Trends in
Ecology & Evolution

CellPress

Opinion

Rebuilding the species pool within a phylogenetic framework

Sijia Yuan (袁思佳)[1], Yuhao Zhao (赵郁豪)[1,2,3], Jean-Philippe Lessard [4], Thomas E. Martin [5,6], Thomas J. Matthews [7,8], and Xingfeng Si (斯幸峰)[1,9,*]

The species pool is a fundamental concept with the potential to connect ecological and evolutionary processes in community ecology and biogeography. However, current definitions of the concept primarily focus on contemporary ecological processes, ignoring the historical and evolutionary processes that generated the species pool. We propose a phylogenetic framework that defines the species pool by inferring evolutionary connectivity between communities in a biogeographic region. Our framework is better suited to addressing questions regarding assembly processes occurring over long time periods, offering deeper evolutionary insights for community ecology and conservation biogeography.

Highlights

The species pool describes the set of species that could potentially disperse to and colonize a region. The way it is delineated can significantly influence the conclusions of ecological studies, thus making it a fundamental concept with broad applications across ecological disciplines.

Considering the potential dispersal of species has improved the delineation of species pools, which establishes a paradigm that takes ecological connectivity into account.

Advances in molecular biology and high-throughput sequencing continue to refine our understanding of phylogenetic relationships and evolutionary histories for many taxa. As a result, phylogenetic trees can be used to estimate biodiversity and the composition of species pools.

The phylogenetic framework we propose has several applications. These include building species pools for islands where all species are endemic, advancing the examination of community structure, and enabling the use of records in biodiversity datasets where taxa cannot be identified to the species level.

Integrating evolutionary history into the definition of the species pool will facilitate greater integration of evolutionary insights into many ecological questions.

The species pool concept in ecology

The **species pool** (see Glossary) is most commonly defined as the set of species capable of dispersing to and establishing in a local community [1–3], making it a central concept in community ecology, island biogeography, and conservation biology [4–6]. With regard to **community assembly** theory, a key question is whether the subset of species in a local community differs from a random subset drawn from a larger pool of species occurring in a biogeographic region [7–9]. However, how the species pool is defined remains debated, and its definition substantially influences biological interpretations of observed patterns [10]. Thus, delineating the species pool is crucial for identifying the predominant processes governing the assembly of ecological communities [11,12].

Delineating a species pool is challenging because it requires knowledge of dynamic species distributions. Communities assemble over thousands or millions of years, yet most species pool definitions focus on contemporary distribution data. For example, an increasingly common approach to delineate the maximum geographic extent of a species pool considers all species occurring in the biogeographic region in which the focal community is nested [13]. This species pool can then be refined by weighting the probability that species could disperse to and persist in the focal community [14,15]. However, both the delineation of the biogeographic extent and the choice of weights are somewhat arbitrary, prompting the development of heuristic methods that explore various thresholds to delineate species pools [16], one of which is the **dispersion field** [12,17,18]. Despite various attempts to refine the definition of species pools, existing definitions rely heavily on the assumption that they are static through time and that dispersal occurs within an ecological time window, ignoring historical factors operating over evolutionary time scales [3,11]. Recent work shows that species within a given contemporarily defined pool tend to share more evolutionary history than expected by chance alone, highlighting the potential role of biogeographic and evolutionary history in shaping species pool composition [19]. Developing a framework reliant on phylogenetic relationships to define the species pool could enhance the rigor of its delineation and advance the concept more broadly.

[1]Center for Global Change and Ecological Forecasting, Zhejiang Zhoushan Island Ecosystem Observation and Research Station, Zhejiang Tiantong Forest Ecosystem National Observation and Research

A phylogenetic framework to incorporate evolutionary history into the species pool definition

Our framework builds upon an established method for delineating species pools—the dispersion field—which is estimated by overlaying the geographic ranges of all species present in a focal community [2,18,20]. This method emphasizes the underlying dispersal process based on the assumption that if a species in a given community has entered the focal community at some point during the assembly process, then all other species in this community could theoretically have done so [17]. The advantages of this approach include its ability to (i) define a community-specific species pool (rather than a region-specific pool) and (ii) provide a spatial representation of the extent and size of the pool from which a focal community can draw species [17,18] (Figure 1B). However, this approach focuses primarily on contemporary dispersal probabilities and may, therefore, misrepresent the pool of potential colonists, given that communities typically assemble over long periods of time [11,21,22]. Here, we extend the existing species-level dispersion-field approach to integrate processes that operate over evolutionary timescales.

First, our approach aims to determine a biologically meaningful geographic extent, or **biogeographic pool**, which should be tailored to a specific ecological or evolutionary

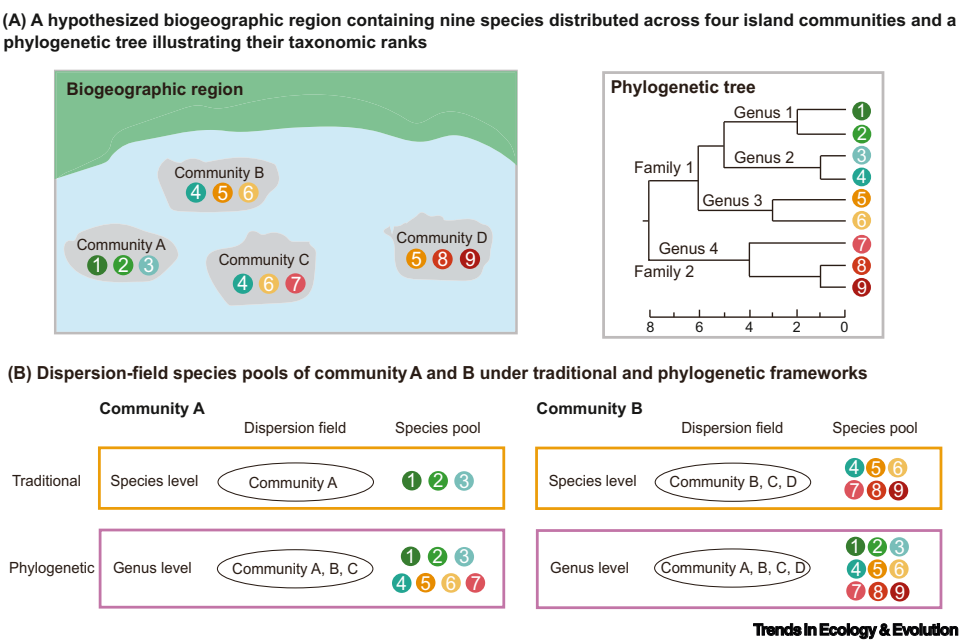

Figure 1. Traditional dispersion-field species pool and the proposed phylogenetic framework integrating evolutionary history into species pool definitions. (A) A hypothesized biogeographic region containing nine species distributed across four island communities and a phylogenetic tree illustrating their taxonomic ranks. (B) Traditional and phylogenetic dispersion-field species pools for Community A (composed entirely of single island endemic species) and Community B (composed of both single island endemic and more widespread species). Under the traditional definition of dispersion fields, the species pool of Community A is restricted to itself because it shares no species with other communities. When building the species pool at the genus level, however, the species pool of Community A expands to include all species in Communities A, B, and C (i.e., Species 1–7), as these three communities share Genus 2. Although Species 4 is absent from Community A, it can be assumed to occupy environments similar to Species 3, given that they belong to the same genus and are close relatives (sister species, assuming no closer relatives are missing from the phylogeny in this case). Although Species 5–7 are more distantly related to the species in Community A (i.e., they belong to different genera and/or families), their inclusion is justified by the assumption that they have the potential to disperse into and establish in Community A through coexistence with Species 4. Species 8 and 9 are excluded because they fall outside the dispersion field (i.e., they are not considered to have the potential to disperse to or establish in the focal community).

Station, Institute of Eco-Chongming, School of Ecological and Environmental Sciences, East China Normal University, Shanghai, China
[2]Institute for Global Change Biology, University of Michigan, Ann Arbor, MI, USA
[3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA
[4]Department of Biology, Concordia University, Montreal, Quebec, Canada
[5]Operation Wallacea, Wallace House, Old Bolingbroke, Spilsby, UK
[6]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada
[7]School of Geography, Earth and Environmental Sciences (GEES) and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, UK
[8]cE3c—Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group, CHANGE—Global Change and Sustainability Institute, Faculty of Agricultural Sciences and Environment, Universidade Dos Açores, Angra do Heroísmo, Portugal
[9]State Key Laboratory for Vegetation Structure, Function and Construction (VegLab), MOE Key Laboratory of Biosystems Homeostasis and Protection, College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China

*Correspondence: six@zju.edu.cn (X. Si).

question. For example, questions focusing on contemporary assembly filters might use a geographic extent smaller than what would be needed for questions related to the deep-time assembly of communities. Subsequently, information on the taxonomic hierarchy of species in the biogeographic pool is used to refine the list of species in the pool (Figures 1 and 2). Specifically, the genera (or other taxonomic ranks) of all species in the focal community are identified. Then, all communities within the biogeographic region containing the same genera (or other taxonomic ranks) as the focal community are included in the (phylogenetic) dispersion field [12] (Figure 1).

This approach traces back to Alfred Russel Wallace, who used hierarchical taxonomic relationships to delineate the biogeographic regions of the world [23], and enables the inference of **evolutionary connectivity**. More recently, Holt *et al.* extended this approach by using time-calibrated molecular phylogenies for the same purpose [24]. The underlying logic is that the deep-time history of gene flow between populations and communities should be reflected in the amount of evolutionary history they share. Thus, taxonomic and phylogenetic relationships can be used to infer the exchange of individuals and their genes between communities, enabling the reconstruction of species pools within a phylogenetic framework [24,25].

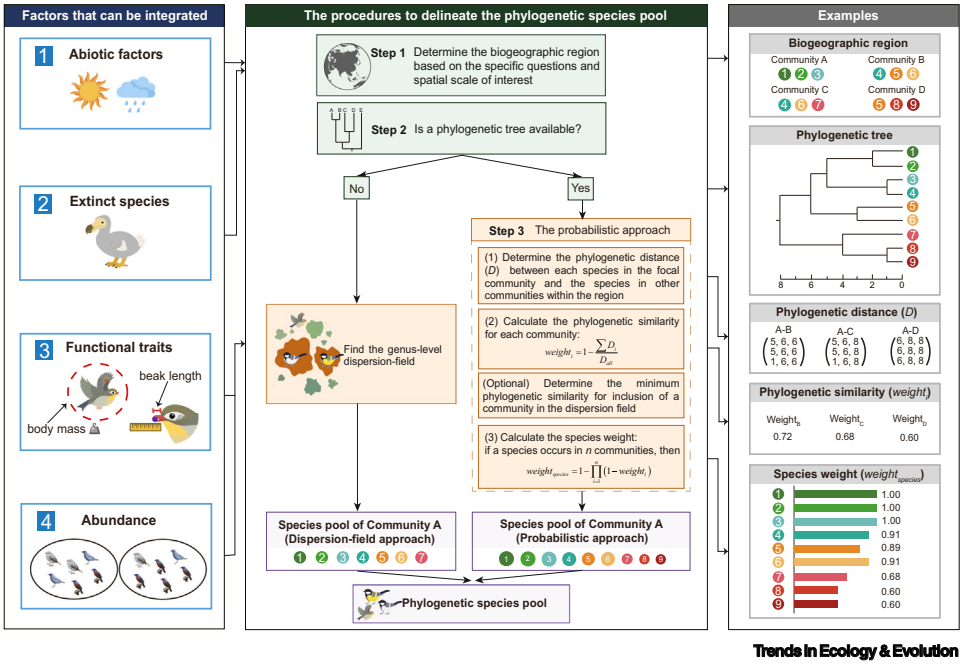

Figure 2. Procedures for delineating the phylogenetic species pool, using the genus-level dispersion field of Community A in Figure 1 as an example. Gray boxes in the right column represent examples corresponding to each step of the probabilistic approach using a phylogenetic tree, while blue boxes in the left column indicate factors that can be integrated into the procedures. Purple boxes in the central column represent the species pool derived from the phylogenetic framework. Arrows between black boxes (i) show different factors and types of data (left column) that can be incorporated into the corresponding steps (central column) and (ii) illustrate examples for those steps (right column). Branch lengths of the phylogenetic tree are used to estimate phylogenetic distances. D_i represents the phylogenetic distance between each species in community i and the focal community, while D_{all} is the total phylogenetic distance across all species pairs between the focal and dispersion-field communities. The subscript i denotes the ith community among the n communities in which a given species occurs.

Glossary

Biogeographic pool: all the species occurring within a particular biogeographic region, the extent of which should be defined according to a specific scientific question.
Cladogenesis: an evolutionary process whereby a single ancestral taxon splits into two or more distinct descendant taxa.
Community assembly: the process by which different species come together to form communities, driven by ecological processes such as dispersal, environmental filtering, and interspecific interactions that determine which species inhabit a particular site.
Community structure: the degree to which the species in a local community are more closely related (clustered) or less closely related (overdispersed) than expected by chance, based on a specific species pool and a null model.
Dispersion field: all communities in the biogeographic region that share a certain number (or proportion) of species with a focal community. In a phylogenetic species pool framework, this concept can be extended to higher taxonomic levels (e.g., the genus level).
Ecological niche similarity: the degree to which species overlap in their niches, including resource utilization, environmental tolerances, and spatiotemporal distribution. The degree of ecological niche similarity between a pair of species can often be inferred from the phylogenetic distance between them.
Evolutionary connectivity: the degree to which gene flow influences evolutionary processes within species across communities; it is typically inferred from the number of conspecific dispersers or the level of genetic similarity between populations.
Null model: a pattern-generating model that produces random samples of ecological data through randomization procedures or by sampling from specified distributions, typically designed to test the role of particular ecological or evolutionary processes.
Species pool: a set of species that can potentially enter a local community, including those that historically existed or originated *in situ*, as well as those capable of dispersing to and establishing within it.
Standardized effect size (SES): a statistical measurement that quantifies

In addition to reflecting historical connectivity, phylogenetic dispersion fields may also reflect **ecological niche similarity**. Phylogenetically closely related species often share similar ecological niches owing to evolutionary conservatism [26], suggesting that they are likely to be adapted to similar environments and, therefore, to co-occur in space [27,28]. We acknowledge, however, that the strength of phylogenetic niche conservatism is unequal across lineages and, therefore, should, where possible, be assessed to facilitate biological interpretations [29]. Nevertheless, the degree of ecological niche similarity among species can often be inferred from phylogenetic distance [30,31], which can then be used to estimate the historical likelihood of dispersal and establishment in the focal community [16].

It should be noted that when a phylogenetic tree for the studied taxon is available, it can be used to reassign species to different taxonomic groups to reflect phylogenetic relationships, if required, as traditional taxonomic groups are often nonmonophyletic. More importantly, a phylogenetic tree can be used, as part of our approach, to estimate the probability of a species being included in the species pool (Figure 2). This probability can be calculated using phylogenetic distances obtained from the phylogenetic tree or estimated from the tree topology if phylogenetic distances are unavailable. These estimates of phylogenetic distance can then serve as proxies for historical connectivity and niche similarity among species [32]. For species already present in the focal community, the relative probability of colonizing and establishing in the focal community can be set as the maximum relative probability value (defined as 1 in this framework). For species not occurring in the focal community, weights can be assigned using the proposed probabilistic phylogenetic approach. First, pairwise phylogenetic distances are computed between all species occurring in the focal community and in communities in the biogeographic region. Second, each community occurring in the biogeographic region is assigned a weight that is positively related to its relative phylogenetic similarity to the focal community (Figure 2). A higher weight implies stronger historical connectivity, making those communities more likely to be part of the species pool of the focal community. Because this approach can be overly liberal, it may be necessary to set a minimum phylogenetic similarity threshold below which communities are excluded from the phylogenetic pool. It is also recommended to explore the effect of threshold choice on the results [18,33]. Finally, the probability that a species in the biogeographic pool colonizes the focal community can be weighted by the divergence-weighted contributions of the communities in which it occurs, given that species vary in their frequencies across the dispersion field (Figure 2).

Our proposed framework can integrate additional factors into the resampling of the phylogenetic species pool. For example, abiotic factors such as climate can be incorporated when delineating biogeographic regions, and species abundances can be considered when calculating species weights. Functional traits, especially those related to dispersal ability, may also be used as a weighting factor for defining the dispersion field [2]. If data on human-driven species extinctions are available, we recommend including them in the species pool because they may, in some cases, include relatively phylogenetically unique species that could influence phylogenetic species pool analyses (Figure 2). We acknowledge that, for some islands, there are likely additional unrecorded (i.e., cryptic) anthropogenic extinctions [34]. However, given their very nature, it is not possible to obtain data on the identities of unrecorded extinct species, meaning they cannot be added to the pool. Future work identifying the number of unrecorded island extinctions across different taxa is needed to better understand the scale of this issue.

Applications of the phylogenetic species pool concept in community ecology and conservation biogeography

Islands serve as a test bed for many important biogeographic hypotheses that rely on the species pool concept, and the use of phylogenetically based species pools could improve our

the degree to which observed patterns of community structure deviate from expectations based on random draws from a predefined pool. SES is calculated as the observed value of a biodiversity metric in the target community minus the mean value of the corresponding metric from a set of randomized values generated using the community's species pool and a specific null model (i.e., the mean of the null distribution), divided by the standard deviation of the null distribution.

Tip-shuffle: a null model algorithm commonly used when inferring phylogenetic community structure that involves randomly shuffling the identities of the terminal tips (e.g., species) of a given phylogenetic tree and fixes site richness.

understanding of insular community assembly and dynamics [35,36]. Previous species pool definitions based on dispersion fields have limited applicability to communities numerically dominated by endemic species, a common characteristic of remote oceanic islands and archipelagos [37]. In extreme cases where all species on an island are single-island endemics, such as mammals on East Falkland Island, approaches that rely on species-level compositional dissimilarity to infer the species pool can only include the island species themselves in the pool [5]. Our method addresses this limitation in two keyways. First, it incorporates a larger set of species that are taxonomically or phylogenetically related to those in the focal community, thereby mitigating the challenges posed by biotas comprised entirely of endemics. Second, it accounts for evolutionary processes such as **cladogenesis** (e.g., Species 3 and Species 4 in Figure 1), which is a common driver shaping species richness patterns on islands [36].

The island species–area relationship (ISAR) is a well-established global ecological pattern whereby island species richness increases with island area [37]. Several drivers of the ISAR have been proposed, including an increase in speciation rate with island area in archipelagos where speciation is an important assembly process [38]. Given this, research on the ISAR has expanded to incorporate evolutionary processes and to evaluate how other dimensions of biodiversity—particularly functional and phylogenetic diversity—scale with island area, so-called island diversity–area relationships (IDARs) [7]. One approach to studying the IDAR involves using **null models** based on a defined species pool to assess how standardized measures of functional and phylogenetic diversity scale with area, thereby providing insights into the relationship between community assembly processes and island area [7]. Thus, our phylogenetic framework can be applied to define a more appropriate island-specific species pool when examining IDARs, helping to identify potential dispersal or environmental filtering and to clarify how these processes scale with island area.

As highlighted in the previous paragraph, species pools are integral to the comparative null model approaches commonly used to infer community assembly processes [39]. This is achieved by testing whether the observed **community structure** deviates from random expectation or from the community structure predicted by the processes included in the species pool definition [5,40,41]. Our phylogenetic framework incorporates evolutionary processes into the definition of the species pool, arguably a more biologically realistic alternative to the traditional definition when dealing with questions involving evolutionary and historical assembly processes. For example, it provides a more rigorous basis for addressing questions such as whether communities are cradles or museums of biodiversity [42], particularly in systems where historical events (e.g., founder effects associated with early colonization) have strongly shaped present-day biodiversity patterns. Box 1 provides a more detailed overview of how this approach can be implemented in practice, and Box 2 provides an illustrative case study applying the phylogenetic framework to infer island community structure.

Our framework extends the scope of community assembly research by overcoming gaps in biodiversity data related to difficulties with species identification, using higher taxonomic ranks as proxies when defining species pools for a given community [53]. One straightforward application lies in the analysis of biodiversity datasets lacking species-level resolution, including those that (i) contain cryptic taxa, (ii) lack species-level phylogenetic trees (e.g., many insect taxa), and (iii) comprise other types of records where species-level identifications are unavailable (e.g., nocturnal rodents recorded in black-and-white images from camera traps). These ambiguities often lead researchers to discard these data, potentially biasing community assembly and species pool analyses. By enabling the inclusion of incomplete or coarsely resolved taxonomic

data, our framework mitigates these biases and expands the scope of analyses for community ecology and conservation assessments in data-limited contexts.

Considerations when applying the phylogenetic framework

Employing our species pool framework requires several important considerations. First, it is based on the assumption that closely related species are generally more likely to exhibit ecological niche similarity [44], and we therefore recommend, where possible, testing for phylogenetic niche conservatism prior to applying this phylogenetic framework. A common approach is to test for phylogenetic signal in various traits when reliable trait data and phylogenetic trees are available [32,54]. A strong phylogenetic signal would indicate that phylogenetic distance is a reliable proxy for ecological niche similarity [19], justifying the use of our probabilistic method. If the phylogenetic signal is weak, we recommend defining the dispersion-field species pool at the genus level without using the probabilistic weighting. When trait data are unavailable, alternative methods, such as ecological niche models, can be used to test for different measures of phylogenetic niche conservatism [54].

Second, in most cases, using the genus level as the primary taxonomic rank for establishing the phylogenetic dispersion field represents a sensible choice. However, where the study question demands it, using the family (or higher) level may be a viable alternative. For example, the use of a family-level species pool could help assess the impact of macroevolutionary processes on the structure of fossil assemblages [55]. In addition, family-level branches in a phylogenetic tree tend to be deeper than those at the genus level, which can create differences in the evaluation of phylogenetic signal. If the phylogenetic signal is weak at the

Box 1. Using the phylogenetic species pool framework to answer key questions in ecology and evolution: Biodiversity cradles or museums?

Our approach offers a new perspective on fundamental questions in ecology and evolution, such as whether regions act as cradles (where species diversity arises from recent rapid speciation) or museums (where species diversity results from the gradual accumulation and preservation of species over time) [42,43]. This classification is often determined by comparing metrics of phylogenetic community structure (e.g., mean pairwise distance) to those derived from various null models [i.e., through the analysis of **standardized effect size (SES)** values] [42] generated with a defined region-specific species pool. In such cases, if a focal community contains a high proportion of young species (i.e., species with shorter terminal branch lengths) relative to the species pool, its phylogenetic structure should be more clustered when compared to null model values (lower SES). Conversely, if the community contains a greater proportion of relatively old species (i.e., species with longer terminal branch lengths), its phylogenetic structure should be more over-dispersed (higher SES) (Figure I). These patterns suggest that historical connectivity and niche conservatism may influence the evaluation of the extent to which ecological processes shape phylogenetic community structure.

Our framework can be used in cradle versus museum classification analyses by defining a region-specific species pool that explicitly accounts for phylogenetic relationships and thus that, in theory, should be better suited for accounting for assembly processes operating at evolutionary timescales. Incorporating deeper branches into the species pool can alter the quantification of community phylogenetic structure, which may in turn affect our inferences regarding the processes driving community assembly and the identification of cradles and museums. For example, if a particular community in a region exhibits a degree of clustering similar to the null expectations under a species-level species pool, but shows a significantly greater degree of clustering than expected under a genus-level species pool, it would provide evidence for a role for environmental filtering as an assembly mechanism that is only apparent after accounting for historical connectivity and phylogenetic niche conservatism. Conclusions on whether this particular region is classified as a museum or a cradle could also differ, depending on the evolutionary history of the additional species in the phylogenetic pool. For example, Figure I provides a hypothetical example of such a scenario. In this example, the genus-level species pools incorporate a number of relatively young/old species that are not in the respective species-level pools, thus strengthening the associated relationship between phylogenetic structure and community evolutionary history, a pattern that should reflect the influence of phylogenetic tree topology on community structure estimation. Further research is needed to determine how this approach works in practice with empirical data for different taxa.

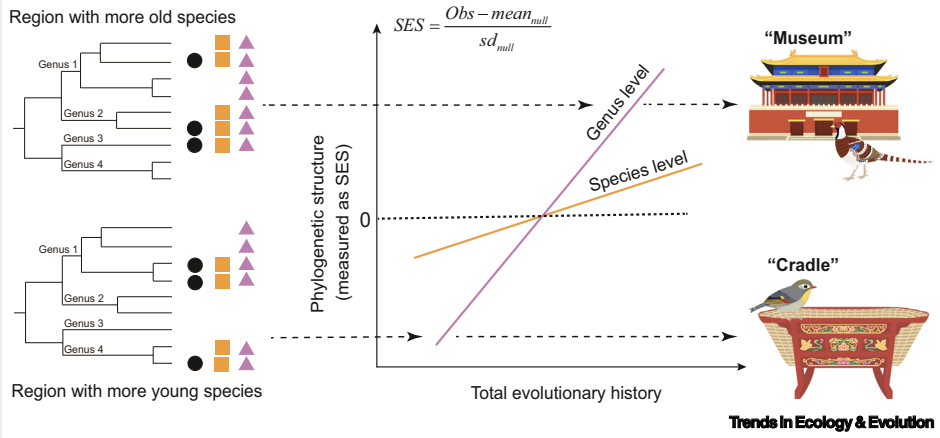

Figure I. Potential applications of the phylogenetic species pool concept when used to assess phylogenetic community structure. A set of simulated focal communities and source pools are used to illustrate hypothetical differences that could emerge when using genus-level and species-level pools, and the resultant effects on the identification of 'museums' and 'cradles'. Black circles represent species within the focal region; orange squares represent the traditional species-level species pool; and purple triangles represent the genus-level species pool inferred from the proposed phylogenetic framework. In the formula for calculating the SES, Obs denotes the phylogenetic metric value (e.g., mean pairwise phylogenetic distance) of the focal region, $mean_{null}$ denotes the mean of the set of corresponding metric values generated by random resampling of a predefined species pool, and sd_{null} represents the standard deviation of the set of metric values derived from random resampling. In this example, total evolutionary history refers to the sum of the length of the terminal branches (in a phylogenetic tree) of all species within a focal region. A 'museum' refers to a region where species diversity results from the gradual accumulation and preservation of species over time, characterized by the presence of many relatively old species, while a 'cradle' refers to a region where species diversity arises from recent rapid speciation (i.e., the presence of many relatively young species). Whether a region is a museum or cradle can be determined by comparing the phylogenetic diversity of a region with values generated from a null model that involves sampling from a predefined pool.

genus level but strong at the family level, it may be appropriate to consider using the family as the primary taxonomic rank.

Third, our phylogenetic framework allows for the inclusion of species that could potentially disperse into the focal community but would otherwise be excluded by traditional species pool definitions. However, there are several considerations regarding the spatial extent of the pool and the inclusion or exclusion of certain species that need to be carefully evaluated based on the specific study question and spatial scale. We illustrate this point with three examples: (i) When the spatial scale (extent) of a study is relatively large, the presence of widespread genera (e.g., the avian genus *Corvus*, which is distributed across almost all continents) in the focal community can result in excessively large and broad dispersion fields [20]. A possible solution is to set a distance threshold based on the dispersal capacity of the study taxon, which can be combined with phylogenetic weights if necessary [18,33]. Alternatively, the spatial extent of the biogeographic region can be adjusted accordingly. (ii) Site-level endemic species (e.g., island endemics) may often be considered unlikely to disperse beyond their native ranges in the absence of human actions, particularly in island systems [56]. Therefore, careful consideration should be given to whether site-level endemic species should be included in the species pool of other sites. (iii) Some closely related species exhibit disjunct geographic distributions (e.g., the family Nelumbonaceae in eastern Asia and eastern North America) [57]. Depending

on the scale of the study, it may be necessary to exclude such cases, for example, by using distance weighting or refining the spatial extent of the source pool region.

Finally, the spatial scale of analysis fundamentally influences perceptions of species distributions and, in turn, how species pools should be defined [11]. At small spatial scales, regionally endemic species may appear widespread because they frequently occur across multiple local communities within the constrained study area. Conversely, at larger scales, the same endemic species often exhibit limited geographic ranges. This scale dependency has important implications for species pool definitions [10]. At local scales, most relatively common species are broadly shared among nearby communities. Thus, the phylogenetically defined species pools will often resemble traditional species-level pools in such cases. As the spatial scale increases, phylogenetic species pools will typically diverge from traditional species-level pools by effectively including species that might otherwise be overlooked.

Box 2. Using the phylogenetic species pool framework to analyses community structure: Null models and extinct species

We use a hypothetical archipelago comprising several study islands to illustrate the use of our proposed phylogenetic species pool framework in inferring community structure and highlight how the choice of null model algorithm and the inclusion or exclusion of recent (i.e., anthropogenic) extinct species may influence the results [44–46]. In this hypothetical example, the biogeographic region is delineated based on the dispersal ability of the focal taxon, allowing for the identification of island-specific species pools. The species on the islands consist primarily of mostly young lineages (relative to the biogeographic pools that include a mix of relatively young and old lineages), with a number of species having been driven extinct from the biogeographic pools by human activities. Our aim is to assess the phylogenetic community structure of the islands in the focal archipelago relative to a given species pool and null model. Due to the presence of numerous relatively young lineages, the island assemblages in the archipelago are expected to exhibit a clustered phylogenetic structure (as shown in Figure I) because, in this case, random draws from the species pools of the focal islands are likely to include some relatively older species. In this example, the genus-level pools contain numerous additional relatively old species that are not present in the species-level pools, and thus, we expect in this case a greater degree of clustering to be apparent when using the former relative to the latter. Analyses using empirical data are required to evaluate whether such a scenario is common in nature, or if the comparison of species-level and phylogenetic pools instead reveals alternative differences in phylogenetic community structure patterns.

Previous studies of phylogenetic structure typically adopt a traditional species pool combined with a **tip-shuffle** null model [7]. However, as with all community assembly analyses of this nature, the choice of null model algorithm can affect the analytical outcomes and biological interpretations [47,48]. In addition to the tip-shuffle null model there are various null model algorithms that randomize the presence values in the species-by-site matrix. Some of these (e.g., those that allow both site richness and species incidences to vary) have been shown to be more prone to Type I error in the analysis of certain biodiversity patterns [49], while others (e.g., those that keep site richness and species incidences fixed or proportional to empirical values) are argued to be more conservative [50]. As the tip-shuffle algorithm is an easy-to-use null model in combination with our phylogenetic species pool approach and is expected to have relatively low Type I error rates when used to infer phylogenetic community structure [5,51,52], it likely represents a sensible general choice for use with our proposed framework. However, further research on appropriate null models to use alongside the proposed framework is needed and, as always, researchers should carefully evaluate which null models are most appropriate for their particular study question. Figure I shows the results of our hypothetical archipelago community structure analysis, comparing the tip-shuffle null model with a liberal null model that randomizes the species-by-site matrix.

In human-modified island systems, the inclusion or exclusion of anthropogenic extinct species in the focal site and species pool can alter the estimation of phylogenetic community structure [7], although the magnitude of the over- or underestimation depends on the evolutionary distinctiveness of the extinct species compared to the extant species. For example, in our hypothetical example, the inclusion of relatively evolutionary distinct extinct species in the species pools (but not on the focal islands themselves) is predicted to result in more clustered community structure of the focal communities. In cases where the phylogenetic species pool already incorporates the genera that the extinct species belong to, the effect of excluding the extinct species on the results should, in many cases, be reduced.

In general, when applying the phylogenetic species pool framework, we recommend simultaneously applying the traditional species-level dispersion-field pool. The results from both the species-level and the phylogenetic dispersion field approaches can then be compared through statistical analyses, with any discrepancies potentially providing insights into the role of evolutionary processes shaping contemporary community structure.

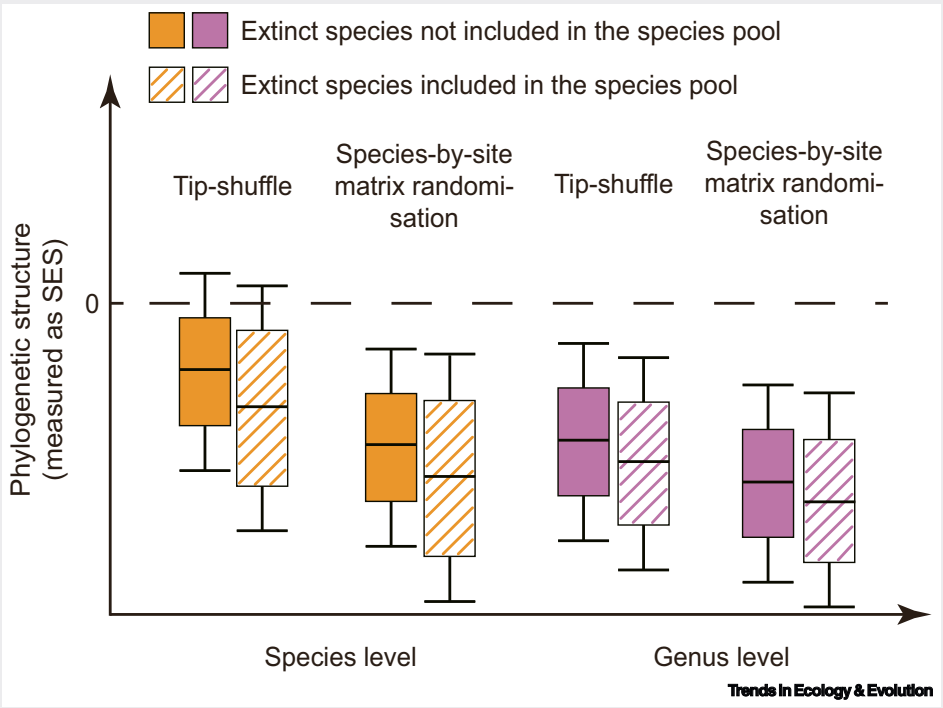

Figure I. Example results of an analysis comparing the traditional and the phylogenetic species pool frameworks (i.e., species vs. genus level pools) when estimating community phylogenetic structure for our hypothetical study islands. Boxplots summarizing the phylogenetic structure measured by the SES values for the focal islands under the traditional (species level) and phylogenetic (genus level) species pools, using two different null model algorithms, one applied to the phylogenetic tree (tip-shuffle) and one applied to the species-by-site matrix (e.g., where site richness and species incidences are allowed to vary in an equiprobable manner) which is known to be prone to Type I error (in this case, providing stronger evidence of clustering). Results are also shown for scenarios with or without the inclusion of relatively evolutionarily distinct extinct species that are part of the wider species pools but are not present on the focal islands themselves.

Outstanding questions

Estimating many biodiversity metrics relies on the definition of species pools. The standardized effect size values we considered here are primarily calculated in relation to α-diversity metrics (e.g., the standardized effect size of mean pairwise phylogenetic distance). Can this approach be used with other biodiversity metrics, such as β-diversity metrics?

The use of null model algorithms can influence the outcomes of community structure analyses, regardless of the choice of species pool. It is likely that analytical results will differ between traditional species-level dispersion-field pools and phylogenetic genus-level dispersion-field pools when inferring community structure. To what extent do these discrepancies provide insights into the role of evolutionary processes shaping contemporary community structure? Are there other important evolutionary processes that are not covered by the phylogenetic species pool approach?

Based on the phylogenetic genus-level species pools proposed here, how can we best integrate functional traits, abundance, and abiotic factors (e.g., contemporary and historical climatic or geographic factors) into this framework to better identify the potential species pool of a given local community? How can we disentangle the effects of evolutionary connectivity from those of phylogenetic niche conservatism on community structure?

Concluding remarks

Incorporating evolutionary processes into the species pool concept within a novel and theoretically grounded phylogenetic framework addresses several limitations associated with traditional species-level pool definitions. In addition, the proposed framework offers new insights and guidance for theoretical ecological research in community ecology and conservation biogeography. Further empirical studies are needed to validate how effectively this framework addresses different ecological and evolutionary questions, as species pools inherently represent simplified approximations of complex metacommunity dynamics influenced by multiple biotic and abiotic factors (see Outstanding questions).

Acknowledgments

The authors acknowledge the helpful feedback from Shaopeng Li and Wande Li, as well as the following funding sources: the National Key Research and Development Program of China (2024YFF1307602), the National Natural Science Foundation of China (3231101666, 32371590, and 32101278), and the NSERC Discovery Grant (RGPIN-2022-04716).

Declaration of interests

The authors declare no competing interests.

References

1. Rahbek, C. and Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4534–4539
2. Lessard, J.-P. *et al.* (2012) Inferring local ecological processes amid species pool influences. *Trends Ecol. Evol.* 27, 600–607
3. Cornell, H.V. and Harrison, S.P. (2014) What are species pools and when are they important? *Annu. Rev. Ecol. Evol. Syst.* 45, 45–67
4. Pearson, D.E. *et al.* (2018) Community assembly theory as a framework for biological invasions. *Trends Ecol. Evol.* 33, 313–325
5. Si, X. *et al.* (2022) Phylogenetic and functional clustering illustrate the roles of adaptive radiation and dispersal filtering in jointly shaping late-Quaternary mammal assemblages on oceanic islands. *Ecol. Lett.* 25, 1250–1262
6. van der Plas, F. *et al.* (2023) Universal beta-diversity-functioning relationships are neither observed nor expected. *Trends Ecol. Evol.* 38, 532–544
7. Matthews, T.J. *et al.* (2023) A global analysis of avian island diversity–area relationships in the Anthropocene. *Ecol. Lett.* 26, 965–982
8. Soares, F.C. *et al.* (2024) Patterns and drivers of taxonomic and functional change in large oceanic island bird assemblages. *Glob. Ecol. Biogeogr.* 33, e13902
9. Hsieh, C. *et al.* (2024) Evolutionary history and environmental variability structure contemporary tropical vertebrate communities. *Glob. Ecol. Biogeogr.* 33, e13829
10. Catano, C.P. *et al.* (2021) Species pool size alters species–area relationships during experimental community assembly. *Ecology* 102, e03231
11. Jarzyna, M.A. *et al.* (2025) Emergence and dynamics of regional species pools. *Glob. Ecol. Biogeogr.* 34, e70046
12. Carstensen, D.W. *et al.* (2013) Introducing the biogeographic species pool. *Ecography* 36, 1310–1318
13. Zhao, Y. *et al.* (2024) Land-use change interacts with island biogeography to alter bird community assembly. *Proc. R. Soc. B Biol. Sci.* 291, 20232245
14. Triantis, K.A. *et al.* (2022) Deterministic assembly and anthropogenic extinctions drive convergence of island bird communities. *Glob. Ecol. Biogeogr.* 31, 1741–1755
15. Coelho, M.T.P. *et al.* (2023) The geography of climate and the global patterns of species diversity. *Nature* 622, 537–544
16. Karger, D.N. *et al.* (2016) Delineating probabilistic species pools in ecology and biogeography. *Glob. Ecol. Biogeogr.* 25, 489–501
17. Lessard, J.-P. *et al.* (2012) Strong influence of regional species pools on continent-wide structuring of local communities. *Proc. R. Soc. B Biol. Sci.* 279, 266–274
18. Borregaard, M.K. *et al.* (2020) Dispersion fields reveal the compositional structure of South American vertebrate assemblages. *Nat. Commun.* 11, 491
19. Loza, M.I. *et al.* (2017) Phylogenetic patterns of rarity in a regional species pool of tropical woody plants. *Glob. Ecol. Biogeogr.* 26, 1043–1054
20. Graves, G.R. and Rahbek, C. (2005) Source pool geometry and the assembly of continental avifaunas. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7871–7876
21. Frishkoff, L.O. *et al.* (2022) Evolutionary opportunity and the limits of community similarity in replicate radiations of island lizards. *Ecol. Lett.* 25, 2384–2396
22. Ricklefs, R.E. (2015) Intrinsic dynamics of the regional community. *Ecol. Lett.* 18, 497–503
23. Wallace, A.R. (1876) *The Geographical Distribution of Animals; with a Study of the Relations of Living and Extinct Faunas as Elucidating the past Changes of the Earth's Surface*, Harper & Brothers, Publishers
24. Holt, B.G. *et al.* (2013) An update of Wallace's zoogeographic regions of the world. *Science* 339, 74–78
25. Mazel, F. *et al.* (2017) Global patterns of β-diversity along the phylogenetic time-scale: the role of climate and plate tectonics. *Glob. Ecol. Biogeogr.* 26, 1211–1221
26. Qian, H. *et al.* (2024) Effects of climate and environmental heterogeneity on the phylogenetic structure of regional angiosperm floras worldwide. *Nat. Commun.* 15, 1079
27. Gómez, J.M. *et al.* (2010) Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature* 465, 918–921
28. Burns, J.H. and Strauss, S.Y. (2011) More closely related species are more ecologically similar in an experimental test. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5302–5307
29. Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol. Lett.* 11, 995–1003
30. Furey, G.N. and Tilman, D. (2023) Plant chemical traits define functional and phylogenetic axes of plant biodiversity. *Ecol. Lett.* 26, 1394–1406
31. Hähn, G.J.A. *et al.* (2024) Global decoupling of functional and phylogenetic diversity in plant communities. *Nat. Ecol. Evol.* 9, 237–248
32. Crisp, M.D. and Cook, L.G. (2012) Phylogenetic niche conservatism: what are the underlying evolutionary and ecological causes? *New Phytol.* 196, 681–694
33. Lessard, J.-P. *et al.* (2016) Process-based species pools reveal the hidden signature of biotic interactions amid the influence of temperature filtering. *Am. Nat.* 187, 75–88
34. Cooke, R. *et al.* (2023) Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. *Nat. Commun.* 14, 8116
35. Hébert, K. *et al.* (2021) Source pool diversity and proximity shape the compositional uniqueness of insular mammal assemblages worldwide. *J. Biogeogr.* 48, 2337–2349
36. Whittaker, R.J. *et al.* (2017) Island biogeography: taking the long view of nature's laboratories. *Science* 357, eaam8326
37. MacArthur, R.H. and Wilson, E.O. (2016) *The Theory of Island Biogeography*, Princeton University Press
38. Valente, L. *et al.* (2020) A simple dynamic model explains the diversity of island birds worldwide. *Nature* 579, 92–96
39. Münkemüller, T. *et al.* (2020) Dos and don'ts when inferring assembly rules from diversity patterns. *Glob. Ecol. Biogeogr.* 29, 1212–1229
40. Gotelli, N.J. and Graves, G.R. (1996) *Null Models in Ecology*, Smithsonian Institution Press
41. Marino, C. *et al.* (2024) The anthropocene biogeography of alien birds on islands: drivers of their functional and phylogenetic diversities. *Ecol. Lett.* 27, e14465
42. Lu, L.-M. *et al.* (2018) Evolutionary history of the angiosperm flora of China. *Nature* 554, 234–238
43. Rangel, T.F. *et al.* (2018) Modeling the ecology and evolution of biodiversity: biogeographical cradles, museums, and graves. *Science* 361, eaar5452
44. Faurby, S. *et al.* (2026) Quantifying the unrecorded loss of avian phylogenetic diversity. *Ecography* 2026, e08267
45. Gotelli, N.J. (2000) Null model analysis of species co-occurrence patterns. *Ecology* 81, 2606–2621
46. Matthews, T.J. *et al.* (2024) The global loss of avian functional and phylogenetic diversity from anthropogenic extinctions. *Science* 386, 55–60
47. Ulrich, W. and Gotelli, N.J. (2010) Null model analysis of species associations using abundance data. *Ecology* 91, 3384–3397
48. Gotelli, N.J. and Ulrich, W. (2012) Statistical challenges in null model analysis. *Oikos* 121, 171–180
49. Ulrich, W. and Gotelli, N.J. (2007) Null model analysis of species nestedness patterns. *Ecology* 88, 1824–1831
50. Ulrich, W. and Gotelli, N.J. (2012) A null model algorithm for presence–absence matrices based on proportional resampling. *Ecol. Model.* 244, 20–27
51. Hardy, O.J. (2008) Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J. Ecol.* 96, 914–926
52. Pigot, A.L. and Etienne, R.S. (2015) A new dynamic null model for phylogenetic community structure. *Ecol. Lett.* 18, 153–163
53. Hortal, J. *et al.* (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* 46, 523–549

54. Pyron, R.A. *et al.* (2015) Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biol. Rev.* 90, 1248–1262
55. He, J. *et al.* (2018) Cenozoic evolution of beta diversity and a Pleistocene emergence for modern mammal faunas in China. *Glob. Ecol. Biogeogr.* 27, 1326–1338
56. Veron, S. *et al.* (2019) Distribution and relative age of endemism across islands worldwide. *Sci. Rep.* 9, 11693
57. Yin, X. *et al.* (2021) Niche overlap and divergence times support niche conservatism in Eastern Asia–eastern North America disjunct plants. *Glob. Ecol. Biogeogr.* 30, 1990–2003