

University of Groningen

What lies beneath?

Janzen, Thijs

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Janzen, T. (2015). *What lies beneath? How patterns in ecology and evolution inform us about underlying processes.* [S.n.].

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Introduction

Thijs Janzen

Introduction

Some biological processes span across scales of space and time and that makes these processes difficult to study. Whereas some processes might be manipulated directly and can be studied in the laboratory, other processes elude this way of investigation, simply by their sheer size or by the timescale they operate on. Evolution operates over timescales that make it impossible to directly observe any large evolutionary changes within a lifetime, and processes that influence species composition within an ecosystem operate within a spatial scale hard to manipulate experimentally. Instead, we can look at the outcome of evolution and community composition, and study patterns caused by underlying processes. For example, traces of evolutionary history can be found in comparing molecular similarity between species, whilst the distribution of traits in a community reveals information about the strength of competition.

It is these traces that I will try to use in order to reverse-engineer the underlying processes. In this thesis I will use modeling techniques to reconstruct these hidden processes and use models to infer the underlying processes driving the origin and maintenance of biodiversity. The thesis consists of two general parts, firstly I will focus on inference in ecology, centered around the maintenance of biodiversity: why do so many species coexist? Why do we find multiple fish species in a lake? Why doesn't a single species overtake the lake and dominate? Secondly I will focus on inference in evolution, centering around the origin of biodiversity: how fast does a new species evolve? Can we link changes in the environment to the evolution of new species? Do rates of speciation slow down over time? What causes such changes?

Inference in Ecology

Understanding how biodiversity is maintained is one of the most important questions in ecology, and it is crucial in understanding how to preserve biodiversity hotspots for future generations. Classic coexistence theory explains the coexistence of species through the overlap of niches (Diamond 1975; Tilman 1981). When two species occupy the same niche, one of them will turn out to be the superior competitor and drive the other to extinction, famously formulated in the competitive exclusion principle (Gause 1934). Niche differences on the contrary allow room for coexistence and niche differences that cause species to more limit themselves than other species, can even promote coexistence (Chesson 2000). Information about species' niches and the differences between these niches can thus inform us to a large extent about potential for coexistence and the maintenance of diversity. A species' niche is however not easily measured, and a good starting point could therefore be assessing a species' traits instead. If a species occupies a specific niche, chances are that it has developed specific traits to exploit its niche. Adaptations to drought might for example suggest that a species' niche includes terrain where water is not readily available. Large leaf area might indicate that a species is typically found in low light conditions, and tries to cope through optimizing its light intake. Trait-based community assembly specifically focuses on the link

between species' traits, their niches and assembly of the community. Community assembly being the process by which species from a regional pool colonize and interact to form local communities (HilleRisLambers *et al.* 2012). If we consider all traits of all species in a community, could we infer community assembly processes from patterns in the distribution of traits? There are two trait-based processes that could leave a profound mark on trait distributions: habitat filtering and limiting similarity.

Habitat filtering occurs when the habitat poses restrictions on the traits of species. The local habitat in a desert for instance selects for traits which promote water retention, whilst species with traits adapted to more moist situations are selected against. As a result, if the habitat has an effect on the trait distribution we expect the trait distribution to be relatively narrow, compared to a trait distribution upon which the habitat has no limiting effects. In a broadleaf evergreen forest in California, USA it was found that trait ranges within sampled sites were more restricted than the overall trait range across all sampled sites (Cornwell *et al.* 2006; Cornwell & Ackerly 2009), suggesting some limiting factors at these sampling sites. Similar patterns, where trait distributions within sites were narrower than trait distributions across sites, were uncovered for a tropical rainforest located in the Ecuadorian Amazon (Kraft *et al.* 2008), for a Neotropical dry-forest, located in Costa Rica (Swenson & Enquist 2009) and for Oak species in Florida (Cavender-Bares *et al.* 2004). For all these tree datasets environmental circumstances at the sites seem to have limited the range of suitable traits, compared to the trait distribution observed across all sites sampled.

Limiting similarity has an exactly opposing effect on trait distributions: it tends to distribute traits more evenly, and increase trait range. Limiting similarity might drive traits of species apart as a result of competition for the same resource, because of shared predation pressures or because of shared susceptibility to pathogens. Five species of Warblers from a forest in North America were shown to be strongly segregated regarding nesting date, nesting height, and feeding position in the canopy (MacArthur 1958). These species actively tried to partition the habitat such that they had minimal overlap, and consequently their traits appeared evenly spaced. Inga trees in Peru and Panama often suffer from considerable herbivore pressure, and defend themselves using an array of anti-herbivory traits including chemical defense, delayed development of young (nutritious) leaves and attraction of leaf-defending ant species (Kursar *et al.* 2009). Species within local sites showed to be more dissimilar in defense strategy than species across sites, whereas traits unrelated to herbivore defense mechanisms showed no discrepancy in distribution between within or across site comparisons. The high dissimilarity within sites seems to indicate that species are selected against having too similar defense strategies, as if to avoid sharing herbivory pressure.

Contrasting trait-based community assembly is the Unified Neutral Theory of Biodiversity, which advocates a strong role for stochasticity (Hubbell 2001). Hubbell argues that although niches are an important part of community assembly, communities could also be viewed as a set of species that is constantly changing due to the influx of migrants, changes in the species pool due to speciation and extinction, and the

turnover generated by the death of individuals and their replacement by newly born individuals. These stochastic events result in a dynamic equilibrium, without assuming any trait-based effects. The Neutral Theory is best applied as a null-expectation: how would a community look like if no trait-based processes would influence its composition? Any remaining patterns can then be attributed to possible trait-based processes. For some communities, the Neutral Model is indeed able to fully explain the observed community composition, suggesting that no other processes influence its composition. Examples of such communities include tropical tree communities in Panama (Hubbell 2001; Volkov *et al.* 2003; Etienne & Olff 2005), bird communities in North America (Mcgill 2003; He 2005), aquatic invertebrate communities in rock pools, Jamaica (Fuller *et al.* 2005), Neotropical freshwater fish communities in Venezuela (Etienne & Olff 2005), aphid communities in the United Kingdom (He 2005) and bacterial communities in tree holes (Woodcock *et al.* 2007).

However, a large number of communities does not perfectly fit the neutral model, and leaves room for trait-based processes. A quick glance reveals coral communities (Dornelas *et al.* 2006), benthic intertidal communities (Wootton 2005), a whole range of marine communities, including mollusk, reef fish, invertebrate, zooplankton and crustacean communities (Connolly *et al.* 2014) and a large set of tropical tree communities, including communities from Ecuador, Colombia, Panama, Cameroon, Malaysia and Thailand (Ricklefs & Renner 2012). For these communities, clearly community composition is not determined solely by stochastic effects, but also influenced by other processes, including trait-based community assembly.

In summary, community assembly can be seen as a combination of both niche-effects exerted by the environment such as habitat filtering, interaction effects such as limiting similarity, and stochastic effects such as dispersal limitation and stochastic turnover. The question remains to what extent each process is responsible for community composition: which process is most important? Does the importance of a process depend on the environment? How does the contribution of each of those processes change over an environmental gradient?

The first part of my thesis revolves around these questions, where I will introduce a new trait-based community assembly model. This new model takes into account stochastic, environmental and between-species effects in order to shed light on community assembly and assess the relative importance of these processes. Furthermore, I will expand the standard Neutral Model by taking into account dispersal syndromes.

Inference in Evolution

The second part of my thesis revolves around the origin of biodiversity: how does biodiversity arise and how do new species originate? Evolution, the change in the characteristics of a population of organisms over successive generations, can generate new species and provides the necessary new species that are at the basis of biodiversity. Research on evolution has come a long way, from inferences from breeding programs and fossils (Darwin 1859) to the reconstruction of the tree of life using molecular tech-

niques (Jones *et al.* 2002; Bininda-Emonds *et al.* 2007; Jetz *et al.* 2012). A tree of life based on genes, a phylogenetic tree, depicts evolution as an ever branching process, where the leaves of the tree represent extant species and the root represents the common ancestor. Using molecular data restricts the analysis to reconstructing the past only for extant species, and phylogenetic trees typically do not include extinctions. However, we know from the fossil record that not all species persist indefinitely: species every now and then go extinct. A high level of extinctions can however leave its mark on a phylogeny. Typically, phylogenies with high levels of extinction show a “pull of the present”, where those species that originated close to present time haven’t had the opportunity yet to have gone extinct, causing an upturn in branching events close to the present time (Nee *et al.* 1994; Etienne & Rosindell 2012). Also, a time-period with an apparent lack of branching events could point towards past extinction events. It is this type of inference that I focus on in this part: looking for patterns in phylogenies in order to reconstruct “hidden” events that have influenced the evolutionary history of a group of species.

An important starting point for the inference of speciation and extinction rates was the mathematical formalization of the most basic speciation and extinction model: the birth-death model (Nee *et al.* 1994). The birth-death model starts a tree with one species, and this species has a probability of speciation, which adds a new species to the species pool (“birth”) or has a probability of going extinct (“death”). Newly born species are added to the tree, and also have a probability of speciation or extinction themselves. Larger birth rates result in larger trees, and larger death-rates result in more profound patterns of extinction in the tree. Given a phylogenetic tree, we can try to estimate the birth and death rate of species in this specific tree: which speciation and extinction rates are most likely to have caused the phylogenetic tree that we have constructed? Given that we can reconstruct a phylogeny from extant species, extinction tends to be smaller than speciation, because high levels of extinction would prevent species from persisting until present time. Conversely, diversification cannot be unlimited either, as we know that planet earth cannot support an unlimited number of species. A number of mechanisms have been proposed that might limit diversification and put a constraint on the number of species (Moen & Morlon 2014). The simplest implementation of a limitation on diversification is assuming that the speciation rate over time has slowly decreased (Rabosky & Lovette 2008b). As a result, the diversification rate decreases as well and the number of species reaches a plateau. Although the time-dependent speciation model provides a good fit to observed slowdowns in speciation, it does not provide any insight into why the speciation rate might decay over time. A more biological explanation of a decrease in the speciation rate over time is provided by the diversity dependent speciation model (Etienne *et al.* 2012). Much akin to logistic growth, where the growth rate of a population diminishes as the population reaches its carrying capacity, the diversity dependent speciation model assumes that a clade has a carrying capacity, and that the speciation rate decreases as the clade accumulates species. The carrying capacity could in this case be interpreted as the maximum

number of available niches, or a limitation to the number of species due to a decrease in range size after speciation. Using the diversity dependence model, the authors embark to explain past diversity in the tribe of *Foraminifera*. *Foraminifera* are a group of plankton that produce a calciferous shell, which easily fossilizes: the White Cliffs of Dover consist (partly) of the remains of *Foraminifera*. As a result, using the excellent fossil record we have a pretty good idea of past diversity in this tribe, in contrast to many other groups of organisms, for which the fossil record only provides a partial account of past biodiversity. As it turns out, the diversity dependent model is able to very closely match past changes in diversity of the *Foraminifera*, whose diversity have remained relatively constant over the past 40 million years. Furthermore, the diversity dependent model allows for independent estimation of the extinction rate (in contrast to the birth-death model, where extinction could only be jointly inferred with speciation), and estimates of extinction using the diversity dependent model are well in line with estimates of extinction using the fossil record.

Speciation might be linked to diversity, or change over time, but it might also be linked to specific traits. Certain traits might provide a competitive advantage and allow for higher speciation rates, or a key innovative trait might give a boost by opening up new niches in the community (Etienne & Haegeman 2012). A whole suite of models has been proposed that uses traits of extant species combined with molecular data to reconstruct the phylogenetic tree. With this phylogenetic reconstruction method we can determine whether differences in traits have contributed to differences in speciation and extinction rates. Models can be applied to either binary traits (e.g. being able to fly or not) (Maddison *et al.* 2007), continuous traits (e.g. body size) (FitzJohn 2010), quantitative traits (FitzJohn 2010) or habitat association (Goldberg *et al.* 2011). For example, FitzJohn and coauthors investigated whether sexual dimorphism contributes to the speciation rate in shorebirds (*Charadriiformes*). The idea behind this is that if males and females differ considerably in size this could generate intersexual conflict, which in turn could yield higher speciation rates. Using a trait-based speciation model they were able to confirm that diversification was generally higher for species that exerted sexual dimorphism, than for species that did not (FitzJohn *et al.* 2009).

All the previously discussed models treat speciation as if happening instantaneously. Although for some polyploid plants instantaneous speciation events do occur (Coyne & Orr 2004), generally speciation takes time to complete. This protracted time period over which speciation has been initiated, but not completed, might leave patterns in phylogenies and could possibly explain the slowdown in speciation events closer to the present. Etienne and Rosindell implemented this protracted speciation in the standard birth-death model (Etienne & Rosindell 2012). Using the protracted speciation model, they were able to show that this waiting time, caused a decrease in speciation events closer to the present. This was directly linked to the time a species required to complete speciation: although there might be many incipient species at present, awaiting speciation completion, these species are not “full” species yet and hence not picked up in the analysis, causing the lack of speciation events close to the present.

In summary, a broad range of models has been proposed to infer past speciation and extinction rates, but generally these models tend to treat past changes in diversity in a phenomenological way, disregarding any underlying mechanics. In the second part of my thesis I will introduce a new model that links changes in the environment with speciation and extinction, in order to include more realism into macro-evolutionary models.

This thesis

This thesis can roughly be divided into two parts: in **part I** (chapters 1, 2 & 3) I focus on inferring processes that have shaped ecological communities, and in **part II** (chapters 4, 5 & 6) I focus on inferring processes that have shaped the evolution of biodiversity.

Part I: Inference in Ecology

Chapter 1 deals with inferring community assembly based on species' traits. Using a dataset consisting of all freestanding trees in a savanna ecosystem we develop a new trait-based community assembly framework. We assess trait based community in the light of dispersal assembly, habitat filtering and limiting similarity. We show that limiting similarity and habitat filtering change across a fire and rainfall gradient and that dispersal assembly is the dominant force determining community composition in savanna tree communities.

Chapter 2 continues along these lines and takes a closer look at how the habitat influences community assembly in cichlid fish in Lake Tanganyika, Zambia. We apply the trait based community assembly model from **chapter 1**, but also look at how species diversity and trait diversity change along a depth, sand cover and habitat complexity gradient. I discover that habitat complexity increases local competition and that sand cover has a strong effect on community composition.

From chapters 1 and 2 it appears that the majority of community assembly seems to be regulated through dispersal assembly – which is very stochastic in nature. In **chapter 3** I explore the effect of differences in dispersal within the Neutral Model. I adapt the standard neutral model to include different modes of dispersal. Using data of tropical trees from Barro Colorado Island, Panama, I test whether including information about dispersal improves our understanding of community assembly for these tropical communities. My analysis reveals that inclusion of information about dispersal improves fit of the model and eliminates previously found ambiguities in the data.

Part II: Inference in Evolution

Before I delve into the inference of speciation rates, extinction rates and past environmental events from phylogenies, I first set out to validate the methods I am about to apply. Firstly, in **chapter 4** we investigate whether the chosen priors when constructing a phylogenetic tree have an influence on inferred rates for these trees. We

test whether the common 2-step approach which first infers the tree using a model prior and secondly infers speciation and extinction rates yields different estimates from a joint inference approach that jointly infers the tree and speciation and extinction rates. It turns out that both methods yield similar estimates and I find that the 2-step approach does not introduce any substantial bias in our estimates, but only if the right clock model is used.

As long as one is able to formulate the likelihood of a phylogeny, given the model, fitting a model to a phylogeny can be fairly straightforward. Nowadays, more and more complex models are being developed for which the likelihood is either computationally too demanding, or simply mathematically intractable. Instead of using the likelihood as a measure of fit, we can substitute the likelihood with one or more summary statistics, and use these summary statistics within an Approximate Bayesian framework. In **chapter 5** I show that three established summary statistics are unable to substitute the likelihood. Furthermore I introduce a novel statistic, the normalized Lineages Through Time statistic, and show that this statistic is successfully able to substitute the likelihood and we recommend its use in Approximate Bayesian Computation.

Chapter 6 builds upon our findings from chapters 4 and 5 and infers past speciation and extinction rates, combined with the inference of past environmental change. I focus on a tribe of cichlids from Lake Tanganyika, the *Lamprologini*. It is generally assumed that the cichlids in Lake Tanganyika have diversified to a great extent due to changing water levels in the lake. These changes in water level caused populations to become geographically isolated, which in turn provided potential for speciation. In this chapter we develop a model that combines extinction, sympatric speciation, allopatric speciation and changes in water level. Using Approximate Bayesian Computation and a given phylogeny, I show that we do not find evidence of water level changes driving diversity in the tribe of *Lamprologini*.

In the last chapter we take a look at what we have learned in the past 6 chapters. How do our approaches between inferences about ecological communities and inferences of evolutionary history differ? Or are they very similar? Can we draw some general conclusions about why species coexists, in both ecological and evolutionary time? Additionally I provide an outlook towards potential new avenues of research.

