all fish. Hence, by (6.1) or (6.2), $d(porcupine, salmon) = d(porcupine, trout)$. This does not seem implausible. If, however, one insists on, say $d(porcupine, salmon) > d(porcupine, trout)$, one is committed to introducing an attribute common to *porcupine* and *trout* but not to *salmon*, thereby violating WNP. Note also that the restriction (6.2) does not apply to the triple *porcupine, salmon, shark*. Indeed, since $lca\{porcupine, shark\} <_{ev} lca\{porcupine, salmon\}$, one would expect $d(porcupine, shark) > d(porcupine, salmon)$.[13]

7 Exploiting Information on the Speed of Evolution

7.1 The Evolutionary Clock Model

The representation in terms of the remoteness-index ρ suggests yet another interpretation of the phylogenetic tree model when there is external information on the speed of evolution. The intuition is as follows. The dissimilarity of a species x from its

---

[13]The intuitive appeal of the latter inequality is confirmed by the fact that mammals and salmons share an important phylogenetic feature, cartilaginous bones, which sharks do not have.

immediate ancestor can be viewed as determined by the "number of steps" taken by evolution to proceed to x from its immediate ancestor. One may also interpret this number as the "speed of evolution." The determination of this number of steps, or of the speed of evolution, is of course an empirical matter, and we will be silent on the issue of how to obtain this information. What is important for our purposes is that the above intuition relies upon the notion of a "step of evolution," and thus yields useful additional structure that one can exploit.

To make this formally precise, consider the tree τ_{ev} augmented by a set of "virtual species," each of which representing one step of evolution (see Figure 17 with x, y, z as actual species).
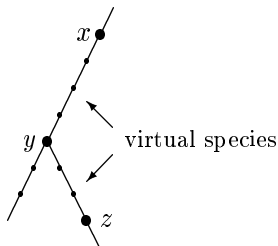

Figure 17: Actual and virtual species

For any x, y with $x \leq_{ev} y$ denote by $\delta(x, y)$ the number of steps needed to go from x to y along the augmented tree. We will refer to $\delta(x, y)$ as the *distance* between x and y. For instance, in Figure 17, $\delta(x, y) = 4$ and $\delta(y, z) = 3$. For simplicity, we will henceforth not distinguish between virtual and actual species and simply consider the augmented tree defined on the union of both. Thus, $\delta(x, y)$ coincides with the graph-distance on the augmented tree. The distance metric δ naturally induces the following index $\delta : X \to \mathbf{R}$,

$$\delta(x) := \delta(x_0, x).$$

Thus $\delta(x)$ is simply the distance of x from the root x_0, which we will refer to as the *evolvedness* of species x.

Given this additional information, it seems natural to consider the phylogenetic tree model $\mathcal{PH}_\delta$, i.e. to take the evolvedness δ as the remoteness-index. We will refer to $\mathcal{PH}_\delta$ as the *evolutionary clock model*. Observe that, in contrast to the derived index ρ which serves pure representation purposes, the index δ is now a primitive datum.

As is easily verified, the model $\mathcal{PH}_\delta$ is characterized by the following set of conditions. For all $A \in \mathcal{PH}_\delta$, A satisfies TR1 (Common Ancestor), TR2 (Unbroken Lineage) as before, and

TR3 If $\delta(x) = \delta(y)$, $z \leq_{ev} x$, $z \leq_{ev} y$ and $z \in A$, then $(x \in A \Leftrightarrow y \in A)$.

Note that if one drops the clause $\delta(x) = \delta(y)$ in TR3, the condition says that any distinctions between x and y must be on cladistic grounds. Condition TR3 requires this only for equally evolved species. This qualified application of the cladistic intuition is compatible with the critiques of the (unqualified) cladistic model given above. Indeed, once the evolutionary speed is taken into account, the given examples loose their force as counterexamples.

An interpretation of δ as speed of evolution requires a calibration of a step of evolution in terms of dissimilarity: Taking one step entails the same dissimilarity between

neighboring species (actual or virtual) anywhere in the tree. Formally, this amounts to the following requirement. For all x, y,

$$d(x^-, x) = d(y^-, y) \text{ and } d(x, x^-) = d(y, y^-), \qquad (7.1)$$

where x^- (resp. y^-) denotes the immediate ancestor of x (resp. y), as before. The requirement (7.1) should be viewed as a convention about what an evolutionary step is.

The evolutionary clock model shows that the cladistic view can be applied to equally evolved species. If one were to apply the cladistic view to all species existing at any point of time, one would implicitly identify evolvedness with time elapsed. But in view of (7.1) this would mean that the speed of evolution is *constant* across the tree: the dissimilarity of a given species at t from the *same* species at $t - k$ is the same as that of another species at t from its ancestor at $t - k$. But this contradicts a fundamental fact about evolutionary history, namely that almost all evolutionary change occurs through speciation, but only very little within species.[14] For a critique of the cladistic model along the same lines, see Faith (1994).

7.2   Translation Invariance

In Section 5 above, we have shown that the diversity of any set of species in a phylogenetic tree model is determined by their pairwise dissimilarities. We have also argued that this greatly reduces the information needed to determine the diversity of arbitrary sets, since with n species there are at most $n \cdot (n - 1)$ non-zero dissimilarities. However, the task of determining all dissimilarities between species still involves the comparison of species that are far apart, a comparison which may be difficult to perform in practice. Can one reduce this task to "local" comparisons, say comparisons that only involve a species and its immediate ancestor? Using the evolutionary clock model, we now offer a proposal to this effect that is applicable in the uniform case, i.e. when $v(\{x\}) = v(\{y\})$ for all x, y. Without loss of generality, we assume the normalization $v(\{x\}) = 1$ in all what follows. Note that uniformity implies symmetry of the dissimilarity metric, i.e. $d(x, y) = d(y, x)$. Recall that $d(x, y)$ quantifies the total weight of all attributes possessed by x but not by y. In particular, if $x \leq_{ev} y$, the dissimilarity $d(x, y)$ measures the weight of all "old" attributes that have been lost in the evolutionary transition from x to y, whereas $d(y, x)$ measures the weight of all "new" attributes that have been gained. The assumption of uniformity is that these measures are in perfect balance. One may thus refer to the uniform case as "evolution without progress." Clearly, this is restrictive; on the other hand, the equivalent assumption of equal valuation of all single species does not seem implausible as a benchmark.

In the uniform case, there are at most $n \cdot (n-1)/2$ different non-zero dissimilarities, and hence an equal number of degrees of freedom. The following proposal further reduces this number to n, the total number of species. Its basic idea is to derive the entire dissimilarity metric from the dissimilarity between neighbors in the tree augmented by virtual species. It is important to realize that although (7.1) fixes the dissimilarity entailed by *one* step of evolution, it does not determine the dissimilarity resulting from a sequence of k steps. Consider, for instance, the species $x_1, ..., x_m$ along

---

[14] A striking example is the famous *coelacanth* which has probably existed in almost unchanged form for more than 100 million years.

24

one line of descent as in Figure 6 above. While $d(x_{k+1}, x_k)$ is constant by (7.1), one has in general $d(x_3, x_1) \neq d(x_4, x_2)$. The following assumption rules this out.

Translation Invariance There exists a transformation f such that for all x, y with $x \leq_{ev} y$,

$$d(x, y) = f(\delta(x, y)).$$

The content of translation invariance is thus that the dissimilarity between two species along a line of descent only depends on their distance, i.e. the number of steps needed by evolution to proceed from one to the other. Note that in the translation invariant case, the transformation f determines via condition (6.1) the entire dissimilarity metric (also for species not on the same lineage of descent).

Say that the function f is *concave* if, for all k, k',

$$1 \leq k \leq k' \Rightarrow f(k') - f(k' - 1) \leq f(k) - f(k - 1).$$

Theorem 4 *Let d be a translation invariant dissimilarity metric satisfying (6.1) with an increasing transformation function f such that $f(0) = 0$ and $f(k) \leq 1$ for all k. There exists a uniform diversity function v with $\Lambda \subseteq \mathcal{PH}_\delta$ that induces the given dissimilarity metric if and only if f is concave. Moreover, v is uniquely determined by f.*

Translation invariance of d implies translation invariance of the attribute weights in the interior of the tree, as follows. Denote by $A_{x,k}$ the attribute starting at x and lasting for k steps. The attribute weighting function λ is called *translation invariant* if $\lambda_{A_{x,k}}$ only depends on k for all attributes $A_{x,k}$ with $x \neq x_0$ and k strictly smaller then the minimal number of steps from x to any terminal node.

Fact 7.1 *If d is translation invariant, then λ is translation invariant.*

The interrelation between the tranformation function f and the attribute weighting function λ which represents the "durability pattern" of attributes is as follows. The more concave f, the more concentrated is λ on attributes $A_{x,k}$ with small k, i.e. on short-lived attributes. Conversely, less concave f means that longer-lived attributes get a larger weight. The following two are the extreme cases. If $f(0) = 0$ and $f(k) = 1$ for all positive k, then λ is concentrated on all singleton attributes; in that case, v is simply the *counting measure*, according to which the diversity of a set of species is given by their number. On the other hand, if f is linear, the dissimilarity metric d coincides with the distance metric δ; in particular, for linear f the dissimilarity metric is additive. This implies, by Corollary 6.1, that the underlying tree τ_{ev} must be a line and that intermediate species have no marginal value.

As a final application, consider once again the "caterpillar" pattern as in Figure 18 below. If $\delta(y_i) \geq \delta(y_j)$ whenever $x_i \geq_{ev} x_j$, then the evolutionary clock model $\mathcal{PH}_\delta$ relativized to $Y = \{y_1, ..., y_m\}$ yields the "line model" on Y. That is, for any attribute $A \in \mathcal{PH}_\delta$ the intersection $A \cap Y$ is an *interval* in the natural ordering $y_1 < y_2 < ... < y_m$. This implies that the diversity $v(Y)$ is given by

$$v(Y) = v(\{y_1\}) + \sum_{i=2}^{m} d(y_i, y_{i-1})$$

(cf. TD (2002, Section 3)). If, in addition, $\delta(x_i, y_i) = \delta(x_j, y_j)$ and $\delta(x_i, x_{i+1}) = \delta(x_j, x_{j+1})$ for all i, j, then translation invariance on the tree entails translation invariance of the induced line model on Y (however, in general not with the same transformation function f).
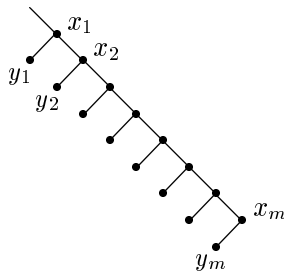

Figure 18: The caterpillar for the last time

Appendix: Proofs

Proof of Proposition 3.1 The uniqueness of the attribute weighting function λ is well-known in the literature (cf. TD (2002, Fact 2.1)). Thus, it suffices to show that v_δ has the form (3.3) if λ satisfies (3.4) and if v_δ is defined from λ via (2.1″). To verify this, observe that, for any S and all x, $S \cap C_x \neq \emptyset$ or $S \cap (X \setminus C_x) \neq \emptyset$. Moreover, if $x \neq x_0$, one has $(S \cap C_x \neq \emptyset$ *and* $S \cap (X \setminus C_x) \neq \emptyset)$ if and only if $\{x^-, x\} \subseteq Sp(S)$. This implies (3.3) by the given weight for λ_X. The last statement in Proposition 3.1 is immediate from (3.4) and the definition of diversity functions as those functions with $\lambda \geq 0$.

Proof of Theorem 1 (Necessity of WNP) We first prove the necessity of WNP. The proof of the sufficiency part relies on the construction of the remoteness-index provided by Facts 4.1 and 4.2, and is given after these results have been proved.

Hence, assume that $\mathcal{PH} \subseteq (\mathcal{H} \cup \mathcal{H}_{ev})^*$ for some hierarchical $\mathcal{H}$. Consider $A, B \in \mathcal{PH}$ and x, y, z with $z \leq_{ev} x$, $z \leq_{ev} y$, $z \in A \cap B$, and $x \in A \setminus B$. We show that $y \notin B \setminus A$ by a contradiction argument. Thus, assume $y \in B \setminus A$. Let $A = A_1 \cap A_2$ and $B = B_1 \cap B_2$ where $A_1, B_1 \in \mathcal{H}$ and $A_2, B_2 \in \mathcal{H}_{ev}$. Since A is not a clade, we have $\{x, z\} \subseteq A_1$ and $y \notin A_1$. Similarly, since B is not a clade, $\{y, z\} \subseteq B_1$ and $x \notin B_1$. But this contradicts the fact that $\mathcal{H}$ is hierarchical.

Proof of Fact 4.1 Irreflexivity of $\succ_{\mathcal{A}}$ is immediate. Asymmetry of $\succ_{\mathcal{A}}$ is a straightforward consequence of WNP. It remains to show that $\succ_{\mathcal{A}}$ is transitive. Hence, suppose that $z \succ_{\mathcal{A}} y$ and $y \succ_{\mathcal{A}} x$; we have to show that $z \succ_{\mathcal{A}} x$. Denote by $w_1 := lca\{x, y\}$ and $w_2 := lca\{y, z\}$. By assumption, there exists $A_1 \in \mathcal{A}$ such that $A_1 \supseteq \{w_1, x\}$ and $y \notin A_1$, and $A_2 \in \mathcal{A}$ such that $A_2 \supseteq \{w_2, y\}$ and $z \notin A_2$. Let $v_1 := lcaA_1$ and $v_2 := lcaA_2$. Since both v_1 and v_2 are ancestors of y, we have $v_1 \geq_{ev} v_2$ or $v_2 \geq_{ev} v_1$.

Case 1: $v_1 \geq_{ev} v_2$

Since A_2 contains the successor y of v_1, and since $y \notin A_1$, we must have $A_2 \supseteq A_1$ by WNP. Since A_2 thus contains both w_1 and w_2 it also contains $lca\{x, z\}$ by the τ_{ev}-connectedness. Hence, we have $A_2 \supseteq \{lca\{x, z\}, x\}$ and $z \notin A_2$. By definition, this implies $z \succ_{\mathcal{A}} x$.

Case 2: $v_2 \geq_{ev} v_1$

If A_1 contains z, which is a successor of v_2, one must have $A_1 \supseteq A_2$ by WNP. But this contradicts the fact that $y \notin A_1$. Hence, $z \notin A_1$. Since A_2 contains $lca\{y, z\}$, also A_1 contains $lca\{y, z\}$; moreover, A_1 contains $lca\{x, y\}$ as well as x by assumption. Hence, A_1 must also contain $lca\{x, z\}$ which yields $z \succ_{\mathcal{A}} x$.

Proof of Fact 4.2 Take any $A \in \mathcal{A}$. By τ_{ev}-connectedness of A, $x := lcaA \in A$; let $r := \max\{\rho(y) : y \in A\}$, and let $\bar{y} \in A$ be such that $\rho(\bar{y}) = r$. Consider any successor z of x with $\rho(z) \leq r$. Since both z and $\bar{y}$ are successors of x, one has $lca\{z, \bar{y}\} \in A$ by τ_{ev}-connectedness of A. But then $z \in A$, since otherwise (4.1) would imply $z \succ_{\mathcal{A}} \bar{y}$ and hence $\rho(z) > r$ by (4.2).

Proof of Theorem 1 (Sufficiency of WNP) Let $\mathcal{A} \subseteq \mathcal{T}_{ev}$ satisfy WNP. By Fact 4.2, there exists ρ such that any $A \in \mathcal{A}$ is of the form $A = A_{x,r}$. But this means that $\mathcal{A} \subseteq (\mathcal{H}_\rho \cup \mathcal{H}_{ev})^*$ with $\mathcal{H}_\rho := \{\{x : \rho(x) \leq r\} : r \in \mathbf{R}\}$ which is hierarchical. By definition, $\mathcal{A}$ is a phylogenetic tree model.

Proof of Theorem 2 As in the proof of Theorem 1, one has $\mathcal{PH}_\rho = (\mathcal{H}_\rho \cup \mathcal{H}_{ev})^*$ with the hierarchical family $\mathcal{H}_\rho := \{\{x : \rho(x) \leq r\} : r \in \mathbf{R}\}$. Hence $\mathcal{PH}_\rho$ is a phylogenetic tree model.

Conversely, let $\mathcal{PH}$ be a phylogenetic tree model. Define ρ via (4.1) and (4.2) and observe that $y \geq_{ev} x$ implies (not $x \succ_{\mathcal{PH}} y$). Therefore, ρ can be chosen such that it satisfies (4.3) in addition. By Fact 4.2, $\mathcal{PH} \subseteq \mathcal{PH}_\rho$.

Proof of Fact 5.1 To verify the stated formula, we have to show that any attribute $A \in \mathcal{PH}_\rho$ that distinguishes x_k from $\hat{x}_k$ in the sense that $x_k \in A$ and $\hat{x}_k \notin A$ also distinguishes x_k from the entire set $\{x_1, ..., x_{k-1}\}$, i.e. $A \cap \{x_1, ..., x_{k-1}\} = \emptyset$. Equivalently, we must show that any attribute that x_k shares with some species in $\{x_1, ..., x_{k-1}\}$ is also possessed by $\hat{x}_k \in \{x_1, ..., x_{k-1}\}$. Thus, assume that $A_{x,r} \supseteq \{x_k, x_i\}$ for some $i < k$. This implies that $x \leq_{ev} lca\{x_i, x_k\}$, hence by the choice of $\hat{x}_k$ also $x \leq_{ev} lca\{\hat{x}_k, x_k\}$. In particular, $\hat{x}_k$ is a successor of x. Moreover, since $x_k \in A_{x,r}$, we have $r \geq \rho(x_k)$, and hence $r \geq \rho(\hat{x}_k)$. Together, these observations imply that $\hat{x}_k \in A_{x,r}$, and hence the stated formula.

Proof of Proposition 5.1 Formula (5.2) follows by induction from Fact 5.1 as shown in the main text.

Proof of Proposition 6.1 By definition, the dissimilarity metric d is additive along lines of descent if for all x, y, z with $x \leq_{ev} y \leq_{ev} z$,

$$\sum_{A : \{x,z\} \cap A \neq \emptyset} \lambda_A = \sum_{A : \{y,z\} \cap A \neq \emptyset} \lambda_A - \sum_{A : y \in A} \lambda_A + \sum_{A : \{x,y\} \cap A \neq \emptyset} \lambda_A. \tag{A.1}$$

It is easily verified that, for any $A \in \mathcal{T}^0$, λ_A occurs as a summand on the left-hand side of (A.1) if and only if it also occurs on the right-hand side (possibly twice with a positive sign and once with a negative sign). This shows that $\Lambda \subseteq \mathcal{T}^0$ implies additivity of d along lines of descent.

Conversely, suppose that $\lambda_{\tilde{A}} > 0$ for some $\tilde{A} \notin \mathcal{T}^0$. Choose $y \in \tilde{A}$ and $z \in E(X) \setminus \tilde{A}$ such that $z \geq_{ev} y$. Such species y and z exist since $\tilde{A}$ is not a clade; also note that the root x_0 is not in $\tilde{A}$ since $\tilde{A}$ is not an anticlade. Now observe that with $x = x_0$, $\lambda_{\tilde{A}}$ does not occur on the left-hand side of (A.1) but twice with a positive sign and once with a negative sign on the right-hand side. By the triangle inquality, this implies $d(x_0, z) < d(x_0, y) + d(y, z)$, hence d is not additive.

Proof of Corollary 6.1 By Proposition 6.1, additivity along lines of descent implies that all relevant attributes are either clades or anticlades; by strict positivity of the dissimilarity metric, *all* anticlades must receive strictly positive weight. By compatibility with a phylogenetic tree model, the family of all anticlades must form a hierarchy. Indeed, no anticlade can be obtained as the intersection of clades with other anticlades. But the hierarchical structure of all anticlades forces τ_{ev} to be a line. Indeed, suppose by way of contradiction that τ_{ev} is not a line; then, there exist three species x_1, x_2, y such that $x_1 \geq_{ev} y$ and $x_2 \geq_{ev} y$ but neither $x_1 \geq_{ev} x_2$, nor $x_2 \geq_{ev} x_1$. In this case the anticlades $X \setminus C_{x_1}$ and $X \setminus C_{x_2}$ are not nested, hence the family of all anticlades cannot form a hierarchy.

For the proof of Theorem 3, we need some additional notation and a lemma. First, observe that any $A \in \mathcal{PH}_\rho$ can be written in the form

$$A = [x, y] := \{z : z \geq_{ev} x \text{ and } \rho(z) \leq \rho(y)\},$$

for some $y \geq_{ev} x$. Indeed, $[x, y] = A_{x,r}$ with $r = \rho(y)$. For each $x \neq x_0$, denote by x^- its immediate ancestor. Furthermore, for any x denote by Y_x^{max} the set of all maximally remote successors of x, i.e. the set of all $y \geq_{ev} x$ such that for no $z \geq_{ev} x$, $\rho(z) > \rho(y)$. Finally, for any pair x, y with $y \geq_{ev} x$ and $y \notin Y_x^{max}$, denote by y^+ any successor of x that is more remote from the root than y and minimal among all these, i.e.

$$y^+ \in \arg\min\{\rho(z) : z \geq_{ev} x \text{ and } \rho(z) > \rho(y)\}.$$

Lemma A.1 *Let $v : 2^X \to \mathbf{R}$ a set function with $v(\emptyset) = 0$, and let $\lambda : 2^X \to \mathbf{R}$ be the unique function with $\lambda_\emptyset = 0$ such that, for all S, $v(S) = \sum_{A : A \cap S \neq \emptyset} \lambda_A$.[15] Suppose that $\Lambda \subseteq \mathcal{PH}_\rho$ for some phylogenetic tree model. Then, for all $x \neq x_0$ and all $y \notin Y_x^{max}$ with $y \geq_{ev} x$,*

$$\lambda_{[x,y]} = [d(x^-, y) - d(x, y)] - [d(x^-, y^+) - d(x, y^+)]. \tag{A.2}$$

Furthermore, for any $y \notin Y_{x_0}^{max}$,

$$\lambda_{[x_0, y]} = d(x_0, y^+) - d(x_0, y),$$

and for any $x \neq x_0$ and any $y^* \in Y_x^{max}$,

$$\lambda_{[x, y^*]} = d(y^*, x^-) - d(y^*, x),$$

and for any $y^* \in Y_{x_0}^{max}$,

$$\lambda_{[x_0, y^*]} = v(\{x_0\}) - d(x_0, y^*).$$

---

[15]For an arbitrary set function v, the function λ is called the *conjugate Moebius inverse*, see TD (2002).

Proof of Lemma A.1 We prove (A.2); the other three formulas follow along the same lines. Thus, take any $x \neq x_0$ and any successor $y \notin Y_x^{max}$; one has

$$d(x, y^+) - d(x, y) = \sum_{z \leq_{ev} x} \lambda_{[z,y]}. \qquad (A.3)$$

Similarly,

$$d(x^-, y^+) - d(x^-, y) = \sum_{z <_{ev} x} \lambda_{[z,y]}. \qquad (A.4)$$

Substracting (A.4) from (A.3) one obtains formula (A.2).

Proof of Theorem 3 Necessity of boundedness, monotonicity, submodularity and condition (6.1) for the existence of an extension is easily verified. The sufficiency part is verified as follows. Any given v^0 and d satisfying (6.1) can be uniquely extended to a set function $v : 2^X \to \mathbf{R}$ with $\Lambda \subseteq \mathcal{PH}_\rho$ via the recursion formula (5.2). Observe that v is well-defined by (6.1) since that condition ensures that $d(x_k, \hat{x}_k) = d(x_k, \hat{\hat{x}}_k)$ for any $\hat{x}_k, \hat{\hat{x}}_k$ satisfying (5.1). By Lemma A.1, submodularity of d guarantees non-negativity of λ at all interior attributes. Similarly, by the other three formulas given in Lemma A.1, monotonicity and boundedness guarantee non-negativity at all other attributes. Hence, v is a diversity function.

Proof of Theorem 4 The result is a simple corollary of Theorem 3. Indeed, it is easily verified that in the translation invariant case, submodularity of d is equivalent to concavity of f. Similarly, d is monotone if and only if f is increasing, and in view of the normalization $v(\{x\}) = 1$, boundedness of d corresponds to $f(k) \leq 1$, for all k.

Proof of Fact 7.1 Consider an interior attribute $[x, y]$ with $\delta(x, y) = k$. If d is translation invariant one has by (A.2), $\lambda_{[x,y]} = [f(k+1) - f(k)] - [f(k+2) - f(k+1)]$, hence $\lambda_{[x,y]}$ only depends on k.

References

[1] ANDO, A., J. CAMM, S. POLASKY and A. SOLOW (1998), Species Distribution, Land Values, and Efficient Conservation, *Science* **279**, 2126-2128.

[2] ARMSWORTH, P.R., B.E. KENDALL and F.W. DAVIS (2002), An Introduction to Biodiversity Concepts for Environmental Economists, *mimeo.*

[3] FAITH, D.P. (1992), Conservation Evaluation and Phylogenetic Diversity, *Biological Conservation* **61**, 1-10.

[4] FAITH, D.P. (1994), Phylogenetic Pattern and the Quantification of Organismal Biodiversity, *Phil. Trans. Roy. Soc. Lond. B* **345**, 45-58.

[5] FAITH, D.P. and P.A. WALKER (1994), Diversity: Reference and User's Guide, Canberra: CSIRO Division of Wildlife and Ecology.

[6] GASTON, K.J. (ed.) (1996), *Biodiversity. A Biology of Numbers and Difference*, Blackwell, Oxford.

[7] NEHRING, K. and C. PUPPE (2002), A Theory of Diversity, *Econometrica* **70**, 1155-1198.

[8] POLASKY, S. (ed.) (2001), *The Economics of Biodiversity Conservation*, Ashgate Publishers.

[9] POLASKY, S., A. SOLOW and J. BROADUS (1993), Searching for Uncertain Benefits and the Conservation of Biological Diversity, *Environmental and Resource Economics* **3**, 171-181.

[10] SOLOW, A., S. POLASKY and J. BROADUS (1993), On the Measurement of Biological Diversity, *Journal of Environmental Economics and Management* **24**, 60-68.

[11] TILMAN, D. (1997), Biodiversity and Ecosystem Functioning, in: Daily, G.C. (ed.), *Nature's Services. Societal Dependence on Natural Ecosystems*, Island Press, Washington, D.C.

[12] TUDGE, C. (2000), *The Variety of Life*, Oxford University Press, Oxford.

[13] VANE-WRIGHT, R.I., C.J. HUMPHRIES and P.H. WILLIAMS (1991), What to Protect? – Systematics and the Agony of Choice, *Biological Conservation* **55**, 235-254.

[14] WEIKARD, H.-P. (2002), Diversity Functions and the Value of Biodiversity, *Land Economics* **78**, 20-27.

[15] WEITZMAN, M. (1992), On Diversity, *Quarterly Journal of Economics* **107**, 363-405.

[16] WEITZMAN, M. (1998), The Noah's Ark Problem, *Econometrica* **66**, 1279-1298.

[17] P.H. WILLIAMS and C.J. HUMPHRIES (1996), Comparing Character Diversity among Biotas, in: Gaston, K.J. (ed.), *Biodiversity. A Biology of Numbers and Difference*, Blackwell, Oxford.