



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The impact of plant functional trait choice in inferences of habitat filtering and competition

Bachelor Thesis

J. Zweidler

February 25, 2023

Advisors: Prof. C. Zhang, Dr. D. Maynard, Dr. A. Paz

Department of Computer Science & Environmental Systems Science, ETH Zürich

Abstract

Global conservation and restoration efforts rely on accurately quantified biodiversity measures to assess ecosystem services. Although species richness is commonly used as a measure of biodiversity, functional diversity provides a more comprehensive understanding of ecosystem functioning and community assembly. Functional diversity poses a challenge because there are dozens of traits one could measure for any given ecosystem and focal organism, with varying combinations resulting in dissimilar patterns. The key is to identify trait combinations that reflect the full range of traits and produce consistent results. Here, we address this challenge by using a global database of tree traits to explore how different trait combinations and relationships affect functional biodiversity patterns. We find that combinations with little to no inter-trait-correlation produce the most reliable results. They tend to outperform other trait selection procedures when it comes to accuracy in trait space representation and result coherence among similar combinations. Collectively, this work describes core principles in the tree functional traits space which can contribute to further establishing functional biodiversity as a dependable instrument.

Contents

Contents	iii
1 Introduction	1
2 Methods	5
2.1 Setup	5
2.1.1 Occurrence Data	5
2.1.2 Trait Data	5
2.1.3 Normalization	7
2.2 Trait combinations	8
2.2.1 Programming Environment	8
2.3 Metric calculations	9
2.3.1 Functional Richness	9
2.3.2 Functional Divergence	11
2.4 Statistics & Analysis	11
3 Results	13
4 Discussion	25
A Appendix	29
Bibliography	35

Chapter 1

Introduction

The term biodiversity refers to the range and variability of life forms that exist on our planet. This encompasses genetics, species, and ecosystem diversity. It is crucial to conduct a precise assessment of biodiversity for various reasons. By accurately evaluating the structural diversity and differences in niches within habitats, we can help maintain the coexistence of different species, promote effective conservation, and better manage natural resources. There are several objective methods available to empirically measure biodiversity, with the traditional approach being species diversity. However, in recent decades, ecologists have increasingly incorporated the measurement of functional traits in their analysis of ecosystems. This is based on the idea of competition and environmental filters, such as climate, disturbance, and biotic interactions, imposing more refined constraints on community composition as more species are added. Therefore, the same amount of functional variation or niche space is more finely divided, making functional diversity a better indicator of habitat heterogeneity and ecosystem functioning. [1] [2].

Functional traits are the set of morphological, physiological or phenological features which represent ecological strategies regarding environmental factors and competition [3]. They have proven to be useful to investigate not only community assembly and ecosystem dynamics, but also ecological diversity [4] [5]. Recent approaches have proven that functional diversity provides stable grounds for analyses either by their own or in combination with preexisting phylogeny-based metrics [6] [7]. The insights gained by these methods can be tremendous, however they do not come without their own downsides. Compared to phylogeny, the field and lab work involved in the data acquisition for functional traits is much bigger. Additional traits considered can incur a much bigger workload to a project, especially taking into account the vast amount of existing or putative traits to choose from. Including the entire trait space would inevitably lead to the most accurate

1. INTRODUCTION

results, yet would be an infeasible task for most project capacities. Therefore an analysis should strive to find the best subspace representation of the trait space by eliminating redundant axes.

Recognizing the degree of complexity of these systems begs the question of how to reduce these dynamics into stable, quantitative and measurable metrics. Previous studies have shown the existence of three distinct axes within the space of functional diversity [8]. These have been described as follows:

1. Functional richness corresponds to the volume of the functional space occupied by species within a community [9].
2. Functional divergence defines how far high species abundances are from the centre of the functional space [10].
3. Functional evenness measures the regularity of the distribution of species abundances and dissimilarities in functional space [11].

Past work has proven that existing metrics qualitatively capture each of the described axes [10] [12]. They underlined the representative correctness of these metrics assuming a given functional trait space. In order to ensure the quality of our results we therefore also should ensure the quality of the trait space. Studies suggests that indeed there exists measurable distinctions in the quality between functional spaces and that these greatly depend on the dimensionality of the trait space [13] [14]. Their research however concludes with no obvious finding of how to predict which functional subspace gives the highest representative accuracy. This however is a crucial step in being able to guarantee qualitative analysis using diversity metrics. We therefore aim to further explore the relations between trait spaces, their meaningfulness and stability.

Our analysis differs from previous general studies as they had to rely on simulated data due to the scarcity of global data availability [15]. Focusing solely on tree functional data enabled us to query large global databases to accumulate our data set. We combined information from multiple databases in order to populate over 21'000 communities world wide. Out of a multitude of possibilities, we selected 18 unique trait measurements to include in our analysis. These were chosen considering their representation in prior trait analyses, their functional uniqueness, their relevance to tree structure and architecture, and the data quality [16].

Concentrating on trait data from trees, recent work has been able to draw comparisons and uncover relations between functional traits [16]. Using hierarchical clustering on a species-level correlation matrix, they were able to group the 18 considered traits into 8 specific clusters. Each of these clusters shows high intra-trait correlation. These observations propose that a possi-

ble bias could indeed exist when comparing subsets of functional traits especially taken into consideration how various diversity metrics are calculated. This inevitably leads to a question of credibility of these functional diversity metrics. Equipped with this knowledge about the tree trait space we asked ourselves, to what extent has conducted field work possibly been based on unreliable trait spaces. Most projects have naturally chosen traits in up to 5 of the 8 clusters described above [7] [17] [18] [19] [20]. The trait spaces tended to be comprised of traits that can easily be measured such as leaf traits or seed mass while root depth or crown diameter is more commonly avoided. We conclude that parts of the conducted field work could have lead to only slight improvements of the obtained results if not being entirely redundant. These processes could possibly be improved by changing the selected subset of traits and therefore either expanding the trait-subspace or by reducing the workload needed for data acquisition and analysis. Choosing similar or more stable trait spaces would also open up the possibility of potential comparisons between studies and ecosystems and could therefore lead to additional insights.

Knowing the significance of reliable biodiversity values, the goal of this thesis is to explore how functional diversity metrics are influenced by the selection of functional trait sub-spaces. Using a global database comprising almost 3'500'000 observations across more than 38'000 species in tandem with recent models of tree functional traits, we compare how the selection of different functional traits affects the resulting global patterns in biodiversity. We expect that choosing only strongly correlated traits will reduce the occupied trait space compared to picking non correlated traits resulting in lower functional diversity metrics. Additionally, we predict that among trait combinations which focus on strongly correlated traits we will see greater variation in their diversity values as with such combinations selecting non correlated traits. Understanding of these underlying dynamics can inform broad-scale biodiversity monitoring by helping to identify and prioritize subsets of traits that will yield consistent values.

Chapter 2

Methods

2.1 Setup

2.1.1 Occurrence Data

In this project we are working with tree community data for ecosystems around the world. This inherently implies working with maps, which for our purposes it suffices to use a coarse grid. As a starting point I was provided with data that has been generated and used in work by Paz et al. [21]. Tree occurrence data was obtained from public databases and filtered using the GlobalTreeSearch data of global tree species [22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. The obtained data set was cleaned to ensure occurrence points occurred in landmasses. No points were considered geographical and environmental outliers for any given species. The curated data set was used to create individual species distribution polygons as alpha hulls. Finally, alpha hulls of species distributions, for which both functional traits and phylogenetic information were available, were combined to obtain the number and identity of species within each 100 * 100 km pixel across the world. The resulting grid map (432 x 208) is very coarse, thus capturing regional gamma-diversity rather than community-level alpha diversity.

2.1.2 Trait Data

Trait data for this analysis comes from the recently published work by Maynard et al. [16], in which they used TRY trait data to impute 18 traits across ca. 50,000 tree species to identify physiological trade-offs [32]. This insight of existing clusters within the trait space will function as a basis for further inferences. While we work off of their assumption that the trait space is grouped into 8 distinct clusters as described by figure 2.1, adapted from Maynard et al. [16], we decided to simplify and merge these into 5 unique clusters for two reasons. First, previous studies commonly picked traits from

2. METHODS

approximately 5 (versus 8) clusters. Additionally, we encountered computational restraints when calculating functional richness metrics using 8 dimensions. Taking figure 2.1 as a basis, we produced our new correlation matrix with 5 functional clusters in following way. The dendrogram on the left of figure 2.1 displays how closely related individual traits are to each other. We simply picked those two clusters which had the strongest affiliation and merged them into one, repeating this process until we ended up with the final 5 clusters. In order to compare the differences in sub-spaces spanned by the above combinations we calculate the selected diversity metrics. Each of these metrics uniquely describe the same community by the means of different traits selected which we take as a prompt to proceed with our cluster analysis.

2.1. Setup

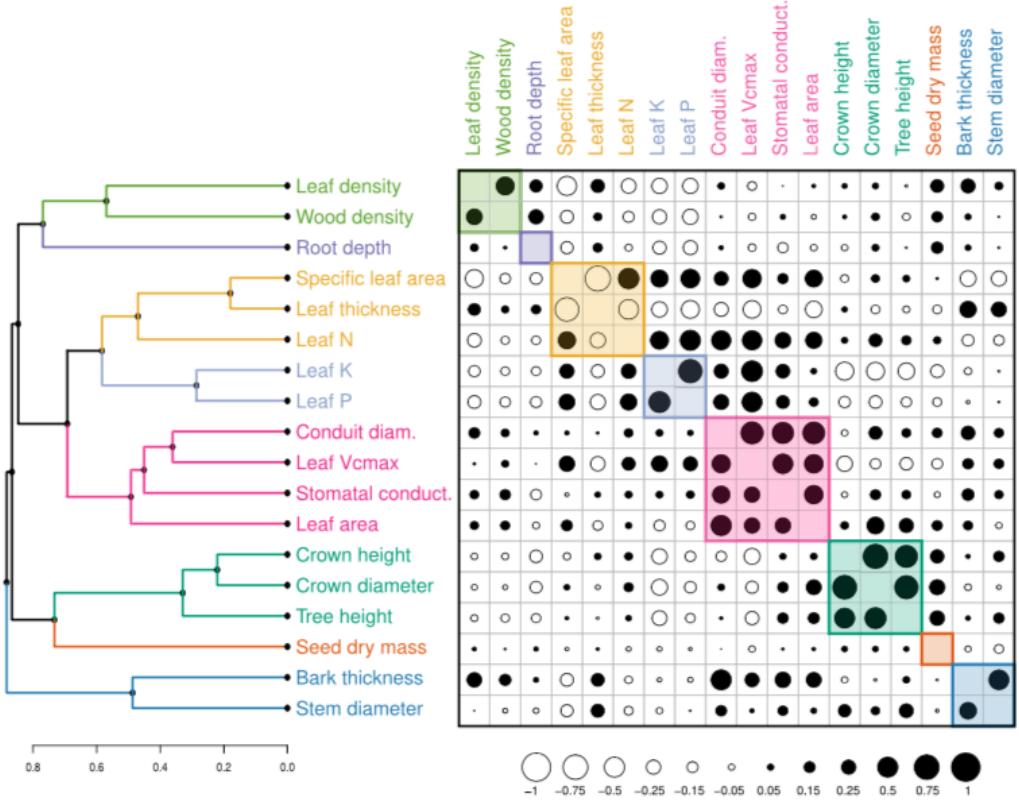


Figure 2.1: Trait correlations and functional clusters. Adapted from “Global relationships in tree functional traits” by Maynard et al. [16]. Trait clusters with high average intra-group correlation. The upper triangle gives the species-weighted correlations incorporating intra-specific variation. The lower triangle gives the corresponding correlations among phylogenetic independent contrasts, which adjusts for pseudo-replication due to the non-independence of closely related species. The size of the circle denotes the relative strength of the correlation, with solid circles denoting positive correlations and open circles denoting negative correlations.

2.1.3 Normalization

Working with very different trait values ranging over different metrics can complicate efforts to appropriately summarize and compare both traits and metrics. Taking a closer look at our data, we generally see a normal distribution with a smaller or larger tail depending on a specific trait. Typically one would proceed by clipping these outliers. This though proved not to be an option as these extreme trait values can reflect ecologically important differences, leading to an artificially restricted trait range if one were to remove these values. That is, if these individual species are absent in a community,

2. METHODS

then that skews the values of the remaining ones, pressing them together, inferring there is little to no difference within the ecosystem. Taking these considerations into account we identified two possible approaches for dealing with skewed traits and metrics. One, we standardize trait values over all possible species. This method highlights the presence or absence of these globally extreme species in local communities. Second, we only consider trait values of species within a certain local community. This method focuses on the spread within an ecosystem but ignores the potential diversity of global species. This could lead to trait measures being very different on a local scale, even though globally they would be almost identical.

2.2 Trait combinations

As mentioned, the aim of this thesis is to compare subsets of the entire trait space and the diversity metrics resulting from their use, in order to identify which trait-selection algorithms yield the most consistent results. In order to compare sub-spaces of our functional trait space we chose trait combinations of specific interest to include in our comparison. We expect them to represent underlying trends and to bring out the behaviour expected in our assumptions. We decided to focus on the following combinations based on the inter-trait correlation analysis performed by Maynard et al. [16]:

1. The 144 unique combinations comprising one trait from each of the five clusters, "one/cluster"
2. Top 150 most negatively correlated trait combinations, "most neg" (Minimum sum of inter trait correlation values)
3. Top 150 most positively correlated trait combinations, "most pos" (Maximum sum of inter trait correlation values)
4. Top 150 least correlated trait combinations, "least" (Minimum absolute sum of inter trait correlation values)
5. Top 150 greatest correlated trait combinations, "most" (Maximum absolute sum of inter trait correlation values)

2.2.1 Programming Environment

For this project we chose to use Julia as our main programming language [33]. Julia is an up and coming environment in data analysis and the development of statistical software. We selected Julia as it provides native and fast parallelism for either CPU or GPU code and also has a strong visualization and plotting interface. As Julia could possibly replace R for computationally costly settings we wanted to future proof our code. This provided enough

motivation to learn a new language and environment. Especially the interactions with Julia's still developing documentation and it's relatively small user base has imposed a steep learning curve upon this project. For this project we had access to a small lab-internal server which enabled us to use CPU-parallelism.

2.3 Metric calculations

Our analysis focuses on two of the most common metrics of functional biodiversity: functional richness (FRic), intended to capture the range of traits occupied by a community (richness); and functional divergence (Rao's Q), otherwise known as mean-pairwise distance, which reflects the degree of functional redundancy and similarity across species. Since we are missing species abundance data we will not be including functional evenness metrics in our analysis. By comparing these indices we hope to gain insight into different underlying mechanisms shaping the space of functional diversity.

2.3.1 Functional Richness

Functional richness corresponds to the volume of the functional space occupied by species within a community. FRic takes a very direct approach of this definition and tries to quantify functional richness by means of the convex hull. It has been shown to be the most accurate representation of functional richness. According to Mouchet et al. (2010)[10], FRic more accurately represents the change in functional space dimensionality caused by community structure than other metrics. It does this especially well since it reacts more sensitively to community assembly rules. So although FRic monotonically increases with species richness, it is more influenced by assembly rules [10]. On the downside, determining the convex hull implies that FRic only considers those species with the most extreme trait values for its calculations. It completely ignores further distributions of species within the convex hull. A further restriction is that we only can geometrically calculate the volume of a figure in n-trait-dimensions when we have at least n available data points. Depending on the amount of traits considered this restriction will cut small communities out of the equation. Addressing these deficiencies we looked at the number and spread of small communities in our data-set and concluded that the amount was negligible. We thus use FRic as a meaningful representation of functional richness while its shortcomings do not influence the overall trend of our analysis.

When it comes to the implementation of FRic there are a few aspects that need to be considered:

1. Which algorithm/library should be used?

2. METHODS

2. What is the run-time of this algorithm?
3. What are other issues that arise during computation?

Since the Julia language is still gaining popularity, depending on the use case there is still a deficit of available libraries and packages. Most implementations involving a convex hull use the Polyhedra.jl package where one can use various Julia wrapped Python packages. After some close observations and various testing, multiple packages could be ruled out due to erroneous calculations of the convex hull. These tests were conducted on simple objects in multiple dimensions where the volume could easily be calculated by hand. In the end we selected Qhull.jl as it gave correct results and the stand-alone variant had by far the fastest run-times.

Bench-marking our primary implementation revealed that we would inevitably run into some computational constraints. Considering the dimensions of our data (e.g., some tropical communities containing more than 1500 unique species, and thus 1500 unique vertices in 8-dimensional trait space), the calculations would go beyond the scope of this project. For that reason we decided to reduce the setting from an 8-dimensional to a 5-dimensional trait space. Apart from giving obvious computational benefits this also enabled us to draw direct comparisons to many conducted field studies as these mostly focus on traits from 5 different clusters. Further realisations about the composition of our data were gained through erroneous messages from Qhull.jl during our calculations. We noticed during our implementation that many species within a community are very close to each other in the trait space, too close for our convex hull calculations (i.e., differing by <0.001 for traits normalized to have a mean of zero and standard deviation of one). This is due to the fact that the trait values we are working with are partially imputed. This resulted in many species with a genus having very similar if not exactly the same values depending on which trait combinations we picked. Thus, further data preparation was needed. In the first step we eliminated all exact duplicates in our species list, and in a second step we added random Gaussian distributed noise with a standard deviation of $\sigma = 0.0001$ to all remaining attributes. This was not done to the original data, but only to local copies for the Qhull algorithm. These measures did not completely eliminate the occurrence of similar data points, but kept it to a minimum. In conclusion we can say that the calculation of FRic was the computationally most costly part of this thesis, particularly given the large number of simulations and combinations used here. This must be considered for future approaches, especially if working in a high-dimensional trait space.

2.3.2 Functional Divergence

The Q metric, otherwise known as Rao's quadratic entropy, is based on a distance matrix of all species in a community. The Euclidean distance matrix is a matrix of squared distances between points which in our case represent species. This can be summarized by following formula,

$$Q = \sum_{i=1}^{S-1} \sum_{j=i+1}^{S-1} d_{ij} p_i p_j$$

where i and j are two species, d_{ij} is the distance between them, S is the number of species in the grid cell, and p_i and p_j describe the relative abundance which we chose to be $1/n$. In our case, we do not have abundance data, and so $p_i = p_j = 1/n$, such that this metric is equivalent to the mean pairwise distance across all species in the grid cell. While being equally one of the most famous and most used metrics, work by Mouchet et al. (2010) [10] suggests Q to be more of a complex metric to interpret. A key feature of Rao's Q is that, unlike functional richness, it is generally independent of species richness [10]. As being sort of a go-between for functional richness and functional divergence, Q depends on both the functional space occupied but also on the interaction of species and their abundances. When it comes to the implementation, this metric is only based on simple matrix operations and therefore it is computationally not intensive. While maybe having less intuitive power than FRic, Rao's Q is fundamental for questions relating to community-assembly processes, and the implementation of Q is straight forward and very fast, which makes it attractive especially for large problem settings [34].

2.4 Statistics & Analysis

Using these functional diversity metrics, the overarching question is to understand which trait groupings and trait-selection procedures result in the smallest amount of variation and most consistent predictions. Typically this could just be done using the standard deviation, however, correlation among traits can constrain the functional space and limit the functional diversity value, leading to auto correlation among the standard deviation and the mean. Thus, in order to measure and compare the standard deviations of different groups we instead relied on the coefficient of variation (CV), which accounts for this difference by scaling the standard deviation by the mean, $CV = \sigma/\mu$. For an arbitrary grid cell g with a species richness equal to N we proceeded as follows: for each trait combination t we selected the corresponding trait data from all of the N species; using a predefined step-wise scaling analysis of a community we then calculated our functional richness and functional divergence metrics for each scaling $s \leq N$ and wrote it into our result matrix.

2. METHODS

Algorithm 1 Metric calculations

```
G ← size(grid_cells)
T ← size(trait_combinations)
S ← size(scaling)
for g in 1:G do
    N ← species richness in grid_cell[g]
    for t in 1:T do
        for s in 1:S do
            if scaling[s] < N then
                c_hull(grid_cell[g], trait_combinations[t], scaling[s])
                rao(grid_cell[g], trait_combinations[t], scaling[s])
            end if
        end for
    end for
end for
```

We then continued by dividing trait combinations into their respective groups and calculated the coefficient of variation for each grid cell for each of the five trait groupings. Projecting the results back onto the world map helped us to visualize the differences in trait-selection procedures. Map visualization was done using the QGIS software which enabled us to easily input our file format, clip values to minimum/maximum values and to fine tune our color encoding. The resulting data let us draw meaningful conclusions regarding the composition and characteristics of our trait groupings. In order to stay within acceptable run-times we needed to parallelize the calculation of the convex hull. Being one of Julia's main features, this process was very straight forward and easy to implement. Taking data movement and memory bandwidth into account we found bottlenecks and could optimise our run-time. Code snippets regarding implementation specifics can be found in the appendix.

Chapter 3

Results

Part of our literature review included studies using functional diversity analysis based on conducted field work and its resulting trait measurements. These were done using various trait combinations as can be seen in the list below. Taking a closer look at the distribution of the chosen traits we can see some clear tendencies. Most studies primarily focused on leaf traits while also including each a key aspect regarding tree form (i.e. tree height) and slow/fast life-history strategies (i.e. wood density). This raises the question of how these compare to the trait combinations that we picked for our analysis. Figure 3.1 we clearly shows that certain traits have been favored by different trait combination groups. We can see that groupings "most neg", "most pos" and "greatest" mainly focus on leaf traits while "least" tends to pick traits which are more evenly distributed among the clusters. The distribution of "one/cluster" does not have equal meaningfulness as the distribution from other groups. This is due to the traits being chosen systematically with traits from small clusters appearing more often than those from larger clusters.

- **Tropical forest:** SLA, Leaf Nitrogen, Leaf Area, Wood Density, Seed Mass, Tree Height [7]
- **Subtropical forest:** SLA, Leaf Area, Wood Density, Seed Mass, Tree Height [17]
- **Subtropical forest:** SLA, Leaf Nitrogen, Leaf Area, Wood Density, Tree Height, Leaf Thickness, Leaf Density, Leaf carbon [18]
- **Continental:** Leaf N, Wood Density, Seed Mass, Tree Height [19]
- **North and South America:** SLA, Leaf N, Leaf P, Wood Density, Seed Mass, Tree Height [20]

3. RESULTS

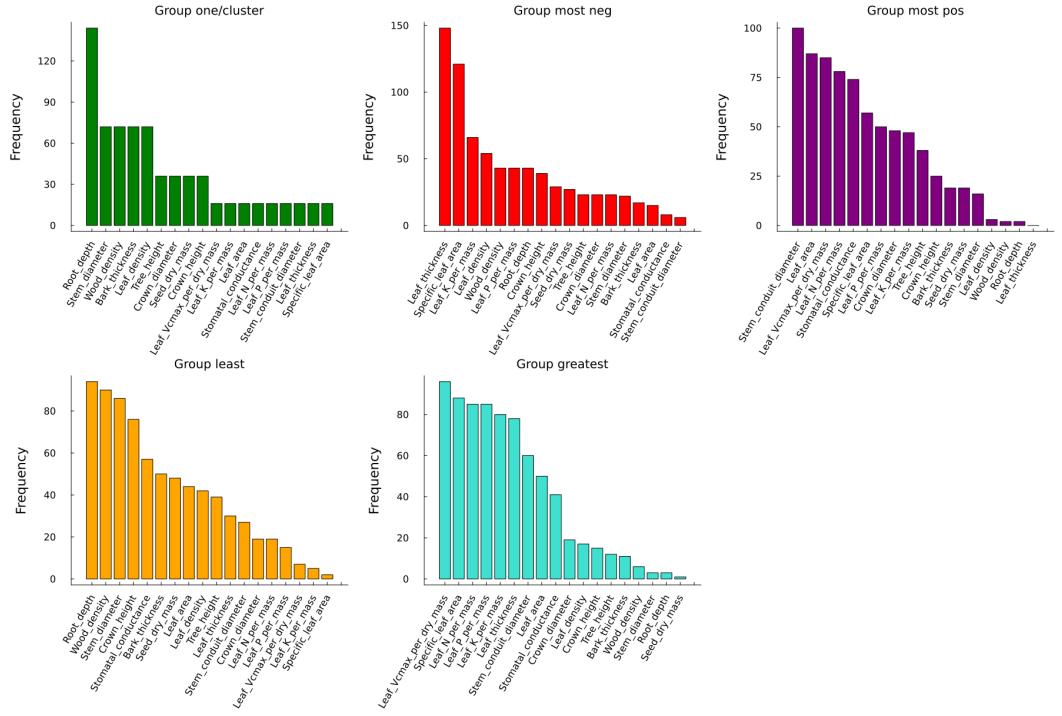
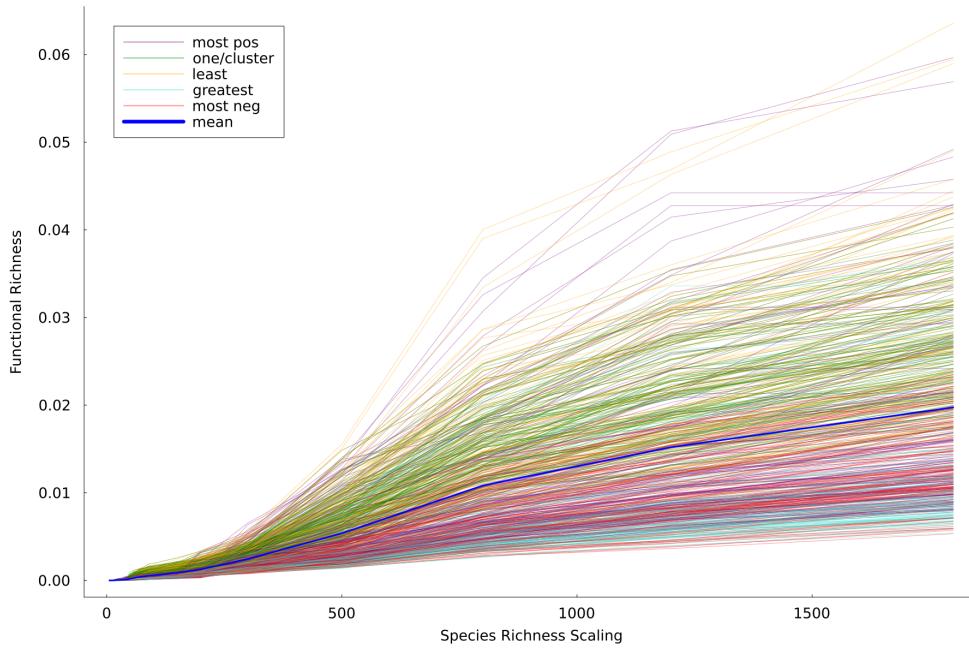
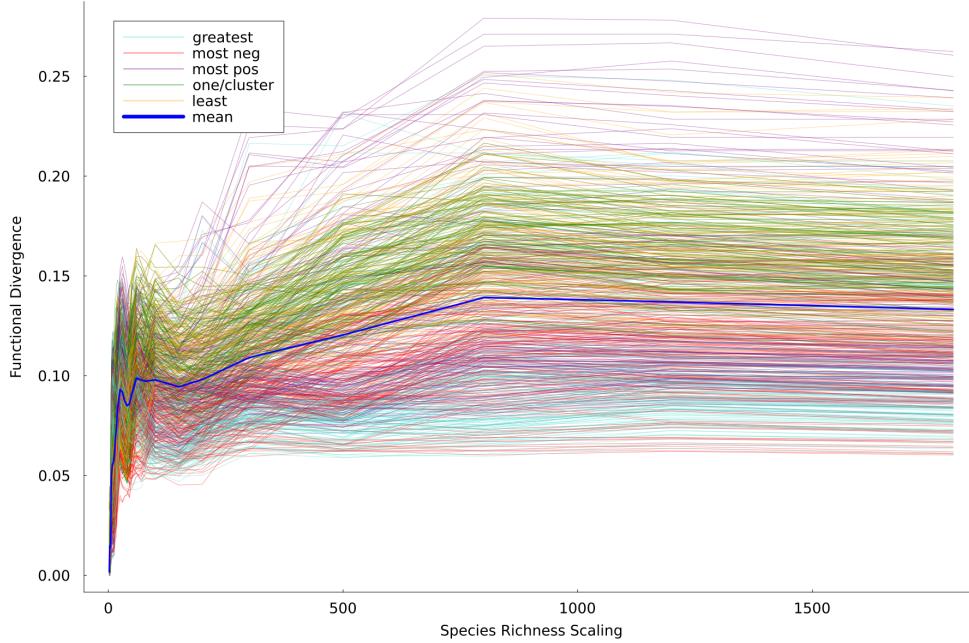


Figure 3.1: **Trait usage per group.** Each sub-figure refers to one of the trait combination groups selected for our comparison. For each group we counted the total occurrences for each trait over all trait combinations. The resulting trait distributions were sorted and plotted so the y-axis refers to the number of representations in different trait combinations within the group.

To illustrate what differences the selection of traits can have, we first conducted a step-wise scaling analysis for each community, its richness and its corresponding diversity metric. This enabled us to approximate how functional diversity varied as a byproduct of species richness, as seen in figure 3.2. The point of these graphics is to give an impression of how different these calculated metrics can be, solely by changing the traits we incorporate. In this specific grid cell we calculated FR values ranging 0.011 - 0.102 and FD values ranging 0.06 - 0.262. While FD calculations show an erratic behavior for low scaling values but converge quickly, we see an opposite trend for FR. Whereas most combinations in the lower-value-range tend to show a convergent behaviour, this does not go for all trait combinations in the high-value-range. Some groupings show signs of future stagnation (group "most pos") while others (group "least") suggest a linear growth. This can strongly influence asymptotic predictions when estimating the functional diversity, for example, when under-sampling a community.



(a) FR

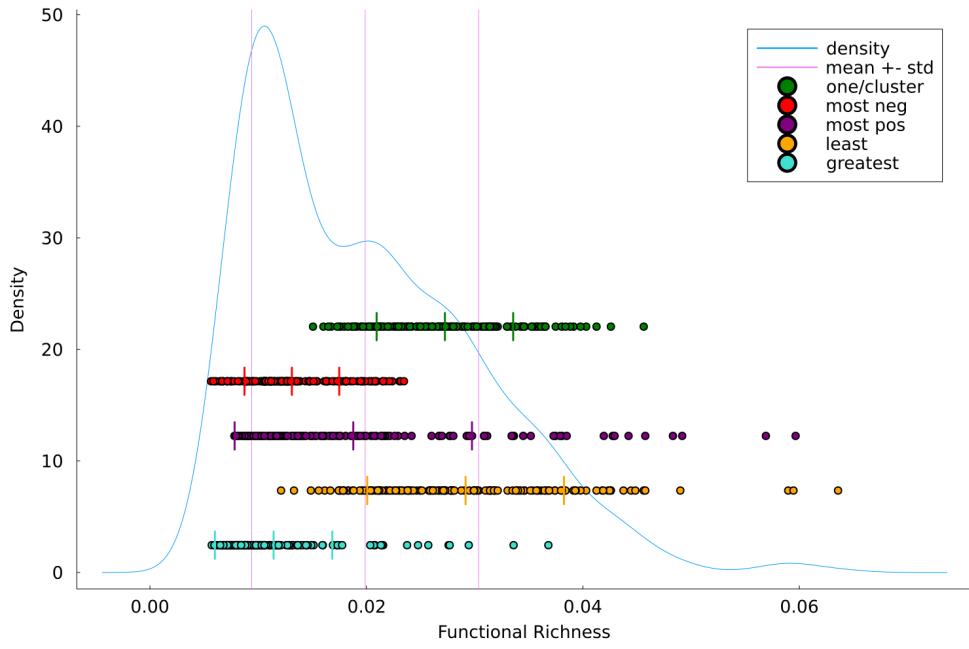


(b) FD

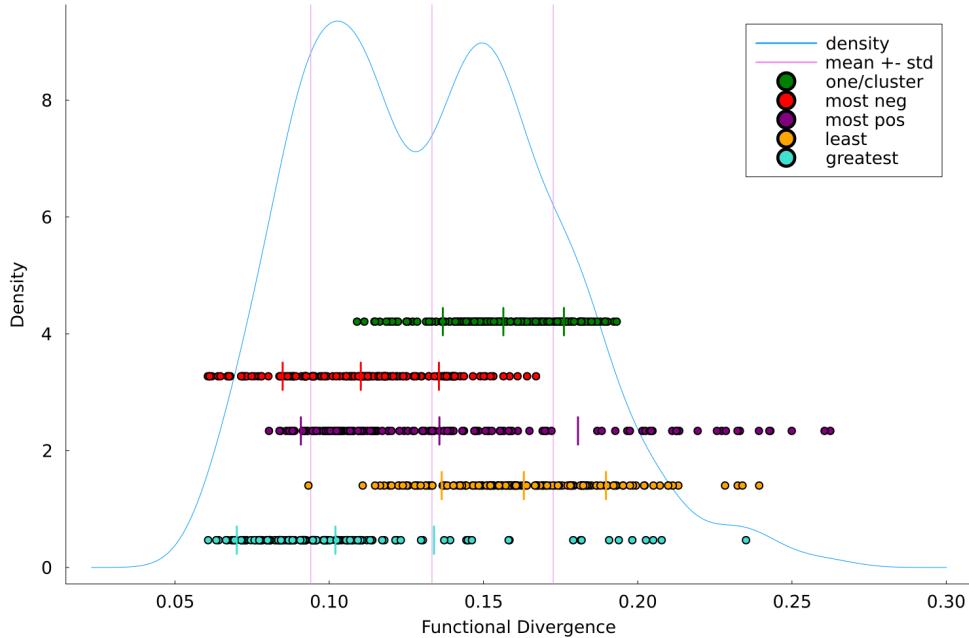
Figure 3.2: Functional diversity behaviour for a single community. We picked a single high richness community (over 1800 species) and calculated FR and FD for 38 different scalings. This enabled us to approximate the asymptotic behaviour for each trait combination in terms of the richness of a community. Each of the 744 trait combinations is represented by a piecewise linear function and colored by group association. The x-axis refers to the scaling of the community size and the y-axis to the functional diversity value.

3. RESULTS

Taking a step back from convergence behaviours, we will untangle this colourful jumble by taking the total richness of the previous community and focusing on the individual distributions of each group of traits. Figure 3.3 clearly highlights the different characteristics between trait combination groups. One being the trend towards higher or lower functional diversity values, which was already apparent from the last figure. The other one being the aforementioned intra group distribution. Since we do not know the "true" FR/FD value we cannot grade the correctness of a group simply by its distance to that value. But we can focus on assessing each group's clustering and therefore its ability to deliver concurrent and consistent results.



(a) FR



(b) FD

Figure 3.3: Density of functional diversity values for a single community.
 Referring to the same community as in Fig. 3.2, we took the functional diversity value for each of the 744 trait combination describing the total community and made a density plot. The pink long vertical lines refer to the total mean +- standard deviation. We plotted each trait combination group containing ca. 150 combinations individually also showing their individual mean +- standard deviation (the y-axis position of each group is solely for improved visualisation).

3. RESULTS

At the global scale, we can clearly see the large differences in CV when comparing different trait combination groups (Figs. 3.4 and 3.5), with both functional richness and functional divergence showing similar trends. We have two distinct clusters consisting of groups "one/cluster", "most neg" and "least", which show consistently low CV and similar spatial patterns; versus the groups "most pos" and "greatest" which yield consistently high CV, with especially high values in the low-richness, high-elevation boreal communities and interior Asia. Apart from these intra-group differences we can also make out trends in terms of community size in general. Larger communities have consistently more stable variances than smaller communities when considering functional richness. For functional divergence we see a tendency towards higher CV values for both very small and very large communities. Figure 3.6 puts the information of our maps into a density distribution to further highlight the differences between trait combination groups.

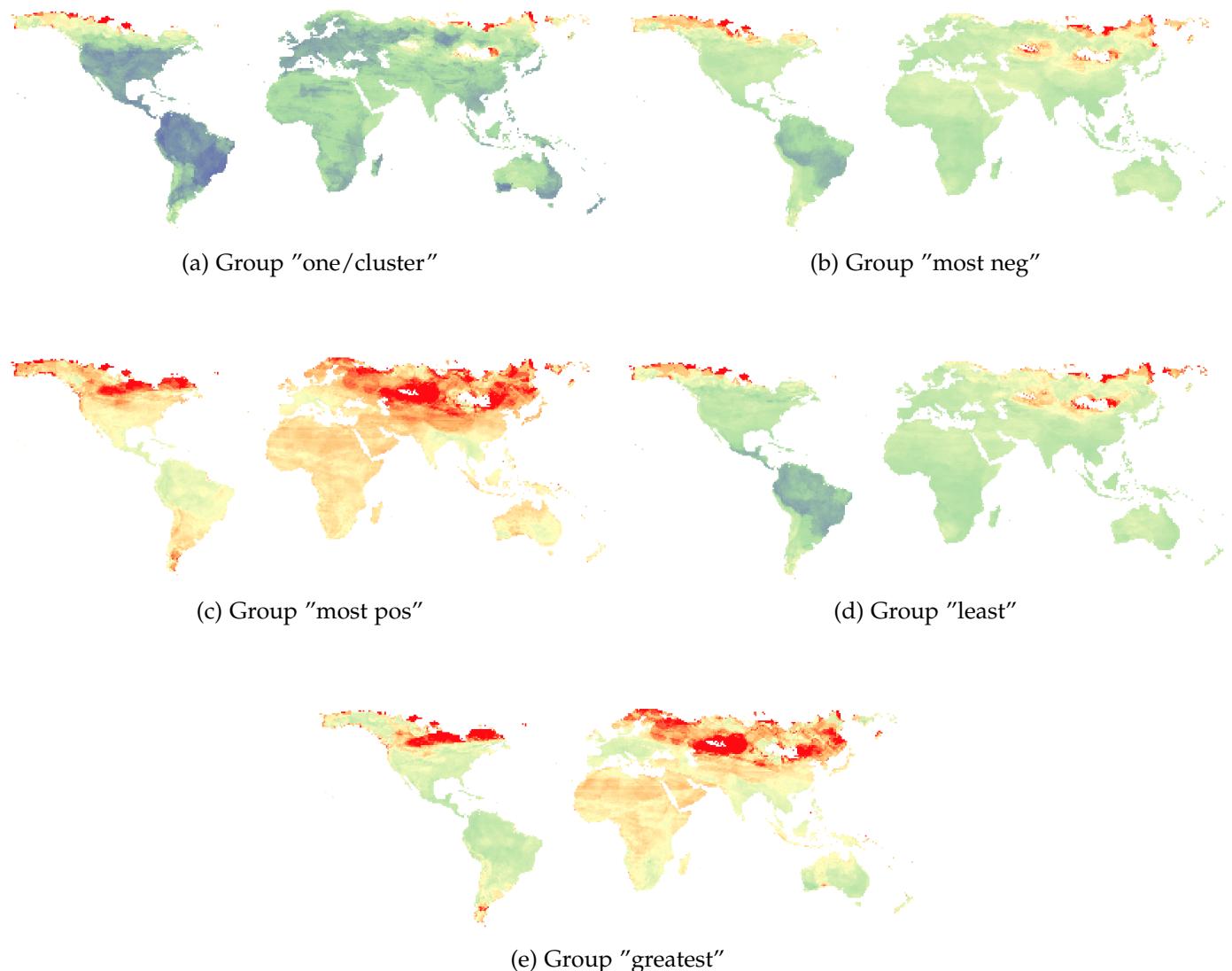


Figure 3.4: **Global mapping of the CV for FR.** We calculated the CV for each community and mapped it to the world grid. We used a heatmap and a linearly distributed color-grading for the CV values. A low CV results in bluish coloring while red represents a high CV. We clipped very high CV values in order to better show the global distribution.

3. RESULTS

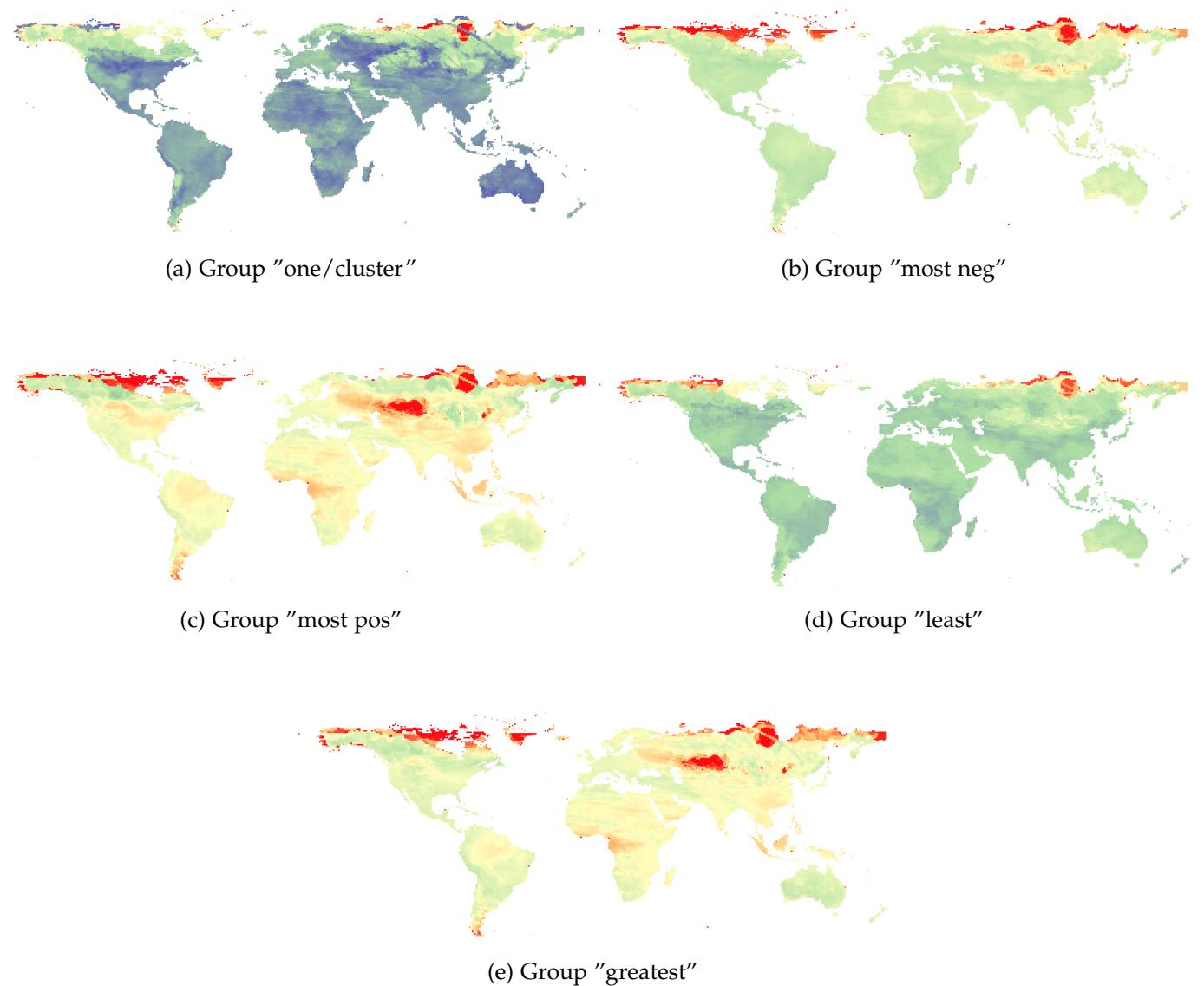
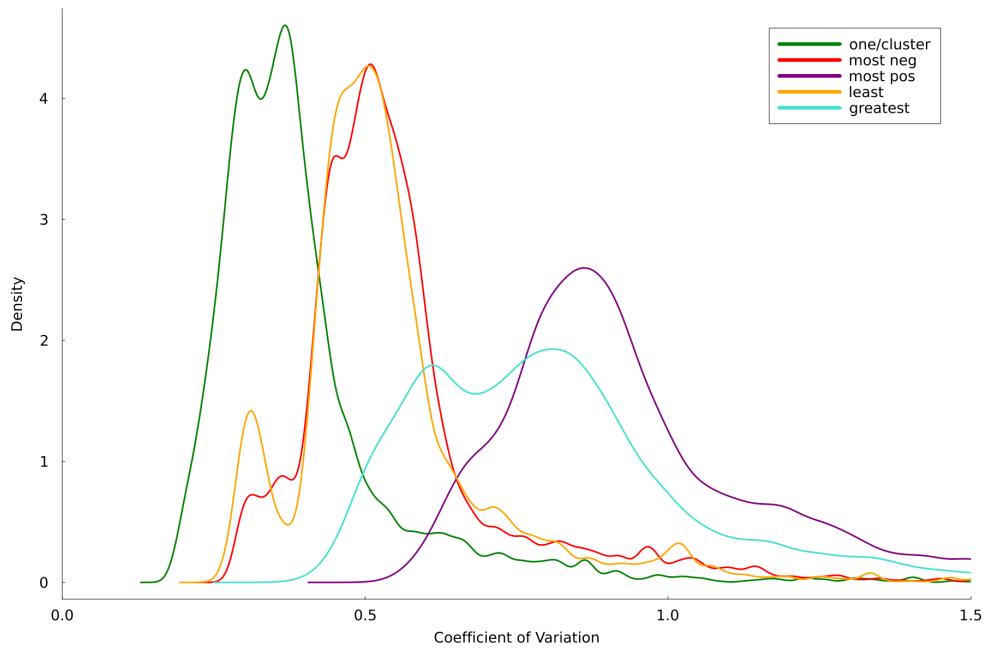
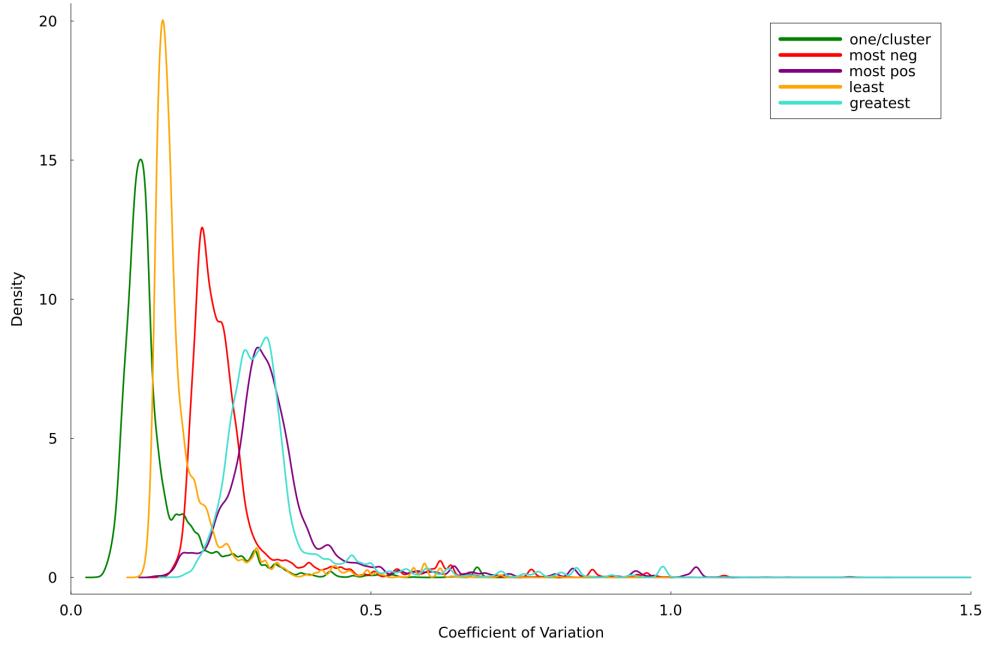


Figure 3.5: **Global mapping of the CV for FD.** For more details see Fig 3.4



(a) FR

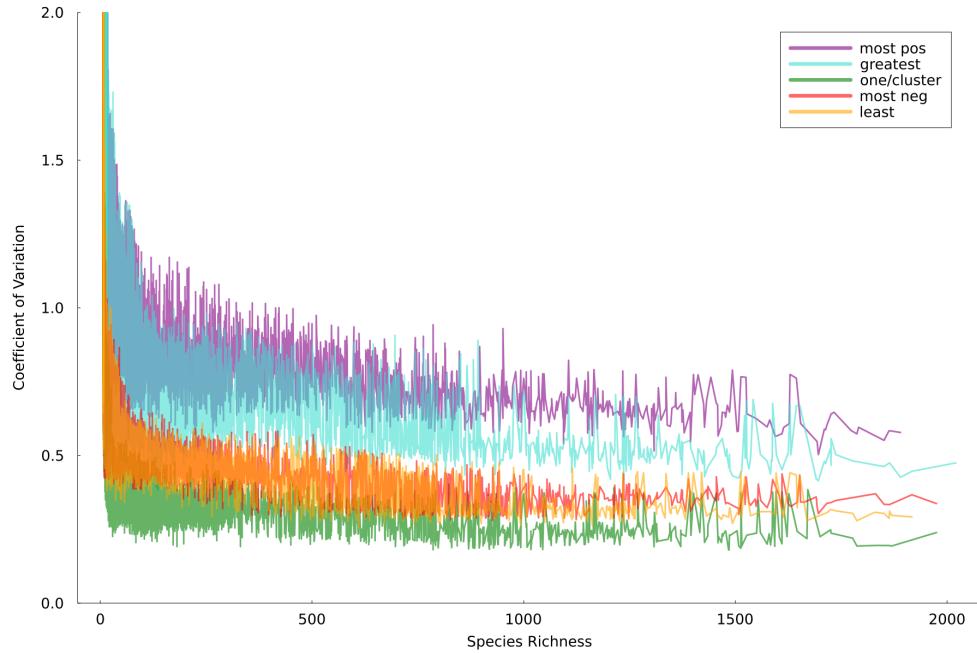


(b) FD

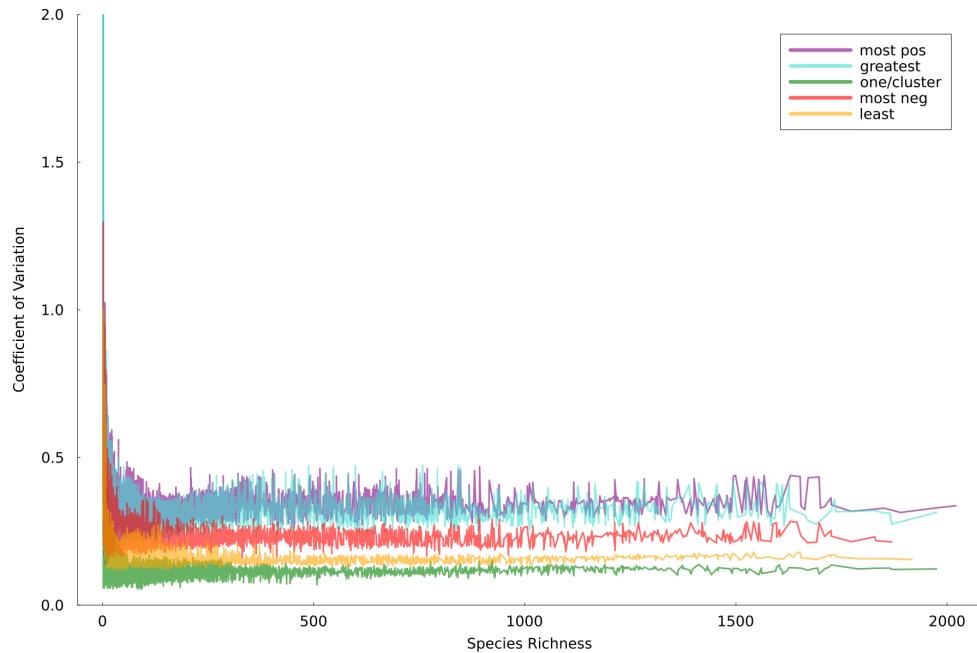
Figure 3.6: Global CV Density. Taking the same data as used for our global mapping, we displayed all global CV values as a density plot split up into trait combination groups.

3. RESULTS

Further links between community size and trait combinations groups can be made by comparing how the CV and the functional diversity values vary across richness (Figures 3.7 and 3.8. As we have seen before, the group "one/cluster" shows overall lowest CV values both for FR and FD. Overall, we see CV values for FR to vary between 25% - 100% while for FD they range from 10% - 45%. Coming back to functional diversity metrics, we see that groups "least" and "one/cluster" on average give the highest FR and FD values over all richness scales while "greatest" and "most neg" give the lowest. FR values reach between 0.017 - 0.043 while FD values range from 0.15 - 0.25 for larger communities.



(a) FR



(b) FD

Figure 3.7: CV values for varying species richness. Taking the same data as used for our global mapping, we put the CV values in relation to the species richness occurring in the corresponding grid cell. We sorted all grid cells by richness and plotted their CV values for each trait combination group.

3. RESULTS

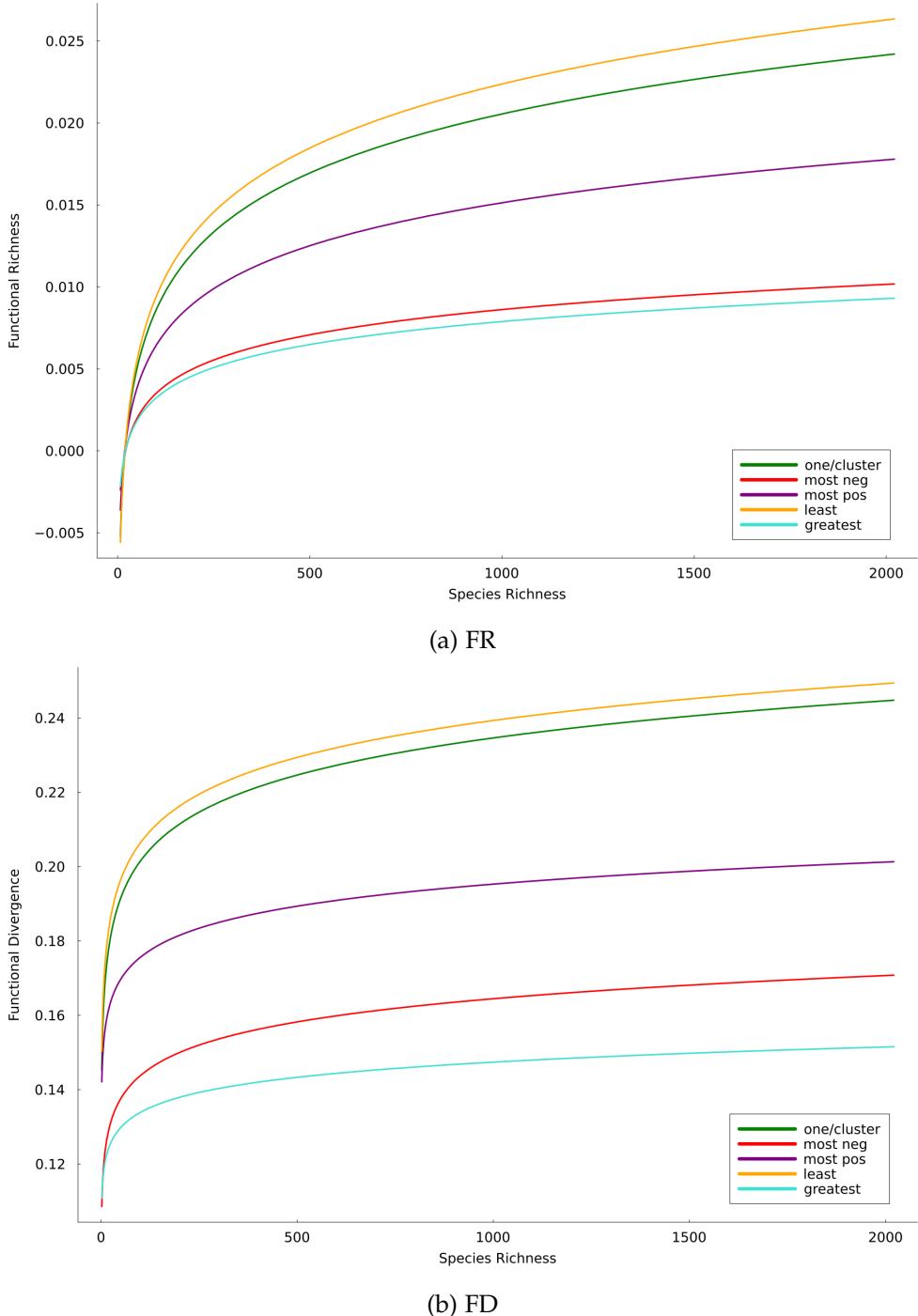


Figure 3.8: **Functional diversity values for varying species richness.** Taking functional diversity values for all grid cells and trait combination resulted in over 15 million individual points. We sorted them by species richness, split them into the trait combination groups and fit a logarithmic function to each of them to enhance visibility.

Chapter 4

Discussion

This thesis provides a general understanding of the linkage between tree functional trait selection and variation in functional diversity metrics. Having knowledge about the correlation between traits and their respective clusters, we analysed various trait combination groups which capture different correlation patterns, and compared these to the resulting variation of the diversity metrics [16]. We could show that there is a vast difference between trait groups, not only in their functional diversity metrics but also when it comes to in-group variation. We found that traits that are least correlated (in absolute value) reliably showed the lowest variance across all communities. On the contrary, high inter-trait correlation resulted in a high coefficient of variation. These results align with our expectations, as our hypothesis was that when picking less correlated traits we gain more stable results and therefore less variance between similar trait combinations. This is reflected by group "one/cluster" having the overall smallest CV, followed by "least", as both identify combinations based on minimal correlation. While somewhat unexpected, we can also observe low CV values for the group "most neg". Considering Fig. 3.1, we see that "most neg" mainly focuses on a few selected traits. Although this gives us a low CV between trait combinations, this result is skewed since it exclusively focuses on a few specific traits, which leads to similar functional diversity values. Regional differences in terms of community richness can be identified by looking at Fig. 3.4 and 3.5. In general, mid to high richness communities show lower CV compared to low richness communities. This in a sense is counter intuitive as low richness communities tend to show less variety in trait characteristics due to strong environmental filtering. Collectively, these results demonstrate that considering trait correlation is critical for the adoption of robust and consistent diversity metrics.

Our secondary hypothesis states, how we expect highly correlated traits to result in lower diversity values. For FR, this is due to the fact that non-

4. DISCUSSION

correlated traits can be viewed as sort of eigenvectors that stand orthogonal to each other and therefore span the trait space. Strongly correlated traits on the other hand will flatten the trait space and therefore result in low functional richness values. Functional Divergence, which is based on a distance matrix, behaves similarly since strongly correlated trait values tend to be closer than non correlated values producing lower FD. In general, we observe this trend (Fig. 3.8), with "least" and "most neg"/"greatest" having the largest and smallest values respectively. However, the only irregularity to this hypothesis is with the group "most pos": according to our theory, strongly correlated traits should show similarly low metrics as groups "most neg" and "greatest". This result likely reflects the fact that some traits can be defined also by their inverse (e.g., leaf mass per area = 1 / specific leaf area), such that most-negative and most-positive groupings are somewhat arbitrary and unstable. As our FR metric depends on a convex hull it is very sensitive to outliers in the trait space. If the trait space is large the influence of outliers, in relation to the size of the trait space, tends to decrease. The opposite goes for small trait spaces where a single outlier can lead to comparatively huge variation. On the other hand, our FD metric takes the average over all distances and is therefore less sensitive to outliers. It converges faster in terms of community size (Fig. 3.2b) and therefore leads to more stable CV for medium sized communities. Thus Rao's quadratic entropy gives better results in cases of under-sampling and approximations. Our results show that selecting trait grouping based on absolute values (e.g., least or one-per-cluster) provides a consistent set of functional diversity values and low coefficient of variation.

Working with vastly different traits, metrics and data certain choices had to be made regarding our methodology. The absence of abundance data for our communities put certain restrictions on the functional diversity metrics we could use. While FRic and Q are popular metrics and their representative power has been proven [10] we totally disregarded functional evenness in our calculations. Additional choices were made during pre-processing of our data. We considered normalizing over traits of species within a single community but this lead to undesired effects for our functional divergence metric as can be seen in figure A.1 in the appendix. This approach would also have taken the existence of other global species out of the equation. This in itself raises a question of desirability but the idea could be further pursued. We chose to normalize over all available species which gave us coherent results by ensuring that all traits are on the same scale. Another, more computational, restriction was due to the calculation of functional richness via the convex hull volume. The calculation time of a convex hull in n-dimensional space grows exponentially with the number of species. We decided that reducing the trait space from the original 8 dimensions down to 5 would better suit our setup and timeline, which also had the benefit of

allowing us to include related work as case studies for comparisons. Naturally it would be interesting and also necessary to conduct the same analysis with the full scaling to see if our conclusions hold or if there are new insights gained by it. Though this might be computationally challenging for convex hull volume, other functional diversity metrics should be less computationally intense but still capture changes in richness (e.g., functional attribute diversity [10]).

Collectively, this work uncovers distinct dynamics inherent to the functional trait space and its effects on functional diversity metrics. We show that highly correlated traits and weakly correlated traits capture different characteristics of communities and lead to a difference in resulting values and their stability. In practice most functional diversity studies restrict their trait selection to only a handful of traits, suggesting that approaches for selecting optimal trait groups must be relatively simple in order to be widely adopted. Thus, although more complex trait selection approaches exist, our results suggest that in practice focusing on weakly correlated traits (or ones from distinct clusters) provides a decent trade-off between simplicity and robustness of results. These insights can potentially help to standardize the set of tree traits chosen for diversity calculations and can provide future research with consistent results while we further broaden our understanding of these complex systems.

Appendix A

Appendix

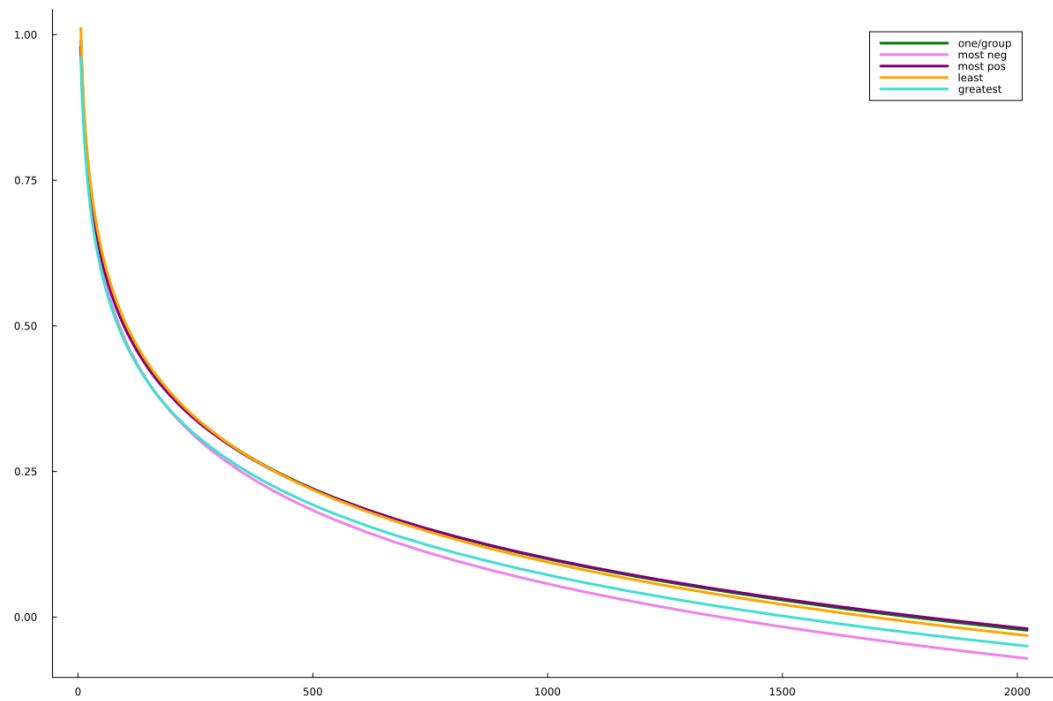


Figure A.1: Undesirable behaviour of FD due to normalization

A. APPENDIX

Listing 1 Process of trait combination selection

```
1  # trait combinations for comparison
2  trait_corr = DataFrame(CSV.File("trait_correlation_table.csv"))
3  # all possible trait combinations
4  combs = collect(combinations(1:18,5))
5  ncombs = length(combs)
6
7  # calculation of specific specs for a trait combination
8  combs_spec = zeros(ncombs, 3)
9  combs_spec[:,1] = collect(1:ncombs)
10
11 for i in 1:ncombs
12     temp = collect(combinations(combs[i],2))
13     temp2 = [trait_corr[x[1],x[2]] for x in temp]
14     combs_spec[i, 2] = sum(temp2)
15     combs_spec[i, 3] = sum(abs.(temp2))
16 end
17
18 # listing of specific trait combinations we are interested in
19 # one from each group
20 trait_comb1 = sort.(collect.(vec(collect(Iterators.product([1,
2], [3], [4, 5, 6, 7, 8, 9, 10, 11, 12], [13, 14, 15, 16], [17,
18])))))
21
22 # sorted by sum
23 combs_spec = combs_spec[sortperm(combs_spec[:, 2]), :]
24 # most negatively correlated
25 trait_comb2 = [combs[Int(x)] for x in combs_spec[1:150, 1]]
26 # most positively correlated
27 trait_comb3 = [combs[Int(x)] for x in combs_spec[end-149:end, 1]]
28
29 # sorted by abs sum
30 combs_spec = combs_spec[sortperm(combs_spec[:, 3]), :]
31 # most not correlated at all
32 trait_comb4 = [combs[Int(x)] for x in combs_spec[1:150, 1]]
33 # greatest sum of absolute values in correlation
34 trait_comb5 = [combs[Int(x)] for x in combs_spec[end-149:end, 1]]
```

Listing 2 Parallelization of our metric calculations. We gained the fastest runtime by using the pmap interface. It's native return option gathers all returns from each individually run into a single result matrix. This eliminates the need of Shared Memory arrays in our use case.

```
1   for t in collect(1:n_trait_comb)
2       @printf("Iteration %d out of %d\n", t, n_trait_comb)
3
4       # create results matrix
5       grid_res_hull = zeros(n_scal_spec+1, n_grid)
6       grid_res_rao = zeros(n_scal_spec+1, n_grid)
7
8       # in parallel run the metric calculations
9       array0fTuples = [(t,y) for y in collect(1:n_grid)]
10      out_vec = @showprogress pmap(array0fTuples) do i
11          ← calc_metrics(traits[Int.(grid_species[i[2]]), trait_comb[i[1]]],
12          ← scal_spec, n_scal_spec, i[2], normed)
13      end
14
15      # fill results matrices
16      out_mat = reduce(hcat, out_vec)
17      grid_res_hull = out_mat[1:n_scal_spec+1,:]
18      grid_res_rao = out_mat[n_scal_spec+2:end,:]
```

A. APPENDIX

Listing 3 This function sets up the final preparations for our metric calculations.

```
1  @everywhere function calc_metrics(spec_traits, scal_spec,
2      ↳ n_scal_spec, grid_idx)
3
4      n_species_total = size(spec_traits, 1)
5
6      if n_species_total > 1
7
8          # filter duplicates and add noise to remaining,
9          ↳ this helps Qhull calculations for points which
10         ↳ are almost identical
11         spec_traits_temp = unique(spec_traits,dims=1)
12         add_gauss!(spec_traits_temp, 0.0001)
13         n_species = size(spec_traits_temp,1)
14
15         # shorten scaling to #species in grid cell + exact
16         ↳ #species in grid cell
17         scal_spec_temp = filter(x -> x < n_species,
18             ↳ scal_spec)
19         push!(scal_spec_temp, n_species)
20
21         ret1 = calc_chull(scal_spec_temp, n_scal_spec,
22             ↳ spec_traits_temp, size(spec_traits, 2),
23             ↳ n_species_total, scal_spec)
24         ret2 = calc_rao(scal_spec_temp, n_scal_spec,
25             ↳ spec_traits_temp, n_species_total, scal_spec)
26         return vcat(ret1,ret2)
27     end
28
29     return zeros((n_scal_spec+1)*2)
30
31 end
```

Listing 4 Step wise scaling calculations of FRic

```
1  @everywhere function calc_chull(scal_spec_temp, n_scal_spec,
2      ↪ temp_spec_traits, n_traits, n_species_total, scal_spec)
3      # initialize the output
4      outvec = zeros(n_scal_spec+1)
5      # we need more points than dims
6      if size(temp_spec_traits, 1) > n_traits
7          n_scal_spec_temp = size(scal_spec_temp,1)
8          # check where to start since QHull sometimes
9          ↪ throws error if points are too close
10         safe = false
11         iter = findfirst(x -> x > n_traits,
12             ↪ scal_spec_temp)
13         ch = []
14         hull = []
15         # find scaling where Qhull does not error
16         while !safe && iter <= n_scal_spec_temp
17             try
18                 ch =
19                     ↪ chull(temp_spec_traits[1:scal_spec_temp[iter],
20                         ↪ :])
21                 hull =
22                     ↪ ch.points[sort(ch.vertices),:]
23                 safe = true
24                 catch e
25                     iter+=1
26                 end
27             end
28             # in case we didnt find a starting point we return
29             if !safe return outvec end
30
31             # we found a starting point
32             outvec[iter+1] = ch.volume
33             # continue with further scalings
34             for i in (iter+1):n_scal_spec_temp
35                 # take computed hull and add new points for next
36                 ↪ scaling
37                 hull = vcat(hull,
38                     ↪ temp_spec_traits[scal_spec_temp[i-1]+1:scal_spec_temp[i],
39                         ↪ :])
40                 ch = chull(hull)
41                 hull = ch.points[sort(ch.vertices),:]
42                 if i == n_scal_spec_temp
43                     outvec[1] = ch.volume
44                 else
45                     outvec[i+1] = ch.volume
46                 end
47             end
48             # filling up outvec to scale to #species before
49             ↪ preprocessing
50             for i in n_scal_spec_temp:n_scal_spec
51                 if scal_spec[i] <= n_species_total
52                     outvec[i+1] = outvec[1]
53                 end
54             end
55         end
56         return outvec
57     end
```

A. APPENDIX

Listing 5 Step wise scaling calculations of Q

```
1  @everywhere function calc_rao(scal_spec_temp, n_scal_spec,
2                                ↳ temp_spec_traits, n_species_total, scal_spec)
3
4      # initialize the results
5      outvec = zeros(n_scal_spec+1)
6      # pairwise distance matrix
7      n_scal_spec_temp = size(scal_spec_temp,1)
8      d1 = pairwise(SqEuclidean(), transpose(temp_spec_traits))
9      # calculate metric for different scalings
10     for i in 1:n_scal_spec_temp
11         if i == n_scal_spec_temp
12             outvec[1] = sum(d1[1:scal_spec_temp[i],
13                               ↳ 1:scal_spec_temp[i]])*(1/scal_spec_temp[i]^2)
14         else
15             outvec[i+1] = sum(d1[1:scal_spec_temp[i],
16                               ↳ 1:scal_spec_temp[i]])*(1/scal_spec_temp[i]^2)
17         end
18     end
19     # filling up outvec to scale to #species before preprocessing
20     for i in n_scal_spec_temp:n_scal_spec
21         if scal_spec[i] <= n_species_total
22             outvec[i+1] = outvec[1]
23         end
24     end
25     return outvec
26 end
```

Bibliography

- [1] F Mermilliod-Blondin, J Roy, E Spehn, and L van Peer. Species diversity, functional diversity, and ecosystem functioning. *Biodiversity and ecosystem functioning: synthesis and perspectives*, page 195, 2002.
- [2] Guang Hu, Yi Jin, Jinliang Liu, and Mingjian Yu. Functional diversity versus species diversity: relationships with habitat heterogeneity at multiple scales in a subtropical evergreen broad-leaved forest. *Ecological research*, 29:897–903, 2014.
- [3] Natalia Perez-Harguindeguy, Sandra Diaz, Eric Garnier, Sandra Lavorel, Hendrik Poorter, Pedro Jaureguiberry, M Syndonia Bret-Harte, William K Cornwell, Joseph M Craine, Diego E Gurvich, et al. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of botany*, 64(8):715–716, 2016.
- [4] Edwin Lebrija-Trejos, Eduardo A Pérez-García, Jorge A Meave, Frans Bongers, and Lourens Poorter. Functional traits and environmental filtering drive community assembly in a species-rich tropical system. *Ecology*, 91(2):386–398, 2010.
- [5] Jennifer L Funk, Julie E Larson, Gregory M Ames, Bradley J Butterfield, Jeannine Cavender-Bares, Jennifer Firn, Daniel C Laughlin, Ariana E Sutton-Grier, Laura Williams, and Justin Wright. Revisiting the holy grail: using plant functional traits to understand ecological processes. *Biological Reviews*, 92(2):1156–1173, 2017.
- [6] Livia Maira Orlandi Laureto, Marcus Vinicius Cianciaruso, and Diogo Soares Menezes Samia. Functional diversity: an overview of its history and applicability. *Natureza & Conservação*, 13(2):112–116, 2015.

BIBLIOGRAPHY

- [7] Nathan JB Kraft and David D Ackerly. Functional trait and phylogenetic tests of community assembly across spatial scales in an amazonian forest. *Ecological monographs*, 80(3):401–422, 2010.
- [8] Norman WH Mason, David Mouillot, William G Lee, and J Bastow Wilson. Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos*, 111(1):112–118, 2005.
- [9] G Legras, Nicolas Loiseau, and J-C Gaertner. Functional richness: Overview of indices and underlying concepts. *Acta Oecologica*, 87:34–44, 2018.
- [10] Maud A Mouchet, Sébastien Villéger, Norman WH Mason, and David Mouillot. Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Functional Ecology*, 24(4):867–876, 2010.
- [11] Legras Gaëlle and Gaertner Jean-Claude. Assessing functional evenness with the feve index: A word of warning. *Ecological Indicators*, 90:257–260, 2018.
- [12] Norman W.H. Mason and David Mouillot. Functional diversity measures. *Academic Press*, pages 597–608, 2013.
- [13] Eva Maire, Gaël Grenouillet, Sébastien Brosse, and Sébastien Villéger. 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(6):728–740, 2015.
- [14] Gaëlle Legras, Nicolas Loiseau, Jean-Claude Gaertner, Jean-Christophe Poggiale, and Nabila Gaertner-Mazouni. Assessing functional diversity: the influence of the number of the functional traits. *Theoretical Ecology*, 13(1):117–126, 2020.
- [15] Linhai Zhu, Bojie Fu, Huoxing Zhu, Cong Wang, Lei Jiao, and Ji Zhou. Trait choice profoundly affected the ecological conclusions drawn from functional diversity measures. *Scientific reports*, 7(1):1–13, 2017.
- [16] Daniel S Maynard, Lalasia Bialic-Murphy, Constantin M Zohner, Colin Averill, Johan Van den Hoogen, Haozhi Ma, Lidong Mo, Gabriel Reuben Smith, Alicia TR Acosta, Isabelle Aubin, et al. Global relationships in tree functional traits. *Nature Communications*, 13(1):3185, 2022.

Bibliography

- [17] Xiaojuan Liu, Nathan G Swenson, Jinlong Zhang, and Keping Ma. The environment and space, not phylogeny, determine trait dispersion in a subtropical forest. *Functional Ecology*, 27(1):264–272, 2013.
- [18] R Maeshiro, B Kusumoto, S Fujii, T Shiono, and Y Kubota. Using tree functional diversity to evaluate management impacts in a subtropical forest. *Ecosphere*, 4(6):1–17, 2013.
- [19] Charles A Nock, Alain Paquette, Matt Follett, David J Nowak, and Christian Messier. Effects of urbanization on tree species functional diversity in eastern north america. *Ecosystems*, 16:1487–1497, 2013.
- [20] Nathan G Swenson, Brian J Enquist, Jason Pither, Andrew J Kerkhoff, Brad Boyle, Michael D Weiser, James J Elser, William F Fagan, Jimena Forero-Montaña, Nikolaos Fyllas, et al. The biogeography and filtering of woody plant functional diversity in north and south america. *Global Ecology and Biogeography*, 21(8):798–808, 2012.
- [21] Maynard DS Paz A, Crowther TW. Functional and phylogenetic dimensions of tree biodiversity. *In prep.*, In prep.
- [22] Botanical information and ecology network. <https://bien.nceas.ucsb.edu/bien/>, accessed 2022.
- [23] A dataset of forest biomass structure for eurasia. <https://www.nature.com/articles/sdata201770#data-records>, accessed 2022.
- [24] Conifers of the world. <https://herbaria.plants.ox.ac.uk/bol/conifers>, accessed 2022.
- [25] Latin american seasonally dry tropical forest floristic network. <http://www.dryflor.info>, accessed 2022.
- [26] Global biodiversity information facility. <https://www.gbif.org>, accessed 2022.
- [27] Globaltree research: The most comprehensive database of the world’s tree species. https://tools.bgci.org/globaltree_search.php, accessed 2022.
- [28] Integrated digitized biocollections. <https://www.idigbio.org>, accessed 2022.
- [29] Tree flora of the neotropical region: a database involving biogeography, diversity and conservation. <http://www.neotrop-tree.info>, accessed 2022.

BIBLIOGRAPHY

- [30] Projecting responses of ecological diversity in changing terrestrial systems. <https://data.nhm.ac.uk/dataset/the-2016-release-of-the-predicts-database>, accessed 2022.
- [31] Rainbio: a mega-database of tropical african vascular plants distributions. <https://gdauby.github.io/rainbio/index.html#dataset>, accessed 2022.
- [32] Jens Kattge, Gerhard Bönisch, Sandra Díaz, Sandra Lavorel, Iain Colin Prentice, Paul Leadley, Susanne Tautenhahn, Gijsbert DA Werner, Tuomas Aakala, Mehdi Abedi, et al. Try plant trait database—enhanced coverage and open access. *Global change biology*, 26(1):119–188, 2020.
- [33] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98, 2017.
- [34] Zoltán Botta-Dukát and Bálint Czúcz. Testing the ability of functional diversity indices to detect trait convergence and divergence using individual-based simulation. *Methods in Ecology and Evolution*, 7(1):114–126, 2016.

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

The impact of plant functional trait choice in inference of habitat filtering and competition.

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

Zweidler

First name(s):

Jeffrey

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zurich, 28.03.23

Signature(s)



For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.