Ecology Letters
September 2021, Volume 24 Issue 9 Pages 1988-2009
https://doi.org/10.1111/ele.13778
https://archimer.ifremer.fr/doc/00694/80643/

The dimensionality and structure of species trait spaces

Mouillot David [1, 2, *] , Loiseau Nicolas [1], Grenié Matthias [3], Algar Adam C. [4], Allegra Michele [5], Cadotte Marc W. [6], Casajus Nicolas [7], Denelle Pierre [8], Guéguen Maya [9], Maire Anthony [10], Maitner Brian [11], McGill Brian J. [12], McLean Matthew [13], Mouquet Nicolas [1, 7], Munoz François [14], Thuiller Wilfried [9], Villeger Sébastien [1], Violle Cyrille [3], Auber Arnaud [15]

[1] MARBEC, University of Montpellier, CNRS, IFREMER, IRD, Montpellier, France
[2] Institut Universitaire de France, IUF, Paris, France
[3] Centre d'Ecologie Fonctionnelle et Evolutive—UMR 5175 CEFE, University of Montpellier, CNRS, EPHE, University of Paul Valéry, IRD, Montpellier, France
[4] Department of Biology, Lakehead University, Thunder Bay, ON, Canada
[5] Institut de Neurosciences de la Timone, Aix Marseille Université, UMR 7289, CNRS, Marseille, France
[6] Department of Biological Sciences, University of Toronto-Scarborough, Toronto, ON, Canada
[7] FRB—CESAB, Institut Bouisson Bertrand, Montpellier, France
[8] Biodiversity, Macroecology & Biogeography, University of Goettingen, Göttingen, Germany
[9] Laboratoire d'Ecologie Alpine, Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LECA, Grenoble, Fr
[10] EDF R&D, LNHE (Laboratoire National d'Hydraulique et Environnement), Chatou, France
[11] Department of Ecology and Evolutionary Biology, University of Connecticut, Mansfield, CT, USA
[12] School of Biology and Ecology and Mitchell Center for Sustainability Solutions, University of Maine, Orono, ME, USA
[13] Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada
[14] LiPhy (Laboratoire Interdisciplinaire de Physique), Université Grenoble Alpes, Grenoble, France
[15] IFREMER, Unité Halieutique Manche Mer du Nord, Laboratoire Ressources Halieutiques, Boulogne-sur- Mer, France

* Corresponding author : David Mouillot, email address : david.mouillot@umontpellier.fr

---

Abstract :

Trait- based ecology aims to understand the processes that generate the overarching diversity of organismal traits and their influence on ecosystem functioning. Achieving this goal requires simplifying this complexity in synthetic axes defining a trait space and to cluster species based on their traits while identifying those with unique combinations of traits. However, so far, we know little about the dimensionality, the robustness to trait omission and the structure of these trait spaces. Here, we propose a unified framework and a synthesis across 30 trait datasets representing a broad variety of taxa, ecosystems and spatial scales to show that a common trade- off between trait space quality and operationality appears between three and six dimensions. The robustness to trait omission is generally low but highly variable among datasets. We also highlight invariant scaling relationships, whatever organismal complexity, between the number of clusters, the number of species in the dominant cluster and the number of unique species with total species richness. When species richness increases, the number of unique species saturates, whereas species tend to disproportionately pack in the richest cluster. Based on these results, we propose some rules of thumb to build species trait spaces and estimate subsequent functional diversity indices.

Keywords : complexity, functional ecology, hypervolume, species clustering, species uniqueness

Introduction

Biodiversity comprises a great variety of organismal forms, functions, diets, physiologies and life histories — hereafter called traits — that have been shaped by large-scale evolutionary and ecological processes (Schluter 1993; Reich *et al.* 1999) and that have important implications for ecosystem functioning (Hector *et al.* 1999; Duffy *et al.* 2001). Thus, quantifying and characterizing trait variation among species is key to understand species assembly rules (Bruelheide *et al.* 2018; Jarzyna *et al.* 2020), evolutionary dynamics (Deline *et al.* 2018; Pigot *et al.* 2020), and ecosystem functioning (Gagic *et al.* 2015; Cadotte 2017) but also to predict biodiversity responses to global changes (McLean *et al.* 2019; Rüger *et al.* 2020) and to guide conservation efforts (Pollock *et al.* 2017; Sala *et al.* 2021). For instance, experiments show that plant communities with higher levels of trait diversity are more productive and have a higher resource use efficiency by intercepting more light, taking up more nitrogen, and occupying more of the available space (Spehn *et al.* 2005) but can also limit plant disease risks (Le Bagousse-Pinguet *et al.* 2021).

Yet, owing to the increasing availability of widespread — but also incomplete and heterogeneous — information on multiple traits collected with various methods across most kingdoms of life (Jones *et al.* 2009; Schneider *et al.* 2017; Perez *et al.* 2019; Kattge *et al.* 2020), the characterization of species ecological strategies and relationships with environmental conditions is becoming more complex and multidimensional than ever (Villeger *et al.* 2011; Bruelheide *et al.* 2018). Reducing this complexity has both theoretical and practical benefits. First, clustering thousands of species into a limited number of entities sharing similar trait values can reveal the amount of functional vulnerability within assemblages (Mouillot *et al.* 2014) or a functional backbone common to separate geographic realms (McLean *et al.* 2021). Second, many traits are strongly correlated owing to life-history trade-offs or adaptive constraints, suggesting that trait diversity within a clade is more limited than expected (Winemiller *et al.* 2015; Díaz *et al.* 2016; Pigot *et al.* 2020). Birds with relatively long, narrow wings, pointed tips, and strong sweep back (such as those of a swallow) fly at high speeds but are energetically inefficient and cannot fly over long distances (Savile 1957). Third, the hyper-dimensionality of trait spaces, where species are placed according to their combinations of traits, prevents the computation of hypervolume-based functional diversity indices or null models to test community assembly hypotheses (Blonder *et al.* 2014; Maire *et al.* 2015). Fourth, predicting biodiversity and ecosystem trajectories under various environmental scenarios needs parsimonious trait-based models (Barros *et al.* 2017; Cooke *et al.* 2019b; Rüger *et al.* 2020) since the use of too many traits may induce overfitting (Bernhardt-Römermann *et al.* 2008).

However, we still lack a unified methodological framework to assess the different aspects of a species trait space. The dimensionality and the structure of a species trait space are indeed two sides of the same coin since they both refer to its complexity, i.e. the way species and their traits are organized in this space. We also lack a synthesis on the main factors shaping the different aspects of species trait spaces. The degree of organismal complexity, which is related to the diversity of cell types (Valentine *et al.* 1994), can indeed influence the complexity of species trait space following key functional innovation in multicellular clades (Knoll 2011; Cox *et al.* 2021; Sosiak & Barden 2021). The environment can also be crucial in determining the course of multicellular evolution and organismal complexity, with aggregative multicellularity evolving more frequently on land while clonal multicellularity is more frequent in water (Fisher *et al.* 2020). On the other hand, the number of species and trait characteristics are likely to influence the complexity of species trait spaces beyond the type of organism and the environment (Zhu *et al.* 2017; Kohli & Jarzyna 2021). Yet, the relative importance of these different potential drivers has never been tested across kingdoms and realms for a vast number and diversity of traits and taxa.

A first critical aspect of a species trait space refers to the well-known dimensionality issue (Laughlin 2014; Maire *et al.* 2015). While dimension reduction is appealing, the devil lies in the details. Indeed, going from a large number of traits to a reduced trait space (Figure 1a-d), that represents meaningful ecological dimensions or axes, is conceptually and methodologically difficult (Maire *et al.* 2015; Winemiller *et al.* 2015; Pigot *et al.* 2020; Sosiak & Barden 2021). High-dimensional spaces might indeed be required to fully capture trait variation among species (Carscadden *et al.* 2017) or clades (Cooney *et al.* 2017). Moreover, the extent to which collected traits, some being potentially uninformative, redundant or incomplete, can be summarized with a few dimensions to reliably represent the diversity of organism forms and functions has not been quantitatively tested across a large set of taxa, ecosystems and traits.

A second key aspect of any species trait space is its robustness to the choice or the omission of traits so its capacity to consistently position species relative to each other whatever the sub-selection of traits for a given goal (environmental filtering, competitive interactions, etc.). This capacity ultimately determines the confidence by which we can estimate metrics like species trait dissimilarity or functional diversity (Carscadden *et al.* 2017; Zhu *et al.* 2017; Kohli & Jarzyna 2021). However, this robustness has been largely overlooked and deserves a dedicated analysis across multiple datasets where the number, completeness, correlation, and type of traits cover a broad range of options.

A third key aspect of any species trait space relates to its structure, and particularly how species are distributed and clustered in that space. Species with very similar traits are likely to play comparable roles in ecosystems (Dehling *et al.* 2016; Pigot *et al.* 2020; Sosiak & Barden 2021), and are packed within a trait space into clusters (Figure 1d). The size (i.e. species richness) of these clusters relates to functional redundancy (Walker 1992; Fonseca & Ganade 2001), which could act as an insurance against the loss of certain combinations of traits and the disruption of ecosystem functioning under disturbance (Sanders *et al.* 2018; McLean *et al.* 2019). The other side of the same coin is functional uniqueness represented by species having no neighbors in the trait space owing to their unique combinations of traits (species B and D in Figure 1d). Several studies suggested that, beyond the positive influence of species trait diversity on ecosystem functioning (Gross *et al.* 2017; Craven *et al.* 2018), these unique species can play key and irreplaceable functional roles (O'Gorman *et al.* 2011; Pigot *et al.* 2016a; Maire *et al.* 2018; Le Bagousse-Pinguet *et al.* 2021). The filling of this trait space through evolutionary history, and more particularly the emergence of species with unique traits, has also motivated numerous studies investigating specialization in clades or competition footprint across the Tree of life (Ricklefs 2010; Cornwell *et al.* 2014; Stubbs & Benton 2016; Phillips *et al.* 2018; Jarzyna *et al.* 2020; Cox *et al.* 2021). Yet, we still lack a flexible framework in which the number and composition of species clusters but also unique species are automatically detected regardless of the shape, the density in terms of species richness and the dimensionality of the trait space in which they are embedded.

Here, we propose a unified and flexible framework to assess (*i*) the optimal number of axes representing species trait diversity (dimensionality), (*ii*) the consistency of the trait space in species placement when sub-setting a limited number of traits (robustness), and (*iii*) the distribution of species among clusters including the proportion of unique species (structure). To better understand the drivers of these three key aspects, we apply our framework on 30 trait datasets spreading across most kingdoms of life (e.g. bacteria, plants, vertebrates) and biomes (terrestrial and marine) at different scales (local to global), and spanning two orders of magnitude in species richness and one order of magnitude in the number of traits with different types (e.g. continuous, categorical, etc..) and varying proportions of missing values (Table 1). To disentangle the drivers of trait space complexity, we then model the dimensionality, the robustness and the structure of these 30 trait spaces as a function of the type of species, the type of ecosystem, the number of species, the number of traits, the type of traits, the correlation between traits and the proportion of missing values. Ultimately, we provide guidance to deal with the heterogeneity and incompleteness of species trait databases when building species trait spaces and assessing trait-based metrics in community ecology, evolution and biogeography.

Materials and Methods

Building species trait space

Among the myriad of methods proposed to reduce the dimensionality of data (Laughlin 2014; Kraemer *et al.* 2018; Nguyen & Holmes 2019), we chose one that is commonly used in ecology, based on well-established ordination techniques, and flexible enough to be adapted to any kind of trait data. Our goal is not to review or compare existing methods but rather to assemble a suite of methods able to extract the main features of any species trait space and test their drivers.

First, we calculated trait dissimilarity between species pairs using the Gower pairwise distance (Gower & Legendre 1986). This metric can handle multiple types of data (e.g., categorical, ordinal and continuous traits) and is also less sensitive to missing values than other distance estimation methods (Podani & Schmera 2006; Pavoine *et al.* 2009). The dissimilarity between two species is only evaluated on traits with known values for both species but this dissimilarity is standardized across all pairs whatever the number of traits considered. This step (Figure 1b) was carried out with the *daisy*() function in the *cluster* R package.

Second, we performed ordination of species in a space of reduced dimensionality by mean of Principal Coordinates Analysis (PCoA), which identifies orthogonal axes along which trait dissimilarity is decomposed (Legendre & Legendre 1998). For this step (Figure 1c) we used the *pcoa()* function in the *ape* R package.

Quality of species trait space

To assess the dimensionality and robustness of species trait spaces, we needed a metric measuring the degree of distortion between the initial trait distance matrix between species pairs (Gower distance on all traits) and the distance matrix after dimensionality reduction (Euclidean distance on PCoA axes) or after removing traits (Gower distance on the sub-selection of traits), respectively. We assumed that a trait space is a high-quality representation of the full dataset if distances between species in that space are close to the initial distances computed with all traits (Maire *et al.* 2015). The approach of comparing the similarity of two distance matrices has precedent in Mantel tests (Legendre & Legendre 1998), although the end goal here is quite different – producing a metric of robustness for low-dimensional trait space. Indeed, Mantel tests only correlate values or ranks between two distance matrices, ignoring the global co-ranking between species and their neighborhood which are key features of species trait space when the ultimate goal is to cluster species and identify functionally unique ones (Pimiento *et al.* 2020a).

Several measures of trait space quality have been proposed (Mérigot *et al.* 2010; Maire *et al.* 2015), but we chose a new one in the field of ecology with five key properties that overcome classical limitations: (*i*) being unitless so independent of the number, range or value of traits, (*ii*) being standardized between 0 and 1 with a clear and intuitive interpretation of these extreme values, *(iii)* avoiding the dilemma of whether or not to square the error, which arises in distance-based quality metrics, (*iv*) being asymmetric by construction so only considering that the lower-dimensional distance matrix is a poorer representation of species distribution in trait space compared to the initial distance matrix, and (*v*) proposing a common, albeit arbitrary, threshold to define quality.

This method is based on the co-ranking matrix Q which compares the ranking of distance between objects in the initial distance matrix and in a lower-dimensional space (Lee & Verleysen 2009). In our case, let us denote by $\delta_{i,j}$ the distance between species i and j in the initial trait matrix (Figure 1a) and $d_{i,j}$ their distance in the lower-dimension matrix (Figure 1c). Then, for any fixed species i, we assessed the ranks of the distances between this species i and all other S-1 species j in both the initial and the lower-dimensional matrix denoted as $\rho_{i,j}$ and $r_{i,j}$, respectively. These ranks varied between 1 and (S-1) with S being the total number of species. The co-ranking matrix Q is of size (S-1) by (S-1) and has for elements the number of species pairs that have the rank k in the initial (all traits) Gower distance matrix and the rank l in the lower-dimensional (PCoA axes) Euclidean distance matrix (Figure 1e). Since the roles played by species i and species j are asymmetric, matrix Q sums at $S(S$-1), so the total number of pairs (S-1) made by each of the S species.

Then, we defined the rank error to be the difference $\rho_{i,j}$ - $r_{i,j}$. If there is no error, i.e. a perfect match in species neighbors between the initial and the lower-dimensional distance matrices, then Q is a diagonal matrix, i.e. ranks k and l will be similar so $\rho_{i,j}$ - $r_{i,j}$=0 for all species pairs. At the opposite, rank mismatches or errors, due to dimensionality reduction or trait omission, induce off-diagonal species pairs in this co-ranking matrix (Figure 1e). These off-diagonal species pairs represent pairs that come at a lower distance rank (intrusion) or at a higher distance rank (extrusion) in the lower-dimensional space compared to the initial space (Lee & Verleysen 2009).

To assess whether the lower-dimensional space was a good representation of the initial space, we needed an asymmetric measure. In other words, a measure that compares the ranks of species pairs in the lower-dimensional matrix to those of the initial matrix and not the way around. A spearman-rank correlation is symmetric (the correlation between A and B equals the correlation between B and A) since it compares the ranks without any primary structure like in Mantel tests. We thus chose the Area Under the Curve (AUC) criteria, which is based on the Somer's *D* statistic, as an asymmetric rank measure (Somers 1962). AUC is unitless and varies between 0 and 1. A value of 1 represents the best-case scenario where the ranking of species pairs would be perfectly preserved between the initial and the lower-dimensional distance matrix (Kraemer *et al.* 2018). A rule of thumb to interpret this metric is that above 0.7 dimensionality reduction can be considered as good or acceptable and above 0.8 as excellent. Below 0.5 the lower-dimensional space is a poor representation of the initial trait space while 0 means as good as random. It corresponds to the null or independence hypothesis in Mantel tests (Legendre & Legendre 1998). More details can be found in Kraemer *et al.* (2018) who developed the *dimRed* and *coRanking* R packages for computing the co-ranking matrix Q with the function *coranking* and then the AUC metric with the function *AUC_lnK_R_NX*.

Complementary to the AUC metric, which is only based on ranks so potentially weakly influenced by some extreme distortion values, we also compared the initial and lower-dimensional distances between species pairs by using the Euclidean distance for multidimensional spaces, also known as the Mean Absolute Deviation (MAD) (Maire *et al.* 2015).

Dimensionality of species trait space

To determine how many dimensions are needed to build a trait space of enough quality that correctly positions species between each other, we used two approaches: a parsimonious one based on the elbow inflection point for the AUC metric and the other one based on a quality threshold for the AUC metric, both tested on 1 to 20 PCoA axes. The idea behind the elbow method is to maximize a given benefit (AUC gain in our case) while reducing the cost (number of dimensions in our case) (Thorndike 1953). Consequently, the inflection point corresponds to the additional PCoA axis above which the benefit becomes lower than the cost (Supplementary Figure 2). This elbow method is classically used in dimensionality analyses (Nguyen & Holmes 2019) but never in combination with AUC.

As a complementary method, we used the AUC quality threshold of 0.7 to determine the dimensionality of the trait space so here the cumulated number of PCoA axes needed to obtain a good or acceptable positioning of species in the lower-dimensional space compared to the initial one based on all traits. This dimensionality assessment is more subjective than the elbow one since based on an arbitrary threshold. However, it has the merit of providing a standardized, so comparable, quality value across datasets for the low-dimensional representations.

The amount of variance explained by the PCoA axes could also be considered as a quality metric of species trait space (Pimiento *et al.* 2020b) like with Principal Components Analyses (PCA) (Pigot *et al.* 2020; Rüger *et al.* 2020). Yet, for non-Euclidean distances like Gower, PCoA axes may obtain negative eigenvalues corresponding to imaginary dimensions (Legendre & Legendre 1998). In that case, the sum of all positive eigenvalues (real axes) is higher than the total variance of data. This intuitive additional piece of information was nonetheless included in our study through the examination of the relationship between the AUC-based dimensionality and the number of axes necessary to explain 50% of trait variation. The proportion of explained variance by PCoA axes was extracted using the ape::pcoa() *R* function.

Robustness to trait omission

To test the robustness, or the lack of sensitivity, of the trait space to trait omission or sub-selection, we randomly removed between 10% and 80% (increments of 10%) of the total number of traits, and then estimated a new Gower distance between all species pairs for each removal percentage; we did not use PCoA axes in this robustness analysis, only traits. Then, we assessed the level of congruence between the initial distance matrix and the lower-dimensional distance matrix by computing the AUC and MAD metrics. These simulations were performed 100 times for each removal percentage. We then extracted an index of robustness defined as the opposite of sensitivity so the mean loss of AUC when 50% of the traits are removed.

Species clustering and uniqueness

To cluster species in the trait space and potentially identify unique species we used the "clustering by fast search and find of density peaks" algorithm which is based on initial pairwise distances and does not require dimensionality reduction (Rodriguez & Laio 2014). Yet, the robustness of the clustering critically depends on the robustness of pairwise species distances to trait omission. Among the many clustering algorithms that have been proposed (Jain & Dubes 1988; Xu & Tian 2015; Condon *et al.* 2016), this one combines the advantages of (*i*) clustering objects regardless of the shape and dimensionality of the space in which they are embedded, (*ii*) detecting isolated objects automatically independently of their number, and (*iii*) making the number and size of clusters emerge with no a priori expectation or arbitrary choice.

In our case, this algorithm first computed the density of neighbors for each species, defined as the number of species that are within a given small distance d_0 (Figure 1f). Given this density, the algorithm then relied on two basic principles: (1) cluster centers were species characterized by a higher density of neighbors than their own neighbors and by a relatively large distance from other species with a higher density of neighbors, and (2) isolated or unique species had no neighbors at maximum d_0 (zero density or redundancy). Once cluster centers and unique species were identified, all remaining species were assigned to a cluster corresponding to the nearest neighbor of higher density (Rodriguez & Laio 2014). We adopted two modifications to reduce arbitrary choices. First, the identification of cluster centers was fully automated: all species with higher neighbor density than their own neighbors and at a distance of at least d_0 from species with higher density were considered as cluster centers. Second, if two clusters were not separated by a "low density valley", i.e. a region of radius d_0 where densities were lower than those of the cluster centers, they were merged.

The whole clustering process thus required only a single free parameter, the threshold d_0, fixed by a rule of thumb by which the minimum distance to the nearest neighbor defining isolation, i.e. species uniqueness in trait space, is the average number of neighbors around each object corresponding to 1 or 2% of the total number of species in the dataset (Rodriguez & Laio 2014). This procedure has the advantage of not fixing a d_0 value a priori for all datasets but instead to define a d_0 value for each dataset only depending on species number. Unique species can thus be considered as relative isolates in the trait space. We chose 1% as a conservative rule to not cluster species being too different in traits so keeping d_0 small. We provide an R implementation of this algorithm along with the code to reproduce all the analyses of this paper (R Core Team, 2021; see section Data and Code availability).

Influence of trait dataset characteristics

To test whether the characteristics of species, ecosystems and traits can influence the dimensionality, robustness, and structure of species trait space we performed General Linear Models (GLMs) with a Gaussian distribution for all response variables, i.e., the elbow-based dimensionality, the threshold-based dimensionality, the robustness to 50% trait removal, the log-transformed number of species clusters, the percentage of species packed in the first cluster and the percentage of unique species (distributions are shown in Supplementary Figure 3). As explanatory factors, we used the type of species life form (plant, invertebrate and vertebrate) and the type of ecosystem (aquatic and terrestrial) to test the potential effects of organismal complexity. We also used the log-transformed number of species and number of traits as the dimensions of the initial species trait matrix. Trait characteristics were then used as potential drivers like the percentage of missing values, the percentage of quantitative traits and the mean pairwise correlation between traits, expressed as the rank-based Kendall index able to mix continuous and categorical traits. Pairwise correlations between quantitative trait dataset characteristics are rather low ($-0.19<r<0.45$) and mainly non-significant (Supplementary Figure 4).

We then used partial regression plots to highlight the effect of each factor while controlling for the others (set at their mean). Statistical analyses were carried out using the function *glm* from the *stats* R package while partial plots were drawn using the function *visreg* from the *visreg* R package.

In addition to the analyses performed on empirical datasets, we also built three simulated datasets to test the effect of species and trait number on the dimensionality of species trait space without changing the type of traits as a controlled experiment. Continuous traits for 1,000 species were generated following a uniform distribution (0-1) with no missing value. In the first dataset we simulated 10 uncorrelated traits, in the second 10 correlated traits ($r = 0.5$) and in the third 20 uncorrelated traits. We then estimated the trait space dimensionality for each level of species number and each dataset using the AUC threshold of 0.7.

Results

Trait space dimensionality

Over the 30 datasets we obtained an optimal reduced dimensionality ranging between 2 and 8 axes (Median=4) using the elbow method and between 2 to 17 axes (Median=6) using the AUC threshold of 0.7 when attained. For all datasets, we could reach the AUC threshold of 0.7 with less than 20 dimensions or PCoA axes, except for plants of the French Alps for which AUC remained low (<0.6) even with many axes (Figure 2). For the remaining 29 datasets, the correlation between the elbow-based and threshold-based dimensionality was positive but weak ($r = 0.3$) and non-significant (*p-value*=0.10) highlighting their complementarity (Supplementary Figure 5). With a more demanding threshold of AUC=0.8 (high quality trait space), up to 24 datasets could reach this value with a maximum of 20 dimensions (Figure 2).

Two first GLMs, including all explanatory factors but only 29 datasets out of 30 (bacteria were excluded since they are the only representative of a kingdom), showed that the type of life form (plant, invertebrate and vertebrate) and the type of ecosystem (aquatic and terrestrial) did not significantly explain the elbow-based and the threshold-based dimensionality (Supplementary Table 1). The partial regression plots illustrate these weak influences while controlling for the other factors (Figure 3). We thus retained only quantitative variables related to the characteristics of the species trait datasets in the following analyses.

`

The elbow-based dimensionality was weakly explained by the five quantitative characteristics of the datasets (R^2=0.15) but the correlation between traits had by far the main effect, albeit non-significant (*p-value*=0.09) (Supplementary Table 2), with a lower optimal number of axes when the correlation between traits increased (Figure 4). The threshold-based dimensionality was well explained by characteristics of the datasets (R^2=0.61) with the log-number of traits and the correlation between traits having the strongest and only significant effects (Supplementary Table 2). The partial regression plots showed that the threshold-based dimensionality strongly increased with the log-number of traits while it decreased with the correlation between traits (Figure 4). As a complementary analysis, our simulated trait datasets confirmed the main influence of the number and the correlation of traits on species trait space dimensionality while the number of species had only an effect for less than 100 species and no effect above 200 species (Supplementary Figure 6).

The number of axes necessary to explain 50% of trait variation was a weak predictor of the elbow-based dimensionality (R^2=0.18) but was a strong predictor of the threshold-based dimensionality (R^2=0.82), albeit underestimated (Supplementary Figure 7).

Robustness to trait omission

The robustness to trait omission was generally low over the 30 datasets with a mean AUC loss of 0.54 (SD=0.12) when 50% of the traits were deleted. In these cases, most low-dimensional trait spaces were poor representations of the initial distances between species. Yet, this robustness was highly heterogeneous among datasets ranging from 0.33 to 0.85 of AUC loss (Figure 5). To stay above the AUC threshold of 0.7, trait omission should not exceed 20% on average when we ignored the five datasets for which even removing 10% of traits induced an AUC loss of more than 0.3 (i.e. AUC < 0.7).

Like for the dimensionality, the robustness to trait omission was not significantly influenced by either the type of species life form or the type of ecosystem (Figure 3, Supplementary Table 1) so these factors were ignored in the following analyses focused on quantitative factors. The robustness to trait omission was strongly dependent on the dataset characteristics (R^2=0.84) with the log-number of traits, the percentage of missing values and the correlation between traits having the strongest and only significant effects (Supplementary Table 2). The partial regression plots revealed quite logically that the robustness to trait omission (opposite to AUC loss) increased with the number of traits but also with the correlation between traits (Figure 4). In contrast, robustness was negatively related to the percentage of missing values, which again makes sense. With many missing values, the trait space is likely to be unstable under trait omission so dimensionality reduction may distort the representation of the initial distances between species.

Species clustering in trait space

Over the 30 datasets, the number of species clusters, delineated by the “fast search and find of density peaks” algorithm, varied between 4 and 434 and was moderately explained by the dataset characteristics (R^2=0.57). The number of clusters was not significantly influenced by either the type of species life form or the type of ecosystem (Figure 6, Supplementary Table 3) so these factors were ignored in the following analyses. The main and only significant drivers were the log-number of species and percentage of missing values (Supplementary Table 4). The number of clusters logically decreased with the percentage of missing values since less trait combinations can be realized but increased with the number of species (Figure 7, Supplementary Figure 8). Yet, the number of clusters increased as a saturating power-law with the number of species owing to a slope much lower than 1 (0.41) in the log-log relationship when we controlled for other effects (Figure 8a).

The proportion of species belonging to the first or dominant cluster was not significantly driven by either the type of species life form or ecosystem (Figure 6, Supplementary Table 3) so these factors were ignored in the following analyses. This species packing into the dominant cluster was mainly driven by the log-number of species with a predictive power of R^2=0.58 while all the other dataset characteristics had non-significant influences (Supplementary Table 4). The slope of the relationship between the proportion of species clustered within the first group and the log-number of species was positive (Figure 7), highlighting that species tended to pack in the richest trait cluster when species richness increased, regardless of the other dataset characteristics. Yet, the log-log relationship between the total species richness and the richness of the first cluster revealed a power law with a slope higher than 1 (1.38) when we controlled for other effects (Figure 8b), suggesting that species packing disproportionately increased with species richness.

Unique species in trait space

The number of unique species, i.e. species that did not belong to any cluster so isolated in the trait space, varied between 27 and 1750 among datasets with a percentage ranging from 2% to 74% (Median=42%). These unique species were widespread in trait space and not just located on the edges, suggesting openings scattered throughout species trait spaces (Figure 9). Yet, well-known unique species appeared clearly far on the edge such as the whale shark (*Rhincodon typus*) which is the largest shark (20 meters long and body mass of 34 tonnes) while being a planktivore, so an ecological outlier among Chondrichthyes.

The proportion of unique species was not significantly influenced by either the type of species life form or ecosystem (Figure 6, Supplementary Table 3) so these categorical factors were ignored in the following analyses only based on quantitative factors. The proportion of unique species was strongly explained by dataset characteristics (R^2=0.82) with the log-number of species and, to a less extent, the percentage of missing values, being the main drivers (Supplementary Table 4).

The partial regression plots revealed that the proportion of unique species had a marked negative relationship with the log-number of species while controlling for other effects (Figure 7), suggesting that species-rich assemblages left less space for ecological uniqueness or that species tended to disproportionately pack into the richest cluster when diversity increased (Figure 8b). This saturating relationship was highlighted by the partial plot linking the total number of species and the number of unique species with a power log-log slope of 0.47 (Figure 8c). The proportion of unique species also decreased with the proportion of missing values since it mechanically reduced the diversity of trait combinations and increased species similarity (Figure 7).

Discussion

The necessary trade-off between trait space quality and operationality

Trait-based approaches have a long tradition in life science since the development of the two-strategy life-history framework from 'fast' (r) to 'slow' (K) organisms (MacArthur & Wilson 1967; Pianka 1972). This oversimplified view was later extended to triangular continuums of plant life-history strategies with the well-known competitive ability - physiological tolerance to stress - adaptation to disturbance (C-S-R) schema introduced by Grime (1977) and the Leaf-Height-Seed (LHS) framework by Westoby (1998). Such meaningful simplifications of trait variability among species have revolutionized functional ecology and inspired similar successful approaches for insects (Greenslade 1983), freshwater fishes (Winemiller & Rose 1992), corals (Darling *et al.* 2012) and microbes (Malik *et al.* 2020). In the case of well-established or experimentally tested causal relationships between traits and environments or functions, the dimensionality issue is of marginal importance when building species spaces with few relevant traits delineating clearly defined ecological strategies. By contrast, when such knowledge is lacking, so when many traits are available with low evidence of particular causal relevance, when big data analyses are performed with many missing values, or when species strategies cannot be summarized by a limited set of traits, ecologists face the challenge of trait space hyper-dimensionality (Blonder *et al.* 2014).

Dimensionality reduction can then be a necessary step since some widely used functional diversity indices (e.g. functional richness) are based on the volume of trait space (convex hull volume) occupied by species of a given ecosystem (Villeger *et al.* 2008; Laliberte & Legendre 2010; Trindade-Santos *et al.* 2020) that can be hardly calculated beyond 6 dimensions, even less (4-5) if null models are required or when pair-wise site measures like β-diversity have to be estimated (Villeger *et al.* 2011; Loiseau *et al.* 2017; Pimiento *et al.* 2020b; Su *et al.* 2021). Since most common functional diversity indices are sensitive to the degree of correlation among traits (Zhu *et al.* 2017), we also suggest to compute these indices from a reduced number of independent PCoA axes to improve the capacity to distinguish between communities along gradients of stress (Trindade-Santos *et al.* 2020).

Beyond practical reasons, this dimensionality value also informs about the extent to which species traits can be reduced to a limited number of ecologically meaningful axes (Díaz *et al.* 2016; Pigot *et al.* 2020). This quest for ecological syndromes or strategies is not new (Westoby 1998; Reich *et al.* 2003) and some previous studies have investigated the intrinsic dimensionality of species traits using various linear and non-linear methods (Westoby 1998; Laughlin 2014; Maire *et al.* 2015). Here, we proposed two complementary ways to estimate linear dimensionality and we applied them to 30 datasets to ultimately identify their main drivers, if any.

Using the parsimonious elbow-based AUC method, we found a median dimensionality of 4 axes which is a rather low value given that we only considered datasets with at least 10 traits in our study (Table 1). Interestingly, for most datasets (25 out of 30) the elbow-based dimensionality is lower than 6 axes (2-5) (Figure 2) suggesting that the calculation of most volume-based functional diversity indices can be performed even with null models. Using the AUC-threshold criteria of 0.7, the dimensionality is higher (median of 6 axes) and generally out of the operational range for calculations of hypervolume-based metrics like functional richness (Villeger *et al.* 2008) or functional β-diversity (Loiseau *et al.* 2017). It reinforces the idea that the diversity of organism forms and functions has a larger dimensionality than previously thought (Pigot *et al.* 2016b; Messier *et al.* 2017) whatever the kingdom and ecosystem. Only poor assemblages (<30 species) can be accurately described with low-dimensionality (<4 axes) as shown in our simulations (Supplementary Figure 6).

This can be partly due to the coexistence of different syndromes related to different sets of traits, corresponding to different ecological strategies, under a given environment (Reich *et al.* 2003; Sosiak & Barden 2021). For instance, landscape filters can shape trait community composition with species sharing some traits (trait syndromes) responding in a similar way under the same environmental conditions (e.g. agricultural intensification) (Gámez-Virués *et al.* 2015). When using large species datasets mixing various environments and many traits like in most our cases (Table 1), the potential multiplication of trait syndromes could explain the relatively high dimensionality in the trait space we have observed, particularly for the plants in the French Alps or stream macroinvertebrates (Figure 2). We may expect lower dimensionality in species trait space built from local communities under severe filters owing to the predominance of a few but highly constrained trait syndromes. We may also expect lower dimensionality when using effect vs. response traits in a more coherent and systematic manner with a clear defined goal (Luck *et al.* 2012).

The most surprising result is the weak positive correlation between the elbow-based and threshold-based dimensionality values showing that a low elbow-based AUC value does not imply passing the 0.7 AUC threshold and vice-versa (Figure 2). This is because the elbow-based method imposes a compromise between the quantity of axes and the quality of the trait space to avoid selecting more poorly informative axes (over-dimensionality) while the threshold-based method only considers quality whatever the quantity of axes. Given this constraint, the elbow-based method provides lower dimensionality values (2-8 axes against 2-17 axes for the threshold method; Supplementary Figure 5) which are also less influenced by dataset characteristics. As a practical guide, we suggest to use the elbow-based method as a first estimate of dimensionality on a given trait dataset and then to increase the number of dimensions to be considered until passing the 0.7 threshold if necessary. With this rule of thumb, we should end-up with an optimal dimensionality comprising between 3 and 6 axes for most datasets, as a trade-off between operationality and quality. Obviously, the operational constraint depends on species number, diversity indices being used and power facilities.

In case a value of AUC=0.5 cannot be reached with a reasonable number of dimensions (<10 axes) like on the French Alps plants (Figure 2) we suggest either to carefully select the most relevant traits given the question being addressed (Thuiller *et al.* 2014) or to avoid indices based on trait space reduction (like functional richness) but instead to use distance-based indices (Rao) only (Laliberte & Legendre 2010; Mouillot *et al.* 2013; Chao *et al.* 2019). For representation purposes, which are classically drawn in 2 or 3 dimensions with PCoA axes 1 to 4 (Stubbs & Benton 2016; Bruelheide *et al.* 2018; Loiseau *et al.* 2020; Pimiento *et al.* 2020b), we suggest to provide the corresponding AUC value as a key information on trait space quality along with the percentage of trait variation explained by axes. Since dimensionality is weakly influenced by dataset characteristics, except trait correlations that decrease dimensionality for both elbow-based and threshold-based criteria (Figure 4), we suggest to pay particular attention to unnecessary or meaningless traits that are strongly independent from the others and would inflate dimensionality potentially biasing biodiversity metrics. Conversely, considering redundant or correlated traits, even if meaningless, has no expected impact on dimensionality so can be very neutral in the building of species trait space and the computation of indices. Yet, using surrogate traits or traits with a coarse resolution to describe a given dimension of ecological strategy can substantially affect the results (Loranger *et al.* 2016; Kohli & Jarzyna 2021).

The low but predictable robustness to trait omissions or choices

Choosing a set of traits always means ignoring some, while important traits can be missed because they are unavailable or unknown. Often traits are ignored for non-biological reasons such as the difficulty of measuring them or the lack of standardization in the research community. The consequences of this sub-selection have been poorly investigated, despite its potential to modify the perceived dissimilarity between species (Carscadden *et al.* 2017) and profoundly affect the estimates of functional diversity (Zhu *et al.* 2017). Here, we randomly reduced our trait datasets to assess the impact of trait omission on AUC loss between the initial distance matrix (all traits) and that based on 90% to 20% of the traits only (Figure 5). When only 10% of traits are removed, AUC is still higher than 0.7 on average across simulations in 21 datasets out of 30, suggesting overall high robustness to low rate of trait omission except for some taxa like palm trees, sharks, thermal fauna and corals which belong to different kingdoms and ecosystems. At 50% of trait removal, AUC severely drops below the 0.7 threshold for all datasets except fishes of the Jakarta Bay (Figure 5).

This overall low but highly variable robustness of species distances to trait omission is very well explained by datasets characteristics (Figure 4). Unsurprisingly, AUC loss at 50% omission rate is negatively related to the number of traits, so that trait-poor datasets (corals, sharks or freshwater fishes) are more sensitive to the removal of traits than their trait-rich counterparts (macro-invertebrates or bacteria). Our statistical model also shows an expected negative relationship between AUC loss and trait correlation, so with more redundant traits the distances between species pairs in a low dimensional space are more strongly preserved. This might explain why dimensionality reduction has been successful for some research fields in functional ecology (e.g. leaf traits and the leaf economic spectrum (Wright *et al.* 2004; Díaz *et al.* 2016)), while other studies such as those spanning many organs of plants have failed to find meaningful reduction in trait dimensionality (Carscadden *et al.* 2017; Messier *et al.* 2017). We also point out that the number of missing values strongly impacts robustness to trait omission so including traits with many missing values (>10%) can be a counterproductive effort, especially with Gower-like metrics which only consider traits with no missing values to assess the distance between two species. We also show a high variability in robustness for a given level of trait omission (Figure 5) suggesting that robustness to trait omission depends on traits being removed, some being more critical than others, independently of their ecological relevance. This reinforces the advice to carefully select traits prior to analyses and pay a particular attention to those being uncorrelated to the others given their disproportionate importance in the structuring of species trait spaces and subsequent analyses.

Taken together these results point out that the robustness of species space to trait omissions or choices is on average lower than previously thought (Douma *et al.* 2012) and that dataset characteristics, not the species life form or ecosystem type, explain this robustness, notably the presence of too many missing values. As a precautionary principle, we suggest to perform sensitivity analyses where traits are removed one by one or until a certain percentage of removal to assess the robustness of the results (Mouillot *et al.* 2014; Pollock *et al.* 2017; McLean *et al.* 2018; Cooke *et al.* 2019a; Loiseau *et al.* 2020). Trait-gap filling through automatic imputation might also be an interesting perspective (Penone *et al.* 2014; Schrodt *et al.* 2015; Goberna & Verdú 2016; Johnson *et al.* 2020). However, given the way most of these approaches work, this is likely that trait imputations will follow the main trends and the main syndromes and will unlikely generate unique species artificially hidden in the space.

Species packing in trait space disproportionally increases with species richness

The species packing in trait space, or so-called over-redundancy (Mouillot *et al.* 2014), provides functional insurance and resilience to ecosystems under disturbances (McLean *et al.* 2019). This packing can be easily assessed with categorical traits since each unique combination of traits, also called functional entity, is a cluster so the clusters with a high number of species, or higher than expected under a null model, are considered as over-packed or over-redundant while those with few species are vulnerable to biodiversity loss (Mouillot *et al.* 2014). With continuous traits or a large mix of traits as in our study, the clustering of species remains an arbitrary decision depending on the methods and thresholds used. We chose a clustering method with the lowest number of arbitrary decisions as possible independently of the shape and structure of species distribution in trait space (Rodriguez & Laio 2014). Surprisingly, this method, despite its attractiveness in other fields (medical and social sciences) and its parsimony (one parameter), has never been applied in ecology and evolution so far.

Using a “fast search and find of density peaks” algorithm (Rodriguez & Laio 2014), we show that the number of clusters increases with the number of species when we control for the other factors (Figure 7) but with a strongly saturating relationship (Figure 8a) suggesting that species tend to over-pack into some clusters instead of creating new clusters in species-rich assemblages as shown for reef fishes (Mouillot *et al.* 2014) or passerine birds (Pigot *et al.* 2016b). With a slope of 0.41 on the log-log scale it means that when species richness doubles, the number of clusters only increases by 30%. As a corollary, the richness of the dominant cluster increases with total species richness on a log-log scale with a slope higher than 1 (Figure 8b) suggesting that additional species disproportionally pack into the most speciose cluster. More precisely two times more species in a given assemblage induces the packing of 2.6 times more species in the dominant cluster. So, biodiversity only reinforces the redundancy of the most common traits instead of providing the level of insurance we should expect from species richness only under a random or proportional distribution of species among clusters (Mazel *et al.* 2014; Mouillot *et al.* 2014). This remarkable trend is observed for all taxa and ecosystem types.

The saturating scaling of uniqueness with species richness

The identification of ecological disparity, gaps, distinctiveness or uniqueness in trait spaces is a long-standing issue in ecology and evolution (Foote 1990; Winemiller 1991; Ricklefs 2005; Bapst *et al.* 2012; Violle *et al.* 2017; Gauzere *et al.* 2020). It contributes, for instance, to estimate the level of functional insurance and vulnerability to species extinction (Mouillot *et al.* 2014) but also to better understand the influence of trait rarity on ecosystem functioning (Maire *et al.* 2018), to set conservation priorities targeting unique species (Loiseau *et al.* 2020), and to illuminate the capacity for innovation in clades (Cornwell *et al.* 2014; Deline *et* *al.* 2018; Reeves *et al.* 2020). Yet, there is no consensus on the way to determine which species are isolated enough in trait spaces to be considered as unique species. Among the myriad of clustering algorithms (Xu & Tian 2015), the method based on fast search and find of density peaks was able to extract unique species in a very intuitive, standard, biodiversity-independent and distribution-free way. We show that the proportion of unique species decreases with species richness (Figure 7) while the number of unique species saturates rapidly with species richness (Figure 8c) suggesting that ecological novelty does not scale proportionally with taxonomic diversity but at a much lower rate whatever the kingdom or ecosystem. With a slope of 0.47 on the log-log scale it means that when species richness doubles, the number of unique species increases by 38%. This result resonates with the saturating link between ecological disparity and species richness across geological periods (Bapst *et al.* 2012) contrary to predictions from theory on adaptive radiations and ecological speciation (Rundell & Price 2009). More precisely, some entire lineages remained ecologically conservative throughout the Mesozoic without exploring vacant portions of trait space and then trait bursts occurred owing to changing abiotic conditions during the Late Jurassic (Reeves *et al.* 2020). Both adaptive radiations due to species interactions and innovative solutions to face new environments are certainly at play to explain the invariant saturating scaling of ecological uniqueness with species richness.

Conclusions

Four take-home messages can be extracted from this analysis. First of all, when no prior selection of traits can be carried out, the minimum dimensionality of trait space is rather large with around 3-6 dimensions. The success of identifying axes of variation, especially when trait correlations are strong, suggests that the research program of finding major trade-off axes grounded in ecological principles shows more promise than the arbitrary selection and removal of traits. Second, most trait spaces are highly sensitive to trait omission, which thus requires careful thinking about which traits might be overlooked, missed and targeted into the future. Third, there are plenty of unique species and the success of the clustering approach suggests that we need to pay more attention to how species pack relative to each other in trait space and not only focus on dimensionality reduction of trait spaces. Fourth, the complexity of multicellular organisms from plants to vertebrates or from aquatic to terrestrial species has little influence on the dimensionality, robustness and structure of trait space. Instead our synthesis suggests that the rate of key functional innovations and the subsequent complexity of trait space are consistent across multicellular clades with multicellularity evolution in plants sharing many features with that leading to animals. Yet, these results are based on only 30 datasets and may lack statistical power to detect some effects. Moreover, these results are only valid for the range of dataset characteristics that we used in our analyses so more than 40 species and ten traits. We obtained different patterns for species-poor assemblages in our simulations but we are confident that our empirical assessment may embrace most species richness conditions encountered in temperate or tropical assemblages for most taxa when building regional or global species trait space.

Acknowledgements

This research is supported by the Fondation pour la Recherche sur la Biodiversité (FRB) and Electricité de France (EDF) in the context of the CESAB project 'Causes and consequences of functional rarity from local to global scales' (FREE)

Supplementary materials

Supplementary Table 1: Statistics of GLM.
Supplementary Table 2: Statistics of GLM.
Supplementary Table 3: Statistics of GLM.
Supplementary Table 4: Statistics of GLM.

Supplementary Figure 1: Number of species and traits in the CESTES database.
Supplementary Figure 2: Theoretical illustration of the Elbow method.
Supplementary Figure 3: Distribution of the six metrics characterizing species trait spaces.
Supplementary Figure 4: Pairwise Pearson correlations between all dataset characteristics.
Supplementary Figure 5: Relation between the dimensionality found with the elbow vs. threshold method.
Supplementary Figure 6: Simulated relationships between dimensionality and species-trait numbers.
Supplementary Figure 7: Relation between dimensionality and the number of PCoA axes explaining 50% of trait variation.
Supplementary Figure 8: Relationships between the log number of species and the log number of clusters, the log number of species in the most dominant clusters and the log number of unique species.
Supplementary Figure 9: Relation between the Area Under the Curve (AUC) metric and the mean absolute deviation (MAD).

References

Bapst, D.W., Bullock, P.C., Melchin, M.J., Sheets, H.D. & Mitchell, C.E. (2012). Graptoloid diversity and disparity became decoupled during the Ordovician mass extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 3428-3433.

Barros, C., Guéguen, M., Douzet, R., Carboni, M., Boulangeat, I., Zimmermann, N.E. *et al.* (2017). Extreme climate events counteract the effects of climate and land-use changes in Alpine tree lines. *Journal of Applied Ecology*, 54, 39-50.

Bernhardt-Römermann, M., Römermann, C., Nuske, R., Parth, A., Klotz, S., Schmidt, W. *et al.* (2008). On the identification of the most suitable traits for plant functional trait analyses. *Oikos*, 117, 1533-1541.

Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography*, 23, 595-609.

Bruelheide, H., Dengler, J., Purschke, O., Lenoir, J., Jiménez-Alfaro, B., Hennekens, S.M. *et al.* (2018). Global trait–environment relationships of plant communities. *Nature Ecology & Evolution*, 2, 1906-1917.

Cadotte, M.W. (2017). Functional traits explain ecosystem function through opposing mechanisms. *Ecology Letters*, 20, 989-996.

Carscadden, K.A., Cadotte, M.W. & Gilbert, B. (2017). Trait dimensionality and population choice alter estimates of phenotypic dissimilarity. *Ecology and Evolution*, 7, 2273-2285.

Chao, A., Chiu, C.-H., Villéger, S., Sun, I.F., Thorn, S., Lin, Y.-C. *et al.* (2019). An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. *Ecological Monographs*, 89, e01343.

Condon, A., Ding, J. & Shah, S. (2016). densityCut: an efficient and versatile topological approach for automatic clustering of biological data. *Bioinformatics*, 32, 2567-2576.

Cooke, R.S.C., Bates, A.E. & Eigenbrod, F. (2019a). Global trade-offs of functional redundancy and functional dispersion for birds and mammals. *Global Ecology and Biogeography*, 28, 484-495.

Cooke, R.S.C., Eigenbrod, F. & Bates, A.E. (2019b). Projected losses of global mammal and bird ecological strategies. *Nature Communications*, 10, 2279.

Cooney, C.R., Bright, J.A., Capp, E.J.R., Chira, A.M., Hughes, E.C., Moody, C.J.A. *et al.* (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature*, 542, 344-347.

Cornwell, W.K., Westoby, M., Falster, D.S., FitzJohn, R.G., O'Meara, B.C., Pennell, M.W. *et al.* (2014). Functional distinctiveness of major plant lineages. *Journal of Ecology*, 102, 345-356.

Cox, D.T.C., Gardner, A.S. & Gaston, K.J. (2021). Diel niche variation in mammals associated with expanded trait space. *Nature Communications*, 12, 1753.

Craven, D., Eisenhauer, N., Pearse, W.D., Hautier, Y., Isbell, F., Roscher, C. *et al.* (2018). Multiple facets of biodiversity drive the diversity–stability relationship. *Nature Ecology & Evolution*, 2, 1579-1587.

Darling, E.S., Alvarez-Filip, L., Oliver, T.A., McClanahan, T.R. & Côté, I.M. (2012). Evaluating life-history strategies of reef corals from species traits. *Ecology Letters*, 15, 1378-1386.

Dehling, D.M., Jordano, P., Schaefer, H.M., Böhning-Gaese, K. & Schleuning, M. (2016). Morphology predicts species' functional roles and their degree of specialization in plant–frugivore interactions. *Proceedings of the Royal Society of London B: Biological Sciences*, 283.

Deline, B., Greenwood, J.M., Clark, J.W., Puttick, M.N., Peterson, K.J. & Donoghue, P.C.J. (2018). Evolution of metazoan morphological disparity. *Proceedings of the National Academy of Sciences*, 115, E8909.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S. *et al.* (2016). The global spectrum of plant form and function. *Nature*, 529, 167-171.

Douma, J.C., Shipley, B., Witte, J.P.M., Aerts, R. & van Bodegom, P.M. (2012). Disturbance and resource availability act differently on the same suite of plant traits: revisiting assembly hypotheses. *Ecology*, 93, 825-835.

Duffy, J.E., Macdonald, K.S., Rhode, J.M. & Parker, J.D. (2001). Grazer diversity, functional redundancy, and productivity in seagrass beds: An experimental test. *Ecology*, 82, 2417-2434.

Fisher, R.M., Shik, J.Z. & Boomsma, J.J. (2020). The evolution of multicellular complexity: the role of relatedness and environmental constraints. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20192963.

Fonseca, C.R. & Ganade, G. (2001). Species functional redundancy, random extinctions and the stability of ecosystems. *Journal of Ecology*, 89, 118-125.

Foote, M. (1990). Nearest-Neighbor Analysis of Trilobite Morphospace. *Systematic Biology*, 39, 371-382.

Gagic, V., Bartomeus, I., Jonsson, T., Taylor, A., Winqvist, C., Fischer, C. *et al.* (2015). Functional identity and diversity of animals predict ecosystem functioning better than species-based indices. *Proceedings of the Royal Society B: Biological Sciences*, 282, 20142620.

Gámez-Virués, S., Perović, D.J., Gossner, M.M., Börschig, C., Blüthgen, N., de Jong, H. *et al.* (2015). Landscape simplification filters species traits and drives biotic homogenization. *Nature Communications*, 6, 8568.

Gauzere, P., Morin, X., Violle, C., Caspeta, I., Ray, C. & Blonder, B. (2020). Vacant yet invasible niches in forest community assembly. *Functional Ecology*, 34, 1945-1955.

Goberna, M. & Verdú, M. (2016). Predicting microbial traits with phylogenies. *The ISME Journal*, 10, 959-967.

Gower, J.C. & Legendre, P. (1986). Metric and Euclidean properties of dissimilarities coefficients. *Journal of Classification*, 3, 5-48.

Greenslade, P.J.M. (1983). Adversity Selection and the Habitat Templet. *The American Naturalist*, 122, 352-365.

Grime, J.P. (1977). Evidence for the Existence of Three Primary Strategies in Plants and Its Relevance to Ecological and Evolutionary Theory. *The American Naturalist*, 111, 1169-1194.

Gross, N., Bagousse-Pinguet, Y.L., Liancourt, P., Berdugo, M., Gotelli, N.J. & Maestre, F.T. (2017). Functional trait diversity maximizes ecosystem multifunctionality. *Nature Ecology & Evolution*, 1, 0132.

Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M.C., Diemer, M., Dimitrakopoulos, P.G. *et al.* (1999). Plant diversity and productivity experiments in European grasslands. *Science*, 286, 1123-1127.

Jain, A.K. & Dubes, R.C. (1988). *Algorithms for clustering data*. Prentice Hall, Englewood Cliffs, New Jersey.

Jarzyna, M.A., Quintero, I. & Jetz, W. (2020). Global functional and phylogenetic structure of avian assemblages across elevation and latitude. *Ecology Letters*, n/a.

Jeliazkov, A., Mijatovic, D., Chantepie, S., Andrew, N., Arlettaz, R., Barbaro, L. *et al.* (2020). A global database for metacommunity ecology, integrating species, traits, environment and space (vol 7, 6, 2020). *Scientific Data*, 7.

Johnson, T.F., Isaac, N.J.B., Paviolo, A. & González-Suárez, M. (2020). Handling missing values in trait data. *Global Ecology and Biogeography*, n/a.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L. *et al.* (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90, 2648-2648.

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P. *et al.* (2020). TRY plant trait database – enhanced coverage and open access. *Global Change Biology*, 26, 119-188.

Knoll, A.H. (2011). The Multiple Origins of Complex Multicellularity. *Annual Review of Earth and Planetary Sciences*, 39, 217-239.

Kohli, B.A. & Jarzyna, M.A. (2021). Pitfalls of ignoring trait resolution when drawing conclusions about ecological processes. *Global Ecology and Biogeography*, n/a.

Kraemer, G., Reichstein, M. & Mahecha, M.D. (2018). dimRed and coRanking-Unifying Dimensionality Reduction in R. *R Journal*, 10, 342-358.

Laliberte, E. & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91, 299-305.

Laughlin, D.C. (2014). The intrinsic dimensionality of plant traits and its relevance to community assembly. *Journal of Ecology*, 102, 186-193.

Le Bagousse-Pinguet, Y., Gross, N., Saiz, H., Maestre, F.T., Ruiz, S., Dacal, M. *et al.* (2021). Functional rarity and evenness are key facets of biodiversity to boost multifunctionality. *Proceedings of the National Academy of Sciences*, 118, e2019355118.

Lee, J.A. & Verleysen, M. (2009). Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing*, 72, 1431-1443.

Legendre, P. & Legendre, L. (1998). *Numerical Ecology*. Second edn. Elsevier, Amsterdam.

Loiseau, N., Legras, G., Gaertner, J.C., Verley, P., Chabanet, P. & Merigot, B. (2017). Performance of partitioning functional beta-diversity indices: Influence of functional representation and partitioning methods. *Global Ecology and Biogeography*, 26, 753-762.

Loiseau, N., Mouquet, N., Casajus, N., Grenié, M., Guéguen, M., Maitner, B. *et al.* (2020). Global distribution and conservation status of ecologically rare mammal and bird species. *Nature Communications*, 11, 5071.

Loranger, J., Blonder, B., Garnier, É., Shipley, B., Vile, D. & Violle, C. (2016). Occupancy and overlap in trait space along a successional gradient in Mediterranean old fields. *Am. J. Bot.*, 103, 1050-1060.

Luck, G.W., Lavorel, S., McIntyre, S. & Lumb, K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *Journal of Animal Ecology*, 81, 1065-1076.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princeton University Press.

Maire, E., Grenouillet, G., Brosse, S. & Villéger, S. (2015). How many dimensions are needed to accurately assess functional diversity? A pragmatic approach for assessing the quality of functional spaces. *Global Ecology and Biogeography*, 24, 728-740.

Maire, E., Villéger, S., Graham, N.A.J., Hoey, A.S., Cinner, J., Ferse, S.C.A. *et al.* (2018). Community-wide scan identifies fish species associated with coral reef services across the Indo-Pacific. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20181167.

Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K. & Allison, S.D. (2020). Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *The ISME Journal*, 14, 1-9.

Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J. *et al.* (2014). Multifaceted diversity-area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Global Ecology and Biogeography*, 23, 836-847.

McLean, M., Auber, A., Graham, N.A.J., Houk, P., Villéger, S., Violle, C. *et al.* (2019). Trait structure and redundancy determine sensitivity to disturbance in marine fish communities. *Global Change Biology*, 25, 3424-3437.

McLean, M., Mouillot, D., Lindegren, M., Engelhard, G., Villéger, S., Marchal, P. *et al.* (2018). A Climate-Driven Functional Inversion of Connected Marine Ecosystems. *Current Biology*, 28, 3654-3660.e3653.

McLean, M., Stuart-Smith, R.D., Villéger, S., Auber, A., Edgar, G.J., MacNeil, M.A. *et al.* (2021). Trait similarity in reef fish faunas across the world's oceans. *Proceedings of the National Academy of Sciences*, 118, e2012318118.

Mérigot, B., Durbec, J.-P. & Gaertner, J.-C. (2010). On goodness-of-fit measure for dendrogram-based analyses. *Ecology*, 91, 1850-1859.

Messier, J., Lechowicz, M.J., McGill, B.J., Violle, C. & Enquist, B.J. (2017). Interspecific integration of trait dimensions at local scales: the plant phenotype as an integrated network. *Journal of Ecology*, 105, 1775-1790.

Mouillot, D., Graham, N.A.J., Villeger, S., Mason, N.W.H. & Bellwood, D.R. (2013). A functional approach reveals community responses to disturbances. *Trends in ecology & evolution*, 28, 167-177.

Mouillot, D., Villeger, S., Parravicini, V., Kulbicki, M., Ernesto Arias-Gonzalez, J., Bender, M. *et al.* (2014). Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 13757-13762.

Nguyen, L.H. & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLOS Computational Biology*, 15, e1006907.

O'Gorman, E.J., Yearsley, J.M., Crowe, T.P., Emmerson, M.C., Jacob, U. & Petchey, O.L. (2011). Loss of functionally unique species may gradually undermine ecosystems. *Proceedings of the Royal Society B: Biological Sciences*, 278, 1886-1893.

Pavoine, S., Vallet, J., Dufour, A.B., Gachet, S. & Daniel, H. (2009). On the challenge of treating various types of variables: application for improving the measurement of functional diversity. *Oikos*, 118, 391-402.

Penone, C., Davidson, A.D., Shoemaker, K.T., Di Marco, M., Rondinini, C., Brooks, T.M. *et al.* (2014). Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods in Ecology and Evolution*, 5, 961-970.

Perez, T.M., Valverde-Barrantes, O., Bravo, C., Taylor, T.C., Fadrique, B., Hogan, J.A. *et al.* (2019). Botanic gardens are an untapped resource for studying the functional ecology of tropical plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374, 20170390.

Phillips, A.G., Töpfer, T., Rahbek, C., Böhning-Gaese, K. & Fritz, S.A. (2018). Effects of phylogeny and geography on ecomorphological traits in passerine bird clades. *Journal of Biogeography*, 45, 2337-2347.

Pianka, E.R. (1972). r and K Selection or b and d Selection? *The American Naturalist*, 106, 581-588.

Pigot, A.L., Bregman, T., Sheard, C., Daly, B., Etienne, R.S. & Tobias, J.A. (2016a). Quantifying species contributions to ecosystem processes: a global assessment of functional trait and phylogenetic metrics across avian seed-dispersal networks. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20161597.

Pigot, A.L., Sheard, C., Miller, E.T., Bregman, T.P., Freeman, B.G., Roll, U. *et al.* (2020). Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology & Evolution*, 4, 230-239.

Pigot, A.L., Trisos, C.H. & Tobias, J.A. (2016b). Functional traits reveal the expansion and packing of ecological niche space underlying an elevational diversity gradient in passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, 283, 20152013.

Pimiento, C., Bacon, C.D., Silvestro, D., Hendy, A., Jaramillo, C., Zizka, A. *et al.* (2020a). Selective extinction against redundant species buffers functional diversity. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20201162.

Pimiento, C., Leprieur, F., Silvestro, D., Lefcheck, J.S., Albouy, C., Rasher, D.B. *et al.* (2020b). Functional diversity of marine megafauna in the Anthropocene. *Science Advances*, 6, eaay7650.

Podani, J. & Schmera, D. (2006). On dendrogram-based measures of functional diversity. *Oikos*, 115, 179-185.

Pollock, L.J., Thuiller, W. & Jetz, W. (2017). Large conservation gains possible for global biodiversity facets. *Nature*, 546, 141.

Reeves, J.C., Moon, B.C., Benton, M.J. & Stubbs, T.L. (2020). Evolution of ecospace occupancy by Mesozoic marine tetrapods. *Palaeontology*, n/a.

Reich, P.B., Ellsworth, D.S., Walters, M.B., Vose, J.M., Gresham, C., Volin, J.C. *et al.* (1999). Generality of leaf trait relationships: A test across six biomes. *Ecology*, 80, 1955-1969.

Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M. *et al.* (2003). The Evolution of Plant Functional Variation: Traits, Spectra, and Strategies. *International Journal of Plant Sciences*, 164, S143-S164.

Ricklefs, R.E. (2005). Small clades at the periphery of passerine morphological space. *American Naturalist*, 165, 43-659.

Ricklefs, R.E. (2010). Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 1265-1272.

Rodriguez, A. & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, 344, 1492.

Rüger, N., Condit, R., Dent, D.H., DeWalt, S.J., Hubbell, S.P., Lichstein, J.W. *et al.* (2020). Demographic trade-offs predict tropical forest dynamics. *Science*, 368, 165.

Rundell, R.J. & Price, T.D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24, 394-399.

Sala, E., Mayorga, J., Bradley, D., Cabral, R.B., Atwood, T.B., Auber, A. *et al.* (2021). Protecting the global ocean for biodiversity, food and climate. *Nature*, 592, 397–402.

Sanders, D., Thébault, E., Kehoe, R. & Frank van Veen, F.J. (2018). Trophic redundancy reduces vulnerability to extinction cascades. *Proceedings of the National Academy of Sciences*.

Savile, O.B.O. (1957). ADAPTIVE EVOLUTION IN THE AVIAN WING. *Evolution*, 11, 212-224.

Schluter, D. (1993). Adaptive Radiation in Sticklebacks - Size, Shape, and Habitat Use Efficiency. *Ecology*, 74, 699-709.

Schneider, F.D., Morsdorf, F., Schmid, B., Petchey, O.L., Hueni, A., Schimel, D.S. *et al.* (2017). Mapping functional diversity from remotely sensed morphological and physiological forest traits. *Nature Communications*, 8, 1441.

Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., Banerjee, A. *et al.* (2015). BHPMF – a hierarchical Bayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. *Global Ecology and Biogeography*, 24, 1510-1521.

Somers, R.H. (1962). A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review*, 27, 799-811.

Sosiak, C.E. & Barden, P. (2021). Multidimensional trait morphology predicts ecology across ant lineages. *Functional Ecology*, 35, 139-152.

Spehn, E.M., Hector, A., Joshi, J., Scherer-Lorenzen, M., Schmid, B., Bazeley-White, E. *et al.* (2005). Ecosystem effects of biodiversity manipulations in European grasslands. *Ecological Monographs*, 75, 37-63.

Stubbs, T.L. & Benton, M.J. (2016). Ecomorphological diversifications of Mesozoic marine reptiles: the roles of ecological opportunity and extinction. *Paleobiology*, 42, 547-573.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S. & Brosse, S. (2021). Human impacts on global freshwater fish biodiversity. *Science*, 371, 835.

Thorndike, R.L. (1953). Who belongs in the family? *Psychometrika*, 18, 267-276.

Thuiller, W., Guéguen, M., Georges, D., Bonet, R., Chalmandrier, L., Garraud, L. *et al.* (2014). Are different facets of plant diversity well protected against climate and land cover changes? A test study in the French Alps. *Ecography*, 37, 1254-1266.

Trindade-Santos, I., Moyes, F. & Magurran, A.E. (2020). Global change in the functional diversity of marine fisheries exploitation over the past 65 years. *Proceedings of the Royal Society B: Biological Sciences*, 287, 20200889.

Valentine, J.W., Collins, A.G. & Meyer, C.P. (1994). Morphological Complexity Increase in Metazoans. *Paleobiology*, 20, 131-142.

Villeger, S., Mason, N.W.H. & Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, 89, 2290-2301.

Villeger, S., Novack-Gottshall, P.M. & Mouillot, D. (2011). The multidimensionality of the niche reveals functional diversity changes in benthic marine biotas across geological time. *Ecology Letters*, 14, 561-568.

Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N.J.B., Cadotte, M.W. *et al.* (2017). Functional Rarity: The Ecology of Outliers. *Trends in Ecology & Evolution*, 32, 356-367.

Walker, B.H. (1992). Biodiversity and Ecological Redundancy. *Conservation Biology*, 6, 18-23.

Westoby, M. (1998). A leaf-height-seed (LHS) plant ecology strategy scheme. *Plant and Soil*, 199, 213-227.

Winemiller, K.O. (1991). Ecomorphological Diversification in Lowland Freshwater Fish Assemblages from Five Biotic Regions. *Ecological Monographs*, 61, 343-365.

Winemiller, K.O., Fitzgerald, D.B., Bower, L.M. & Pianka, E.R. (2015). Functional traits, convergent evolution, and periodic tables of niches. *Ecology Letters*, 18, 737-751.

Winemiller, K.O. & Rose, K.A. (1992). Patterns of Life-History Diversification in North American Fishes: implications for Population Regulation. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 2196-2218.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F. *et al.* (2004). The worldwide leaf economics spectrum. *Nature*, 428, 821-827.

Xu, D. & Tian, Y. (2015). A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2, 165-193.

Zhu, L., Fu, B., Zhu, H., Wang, C., Jiao, L. & Zhou, J. (2017). Trait choice profoundly affected the ecological conclusions drawn from functional diversity measures. *Scientific Reports*, 7, 3643.

Table 1: We compiled 30 published trait datasets including 17 assembled in the CESTES database (Jeliazkov *et al.* 2020). We selected datasets with at least 40 species and 10 traits for high-dimensional space and perform sub-selection analyses (Supplementary Figure 1). Further, to allow the calculation of trait dissimilarity between all species pairs and test of robustness to trait omission, we removed traits with missing information for more than 60% of species and species with more than 50% of missing information for their traits. This procedure altered 4 out of 30 datasets and on average removed 15% (7-25%) of species and 57% (22-79%) of traits per dataset. We provide below a description and a reference of each dataset with the geographic extent and location (Area), species richness (S), the number of traits (T), the percentage of missing values (%NA), the percentage of quantitative traits (%Q) (the others, 1-%Q, being categorical), the mean pairwise Kendall correlation between traits (Cor) and a unique icon for each taxon used in some other Figures.

Datasets	**Taxon**	**Area**	**S**	**T**	**%NA**	**%Q**	**μ Cor**	**Icons**
Biolog	Bacteria	Global	865	97	0	99	0.17	
Bartonova *et al.* 2016	Butterfly	Czech Republic	128	13	0	100	0.22	
BirdLife	Bird	Global	9297	20	0	100	0.12	
Carvalho *et al.* 2015	Stream fishes	Amazonia. Brazil	65	26	0	4	0.17	
Charbonnier *et al.* 2016	Bird	Europe	73	10	0	40	0.15	
Chmura *et al.* 2016	Plant	Poland	46	17	0	94	0.18	
Fish Base	Chondrichthyes	Global	969	14	23	79	0.21	
Clearly *et al.* 2016	Vertebrate	Jakarta Bay. Indonesia	165	15	0	87	0.43	
Coral Trait Database	Invertebrate	Global	802	12	25	42	0.12	
Diaz *et al.* 2008	Invertebrate	Segura River. Spain	208	62	0	0	0.12	
Eallonardo *et al.* 2013	Plant	New York State. US	41	11	0	55	0.24	
Toussaint *et al.* 2016	Freshwater-fish	Global	8134	10	3	100	0.10	
Fried *et al.* 2012	Plant	France	75	10	0	30	0.17	
Gibb *et al.* 2015	Spider	South-Eastern Australia	86	10	0	100	0.41	
Goncalves *et al.* 2014	Spider	Brazilian coast	112	21	0	95	0.32	
Jeliazkov *et al.* 2013	Macro-invertebrate	France	112	89	0	0	0.14	

Krasnov *et al.* 2015	Ectoparasite	Palearctic	177	12	0	100	0.17	
Loiseau *et al.* 2020	Terrestrial mammals	Global	4675	15	0	73	0.14	
McLean *et al.* 2018	Fish	North Sea. Atlantic	138	14	3	64	0.13	
Doledec *et al.* 2011	Stream macro-invertebrate	New-Zealand	495	59	0	0	0.14	
Pakeman *et al.* 2011	Plant	Scotland	148	28	0	36	0.14	
Kissling *et al.* 2019	Plant	Global	2557	22	28	82	0.17	
Pavoine *et al.* 2011	Plant	Algeria	56	14	0	29	0.15	
Rimet & Druart 2018	Phytoplankton	Temperate lakes	1222	15	0	40	0.30	
Thuiller *et al.* 2014	Plant	French Alps	3718	33	16	12	0.12	
Riberta *et al.* 2001	Beetle	Scotland	68	20	0	50	0.17	
Chapman *et al.* 2019	Thermal vent	Global	646	16	15	31	0.18	
USDA 2020	Plant	US	1876	20	6	90	0.09	
Villéger *et al.* 2012	Fish	Mexico	46	16	0	100	0.19	
Yates *et al.* 2014	Ant	New South Wales Australia	123	11	0	91	0.18	

Figure 1: Theoretical example showing the different steps of our framework from species trait matrix (a) to species trait space (d) after calculating species pairwise distances (b) and extracting synthetic axes providing new species coordinates in a low-dimensional space (c). Then the ranking of species pairs in both high-dimensional (i.e. considering all traits so distance matrix b) and low-dimensional space (i.e. considering coordinates on few axes in c) can provide a Q matrix where the diagonal corresponds to all species pairs with a perfect match in their ranking in both spaces while off diagonal values correspond to mismatching species pairs in the co-ranking, i.e. species get closer in low-dimensional space (intrusion) or farther (extrusion) compared to their relative position in the high-dimensional space. A clustering algorithm isolates two unique species (species B and D) in the trait space (no neighbors within a given radius d_0) and creates two clusters with 2 (green) and 3 (red) species (f). See Methods for details.

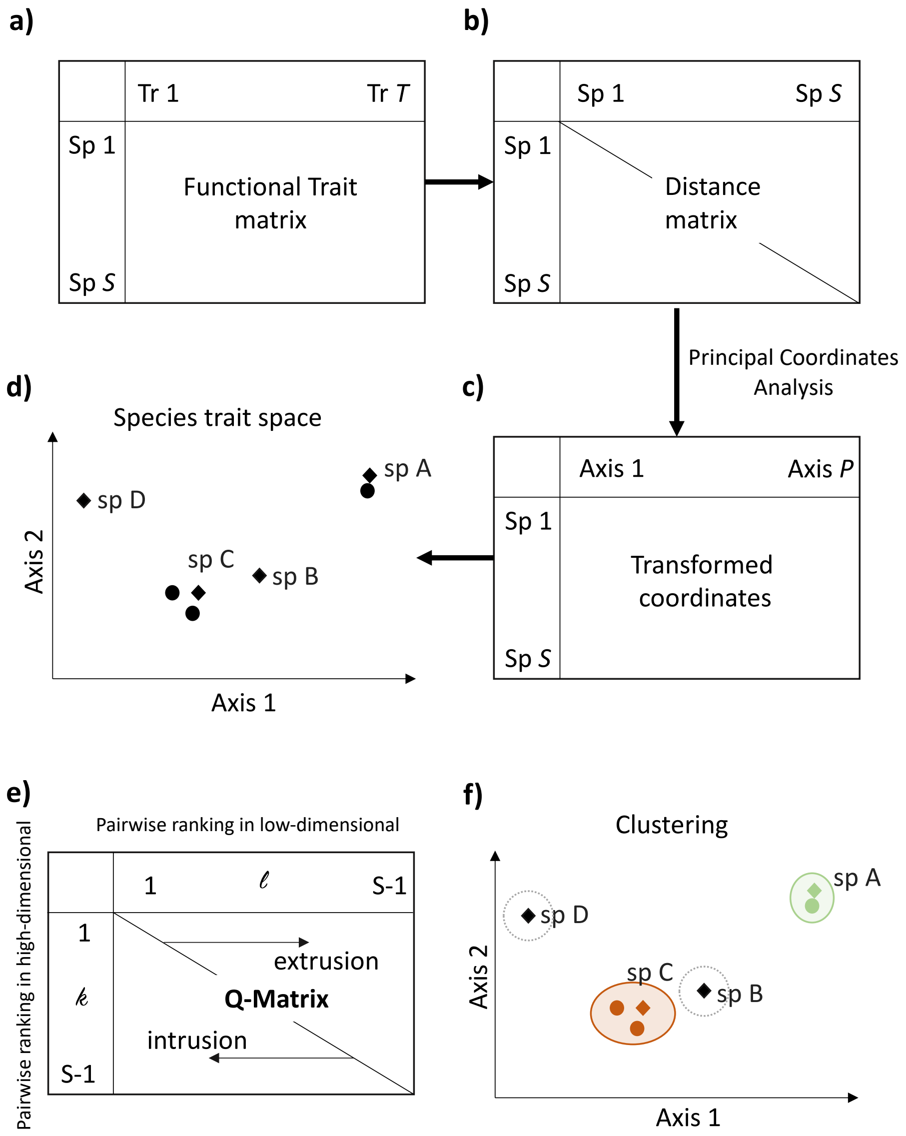

Figure 2: Influence of the number of dimensions (number of retained PCoA axes) used to build the 30 species trait spaces on the space quality assessed by the Area Under the Curve (AUC) criteria. The black dots and dotted lines correspond to the elbow-based optimal dimensionality for each dataset. The values indicate the elbow-based dimensionality, the total species richness (#*S*) and the total number of traits (#*T*) in each dataset. Datasets are ranked (top-left to bottom-right and from dark green to dark red) following the number of species.

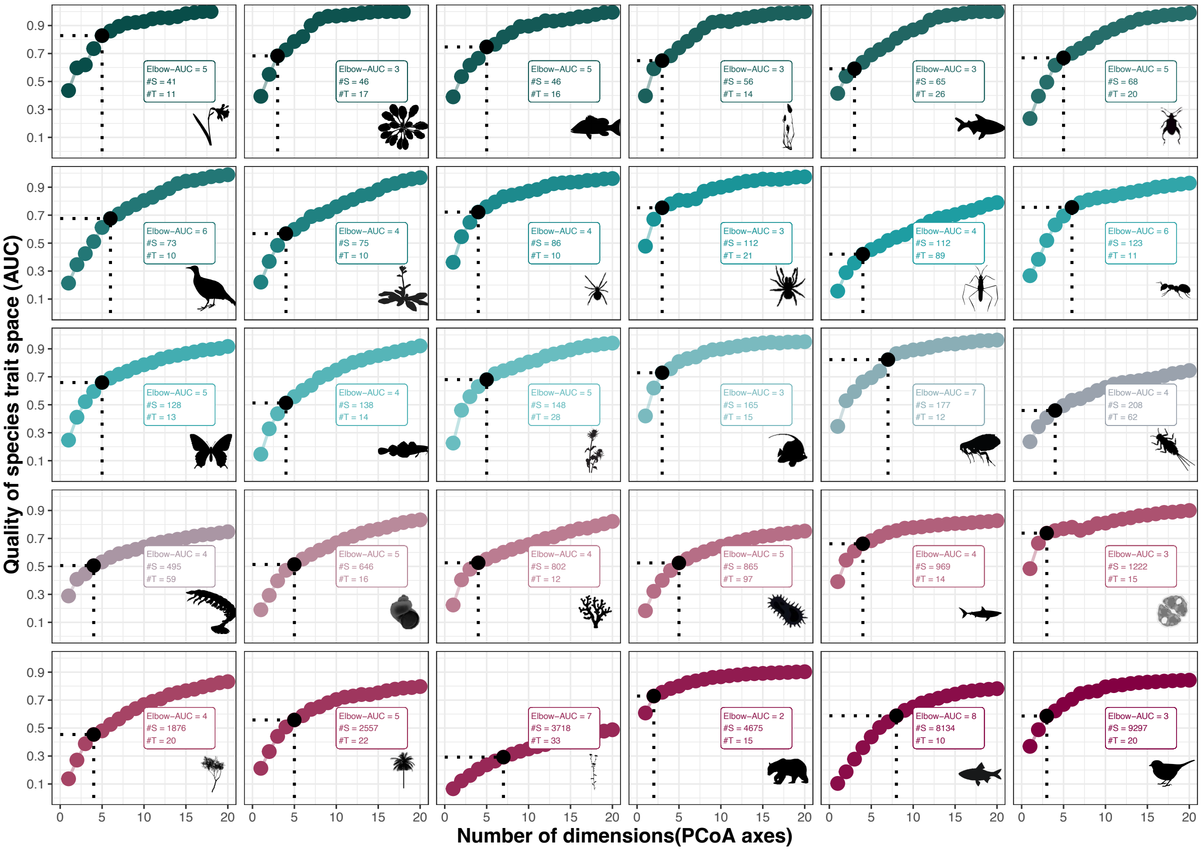

Figure 3: Partial plots showing the influence of the species life form (Plant, Invertebrate, Vertebrate) and ecosystem type (Aquatic, Terrestrial), while controlling for the five dataset quantitative characteristics, on species trait space dimensionality measured with the elbow-based (first row) or threshold-based (second row) AUC criteria. The third row shows trait space robustness, in terms of AUC loss, to trait removal or omission (50%) according to the two factors being tested. Related statistics are reported in the Supplementary Table 1, the effects are all non-significant.

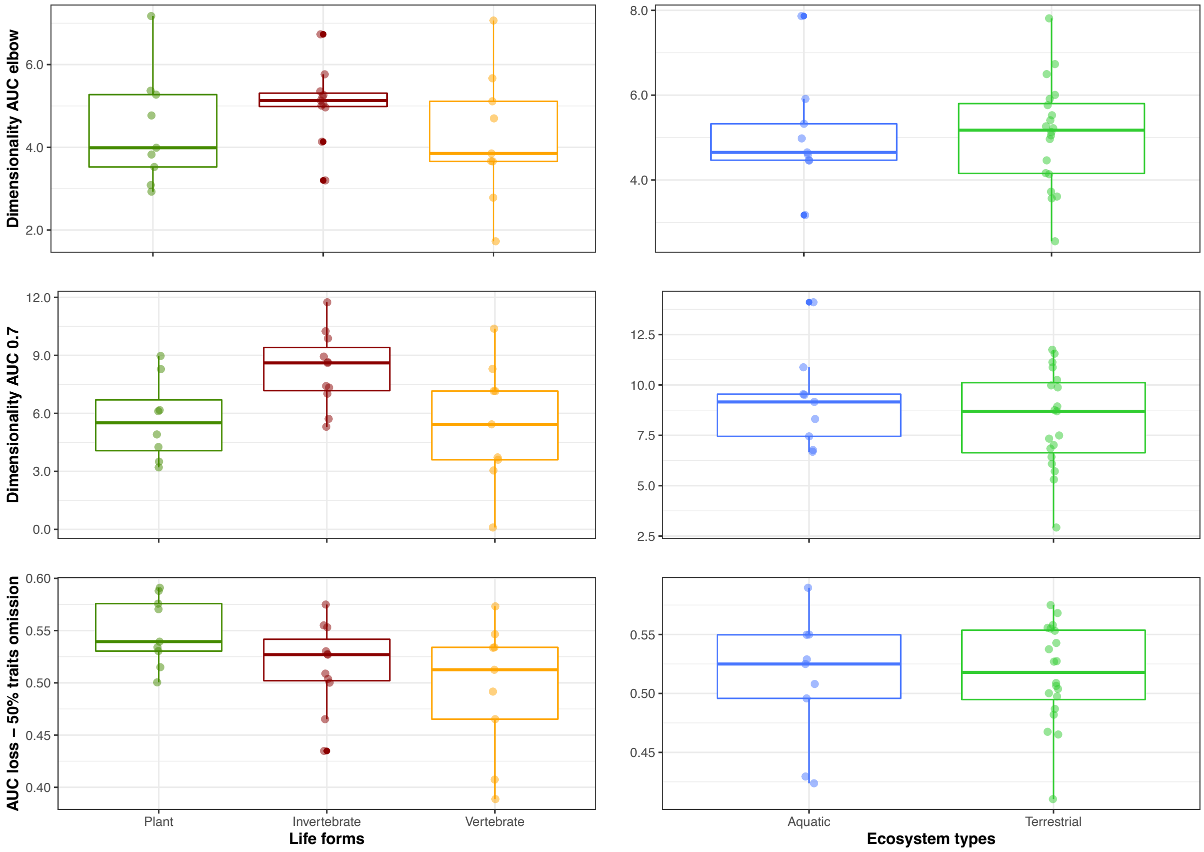

Figure 4: Partial plots showing the influence of the five trait dataset characteristics on species trait space dimensionality measured with the elbow-based (first row) or threshold-based (second row) AUC criteria. The third row shows trait space robustness, in terms of AUC loss, to trait removal or omission (50%) according to the five characteristics. Only significant (*p*<0.05) relationships are colored the others are grey. Related statistics are reported in Supplementary Table 2.

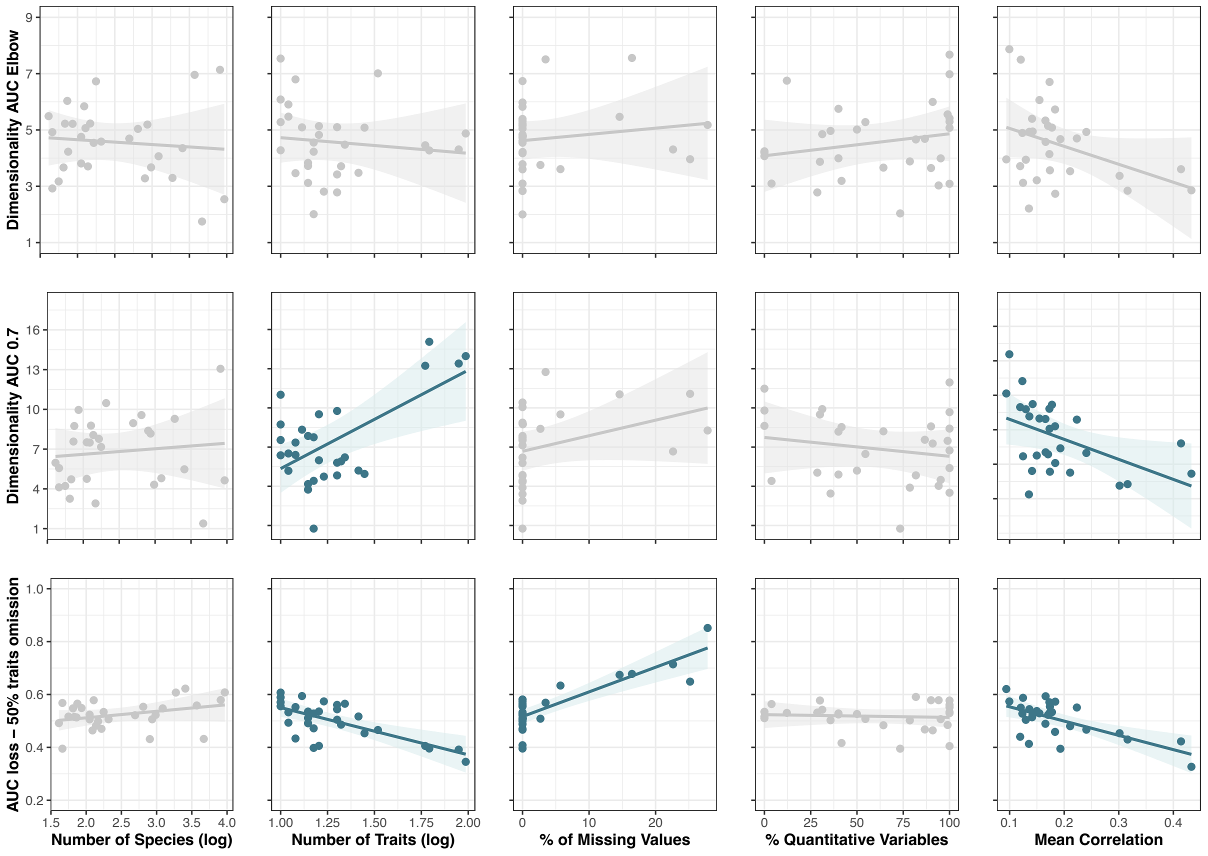

Figure 5: Influence of the percentage of traits omission (between 10% and 80%) on the quality of the trait space in terms of AUC when representing species in a trait space of lower-dimensionality. For this, we randomly removed traits 100 times for each level of omission to obtain the boxplots across the 30 datasets ranked by the total number of species (top-left to bottom right). For 0% of trait omission AUC is 1.

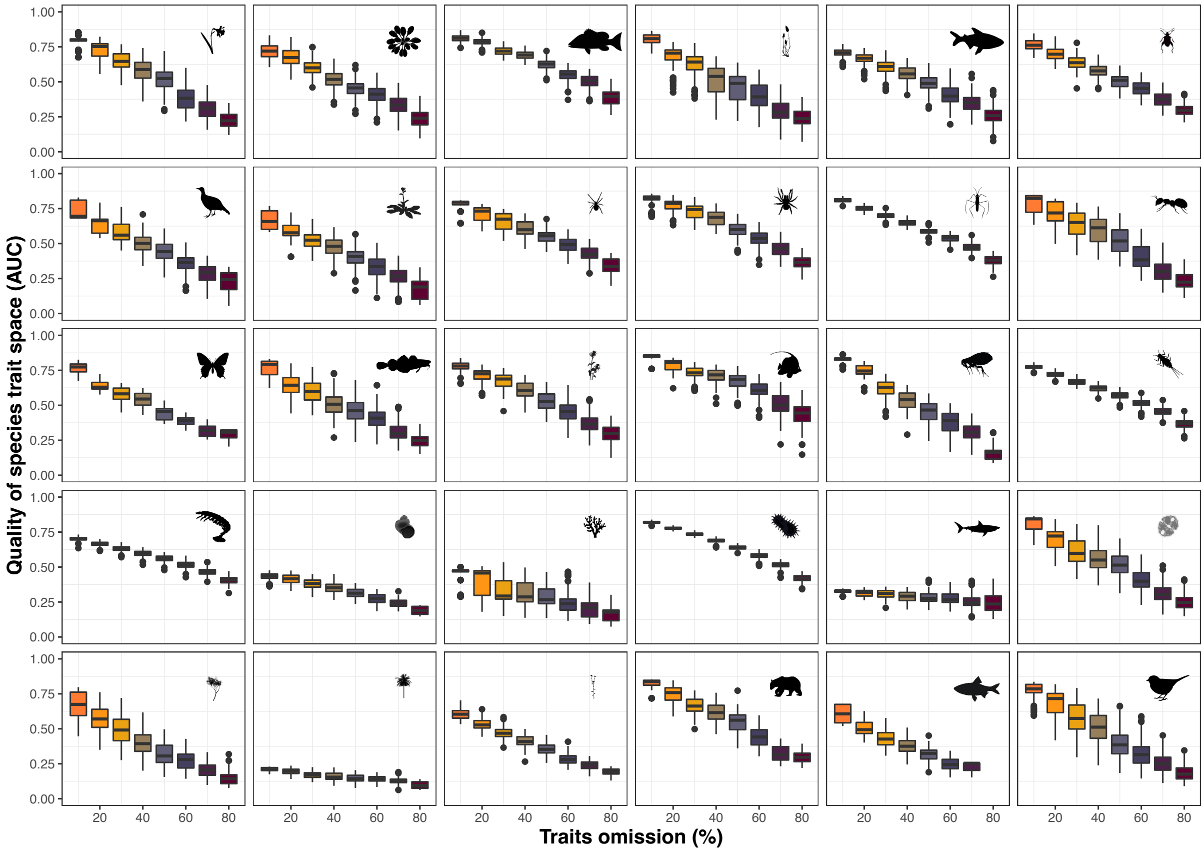

Figure 6: Partial plots showing the influence of the species life form (Plant, Invertebrate, Vertebrate) and ecosystem type (Aquatic, Terrestrial), while controlling for the five dataset quantitative characteristics, on the log-number of species clusters (first row), the proportion of species packed in the first or dominant cluster (second row) and the proportion of unique species so those isolated in the trait space (third row). Related statistics are reported in the Supplementary Table 3, the effects are all non-significant.

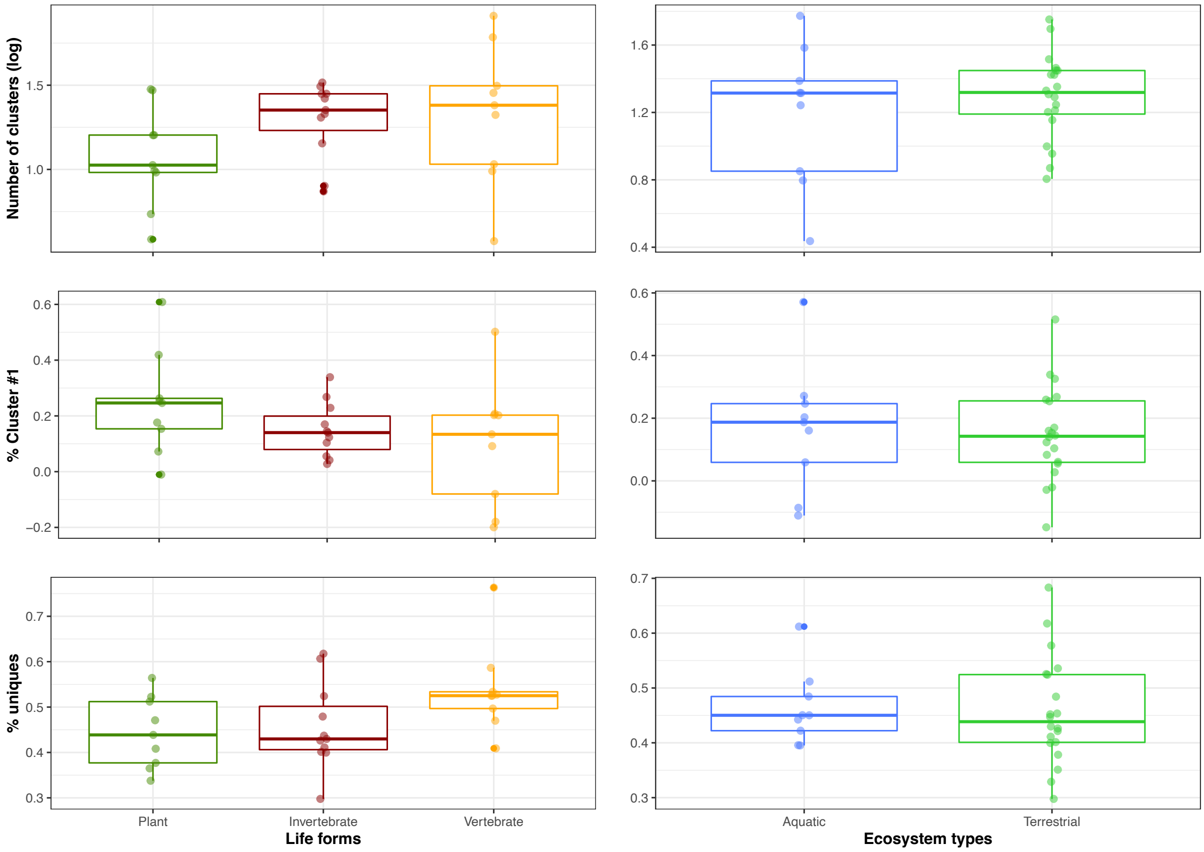

Figure 7: Partial plots showing the influence of the five trait dataset characteristics on the log-number of species clusters (first row), the proportion of species packed in the first or dominant cluster (second row) and the proportion of unique species so those isolated in the trait space (third row). Only significant (*p*<0.05) partial relationships are blue plain dots and lines, others are in grey. Related statistics are reported in Supplementary Table 4.

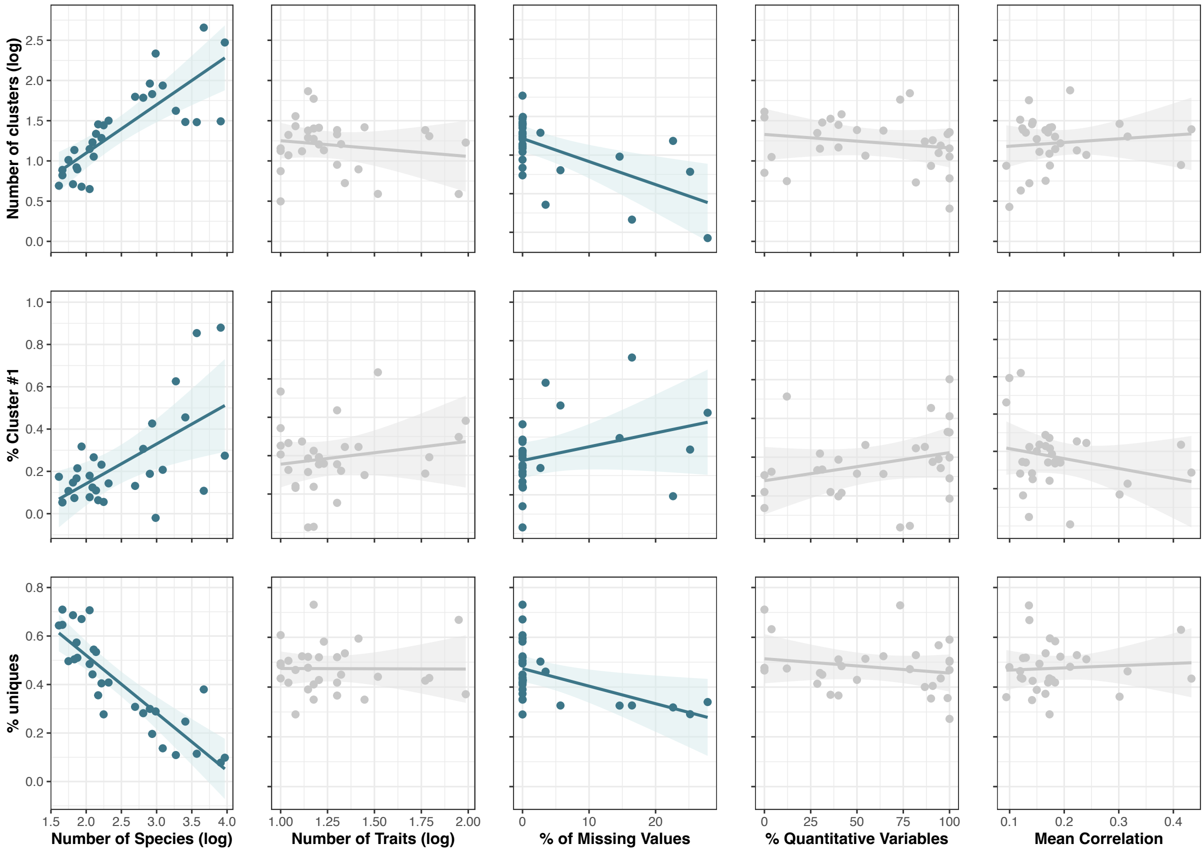

Figure 8: Partial log-log relationships between the number of groups clustered by the fast search and find of density peaks algorithm (a), the number of species in the most dominant cluster (b) and the total number of unique species so those not being part of any group (c), and the number of species in the 30 datasets. Slopes of the log-log relationships, so exponents of power laws, are reported.

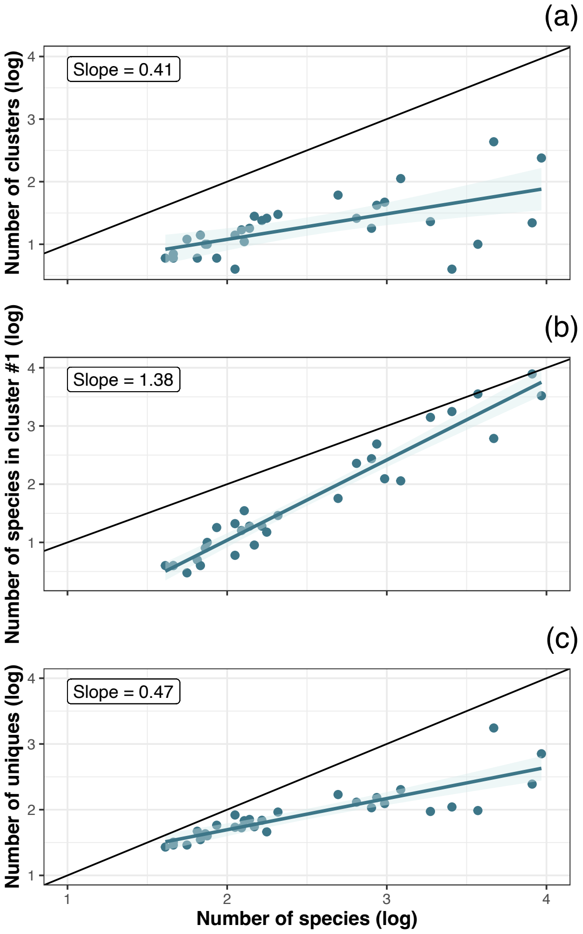

Figure 9. Trait spaces for the 30 datasets where the two axes come from Principal Coordinates Analyses (PCoA) representing the distribution of species according to their trait values. Species colored in dark are detected as statistically and ecologically unique species by the fast search and find of density peaks algorithm. The whale shark (*Rhincodon typus*) is highlighted in blue being highly distinct and unique in its clade. Datasets are ranked (top-left to bottom-right and from dark green to dark red) following the number of species.