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.].

Summary

Both community ecology and evolution span across spatial scales and time-scales that makes them inherently difficult to study. Macro-evolutionary processes that generate new species typically span over millions of years, well outlasting the typical time of a PhD research project (4 years). Community ecology processes involve changes in ecosystems typically well outside the control of the researcher, either due to large temporal (e.g. long generation times and slow turnover of species in a community), or large spatial scales (e.g. long-distance dispersal). Instead of directly manipulating these processes, or studying these processes in the lab, it is often only feasible to look at the outcome of these processes in the current composition of communities, and study patterns caused by underlying processes. Using patterns as a proxy for processes resembles that of a black box approach: imagine having a black box at your disposal that you cannot open to look inside. If you feed the box input, it readily provides output, but you do not know how the internal mechanics operate. The only way of finding out how the inner mechanics of the black box work, is by comparing changes in input and output, and by looking at patterns in already produced output that give away the working of its inner mechanics.

In this thesis I have considered evolution and community assembly as the "black box", patterns in DNA or species abundance distributions as output, and species pools or a single ancestor as the input. I have focused on trying to understand why so many species coexist in savanna tree communities in South Africa, in cichlid fish communities in Lake Tanganyika and in tropical tree communities in Panama. Secondly I have focused on testing whether changes in the environment have driven diversification in cichlid fish in Lake Tanganyika, and tested a number of underlying assumptions regarding diversification analysis.

In the first two chapters of my thesis I have focused on patterns in the distribution of traits of species within a community. Patterns in the trait distribution might point towards the underlying driving process: if the environment poses strong restrictions upon traits, we expect the trait distribution to be narrow because species with extreme trait values are filtered out of the community. Conversely, if species with similar traits experience higher levels of competition, higher predation pressure or higher pathogen load, we expect species with similar traits to be selected against and consequently traits to be more evenly distributed and the trait distribution to be relatively broad. Lastly, the observed trait distributions could simply be a random subset of the traits of all available species able to disperse to the community, regardless of their traits, resulting in a trait distribution without clear patterns. We have combined these three mechanisms into a unifying framework, where species are removed from the metacommunity in a stepwise fashion until we obtain the composition of the local community: STEPwise Community Assembly Model (STEPCAM). Stepwise removal occurs either due to environmental filtering, limiting similarity or dispersal assembly. Using Approximate Bayesian Computation we fitted STEPCAM upon community composition data, combined with species trait data and inferred the relative contribution of these three processes. We applied STEPCAM to two different communities: savanna trees in South-

Africa, and cichlid fish in Lake Tanganyika. We found that dispersal assembly was the most dominant process in determining community composition in savanna trees. Apart from dispersal assembly, habitat filtering turned out to be more important in areas with high fire frequency, whilst limiting similarity was more profound in areas with low fire frequency and high rainfall. In cichlid fish we found similar patterns, where dispersal assembly composed the majority of community assembly, but limiting similarity became more pronounced with an increase in habitat complexity and habitat filtering became more important with increasing sand cover.

In both savanna tree communities and cichlid communities, the majority of community assembly was determined by random dispersal. In the third chapter of my thesis I have further investigated the impact of differences in dispersal on community assembly. Differences in dispersal between functionally equivalent species might arise due to differences in seed size (Muller-Landau and Hardesty 2005), differences in fruit size (Seidler and Plotkin 2006) flight prowess (Valtonen *et al.* 2013) or differences in the duration of the pelagic larval stage in coral reef fish (Victor and Wellington 2000, Almany *et al.* 2007). In this chapter I have focused on tropical tree species, of which 73% disperse via animals (e.g. via bats, birds and small mammals) and the remaining 27% disperse via wind, water or ballistics (Muller-Landau and Hardesty 2005). These differences in seed dispersal might change dispersal ranges and impact dispersal rates, and could impact migration rates between the local and metacommunity. I have modified the standard neutral model of biodiversity to include two guilds that differ in dispersal and fit the guild-structured model on a neotropical tree dataset from Barro Colorado island, Panama. We find that trees that disperse through biotic means (e.g. via birds, mammals, insects etc.) have a higher dispersal rate than trees that disperse through abiotic means (e.g. via wind or water). This could partly be due to the closed canopy of the rainforest, making dispersal via wind less effective. Furthermore we find that adding ecological information about dispersal improves the fit of the neutral model, and partially resolves the two-optima problem previously faced when fitting the neutral model.

In the next three chapters in my thesis I have focused on the inference of macro-evolutionary processes using phylogenetic trees. I have relied on previously published phylogenetic trees, but the construction of a phylogenetic tree is an intricate process that relies on a range of assumptions, which might affect the outcomes of diversification analysis. In chapter 4 I have set out to test the effect of one of these assumptions: the effect of the prior tree model. When reconstructing a phylogenetic tree from molecular data, a tree branching model is used as prior model to reconstruct the tree. The specifics of this model could affect any inferences made at a later stage using the reconstructed tree. If the tree was reconstructed assuming low levels of extinction, perhaps inference made at a later stage has a lower probability of detecting extinction. In this chapter we have compared two different approaches in inferring diversification rates, either using a two-step approach, where first the tree was reconstructed and secondly diversification rates were estimated or a joint approach, where during tree

reconstruction, diversification rates are simultaneously fitted. We found that both approaches yield similar estimates for diversification rates, and did not find evidence that the prior tree model influences diversification rate estimates.

As diversification models become more complex and capture a larger part of the underlying complexity of included processes, it becomes more and more challenging to derive a tractable likelihood function, or computationally demanding to evaluate this likelihood function. Over the past decade, a new Bayesian inference method has been developed that does not rely on the likelihood, but uses summary statistics to attain parameter estimates for the fitted model: Approximate Bayesian Computation (ABC) (Beaumont *et al.* 2002, Csilléry *et al.* 2010, Beaumont 2010, Sunnåker *et al.* 2013). In this chapter I have looked at three established summary statistics – tree size, the gamma statistic and phylogenetic diversity and introduced a novel summary statistic: the normalized Lineage-Through-Time statistic. I evaluated the performance within ABC of these summary statistics for three models of increasing complexity: the standard birth-death model (Nee *et al.* 1994), the time-dependent speciation model (Rabosky and Lovette 2008) and the diversity-dependent speciation model (Etienne *et al.* 2012). As it turns out, only the normalized Lineage-Through-Time statistic is able to adequately substitute the likelihood within an ABC framework.

In the last chapter I have taken a closer look at the interaction between changes in the environment and speciation. As geographical barriers build, populations might get separated, which in turn might provide potential for allopatric speciation. Disappearance of geographical barriers can reunite separated populations and counteract any acquired divergence. These geographical changes can take the form of the formation of mountains or the movement of continents, processes that stretch over long time periods. However, some landscape changes might be on much shorter timescales, such as the formation of rivers, changes in water level of a lake or changes in connectivity between islands as a result of changes in sea level. These relatively rapid changes, coined "dynamical landscapes" (Aguilée *et al.* 2011) might act as a "species pump" (Rossiter 1995) by continually separating populations, providing potential for allopatric speciation, and reshuffling these populations as the separation is lifted. In this chapter I have developed a model that captures this waxing and waning of geographical isolation and have fitted this model upon a phylogeny of a tribe of cichlid fish from Lake Tanganyika, Africa. Lake Tanganyika is known to have undergone several water level changes over the past few million years. Furthermore, the bathymetry of Lake Tanganyika causes the lake to split up into two smaller lakes if the water level drops far enough. These continuous changes in water level could have contributed to the large diversity of cichlid fish we observe nowadays in the lake. Using Approximate Bayesian Computation (using the summary statistics validated in chapter 5) I found no evidence for water level changes and associated allopatric speciation. This needs not indicate that water level changes are irrelevant, rather it suggests that water level changes at such a large scale do not impact diversity. Interactions on a smaller scale, where the fragmented shore line alters due to relatively small water level changes

could still impact allopatric speciation and hence diversity. Alternatively the constant rates pure-birth tree model used to reconstruct the phylogeny of the Lake Tanganyika tribe that I used in my analysis, could have biased our findings. Although my findings from chapter 4 suggest that such biases are typically small, the specific branch rate model used to reconstruct the phylogeny was not assessed in chapter 4 and could still bias our findings.

Conclusion

In the past six chapters I have shown that even though we cannot observe some processes in real time, the patterns they leave in either community composition, or in a phylogenetic tree, allow us to reverse-engineer these processes. After the mathematical and computational dust has settled, it appears that there are a number of conclusions we can draw: first of all stochasticity appears to be an important determining factor in shaping community composition, and trait-based processes are much less important than expected. Either effects are truly stochastic, or stochasticity captures underlying, but not yet identified, mechanisms. Furthermore, the habitat does not only act upon community composition, but can also have important interactions with speciation and drive diversity within a clade, especially when we consider that the habitat also changes over time. Lastly I have demonstrated that using computational methods we are not limited to likelihood methods, but we can allow for more complex models that were previously outside the scope of analysis. With the current ongoing increase in computation power, I can only expect these computational methods to increase in importance and open up new avenues of research in ecology and evolution.