SUMMARY

Species are unevenly spread across time, space and hierarchical levels which can be the result of lineages differing in rates of origination and extinction. This difference can be associated to 1) variation in the extent of species' geographic distribution due to abiotic constraints, 2) traits or features that vary across species and are subject to selection or 3) species inhabiting particular regions whose environmental conditions spur species diversification.

In Chapter 1, I explore the role of ecological limits in regulating the distribution and diversification of species. Although such limits must ultimately arise from constraints on local species coexistence, this spatial context is missing from most macroevolutionary models. I develop a stochastic, spatially explicit model of species diversification to explore the phylogenetic and biogeographic patterns expected when local diversity is bounded. I show how local ecological limits, by regulating opportunities for range expansion and thus rates of speciation and extinction, lead to temporal slowdowns in diversification and predictable differences in equilibrium diversity between regions. However, this model also shows that even when regions have identical diversity limits, the dynamics of diversification and total number of species supported at equilibrium can vary dramatically depending on the relative size of geographic and local ecological niche space: small regions with higher local ecological limits support a higher standing diversity and more balanced phylogenetic trees than large geographic areas with more stringent constraints on local coexistence. These findings highlight how considering the spatial context of diversification can provide new insights into the role of ecological limits in driving variation in biodiversity across space, time, and clades. The neutral nature of this model makes species similar in range size (hence in diversification rates) regardless of their phylogenetic relatedness as long as they are born around the same evolutionary time. However, species might not be equivalent to each other.

When a trait is ecologically/evolutionarily relevant, selection can act on it which leads to lineages differing in diversification rates according to trait variation. Traits change over time and therefore trait evolution and diversification are coupled and should be analyzed together. The state-dependent speciation and extinction (SSE) framework contains methods to detect the dependence of diversification on these traits. For the analysis of traits with multiple states, MuSSE (multiple-states dependent speciation and extinction) was developed. However, MuSSE and other SSE models have been shown to yield false positives, because they cannot separate differential diversification rates from dependence of diversification on the observed traits. The recently introduced method HiSSE (hidden-state-dependent speciation and extinction) resolves this problem by allowing a hidden state to affect diversification rates. Unfortunately, HiSSE does not allow traits with more than two states, and, perhaps more interestingly, the simultaneous action
of multiple traits on diversification. In Chapter 2, I introduce an R package (SecSSE: several examined and concealed states-dependent speciation and extinction) that combines the features of HiSSE and MuSSE to simultaneously infer state-dependent diversification across two or more examined (observed) traits or states while accounting for the role of a possible concealed (hidden) trait. Moreover, SecSSE also has improved functionality when compared with its two “parents.” First, it allows for an observed trait being in two or more states simultaneously, which is useful for example when a taxon is a generalist or when the exact state is not precisely known. Second, it provides the correct likelihood when conditioned on non-extinction, which has been incorrectly implemented in HiSSE and other SSE models. To illustrate this method, I apply SecSSE to seven previous studies that used MuSSE, and find that in five out of seven cases, the conclusions drawn based on MuSSE were premature. I test with simulations whether SecSSE sacrifices statistical power to avoid the high Type I error problem of MuSSE, but I find that this is not the case: for the majority of simulations where the observed traits affect diversification, SecSSE detects this.

Evolutionarily relevant traits include the specialization to different resources (e.g., diet) or habitats. Morphological and physiological characteristics need to be adjusted to specialize in a given habitat and the contribution of such a specialization to patterns of diversity of is not well studied. For instance, cichlid fish species live and specialize at different water depths: light spectrum, temperature, dietary resources and parasites all vary along the water column, generating depth-dependent selective regimes that may have macro-evolutionary consequences. In Chapter 3, I apply and extend the SecSSE model to establish whether 1) diversification rates depend on species depth ranges and 2) depth distributions change during speciation in the Lamprologini clade (endemic to Lake Tanganyika, East Africa). I compare contrasting hypotheses that explicitly account for different models of trait evolution and modes of speciation. I do not find evidence for depth shifts during speciation. Instead, depth shifts occur along the branches of the tree. I do find an association between depth range and speciation rate: depth range generalists (i.e. species distributed along the entire water column) have higher rates of speciation than depth range specialists (i.e. species occupying either shallow or deep water). I show that transitions between shallow water and deep water only occur through a generalist phase, and that shallow-water specialization is a macro-evolutionary endpoint: it is unlikely to change to another state. To explain these findings, I hypothesize that specialization to a given depth range affects dispersal capacities, which could cause differences in speciation rates. An interesting insight of this chapter is that the evolution of a trait can be tightly linked to speciation rates, but does not change during the speciation events themselves.

Besides the interaction between a trait and diversification rates, species inhabit different areas where regimes of diversification could vary and explain the heterogeneous distribution of species. One of well-known biogeographic patterns, is the altitudinal species richness gradient which ultimately results from the interplay of speciation and dispersal, but the relative contribution of these processes is still unclear. In Chapter 4, I use the entire songbird’s (Passeriformes) phylogenetic tree as well as distributional data to assess a comprehensive set of contrasting hypotheses on homogeneity or heterogeneity of diversification rates along the elevational gradient jointly with differential rates of altitudinal dispersal. I employed SecSSE to compare these hypotheses, while accounting for potential differences across geographic and biome realms. I found no support for differential diversification rates being responsible for the species richness gradient. Instead, this gradient seems primarily due to rates of dispersal from high to low elevations being higher than the rates in the reverse direction.

In the Synthesis, I show that a trait can be linked to ecological dominance in local communities, but this advantage in fitness is not enough to increase their speciation (or decrease extinction) rate: an ecologically relevant trait is not necessarily subject to species selection. Finally, I claim that species selection is still a promising area of research, especially in the light of new tools and that phylogenetic information can help understand spatial patterns of diversity.
INTRODUCTION

Life is not homogeneously distributed. This simple statement might be one of the few generalizations in natural sciences and holds true in different dimensions: life is unevenly spread across time, space and hierarchical levels. When comparing the fossil record with the current diversity of life, the extent of the change in composition of flora and fauna over time is enormous, where some groups reigned in the past but have only few delegates surviving to the present. Spatial variation in the number of organisms is also astonishing, with the exuberance of jungles contrasting with plain arid ecosystems, one of many gradients of diversity. With a more detailed examination, one can also find examples of groups of animals or plants which have tremendously flourished producing endless forms while other groups have no more than a handful of members. I cannot explain all the mechanisms or natural forces that have created and maintained this variation in the distribution of life, but in this thesis I will provide some insights into how the spatial distribution of life is the result of natural forces and stochastic mechanisms operating over geological time. At the same time, I will also show that the distribution of life can regulate natural forces and stochastic mechanisms.

The nature of traits at different hierarchical levels

A trait is a characteristic or feature of a biological entity, such as height and body mass, but also dispersal capacity, diet, reproduction mode or habitat preference, to name a few. Organisms interact with the environment through their traits, and often it is possible to describe the ecological role of an organism by looking only at selected and informative traits. These traits are ecologically relevant and their study has been fundamental in biology.

By grouping organisms, many individual trait values (e.g., body masses) can be summarized by averaging them or using another summary statistic. In this way, a species can be described with a mean body size. This is called species’ aggregate trait because body mass is a trait that exists and is measured at the individual level, but it is just scaled up to a higher hierarchal level. In contrast, there is a set of traits that cannot be found at the organism level; these are called emergent traits. Examples of emergent traits at the species level are genotypic variability and geographic range size.

The extension of selection to a higher level than organism

One of the central notions of the process of natural selection is that there is a relationship between a trait and fitness (Lloyd and Gould 1993). Individuals having the right traits for their environment will survive and reproduce. If this idea is extended to species, species with the right traits will be able to persist longer and perhaps create more species. The idea of selection acting at a higher hierarchical level might not seem very Darwinian but it contains the main elements of that theory (Gould 1982). Elisabeth Vrba (1984) proposed
a definition to species selection: “species selection is that interaction between heritable, emergent character variation and the environment which causes differences in speciation and/or extinction rates among the variant species within a monophyletic group”. According to Vrba, species selection acts on traits that are emergent at the species level and heritable.

Whether selection at species levels operates on traits that emerge only at this level has been a matter of philosophical discussion (Grantham 2007). This goes hand-in-hand with the theoretical idea of whether species-level fitness is defined as an average/sum of the organismic fitnesses within a species or not (Lloyd and Gould 1993; Powell and MacGregor 2011). In practice, evolutionary biologists assessing species selection have used either emergent (Eastman and Storfer 2011) or aggregate traits (Goldberg et al. 2010) without paying much attention to the nature of the trait. In the case of aggregate traits, the means by which selection acts on that trait at the individual level might be different from the selection pressures taking place at the species level. In the case of aggregate traits, whether selection acts on traits that emerge only at this level has been a matter of philosophical discussion (Grantham 2007). This goes hand-in-hand with the theoretical idea of whether species-level fitness is defined as an average/sum of the organismic fitnesses within a species or not (Lloyd and Gould 1993; Powell and MacGregor 2011). In practice, evolutionary biologists assessing species selection have used either emergent (Eastman and Storfer 2011) or aggregate traits (Goldberg et al. 2010) without paying much attention to the nature of the trait. In the case of aggregate traits, the means by which selection acts on that trait at the individual level might be different from the selection pressures taking place at the species level (Sargent 2004) but there is no consensus on this. For instance, Van Valen (1975) claims that selection at the species level (e.g., mammal body size) parallels selection at the individual level. So it is suggested that the use of either emergent or aggregate traits might not be fundamental (Rabosky and McCune 2009) as both types of traits have shown to be evolutionarily relevant.

However, aggregate and emergent traits may behave differently during speciation. During a speciation event, the heritability of an aggregate trait seems guaranteed (unless the trait is also relevant ecologically, which might cause ecological divergence) but this does not seem straightforward for emergent traits. Consider the case of geographic range size. On the one hand, the traits that determine the extent of a species’ distribution seem to be inherited which implies that geographic range is inherited (Waldron 2007). On the other hand, events of vicariance or other modes of speciation strongly affect range size when species split. Intra-species genetic variability is another emergent trait that could be subject to selection, and at least for an orchid clade, the evidence on its heritability is weak (Kisel et al. 2012).

**Trait evolution: punctuated equilibrium and gradualism**

Extant species within a clade can show variation in a character which suggests that the trait has not remained static but evolved over evolutionary time. Quantitative traits evolve by increasing or decreasing their value: body mass evolution represents a lineage attaining a larger or smaller size within biological, physical or metabolic boundaries (Schramski et al. 2015). For categorical characters such as diet, trait evolution means the switch from one diet guild to another (e.g., from carnivore to herbivore).

Once one accepts the notion of trait evolution, the question that follows is when the change takes place. Does this occur during speciation events or between speciation events? This dichotomy has fueled a long and intense debate, and as a result, punctuated equilibrium and gradualism have emerged as two main lines of thought. Researchers can rely on the (incomplete) fossil record and/or also use mathematical modelling to study the contribution of these two processes to explain macroevolution. Out of DNA molecules of living species and using probabilistic models, it is possible to formulate a hypothesis of how a clade evolved over time, in other words, the relationships between ancestors and descendants can be reconstructed (i.e., a phylogenetic tree). The branch lengths in these trees represent the time between two speciation events which are the nodes (note: speciation could also occur between the nodes if the phylogenetic tree only reflects extant species, and extinct species are not represented). Support for gradualism would be found when most of the trait evolution occurs along the branches whereas a punctuated theory would be favored if the trait remains mostly the same during these stretches of time and abruptly changes at the nodes.

**Two examples of traits potentially linked to species selection**

The idea of species having differential extinction rates due to character variation might look intuitive i.e., certain features can increase species’ survival by making them highly resistant to extinction. However, species selection could also affect the probability of originating new species by acting on a wide variety of traits (Jablonski 2008). Although these traits could substantially be different from each other, their variation is thought to ultimately affect chances of reproductive isolation and therefore cause species divergence. For instance, flower morphology in angiosperm plants is a core trait for vegetal reproduction which species selection acts upon (Sargent 2004; Goldberg et al. 2010). In this regard, Rabosky (2016) points out that even those reproduction-related traits that are proven to be important for species selection are not necessarily linked to reproductive isolation (see also Etienne et al. 2014). Ellis and Oakley (2016) studied whether species with bioluminescence have diversified more; they separated species using bioluminescence as a strategy to hunt or hide. They found that only those species using bioluminescence as reproductive signals have diversified faster, and concluded that sexually selected traits are an engine of speciation.

As mentioned above, the geographic range size of species is an emergent trait that’s important for how species accumulate over time. The geographic extent of a species’ distribution is tightly associated with chances of extinction: small-ranged species are more likely to become extinct. On the other hand, a large range size provides high chances of speciation as individuals are more likely to encounter new habitats and eventually diverge. An important feature of this trait, is that during speciation, the size of the range is usually not inherited by the descendants. It is likely that one of the daughter species will start with a rather small range and expand over time. This means that geographic
range has an important effect on dynamics of speciation and extinction, but it is also influenced by speciation events.

Spatial distribution of diversity and potential explanations

Species richness often varies with elevation or altitude which produces the so-called gradients of diversity, due to an environmental variable that changes alongside elevation or altitude which is responsible for the differential accumulation of species. It is important to keep in mind that the complex patterns of species distribution should be the outcome of some basic processes: species originate by speciation, can disperse across space, make new species and die (if that) by extinction. Even though there is a number of possible explanations on why there are areas with a remarkable number of species (Etienne et al. 2019; Pontarp et al. 2019), one can group them in two large families of mechanisms: mechanisms based on ecological processes or historical contingencies and mechanisms based on diversification dynamics.

1) Mechanisms based on ecological processes and historical contingencies

Under this view, the rates of origination and extinction in species-rich areas are not particularly higher than in any other region. What makes these hotspots special is that they have special properties that sustain more species (e.g., due to higher primary productivity). This means that there are limits in the number of species/individuals that can live together but this constraint is relaxed in these areas (Rabosky and Hurlbert 2015). This does not explain yet how species accumulate. This requires that species tend to disperse to a species-rich area more often that they leave this area. If we reject the idea that there is a cap on the number of coexisting species (see Hamon and Harrison 2015), the explanation at hand is that species-rich areas are older, providing more time for diversity to accumulate.

2) Mechanisms based on diversification dynamics

The mere fact of inhabiting certain areas could cause lineages to increase their chances of undergoing speciation. For instance, rates of mutation or physiological processes in warm areas are higher (Hoekstra et al. 2013; Porcelli et al. 2015) which can lead to higher rates of species origination. Environmental conditions in a given region (e.g., high climatic stability) might increase the chances of species’ survival that amounts to a decrease in extinction risk and therefore a fast growth of diversity.

A third scenario is a mixture of both families of mechanisms: high species coexistence is allowed locally and this increases the per-lineage rate of diversification by modifying the dynamics of range expansion-contraction (see the last part of this Introduction).

The above mechanisms to explain the distribution of diversity are based on the idea that species are intrinsically similar to one another both evolutionarily and ecologically, and the differences that exist, are not relevant for their rates of diversification or range expansion. If any difference is present, it is acquired by the species when inhabiting a given region. These hypotheses are therefore neutral in essence (Hubbell 2001). In contrast, there are studies where species are assumed to have differences in rates of diversification, so the overall diversification of a site ultimately depends on what species live in it (Quintero and Jetz 2018; Rabosky et al. 2018). These studies argue that species-rich areas are the result of the accumulation of highly-diversifying species that for some reason tend to live together.

Overview of this thesis

In this thesis, I aim to show how the spatial distribution of species is dynamic over time and has important consequences for patterns of species origination and extinction. I show that species do not have the same rates of diversification, and this can be the result of 1) regional equilibrium or out-of-equilibrium dynamics or 2) spatial distribution across gradients. To this end, I developed methods based on simulations and likelihood calculations. They include predictive theoretical models, and methods that are applied to case studies varying in taxonomic scope (songbirds; fish) and geographic scale (worldwide elevational distribution; lake depth gradient).

Are there limits to biodiversity? If so, where do they come from? In Chapter 1, I simulate a clade since its origin until an equilibrium in accumulation of richness is reached; the onset of this stage depends on the number of species that every local community can sustain. I find that a simple individual-based model that includes speciation, colonization and local extinction with different extents of species coexistence, provides insightful expectations on 1) the relationship between local and regional richness, 2) the decrease of diversification rates as the equilibrium is reached, 3) the relative contributions of ecological and geographic constraints to the accumulation of species in a region and 4) how informative reconstructed phylogenies and range size distributions are to infer underlying processes.

Are the traditional methods that assess differential diversification rates based on trait states reliable? Has the process of trait-dependent diversification left a trace in phylogenetic trees? In Chapter 2, I extend existing likelihood approaches (Beaulieu and O’Meara 2016) to infer the dependence of diversification rates on traits. This model is implemented in an R package (SecSSE), available for other evolutionary biologists. With this tool, I revisit seven studies that used the existing approaches and found that their conclusions were premature in most cases. I also show the capabilities and limitations of this tool in rate estimation and model comparison by exploring a number of simulation scenarios with different parameter combinations and extents of data completeness.

Do cichlid fish in Lake Tanganyika switch from one water depth to another? Are there different diversification regimes across the water column? Does speciation induce a shift
from one depth to another? In Chapter 3, I use SecSSE to explore how diversification rates in cichlids depend on the evolution of depth preferences. I also upgrade SecSSE to handle contrasting hypotheses on how speciation is associated with depth-preference shifts. I demonstrate that the chances of moving from shallow water to deep water and the reverse are not the same. Furthermore I show that the evolution of a trait can be tightly linked to speciation rates but is independent of the speciation events themselves.

Most species live in lowlands; is this the result of higher rates of origination or higher species survival in these areas? Do species disperse to a different elevation at some time in their lifetimes? The general question is then, how does biological diversity accumulate along the elevational gradient? In Chapter 4, I explore species diversification and dispersal in the well-known elevational gradient of songbird richness. I feed SecSSE with phylogenetic and geographic information to assess the homogeneity of diversification rates along the gradient. Because of the global scope of this study, I controlled the analyses by considering potential differences in latitude and longitude. I show that 1) differences in richness cannot solely be explained by differential rates of speciation or extinction, 2) lineages are more likely to disperse towards the lowlands and 3) the rates of speciation are not homogeneous across the globe.