diversity

MDPI

Article

Phylogenetic Distance Metrics for Studies of Focal Species in Communities: Quantiles and Cumulative Curves

Gregory S. Gilbert [1,*] and Ingrid M. Parker [2]

[1] Environmental Studies Department, University of California, Santa Cruz, CA 95064, USA
[2] Ecology & Evolutionary Biology Department, University of California, Santa Cruz, CA 95064, USA; imparker@ucsc.edu
* Correspondence: ggilbert@ucsc.edu; Tel.: +1-831-459-5002

Abstract: The phylogenetic distance between species often predicts differences in ecologically important traits. The phylogenetic diversity and structure of biological communities can inform our understanding of the processes that shape those communities, and there is a well-developed framework for comparing phylogenetic structures of communities. However, particularly in studies of phylogenetic distances from one focal species to other members of its assemblage (a one-to-many framework), the standard metrics of community-wide studies encounter significant limitations due to the left-skewed distribution of pairwise phylogenetic distances in most biological communities. For studies that require estimating the degree of phylogenetic isolation of a focal taxon, the mean phylogenetic distance (MPD) usually provides little power to distinguish among taxa because it is heavily weighted by the many ways to be distantly related, whereas the nearest taxon distance (NTD) is highly idiosyncratic and ignores cases where multiple close relatives may contribute equally strongly to influence the focal species. Here we highlight the value of examining the cumulative distribution of phylogenetic distances in studies that take a focal-species approach. We describe and discuss the benefits of two new metrics. An integrated metric of phylogenetic distances (AU-PhyDC) uses information from the whole cumulative distribution, whereas the tenth quantile (PD10) is an extremely simple metric that improves on NTD by capturing the influence of multiple close relatives on ecological interactions. Several recent examples found that PD10 did a better job of revealing ecological patterns than NTD or MPD. We provide R code to facilitate the use of these approaches and advocate for the inclusion of PD10 along with NTD and MPD in statistical packages for phylogenetic ecology.

Keywords: phylogenetic distance; community structure; evolutionary ecology

check for updates

Citation: Gilbert, G.S.; Parker, I.M. Phylogenetic Distance Metrics for Studies of Focal Species in Communities: Quantiles and Cumulative Curves. *Diversity* **2022**, *14*, 521. https://doi.org/10.3390/d14070521

Academic Editor: Michael Wink

Received: 3 June 2022
Accepted: 26 June 2022
Published: 28 June 2022

1. Introduction

The consideration of evolutionary relationships among species has transformed the way we look for patterns in ecology and the way we think about ecological communities [1,2]. Shared evolutionary history produces correlations in trait values, which are expressed as a phylogenetic signal in the trait [3,4]. Although convergent evolution and strong stabilizing selection can reduce the phylogenetic signal [5], a wide array of ecologically important physiological or behavioral traits show a significant phylogenetic signal [6–9], and niche conservatism is now recognized as an important feature of a species' ecology [10]. The combination of readily available DNA sequence data, taxonomic supertrees [11–14], and new tools to estimate evolutionary divergence among pairs of species [15–17] has prompted many ecologists and conservation biologists to adopt a phylogenetic perspective.

When ecologically important traits show a phylogenetic signal, we expect that evolutionary history will influence the interactions of species with each other and their environment. For instance, because closely related species are likely to share traits that are favored under similar conditions (habitat filtering), species that share a habitat are predicted to be more closely related to each other than would be expected at random [11,12]. On the other hand, individuals of very closely related species that are living close together within a suitable habitat may compete for the same limiting resources. At such local scales, we expect limiting similarity to result in longer phylogenetic distances than expected by chance [13–16]. Tests of these hypotheses consider averages across species in a community and community-level outcomes such as phylogenetic diversity.

Separate from these community-level questions, we may be interested in one focal species' phylogenetic relationship to the other species in its assemblage. For instance, Darwin's naturalization hypothesis posits that an introduced plant species will be more successful when it is less closely related to the species in its new range [17,18]. Many tests of this hypothesis have utilized the quantitative tools of phylogenetic ecology in plants, animals, and microbes e.g., [19,20].

Another interesting application of phylogenetic distance metrics is in the study of species interactions such as plant–herbivore, plant–pollinator, and plant–pathogen interactions. Because many plant traits involved in pathogen attack or resistance show phylogenetic conservatism [16], we should expect the host ranges of plant pathogens to show a phylogenetic signal, and they do [9,21,22]. The same is often true for herbivores e.g., [23]. The ecological implication of phylogenetic signal in host range is that co-occurring plants that are more closely related will share more pathogens [24] and herbivores [23], and a species that is more phylogenetically isolated from its community should escape disease [25] and herbivory [26]. In addition, a similar phylogenetic signal is often also seen in plant–pollinator interactions e.g., [27], and plant–fungal mutualisms e.g., [28], where close relatives are more likely to share mutualists.

A large number of metrics of phylogenetic diversity and distance have been developed for biological communities and are summarized in several recent reviews [29–31]. For describing phylogenetic divergence [29] across a community or an assemblage of taxonomic units, the two most commonly used metrics are the mean pairwise distance (MPD) and the mean nearest taxon distance (MNTD). The MNTD takes the phylogenetic distance to the closest relative for each taxon and calculates the mean across all taxa. The MNTD provides an estimate of how phylogenetically distinct, on average, is each member of the community from its closest relative. In contrast, the MPD is the mean of the pairwise distances from each taxon to each other taxon (mathematically equivalent to average taxonomic diversity (AvTD or $\triangle+$, [32]). One way MPD and MNTD improved upon previous metrics is that they do not depend on the size of the assemblage and, therefore, can be compared across studies [32]. Although both MPD and MNTD measure phylogenetic divergence, MPD reflects branching deeper in the phylogenetic tree, whereas MNTD reflects structure at the tips of the branches, and the information provided by the two metrics is often distinct and complementary [29].

We have found that these most common metrics for analyzing phylogenetic distances at the community level have problematic limitations when applied to questions about focal species utilizing a one-to-many approach. In a focal-species context, phylogenetic distances are measured from one species to many neighbors, thus the nearest taxon distance (NTD) is restricted to a single pairwise distance. This makes the NTD measure strongly dependent on the presence of even a single close relative in a diverse community, highly sensitive to any phylogenetic or taxonomic errors that may appear in the data (in ways that MNTD is not), and insensitive to whether the focal species has one or many close relatives as neighbors. Because in any assemblage of species there are more ways to be distantly related to neighbors than closely related [33], the mean taxon distance (MTD) in a focal-species context will be shaped strongly by the numerous long phylogenetic distances in the assemblage, which limits its sensitivity. Metrics other than the mean and nearest taxon distances could improve studies of focal species in communities.

Here, we use empirical examples from real plant communities to outline and illustrate some of these limitations of commonly used mean and nearest phylogenetic distance metrics when applied to focal-species analyses of biological communities. We then show

how inspecting the cumulative distribution of phylogenetic distances provides informative insights, and how information across the distribution can be collapsed into a single metric. We advocate for the usefulness of phylogenetic distance quantiles and share our experience with one particularly useful metric, the 10^{th} quantile phylogenetic distance (PD10). Finally, we briefly describe how this measure can be weighted for species abundance and can be extended to whole-community analysis.

2. Materials and Methods

2.1. Phylogenetic Structure of Example Plant Communities

To illustrate the behavior of standard metrics, as well as to show how cumulative distribution and quantile-based metrics can be applied for focal-species analyses of communities, we drew on the example of a plant community that we have studied extensively: a non-grazed meadow in coastal California with 43 angiosperm species, dominated by non-native annuals (Figure 1) [25]. We collected plant community data as part of a previous study [25] on the campus of the University of California Santa Cruz. We used this data set simply as a representative example of a well-studied, wild plant community that consists of a mix of native and introduced species. Our previous published work in this system allowed us to draw on species composition, phylogenetic relationships including focal-species (one-to-many) analyses, and a dependent variable (disease intensity) to illustrate the use of the metrics [25].

To test the generality of our findings related to the phylogenetic structure of angiosperm communities, we inspected plant species composition data for eight diverse plant communities across the state of California, USA. We identified 32 datasets available through the University of California Natural Reserves System (https://ucnrs.org/plant-list/, accessed on 25 November 2020), and of these we selected 8 sites, ranging in size from 159 to 286 plant species, that encompass a broad range of types of communities (e.g., desert, maritime chaparral, montane forest, salt marsh) and geographic regions (e.g., Sierra Nevada Mountains, Mojave Desert, Coast Range Foothills, Central Coast) (Table S1).

2.2. Calculation of Community Phylogenetic Distance Metrics

The first step in calculating community phylogenetic distance metrics is to calculate a matrix of all pairwise phylogenetic distances from a phylogenetic tree of the community. Units of phylogenetic distance are Myr = Millions of years of independent evolution, which is twice the time to the most recent common ancestor (given as Ma in Figure 1). The R code to calculate the distance matrix and metrics is given in File S1, including the phylogenetic tree for the meadow community in Figure 1. File S2 includes code for creating a phylogenetic tree from a list of genus and species names of plants in any community.

When a supertree is used to create a phylogenetic tree for a community that includes more than two species in the same genus (or family), the resulting tree includes polytomies if the supertree is not resolved below the genus (or family) level. Within a polytomy clade, all of the reciprocal phylogenetic distances among species would be identical, even though some species are more closely related than others. In such cases, the polytomy can be resolved by grafting a finer resolution phylogeny of that clade (from the literature or prepared directly from sequence data) onto that branch and dated using the bladj function in the same way as the tree is dated in File S2.

A community of n species is represented by a symmetric distance matrix with a diagonal of n zeros, which are distances of each species to itself, and a lower triangle of $(n^2 - n)/2$ distance elements. The MPD is the mean of all pairwise distances from each taxon to each other taxon, and the MNTD is the mean of the closest pairwise distance for each taxon, averaged across all taxa. Because the distance from the species to itself (0 Myr) is excluded, the MNTD is the average 0^{th} quantile of the phylogenetic distances. For the species in our example meadow community (Figure 1), these measures are MPD = 244.1 Myr and MNTD = 71.3 Myr.

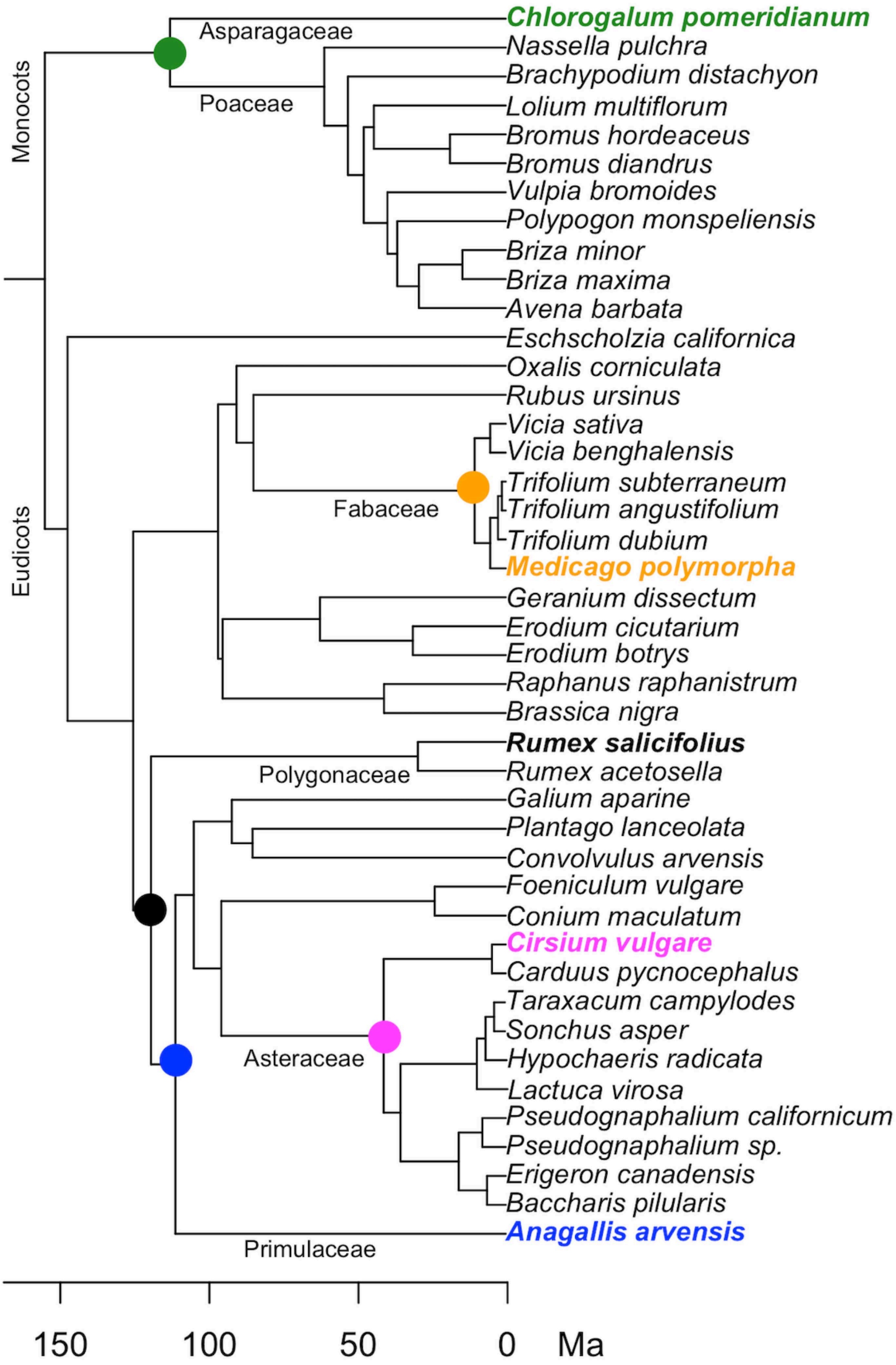

Figure 1. Phylogenetic tree of the 43 angiosperm plant species recorded in ten 20 m radius plots in a meadow on the UC Santa Cruz campus in coastal California. Units (Ma) are millions of years ago. See details of plant sampling and tree construction in Parker et al. [25]. Colored circles indicate the node distance that corresponds to the PD10 phylogenetic distances discussed in Section 3.2. The dated Newick file for the tree and code to visualize and analyze the tree are given in File S1.

2.3. Examination of Phylogenetic Distances from Focal Species in a Community

Many ecological questions are best addressed by examining how phylogenetic distances shape how focal species interact with neighboring species, rather than examining the phylogenetic structure of the community overall. Each focal plant species in a community of n species has $n - 1$ phylogenetic distances to neighboring species. We used the UCSC Meadow community to examine the structure of phylogenetic distances from the focal-species perspective using cumulative distributions of phylogenetic distances as quantiles (in 2% increments); thus, the 0th quantile is the nearest phylogenetic neighbor to the focal species (nearest taxon distance, NTD), the 10th quantile (the distance at which 10% of all distances are less than or equal to that phylogenetic distance; PD10), the 50th quantile is the median distance, and the 100th quantile is the maximum phylogenetic distance. We also calculated the mean phylogenetic distance from the focal species to all other community members (meanPD). We heuristically compared the information provided by each metric, inspecting how well each metric discriminated among different focal species in the context of the community.

2.4. Calculation of the AUPhyDC: Area under the Phylogenetic Distance Curve

The limitations of traditional phylogenetic distance metrics led us to explore a new way to synthesize all phylogenetic distance information for a given focal species by calculating the area under the phylogenetic distance curve (AUPhyDC), analogous to the area under the disease progress curve (AUDPC) that is commonly used in studies of plant disease spread [34]. The AUPhyDC is calculated as a Riemann sum of the areas of rectangles with height at the midpoint phylogenetic distances of each two successive quantiles. We used 2% quantile intervals; with the quantile scale in decimals rather than percents, the scale ranges from 0 to 1 with units of width of 0.02. The AUPhyDC has larger values for focal taxa that are phylogenetically isolated from their neighbors, and smaller values for those with many close relatives in the community.

2.5. Abundance Weighting of PD10 or AUPhyDC Metrics

Phylogenetic distance metrics at the community level such as MPD and MNTD can be weighted by species abundances to provide, for example, a scaled measure of evolutionary diversity of a community [35]. The PD10 and AUPhyDC metrics can also be weighted by relative abundance. Functionally, a vector of one-to-many phylogenetic distances can be expanded so that each species pair is replicated in proportion to the relative abundance of the target species. The cumulative distribution of phylogenetic distances is then generated based on the individuals (or units of percent cover). Weighted PD10 (wPD10) and weighted AUPhyDC (wAUPhyDC) are calculated from this distribution (R Code in File S1).

3. Results

3.1. Distributions of Phylogenetic Distances in Communities Are Skewed toward Long Distances

The distribution of phylogenetic distances in the UCSC Meadow community was left-skewed with a few short distances and many long distances (Figure 2). This reflects that there are many more ways for species to be distantly related than to be closely related. For example, there are 10 *Poaceae* species that have relatively short distances to each other, but each also has a 309 Myr distance to all 32 of the Eudicot species in the community. Of the eight distinct plant communities in the UC Natural Reserve System, all showed this same left-skewed structure (Figure 2). This means that broadly across plant communities, most plant–plant interactions will be with distantly related neighbors, even for species that have a number of close relatives in the community.

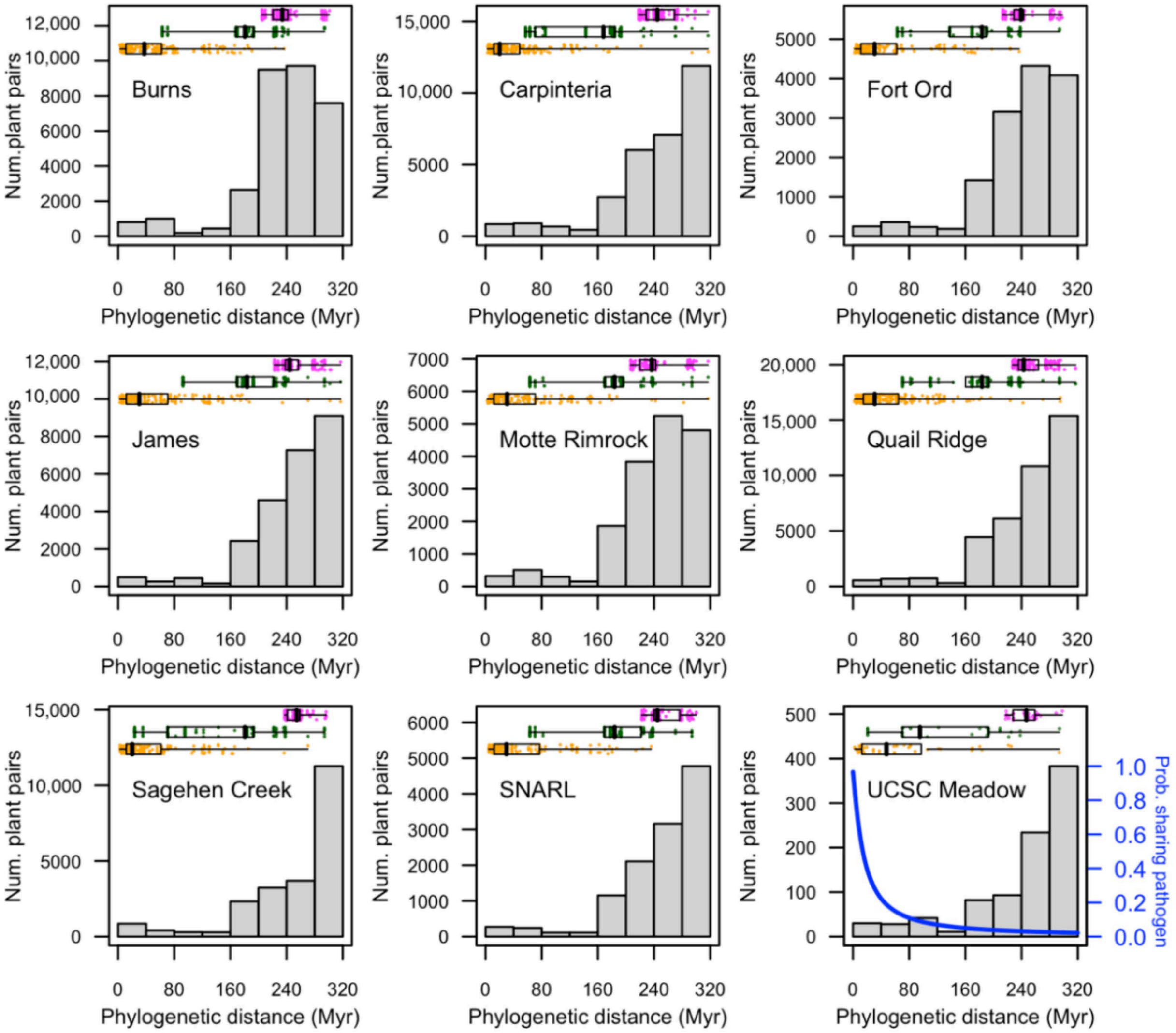

Figure 2. Features of the cumulative distribution of pairwise phylogenetic distances in plant communities. All pairwise phylogenetic distances were calculated for angiosperm plants in the UCSC Meadow (final panel, phylogenetic tree shown in Figure 1), and in each of eight UC Natural Reserves (species list in Table S1, and the R code to create phylogenetic trees from lists in). Phylogenetic distances are given as Myr of independent evolution between pairs of plant taxa (2× the time to the most recent common ancestor). Clouds of dots show the distributions of NTD (orange), PD10 (green), and meanPD (magenta) phylogenetic distances for all focal species in that community; associated box plots show the 25th and 50th quantile box, median thick line, and min and max whiskers. The blue curve in the UCSC Meadow panel indicates the empirically measured phylogenetic signal of host sharing, given as the probability that two plant species share a fungal pathogen at a given phylogenetic distance, where $\mathrm{logit}(pS) = 3.35861 - 2.857656 \times \log_{10}(\mathrm{Myr} + 1)$ [21,25].

3.2. Limitations of Standard Phylogenetic Metrics for Focal-Species Analyses

We examined the cumulative distribution of phylogenetic distances from a focal species to all the other species in a community, and variation among these curves revealed important differences among species (Figure 3). For instance, in the UCSC Meadow community, *Medicago polymorpha* (orange) has a number of close relatives and then a big gap without intermediate relatives, whereas for *Cirsium vulgare* (magenta), intermediate relatives make up a full quarter of the community (Figures 1 and 3). Meanwhile, the closest relative of *Anagallis arvensis* (blue) jumps straight to 221 Myr. The curves in Figure 3 illustrate the structural limitations of the common phylogenetic metrics for focal-species (one-to-many) applications. The nearest phylogenetic (taxonomic) distance (NTD) does not reveal when a focal species has multiple close relatives in the assemblage. Within our meadow community, *Medicago polymorpha* and *Cirsium vulgare* share an NTD of about 10 Myr (Figures 1 and 3). However, *M. polymorpha* has five close relatives, such that 14% of the species in the community are <21 Myr distant. In contrast, *C. vulgare* has just one close relative, with its next closest species at >80 Myr (Figure 3). Similarly, *Rumex salicifolius* has a single relative at 59 Myr, but its next closest species is a very distant 238 Myr. The distribution of NTD across all species is heavily right-skewed (Figure 4a). Across all the species in the UCSC Meadow community with an NTD < 20 Myr, half of those have a second neighbor equally closely related, but this is not captured in the NTD. Any ecological process that shows a continuous response to phylogenetic relatedness would be expected to be influenced by multiple close relatives of a focal species; host sharing by both pathogens and herbivores is just one example. By focusing only on the single closest taxon, NTD omits potentially important information from other relatives in the community.

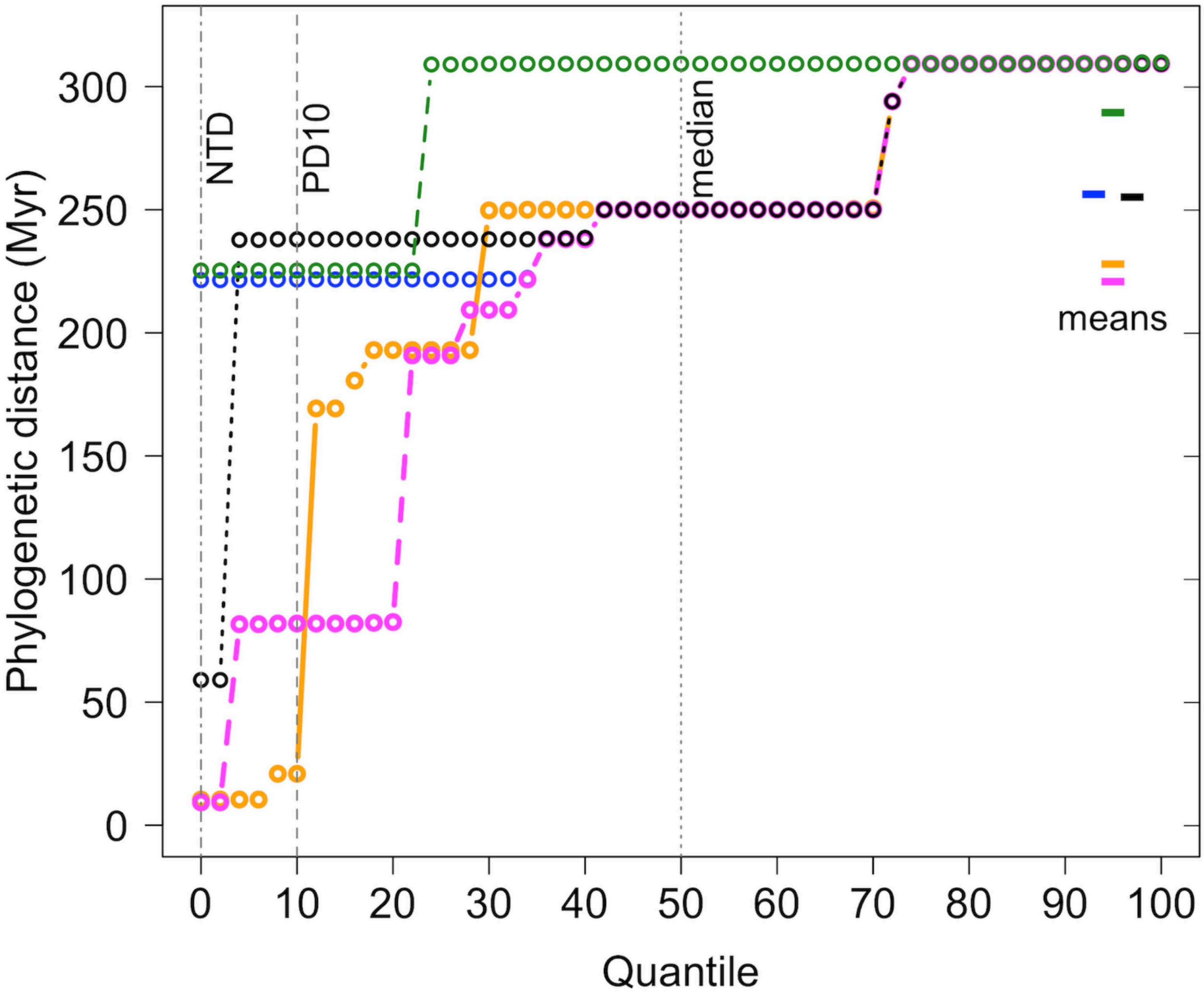

Figure 3. Cumulative phylogenetic distance curves for five focal UCSC Meadow plant species to other species in the community. Phylogenetic distances from the focal species to every other species in the meadow community are based on the phylogenetic tree in Figure 1. Focal species are *Medicago polymorpha* (orange); *Cirsium vulgare* (magenta); *Rumex salicifolius* (black); *Anagallis arvensis* (blue); *Chlorogalum pomeridianum* (green). Vertical dashed lines indicate the thresholds for nearest taxon distance (NTD), the 10th quantile (PD10), the 50th quantile (median); the mean phylogenetic distance for each focal species is shown as a correspondingly colored bar at right.

On the other hand, mean phylogenetic distance (called meanPD here for focal-species analyses, to distinguish it from the community-wide MPD) is disproportionately influenced by the long distances that dominate most communities (Figures 2 and 4c). This swamping effect diminishes the contribution of close relatives, those species most likely to have strong ecological interactions (Figure 4). As such, meanPD is unlikely to provide a sensitive measure of how focal species interact ecologically with their neighbors. In Figure 3, the meanPD shows some variation among the five species (e.g., *M. polymorpha* and *C. vulgare*, with more close relatives, have lower meanPD values than those species without close relatives), but the measure is not very sensitive to the short phylogenetic distances that most reflect ecological importance. Note that the mean distance for *M. polymorpha* is actually greater than that for *C. vulgare*, despite having five-fold more close relatives. Using the median phylogenetic distance instead of the mean is not an effective alternative. Four of the five species have exactly the same median phylogenetic distance to their neighbors, irrespective of how many close relatives they have in the community. Such limitations are not restricted to the exemplar species of Figure 3; the mean phylogenetic distances for all the species are distributed in a narrow range from 217 to 298 Myr (Figure 4c), all within the range of distances where a phylogenetic signal suggests ecological interactions to be weak. Given these limitations, we suggest that focal-species studies may not be well served by either meanPD or NTD.

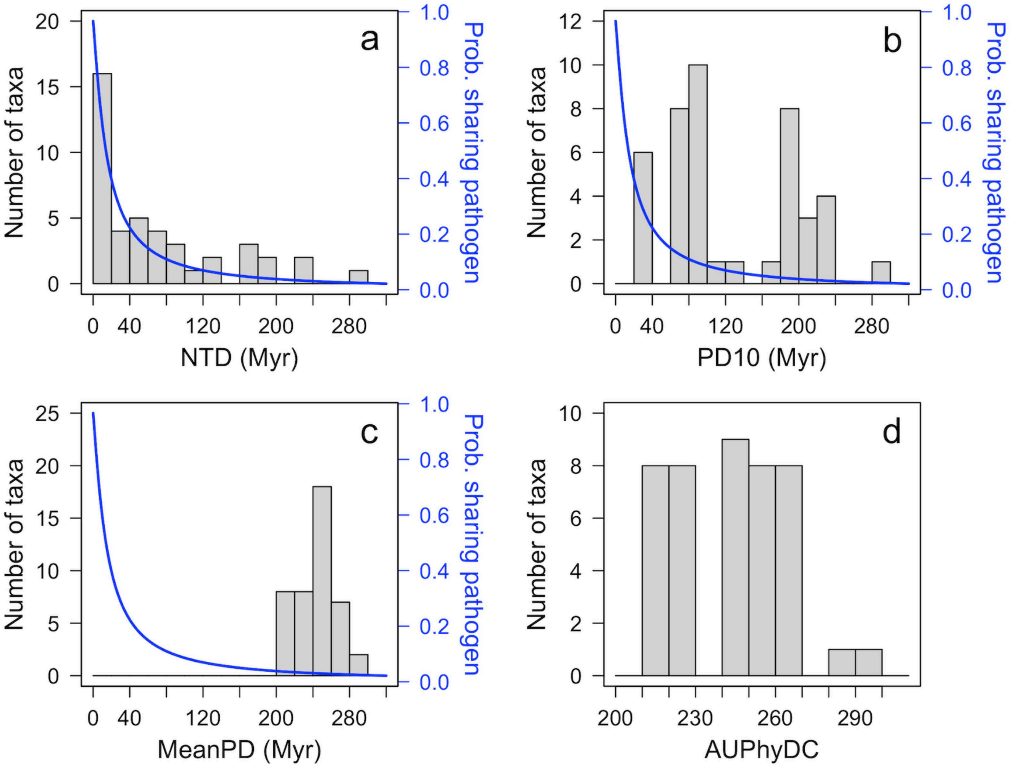

Figure 4. Distributions of four one-to-many metrics of phylogenetic distance. (**a**) Phylogenetic distance to the most closely related taxon (NTD); (**b**) the 10th quantile distance from the focal species to the remaining taxa (PD10); (**c**) mean phylogenetic distance from each focal taxon to the remaining taxa (meanPD); (**d**) area under the phylogenetic distance curve (AUPhyDC) for each focal taxon, as described in Figure 5. Metrics were calculated for each of the 43 angiosperm focal taxa shown in Figure 1. For all metrics, phylogenetic distance from a taxon to itself is excluded. The blue curve indicates the empirically measured phylogenetic signal of two plant species sharing a fungal pathogen, where logit (probability of sharing) = 3.35861 − 2.857656 × $\log_{10}$(1 + phylogenetic distance in Myr) [21,25].

3.3. AUPhyDC: Area under the Phylogenetic Distance Curve

Given the limitations of meanPD and NTD for focal-species studies, we explored the use of a novel metric, the area under the phylogenetic distance curve (AUPhyDC), to synthesize phylogenetic distance information across all taxa in the community (Figure 5). The AUPhyDC has a larger value for those species that are more phylogenetically isolated from the rest of the community (i.e., where most distances are long). Calculating the AUPhyDC for each of the 43 species in the UCSC Meadow community showed a relatively flat distribution of AUPhyDC values across species (Figure 4d).

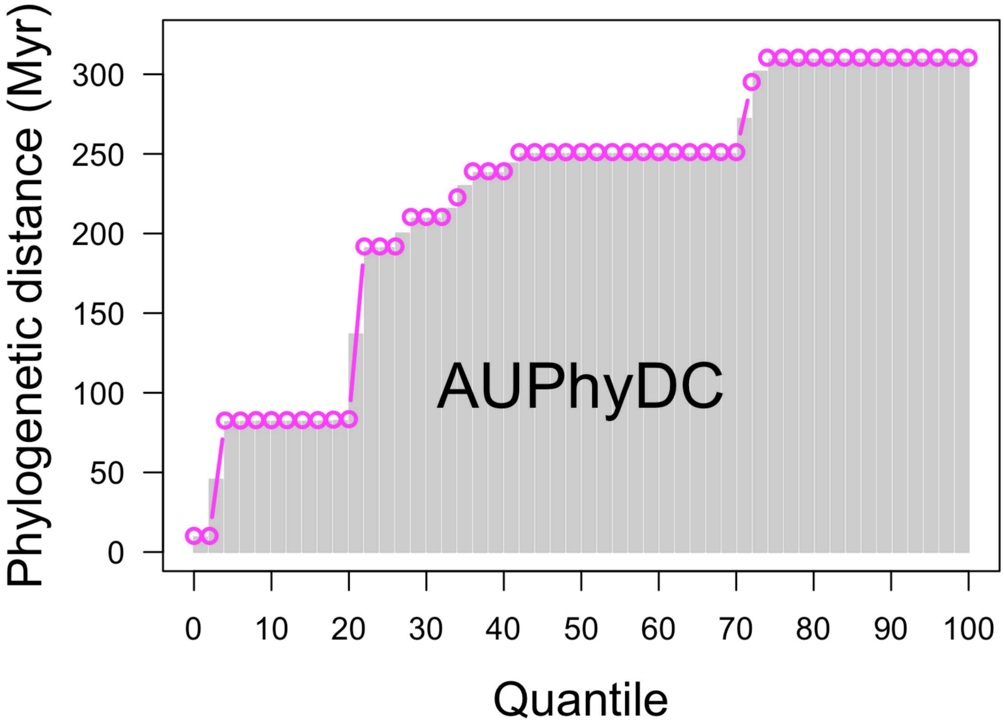

Figure 5. Area under the phylogenetic distance curve (AUPhyDC). The AUPhyDC provides a flexible, integrative measure of the overall distribution of phylogenetic distances from a focal taxon to all the other taxa in its community. The magenta line and circles show the cumulative phylogenetic distance curve of *Cirsium vulgare*, as shown in Figure 3. Each circle is the phylogenetic distance associated with that quantile, in increments of 2%. The area under that curve (AUPhyDC: area under the phylogenetic distance curve, in grey) can be estimated using a Reimann sum of the area of each of the grey rectangles (see Section 2.4 and File S1). The AUPhyDC is larger for taxa that are more evolutionarily isolated from their surrounding community because long phylogenetic distances are reached at lower quantiles.

3.4. PD10: The 10^{th} Quantile Phylogenetic Distance

Inspection of the cumulative distribution curves in Figure 3 suggests an important pattern: the greatest differences among focal species occur between the 5^{th} and 20^{th} quantiles of phylogenetic distances. As described above, the influence of the many long phylogenetic distances obscures variation among species beyond the 20^{th} quantile of phylogenetic distances. Between the 5^{th} and 20^{th} quantiles, focal taxa that are part of a diverse clade of close relatives show phylogenetic distances quite distinct from those for taxa that are more phylogenetically isolated. In the plant assemblages examined to date [25,36,37], we and others have found consistently that the 10^{th} quantile is a superior metric for examining how focal species differ in the structure of their phylogenetic relationships.

To illustrate, we compare NTD, meanPD, and PD10 for their descriptive (and predictive) power in the 43-species meadow community described above (Figure 1). We studied whether phylogenetic isolation decreased the vulnerability of a focal species to pathogen attack [25] (Figure 6). We found that diseased leaf area declined significantly with PD10 (Figure 6b). However, this relationship was obscured when using either NTD (Figure 6a) or meanPD (Figure 6c). In that system, PD10 did a better job of characterizing phylogenetic rarity in a way that was ecologically relevant, integrating the contributions of pathogens from multiple potential shared hosts. Boxplots for NTD, PD10, and meanPD for each of the focal species in the eight additional plant communities show similar patterns (Figure 2); PD10 consistently spans a broad range of variation in phylogenetic isolation of focal species, including the short to intermediate phylogenetic distances associated with ecologically relevant differences, without the idiosyncratic patterns associated with the NTD and with less compression of distances seen in the meanPD.

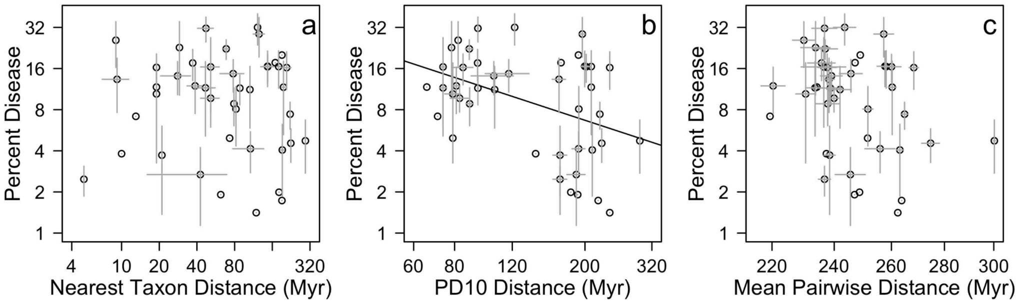

Figure 6. Phylogenetic isolation as a predictor of plant disease severity: comparison of three phylogenetic distance metrics. Disease severity on the meadow species in Figure 1, as a function of three phylogenetic distance metrics for focal plants to their surrounding community. The PD10 metric (**b**) was a better predictor than either NTD (**a**) or meanPD (**c**). Panel b for PD10 is adapted from Figure2b in Parker et al. [25] and we have added the equivalent panels for NTD and meanPD based on the same data. The regression analysis for PD10 showed a highly significant effect of phylogenetic isolation on reducing disease severity ($\log_{10}$(mean % Disease) = 2.638 − 0.788 × $\log_{10}$(phylogenetic distance); $R^2 = 0.171$, $p = 0.006$), but such a pattern was not revealed when using either NTD ($R^2 = 0.001$, $p = 0.82$) or meanPD ($R^2 = 0.081$, $p = 0.065$) as measures of phylogenetic distance. Each circle represents the mean % foliar disease and mean phylogenetic metric for a species of plant across 10 replicate plots in the meadow; whiskers indicate standard error of the mean. Disease severity data previously published in [25].

3.5. Abundance Weighting of AUPhyDC and PD10

Weighting phylogenetic distance by abundance in a focal-species context is conceptually equivalent to focusing on the individual plant and the probability of interactions with other species (or individuals, including conspecifics) in its community. In contrast to the unweighted metrics, weighting by abundance allows the incorporation of the abundance of the focal species itself. The weighted PD10 and AUPhyDC metrics provide a meaningful and straightforward way to integrate conspecific and heterospecific interactions in terms of their contributions to the ecological interactions of a particular individual in a community. An example from the UCSC Meadow community of the effects of weighting on these measures is provided in Figures S1 and S2.

4. Discussion

4.1. Distributions of Phylogenetic Distances in Communities Are Skewed toward Long Distances

Plant communities show a left-skewed distribution of phylogenetic distances between species. We found this pattern for our intensively-studied meadow community, as well as for eight diverse plant communities across California. It has also been noted by others [33]. The left-skewed structure of phylogenetic distances between plant species means that most potential ecological interactions between plant species are between distant relatives. However, the phylogenetic signal decays rapidly for many ecologically important traits, which means that species pairs at short phylogenetic distances are expected to have the strongest ecological interactions. For instance, the probability that two plant species share a fungal pathogen declines from >0.9 if the species are less than 10 Myr apart, and to less than 0.1 by 100 Myr (Figure 2, UCSC Meadow) [25]. Similarly, oak trees suffered ten-fold more herbivory if their neighbor was a close relative than if the neighbor was >100 Myr phylogenetically distant [26]. Besides interactions with plant enemies, a range of other traits important in ecological interactions may also show a phylogenetic signal: e.g., competitive asymmetry among plant pairs [38], pollinator preferences [27], ecological optima in diatoms [39], and animal prey ranges [40].

4.2. Cumulative Phylogenetic Distance Curves for Analysis of Focal Species

As we have shown, the cumulative phylogenetic distance curves can be a rich source of information. The curves themselves can be compared to test predictions about how ecological processes respond to phylogenetic relatedness. For example, Lynch et al. [41] compared the cumulative phylogenetic distance curves with bootstrapped 95% confidence intervals for sets of tree species that differed in their susceptibility to an invasive pest–pathogen complex. Those species that were vulnerable to attack were a phylogenetic subset of the overall tree community, showing shorter than expected phylogenetic distances, and those species for which attack was lethal were an even more restricted subset of closely related species. Another approach is to do comparisons across the full range of phylogenetic distance quantiles to inspect observed patterns in the context of a null model. For example, in an earlier study we drew assemblages randomly from a regional species pool, calculated cumulative phylogenetic distance curves, and generated null distributions for quantile segments from 1% to 50% (extended data Figure 4 in [25]). Then, we compared the observed assemblage to the null distributions, allowing a test for significant under- or over-dispersion at each quantile segment (extended data Figure 5 in [25]).

Although inspection across the full range of quantiles may be most informative, simpler metrics are useful and often necessary for the purposes of comparing large numbers of samples or species. We explored two new metrics for analysis of the structure of one-to-many phylogenetic distances for ecological studies using a focal-species approach: the area under the phylogenetic distance curve (AUPhyDC) and the 10th quantile distance (PD10).

4.3. Uses and Contextual Limits of the AUPhyDC

The AUPhyDC provides a highly integrative measure of phylogenetic distance from a focus species to all other members of the community. The values of AUPhyDC are smaller for species with numerous close relatives and large for species that are phylogenetically isolated. However, there are challenges to the application of AUPhyDC as a general approach. For instance, the maximum possible AUPhyDC value in a community is always equivalent to the value of the longest distance between two species in that community. A focal species for which the distance to its closest relative is also the greatest observed phylogenetic distance in the system would have both the 0th and the 100th quantile equal to that maximum phylogenetic distance, and its AUPhyDC would be the maximum distance times 1. For example, in an assemblage of one monocot species and ten dicot species, all the pairwise distances from the monocot would be 309 Myr (time to most recent common ancestor between monocots and dicots, and the maximum phylogenetic distance for this assemblage). In this case, the AUPhyDC for the monocot would be 309. However, if the assemblage included nonvascular plants, the maximum phylogenetic distance would be 940 Myr (twice the time since the estimated appearance of the earliest land plants). A universal standard for plant communities could be established as the complement area of the AUPhyDC; 940 minus the observed AUPhyDC. This complement AUPhyDC (cAUPhyDC) would have larger values when the focal species had larger numbers of close relatives, and smaller values when phylogenetically isolated. Communities of different types of organisms—fungi, mammals, insects, etc.—have different maximum phylogenetic distances and would require different universal standards. Currently, AUPhyDC is most valuable as a tool for comparing species within a community or within a bounded phylogenetic range.

4.4. Ecological Value of the 10th Quantile Phylogenetic Distance PD10

Although we adopted PD10 because of its superior performance in focal-species (one-to-many) studies, the mean PD10 (MPD10) may also be useful as a community-wide metric in the same way that MPD and MNTD are currently used. For example, in a study of habitat filtering in the pygmy forest community on nutrient-deficient sites in Northern California, Cary [37] found that the degree of nutrient deficiency and the stature (pygmy-ness) of trees was closely related to MPD10 of the assemblage, but that no pattern was detected using either MPD or MNTD. Similarly, Lynch et al. [41] found that the MPD10 was a useful indicator of phylogenetic clustering and strongly predicted susceptibility of tree host species to a novel pest–pathogen complex. The subset of regional tree species that could be attacked by the beetle had a MPD10 ≅ 130 Myr, whereas the more restrictive subset of hosts on which the beetle and fungal pathogen could reproduce were more closely related (MPD10 ≅ 110 Myr), and the smaller subset of tree hosts for which attack was lethal showed the lowest MPD10 (≅65 Myr). These are just two examples of where PD10 may be useful in the analysis of the phylogenetic structure of biological communities, where the more traditional measures of MPD and MNTD have dominated, including patterns and maintenance of diversity, community assembly, and biological invasions [42–44].

Our PD10 metric reflects evolutionary isolation of a focal species from its neighbors. This is analogous to the use of measures of evolutionary distinctiveness (ED) that are commonly applied in discussions of conservation priorities and extinction threats [45,46], but which are seldom used in studies of community structure and species interactions. For the nine plant communities presented here (Figure 2), PD10 and ED are positively correlated (Pearson's r 0.46 to 0.91; Table S2). An advantage of ED is that it provides a single value for evolutionary distinctiveness of each focal species from all other members of an assemblage or clade without reference to particular thresholds; this contrasts to the heuristically selected 10th quantile for PD10, a somewhat arbitrary threshold similar to the median (50th quantile) or 95% confidence intervals. However, PD10 has the important advantage of having a clear conceptual and intuitive linkage to the distribution of phylogenetic distances from a focal species to each of its neighbors, conveying clear information about how community structure can shape species interactions. Such insights are particularly important for comparisons across different communities, or when considering management options, such as manipulating the structure of community composition to effect a desired outcome, e.g., reducing disease intensity. Although ED measures may be useful analytically, the scaling of PD10 is linked to phylogenetic distance itself and is thus more intuitive.

In summary, because it is a simple, comparable, and useful metric, we advocate for the consideration of the 10th quantile of phylogenetic distance (PD10) in focal-species analyses of phylogenetic structure (Figures 3 and 4b). PD10 has the twin benefits of being unaffected by the preponderance of long phylogenetic distances in most communities (unlike meanPD), and integrating important information from multiple close relatives (unlike NTD). For communities with fewer than 10 species, by definition, PD10 is the same as NTD, thus it only provides additional information for larger assemblages. Like both meanPD and NTD (but unlike AUPhyDC), PD10 can be readily compared across studies.

4.5. Abundance Weighting of AUPhyDC and PD10 Permits Inclusion of Conspecific Interactions

Because interactions between conspecifics are often ecologically important, many interactions (e.g., competition, herbivory, pathogen attack, facilitation) are best examined by integrating the effects of both conspecifics and close relatives [24]. Whereas unweighted PD10 and AUPhyDC, such as MNTD and MPD, do not include the distance of a focal species to conspecifics (i.e., where PD = 0 Myr), the weighted versions provide the option of doing so. For a focal species that is numerically dominant, intraspecific interactions will be common and likely ecologically important. A weighted PD10 for that species would be (appropriately) small. In fact, when a focal species makes up more than 10% of the community, the PD10 would be zero. The use of weighted metrics to combine the effects of conspecific and heterospecific neighbors is a powerful feature of applying an individual-centered perspective to phylogenetic ecology.

5. Conclusions

The widespread availability of phylogenetic information has transformed the evolutionary ecology of species interactions. Ecological questions about the performance of focal species in the context of an assemblage of neighboring species of different degrees of relatedness take the form of one-to-many comparisons of phylogenetic distances. Because ecologically important traits often show a phylogenetic signal, the degree of phylogenetic isolation from neighbors may influence the intensity of disease, herbivory, competition, or facilitation experienced by a species. Analysis of the cumulative distribution curves of one-to-many phylogenetic distances, including the area under such curves (AUPhyDC), can help uncover structural differences among species in phylogenetic isolation from their communities. Quantiles such as PD10 provide simple and effective measures of phylogenetic isolation for focal-species analysis while avoiding the biases generated by the left-skewed distributions of phylogenetic distance common to most biological communities. We envision the use of PD10 alongside NTD and meanPD as a standard, potentially more informative metric that can be compared across studies. More focal-species studies are needed to evaluate the effectiveness of these metrics, and we hope future researchers will explore their behavior in their own systems.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d14070521/s1, File S1: Reproducible R code for calculating the one-to-many phylogenetic distance metrics NTD, PD10, meanPD, and AUPhyDC, with example phylogenetic tree from UCSC Meadow community; File S2: Generic R code to create a dated phylogenetic tree and phylogenetic distance matrix from any list of plant species names, based on the supertree R2G2_20140601.new; Table S1: Species lists of angiosperm plants documented at eight University of California Natural Reserves; Figure S1: Comparison of unweighted and weighted PD10 and AUPhyDC metrics, with an example from the UCSC Meadow community; Figure S2: Cumulative abundance-weighted phylogenetic distance curves from the UCSC Meadow community; Table S2: Pearson correlation coefficients among PD10 and evolutionary distinctiveness metrics.

Author Contributions: The authors contributed equally to all aspects of the manuscript. Conceptualization, G.S.G. and I.M.P.; methodology, G.S.G. and I.M.P.; validation, G.S.G. and I.M.P.; formal analysis, G.S.G. and I.M.P.; investigation, G.S.G. and I.M.P.; resources, G.S.G. and I.M.P.; data curation, G.S.G. and I.M.P.; writing—original draft preparation, G.S.G. and I.M.P.; writing—review and editing, G.S.G. and I.M.P.; visualization, G.S.G. and I.M.P.; project administration, G.S.G. and I.M.P.; funding acquisition, G.S.G. and I.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the US National Science Foundation, grants number DEB-0842059, DEB-1136626, and DEB-1655896.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All data used in this study are available in the Supplementary Materials.

Acknowledgments: We thank Pedro Jordano and Montse Vilà at the Spanish National Research Council–Doñana Biological Station (CSIC-EBD) for hosting our sabbatical visits to develop this work. We thank the Jordano, Vilà, Parker, and Gilbert labs for helpful discussions, and S. Grove, N. Lustenhouwer, S. Lynch, M. Melen, E. Aiello, J. Detka and T. Farber for critical comments on the manuscript. We acknowledge that the land on which this research was conducted is the unceded territory of the Awaswas-speaking Uypi Tribe. The Amah Mutsun Tribal Band, comprising the descendants of indigenous people taken to missions in Santa Cruz and San Juan Bautista during Spanish colonization of the Central Coast, is today working hard to restore traditional stewardship practices on these lands and heal from historical trauma.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Cavender-Bares, J.; Kozak, K.H.; Fine, P.V.A.; Kembel, S.W. The merging of community ecology and phylogenetic biology. *Ecol. Lett.* **2009**, *12*, 693–715. [CrossRef] [PubMed]
2. Vamosi, S.M.; Heard, S.B.; Vamosi, J.C.; Webb, C.O. Emerging patterns in the comparative analysis of phylogenetic community structure. *Mol. Ecol.* **2009**, *18*, 572–592. [CrossRef] [PubMed]
3. Harvey, P.H.; Pagel, M.D. *The Comparative Method in Evolutionary Biology*; Oxford University Press: Oxford, UK, 1991.
4. Blomberg, S.P.; Garland, T.; Ives, A.R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **2003**, *57*, 717–745. [CrossRef] [PubMed]
5. Revell, L.J.; Harmon, L.J.; Collar, D.C. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **2008**, *57*, 591–601. [CrossRef]
6. Losos, J.B. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* **2008**, *11*, 995–1003. [CrossRef]
7. Alcantara, S.; Lohmann, L.G. Contrasting phylogenetic signals and evolutionary rates in floral traits of Neotropical lianas. *Biol. J. Linn. Soc.* **2011**, *102*, 378–390. [CrossRef]
8. Baraloto, C.; Hardy, O.J.; Paine, C.E.T.; Dexter, K.G.; Cruaud, C.; Dunning, L.T.; Gonzalez, M.A.; Molino, J.F.; Sabatier, D.; Savolainen, V.; et al. Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *J. Ecol.* **2012**, *100*, 690–701. [CrossRef]
9. Gilbert, G.S.; Webb, C.O. Phylogenetic signal in plant pathogen-host range. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4979–4983. [CrossRef]
10. Wiens, J.J.; Ackerly, D.D.; Allen, A.P.; Anacker, B.L.; Buckley, L.B.; Cornell, H.V.; Damschen, E.I.; Jonathan Davies, T.; Grytnes, J.A.; Harrison, S.P. Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol. Lett.* **2010**, *13*, 1310–1324. [CrossRef]
11. Cavender-Bares, J.; Keen, A.; Miles, B. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* **2006**, *87*, S109–S122. [CrossRef]
12. Willis, C.G.; Halina, M.; Lehman, C.; Reich, P.B.; Keen, A.; McCarthy, S.; Cavender-Bares, J. Phylogenetic community structure in Minnesota oak savanna is influenced by spatial extent and environmental variation. *Ecography* **2010**, *33*, 565–577. [CrossRef]
13. Cavender-Bares, J.; Ackerly, D.D.; Baum, D.A.; Bazzaz, F.A. Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.* **2004**, *163*, 823–843. [CrossRef]
14. Maherali, H.; Klironomos, J.N. Influence of phylogeny on fungal community assembly and ecosystem functioning. *Science* **2007**, *316*, 1746–1748. [CrossRef]
15. Webb, C.O.; Gilbert, G.S.; Donoghue, M.J. Phylodiversity-dependent seedling mortality, size structure, and diseases in a Bornean rain forest. *Ecology* **2006**, *87*, S123–S131. [CrossRef]
16. Gilbert, G.S.; Parker, I.M. The evolutionary ecology of plant disease: A phylogenetic perspective. *Annu. Rev. Phytopathol.* **2016**, *54*, 549–578. [CrossRef]
17. Daehler, C.C. Darwin's naturalization hypothesis revisited. *Am. Nat.* **2001**, *158*, 324–330. [CrossRef]
18. Duncan, R.P.; Williams, P.A. Darwin's naturalization hypothesis challenged. *Nature* **2002**, *417*, 608–609. [CrossRef]
19. Strauss, S.Y.; Webb, C.O.; Salamin, N. Exotic taxa less related to native species are more invasive. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5841–5845. [CrossRef]
20. Ma, C.; Li, S.-p.; Pu, Z.; Tan, J.; Liu, M.; Zhou, J.; Li, H.; Jiang, L. Different effects of invader–native phylogenetic relatedness on invasion success and impact: A meta-analysis of Darwin's naturalization hypothesis. *Proc. Soc. B Biol. Sci.* **2016**, *283*, 20160663. [CrossRef]
21. Gilbert, G.S.; Magarey, R.; Suiter, K.; Webb, C.O. Evolutionary tools for phytosanitary risk analysis: Phylogenetic signal as a predictor of host range of plant pests and pathogens. *Evol. Appl.* **2012**, *5*, 869–878. [CrossRef]
22. Gilbert, G.S.; Briggs, H.M.; Magarey, R. The impact of plant enemies shows a phylogenetic signal. *PLoS ONE* **2015**, *10*, e0123758. [CrossRef]
23. Pearse, I.S.; Hipp, A.L. Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18097–18102. [CrossRef]
24. Liu, X.B.; Liang, M.X.; Etienne, R.S.; Wang, Y.F.; Staehelin, C.; Yu, S.X. Experimental evidence for a phylogenetic Janzen-Connell effect in a subtropical forest. *Ecol. Lett.* **2012**, *15*, 111–118. [CrossRef]
25. Parker, I.M.; Saunders, M.; Bontrager, M.; Weitz, A.P.; Hendricks, R.; Magarey, R.; Suiter, K.; Gilbert, G.S. Phylogenetic structure and host abundance drive disease pressure in communities. *Nature* **2015**, *520*, 542–544. [CrossRef]
26. Yguel, B.; Bailey, R.; Tosh, N.D.; Vialatte, A.; Vasseur, C.; Vitrac, X.; Jean, F.; Prinzing, A. Phytophagy on phylogenetically isolated trees: Why hosts should escape their relatives. *Ecol. Lett.* **2011**, *14*, 1117–1124. [CrossRef]
27. Cirtwill, A.R.; Dalla Riva, G.V.; Baker, N.J.; Ohlsson, M.; Norström, I.; Wohlfarth, I.M.; Thia, J.A.; Stouffer, D.B. Related plants tend to share pollinators and herbivores, but strength of phylogenetic signal varies among plant families. *New Phytol.* **2020**, *226*, 909–920. [CrossRef]

28. Jacquemyn, H.; Merckx, V.; Brys, R.; Tyteca, D.; Cammue, B.P.A.; Honnay, O.; Lievens, B. Analysis of network architecture reveals phylogenetic constraints on mycorrhizal specificity in the genus *Orchis* (Orchidaceae). *New Phytol.* **2011**, *192*, 518–528. [CrossRef]
29. Tucker, C.M.; Cadotte, M.W.; Carvalho, S.B.; Davies, T.J.; Ferrier, S.; Fritz, S.A.; Grenyer, R.; Helmus, M.R.; Jin, L.S.; Mooers, A.O.; et al. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol. Rev.* **2017**, *92*, 698–715. [CrossRef]
30. Pavoine, S. A guide through a family of phylogenetic dissimilarity measures among sites. *Oikos* **2016**, *125*, 1719–1732. [CrossRef]
31. Washburne, A.D.; Morton, J.T.; Sanders, J.; McDonald, D.; Zhu, Q.; Oliverio, A.M.; Knight, R. Methods for phylogenetic analysis of microbiome data. *Nat. Microbiol.* **2018**, *3*, 652–661. [CrossRef]
32. Clarke, K.R.; Warwick, R.M. A taxonomic distinctness index and its statistical properties. *J. Appl. Ecol.* **1998**, *35*, 523–531. [CrossRef]
33. Cadotte, M.W.; Davies, T.J.; Peres-Neto, P.R. Why phylogenies do not always predict ecological differences. *Ecol. Monogr.* **2017**, *87*, 535–551. [CrossRef]
34. Shaner, G.; Finney, R. The effect of nitrogen fertilization on the expression of slow-mildewing resistance in Knox wheat. *Phytopathology* **1977**, *67*, 1051–1056. [CrossRef]
35. Cadotte, M.W.; Dinnage, R.; Tilman, D. Phylogenetic diversity promotes ecosystem stability. *Ecology* **2012**, *93*, S223–S233. [CrossRef]
36. Cobb, R.C.; Meentemeyer, R.K.; Rizzo, D.M. Apparent competition in canopy trees determined by pathogen transmission rather than susceptibility. *Ecology* **2010**, *91*, 327–333. [CrossRef]
37. Cary, K.L. *Physiology and Community Assembly in Mendocino's Pygmy Forest*; University of California: Santa Cruz, CA, USA, 2019.
38. Godoy, O.; Kraft, N.J.B.; Levine, J.M. Phylogenetic relatedness and the determinants of competitive outcomes. *Ecol. Lett.* **2014**, *17*, 836–844. [CrossRef]
39. Keck, F.; Rimet, F.; Franc, A.; Bouchez, A. Phylogenetic signal in diatom ecology: Perspectives for aquatic ecosystems biomonitoring. *Ecol. Appl.* **2016**, *26*, 861–872. [CrossRef]
40. Naisbit, R.E.; Kehrli, P.; Rohr, R.P.; Bersier, L.-F. Phylogenetic signal in predator–prey body-size relationships. *Ecology* **2011**, *92*, 2183–2189. [CrossRef]
41. Lynch, S.C.; Eskalen, A.; Gilbert, G.S. Host evolutionary relationships explain tree mortality caused by a generalist pest-pathogen complex. *Evol. Appl.* **2020**, *14*, 1083–1094. [CrossRef]
42. Honorio Coronado, E.N.; Dexter, K.G.; Pennington, R.T.; Chave, J.; Lewis, S.L.; Alexiades, M.N.; Alvarez, E.; Alves de Oliveira, A.; Amaral, I.L.; Araujo-Murakami, A. Phylogenetic diversity of Amazonian tree communities. *Divers. Distrib.* **2015**, *21*, 1295–1307. [CrossRef]
43. Liendo, D.; Biurrun, I.; Campos, J.A.; García-Mijangos, I.; Pearman, P.B. Effects of disturbance and alien plants on the phylogenetic structure of riverine communities. *J. Veg. Sci.* **2021**, *32*, e12933. [CrossRef]
44. Montaño-Centellas, F.A.; Loiselle, B.A.; Tingley, M.W. Ecological drivers of avian community assembly along a tropical elevation gradient. *Ecography* **2020**, *44*, 574–588. [CrossRef]
45. Redding, D.W.; Mooers, A.Ø. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **2006**, *20*, 1670–1678. [CrossRef]
46. Isaac, N.J.; Turvey, S.T.; Collen, B.; Waterman, C.; Baillie, J.E. Mammals on the EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE* **2007**, *2*, e296. [CrossRef]