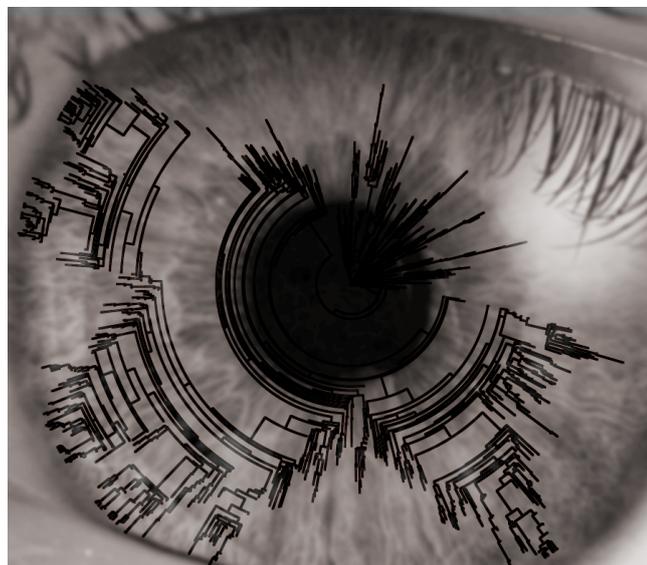


# Can we identify community assembly from phylogenies?

**A test with a biogeographic simulation for predicting phylogenetic community structure and environmental niche evolution**

Jan Paul Bauche

Master thesis (Student ID 2915784)  
submitted to  
the Faculty of Environment & Natural Resources  
at the Albert-Ludwigs-University Freiburg



Supervisor: Dr. Florian Hartig, Department of Biometry and Environmental System Analysis  
Co-supervisor: Priv.-Doz. Dr. H. Martin Schaefer, Department of Evolutionary Biology and Animal Ecology

Freiburg, December 2014

Cover photograph: "A focus on phylogenies"  
Background image by : "Petr Novák, Wikipedia"  
Idea and design by : Paul Bauche

## **Abstract**

Community assembly processes shape species communities in terms of biogeographical distribution and local coexistence. Such processes can be for example environmental niche adaptation which leads to habitat filtering or local competition resulting from inter- and intra-species similarities. Such similarities would, following the idea of limiting similarity, lead to the co-occurrence of less similar species. The question which mechanisms are responsible for the assembly of a specific community and how they can be identified is a long discussed topic in evolutionary ecology and has recently received renewed attention due to new types of data becoming available. The idea of identifying these mechanisms via the phylogenetic structure of a community has moved into focus. However it remains unclear and strongly debated how reliable inferences of community assembly processes based on phylogenetic data actually are. In this study I produce patterns of phylogenetic clustering and spatial species distribution, derived from simulations of community assembly processes to identify the resulting combinations. This allows me to assess the suitability of phylogenetic clustering or overdispersion as an indicator for community assembly processes and to show which patterns are best suited for the identification of the mechanisms which lead to the assembly of a species community. I find that phylogenetic clustering alone does not provide sufficient information to identify the underlying processes. A combination with spatial patterns such as global and local rank-abundance curves does not result in a clearer picture. I show that the combination of phylogenetic and spatial patterns of different scales is not sufficient to predict species community assembly processes.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Methods</b>	<b>7</b>
2.1	State variables and scales . . . . .	7
2.1.1	The landscape . . . . .	7
2.1.2	The individuals . . . . .	7
2.2	The processes . . . . .	9
2.2.1	Dispersal . . . . .	9
2.2.2	Density dependence . . . . .	10
2.2.3	Trait evolution . . . . .	10
2.2.4	Speciation . . . . .	11
2.3	Spatial and phylogenetic community patterns . . . . .	11
2.3.1	Phylogenetic patterns . . . . .	12
2.3.2	Spatial patterns . . . . .	12
2.4	Virtual experiment . . . . .	13
<b>3</b>	<b>Results</b>	<b>15</b>
3.1	The metacommunity . . . . .	15
3.2	The local community . . . . .	18
<b>4</b>	<b>Discussion</b>	<b>20</b>
	<b>Acknowledgments</b>	<b>22</b>

# Chapter 1

## Introduction

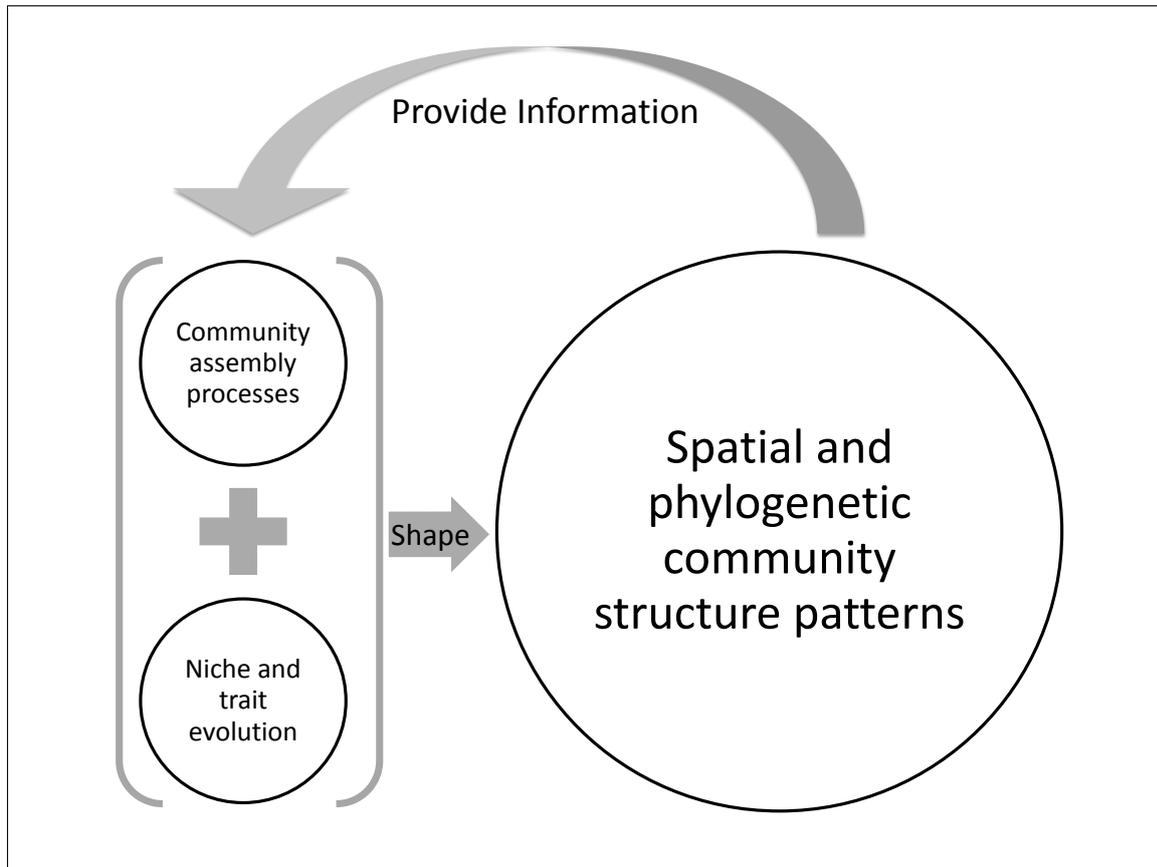
How or whether it is possible to infer the shaping process of species communities, or the other way around what is the influence of certain mechanisms on the species community patterns is a question that has received a high amount of attention for a long time in theoretic as well as applied ecology (e.g. Darwin, 1859; Hutchinson, 1959; Sugihara, 1980; Lawton and Strong, 1981). Here the focus was mostly on co-occurrence and therefore species abundance (e.g. Hyman, 1955) or community stability (e.g. MacArthur, 1955). With the focus of the scientific community on the impacts of anthropogenic climate change on ecosystems there has also been an increase of interest in the effect climate change has on the community assembly processes (e.g. Walther et al., 2002; Lavergne et al., 2010).

In ecology the stability of a community is often measured by its potential species richness, respectively the amount of species which can coexist in the equilibrium state. Or in other words the ability of species within the community to recover from small abundances (Chesson, 2000). For experimental or model cases Adler et al. (2007) showed that one could retrieve information about the stability of a species community by observing the change of the overall reproduction rate against the mean relative frequency. Another model approach concerning niche stabilization and stochasticity was conducted by Pigolotti and Cencini (2013). Münkemüller et al. (2012) and Weinstein et al. (2014) found that diversity indices can be useful in determining the processes shaping a species community. In a modeling approach concerned with the spatial patterns Chave et al. (2002) were able to show that a conspecific density dependence has an effect on the community, increasing its species richness.

The total number of species that can be present within a community depends on several factors. These factors can be for example the dispersal range of a species, the biogeography of the region or species specific factors such as a conspecific density dependence. This conspecific density dependence can also be replaced by a density dependence based on the degree of phylogenetic connectivity which leads to a more precise measure of relatedness (Webb et al., 2002).

Lately, due to an increase in availability of genetic data the question arose whether it is possible to infer the community assembly processes via the communities phylogeny or other spatial or temporal patterns (HilleRisLambers et al., 2012). This idea is visualized in figure 1.1.

Patterns of relatedness such as phylogenetic clustering and overdispersion have moved into the focus of studies concerned with species community assembly (Vamosi et al., 2009; Webb et al., 2002). The approach by Webb et al. (2002) suggests that there is a correlation between the structure of the phylogeny and the underlying biogeography and the processes that shaped the community. Due to the effect of niche-conservatism the species' traits and their resulting patterns should reflect the community assembly. The two major processes being competition, attributed to an over dispersed phylogeny since more similar individuals or species compete more strongly for the same resources, and habitat filtering, attributed to a clustered phylogeny



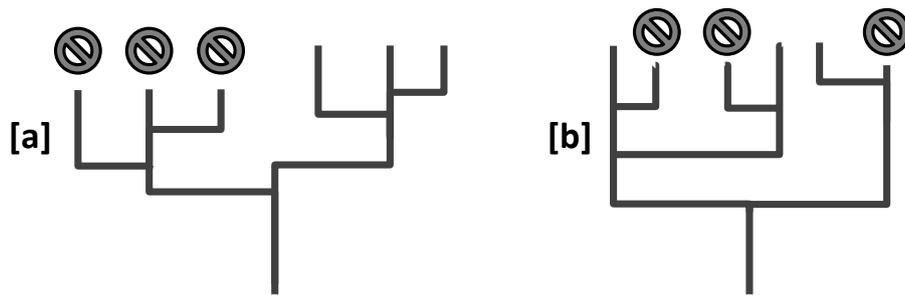
**Figure 1.1:** Concept of inference. Since trait and niche evolution in combination with the community assembly processes shape the community patterns I can try to infer these process once I record the patterns.

because the adaptation to specific environmental conditions forces similar species into similar habitats (Weiher and Keddy, 1995; Webb, 2000). Several other authors however argue that there is no direct correlation to be found between phylogenetic clustering and habitat filtering. They state that many different processes, or combinations thereof, such as dispersal or combinations of competition processes, can lead to similar patterns of community assembly (Mayfield and Levine, 2010; Kembel, 2009).

Despite the numerous papers published on this topic the question whether and to which degree one can infer community assembly processes from their resulting spatial and phylogenetic patterns is still debated. Most of the attempts at identifying community assembly processes have been made for only very specific parameters e.g. neutral conditions (Jabot and Chave, 2009). Also the mechanisms themselves are strongly simplified using for example a conspecific density dependence solely on the species level not taking relatedness of species into consideration. Chave et al. (2002) looked at the differences between local and global dispersal but only on the spatial scale.

Due to such one-sided approaches it is unclear how well the inference of species assembly processes from phylogenetic community patterns really works. Currently there are many different explanations as to how species communities are shaped based on the emerging spatial or phylogenetic patterns. Combining this information, however, may yield pattern combinations that are much more distinct than phylogenetic or spatial patterns alone.

To determine the reliability of phylogenetic data for the inference of species community as-



**Figure 1.2:** Phylogenetic clustering (a) occurs mainly due to habitat filtering while phylogenetic overdispersion (b) is attributed to competitive exclusion (Mayfield and Levine, 2010)

sembly processes, I simulate dispersal limitation, environmental preferences and local competition and produce phylogenetic and spatial patterns. In the analysis I simulate a field application by establishing local plots and analyze if the data from these plots is suited to distinguish the simulated assembly processes. I then link the resulting patterns to the biogeography and the evolutionary processes of the community as well as to dispersal limitation and density dependence.

The spatial patterns I produce from the plots and the metacommunity are species-area curves (SACs) and rank-abundance curves (RACs). The SACs display the accumulated species richness as a function of plot size or the amount of equally sized plots. They serve as an indicator for the clustering of a species community. A positively bent curve usually indicates clustering since an increase in plot size or number leads to an increase in species richness while a negatively bent curve indicates a more neutral distribution of species within the community.

RACs display the abundance of species as a function of their rank, derived from the abundance. They serve as an indicator for the amount of equally abundant species a community can support, and therefore as an indicator for the alpha diversity of a community. An S-shaped curve with a pronounced table is indicative of a more stable community since it means several species of the same abundance can be supported. A more linear curve indicates a less stable or neutral community which supports few highly abundant species and other than that consists of many species with low abundance (Chave et al., 2002). In the analysis I will link these patterns to the underlying processes to assess their usefulness as indicators for the prediction of community assembly processes.

In case of the phylogeny I am looking at patterns of overdispersion and clustering (Figure 1.2). These patterns are derived from the degree of phylogenetic relatedness and can be expressed in indices ranging from one to zero or in total values of phylogenetic branch lengths. Their difference from a chosen null model gives me information about the degree of clustering or overdispersion. Table 2.2 summarizes the expected outcomes for eight virtual experiments which will be conducted with this model. Especially interesting are the ones which are marked as unknown since they might create new unique patterns or combinations thereof or they might mimic other model outcomes and create new options for the interpretation of species community patterns.

I want to know 1.) whether phylogenetic clustering or overdispersion can be used to identify environmental filtering or respectively competitive exclusion, 2.) if dispersal limitation has an influence on the resulting community patterns which might affect the interpretation and 3.) how spatial patterns can aid the identification of community assembly processes.

Answering these questions will make it easier to choose community patterns for an analysis

of assembly processes as well as niche- and trait evolution. This will enable field- and theoretical ecologists to better choose and design their plots or respectively to increase the efficiency in the design of their experiments which will help advance the understanding of the evolution of species communities and at the same time aid conservation efforts.

# Chapter 2

## Methods

The aim of this study is the identification of species community assembly processes from local plots via the phylogenetic patterns and possibly their combination with spatial patterns. For this purpose I conduct virtual experiments in which I simulate a large biogeographical region with different assembly processes and record the resulting phylogenetic and spatial patterns from a number of local plots.

### 2.1 State variables and scales

The Model I used for simulating the desired species community assembly processes, is a spatially discrete model of a biogeographical region written in C++. The complete code can be found in the electronic supplement 1.

#### 2.1.1 The landscape

The Landscape is represented by a square grid of  $L^2$  cells with  $L = 256$ . Respectively it consists of 65,536 grid cells inhabited by one individual per grid cell at any time (no empty space). It has an air temperature gradient with values bounded between zero and one (Figure 2.1). The boundaries are warped to form a torus to avoid boundary effects.

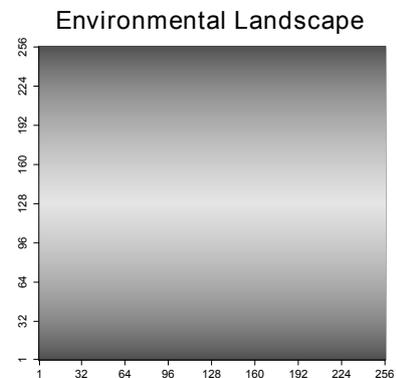
#### 2.1.2 The individuals

Each individual belongs to a species. It has two active traits bounded between one and zero. One for the interaction with other individuals deemed the competition trait which describes an individuals preferences apart from the simulated environment. The more similar the traits of two individuals are the stronger they compete with each other.

The other trait simulates an individuals interaction with the environment. It is defined as the mean of a Gaussian distribution, deemed the environmental trait.

Assuming the Gaussian function:

$$a_{\mu,\sigma}(x) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \exp \left[ \frac{-(x - \mu)^2}{2 \cdot \sigma^2} \right] \quad (2.1)$$



**Figure 2.1:** The thermal environment of the model with values between one and zero, boundaries are warped

describes the resource utilization  $a$  depending on the resource  $x$  and therefore the niche of an individual  $\mu$ ,  $\sigma$ , the exponent describes the relation of  $\sigma$  to  $\mu$ . A species can utilize the resource  $x$  best at location  $\mu$ .

If  $\mu$  does not equal  $x$  then  $\sigma$  determines the width of the niche and therefore the degree of specialization or generalization (Ackermann and Doebeli, 2004). By defining a degree of resource utilization  $u$  at a given point apart from  $\mu$  and resolving the second part of the Gaussian distribution for  $\sigma$  I can determine the desired niche width for my model.

$$\sigma = \sqrt{\frac{-(x - \mu)^2}{2 \cdot \ln(u)}} \quad (2.2)$$

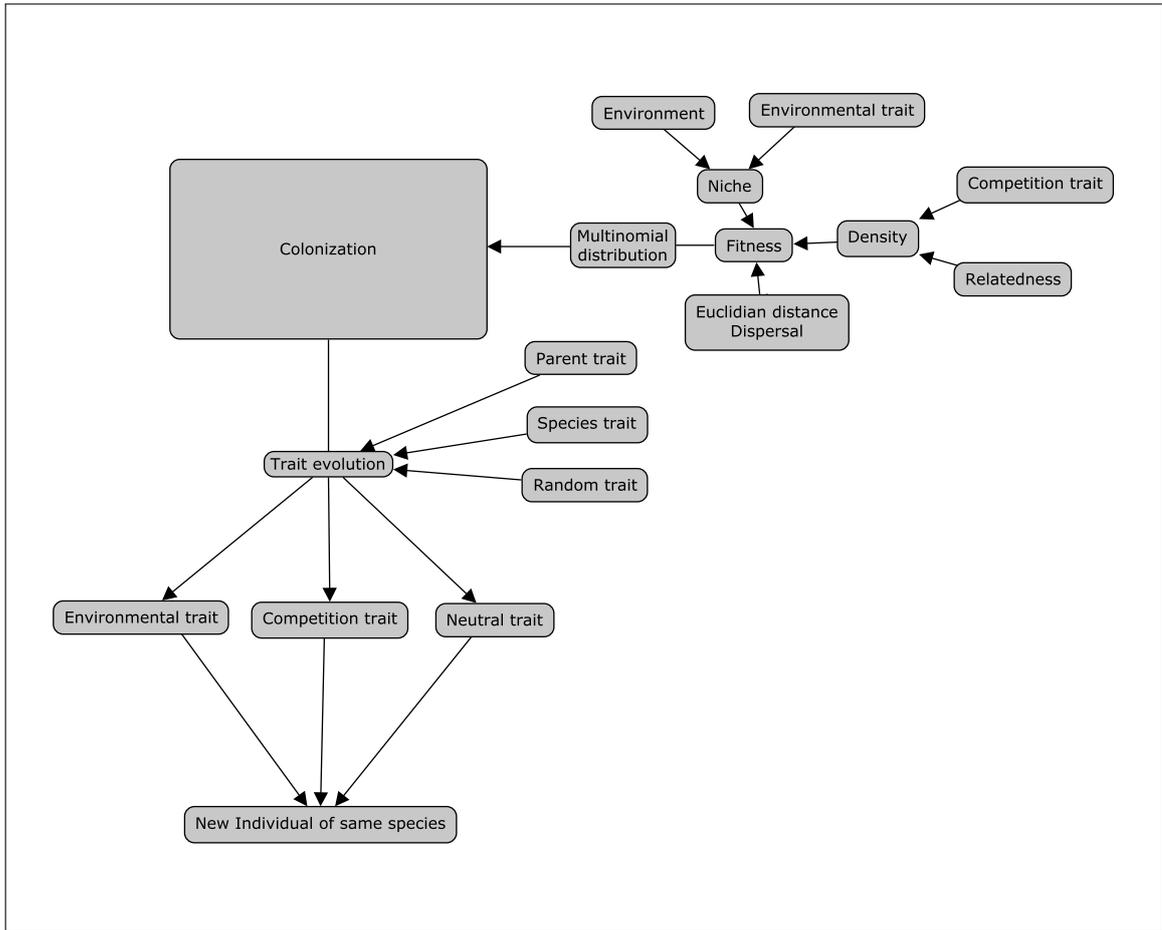
For a resource utilization of ten percent at a diversion of ten percent of the total data range from the optimum I get:

$$\begin{aligned} \sigma &= \sqrt{\frac{-(0.1)^2}{2 \cdot \ln(0.1)}} \quad (2.3) \\ &= 0.04659906 \end{aligned}$$

This value is used in the model to set the niche width of the individuals.

Like its individuals a species has trait values for the interaction with others and the environment. These traits do not evolve on their own but are instead defined as the mean over all individuals of the species. They serve as an attractor during the individual trait evolution process. Each species also contains information about its ancestral species and the species which emerged from it. This information is used to construct the global phylogeny.

## 2.2 The processes



**Figure 2.2:** Flowchart of the dispersal process as implemented in the model. The fitness of each individual is calculated by its environmental adaptation, its relatedness to its neighbors and its distance (euclidean) to the target cell.

The simulation runs over several generations. Each generation consists of two major steps. First the individuals reproduce and disperse over the landscape by colonizing cells of the grid. This dispersal is, depending on the experiment, controlled by either all or some of the following factors: Euclidean distance to the cell to be colonized, environmental adaptation, relatedness based density dependence (Figure 2.2). Then a number of individuals of new species set by the speciation rate are introduced into the community by replacing an existing individual with one of a new species. At the end of the dispersal- speciation processes, before the colonization, the traits of the new individuals are subject to evolution.

### 2.2.1 Dispersal

The dispersal process describes the attempt of the individuals in a given area to colonize an available and suited location. The individual chance for colonization is determined by the individuals fitness, which in turn depends on differing internal and external factors.

In my simulation, each generation  $L^2$  cells are sequentially chosen at random. The individual within the chosen cell dies and is replaced by the offspring of another individual. The dispersal is based on an exponential distribution kernel and can be limited in terms of its range. It is also

based on the individuals niche preference towards the implemented environmental factors and the relatedness to the individuals in its neighborhood.

The controlling mechanisms for the colonization are therefore the individuals distance to the empty cell, the local conspecific density and the environmental adaptation of the individual. These parameters lead to a certain weight. A multinomial distribution consisting of all the individuals within the range of the kernel is then used to choose the individual that eventually produces its offspring and disperses it to the now empty grid cell (Figure 2.2). The dispersal kernel is defined as:

$$w_d = \exp\frac{1}{-\Delta s/2} \quad (2.4)$$

with  $w_d$  as the weight derived from the euclidean distance  $\Delta s$  between the cell to be populated and the potential parent. This dispersal kernel leads to an exponential decline in reproductive fitness for an individual with increasing distance from the cell to be populated. For my model experiments I chose to cap the dispersal at a radius of four cells.

### 2.2.2 Density dependence

The density dependence can have a number of reasons. For example it can be the result of an increase in species specific pests (Janzen, 1970; Connell, 1971) or of competition for resources as described by Darwin (1859) apart from the defined environmental niche or as a combination of the former.

In my simulation each individual and respectively each species has a so called competition trait that defines its preference. The closer the competition trait values of two individuals are the stronger they compete with each other. Numerically this competition is resolved by calculating the sum total of all absolute competition trait differences in a certain area around the individual and including the normalized value into the multinomial distribution which determines who gets to reproduce during that event. It can be written as:

$$r_i = \frac{\sum_{j=1}^n \|(c_i - c_j)\|}{n} \quad (2.5)$$

With  $r_i$  the relatedness index for the potential parent  $i$ ,  $c$  the competition trait values for  $i$  and the respective neighbor  $j$  and  $n$  the number of individuals in the neighborhood of the potential parent. The closer  $r_i$  is to one the less related individuals are in the neighborhood of  $i$  and vice versa.

### 2.2.3 Trait evolution

The environmental preference of an individual and subsequently a species is defined via its environmental trait. During speciation and dispersal this trait is subject to evolution, in a real life setting this would be genetic recombination, which may alter the preference of the individual which in turn has an effect on the species overall niche preference. The intensity of this effect is determined by the species' abundance and the individuals degree of divergence from the species mean.

In my simulation this trait is modeled as the mean of a Gaussian function (see equation 2.1). During each dispersal event the mean is slightly changed. I assume that the traits of a new individual are influenced by the trait of the parent, a weighted random value and the attraction to the species mean (equation 2.6). The trait value of the parent has the highest influence in this case while the species and the random effect have only slight influence on the trait of the offspring. The trait of the species is determined by the arithmetic mean over the traits of all its individuals. This mean is calculated on the fly during each dispersal event. The results in an Ornstein-Uhlenbeck (OU) process for the evolution of the species traits which ensures that a species does not have a variance of traits which covers the complete trait range. In the case of speciation the species does of course have no influence on the trait value. It is instead calculated from the parental trait value and a random factor. This factor has the same weight as during the dispersal stage of the model. The trait values are calculated as follows.

During dispersal:

$$newtrait = (1 - w_s \cdot p) + (w_s \cdot s) + (w_r \cdot r) \quad (2.6)$$

During speciation:

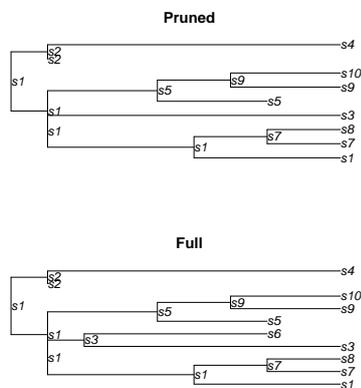
$$newtrait = (w_p \cdot p) + (w_r \cdot r) \quad (2.7)$$

With  $p$  as the trait value of the parent,  $s$  the trait value of the species,  $r$  a random trait value  $-0.5 \leq r \leq 0.5$  and  $w$  the respective weight.

## 2.2.4 Speciation

Speciation in the context of species community describes the introduction of new species into the community via rare point mutations that can become more abundant or go extinct (Hubbell, 1997). In my model speciation is controlled by the speciation rate. During each generation a fixed number of species is introduced into the community by replacing a single, randomly chosen individual with an individual of a new species. Their trait evolution is based on the trait of their ancestor and a randomly generated value (equation 2.7). This process marks the end of one generation. In the following one the newly introduced species have the chance to establish themselves or go extinct.

## 2.3 Spatial and phylogenetic community patterns



**Figure 2.3:** The Phylogeny was pruned removing all extinct leaves without branching events. See species 6 as an example in my test dataset.

The model output consist of three matrices recorded every 100 generations representing the landscape, one for the species, one for the environment which is recorded just once if there are no further changes to the environment, and one for the environmental trait of the species, a table containing all information about each species and the phylogeny recorded directly within the model in the newick format originally based on the observations by Cayley (1857). It is represented as a rooted tree with branch lengths corresponding with the difference in generations

between branching events. Since the complete phylogeny is too large to be analyzed by regular software in a practical time frame I pruned the phylogeny internally by removing all extinct leaves without any branching events (See figure 2.3). This leaves me with considerably less species in my phylogeny since many go extinct fast due to stochastic effects. This output is then imported into R and analyzed further.

### 2.3.1 Phylogenetic patterns

The phylogenetic analysis of the model output was conducted in R using the packages "picante" (Kembel et al., 2010), "phytools" (Revell, 2012), "geiger" (Harmon et al., 2008) and "phylobase" (Bolker et al., 2011). I calculated the phylogenetic species richness and clustering as well as balance. In addition I calculated the phylogenetic SACs over 100 sample plots.

To detect clustering evenness or overdispersion I calculated the standardized effect size (SES) for the phylogenetic distance (PD) (Faith, 1992) the mean pairwise phylogenetic distance (MPD) which is  $-1 \cdot NRI$  with NRI the net relatedness index and the mean nearest taxon distance (MNTD) which is  $-1 \cdot NTI$  with NTI the nearest taxon distance (for further information on NRI and NTI see Webb (2000)). When evaluating the SESs I took into account that they can be influenced by the size of the phylogeny and its general structure (Swenson, 2009).

I decided to conduct the phylogenetic analysis using the mean nearest taxon distance (MNTD). I tested two other indices. The phylogenetic distance (Faith's PD) and the mean phylogenetic distance (MPD). These are however inconsistent over my results due to a strong influence of species richness and phylogenetic structure. The results of the other metrics can be found in the electronic supplementary 2. Except for the *independent swap* which always yielded a p-value of 0.5, the tested null models I compared the calculated dispersion indices to, did not yield varying results. I sample over the extant phylogeny. To compare the overall extant phylogenies I calculated the Colless' imbalance (Colless, 1982) which is defined as:

$$C_{imb} = 2 \cdot \frac{\sum_{i=1}^{n_{nodes}} |(n_{li} - n_{ri})|}{(n - 1) \cdot (n - 2)} \quad (2.8)$$

With  $n_{li}$  the number of offspring of the left branch of the node  $i$  and  $n_{ri}$  the number of offspring of the right branch of the node  $i$  and  $n$  the total number of taxa in the tree. The index is bounded between one and zero. zero representing a completely balanced tree and one respectively representing a completely imbalanced tree.

### 2.3.2 Spatial patterns

To test for diversity indices and spatial patterns, rank-abundance curves (RACs) are calculated and the total species richness is compared to the expected species richness as calculated by Hubbell (1997). Species-area curves (SACs) are calculated for all scenarios to show the resulting differences in spatial species clustering and richness for the eight scenarios.

Plotting the species- and trait landscapes gives a visual impression of the degree of clustering and RACs for the sample plots from the whole species landscape show the localized species richness and species distribution. The combination of those patterns allows a clear distinction of spatial clustering.

**Table 2.1:** Model parameters for the scenarios. Speciation rate ( $\Theta$ ), dispersal distance ( $\Delta disp$ ), density calculation cutoff ( $\Delta dens$ ), environmental niche width( $\sigma$ ), grid size ( $L^2$ )

CASE	$\Theta$	$\Delta dens$	MODEL RUNS	$\Delta disp$	$\sigma$	$L^2$
Nearest neighbor dispersal						
Neutral	2	1	$1e + 5$	1	<i>na</i>	$256^2$
Competition	2	1	$1e + 5$	1	<i>na</i>	$256^2$
Environmental niche	2	1	$1e + 5$	1	0.0466	$256^2$
Competition and environ- mental niche	2	1	$1e + 5$	1	0.0466	$256^2$
Intermediate distance dispersal						
Neutral	2	2	$1e + 5$	4	<i>na</i>	$256^2$
Competition	2	2	$1e + 5$	4	<i>na</i>	$256^2$
Environmental niche	2	2	$1e + 5$	4	0.0466	$256^2$
Competition and environ- mental niche	2	2	$1e + 5$	4	0.0466	$256^2$

## 2.4 Virtual experiment

In order to link the different emerging patterns to the underlying mechanisms I create eight scenarios of different combinations of community assembly processes (see table 2.2). The parameters for those scenarios can be seen in table 2.1. To be able to get a representation of field conditions (Zurell et al., 2010) I additionally sample 100 plots of 100 cells each out of the metacommunity and use those to calculate the standardized effect size of the mean nearest taxon distance (MNTD) for phylogenetic clustering by calculating it for the local plots and comparing it against a null-model generated from the phylogeny pool. Additionally I calculate rank abundance curves from the local plots by plotting the abundance of each species against the rank derived from the abundance. The most abundant species receives rank one and the least abundant species is ranked at the total number of species in the plot.

These scenarios are run simultaneously and later on compared against each other in terms of the resulting patterns.

**Table 2.2:** Spatial and phylogenetic community patterns as expected from the eight experiments, RAC = rank-abundance curve, SAC = species-accumulation curve

CASE	INTERMEDIATE	NEAREST NEIGHBOR
Density Dependence	<ul style="list-style-type: none"> <li>• RAC: Negative curve</li> <li>• SAC: Negative curve</li> <li>• Phylogeny: Over dispersed</li> </ul>	<ul style="list-style-type: none"> <li>• RAC: ??</li> <li>• SAC: ??</li> <li>• Phylogeny: ??</li> </ul>
Environmental Dependence	<ul style="list-style-type: none"> <li>• RAC: S-shaped</li> <li>• SAC: Positive curve</li> <li>• Phylogeny: Clustered</li> </ul>	<ul style="list-style-type: none"> <li>• RAC: S-Shaped</li> <li>• SAC: Positive curve</li> <li>• Phylogeny: Clustered</li> </ul>
Neutral Conditions	<ul style="list-style-type: none"> <li>• RAC: Linear</li> <li>• SAC: Negative Curve</li> <li>• Phylogeny: Even</li> </ul>	<ul style="list-style-type: none"> <li>• RAC: S-shaped</li> <li>• SAC: Positive curve</li> <li>• Phylogeny: Clustered</li> </ul>
Density and Environmental Dependence	<ul style="list-style-type: none"> <li>• RAC: ??</li> <li>• SAC: ??</li> <li>• Phylogeny: ??</li> </ul>	<ul style="list-style-type: none"> <li>• RAC: ??</li> <li>• SAC: ??</li> <li>• Phylogeny: ??</li> </ul>

# Chapter 3

## Results

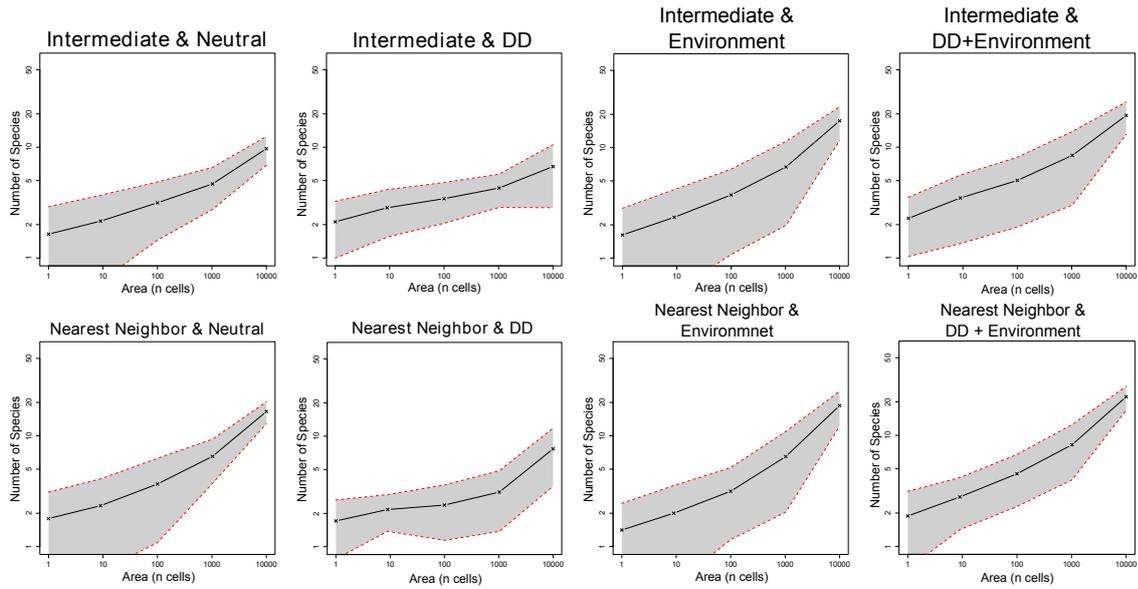
### 3.1 The metacommunity

I produce global phylogenetic and spatial patterns from the simulated assembly processes which give an impression of the metacommunity. The Colless' imbalance (Table 3.1) did not yield results for all the extant phylogenies since they were not completely dichotomous. Therefore the comparison is only possible for the neutral and environmentally dependent cases. Here I see a large difference in balance between the nearest neighbor and the intermediate dispersal scenario. The nearest neighbor dispersal yields a much more balanced phylogeny than the intermediate dispersal. For the environmentally dependent scenarios the difference is not as big and has reversed effect. Here the intermediate dispersal leads to a more balanced phylogeny. This corresponds with the effect I see in the phylogenetic SACs (Figure 3.5) where the intermediate dispersal yields a higher cumulated phylogenetic species richness than the nearest neighbor dispersal in the environmentally dependent scenario.

**Table 3.1:** Colless' imbalance for the total extant phylogeny of each scenario. 0 = fully balanced, 1 = fully imbalanced (see equation 2.8)

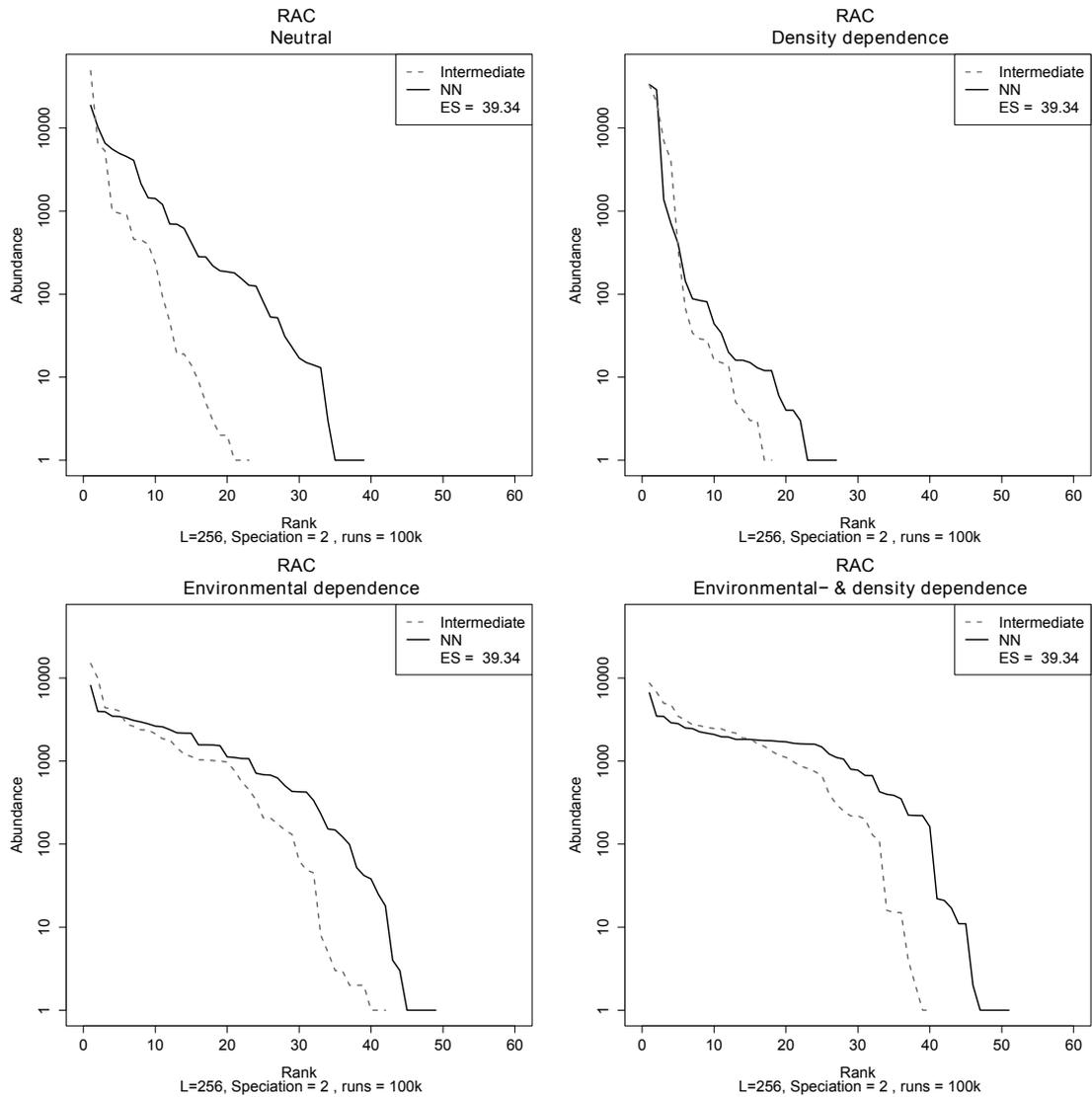
CASE	NEAREST NEIGHBOR	INTERMEDIATE
Neutral	0.122	0.541
Density dependence	<i>na</i>	<i>na</i>
Environmental niche	0.241	0.179
Density dependence and environmental niche	<i>na</i>	0.247

The species-area curves (SACs) for the intermediate and nearest neighbor dispersal experiments follow a more or less linear pattern with a slight positive curve (Figure 3.1 ). This is indicative for a clustered species landscape which can also be observed in the plotted species landscape (Figure 3.3). For the intermediate and the nearest neighbor dispersal the SACs do not offer distinguishable patterns. Only the overall accumulated species richness is increased for all environmentally dependent cases.



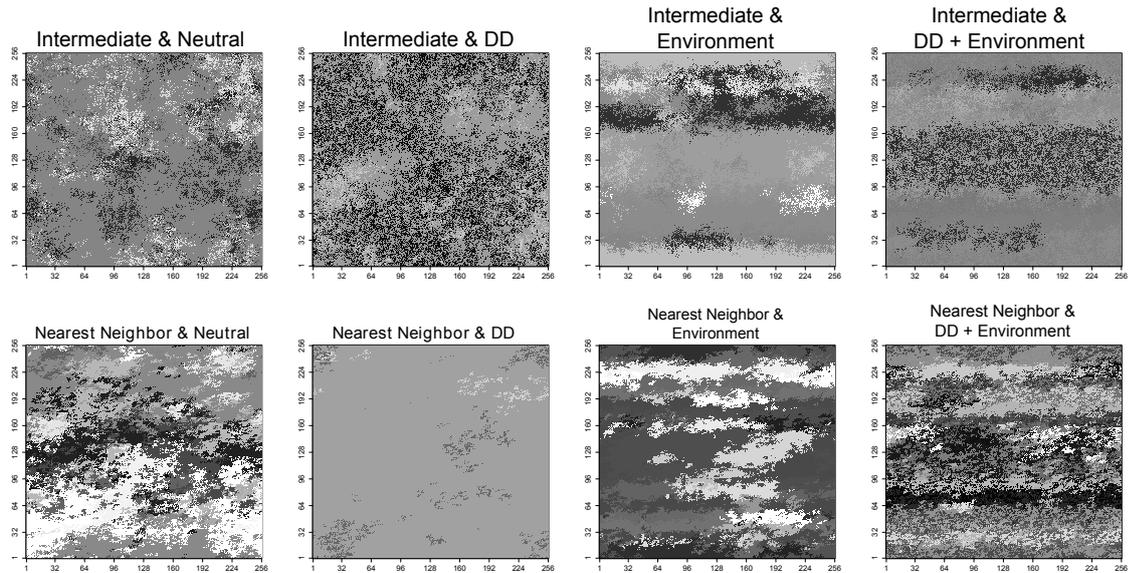
**Figure 3.1:** Species-area curves SACs for all eight experiments. Calculated from the mean out of 100 repetitions for plots of increasing size displayed on a log-log scale. All calculations were done after  $1e + 5$  model runs with a grid size of  $256^2$ .

The rank-abundance curves (RACs) differ more in terms of their patterns. While the neutral case is fairly linear the density dependent case has a negative curve the environmentally dependent as well as the environmentally- and density dependent case are positively curved with a pronounced s-shape, indicating a higher amount of similarly abundant species and, like the SACs, exhibit a higher species richness than the cases without environmental dependence. The lowest species richness occurs for the density dependent case. (Figure 3.2).

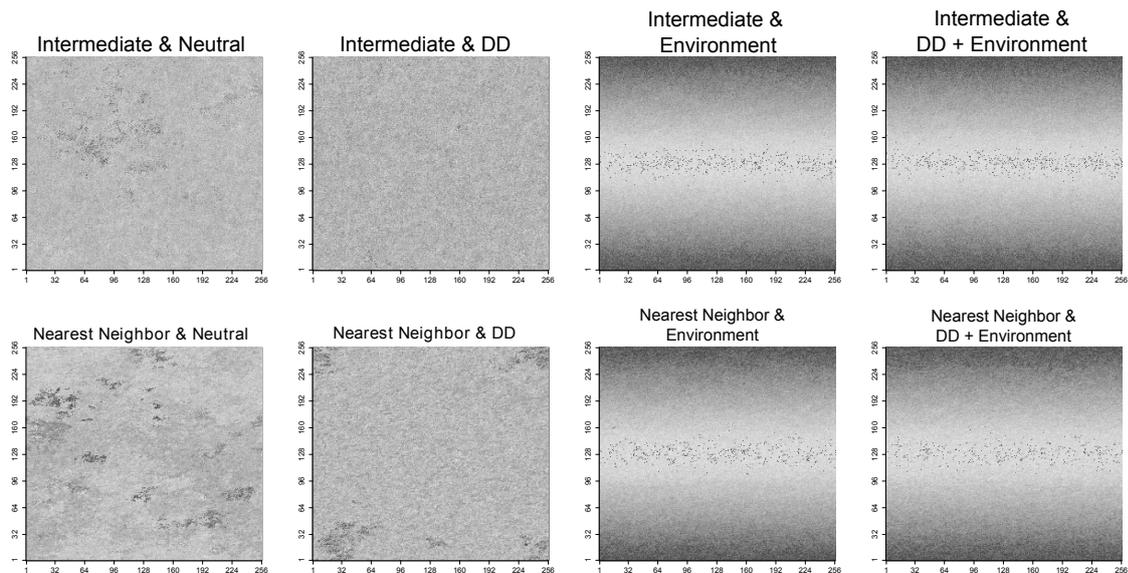


**Figure 3.2:** Global rank-abundance curves (RACs) for all eight experiments, ES = expected species richness (Hubbell, 1997), NN = nearest neighbor, displayed on a semi-log scale

In the environmentally influenced scenarios the landscape-plots of the distribution of the environmental niche trait (Figure 3.4) follow the structure of the environment (Shown in figure 2.1) while the spatial species distribution (Figure 3.3) is not as tightly bound to the environment. This is to be expected since the individuals of a species have a range of optimal niche values. In all cases of local dispersal limitation I observe an increase in spatial clustering due to the implementation of environmental dependence and an increase in dispersion due to density dependence. They are however not equally strong since the case of both dependencies displays still more spatial clustering than the neutral case.



**Figure 3.3:** Spatial distribution of the species on individual level for all eight experiments, calculated after  $1e + 5$  simulation runs on a grid of  $256^2$  cells. DD = density dependence



**Figure 3.4:** Spatial distribution of the environmental trait on individual level for all eight experiments, calculated after  $1e + 5$  simulation runs on a grid of  $256^2$  cells. DD = density dependence

### 3.2 The local community

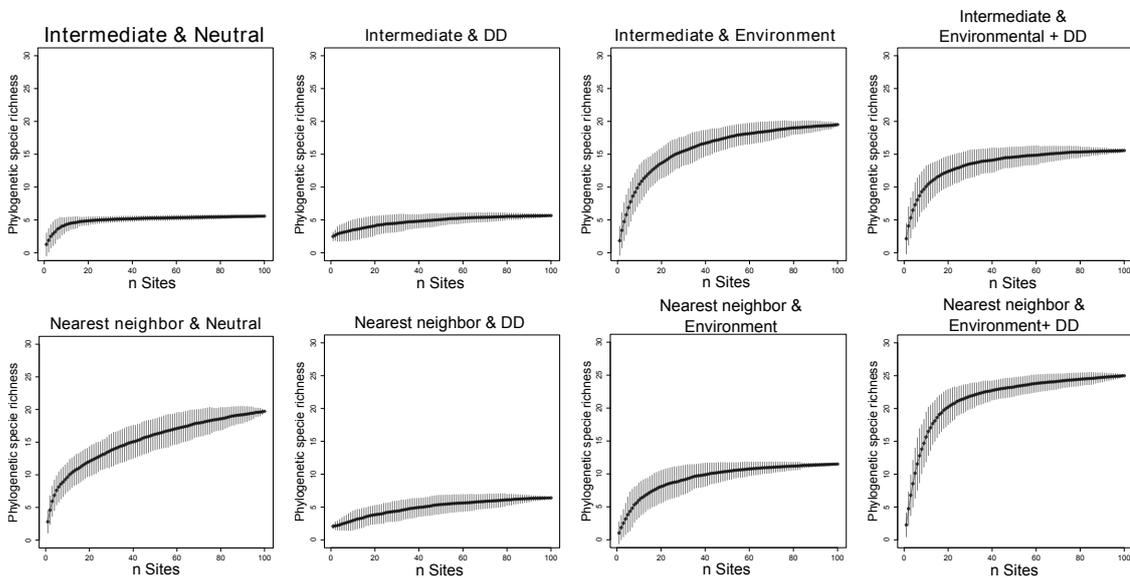
My main focus, however, was to test whether it is possible to record the whole metacommunity in a real live application. I therefore produce patterns on a local scale, as they would be recorded by field ecologists. The results are calculated from 100 randomly sampled plots out of the metacommunity, each consisting of 100 grid cells. The comparison of the average p-values of the mean nearest taxon distance (MNTD) of the different scenarios (Table 3.2) shows that the density dependence as implemented in my model leads to an overdispersion (high p-values) while the introduction of environmental dependence leads to clustering (low p-values). For a completely

neutral scenario I would expect p-values of 0.5. In the case of a limited kernel dispersal these effects are generally weaker than in the case of nearest neighbor dispersal. There is a difference in the effect in the case of the combination of density and environmental dependence. While it leads to strong clustering in the case of nearest neighbor dispersal it yields an even phylogeny for the limited kernel dispersal. The effect of density dependence is much stronger than the effect of environmental dependence in both cases.

**Table 3.2:** Average of p-values of the mean nearest taxon distance (MNTD) for 100 sample plots out of the total landscape with 100 cells each. High values indicate clustering and vice versa.

CASE	NEAREST NEIGHBOR	INTERMEDIATE
Neutral	0.5505	0.6978
Competition	0.8113	0.8762
Environmental niche	0.455	0.4172
Competition and environmental niche	0.254	0.5335

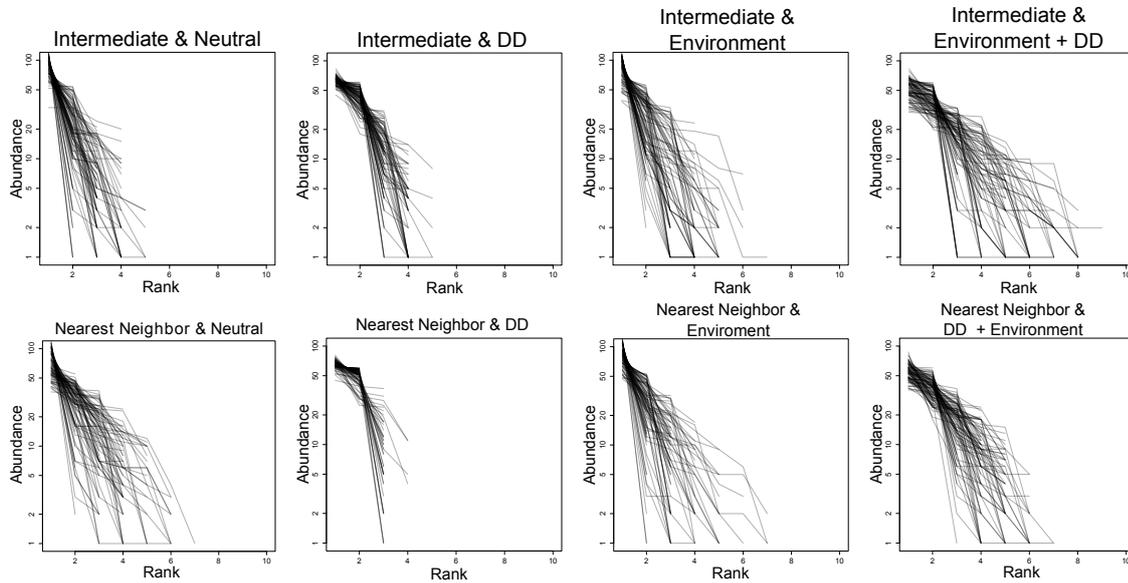
The differences of species accumulation curves (SACs) for the phylogenetic richness (Figure 3.5) follow the same pattern as the p-values of the MNTD. The only difference lies in the environmentally dependent scenarios. Here I can see a decline in phylogenetic richness in comparison to the neutral model for the nearest neighbor dispersal while there is a strong increase for the intermediate dispersal. For the combination of environmental and density dependence this effect is reversed and I see a strong increase in the phylogenetic species richness for the nearest neighbor dispersal while there is a decline for the intermediate dispersal.



**Figure 3.5:** Phylogenetic species accumulation curves (SACs) for all eight experiments. Calculated over 100 sample plots, randomly drawn from of the metacommunity, consisting of 100 cells each

The positive effect the environmental dependence has on clustering and therefore species richness can also be observed when plotting the RACs for 100 sample plots each consisting of 100 cells. These plots are randomly drawn from the whole species landscape and their individual species richness is calculated. They reflect the patterns of the RACs for the total landscape and show an increase in equally sized plots for the environmentally dependent scenarios. This means

I am able to compare field like results to answer my questions about the reliability of phylogenetic clustering indices, the influence of dispersal limitations and the value of spatial patterns in determining community assembly processes.



**Figure 3.6:** Multiple local rank-abundance curves (RACs) for all eight experiments. Calculated over 100 sample plots, randomly drawn from of the metacommunity, consisting of 100 cells each. Displayed on a semi-log scale.

## Chapter 4

# Discussion

In this study I ask whether and to what degree it is possible to infer species community assembly mechanisms and trait evolution from phylogenetic community patterns and if spatial patterns are able to provide additional insights and information towards this question. For this purpose I simulate competition, environmental preferences and dispersal limitation in a large biogeographic region and link them to patterns of phylogenetic clustering or overdispersion as well as spatial patterns such as rank-abundance curves (RACs) for local plots within this meta-community.

As expected a limited dispersal range leads to a more clustered community and this effect is intensified by the introduction of an environmental niche. The inverse can be observed when introducing a density dependence based on a competition marker.

Surprising is the fact that the neutral scenarios present an overdispersed state since I would have expected a clustered or neutral phylogeny. Also the interaction between the environmental adaptation and dispersal limitation is not intuitive. Combining environmental- and density

dependence leads to a more clustered state than neutral but more dispersed than with only environmental dependence for an intermediate dispersal distance, while it leads to even stronger clustering for a nearest neighbor dispersal distance. I was able to reproduce these results in a second model run over  $1e + 5$  generations.

The spatial patterns reflect the effect I observe in the phylogenetic clustering. While the addition of density dependence to the environmentally dependent scenario leads to more clustering for the nearest neighbor dispersal it results in a decline for the intermediate dispersal. In case of the spatial patterns the same effect occurs. While species richness increases for the nearest neighbor dispersal it declines for the intermediate distance dispersal.

In a recent study conducted on empirical data, [Chisholm et al. \(2014\)](#) found that environmental influences played a key role in community assembly processes and dynamics in some cases but were unable to establish the driving factors for all cases. Meanwhile [Comita et al. \(2010\)](#) observed a strong effect of conspecific density dependence on species richness with an asymmetry leaning towards species with low abundance. In a study on palm trees in the Neotropics, [Eiserhardt et al. \(2013\)](#) identified a joint effect of dispersal limitation and environmental niche adaptation on community assembly.

This indicates that, when observing and analyzing species community patterns, it is important not to focus on a single pattern but to look at their combinations if one wants to infer the underlying community assembly processes, as it was attempted for a semi neutral state, with a non-neutral metacommunity and neutral local communities, by [Jabot et al. \(2008\)](#). Here they found a combined effect of dispersal limitation and environmental filtering on the phylogenetic structure of a species community. This is especially important for field data since the most common case is a community in which environmental niche adaptation, dispersal limitation and competitive mechanisms co-occur.

However, the non-linearity in the interaction of the environmental dependence and dispersal limitation I observed in my simulation implies that interpreting these patterns without a reference or prior knowledge about the species traits will not yield conclusive results. Since this kind of reference is not available in a field application there is need for a pattern or a combination of patterns that does not require a comparison but stands for itself.

Despite the advances of my findings I have to take into account that I am still lacking detail in some cases. In terms of trait adaptation my model skips the individual adaptation process and is concerned only with the evolutionary trait adaptation. Even though this may be the most crucial part of trait adaptation ([Mayr, 1982](#)) it is only one of many processes contributing to phenotypic trait adaptation ([Ackerly, 2003](#)). Simulating this process in more detail might yield a clearer picture of individual trait adaptation. In retrospect the density dependence is very strong in my model, severely reducing species richness, and might require a controlling factor to limit its influence. It has an even higher impact on the nearest neighbor dispersal due to the fact that the radius for density calculation is supposed to be smaller than the dispersal range which in case of the nearest neighbor dispersal can not be achieved since it already is the lowest possible range. In their paper [Chave et al. \(2002\)](#) found that the introduction of a conspecific density dependence lead to an increase in species richness. Their density was however based on species identity while, in my study, I base it on an interspecies competition trait as suggested by [Webb et al. \(2002\)](#).

Adding a temporal dimension, such as a stability measure as introduced by [Adler et al. \(2007\)](#), to the observations could increase the precision and might allow a better distinction of as of yet diffuse patterns. To be able to infer community patterns of mobile species it would be interesting to include movement patterns sexual reproduction into the simulation. This would open ways to also look at the evolution of movement of individuals or groups e.g. herds. Unfortunately for the case of this master thesis there is a time limit which restricts the model dimensions and the

number of possible repetitions and variations due to computation time.

In conclusion I find 1.) that it is not possible to simply use patterns of phylogenetic clustering or overdispersion to precisely determine whether the assembly processes were dominated by competitive exclusion or environmental filtering because 2.) dispersal limitation has a strong, and non-linear influence on the patterns resulting from community assembly processes and has to be taken into consideration when interpreting them and 3.) while spatial patterns help distinguishing community assembly in a model scenario they are not sufficient in a field application. The resulting implication is that it is not possible to identify community assembly from the recorded patterns in a field application, since the interaction between dispersal limitation and environmental dependence is not a linear one. Model studies concerned with species community assembly processes should not focus on patterns of a single scale but rather take advantage of the scales available. That is what I achieve with my approach at identifying assembly mechanisms from combined community patterns.

## Acknowledgments

I want express my gratitude to all those who supported me throughout my studies and my thesis. Thanks to my supervisor Dr. Florian Hartig for his support during my master thesis. Thanks to the members of the department of "Biometry and Environmental System Analysis" who were very helpful in developing and focusing the hypothesis through in several discussions. Thanks to Joachim Maack and Alexander Niederbühl for proofreading my thesis. Thanks to my parents for their continuous moral and financial support throughout my studies. And finally I want to thank my girlfriend Lotta Kiesow, simply for always being there for me.

## Bibliography

- Ackerly, D. D., 2003. Community Assembly, Niche Conservatism, and Adaptive Evolution in Changing Environments. *International Journal of Plant Sciences* 164 (S3), pp. S165–S184.
- Ackermann, M., Doebeli, M., 2004. Evolution of Niche Width and Adaptive Diversification. *Evolution* 58 (12), 2599–2612.
- Adler, P. B., HilleRisLambers, J., Levine, J. M., 2007. A niche for neutrality. *Ecology letters* 10 (2), 95–104.
- Bolker, B., Butler, M., Cowan, P., de Vienne, D., Eddelbuettel, D., Holder, M., Jombart, T. and Kembel, S., Michonneau, F., Orme, D., O'Meara, B., Paradis, E., Regetz, J., Zwickl, D., 2011. phylobase: Base Package for Phylogenetic Structures and Comparative Data. R Package. Version 0.6.3. <http://cran.r-project.org/package=phylobase>.

- Cayley, A., 1857. XXVIII. On the theory of the analytical forms called trees. *Philosophical Magazine Series 4* 13 (85), 172–176.
- Chave, J., Muller-Landau, H. C., Levin, S. A., 2002. Comparing Classical Community Models: Theoretical Consequences for Patterns of Diversity. *The American Naturalist* 159 (1), pp. 1–23.
- Chesson, P., 2000. Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics* 31 (1), 343–366.
- Chisholm, R. A., Condit, R., Rahman, K. A., Baker, P. J., Bunyavejchewin, S., Chen, Y.-Y., Chuyong, G., Dattaraja, H., Davies, S., Ewango, C. E., et al., 2014. Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecology Letters*.
- Colless, D. H., 1982. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. by E. O. Wiley, Review by: Donald H. Colless. *Systematic Zoology* 31 (1), pp. 100–104.
- Comita, L. S., Muller-Landau, H. C., Aguilar, S., Hubbell, S. P., 2010. Asymmetric density dependence shapes species abundances in a tropical tree community. *Science* 329 (5989), 330–332.
- Connell, J. H., 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. *Dynamics of populations* 298, 312.
- Darwin, C., 1859. *On the origins of species by means of natural selection*. London: Murray.
- Eiserhardt, W. L., Svenning, J.-C., Baker, W. J., Couvreur, T. L., Balslev, H., 2013. Dispersal and niche evolution jointly shape the geographic turnover of phylogenetic clades across continents. *Scientific reports* 3.
- Faith, D. P., 1992. Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61 (1), 1 – 10.
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., Challenger, W., 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24 (1), 129–131.
- HilleRisLambers, J., Adler, P., Harpole, W., Levine, J., Mayfield, M., 2012. Rethinking community assembly through the lens of coexistence theory. *Annual Review of Ecology, Evolution, and Systematics* 43 (1), 227.
- Hubbell, S. P., 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* 16 (1), S9–S21.
- Hutchinson, G. E., 1959. Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *The American Naturalist* 93 (870), pp. 145–159.
- Hyman, L. H., 1955. How Many Species? *Systematic Zoology* 4 (3), pp. 142–143.
- Jabot, F., Chave, J., 2009. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecology Letters* 12 (3), 239–248.
- Jabot, F., Etienne, R. S., Chave, J., 2008. Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* 117 (9), 1308–1320.
- Janzen, D. H., 1970. Herbivores and the Number of Tree Species in Tropical Forests. *The American Naturalist* 104 (940), pp. 501–528.

- Kembel, S. W., 2009. Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters* 12 (9), 949–960.
- Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., Webb, C. O., 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26 (11), 1463–1464.
- Lavergne, S., Mouquet, N., Thuiller, W., Ronce, O., 2010. Biodiversity and Climate Change: Integrating Evolutionary and Ecological Responses of Species and Communities. *Annual Review of Ecology, Evolution, and Systematics* 41 (1), 321–350.
- Lawton, J. H., Strong, D. R., J., 1981. Community Patterns and Competition in Folivorous Insects. *The American Naturalist* 118 (3), pp. 317–338.
- MacArthur, R., 1955. Fluctuations of Animal Populations and a Measure of Community Stability. *Ecology* 36 (3), pp. 533–536.
- Mayfield, M. M., Levine, J. M., 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology letters* 13 (9), 1085–1093.
- Mayr, E., 1982. *The growth of biological thought: diversity, evolution, and inheritance*. Harvard University Press.
- Münkemüller, T., de Bello, F., Meynard, C. N., Gravel, D., Lavergne, S., Moullot, D., Mouquet, N., Thuiller, W., 2012. From diversity indices to community assembly processes: a test with simulated data. *Ecography* 35 (5), 468–480.
- Pigolotti, S., Cencini, M., 2013. Species abundances and lifetimes: From neutral to niche-stabilized communities. *Journal of theoretical biology* 338, 1–8.
- Revell, L. J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3 (2), 217–223.
- Sugihara, G., 1980. Minimal Community Structure: An Explanation of Species Abundance Patterns. *The American Naturalist* 116 (6), pp. 770–787.
- Swenson, N. G., 2009. Phylogenetic Resolution and Quantifying the Phylogenetic Diversity and Dispersion of Communities. *PLoS ONE* 4 (2), e4390.
- Vamosi, S. M., Heard, S. B., Vamosi, J. C., Webb, C. O., C. O., 2009. Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology* 18 (4), 572–592.
- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T. J. C., Fromentin, J.-M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. *Nature* 416 (6879), 389–395.
- Webb, C. O., 2000. Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *The American Naturalist* 156 (2), pp. 145–155.
- Webb, C. O., Ackerly, D. D., McPeck, M. A., Donoghue, M. J., 2002. Phylogenies and community ecology. *Annual review of ecology and systematics*, 475–505.
- Weihner, E., Keddy, P. A., 1995. The Assembly of Experimental Wetland Plant Communities. *Oikos* 73 (3), pp. 323–335.

- Weinstein, B. G., Tinoco, B., Parra, J. L., Brown, L. M., McGuire, J. A., Stiles, F. G., Graham, C. H., 2014. Taxonomic, Phylogenetic, and Trait Beta Diversity in South American Hummingbirds. *The American Naturalist* 184 (2), pp. 211–224.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., T., Nehrbaß, N., Pagel, J., Reineking, B., Schröder, B., Grimm, V., 2010. The virtual ecologist approach: simulating data and observers. *Oikos* 119 (4), 622–635.

# Selbstständigkeitserklärung

## Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht.

(I hereby declare that I have composed this document unassistedly and that I only used the sources and devices I declared. Passages taken verbatim or in meaning from other sources are identified as such and the sources are acknowledged and cited.)

Freiburg, December 2014