UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA

Tese de Doutorado

# Modelagem hierárquica multi-espécies para dados de comunidades: benefícios e aplicações

Aline Richter

Porto Alegre, janeiro de 2023

**Aline Richter**

# Modelagem hierárquica multi-espécies para dados de comunidades: benefícios e aplicações

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ecologia, do Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título Doutora em Ciências com ênfase em Ecologia.

Orientador: Prof. Dr. Leandro Duarte

Coorientador: Prof. Dr. Cristiano A. Iserhard

Porto Alegre, janeiro de 2023

CIP - Catalogação na Publicação

Richter, Aline
Modelagem hierárquica multi-espécies para dados de comunidades: benefícios e aplicações / Aline Richter. -- 2023.
190 f.
Orientador: Leandro da Silva Duarte.

Coorientador: Cristiano Agra Iserhard.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Biociências, Programa de Pós-Graduação em Ecologia, Porto Alegre, BR-RS, 2023.

1. Ecologia de Comunidades. 2. Modelos Hierárquicos Bayesianos. 3. Borboletas frugívoras. 4. Mata Atlântica. 5. Pampa. I. Duarte, Leandro da Silva, orient. II. Iserhard, Cristiano Agra, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

**Aline Richter**

# Modelagem hierárquica multi-espécies para dados de comunidades: benefícios e aplicações

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ecologia, do Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título Doutora em Ciências com ênfase em Ecologia.

Orientador: Prof. Dr. Leandro Duarte

Coorientador: Prof. Dr. Cristiano A. Iserhard

Aprovada em: Porto Alegre, 20 de janeiro de 2023

**Banca Examinadora**

Profa. Dra. Maria João Ramos Pereira
Universidade Federal do Rio Grande do Sul (UFRGS)

Prof. Dr. Thadeu Sobral-Souza
Universidade Federal de Mato Grosso (UFMT)

Prof. Dr. Lucas Augusto Kaminski
Universidade Federal de Alagoas (UFAL)

# Agradecimentos

Como nada se constrói sozinho, gostaria de deixar aqui meu muito obrigada a algumas pessoas que tive a felicidade de conhecer durante minha trajetória acadêmica e algumas que me acompanham desde antes.

Aos meus pais, Elvira e Itamar que da forma deles sempre me apoiaram e me incentivaram a buscar meu caminho. Amo vocês incondicionalmente.

A minha família de origem, pelo apoio e curiosidade sobre as maravilhosas borboletas.

Ao Gabriel, pelo carinho e paciência, por ser meu companheiro de profissão, de campo, e de vida. Obrigada pelo apoio nos momentos que eu mais precisei e por estar sempre ao meu lado.

A minha família de Porto Alegre, Gabriel, Dani, Ti, Gabi e JP. Não tenho palavras para descrever o quanto vocês foram importantes para mim, me acolhendo em suas casas por longos períodos pandêmicos e após também. Obrigada pelas comidas maravilhosas e as risadas de doer a barriga. Amo muito vocês!

A minha família de Pelotas, Tuca, Pri, Isa, Kaká, que apesar das distâncias impostas pela vida, a gente sempre dá um jeito de se reencontrar e celebrar nossa amizade. Obrigada, amo vocês!

Ao pessoal do vôlei do Lamef e IPH, obrigada por acontecerem na minha vida e me mostrarem que as vezes sair da nossa bolha é preciso. Vamos mulheres!

Aos meus amigos do PPG ecologia da UFRGS, pelas risadas, conversas e festinhas, em especial a Mi e a Vivi que me acolherem tão carinhosamente. Ao pessoal do Roleco, pelos nossos rolezinhos pré-pandêmicos, sempre com muita diversão e piña coladas. Obrigada de coração!

Aos meus colegas de LEFF, que também viraram amigos, obrigada pelas conversas, desabafos, incentivos e café! Foi muito bom ter tido a oportunidade de compartilhar esse momento com vocês!

Aos meus ajudantes de campo, Gabriel, Arthur, Dani, Berga, Vivi, Rômulo, Tuca, Kaká e Iury, que de alguma forma acharam um tempinho pra me ajudar e vivenciar essa experiência comigo! Muito obrigada.

Aos meus orientadores Leandro e Cristiano, muito obrigada pelo aprendizado e pelo apoio no desenvolvimento da tese.

A todos os professores e professoras que me ensinaram tanto ao longo da minha trajetória academia, muito obrigada!

Ao PPG Ecologia da UFRGS, pela oportunidade de desenvolver meu doutorado em uma universidade pública e de qualidade.

Aos proprietários e funcionários das propriedades particulares e funcionários da Estação Experimental Agronômica da UFRGS, que possibilitaram a coleta de dados.

A CAPES pelo concedimento da bolsa e ao INCT – Ecologia, Evolução e Biodiversidade pelo financiamento do projeto que propiciou a coleta de dados da tese.

# Resumo

A biodiversidade é moldada por distintos processos ecológicos e evolutivos, atuais e passados, e entender como ela se distribui no espaço e no tempo é uma das questões centrais da ecologia de comunidades. Para entender a biodiversidade, precisamos formas de representá-la, para tanto utilizamos métricas de diversidade. As métricas de diversidade descrevem um padrão, seja na ocorrência das espécies ou de riqueza, e é a partir desses padrões que inferimos os principais processos que moldam as comunidades. Distintos processos podem gerar um mesmo padrão, e discriminar quais processos são mais relevantes nos permite compreender a montagem das comunidades, além de aumentar nossa capacidade de predizer como alterações ambientais afetam a biodiversidade. Independente do processo que estrutura as comunidades, quando medimos a biodiversidade contamos com um erro inerente a qualquer estudo ecológico: a detecção imperfeita. Nessa tese, usamos modelos hierárquicos multi-espécies tanto para acessar como a detecção imperfeita pode afetar os padrões de diversidade (Capítulo 1), como para avaliar processos que estruturam as comunidades em escalas distintas (Capítulo 2 e Capítulo 3). Como organismo modelo, utilizamos a guilda de borboletas frugívoras, as quais são representativas das respostas da diversidade às alterações ambientais. Utilizando dados de comunidades de borboletas frugívoras em escala local, no Capítulo 1, avaliamos como falhas na detecção de indivíduos podem confundir o padrão observado em distintas métricas de diversidade, ressaltando a importância do uso desses modelos para avaliar a diversidade sempre que possível. Nos capítulos 2 e 3, apesar de não considerarmos a detecção imperfeita, acessamos como o filtro ambiental (determinado por variáveis climáticas e de paisagem) e biótico modelam a distribuição das espécies. Enquanto no capítulo 2, avaliamos esses processos para comunidades de borboletas frugívoras do Pampa gaúcho, um dos biomas mais desprotegidos do Brasil, no Capítulo 3 usamos um conjunto de dados de comunidades de borboletas frugívoras da Mata Atlântica, a qual é considerada *hotspot* de diversidade. De maneira geral, demonstro nessa tese a importância da utilização de ferramentas de modelagem que considerem a detecção imperfeita, bem como a não independência (coocorrência) das espécies em modelos de ecologia de comunidade, como os modelos hierárquicos multi-espécies. Dentre as principais vantagens desses modelos destaco a propagação de erro nas estimativas dos parâmetros e a resposta compartilhada das respostas entre as espécies, que permite tanto modelar espécies raras e melhorar a estimativa dos parâmetros quando a confiabilidade da estimativa. Dentre as limitações, destaco a complexidade que esses modelos podem assumir, sendo dispendiosos de tempo e de informações a priori, e também da natureza dos dados, já que modelos como o de detecção precisam de réplicas temporais ou espaciais. Apesar disso, o desenvolvimento de abordagens mais generalizadas e a popularização dos modelos hierárquicos multi-espécies têm muito a contribuir para o entendimento da biodiversidade e dos principais processos que a mantém no tempo e no espaço.

**Palavras chave**: Atributos funcionais, borboletas frugívoras, detecção imperfeita, diversidade taxonômica, filogenia, filtro ambiental, JSDM, Mata Atlântica, Pampa.

# Abstract

Biodiversity is shaped by current and past ecological and evolutionary processes, and understanding how it is distributed in space and time is one of the central questions of community ecology. To understand biodiversity, we need ways of representing it, such as diversity metrics. Diversity metrics describe a pattern, either in species occurrence or richness, and it is from these patterns that we infer the main processes that shape communities. Different processes can generate the same pattern, and decoupling which processes are most relevant gives us a better understanding of community assembly and increases our ability to predict how environmental changes affect biodiversity. Regardless of the process that structures communities, when we measure biodiversity, we rely on an inherent error in all ecological work: imperfect detection. In this thesis, I employ hierarchical multi-species models to assess how imperfect detection affects diversity patterns (Chapter 1) and also to evaluate which processes are responsible for structuring communities in different scales (Chapter 2 and Chapter 3). For this, I use the guild of fruit-feeding butterflies as a model organism, which is highly diverse and sensitive to environmental changes. Using data from local-scale fruit-feeding butterfly communities, in Chapter 1, we evaluated how failures to detect individuals can bias diversity patterns observed in taxonomic, functional, and phylogenetic diversity, highlighting the importance of using these models to obtain more reliable estimates of diversity whenever possible. In Chapters 2 and 3, we assessed how the environmental (determined by climate and landscape variables) and biotic filter shape species distribution. While in Chapter 2, we evaluated these processes for distinct communities of fruit-feeding butterflies in Pampa gaucho, one of the most unprotected biomes in Brazil, in Chapter 3, we used a dataset of fruit-feeding butterfly communities in the Atlantic Forest, which is among the world’s top five biodiversity hotspots. Overall, I demonstrate in this thesis the importance of using modeling tools that consider imperfect detection and species' non-independence (cooccurrence) in community ecology models, such as hierarchical multi-species models. We demonstrated that hierarchical multi-species models are flexible and robust tools and provide an important direction for community analyses. Among its main advantages is the propagation of errors in parameter estimates and the sharing of information among species, allowing both modeling of rare species and improving parameter estimates and estimate uncertainty. Limitations include the complexity that these models can take on, being costly in time and prior information, and the nature of the data, since models, such as *N-mixture* models, need temporal or spatial replicates. Nevertheless, the development of more generalized approaches and the popularization of hierarchical multi-species models has much to contribute to understanding biodiversity and the main processes that maintain it in time and space.

**Key-words**: Fruit-feeding butterflies, Atlantic Forest, taxonomic diversity, phylogeny, functional traits, imperfect detection, Pampa, environmental filtering, JSDM.

# Sumário

# Introdução Geral

A biodiversidade como vemos atualmente é o resultado de bilhões de anos de evolução, moldadas por processos naturais e, mais recentemente, por ações antrópicas (TSCHARNTKE et al., 2005; GÁMEZ-VIRUÉS et al., 2015; RAWAT; AGARWAL, 2015). Frente ao declínio da biodiversidade global, é urgente entendermos como as espécies variam no espaço e no tempo, principalmente em sistemas sob altas pressões antrópicas (JARZYNA; JETZ, 2016). A Ecologia de comunidades busca compreender a geração, manutenção e distribuição da diversidade no espaço e no tempo (MITTELBACH; MCGILL, 2019). Embora hoje dispomos de diversas formas de medir a diversidade, essas medidas representam simplificações da realidade e explicam apenas um ou poucos aspectos da estrutura das comunidades. Além disso, por se tratarem de estimativas provenientes de amostras, estão sujeitas a erros, sendo que o acúmulo desses erros pode resultar em uma representação equivocada da diversidade, afetando a nossa capacidade de detectar padrões de biodiversidade, bem como desvendar os mecanismos responsáveis por tais padrões.

Muitos estudos de ecologia de comunidades usam métricas de diversidade para descrever os padrões e inferir os processos que estruturam as comunidades. Porém, é reconhecido que um mesmo padrão pode ser gerado por distintos processos que atuam em múltiplas escalas. Esses processos podem ser determinísticos ou estocásticos e podem ser vistos como filtros que operam em diferentes escalas. O filtro ambiental está relacionado ao nicho das espécies, sendo que a relação entre espécie e ambiente pode ser mediada por alguma característica funcional ou pelas relações filogenéticas entre as espécies. Além da interação com o ambiente, as espécies interagem entre si (filtro biótico), fazendo com que algumas espécies tendam a coocorrer mais ou menos do que o esperado ao acaso, indicando um possível sinal de interação competitiva (menor coocorrência) ou de facilitação (maior coocorrência). Eventos estocásticos como a deriva ecológica ou desastres ambientais podem adicionar variação às comunidades, principalmente naquelas comunidades com poucos indivíduos (SIQUEIRA et al., 2020), afetando a relação entre organismo e ambiente (CHASE; MYERS, 2011; OVASKAINEN et al., 2017). Todos esses processos estão atuando simultaneamente sobre os indivíduos, e desvendar sua importância relativa para determinar a distribuição das espécies pode ser uma tarefa complicada e muitas vezes inalcançável.

Desde seu surgimento, a ecologia de comunidades passou por grandes mudanças teóricas e analíticas, passando de uma ciência descritiva para uma ciência preditiva (OVASKAINEN; ABREGO, 2020; POGGIATO et al., 2021). As primeiras análises de comunidade se limitavam a descrever qualitativamente os padrões de riqueza ou abundância e os primeiros métodos multivariados só emergiram nos anos 60, juntamente com o desenvolvimento do conceito de diversidade beta. Avaliar as comunidades de forma multivariada nos permite capturar os padrões de variação na composição de espécies entre locais ou ao longo do tempo. O desenvolvimento de métodos de ordenação restrita (*constrained*) permitiu testar diretamente os efeitos de processos ecológicos e espaciais, também chamados de processos de montagem, na variação da ocorrência e abundância das espécies bem como a hipóteses relacionadas a padrões de coocorrência (OVASKAINEN; ABREGO, 2020). Descrever a diversidade na forma univariada (riqueza) ou multivariada (composição) depende do objetivo de cada estudo, porém os dois cenários podem ser abordados dentro de um contexto Bayesiano ou Frequentista.

Embora a abordagem Frequentista ainda seja a inferência estatística mais utilizada em estudo de comunidades, a uso da inferência Bayesiana associada à modelagem hierárquica tem crescido dentro da ecologia de comunidades. A inferência bayesiana trata os parâmetros como variáveis aleatórias governadas por uma distribuição teórica conhecida, e então a inferência é feita com base em uma distribuição posterior (KÉRY; ROYLE, 2016). Já os modelos hierárquicos estimam os parâmetros seguindo uma estrutura hierárquica, e associados com a inferência bayesiana, permitem que espécies possam compartilhar informações entre si e nos distintos níveis hierárquicos. Dessa forma, os modelos hierárquicos multi-espécies possibilitam a estimativa de parâmetros para espécies raras, as quais geralmente são desconsideradas nos modelos de distribuição (tanto frequentista quanto em alguns modelos bayesianos) e ainda permitem a propagação da incerteza durante a estimativa dos parâmetros em todos os níveis hierárquicos do modelo. Considerando que a forma como as espécies variam nas suas respostas as condições ambientais está muito mais para uma coleção de histórias relacionadas do que trajetórias independentes, cada vez mais se torna urgente o uso de ferramenta analíticas que incorporem a natureza multivariada dos dados.

Dado a importância da descrição e compreensão dos fatores que geram padrões ecológicos para ecologia de comunidade e biologia da conservação, é essencial o uso de ferramentas capazes de produzir estimativas acuradas de padrões em biodiversidade (ex. inclusão de detecção imperfeita), e que permitam testar hipóteses sobre processos ecológicos (ex. considerando coocorrência de espécies). Nesse sentido, os modelos hierárquicos multi-espécies se mostram mais promissores que os métodos convencionais frequentistas, e ainda abre um novo campo para o desenvolvimento e teste de novas teorias (ELLISON, 2004). Portanto, nos próximos parágrafos, descreverei brevemente como os modelos hierárquicos multi-espécies foram utilizados nesta tese para responder diferentes questões associadas com a resposta das borboletas frugívoras a fatores ambientais em distintas escalas. Borboletas frugívoras são consideradas como ótimos modelos de estudo para representar a relação entre biodiversidade e ambiente (BROWN; FREITAS, 2000). Essas borboletas se alimentam exclusivamente de frutos fermentados, exsudatos de plantas entre outros compostos orgânicos em decomposição (DEVRIES, 1988), facilitando a amostragem dessa guilda através de métodos passivos e padronizados (FREITAS et al., 2014). Além disso, elas representam entre 50 e 75% da diversidade de borboletas da região neotropical (BROWN, 2005), e devido ao grande número de estudos com esse grupo, hoje dispomos de uma boa resolução taxonômica e filogenética (CHAZOT et al., 2021), bem como informações sobre características morfológicas e comportamentais (BECCALONI et al., 2008; SHIREY et al., 2022).

No capítulo 1, analisamos uma comunidade de borboletas frugívoras em uma escala local, onde buscamos entender como a detecção imperfeita, ou seja, a falha em observar todos os indivíduos de uma comunidade (MACKENZIE et al., 2002), pode afetar os padrões de diversidade observados entre dossel e sub-bosque de uma floresta com Araucária no Sul do Brasil. Para isso, usamos um modelo multi-espécies para dados de abundância (*N-mixture*), o qual permite separar o processo biológico do processo amostral e estimar a verdadeira abundância das espécies. A partir desta estimativa comparamos o padrão de diversidade resultante, considerando ou não a deteção imperfeita (*Hidden Diversity Framework*). Observamos que a detecção imperfeita gerou um ruído ou um viés nos padrões de diversidade e que a intensidade de viés depende da métrica avaliada (taxonômica, funcional ou filogenética). Concluímos este capítulo destacando a necessidade que ferramentas analíticas que modelem os erros na detecção sejam usadas em estudos de comunidades locais, garantindo uma maior acurácia no padrão observado bem como na inferência dos padrões que mantêm essas comunidades no tempo e no espaço.

Embora o problema da detecção imperfeita seja considerado nos modelos *N-mixture*, esse tipo de modelo não costuma incluir a coocorrência das espécies na estimativa dos parâmetros oriundos do processo biológico, além de terem sua aplicação limitada a delineamentos amostrais específicos que nem sempre são comuns em dados de ecologia de comunidades (ex. réplicas amostrais). Dessa forma, além da limitação em aplicação, não conseguimos distinguir a importância relativa de filtros ambientais e processos bióticos na distribuição das espécies. Modelos de distribuição conjunta de espécies (*joint species distribution models*, JSDM) surgem como uma importante ferramenta nesse sentido, pois modelam as coocorrência residuais, podendo indicar interações bióticas (POGGIATO et al., 2021), além de aumentar o poder estatístico destes modelos. A Modelagem Hierárquica de Comunidades (conhecido como HMSC em inglês) é um *framework* para análise de comunidades, onde os componentes do modelo são conceitualmente ligados aos processos de montagem, como filtros ambientais e bióticos. Embora o padrão de coocorrência residual não possa ser interpretado puramente como interação biótica (OVASKAINEN; ABREGO, 2020; POGGIATO et al., 2021) essa informação pode melhorar a capacidade de modelos em predizer a ocorrência ou abundância das espécies. Portanto, visto a dificuldade de construir modelos que incorporem a detecção imperfeita em macro escala, tal qual o utilizado no primeiro capítulo 1, nos capítulos 2 e 3 utilizei o arcabouço analítico do HMSC para estimar como as espécies respondem ao filtro ambiental, considerando também como atributos funcionais e relações filogenéticas são capazes de afetar tais relações.

Variáveis ambientais como características de paisagem e condições climáticas podem afetar a ocorrência e abundancia de formas distintas (CHECA et al., 2014, 2019; SANTOS et al., 2020). Variações microclimáticas podem ter um papel fundamental na dinâmica de metapopulações, afetando o fitness e controlando o tamanho das populações (abundância das espécies) (CHECA et al., 2019; XING et al., 2016), enquanto que no nível de comunidade, o microclima pode ser crucial na segregação de nicho, levando ao aumento na diversidade das comunidades (WALLISDEVRIES; VAN SWAAY, 2006). Por outro lado, paisagens heterogêneas podem ser importantes para manter a biodiversidade, pois possibilitam a movimentação das espécies entre distintas estruturas de habitat permitindo a coocorrência das espécies (BÖRSCHIG et al., 2013; GÁMEZ-VIRUÉS et al., 2015). Considerando isso, no capítulo 2, avaliamos como variáveis microclimáticas e de paisagem afetam a distribuição e ocorrência de espécies de borboletas frugívoras amostradas no Pampa gaúcho, e como características intrínsecas da comunidade (tamanho das comunidades baseado no número de indivíduos) atuam para determinar a ocorrência e abundância destas espécies no espaço. Além de saber qual a contribuição do filtro ambiental na variação da ocorrência e abundância das espécies, investigamos como padrões de riqueza de espécies e composição funcional foram afetados pelas variáveis ambientais. Observamos que as variáveis de paisagem tenderam a determinar a variação na ocorrência enquanto o micro-habitat determinou a variação na abundância das espécies. No geral, as comunidades responderam negativamente ao aumento da temperatura, indicando que habitats campestres tendem a ser menos ricos e abundantes que habitat de floresta ripária. No entanto, a heterogeneidade da paisagem foi importante para manter a diversidade, uma vez que aumento na homogeneização da paisagem levou a um aumento da ocorrência e abundância de espécies mais generalistas. Além disso, observamos que em comunidades maiores (compostas por mais indivíduos) os efeitos do filtro ambiental tenderam a ser mais fortes, expondo a fragilidade de comunidades pequenas, tanto no entendimento dos processos que as estruturam, quanto na nossa capacidade de predizer como essas comunidades irão responder as mudanças climáticas.

Por fim, no capítulo 3, avaliamos como as comunidades de borboletas frugívoras na Mata Atlântica respondem a variáveis climáticas e de paisagem e como essas respostas são distribuídas espacialmente em uma escala biogeográfica. Observamos que as variáveis climáticas foram mais importantes do que as variáveis de paisagem para determinar a ocorrência das espécies, e que os atributos e a filogenia influenciaram o nicho das espécies. Ainda, observamos que riqueza e a forma média da asa tiveram respostas similares as variáveis ambientais, refletindo em um padrão sobreposto nas predições espaciais, onde locais ricos e espécies que apresentam alta capacidade dispersiva estão geralmente associadas com centros endemismo e zonas de transição. Assim, sugerimos que estes locais são importantes no fornecimento de espécies para áreas adjacentes devido a eventos de dispersão. Além disso, observamos que o tamanho do corpo nesse grupo segue a regra inversa de Bergmann (SHELOMI, 2012), uma vez que comunidades em altas elevações tendem a ser compostas por espécies com menores tamanhos de corpo.

De maneira geral demonstrei que os modelos bayesianos multi-espécies representam uma ferramenta promissora para estudos em ecologia de comunidades, permitindo a inclusão de fatores importante como a detecção imperfeita e a coocorrência das espécies, fornecendo uma compreensão mais acurada dos processos que governam a montagem das comunidades. Além das contribuições teóricas e práticas provenientes dos três capítulos dessa tese para o entendimento de comunidade de borboletas frugívoras, usando os dados coletados para o capítulo 2, foi desenvolvido um guia para borboletas frugívoras do Pampa gaúcho (MACHADO et al., 2022). O guia está disponível em formato e-Book e pode ser acessado no link https://www.ufrgs.br/leff/wp-content/uploads/2022/12/Guia-Ilustrativo-de-Borboletas-Frugivoras-do-Pampa-Gaucho-1.pdf. Este guia foi concebido com o intuito de aumentar o conhecimento sobre as borboletas, além de conscientizar a população local sobre a diversidade de espécies que ocorrem nesse bioma rico, mas também pouco protegido.

# Capítulo 1 – Efeito da detecção imperfeita sobre a diversidade

## THE HIDDEN SIDE OF DIVERSITY: EFFECTS OF IMPERFECT DETECTION ON MULTIPLE DIMENSIONS OF BIODIVERSITY

1- Capítulo publicado na Revista *Ecology and Evolution* em agosto de 2021

https://doi.org/10.1002/ece3.7995

2- Este capítulo tem como coautores: Gabriel Nakamura, Cristiano Agra Iserhard e Leandro Duarte

## Abstract

1. Studies on ecological communities often address patterns of species distribution and abundance, but few consider uncertainty in counts of both species and individuals when computing diversity measures.
2. We evaluated the extent to which imperfect detection may influence patterns of taxonomic, functional, and phylogenetic diversity in ecological communities.
3. We estimated the true abundance of fruit-feeding butterflies sampled in canopy and understory strata in a subtropical forest. We compared the diversity values calculated by observed and estimated abundance data through the hidden diversity framework. This framework evaluates the deviation of observed diversity when compared with diversities derived from estimated true abundances, and if such deviation represents a bias or a noise in the observed diversity pattern.
4. The hidden diversity values differed between strata for all diversity measures, except for functional richness. The taxonomic measure was the only one where we observed an inversion of the most diverse stratum when imperfect detection was included. Regarding phylogenetic and functional measures, the strata showed distinct responses to imperfect detection, despite the tendency to overestimate observed diversity. While the understory showed noise for the phylogenetic measure, since the observed pattern was maintained, the canopy had biased diversity for the functional metric. This bias occurred since no significant differences were found between strata for observed diversity, but rather for estimated diversity, with the canopy being more clustered.

5. We demonstrate that ignore imperfect detection may lead to unrealistic estimates of diversity and hence, to erroneous interpretations of patterns and processes that structure biological communities. For fruit-feeding butterflies, according to their phylogenetic position or functional traits, the undetected individuals triggered different responses in the relationship of the diversity measures to the environmental factor. This highlights the importance to evaluate and include the uncertainty in species detectability before calculating biodiversity measures to describe communities.

**Keywords:** assemblage structure, community hierarchical models, detection probability, environmental gradients, fruit-feeding butterflies, functional traits, phylogenetic diversity.

## Introduction

Estimating the whole biodiversity in a community is a key challenge for ecologists. First, because we do not have time and resources to sample all species and individuals that are present in a community. Second, even focusing on a target group, there are large proportions of species or individuals that remain “hidden” (Chao et al., 2017; Devarajan, Morelli, & Tenan, 2020; Guillera-Arroita, Kéry, & Lahoz-Monfort, 2019; Yoccoz, Nichols, & Boulinier, 2001). This occurs since both species and individuals are not perfectly observed in the field (i.e. they are undetected during sampling), and different species have distinct probabilities of being detected (Boulinier, Nichols, Sauer, Hines, & Pollock, 1998; Ribeiro, Williams, Specht, & Freitas, 2016). Classical community analyses commonly ignore imperfect detection, for both incidence and abundance-based approaches, as well as its effects on diversity measures (DeVries, Alexander, Chacon, & Fordyce, 2012; Pillar & Duarte, 2010; Wiens & Donoghue, 2004). Identify the effects of imperfect detection in classical diversity measures might improve our understanding of relationships between diversity and environmental gradients (Roth, Allan, Pearman, & Amrhein, 2018), and ultimately the processes that structure the biological communities (Dorazio, Connor, & Askins, 2015).

A considerable portion of community studies that employed models that account for imperfect detection (e.g., Multi-Species Hierarchical Models) are interested in evaluating the true occurrence or abundance, aiming to guide management practices (Ruiz-Gutiérrez, Zipkin, & Dhondt, 2010; Yamaura et al., 2012; Zipkin, Andrew Royle, Dawson, & Bates, 2010). But, the effects of imperfect detection are not restricted only to the taxonomic aspect of diversity (e.g. species richness), and our ability in detecting biodiversity patterns may differ among different components of diversity (Iknayan, Tingley, Furnas, & Beissinger, 2014; Jarzyna & Jetz, 2016). Species co-occurring in communities exhibit different levels of shared evolutionary history and variation in phenotypic traits. These features of species are widely used to infer historical and/or ecological mechanisms determining community assembly patterns (Duarte, Debastiani, Carlucci, & Diniz-Filho, 2018; Graham & Fine, 2008; Webb, Ackerly, McPeek, & Donoghue, 2002). Despite the increase in studies that quantified phylogenetic or functional diversity (de Bello et al., 2015; Tucker et al., 2017), few consider the imperfect detection in species count for calculate it (Chao et al., 2017; Frishkoff, De Valpine, & M’Gonigle, 2017) or have quantified the role and magnitude of the effects of imperfect detection on distinct facets of diversity (Jarzyna & Jetz, 2016; Si et al., 2018). If undetected species have unique phylogenetic information or functional traits, by underestimating their contribution to diversity estimate, we are neglecting an ecologically important part of the assemblages (Jarzyna & Jetz, 2016). Consequently, we would observe a more clustered assemblage than they really are (Si et al., 2018). The opposite can also occur when undetected species are phylogenetically or functionally redundant (Jarzyna & Jetz, 2016), and the observed assemblages will overestimate phylogenetic and functional diversity. Furthermore, the detection of species can be biased at some part of the environmental gradient evaluated (Roth et al., 2018). If this occurs, not only the observed diversity pattern can be affected, but also our interpretation of the relationship among diversity and environmental gradients (Fig. 1a-b).

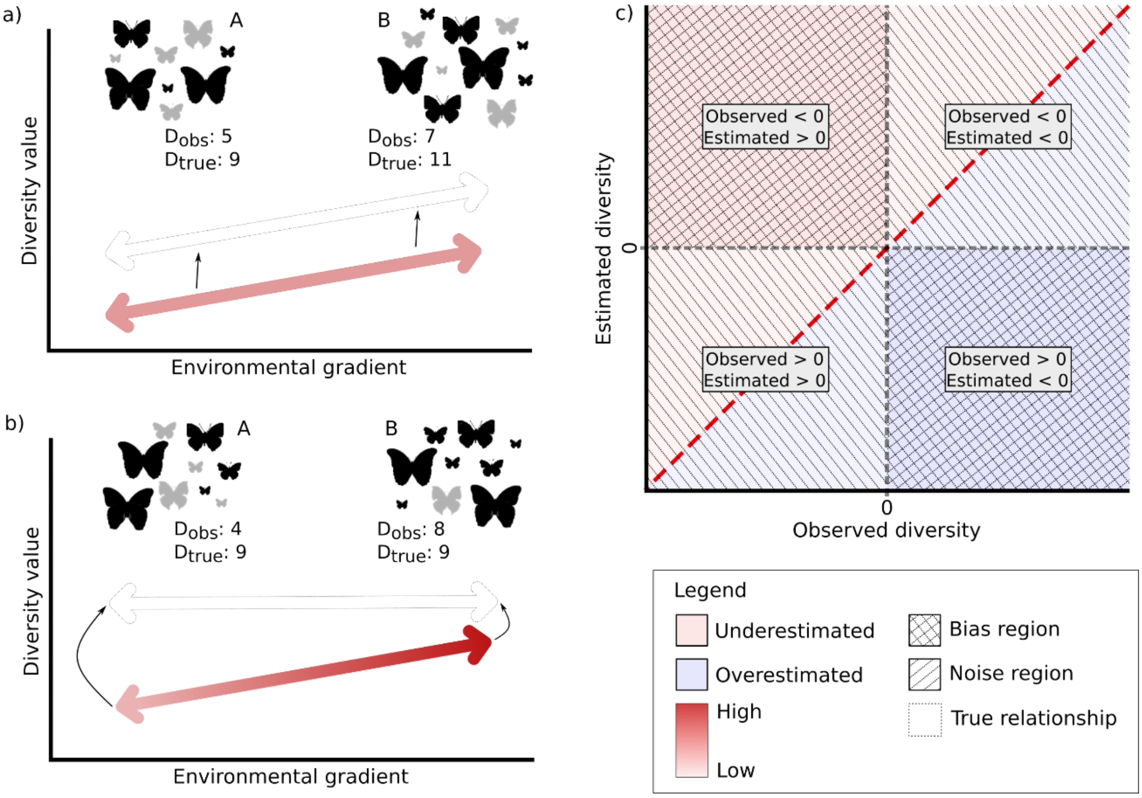

**Figure 1**. Schematic representations of the hidden diversity framework, demonstrating how imperfect detection can influence the relationship between an environmental factor and a diversity descriptor. Suppose that each set of butterflies represents a community, called A and B. $D_{obs}$ is the value of a given diversity measure calculated from an observed community (dark butterflies), which has imperfectly detected species (gray butterflies, probability of detection (p) < 1). $D_{true}$ represents the real value of this diversity if all species in the community were sampled (p = 1). For the sake of simplicity, we will call the difference between the true and observed values hidden diversity (HD). Note that in a) despite the 4 units decrease in diversity for both communities ($HD_A = HD_B = -4$), B remained more diverse than A, and the error associated with imperfect detection was constant along the environmental gradient. On the other hand, in b) the detection probability is not equal along the gradient, which might lead to a bias in the observed relationship between diversity measures and

environmental factors, once that $D_{true}$ is the same for A and B, but when only observed data is employed to calculate diversity B more diverse than A ($HD_A$ = -5, $HD_B$ = -1). In c) we show a way to interpret the hidden diversity, which takes into account the signal of the observed and estimated diversity value. The blue and red colors are associated with positive and negative values of HD, respectively. If the observed value (x-axis) is positive and the estimated value (y-axis) is negative we have an overestimation in the diversity value, while if the observed value is negative and the estimated value is positive we have an underestimation in the diversity, and both are in the critical bias region. When the observed and estimated values have the same sign, the observed patterns tend to hold despite the noise added by imperfect detection.

Insects are the most species-rich taxa in the world, which poses a major challenge for ecologists interested in evaluating insect diversity patterns (Thomas, 2005). Among insect groups, butterflies are considered important biological indicators due to their short life-cycle and high sensibility to changes in environmental features (New, 1997, Brown & Freitas, 2000). Fruit-feeding butterflies are a conspicuous guild of tropical butterflies that feed on rotting fruit, carrion, or plant exudates (DeVries, 1988) and represent about 50 – 75% of nymphalid diversity in the Neotropical region (Brown, 2005). Assemblages of fruit-feeding butterflies show high vertical stratification (Devries, 1988; DeVries, Alexander, Chacon, & Fordyce, 2012; Ribeiro & Freitas, 2012; Santos, Iserhard, Carreira, & Freitas, 2017), with the canopy generally being taxonomically more diverse than understory. These strata exhibit a large difference in their microclimatic conditions, habitat structure and, hence in their taxonomic composition (Araujo, Freitas, Souza, & Ribeiro, 2020; DeVries et al., 2012; Santos et al., 2017). Whereas Charaxinae, Biblidinae, and Nymphalinae are recognized as canopy-dwellers, Satyrinae is generally associated with understory sites (Schulze, Linsenmair, & Fiedler, 2001). In a phylogenetic or functional perspective, the composition of those strata could be also dissimilar, once that lineages of fruit-feeding butterflies exhibit habitat preferences (Fordyce & DeVries, 2016) and individuals tend to show traits that varying according to characteristics and preferences (Graça, Pequeno, Franklin, & Morais, 2017).

Due to their feeding habit, these butterflies can be sampled with passive and standardized methodologies such as bait traps (Freitas et al., 2021). Unlike other methods to sample butterflies (entomological nets or transect counts), bait traps avoid bias related to variation in the observer or personal expertise about species detection (Boulinier et al., 1998, Kéry & Plattner, 2007, Ribeiro et al., 2016). However, the detection of individuals might be biased by bait attractiveness in different habitats and by the individual ability to find the trap. Weather conditions as wind speed, rain, and temperature, can influence the bait volatiles, leading to decreased attractiveness, especially in open habitats (Marini-Filho & Martins, 2010). Fruit-feeding butterflies typically use odor cues to locate food, and some groups, such as Charaxinae, can find more accurately their preferred food (Molleman, Alphen, Brake, & Zwaan, 2005). Further, individuals that have high mobility, may often be undetected in a sampling site because it is visiting other sites within their home range (Joseph, Elkin, Martin, & Possingham, 2009). Therefore, bearing in mind the intrinsic challenges of sampling in the canopy together with the characteristics of individuals that inhabit this stratum, it is more likely that the canopy has a higher number of undetected individuals than understory, yielding a bias in diversity measured in this stratum.

In this study, we aimed to analyze the extent to which imperfect detection, assessed by the estimates of the true abundance of species, can lead to changes in observed patterns of taxonomic, functional, and phylogenetic diversities of butterflies living in different forest strata (canopy *vs.* understory). We expect that: (i) canopy will show lower individual detection than understory, leading to a source of bias that hides the true diversity value for this stratum. Consequently, this bias induces an erroneous inference when we compare diversity values between canopy and understory. (ii) The effect of imperfect detection will be lower for phylogenetic and functional measures concerning taxonomic diversity. In this case, an increment in species number will not be followed by an increment in both phylogenetic and functional diversity, indicating that undetected species are redundant with species sampled in the observed community.

## Material and Methods

### *Study sites and sampling procedures*

The study site was located in Floresta Nacional de São Francisco de Paula (FLONA-SFP; centered at 29°25’22’’S, 50°23’11’’W) in Southern Brazil. FLONA-SFP comprises an area of 1,615 ha in the Atlantic Forest biome and is composed of Mixed Ombrophilous Forest with the presence of *Araucaria angustifolia* (Bertol.) Kuntze, as well as patches with *Pinus* sp. and *Eucalyptus* sp. plantations (ICMBio, 2020). The climate of the region is temperate without a dry season, and with annual mean rainfall near 2,000 mm and an annual average temperature of 14.5°C (Sonego, Backes, & Souza, 2007).

Fruit-feeding butterfly assemblages were sampled between November 2016 and March 2017, which correspond to the summer season in the Southern Hemisphere and which is the best period of the year for sampling butterflies in the Atlantic Forest (Iserhard, Romanowski, Richter, & Mendonça, 2017). We adopted standardized methods for sampling fruit-feeding butterflies in the Neotropical region (Freitas et al., 2014), which consisted in install five traps per sampling unit, which were baited with a mixture of mashed banana and sugarcane juice (Freitas et al., 2021). We performed monthly surveys at six sites of native forest within FLONA-SFP for five months. In each month, the traps remained open for eight to ten consecutive days and every 48h the traps were checked and the bait was replaced. This totalizes a sampling effort of 2,520 trap-days (10 traps × 6 sampling units × 42 sampling days). In each site, we sampled the assemblages of fruit-feeding butterflies in the canopy (~15 m above the ground, inside canopy tree crowns) and in the understory (1.5 m above the ground) and each stratum was considered as one independent sampling unit. In every trap checking, we measured the temperature of the base of each trap using an infrared thermometer (GM-300, Benetech®).

### *Community model for abundance data*

We employed a modification of the Dorazio-Royle-Yamaura model (DRY) (Kéry & Royle, 2016; Yamaura et al., 2011; Yamaura, Kéry, & Andrew Royle, 2016) to estimate uncertainties in the individual counts for fruit-feeding butterflies. The modifications allow the model to estimate the mean abundance ($\lambda_{ik}$) and detection probability ($p_{ijk}$) for each stratum (Zipkin et al., 2010). We assumed that local abundance remained unchanged during the survey (i.e. closure assumption, Kéry, Royle, & Schmid, 2005) since we sampled in a narrow time window, and that mean abundance and detection probability were independent among species. Abundance for each species $k$ at each site $i$ is a latent variable (i.e. imperfectly observed) called $N_{ik}$, which follows a Poisson distribution:

$$N_{ik} \sim \mathrm{Poisson}(\lambda_{ik})$$

where $\lambda_{ik}$ is the mean or expected abundance. We assumed that $\lambda_{ik}$ varied among sites depending on species random effects and if point *i* was in the canopy (Strata = 0) or the understory (Strata = 1), thus allowing species-level effects to differ between the two strata (Zipkin et al., 2010). We also included a slope for the mean temperature obtained from the base of the traps of each site *i* (Temp) and add two random site effects, because samplings were repeated in time (sampling months, SM) for each sampling units (SU), and hence their measures are not independent within them. We fit the model for biological process using a log-link function, as follows:

$$\log(\lambda_{ik}) = \beta.\text{can}_k \times (1 - \text{Strata}_i) + \beta.\text{und}_k \times \text{Strata}_i + \beta 1_k \times \text{Temp}_i + s_{\text{SU}_i k} + m_{\text{SM}_i k}$$

where β.can and β.und are the species-specific intercepts for canopy and understory, respectively, β1 is the species-specific slope for the temperature effect, *s* and *m* are the random effects for six sampling units and five sampling months.

We describe the detection process as:

$$y_{ijk} \sim \text{Binomial}(N_{ik}, p_{ijk})$$

where the number of detected individuals $y_{ijk}$ during visit *j* was obtained with $N_{ik}$ trials and a probability of successful detection $p_{ijk}$. The detection history $y_{ijk} > 0$ indicates that the species *k* (1, 2, ..., 35) was observed in site *i* (1, 2, ..., 12) during the sampling occasion *j* (1, 2, …, 5), while $y_{ijk} = 0$ implies the species was undetected. We modeled detectability as a logit-linear combination of species-specific detection probabilities dependent on the stratum and two covariates:

$$\text{logit}\left(p_{ijk}\right) = \alpha.\text{can}_k \times (1 - \text{Strata}_i) + \alpha.\text{und}_k \times \text{Strata}_i + \alpha 1_k \times \text{Date}_{ij} + \alpha 2_k \times \text{Temp}_{ij}$$

where α.can and α.und are the species-specific intercepts for canopy and understory, respectively, and α1 is the linear effects of the sampling day (transformed to Julian date) and α2 is the linear effects of the temperature by day.

All covariates for the biological and observation process were standardized before perform the Bayesian model. The effect of predictors was corroborated whenever 95% of the Credible Interval (CRI) did overlap zero. We defined species-specific parameters for each stratum and for covariates as coming from normal hyper-distributions, e.g., $\beta.can_k$ ~ Normal ($\mu_{\beta.can}$, $\tau_{\beta.can}$), being that these priors describe the heterogeneity among species. We determined vague priors for the hyper-parameters that define the mean ($\mu$) and precision ($\tau$) at the community-level, such that $\mu$ ~ Normal (0, 0.001) and $\tau$, that is the inverse of variance ($\tau = sd^{-2}$), where sd ~ Uniform (0, 10), and these hyper-parameters are shared by all species in each stratum (Yamaura et al., 2016). Considering that the mean detection probability must vary between 0 and 1, we defined $\mu_{\alpha} = logit(\mu_{\alpha.pre})$, when $\mu_{\alpha.pre}$ ~ Uniform (0, 1), and then, $\alpha_k$ ~ Normal ($\mu_{\alpha}$, $\tau_{\alpha}$). The model was run using the package *jagsUI* (v. 1.4.9, Kellner, 2021) with three Markov Chains Monte Carlo (MCMC), 150,000 iterations with the first 50,000 iterations discarded, and a thinning rate of 100. The model code is given in Appendix A (N-mixture model). These settings of MCMC results in a posterior sampling with 3,000 iterations. We also defined initial values for parameter N and monitored the community mean and species-level parameters. We checked the convergence of MCMC by R-hat statistics (Gelman & Rubin, 1992) and graphical visualization.

In addition, we checked and validated the N-mixture model through simulation of metacommunities (Appendix B – Model Validation). For each simulation, we set the mean expected abundance for canopy and understory ($\beta_{s1}$ and $\beta_{s2}$) or the mean probability for canopy and understory ($\alpha_{s1}$ and $\alpha_{s2}$) to vary, while all other parameters were kept constant. For each parameter, we defined true mean values, which we consider low, intermediate, and high, resulting in 12 simulated metacommunities (hereafter treated as setting code A to L). The output of the simulation provided two main information: the true abundance of species for each community ($N_s$) and the imperfect observed community ($yobs_s$). The $yobs_s$ was then subjected to the N-mixture model, and we monitored all parameters estimated. For the biological model, all true values of parameters and hyper-parameters fall within 95% of the credible interval of the posterior distribution (Appendix B – Fig. B1 to B3), indicating that model was able to recovery true parameters values.

### *Phylogenetic and functional data*

We collected at least one specimen of each butterfly species captured in bait traps for subsequent measurement of functional traits. We selected 12 functional traits to characterize functional diversity in each community, including traits related to flight performance, habitat use, and ecological behavior (Table 1) (Chai & Srygley, 1990; Dudley, 2002; Spaniol, Duarte, Mendonça, & Iserhard, 2019). Using the recently proposed phylogeny of Chazot et al. (2019) for Nymphalidae, we obtained the phylogenetic relationships among the 35 species of fruit-feeding butterflies recorded in this study. We pruned the complete tree to calculate measures of phylogenetic diversity and structure of communities. We used the packages *ape* (v. 5.3, Paradis & Schliep, 2019), and *phytools* (v. 0.6-44, Revell, 2012) to prune the tree.

**Table 1.** Description for the functional traits measured for fruit-feeding butterflies sampled at FLONA-SFP, southern Brazil. C – continuous traits, B – binary traits.

| | Trait name | Type | Measure | Description | References |
|---|---|---|---|---|---|
| FWL | Forewing length (mm) | C | Forewing base to apex | Used as a proxy for body size and related with dispersion capacity | Chai and Srygley 1990, Sekar 2012 |
| TM:TDM | Thorax mass to total body mass ratio | C | The ratio between thorax mass and total body mass | The proportion that represents the investment in thorax mass; related to flight capacity due that thorax allocates the flight muscles | Chai and Srygley 1990 |
| AM:TDM | Abdomen mass to total body mass ratio | C | The ratio between abdomen mass and total body mass | The proportion that represents the investment in abdomen mass; related to investment in reproductive tissues | Srygley and Chai 1990 |
| FEA | Functional eye area (mm²) | C | Set of linear eye measurements | Represent the functional visual field; associated with habitat perception | Rutowski 2000, Turlure et al. 2016 |

| | | | | | |
|---|---|---|---|---|---|
| WL | Wing loading (N/m$^2$) | C | Amount of body mass sustained by wing area unit | Related with flight speed and agility and can be associated with adaptative response to environmental gradients | Chai and Srygley 1990, Berwaerts et al. 2002, Turlure et al. 2016 |
| AR | Aspect Ratio | C | The ratio between forewing span squared to forewing area | Express the wing shape; related to flight speed and agility | Chai and Srygley 1990, Berwaerts et al. 2002 |
| FS | Food specialization | C | Amount of host plants used by immature stages | Express the food habit; lower values represent specialists and higher values represent more generalists species. | Graça et al. 2017 |
| Iridescence | Wing Iridescence | B | Presence or absence of iridescence coloration | Related with intra and interspecific visual recognition | Pinheiro et al. 2016, Spaniol et al. 2019 |
| Eyespots | Wing Eyespot | B | Presence or absence of wing eyespots | Related with defense strategies to avoid or deflect attacks of visually hunting predators | Stevens 2005, Olofsson et al. 2010 |

| | | | | | |
|---|---|---|---|---|---|
| Rings | Mimetic Ring | B | Member or not of mimetic rings complex | Indicate if species are a member of mimetic rings; related to Mullerian, Batesian or scape mimetic rings | Su et al. 2015, Spaniol et al. 2019 |
| Camouflage | Camouflage strategies | B | Colorations and shapes that resemble background or environmental structures | Related to capacity to avoid predators | Ruxton et al. 2004; Skelhorn et al. 2010 |
| Disruptive | Disruptive Coloration | B | Conspicuous colorations in the wing's periphery that disguises the body outline of the animal | Related to capacity to avoid predators, by preventing prey recognition | Schaefer and Stobbe 2006 |

*Incorporating imperfect detection in diversity measures: The Hidden Diversity framework*

To evaluate the magnitude of the effects of imperfect detection on diversity measures we developed an R function called *hidden.diversity* (HD) (Appendix C – Hidden diversity framework). This function returns, for each site *i*, the deviation of observed diversity from the estimated diversity, given imperfect detection, and this difference is divided by the standard deviation of the estimated diversity as follow:

$$\text{hidden.diversity}_i = \frac{\text{div.obs}_i - \overline{\text{div.est}_i}}{\text{sd.div.est}_i}$$

where $div.obs_i$ is the taxonomic, functional or phylogenetic diversity value obtained with observed count data for each site, $\overline{div.est_i}$ is the mean diversity value obtained from $N_{ik}$ posterior sampling in each site and $sd.div.est_i$ is the standard deviation of $div.est_i$. Positive and negative values of HD indicate, respectively, an overestimation and underestimation of observed diversity to estimated diversity values. Overestimation of diversity can only occur for phylogenetic or functional measures, since that species included can be functionally or phylogenetically redundant, and the N-mixture model only accounts for false negatives. However, distinct scenarios can generate positive or negatives HD values, and we disentangle these possible scenarios by plotting the relationship between observed and estimated diversity values (Fig. 1c). We called noise when observed and estimated diversity has the same signal, in other words, the observed pattern (overdispersion or clustering) does not change after corrected by imperfect detection, but still can be overestimated or underestimated in comparison with the estimated true diversity. On the other hand, a bias will occur if the observed and estimated diversity have opposite signs, and for these cases, an erroneous pattern in phylogenetic and/or functional structure of communities will be observed when undetected species are not considered.

The input of the *hidden.diversity* function is the observed community data, a phylogenetic tree, a matrix containing the mean traits for each species, and the matrix $N_{ik}$ estimated by the N-mixture model which represents the detection-corrected abundance. The function internally always calculates taxonomic diversity (TD) and abundance, and allows the user to choose the diversity metric that should be calculated – "*pd*" for branch length and "*mpd*" for distance-based approach - for both phylogenetic and functional diversity. The function will calculate the standardized effect size for phylogenetic diversity (SES.PD), and functional diversity (SES.FD), if only "*pd*" is informed, and the SES for phylogenetic and functional structure (SES.MPD and SES.MFD respectively) if only "*mpd*" is informed, or both if the user wishes. Also, the function allows indicating if there are binary data in the trait matrix, if the diversity measures should be weighted by abundance, the type of null model, the number of permutations used to calculate the null models, which aimed. Null models allow removing the effect of species richness on diversity measures by randomizing communities, permuting by permuting the positions of species in the phylogenetic tree or functional dendrograms, or by permuting the sampling units (rows) or species identities (columns) in the community matrix. Null models are implemented in the package *picante* (Kembel et al., 2010). The function output is a data frame containing SES values of diversity measures for each site (observed and estimated) and the value of hidden diversity.

We employed the HD for each diversity measure to evaluate differences between canopy and understory in the bias yielded by imperfect detection. For this, we performed a linear mixed model (LMM) using the HD values for each diversity measure as the response variable, the strata as a fixed predictor, and the sampling months and sites as random factors. We used the *lme4* package (v. 4.0.5, Bates, Mächler, Bolker, & Walker, 2015) to perform the LMM and the *ggplot2* package (v. 3.3.4 Wickham, 2016) to graphical visualization of the results.

## Results

Our database contained 35 species and 914 individuals of fruit-feeding butterflies. We found that canopy had lower community-level mean abundance than understory (values in the natural scale, $\mu_{\beta.\mathrm{can}} = 0.166$ CRI$_{95\%}$ = 0.008 to 0.104, $\mu_{\beta.\mathrm{und}} = 2.655$, CRI$_{95\%}$ = 0.001 to 0.117). Moreover, understory assemblages had a higher mean detection probability ($\mu_{\alpha.\mathrm{can}} = 0.032$, CRI$_{95\%}$ = 0.025 to 0.038, $\mu_{\alpha.\mathrm{und}} = 0.497$, CRI$_{95\%}$ = 0.033 to 0.964) (Fig. 2). We do not explore the effects of predictor variables on abundance and detection probability because these results are not crucial for this study, but the values for hyper-parameters for community-level are shown in Appendix A (Fig. A1 and A2, Table A1).

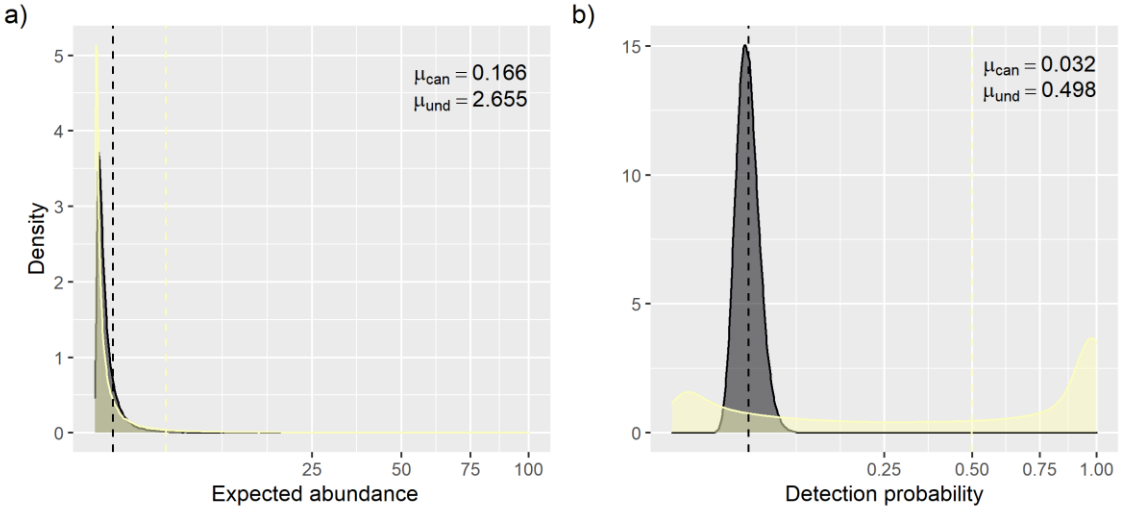

**Figure 2.** Community-mean distribution for expected abundance (a) and detection probability (b) for fruit-feeding butterflies sampled at FLONA-SFP, Southern Brazil.

These distributions were generated using the community hyperparameters for canopy ($\mu_{can}$ and $sd_{can}$, black color) and understory ($\mu_{und}$ and $sd_{und}$, light-yellow color). The dashed line represents the mean for each stratum (μ). We apply a square root transformation on the *x*-axis to better improve the visualization.

Hidden diversity (HD) demonstrated that there was an underestimation for both strata when only the species richness was evaluated (TD), and for this diversity measure, the HD differed between strata (Fig. 3a, Table 2). All other diversity measures tended to be overestimated (positive HD values). Phylogenetic and functional measures had opposite responses concerning the most overestimated stratum, while for phylogenetic measures understory was more overestimated than the canopy, for functional measures canopy tended to show higher overestimation than understory. Only for functional richness (SES.FD) we did not observe a difference in the HD between strata (Table 2). However, observing the relationship among observed and estimated diversity, we found that for most sites, the pattern of positive or negative SES value was maintained. This implies that, despite the error associated with not accounting for imperfect detection, for the fruit-feeding butterfly assembly, imperfect detection acts more like a noise than a bias (Fig. 3b).

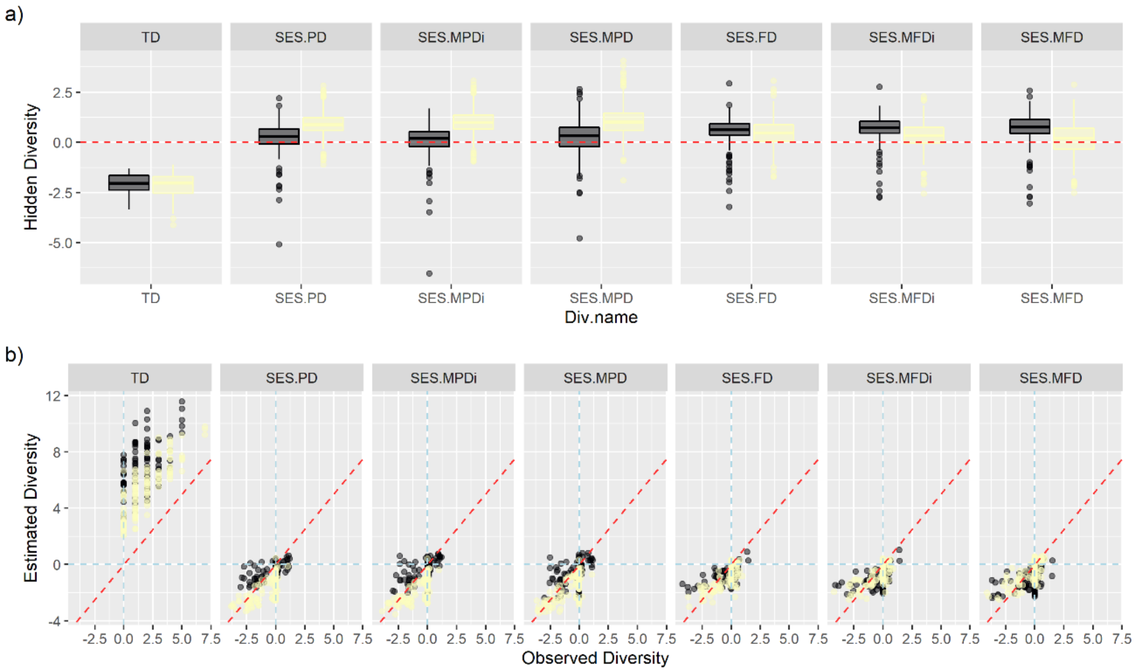

**Figure 3**. The effects of imperfect detection on multiple dimensions of biodiversity, evaluated by the hidden diversity framework for an assemblage of fruit-feeding butterflies sampled at FLONA-SFP, southern Brazil. a) Response of each stratum – Canopy (dark boxplots) and Understory (light yellow boxplots) – to the imperfect detection and their variation among the diversity measures. TD – taxonomic diversity, SES – standardized effect size, PD/FD – phylogenetic/functional diversity, MPD/MFD – abundance-based mean pairwise phylogenetic/functional distance, MPDi/MFDi – incidence-based mean pairwise phylogenetic/functional distance. The red dashed line indicates no difference in diversity value between observed and estimated data. b) Visual evaluation of the effect of the imperfect detection by sampling unit (points) and environmental factors (colors, dark – Canopy; yellow – Understory). Points above the dashed red line indicate an underestimate of the diversity and negative values of hidden diversity; points below this line indicate an overestimation of diversity and positive values of hidden diversity.

**Table 2**. Relationship of hidden diversity values for taxonomic, phylogenetic, and functional measures (HD.TD, HD.PD/MPD, HD.FD.MFD) and vertical stratification for the assemblage of fruit-feeding butterflies sampled at FLONA-SFP, southern Brazil. Bold values indicate a statistical significance at a threshold of 0.05. The asterisk indicates values of mean pairwise distances calculated with incidence instead of abundance.

| | Estimate | SE | t value | p |
|---|---|---|---|---|
| HD.TD | | | | |
| intercept | -2.064 | 0.163 | -12.679 | **0.000** |
| Slope | -0.099 | 0.046 | -2.163 | **0.031** |
| HD.PD | | | | |
| intercept | 0.174 | 0.090 | 1.931 | 0.092 |
| Slope | 0.728 | 0.090 | 8.060 | **0.000** |
| HD.FD | | | | |
| intercept | 0.509 | 0.064 | 7.955 | **0.000** |
| Slope | -0.032 | 0.090 | -0.349 | 0.727 |
| HD.MPDi* | | | | |
| intercept | 0.076 | 0.128 | 0.595 | 0.572 |
| Slope | 0.938 | 0.095 | 9.849 | **0.000** |
| HD.MPD | | | | |
| intercept | 0.226 | 0.177 | 1.279 | 0.240 |
| Slope | 0.848 | 0.104 | 8.164 | **0.000** |
| HD.MFDi* | | | | |
| intercept | 0.619 | 0.063 | 9.838 | **0.000** |
| Slope | -0.312 | 0.089 | -3.511 | **0.001** |
| HD.MFD | | | | |
| intercept | 0.680 | 0.094 | 7.212 | **0.000** |
| Slope | -0.564 | 0.101 | -5.563 | **0.000** |

## Discussion

Our results demonstrate that neglect imperfect detection can produce unrealistic estimates of diversity, which can be unbalanced between treatment levels or environmental gradients. Considering that several community studies are pattern-based, ignoring the effect of imperfect detection can leads to spurious interpretations of the mechanisms driving community assembly (Joseph et al., 2009), mainly when inversion in the observed pattern occurs (critical bias regions, Fig. 1c). For the assemblage of fruit-feeding butterflies studied, we found a noise for site scale (the majority of points are in the noise region, Fig. 3b), typically produced by the inclusion of redundant species at understory for phylogenetic measures, and redundant species at the canopy for functional measures. This occurs because the capacity to detect distinct lineages or functional traits in both strata was higher than the ability to detect new species (Jarzyna & Jetz, 2016), leading to an increase in phylogenetic or functional clustering in relation to the observed data. However, since there is a difference in the detection of individuals between strata (reached by hidden diversity), the relationship between diversity and the environment is biased.

Canopy and understory have distinct features including microclimatic conditions, forest structure, and resource availability (Grimbacher & Stork, 2007; Sobek, Tscharntke, Scherber, Schiele, & Steffan-Dewenter, 2009). Such differences are commonly associated with the vertical stratification of animal taxa, especially for insects (Ashton et al., 2016; Basset et al., 2015). For fruit-feeding butterflies, is recognized that some families or tribes are more related with canopy or understory (DeVries et al., 2012; Hill, Hamer, Tangah, & Dawood, 2001), including the detection probability of species can differ among strata (Ribeiro et al., 2016). In addition to the lack of studies evaluating phylogenetic and functional diversity for this group, for the

Neotropical region, there is no clear pattern as to which is the most diverse stratum from a taxonomic perspective (understory - Araujo et al., 2020; Barlow, Overal, Araujo, Gardner, & Peres, 2007; Lourenço et al., 2019; Whitworth et al., 2016; canopy - Devries, 1988; DeVries et al., 2012; Ribeiro & Freitas, 2012; Santos et al., 2017). In our study, we show that there was a large underestimation in species richness, providing evidence that there is a bias for observed taxonomic diversity in canopies sites. This was the only case where there was an inversion in the observed pattern: understory was richest than canopy employing the observed data, but the canopy has a higher richness than understory when we used the estimated data (Appendix C – Fig. C1, Table C1). For phylogenetic measures, despite the difference in HD values between stratum, the observed pattern was maintained and only the magnitude of the effect was adjusted. However, for functional measures based on distances (SES.MFD), the inclusion of undetected individuals revealed a significant difference (understory was more diverse than canopy), unobserved when only observed data was used (Appendix C – Table C1).

As expected, the inclusion of undetected species had a larger effect on taxonomic diversity measures than on phylogenetic or functional ones. While for taxonomic diversity, each undetected species leads to an increment of the estimated diversity value, for phylogenetic and functional measures undetected species may be redundant, i.e., contain evolutionary or functional information, respectively, that was already covered in the observed data. Furthermore, we observed that the understory had a large number of species belonging to the same lineage that were undetected in the field. Generally, fruit-feeding butterflies that inhabit the understory belong to Satyrinae (particularly to the tribes Morphini and Brassolini). These species tend to be more abundant during the summer months (December to February in Southern Hemisphere) (Iserhard et al., 2017), and hence more individuals are available to be detected. But at the beginning or end of this season, a smaller number of individuals are active, hindering its detection. Such features could explain the clustered pattern observed in the understory when we include imperfect detection to perform phylogenetic measures. Similarly, a clustered pattern was revealed for functional measures for canopy. Species that occupy this stratum generally exhibit traits related to flight performance (Chai & Srygley, 1990; Graça et al., 2017), given high mobility to looking for resources and favorable conditions (Shahabuddin & Ponte, 2005). Thus, a simple explanation for the inclusion of redundant traits in the canopy could be that individuals were absent because they were visiting a part of their home area that was not covered by the survey (Joseph et al., 2009; Ribeiro et al., 2016). Future investigations should be conducted in this context to understand whether high flight mobility leads to a low probability of butterfly detection.

Biodiversity measures are important tools to guide species conservation decisions, as well as to infer about the ecological and evolutionary process that structure assemblages. Since accounting for imperfect detection improves the accuracy of estimates of diversity patterns, in some circumstances, it is strongly recommended (Fig. 1), because it may lower the risk of erroneously inferring biological processes that are implied by sampling uncertainty (Joseph et al., 2009). Several models have been proposed in recent years to incorporate imperfect detection in order to improve the efficiency of estimating parameters in community studies (Abrams, Sollmann, Mitchell, Struebig, & Wilting, 2021; Broms et al., 2015; Frishkoff et al., 2017; Jarzyna & Jetz, 2016; Tingley, Nadeau, & Sandor, 2020; Zipkin et al., 2010). Further, these models allow us to propagate the uncertainty in species-specific detectability to biodiversity measures, as we demonstrated here. We expect that the framework developed in this study help researchers to better understand and describe diversity patterns and the mechanisms that assembly ecological communities.

**Acknowledgments** – We thank anonymous reviewers that made important suggestions to improve our manuscript. AR thanks Viviane Zulian and Gonçalo Ferraz for discussions about hierarchical models. AR thanks colleagues that helped with the butterfly surveys. The authors thank the staff of Floresta Nacional de São Francisco de Paula for providing logistic assistance during the sampling periods. AR research activities have been supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (graduate fellowship). LSD research activities have been supported by CNPq Productivity Fellowship (grant 307527/2018-2). AR, GN, CAI and LSD are members of the National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, supported by MCTIC/CNPq (proc. 465610/2014-5) and FAPEG (proc. 201810267000023).

**Authors' contributions** – AR, CAI and LSD conceived the ideas of the study and AR wrote the first draft of the manuscript. AR and CAI collected the data. AR and GN performed the analyses and data simulations. AR, GN and LSD developed the *hidden.diversity* framework. All authors contributed critically to the drafts and gave final approval for publication.

**Conflicts of interest** – The authors declare that they have no conflict of interest.

**Data Availability Statement** – The code and all necessary data to perform all analyses made in this manuscript, as well the code for the hidden diversity framework is available on Zenodo:10.5281/zenodo.5132227.

## References

Abrams, J. F., Sollmann, R., Mitchell, S. L., Struebig, M. J., & Wilting, A. (2021). Occupancy-based diversity profiles: capturing biodiversity complexities while accounting for imperfect detection. *Ecography*, 1–12. doi: 10.1111/ecog.05577

Araujo, P. F., Freitas, A. V. L., Souza, G. A. De, & Ribeiro, D. B. (2020). Studies on Neotropical Fauna and Environment Vertical stratification on a small scale : the distribution of fruit-feeding butterflies in a semi- deciduous Atlantic forest in Brazil. *Studies on Neotropical Fauna and Environment*, *00*(00), 1–30. doi: 10.1080/01650521.2020.1728033

Ashton, L. A., Nakamura, A., Basset, Y., Burwell, C. J., Cao, M., Eastwood, R., … Kitching, R. L. (2016). Vertical stratification of moths across elevation and latitude. *Journal of Biogeography*, *43*(1), 59–69. doi: 10.1111/jbi.12616

Barlow, J., Overal, W. L., Araujo, I. S., Gardner, T. A., & Peres, C. A. (2007). The value of primary, secondary and plantation forests for fruit-feeding butterflies in the Brazilian Amazon. *Journal of Applied Ecology*, *44*(5), 1001–1012. doi: 10.1111/j.1365-2664.2007.01347.x

Basset, Y., Cizek, L., Cuénoud, P., Didham, R. K., Novotny, V., Ødegaard, F., … Leponce, M. (2015). Arthropod distribution in a tropical rainforest: Tackling a four dimensional puzzle. *PLoS ONE*, *10*(12), 1–22. doi: 10.1371/journal.pone.0144110

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). doi:

10.18637/jss.v067.i01

Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E., & Pollock, K. H. (1998). Estimating species richness: The importance of heterogeneity in species detectability. *Ecology*, *79*(3), 1018–1028. doi: 10.1890/0012-9658(1998)079[1018:ESRTIO]2.0.CO;2

Broms, K. M., Hooten, M. B., & Fitzpatrick, R. M. (2015). Accounting for imperfect detection in Hill numbers for biodiversity studies. *Methods in Ecology and Evolution*, *6*(1), 99–108. doi: 10.1111/2041-210X.12296

Brown, K. S., & Freitas, A. V. L. (2000). Atlantic Forest Butterflies: Indicators for Landscape Conservation. *Biotropica*, *32*(4b), 934–956. doi: 10.1111/j.1744-7429.2000.tb00631.x

Brown, K. S. J. (2005). Geologic, evolutionary, and ecological bases of the diversification of neotropical butterflies: implications for conservation. *Tropical Rainforests: Past, Present, and Future*, pp. 166–201.

Chai, P., & Srygley, R. B. (1990). Predation and the flight, morphology, and temperature of Neotropical rain-forest butterflies. *American Naturalist*, *135*(6), 748–765. doi: 10.1086/285072

Chai, Peng, & Srygley, R. B. (1990). Predation and the Flight, Morphology, and Temperature of Neotropical Rain-Forest Butterflies. *The American Naturalist*, *135*(6), 748–765.

Chao, A., Chiu, C. H., Colwell, R. K., Magnago, L. F. S., Chazdon, R. L., & Gotelli, N. J. (2017). Deciphering the enigma of undetected species, phylogenetic, and functional diversity based on Good-Turing theory. *Ecology*, *98*(11), 2914–2929. doi: 10.1002/ecy.2000

Chazot, N., Wahlberg, N., Freitas, A. V. L., Mitter, C., Labandeira, C., Sohn, J. C., …

Heikkilä, M. (2019). Priors and Posteriors in Bayesian Timing of Divergence Analyses: The Age of Butterflies Revisited. *Systematic Biology*, *68*(5), 797–813. doi: 10.1093/sysbio/syz002

de Bello, F., Berg, M. P., Dias, A. T. C., Diniz-Filho, J. A. F., Götzenberger, L., Hortal, J., … Lepš, J. (2015). On the need for phylogenetic ‘corrections’ in functional trait-based approaches. *Folia Geobotanica*, *50*(4), 349–357. doi: 10.1007/s12224-015-9228-6

Devarajan, K., Morelli, T. L., & Tenan, S. (2020). Multi-species occupancy models: review, roadmap, and recommendations. *Ecography*, *43*(11), 1612–1624. doi: 10.1111/ecog.04957

Devries, P. J. (1988). Stratification of fruit-feeding nymphalid butterflies in a Costa Rican rainforest. *Journal of Research on the Lepidoptera*, *26*(1–4), 98–108.

DeVries, P. J., Alexander, L. G., Chacon, I. A., & Fordyce, J. A. (2012). Similarity and difference among rainforest fruit-feeding butterfly communities in Central and South America. *Journal of Animal Ecology*, *81*(2), 472–482. doi: 10.1111/j.1365-2656.2011.01922.x

Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the Effects of Habitat and Biological Interactions in an Avian Community. *PLOS ONE*, *10*(8), e0135987. doi: 10.1371/journal.pone.0135987

Duarte, L. D. S., Debastiani, V. J., Carlucci, M. B., & Diniz-Filho, J. A. F. (2018). Analyzing community-weighted trait means across environmental gradients: should phylogeny stay or should it go? *Ecology*, *99*(2), 385–398. doi: 10.1002/ecy.2081

Dudley, R. (2002). Mechanisms and implications of animal flight maneuverability. *Integrative and Comparative Biology*, *42*(1), 135–140. doi: 10.1093/icb/42.1.135

Fordyce, J. A., & DeVries, P. J. (2016). A tale of two communities: Neotropical butterfly assemblages show higher beta diversity in the canopy compared to the understory. *Oecologia*, *181*(1), 235–243. doi: 10.1007/s00442-016-3562-0

Freitas, André V. L., Santos, J. P., Rosa, A. H. B., Iserhard, C. A., Richter, A., Siewert, R. R., … Lourenço, G. M. (2021). Sampling Methods for Butterflies (Lepidoptera). In *Measuring Arthropod Biodiversity* (pp. 101–123). Cham: Springer International Publishing. doi: 10.1007/978-3-030-53226-0_5

Freitas, André Victor Lucci, Iserhard, C. A., Santos, J. P. dos, Carreira, J. Y. O., Ribeiro, D. B., Melo, D. H. A., … Uehara-Prado, M. (2014). Studies with butterfly bait traps: An overview. *Revista Colombiana de Entomología*, *40*(2), 203–212. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-04882014000200013&lng=en&nrm=iso&tlng=en

Frishkoff, L. O., de Valpine, P., & M’Gonigle, L. K. (2017). Phylogenetic occupancy models integrate imperfect detection and phylogenetic signal to analyze community structure. *Ecology*, *98*(1), 198–210. doi: 10.1002/ecy.1631

Frishkoff, L. O., De Valpine, P., & M’Gonigle, L. K. (2017). Phylogenetic occupancy models integrate imperfect detection and phylogenetic signal to analyze community structure. *Ecology*, *98*(1), 198–210. doi: 10.1002/ecy.1631

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. doi: 10.1214/ss/1177011136

Graça, M. B., Pequeno, P. A. C. L., Franklin, E., & Morais, J. W. (2017). Coevolution between flight morphology, vertical stratification and sexual dimorphism: what can we learn from tropical butterflies? *Journal of Evolutionary Biology*, *30*(10), 1862–1871. doi: 10.1111/jeb.13145

Graham, C. H., & Fine, P. V. A. (2008). Phylogenetic beta diversity: Linking ecological and evolutionary processes across space in time. *Ecology Letters*, *11*(12), 1265–1277. doi: 10.1111/j.1461-0248.2008.01256.x

Grimbacher, P. S., & Stork, N. E. (2007). Vertical stratification of feeding guilds and body size in beetle assemblages from an Australian tropical rainforest. *Austral Ecology*, *32*(1), 77–85. doi: 10.1111/j.1442-9993.2007.01735.x

Guillera-Arroita, G., Kéry, M., & Lahoz-Monfort, J. J. (2019). Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation. *Ecology and Evolution*, *9*(2), 780–792. doi: 10.1002/ece3.4821

Hill, J., Hamer, K., Tangah, J., & Dawood, M. (2001). Ecology of tropical butterflies in rainforest gaps. *Oecologia*, *128*(2), 294–302. doi: 10.1007/s004420100651

ICMBio, I. C. M. de C. da B. *PLANO DE MANEJO DA FLORESTA NACIONAL DE SÃO FRANCISCO DE PAULA.* , (2020).

Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology and Evolution*, *29*(2), 97–106. doi: 10.1016/j.tree.2013.10.012

Iserhard, C. A., Romanowski, H. P., Richter, A., & Mendonça, M. de S. (2017). Monitoring Temporal Variation to Assess Changes in the Structure of Subtropical Atlantic Forest Butterfly Communities. *Environmental Entomology*, *46*(4), 804–813. doi: 10.1093/ee/nvx115

Jarzyna, M. A., & Jetz, W. (2016). Detecting the Multiple Facets of Biodiversity. *Trends in Ecology & Evolution*, *31*(7), 527–538. doi: 10.1016/j.tree.2016.04.002

Joseph, L. N., Elkin, C., Martin, T. G., & Possingham, H. P. (2009). Modeling abundance using N-mixture models: The importance of considering ecological mechanisms. *Ecological Applications*, *19*(3), 631–642. doi: 10.1890/07-2107.1

Kellner, K. (2021). *jagsUI: A Wrapper Around "rjags" to Streamline "JAGS" Analyses*. Retrieved from https://cran.r-project.org/package=jagsUI

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., … Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. doi: 10.1093/bioinformatics/btq166

Kéry, M., & Plattner, M. (2007). Species richness estimation and determinants of species detectability in butterfly monitoring programmes. *Ecological Entomology*, *32*(1), 53–61. doi: 10.1111/j.1365-2311.2006.00841.x

Kéry, M., & Royle, J. A. (2016). *Applied Hierarchical Modeling in Ecology Analysis of distribution, abundance and species richness in R and BUGS*. Elsevier. doi: 10.1016/B978-0-12-801378-6.01001-8

Kéry, M., Royle, J. A., & Schmid, H. (2005). MODELING AVIAN ABUNDANCE FROM REPLICATED COUNTS USING BINOMIAL MIXTURE MODELS. *Ecological Applications*, *15*(4), 1450–1461. doi: 10.1890/04-1120

Lourenço, G. M., Soares, G. R., Santos, T. P., Dáttilo, W., Freitas, A. V. L., & Ribeiro, S. P. (2019). Equal but different: Natural ecotones are dissimilar to anthropic edges. *PLoS ONE*, *14*(3). doi: 10.1371/journal.pone.0213008

Marini-Filho, O. J., & Martins, R. P. (2010). Nymphalid butterfly dispersal among forest fragments at Serra da Canastra National Park, Brazil. *Journal of Insect Conservation*, *14*(4), 401–411. doi: 10.1007/s10841-010-9271-9

Molleman, F., Alphen, M. E. Van, Brake, P. M., & Zwaan, B. J. (2005). Preferences and Food Quality of Fruit-Feeding Butterflies in Kibale Forest, Uganda 1. *Africa*, *37*(4), 657–663.

New, T. R. (1997). Are Lepidoptera an effective "umbrella group" for biodiversity conservation? *Journal of Insect Conservation*, *1*(1), 5–12. doi:

10.1023/A:1018433406701

Olofsson, M., Vallin, A., Jakobsson, S., & Wiklund, C. (2010). Marginal Eyespots on Butterfly Wings Deflect Bird Attacks Under Low Light Intensities with UV Wavelengths. *PLoS ONE*, *5*(5), e10798. doi: 10.1371/journal.pone.0010798

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526–528. doi: 10.1093/bioinformatics/bty633

Pillar, V. D., & Duarte, L. D. S. (2010). A framework for metacommunity analysis of phylogenetic structure. *Ecology Letters*, *13*(5), 587–596. doi: 10.1111/j.1461-0248.2010.01456.x

Pinheiro, C. E. G., Freitas, A. V. L., Campos, V. C., Devries, P. J., & Penz, C. M. (2016). Both Palatable and Unpalatable Butterflies Use Bright Colors to Signal Difficulty of Capture to Predators. *Neotropical Entomology*, *45*(2), 107–113. doi: 10.1007/s13744-015-0359-5

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Ribeiro, Danilo B., & Freitas, A. V. L. (2012). The effect of reduced-impact logging on fruit-feeding butterflies in Central Amazon, Brazil. *Journal of Insect Conservation*, *16*(5), 733–744. doi: 10.1007/s10841-012-9458-3

Ribeiro, Danilo Bandini, Williams, M. R., Specht, A., & Freitas, A. V. L. (2016). Vertical and temporal variability in the probability of detection of fruit-feeding butterflies and moths (Lepidoptera) in tropical forest. *Austral Entomology*, *55*(1), 112–120. doi: 10.1111/aen.12157

Roth, T., Allan, E., Pearman, P. B., & Amrhein, V. (2018). Functional ecology and imperfect detection of species. *Methods in Ecology and Evolution*, *9*(4), 917–928. doi: 10.1111/2041-210X.12950

Ruiz-Gutiérrez, V., Zipkin, E. F., & Dhondt, A. A. (2010). Occupancy dynamics in a tropical bird community: unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology*, *47*(3), 621–630. doi: 10.1111/j.1365-2664.2010.01811.x

Rutowski, R. L. (2000). Variation of eye size in butterflies: inter- and intraspecific patterns. *Journal of Zoology*, *252*(2), 187–195. doi: 10.1111/j.1469-7998.2000.tb00614.x

Santos, J. P. Dos, Iserhard, C. A., Carreira, J. Y. O., & Freitas, A. V. L. (2017). Monitoring fruit-feeding butterfly assemblages in two vertical strata in seasonal Atlantic Forest: Temporal species turnover is lower in the canopy. *Journal of Tropical Ecology*, *33*(5), 345–355. doi: 10.1017/S0266467417000323

Schaefer, H. M., & Stobbe, N. (2006). Disruptive coloration provides camouflage independent of background matching. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1600), 2427–2432. doi: 10.1098/rspb.2006.3615

Schulze, C. H., Linsenmair, K. E., & Fiedler, K. (2001). Understorey versus canopy: patterns of vertical stratification and diversity among Lepidoptera in a Bornean rain forest. *Plant Ecology*, *153*(1), 133–152. doi: 10.1023/A:1017589711553

Sekar, S. (2012). A meta-analysis of the traits affecting dispersal ability in butterflies: Can wingspan be used as a proxy? *Journal of Animal Ecology*, *81*(1), 174–184. doi: 10.1111/j.1365-2656.2011.01909.x

Shahabuddin, G., & Ponte, C. A. (2005). Frugivorous butterfly species in tropical forest fragments: correlates of vulnerability to extinction. *Biodiversity and Conservation*, *14*(5), 1137–1152. doi: 10.1007/s10531-004-7842-3

Si, X., Cadotte, M. W., Zhao, Y., Zhou, H., Zeng, D., Li, J., … Tingley, M. W. (2018). The importance of accounting for imperfect detection when estimating functional and phylogenetic community structure. *Ecology, 99*(9), 2103–2112. doi: 10.1002/ecy.2438

Skelhorn, J., Rowland, H. M., Speed, M. P., & Ruxton, G. D. (2010). Masquerade: Camouflage Without Crypsis. *Science, 327*(5961), 51–51. doi: 10.1126/science.1181931

Sobek, S., Tscharntke, T., Scherber, C., Schiele, S., & Steffan-Dewenter, I. (2009). Canopy vs. understory: Does tree diversity affect bee and wasp communities and their natural enemies across forest strata? *Forest Ecology and Management, 258*(5), 609–615. doi: 10.1016/j.foreco.2009.04.026

Sonego, R. C., Backes, A., & Souza, A. F. (2007). Descrição da estrutura de uma Floresta Ombrófila Mista, RS, Brasil, utilizando estimadores não-paramétricos de riqueza e rarefação de amostras. *Acta Botânica Brasílica, 21*(4), 943–955. doi: 10.1590/S0102-33062007000400019

Spaniol, R. L., Duarte, L. da S., Mendonça, M. de S., & Iserhard, C. A. (2019). Combining functional traits and phylogeny to disentangling Amazonian butterfly assemblages on anthropogenic gradients. *Ecosphere, 10*(8). doi: 10.1002/ecs2.2837

Stevens, M. (2005). The role of eyespots as anti-predator mechanisms, principally demonstrated in the Lepidoptera. *Biological Reviews, 80*(4), 573–588. doi: 10.1017/S1464793105006810

Su, S., Lim, M., & Kunte, K. (2015). Prey from the eyes of predators: Color discriminability of aposematic and mimetic butterflies from an avian visual perspective. *Evolution, 69*(11), 2985–2994. doi: 10.1111/evo.12800

Thomas, J. . (2005). Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1454), 339–357. doi: 10.1098/rstb.2004.1585

Tingley, M. W., Nadeau, C. P., & Sandor, M. E. (2020). Multi-species occupancy models as robust estimators of community richness. *Methods in Ecology and Evolution*, *11*(5), 633–642. doi: 10.1111/2041-210X.13378

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., … Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, *92*(2), 698–715. doi: 10.1111/brv.12252

Turlure, C., Schtickzelle, N., Van Dyck, H., Seymoure, B., & Rutowski, R. (2016). Flight Morphology, Compound Eye Structure and Dispersal in the Bog and the Cranberry Fritillary Butterflies: An Inter- and Intraspecific Comparison. *PLOS ONE*, *11*(6), e0158073. doi: 10.1371/journal.pone.0158073

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505. doi: 10.1146/annurev.ecolsys.33.010802.150448

Whitworth, A., Villacampa, J., Brown, A., Huarcaya, R. P., Downie, R., & MacLeod, R. (2016). Past human disturbance effects upon biodiversity are greatest in the canopy; A case study on rainforest butterflies. *PLoS ONE*, *11*(3). doi: 10.1371/journal.pone.0150520

Wickham, H. (2016). *ggplot2*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-24277-4

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, *19*(12), 639–644. doi:

10.1016/j.tree.2004.09.011

Yamaura, Y., Andrew Royle, J., Kuboi, K., Tada, T., Ikeno, S., & Makino, S. (2011). Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology*, *48*(1), 67–75. doi: 10.1111/j.1365-2664.2010.01922.x

Yamaura, Y., Kéry, M., & Andrew Royle, J. (2016). Study of biological communities subject to imperfect detection: bias and precision of community N-mixture abundance models in small-sample situations. *Ecological Research*, *31*(3), 289–305. doi: 10.1007/s11284-016-1340-4

Yamaura, Y., Royle, J. A., Shimada, N., Asanuma, S., Sato, T., Taki, H., & Makino, S. (2012). Biodiversity of man-made open habitats in an underused country: a class of multispecies abundance models for count data. *Biodiversity and Conservation*, *21*(6), 1365–1380. doi: 10.1007/s10531-012-0244-z

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, *16*(8), 446–453. doi: 10.1016/S0169-5347(01)02205-4

Zipkin, E. F., Andrew Royle, J., Dawson, D. K., & Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation*, *143*(2), 479–484. doi: 10.1016/j.biocon.2009.11.016

## Appendix A – N-Mixture Model

Code for the hierarchical Bayesian N-Mixture model (Static_Nmixture_P_RE.txt) and visual evaluation of parameters. This model is a modification of the proposed model for The Swiss Breeding Bird Survey (MHB 2014) in the Chapter 11 of the Applied Hierarchical Modelling in Ecology (Kéry & Royle, 2016). Furthermore, we assumes that the species-level effects can varies between strata, according whit Zipkin, Andrew Royle, Dawson, & Bates (2010). The code for the model is available on [https://github.com/richterbine/TheHiddenSideofDiversity/tree/main/R/Bayesian_models](https://github.com/richterbine/TheHiddenSideofDiversity/tree/main/R/Bayesian_models)

```
model {
  # Community priors (with hyperparameters) for species-specific parameters
  for(k in 1:nspec){
    beta.can[k] ~ dnorm(mu.beta.can, tau.beta.can)     # Species-specific random intercept for abundance - canopy
    beta.und[k] ~ dnorm(mu.beta.und, tau.beta.und)     # Species-specific random intercept for abundance - understory
    alpha.can[k] ~ dnorm(mu.alpha.can, tau.alpha.can)  # Species-specific random intercept for detectability - canopy
    alpha.und[k] ~ dnorm(mu.alpha.und, tau.alpha.und)  # Species-specific random intercept for detectability -
understory
```

```
    beta1[k] ~ dnorm(mu.beta1, tau.beta1)         # Species-specific random slope for temperature in biologic process
    alpha1[k] ~ dnorm(mu.alpha1, tau.alpha1)    # Species-specific random slope for sampling process in sampling process
    alpha2[k] ~ dnorm(mu.alpha2, tau.alpha2)    # Species-specific random slope for temperature in sampling process

    for(n in 1:5) {
      month[n, k] ~ dnorm(0, tau.month)
    }
    for(n in 1:6){
      area[n, k] ~ dnorm(0, tau.area)
    }
  }
  # Hyperpriors for community hyperparameters
# abundance model – intercept
  mu.beta.can ~ dnorm(0, 0.001)
  tau.beta.can <- pow(sd.beta.can, -2)
  sd.beta.can ~ dunif(0, 10)
```

```
 mu.beta.und ~ dnorm(0, 0.001)
 tau.beta.und <- pow(sd.beta.und, -2)
 sd.beta.und ~ dunif(0, 10)

# abundance model - slope for temperature by sites
 mu.beta1 ~ dnorm(0, 0.001)
 tau.beta1 <- pow(sd.beta1, -2)
 sd.beta1 ~ dunif(0, 10)

# detection model - intercept
 mu.alpha.can.pre ~ dunif (0, 1)  # Detection can have any value between 0 and 1 with equal probability
 mu.alpha.can <- logit(mu.alpha.can.pre)  # Inverse logit – values from -inf to inf as in norm
 tau.alpha.can <- pow(sd.alpha.can, -2)
 sd.alpha.can ~ dunif(0, 10)

 mu.alpha.und.pre ~ dunif(0, 1)
 mu.alpha.und <- logit(mu.alpha.und.pre)
```

```
  tau.alpha.und <- pow(sd.alpha.und, -2)
  sd.alpha.und ~ dunif(0, 10)
# detection model - slope for julian dates
  mu.alpha1 ~ dnorm(0, 0.001)
  tau.alpha1 <- pow(sd.alpha1, -2)
  sd.alpha1 ~ dunif(0, 10)

  # detection model - slope for temperatures per day
  mu.alpha2 ~ dnorm(0, 0.001)
  tau.alpha2 <- pow(sd.alpha2, -2)
  sd.alpha2 ~ dunif(0, 10)

  tau.month <- pow(sd.month, -2)
  sd.month ~ dunif(0, 10)

  tau.area <- pow(sd.area, -2)
  sd.area ~ dunif(0, 10)
```

```
  # Ecological model for true abundance (process model)

  for(k in 1:nspec) {

    for (i in 1:nsite) {

      N[i,k] ~ dpois(lambda[i,k])        # latent abundance of each species in each site

      log(lambda[i,k]) <- beta.can[k] * (1 - Strata[i]) + beta.und[k] * Strata[i] + beta1[k] *
Temp[i] + month[Month[i], k] + area[Area[i], k]

    }

  }


  # Observation model for replicated counts

  for(k in 1:nspec) {

    for (i in 1:nsite) {

      for (j in 1:nrep) {

        yc[i,j,k] ~ dbin(p[i,j,k], N[i,k])

        logit(p[i,j,k]) <- alpha.can[k] * (1 - Strata[i]) + alpha.und[k] * Strata[i] + alpha1[k]
* Date[i, j] + alpha2[k] * Temp_det[i, j]
```

```
        }
      }
    }

    # Other derived quantities
    for(k in 1:nspec) {
      mlambda.can[k] <- exp(beta.can[k])      # Expected abundance on natural scale for canopy
      mlambda.und[k] <- exp(beta.und[k])      # Expected abundance on natural scale for understory
      logit(mp.can[k]) <- alpha.can[k]        # Mean detection on natural scale for canopy
      logit(mp.und[k]) <- alpha.und[k]        # Mean detection on natural scale for understory
    }
  }
```

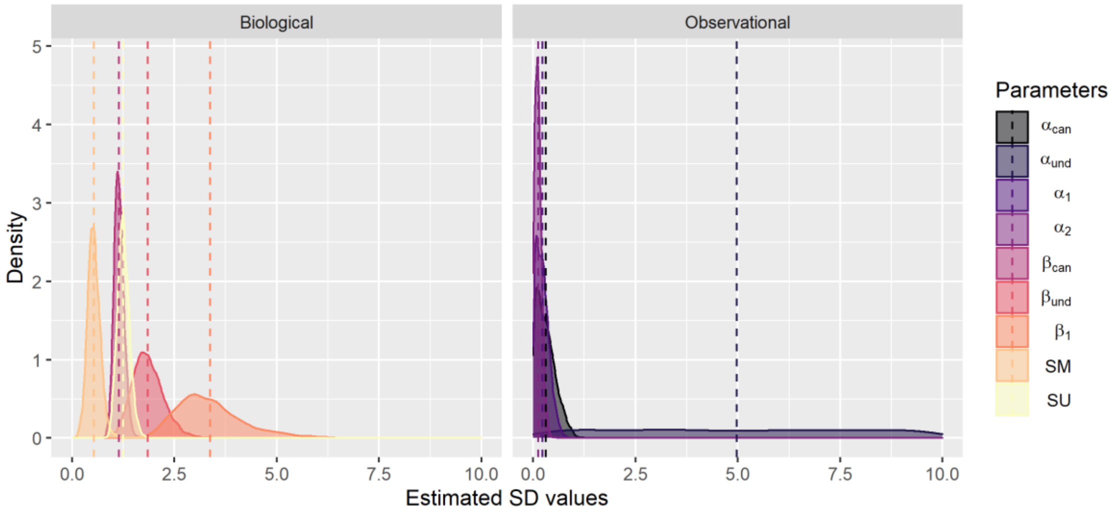

**Figure A 1**. Community distribution of the standard deviation for mean parameters, estimated by N-mixture models for biological and observational processes, using the community data for fruit-feeding butterflies sampled at FLONA-SFP, Southern Brazil.

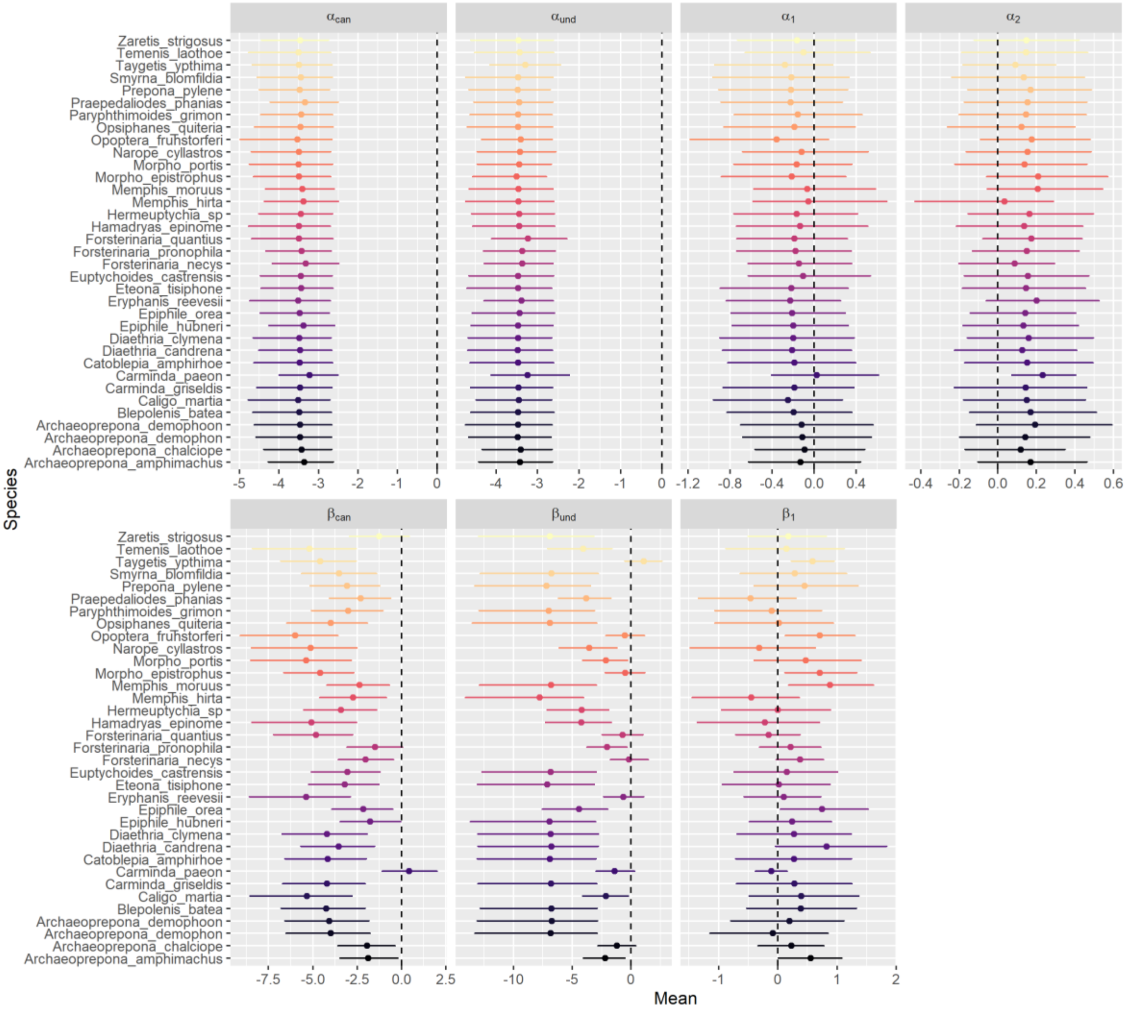

**Figure A 2**. Specie-specific mean parameters were estimated by the N-mixture model for detection probability (upper panel) and abundance (lower panel). Points represent the estimated parameter for each species and horizontal bars indicate the 95% of credible interval (CRI). If the CRI overlaps the zero (vertical dashed line), we conclude that there was no effect of the predictor in species-specific responses. For detection probability, sampling day ($\alpha_1$) does not affect the detection of any species, while the temperature at trap base ($\alpha_2$) only affected positively the species *Carminda paeon* (Godart, 1824). For expected abundance, temperature ($\beta_1$) has a positive effect for *Taygetis ypthima* Hübner, [1821], *Opoptera fruhstorferi* (Röber, 1896), *Morpho epistrophus* (Fabricius, 1796), *Memphis moruus* (Fabricius, 1775), *Epiphile orea* (Hübner, [1823]).

**Table A1**. The mean and the 95% Bayesian credible interval (CRI) for community-level summaries of the hyper-parameters for expected abundance (β's) and detection probability (α's) estimated by the N-mixture model for fruit-feeding butterflies sampled in FLONA-SFP, southern Brazil. Asterisk represents hyper-parameters that overlap zero.

| | Mean | Low CRI | Upper CRI |
|---|---|---|---|
| $\mu\ \beta_{can}$ | -3.4991 | -4.4792 | -2.5917 |
| SD $\beta_{can}$ | 1.8470 | 1.1903 | 2.6484 |
| $\mu\ \beta_{und}$ | -4.4248 | -6.3377 | -2.9100 |
| SD $\beta_{und}$ | 3.3726 | 2.1065 | 5.3857 |
| $\mu\ \beta_1$ | 0.2237 | -0.0740 | 0.4975* |
| SD $\beta_1$ | 0.5334 | 0.2564 | 0.8592 |
| $\mu\alpha_{can}$ | -3.4379 | -4.0621 | -2.9272 |
| SD $\alpha_{can}$ | 0.3105 | 0.0097 | 0.8413 |
| $\mu\ \alpha_{und}$ | -0.0359 | -3.9749 | 3.6357* |
| SD $\alpha_{und}$ | 4.9700 | 0.2257 | 9.7353 |
| $\mu\alpha_1$ | -0.1693 | -0.4526 | 0.1057* |
| SD $\alpha_1$ | 0.2276 | 0.0118 | 0.5851 |
| $\mu\alpha_2$ | 0.1520 | 0.0094 | 0.2878 |
| SD $\alpha_2$ | 0.1233 | 0.0048 | 0.3062 |
| SD SM | 1.2553 | 0.9940 | 1.5692 |
| SD SU | 1.1465 | 0.9365 | 1.3983 |

## Appendix B – Model Validation

Visual evaluation of the performance of the N-mixture model built. We used 12 simulated communities, which we vary parameters related to biological process ($\beta_1$ and $\beta_2$) and observational process ($\alpha_1$ and $\alpha_2$). We consider that model is valid if the values of the real parameters are within the posterior distribution retuned by the N-mixture model. The function for simulate metacommunities is a modified version of the *simComm* function developed by Kéry & Royle (2016) and available in the AHMBook package (v. 0.2.2). The code to simulate metacommunities are available on https://github.com/richterbine/TheHiddenSideofDiversity/tree/main/R/Model_validation

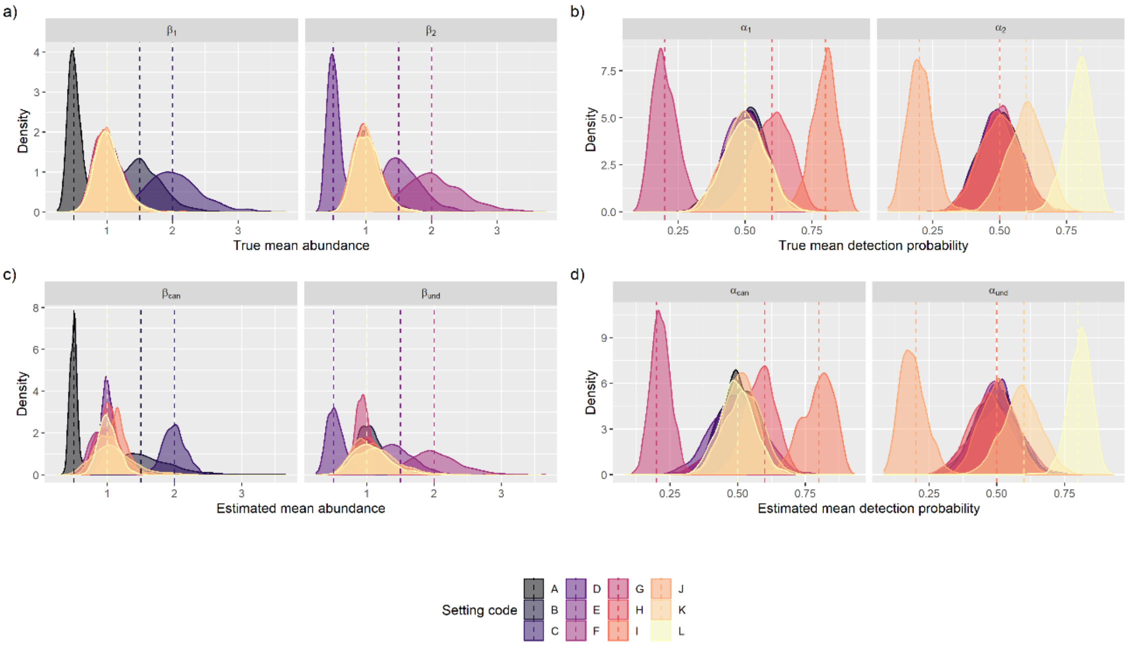

**Figure B3**. Community distribution of mean abundance and detection probability for simulated communities. a) and b) are the distribution using the true values for intercept parameters ($\mu_{\beta1}$, $sd_{\beta1}$, $\mu_{\beta2}$, $sd_{\beta2}$, $\mu_{\alpha1}$, $sd_{\alpha1}$, $\mu_{\alpha2}$, $sd_{\alpha2}$). c) and d) are the distribution using the mean values estimated by the Bayesian model ($\mu_{\beta can}$, $sd_{\beta can}$, $\mu_{\alpha can}$, $sd_{\alpha can}$, $\mu_{\beta und}$, $sd_{\beta und}$, $\mu_{\alpha und}$, $sd_{\alpha und}$). The color indicates the 12 scenarios simulated, when we vary the values for $\mu_{\beta1}$, $\mu_{\beta2}$, $\mu_{\alpha1}$, $\mu_{\alpha2}$ for three levels: $\mu_\beta = 0.5, 1.5, 2.0$, the default of function is 1.0; $\mu_\alpha = 0.2, 0.6, 0.8$, the default of function is 0.5. We allow only one parameter to vary at a time, while others remained fixed to simulate the community. For A, B, and C the $\mu_{\beta1}$ varies, for D, E, and F the $\mu_{\beta2}$ varies, for G, H, and I the $\mu_{\alpha1}$ varies and for J, K and L the $\mu_{\alpha2}$ varies.

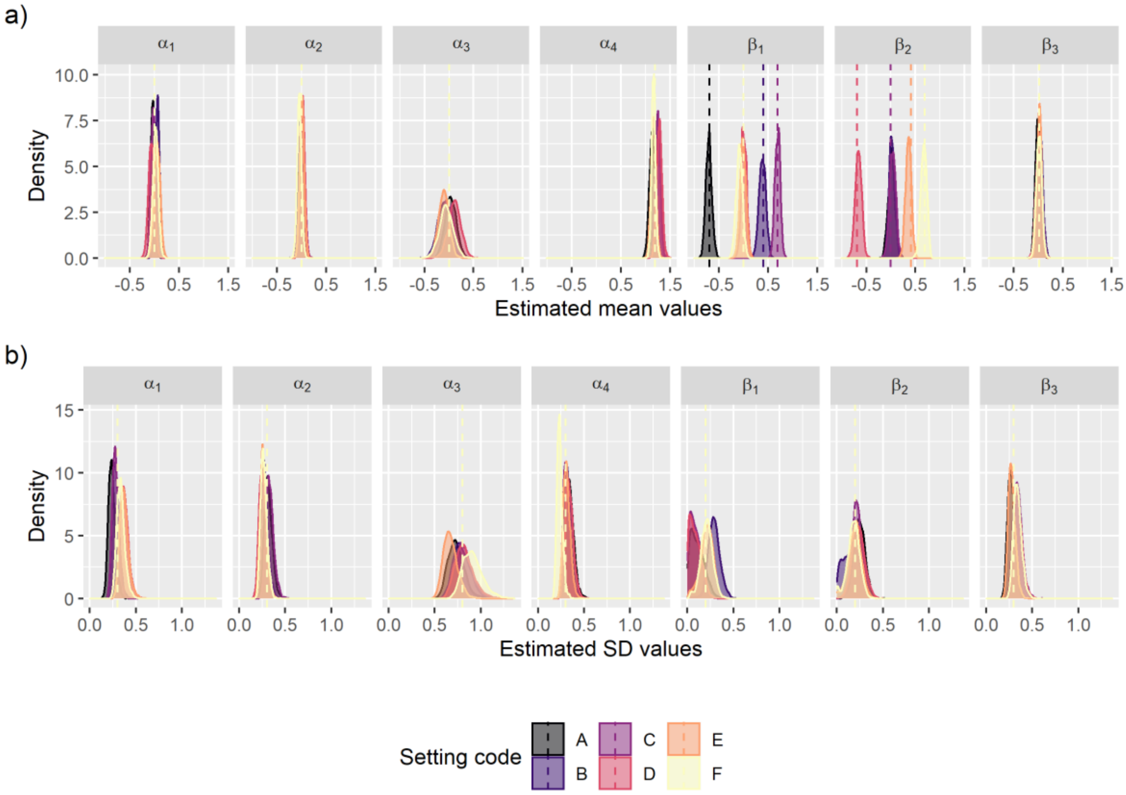

**Figure B4**. Posterior distribution of estimated hyper-parameters (μ and sd) for simulated communities. a) and b) are the estimated mean values for all parameters considering the models simulated varying the mean abundance (β values, setting A to F). The dashed line represents the true values used to simulate the community for each set.

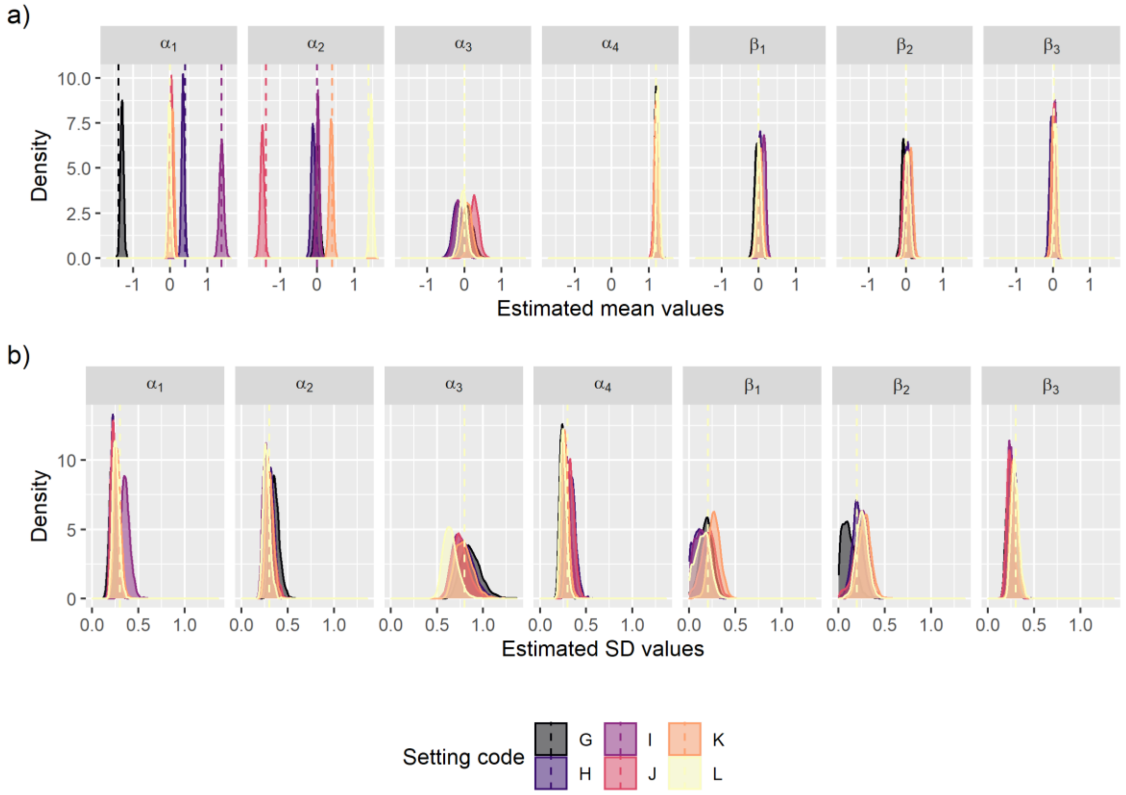

**Figure B5**. Posterior distribution of estimated hyper-parameters (μ and sd) for simulated communities. a) and b) are the estimated values considering the communities simulated varying the mean detection probability (α values, setting G to L). The dashed line represents the true values used to simulate the community for each set.

## Appendix C - Hidden diversity framework

Function to evaluate the extent to which imperfect detection may affect patterns of taxonomic, functional, and phylogenetic diversities in ecological communities This function allows the user to estimate, in the form of a Standardized Effect Size (SES), how much of the diversity was hidden when we do not account with the imperfect detection. Further, the user can calculate the hidden diversity only for taxonomic diversity (TD and Abundance), for functional (SES.FD and SES.MFD), and taxonomic diversity, for phylogenetic (SES.PD and SES.MPD) and taxonomic diversity, or for all measures.

### Arguments

comm = Community data, with sites in the rows and species in the columns.

N = an array when each “slice” is a matrix of the true-abundance (sites in rows and species in columns) for one posterior sampling of the N-mixture model. This object represents the detection-corrected abundance.

phy = a phylogenetic tree, with branch length.

trait = a trait community matrix, with species in rows and traits in columns.

Metric = a string with one or two options c(“pd”, “mpd”). If only “pd” is provided, the function will calculate the SES (standardized effect size) of the sum of the branch length for each community. If only “mpd” is provided, the function will calculate the SES for the mean pairwise distance of species that compose each community. Default is c(“pd”, “mpd”). This argument only works if a phylogenetic tree or a trait matrix is provided.

binary = logical. Default is FALSE. Only necessary when a trait matrix is present. If binary = TRUE, the function is taken into account binary traits for construct a functional dendrogram.

abundance.weighted = logical. Default is FALSE. In this case, for SES.MPD and SES.MFD only the occurrence/incidence/richness of species is accounted for by calculating the diversity values. If abundance.weighted = TRUE, the SES.MPD and SES.MFD will be calculated for an abundance-based community matrix.

Null.model = a string with the null models allowed by ses.pd and ses.mpd function from *picante* package.

Runs = number of permutations used to calculate the null models.

Parallel = number of cores used to process the function

**Value**

The function returns a list with two to six data frames (dependent on imputed diversity):

TD, Abund = always returned. Each object is a data frame with four columns: the observed richness/abundance, the mean richness/abundance for estimated data (N), the standard deviation of the estimated richness/abundance, and the hidden diversity.

sesPD, sesFD, sesMPD and sesMFD = only returned if the user informed a phylogenetic tree and/or a functional traits matrix. Each object has four columns: the SES value for observed diversity, the SES values for mean estimated diversity, the standard deviation for estimated SES, and the hidden diversity.

```
hidden.diversity <- function(comm, N, phy = NULL, trait =
             NULL, metrics = c("pd", "mpd"), binary =
             FALSE, abundance.weighted = FALSE,  null.model
             = "taxa.labels", runs = 499, parallel = 3) {


  n.site <- dim(N)[1] # n.site: the number of sampling sites

  n.samp <- dim(N)[3] # n.samp: the number of posterior sampling


 # transforming N in occurrence data (y)

  y <- N

  for (i in 1: dim(y)[3]) {

    b = which(y[,,i] > 0)

    y[,,i][b] = 1

    y[,,i][-b] = 0

  }


 # calculating the observed and estimated richness (TD)

  TD.df <- data.frame(TD.obs = apply(vegan::decostand(x =
       comm, method = "pa"), 1, sum), TD.est =
       apply(apply(y, c(1,3), sum), 1, mean), TD.sd =
       apply(apply(y, c(1,3), sum), 1, sd))

  TD.df$HD.TD <- (TD.df$TD.obs - TD.df$TD.est)/TD.df$TD.sd


 # calculating the observed and estimated abundance

  N.df <- data.frame(N.obs = apply(comm, 1, sum), N.est =
      apply(apply(N, c(1,3), sum), 1, mean), N.sd =
      apply(apply(N, c(1,3), sum), 1, sd))

  N.df$HD.N <- (N.df$N.obs - N.df$N.est) / N.df$N.sd


  div_measures <- c("rich", "abund", "pd", "mpd")

  hd_metric <- pmatch(metrics, div_measures)


 # if a phylogenetic tree is provided
```

```
  if (!is.null(phy)){
    if(any(hd_metric == 3)){ # calculation of PD
  # observed data
      pd.obs <- picante::ses.pd(samp = comm, tree = phy,
          null.model = null.model,      runs = runs,
          include.root = F)


  # estimated data
      if (is.numeric(parallel)) {
        CL1 <- parallel::makeCluster(parallel, type =
"PSOCK")
        newClusters <- TRUE
      }
      if (!inherits(CL1, "cluster")) {
        pd.ses <- array(NA, dim = c(n.site, 2, n.samp))
        for (i in 1:n.samp){
          temp_pd <- picante::ses.pd(samp = N[,,i], tree =
             phy, null.model = null.model, runs = runs,
             include.root = F)
          pd.ses[ , 1, i] <- cbind(temp_pd[,6])
          }
        PD.df <- data.frame(SES.PD.obs = pd.obs[
          ,"pd.obs.z"],                             SES.PD.est
          = apply(pd.ses[,1,], 1, mean, na.rm = T),
          SES.PD.sd = apply(pd.ses[,1,], 1, sd, na.rm = T))
      }
      else {
        res_sesPD_samp <- parallel::parApply(cl = CL1,
                 MARGIN = 3, X = y, FUN = picante::ses.pd,
                 tree = phy, null.model = null.model,
                 include.root = F, runs = runs)


        HD.comm <- list(SES.PDest = res_sesPD_samp)
```

```
        PD_est <- matrix(unlist(lapply(HD.comm$SES.PDest,
           function(x) x$pd.obs.z)), nrow = nrow(comm),
           ncol = dim(y)[3], dimnames =
           list(rownames(comm), paste("samp", 1:dim(y)[3],
           sep = "_")))

        matrix_mean_SES_PD <-
                      data.frame(matrix(c(apply(PD_est,
                      MARGIN = 1, mean), apply(PD_est,
                      MARGIN = 1, sd)), nrow = nrow(comm),
                      ncol = 2, dimnames =
                      list(rownames(comm), c("mean_ses.pd",
                      "sd_ses.pd")), byrow = FALSE))


        PD.df <- data.frame(SES.PD.obs = pd.obs$pd.obs.z,
          SES.PD.est = matrix_mean_SES_PD$mean_ses.pd,
          SES.PD.sd = matrix_mean_SES_PD$sd_ses.pd)

        }

  # calculating the Hidden Diversity for SES.PD

      PD.df$HD.PD <- (PD.df$SES.PD.obs - PD.df$SES.PD.est)
/ PD.df$SES.PD.sd

    }


    if(any(hd_metric == 4)){ # calculation of MPD

  # observed data

      mpd.obs <- picante::ses.mpd(samp = comm, dis =
          cophenetic(x = phy), null.model = null.model,
          runs = runs)


  # estimated data

      mpd.ses <- array(NA, dim = c(n.site, 2, n.samp))


      if (is.numeric(parallel)) {

        CL1 <- parallel::makeCluster(parallel, type =
"PSOCK")

        newClusters <- TRUE

      }
```

```
if (!inherits(CL1, "cluster")) {

  mpd.ses <- array(NA, dim = c(n.site, 1, n.samp))

  for (i in 1:n.samp){

    temp_mpd <- picante::ses.mpd(samp = N[,,i], dis =
        cophenetic(x = phy), null.model = null.model,
        runs = runs)


    mpd.ses[,1,i] <- cbind(temp_mpd[,6])

  }


  MPD.df <- data.frame(SES.MPD.obs = mpd.obs[ ,
     "mpd.obs.z"], SES.MPD.est = apply(mpd.ses[,1,],
     1, mean, na.rm = T), SES.MPD.sd =
     apply(mpd.ses[,1,], 1, sd, na.rm = T))

}

else {

  res_sesMPD_samp <- parallel::parApply(cl = CL1,
            MARGIN = 3, X = y, FUN =
            picante::ses.mpd, dis = cophenetic(phy),
            abundance.weighted = abundance.weighted,
            null.model = null.model, runs = runs)

  HD.comm <- list(SES.MPDest = res_sesMPD_samp)


  MPD_est <- matrix(unlist(lapply(HD.comm$SES.MPDest,
      function(x) x$mpd.obs.z)), nrow = nrow(comm),
      ncol = dim(y)[3], dimnames =
      list(rownames(comm), paste("samp", 1:dim(y)[3],
      sep = "_")))


  matrix_mean_SES_MPD <-
               data.frame(matrix(c(apply(MPD_est,
               MARGIN = 1, mean), apply(MPD_est,
               MARGIN = 1, sd)), nrow = nrow(comm),
               ncol = 2, dimnames =
               list(rownames(comm),
               c("mean_ses.pd", "sd_ses.mpd")),
               byrow = FALSE))
```

```
        MPD.df <- data.frame(SES.MPD.obs =
           mpd.obs$mpd.obs.z, SES.MPD.est =
           matrix_mean_SES_MPD$mean_ses.pd, SES.MPD.sd =
           matrix_mean_SES_MPD$sd_ses.mpd)

        }


  # hidden for mpd phylo

    MPD.df$HD.MPD <- (MPD.df$SES.MPD.obs -
MPD.df$SES.MPD.est) / MPD.df$SES.MPD.sd

    }

  }


 # calculation for traits

  if (!is.null(trait)){

    if(binary == TRUE){

      bin <- vector()

      for(i in 1:ncol(trait)){

        bin[i] <- is.integer(trait[, i]) |
is.factor(trait[, i])

      }

      con.t <- which(bin == F)

      bin.t <- which(bin == T)

      t.dist <-
        ade4::dist.ktab(ade4::ktab.list.df(list(log(trait[,
        con.t]), ade4::prep.binary(trait[, bin.t],
        col.blocks = ncol(trait[, bin.t])))), type = c("Q",
        "B")) # create a dist matrix, considering mixed-
        variables

    }

      else {

      t.dist <-
ade4::dist.ktab(ade4::ktab.list.df(list(log(trait))), type
= "Q")
```

```
    }

    tree.func <- hclust(d = t.dist, method = "average") #
clustering using UPGMA

    tree.func <- ape::as.phylo(tree.func)


    if(any(hd_metric == 3)){ # calculation of FD

  # observed data

      fd.obs <- picante::ses.pd(samp = comm, tree =
         tree.func, null.model = null.model, runs = runs,
         include.root = F)



  # estimated data

      if (is.numeric(parallel)) {

        CL1 <- parallel::makeCluster(parallel, type =
"PSOCK")

        newClusters <- TRUE

      }


      if (!inherits(CL1, "cluster")) {

        fd.ses <- array(NA, dim = c(n.site, 1, n.samp))

        for (i in 1:n.samp){

          temp_fd <- picante::ses.pd(samp = N[,,i], tree =
             tree.func, null.model = null.model, runs =
             runs, include.root = F)

          fd.ses[ , 1, i] <- cbind(temp_fd[,6])

        }

        FD.df <- data.frame(SES.FD.obs = fd.obs[ ,
          "pd.obs.z"],SES.FD.est = apply(fd.ses[,1,], 1,
          mean, na.rm = T), SES.FD.sd = apply(fd.ses[,1,],
          1, sd, na.rm = T))

      }

        else {

        res_sesFD_samp <- parallel::parApply(cl = CL1,
                 MARGIN = 3, X = y, FUN = picante::ses.pd,
```

```
                tree = tree.func, null.model = null.model,
                include.root = F, runs = runs)


        HD.comm <- list(SES.FDest = res_sesFD_samp)

        FD_est <- matrix(unlist(lapply(HD.comm$SES.FDest,
           function(x) x$pd.obs.z)), nrow = nrow(comm),
           ncol = dim(y)[3], dimnames =
           list(rownames(comm), paste("samp", 1:dim(y)[3],
           sep = "_")))

        matrix_mean_SES_FD <-
                    data.frame(matrix(c(apply(FD_est,
                    MARGIN = 1, mean), apply(FD_est,
                    MARGIN = 1, sd)), nrow = nrow(comm),
                    ncol = 2, dimnames =
                    list(rownames(comm), c("mean_ses.fd",
                    "sd_ses.fd")), byrow = FALSE))


        FD.df <- data.frame(SES.FD.obs = fd.obs$pd.obs.z,
          SES.FD.est = matrix_mean_SES_FD$mean_ses.fd,
          SES.FD.sd = matrix_mean_SES_FD$sd_ses.fd)

        }
```

# hidden FD

```
      FD.df$HD.FD <- (FD.df$SES.FD.obs - FD.df$SES.FD.est)
/ FD.df$SES.FD.sd

    }


    if(any(hd_metric == 4)){

      dist.func <- cophenetic(x = tree.func)

      mfd.obs <- picante::ses.mpd(samp = comm, dis =
           dist.func, null.model = null.model, runs = runs)

      mfd.ses <- array(NA, dim = c(n.site, 1, n.samp))


      if (is.numeric(parallel)) {

        CL1 <- parallel::makeCluster(parallel, type =
"PSOCK")

        newClusters <- TRUE
```

```
}

if (!inherits(CL1, "cluster")) {

  for (i in 1:n.samp){

  temp_mfd <- picante::ses.mpd(samp = N[,,i], dis =
       dist.func, null.model = null.model, runs =
       runs)



    mfd.ses[ , 1, i] <- cbind(temp_mfd[, 6])

  }

  MFD.df <- data.frame(SES.MFD.obs = mfd.obs[ ,
     "mpd.obs.z"], SES.MFD.est = apply(mfd.ses[,1,],
     1, mean, na.rm = T), SES.MFD.sd =
     apply(mfd.ses[,1,], 1, sd, na.rm = T))

  }

  else {

  res_sesMFD_samp <- parallel::parApply(cl = CL1,
            MARGIN = 3, X = y, FUN =
            picante::ses.mpd, dis =
            cophenetic(tree.func),
            abundance.weighted = abundance.weighted,
            null.model = null.model, runs = runs)

  HD.comm <- list(SES.MFDest = res_sesMFD_samp)



  MFD_est <-
      data.frame(matrix(unlist(lapply(HD.comm$SES.MFD
      est, function(x) x$mpd.obs.z)), nrow =
      nrow(comm), ncol = dim(y)[3], dimnames =
      list(rownames(comm), paste("samp", 1:dim(y)[3],
      sep = "_"))))



  matrix_mean_SES_MFD <-
                data.frame(matrix(c(apply(MFD_est,
                MARGIN = 1, mean), apply(MFD_est,
                MARGIN = 1, sd)), nrow = nrow(comm),
                ncol = 2, dimnames =
                list(rownames(comm),
                c("mean_ses.mfd", "sd_ses.mfd")),
                byrow = FALSE))
```

```
        MFD.df <- data.frame(SES.MFD.obs =
           mfd.obs$mpd.obs.z, SES.MFD.est =
           matrix_mean_SES_MFD$mean_ses.mfd, SES.MFD.sd =
           matrix_mean_SES_MFD$sd_ses.mfd)

            }
```

# hidden mfd

```
      MFD.df$HD.MFD <- (MFD.df$SES.MFD.obs -
MFD.df$SES.MFD.est) / MFD.df$SES.MFD.sd

    }

  }


  if (newClusters) {

    parallel::stopCluster(CL1)

  }


  if(!is.null(trait) & !is.null(phy)){

    if(all(hd_metric == c(3, 4))){

      list_res <- vector(mode = "list", length = 6)

      names(list_res) <- c("TD", "Abund", "sesPD",
"sesMPD", "sesFD", "sesMFD")

      list_res$TD <- TD.df

      list_res$Abund <- N.df

      list_res$sesPD <- PD.df

      list_res$sesMPD <- MPD.df

      list_res$sesFD <- FD.df

      list_res$sesMFD <- MFD.df


      for (i in 3:length(list_res)) {

        pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
```

```
        list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
      }

      return(list_res)
    }
      else {
      if(hd_metric == 3){
        list_res <- vector(mode = "list", length = 4)
        names(list_res) <- c("TD", "Abund", "sesPD",
"sesFD")
        list_res$TD <- TD.df
        list_res$Abund <- N.df
        list_res$sesPD <- PD.df
        list_res$sesFD <- FD.df

        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }

        return(list_res)
      }

      if(hd_metric == 4){
        list_res <- vector(mode = "list", length = 4)
        names(list_res) <- c("TD", "Abund", "sesMPD",
"sesMFD")
        list_res$TD <- TD.df
```

```
        list_res$Abund <- N.df
        list_res$sesMPD <- MPD.df
        list_res$sesMFD <- MFD.df

        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }

        return(list_res)
      }
    }
  }

  if(is.null(phy) & !is.null(trait)){
    if(all(hd_metric == c(3, 4))){
      list_res <- vector(mode = "list", length = 4)
      names(list_res) <- c("TD", "Abund", "sesFD",
"sesMFD")
      list_res$TD <- TD.df
      list_res$Abund <- N.df
      list_res$sesFD <- FD.df
      list_res$sesMFD <- MFD.df

      for (i in 3:length(list_res)) {
        pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
        list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
```

```
      }

      return(list_res)
    }
      else{
      if(any(hd_metric == 3)){
        list_res <- vector(mode = "list", length = 3)
        names(list_res) <- c("TD", "Abund", "sesFD")
        list_res$TD <- TD.df
        list_res$Abund <- N.df
        list_res$sesFD <- FD.df

        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }

        return(list_res)
      }
      if(any(hd_metric == 4)){
        list_res <- vector(mode = "list", length = 3)
        names(list_res) <- c("TD", "Abund", "sesMFD")
        list_res$TD <- TD.df
        list_res$Abund <- N.df
        list_res$sesMFD <- MFD.df

        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
```

```
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }

        return(list_res)
      }
    }
  }
  if(!is.null(phy) & is.null(trait)){
    if(all(hd_metric == c(3, 4))){
      list_res <- vector(mode = "list", length = 4)
      names(list_res) <- c("TD", "Abund", "sesPD",
"sesMPD")
      list_res$TD <- TD.df
      list_res$Abund <- N.df
      list_res$sesPD <- PD.df
      list_res$sesMPD <- MPD.df

      for (i in 3:length(list_res)) {
        pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
        list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
      }

      return(list_res)
    }
      else {
      if(hd_metric == 3){
        list_res <- vector(mode = "list", length = 3)
        names(list_res) <- c("TD", "Abund", "sesPD")
```

```
        list_res$TD <- TD.df
        list_res$Abund <- N.df
        list_res$sesPD <- PD.df
        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }

        return(list_res)
      }

      if(hd_metric == 4){
        list_res <- vector(mode = "list", length = 3)
        names(list_res) <- c("TD", "Abund", "sesMPD")
        list_res$TD <- TD.df
        list_res$Abund <- N.df
        list_res$sesMPD <- MPD.df

        for (i in 3:length(list_res)) {
          pos_obs_na <- which(is.na(list_res[[i]][,1]) ==
TRUE & is.na(list_res[[i]][,2]) == FALSE)
          list_res[[i]][pos_obs_na, paste("HD",
gsub("ses","", names(list_res)[i]), sep = ".")] <-
list_res[[i]][pos_obs_na, 2]/list_res[[i]][pos_obs_na, 3]
        }
                return(list_res)
      }
    }
  }
}
```

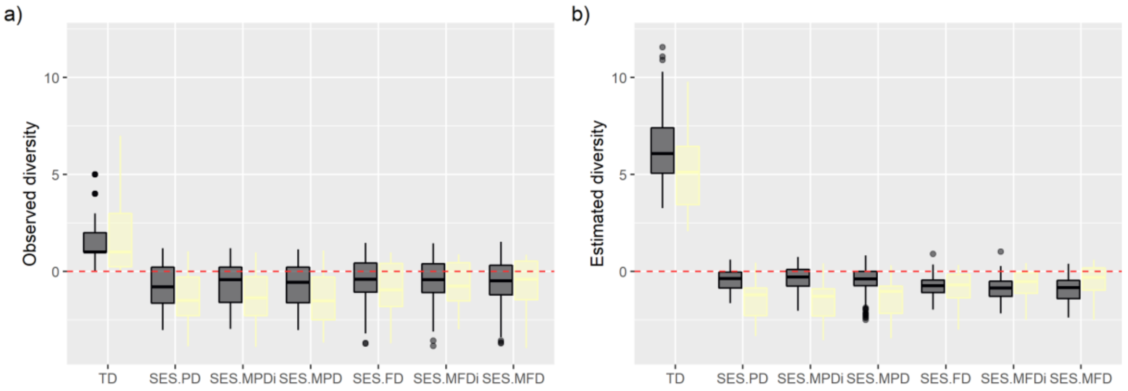

**Figure C1**. Relationship among diversity measures and canopy (dark boxplots) and understory (light yellow boxplots). a) Diversity pattern calculated using only observed data. b) Diversity pattern calculated using 100 matrices of estimated true abundance. TD – taxonomic diversity, SES – standardized effect size, PD/FD – phylogenetic/functional diversity, MPD/MFD – abundance-based mean pairwise phylogenetic/functional distance, MPDi/MFDi – incidence-based mean pairwise phylogenetic/functional distance.

**Table C1**. Relationship between distinct facets of biodiversity and vertical stratification for fruit-feeding butterflies community sampled at FLONA-SFP, southern Brazil. The first four columns show the relationship of diversity measures obtained by observed data with canopy and understory, and the last four columns show the relationship of diversity measures obtained by estimated data (corrected-by-detection). Bold values indicate a statistical significance at a threshold of 0.05. Asterisk indicates the unique case where there was an inversion of the most diverse stratum. TD – taxonomic diversity, SES – standardized effect size, PD/FD – phylogenetic/functional diversity, MPD/MFD – abundance-based mean pairwise phylogenetic/functional distance, MPDi/MFDi – incidence-based mean pairwise phylogenetic/functional distance.

| | Observed data | | | | Estimated data | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | t value | p | Estimate | SE | t value | p |
| TD | | | | | | | | |
| Canopy | 1.440 | 0.373 | 3.860 | **0.004** | 6.296 | 0.785 | 8.021 | **0.000** |
| Understory* | 0.280 | 0.140 | 2.003 | **0.046** | -1.169 | 0.088 | -13.35 | **0.000** |
| SES.PD | | | | | | | | |
| Canopy | -0.547 | 0.297 | -1.841 | 0.108 | -0.459 | 0.264 | -1.739 | 0.132 |
| Understory | -0.584 | 0.194 | -3.003 | **0.003** | -1.017 | 0.048 | -21.33 | **0.000** |
| SES.FD | | | | | | | | |
| Canopy | -0.589 | 0.315 | -1.873 | 0.094 | -0.772 | 0.213 | -3.619 | **0.009** |
| Understory | 0.009 | 0.200 | 0.047 | 0.963 | -0.038 | 0.049 | -0.764 | 0.446 |
| SES.MPDi | | | | | | | | |
| Canopy | -0.506 | 0.304 | -1.668 | 0.137 | -0.382 | 0.277 | -1.379 | 0.212 |
| Understory | -0.580 | 0.186 | -3.118 | **0.002** | -1.140 | 0.051 | -22.29 | **0.000** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SES.MPD | | | | | | | | |
| Canopy | -0.546 | 0.318 | -1.719 | 0.128 | -0.480 | 0.315 | -1.523 | 0.172 |
| Understory | -0.568 | 0.184 | -3.084 | **0.003** | -0.930 | 0.050 | -18.48 | **0.000** |
| SES.MFDi | | | | | | | | |
| Canopy | -0.626 | 0.306 | -2.044 | 0.072 | -0.914 | 0.218 | -4.199 | **0.004** |
| Understory | 0.087 | 0.201 | 0.433 | 0.666 | 0.276 | 0.045 | 6.188 | **0.000** |
| SES.MFD | | | | | | | | |
| Canopy | -0.680 | 0.341 | -1.996 | 0.076 | -0.922 | 0.263 | -3.505 | **0.008** |
| Understory | 0.175 | 0.202 | 0.866 | 0.388 | 0.463 | 0.044 | 10.61 | **0.000** |

# Capítulo 2 – Processos que moldam a diversidade local

## UNVEILING THE ECOLOGICAL PROCESSES THAT SHAPE THE DIVERSITY OF PAMPAS' FRUIT-FEEDING BUTTERFLIES

1- Capítulo escrito na forma de artigo cientifico, em inglês e formatado de acordo com as normas da revista *Journal of Animal Ecology*.

2- Este capítulo tem como coautores: Gabriel Nakamura, Cristiano A. Iserhard e Leandro da Silva Duarte

## Abstract

1. The Uruguayan savanna biome comprises one of the most extensive areas of natural grassland vegetation on the Earth. This biome has significantly changed land cover in recent decades, where croplands have replaced natural grasslands. Modification of landscape characteristics can affect biodiversity changing species occurrence and abundance, which modify the community structure and hence the ecosystem functioning in this fragile biome.
2. Here, we evaluated how microhabitat conditions and landscape affect the occurrence and abundance of fruit-feeding butterflies in different communities in the Brazilian Pampa biome. Furthermore, we assessed whether intrinsic community characteristics, such as community size, can affect the relationship between species and the environment.
3. We found that landscape variables contributed more to explain variation in species occurrence, while microhabitat and dispersal trait were more critical to explain variation in species abundance. Richness and abundance were generally lower under higher temperatures (grassland sites), and this habitat is characterized by species with less flight ability. Further, more mobile species were benefited under an increasing homogenization of the landscape. Finally, we observed that larger communities were more driven by environmental filters than smaller communities.
4. The Pampa naturally comprises a grassland-forest mosaic maintained by low-intensity farm activities. Heterogenous landscapes can support more species than simplified landscapes because they provide distinct microhabitats and potential

sources of diversity to surrounding regions. The homogenization of fauna can profoundly affect species’ interaction and even ecosystems’ functioning. Further, we draw attention to the fragility of communities with few individuals since their relationships with the environmental filter are less robust. This mismatch between species and environment can reduce our understanding of the processes that shape communities and our ability to predict community responses to climate and land use change.

**Keywords**: Occurrence, abundance, natural grasslands, functional traits, environmental filter, JSDM, community size.

## Introduction

The Uruguayan savanna biome (*Pampa*) comprises one of the most extensive areas of natural grassland vegetation on the Earth, covering approximately 70 million hectares in Southern Brazil, central-eastern Argentina, and Uruguay (Baeza & Paruelo, 2020). Since the introduction of cattle raising in the 17th century, grazing has represented the main economic activity in Pampa and is recognized as a key factor in maintaining the ecological features of grasslands, including the prevention of forest expansion (Lima et al., 2020; Overbeck et al., 2007; Pillar & Quadros, 1999). Unfortunately, in recent decades this biome experienced a significant change in land use/land cover, where the natural grasslands were replaced mainly by crops, silviculture, and artificial pastures (Baeza et al., 2014; Baeza & Paruelo, 2020). Increasing land use intensification is expected to affect biodiversity by decreasing the population size and modifying the community structure and the ecosystem functioning (Gámez-Virués et al., 2015; Medan et al., 2011). Furthermore, the loss of natural coverage, combined with the lower percentages of protected areas, places Pampa as one of the most endangered biomes of the world (Hoekstra et al., 2004; Overbeck et al., 2007; Roesch et al., 2009).

The use of well-known groups of invertebrates could be critical to identify priority areas for conservation because they exhibit high sensitivity to environmental changes (Brown & Freitas, 2000; New, 1997). For the Pampa biome, there is a gap in knowledge of species distribution, especially of invertebrates (Lima et al., 2020; Renner et al., 2018), jeopardizing our ability to predict how species can respond to environmental change. Within insects, butterflies are used worldwide as biological models to understand underlying processes that structure communities at local and landscape scales (Börschig et al., 2013; Pavoine et al., 2014) and generally reflect the responses of non-target groups (New, 1997). Both landscape and climatic conditions can affect the occurrence and abundance of butterflies (Checa et al., 2019; Santos et al., 2020). Microclimatic variation can play a significant role in metapopulation dynamics controlling abundance (Checa et al., 2019; Xing et al., 2016), and at the community level, microclimate can be critical for niche segregation, favoring the persistence of diverse communities (WallisDeVries & van Swaay, 2006). On the other hand, low-intensity farming is important in maintaining local diversity and enhancing biodiversity in the surrounding landscape (Ekroos et al., 2013), while increasing land use intensity can reduce species diversity and leads to biotic homogenization (Börschig et al., 2013; Gámez-Virués et al., 2015). Furthermore, species' response to microhabitat and landscape variables can depend on their life-history traits, such as dispersal ability (Öckinger et al., 2010; van Halder et al., 2017).

Dispersion is a critical trait in determining the survival and persistence of species, mainly when there are shifts in climate niches or to connect populations in a fragmented landscape (Freire et al., 2021). In butterflies, traits related to wing or thorax measures can be related to dispersal capacity because they affect flight performance (Chai & Srygley, 1990; Srygley & Chai, 1990). The wing-thorax ratio (WTR) is a measure of body allometry, where low WTR values describe a stouter body and faster flights, and high WTR values indicate slender bodies and slow flights (Graça et al., 2017; Hall & Willmott, 2000). Butterflies inhabiting extensively managed grasslands experience more stable habitats than intensively managed ones, which allows habitat specialization (Börschig et al., 2013). Species specialists have narrow feeding niches, lower mobility, and are generally smaller than generalist species (Börschig et al., 2013; Freire et al., 2021). Due to their lower dispersal ability, specialized species are more affected by landscape changes, both by altering host plant availability and changing biotic interactions. Identifying how dispersal-related traits are linked to environmental features could improve our understanding of how species respond to habitat alteration in threatened biomes such as Pampa, as well as the identification of species under high extinction risk (Börschig et al., 2013; Kotiaho et al., 2005; Pavoine et al., 2014).

Trait-based approaches benefit from the framework of Assembly Rules (Ackerly, 2003; McGill et al., 2006). Assembly rules postulate that deterministic and stochastic processes, viewed as filters (Zobel, 1997), can operate at different scales until determining the local species composition (Ovaskainen & Abrego, 2020; Zobel, 1997). Deterministic processes are related to the niche, including the response of species to the environmental gradients (environmental filter) or species interactions (biotic filter), and actuate as selective forces, filtering species that can survive and persist in local communities (Kraft et al., 2015; Ovaskainen & Abrego, 2020). By considering the traits that influence species' responses to the environment rather than focusing on the species itself, we can gain a deeper understanding of assembly processes, which can be linked with speciation and adaptation, if we also take into account how these traits evolved (Ovaskainen & Abrego, 2020). On the other hand, the stochastic processes can provide variation in the local communities by altering the relative abundance of species or their occurrences at random (Siqueira et al., 2020). Mechanistic and observational studies indicate that stochasticity can play a critical role in communities composed of small or isolated populations (Siqueira et al., 2020). Investigating how species respond to environmental, biotic, and stochastic filters in different environmental contexts and community sizes may be relevant from an applied perspective since many types of environmental disturbances, such as land use intensification, can reduce the size of local communities, compromising their functioning (Siqueira et al., 2020).

Fruit-feeding butterflies (Lepidoptera, Nymphalidae) compose a guild that feeds mainly on rotten fruit rather than nectar (DeVries, 1987), and due to this characteristic, they can be sampled by standardized and passive methods, allowing comparison among different sites (Freitas et al., 2014). Thus, this guild represents an excellent study model for evaluating how processes act to assemble distinct communities in the Brazilian Pampa. Considering the high threat to the Pampa biome, we aimed to unravel the ecological processes that shape the diversity of fruit-feeding butterflies of the Pampas. Specifically, we evaluated 1) how the environment filter defined by microclimatic conditions and landscape use, dispersal traits, and the phylogenetic relationships act to shape the occurrence and abundance of fruit-feeding butterflies, 2) how the community-level responses (taxonomic and functional) depend on the environmental variation, and 3) the relationship between the community size and the strength of the environmental filter to species distributions. We expect that 1) landscape variables will be more important to determine variation in fruit-feeding butterflies' occurrence than microclimate variables, which will be more important to structure the abundance variation, and both the trait and the phylogeny will be important in defining the niche of species (Freire et al., 2021; Graça et al., 2017). For community level, we expect that 2) taxonomic structure will be more affected by microhabitat while functional composition will respond more to landscape features, leading to an increase of butterflies with high flight mobility at homogeneous sites. Further, we expect 3) that larger communities (high abundance) will have a stronger signal of environmental selection processes and less effect of stochasticity than small communities.

## Material and Methods

### *Study area*

We sampled the fruit-feeding butterflies in four municipalities of the Rio Grande do Sul State, southern Brazil: Eldorado do Sul (EEA – 29° 59′ 49.7″ S, 51° 18′ 28.3″ W), Jaguarão (JG – 32° 33′ 58″ S, 53° 22′ 33″ W), São Gabriel (SG – 30° 20′ 9″ S, 54° 19′ 12″ W), and Quaraí (QR – 30° 23′ 16″ S, 56° 27′ 3″ W) (Fig. 1). For JG, SG, and QR we sampled private proprieties in the rural areas of the municipalities, while the EEA (Estação Experimental Agronômica) is a public area from the Universidade Federal do Rio Grande do Sul, with a total area of 1560ha. All the sampling points are located under the Pampa biome domain (Fig 1), consisting of a landscape dominated by grasslands, and the forest formation is restricted to slopes and near water bodies (riparian forests), composing a natural mosaic (Luza et al., 2014). As the Pampa is in a transitional zone between tropical and temperate climates, it has hot summers and cool winters and no dry season (Overbeck et al., 2007), the annual precipitation ranges from 1200–1600mm, and the mean annual temperatures range from 13 – 17 °C. The sampled sites have a high heterogeneity in geomorphological composition, which could affect the floristic composition and consequently alter the butterflies' composition (Curtis et al., 2015; Roesch et al., 2009). The EEA is located in the Central Depression, where the soil is primarily formed from arenitic substrates (sandy soils). São Gabriel and Jaguarão belong to the Sul-Riograndense Shield, this formation is dominated by outcrop rocks of igneous rock, making the landscape with moderate to strong slopes (Filho et al., 2017; Ramos & Koester, 2014), and Quaraí belongs to Serra Geral Formation, composed predominantly by basaltic rocks (Silva et al., 2004).

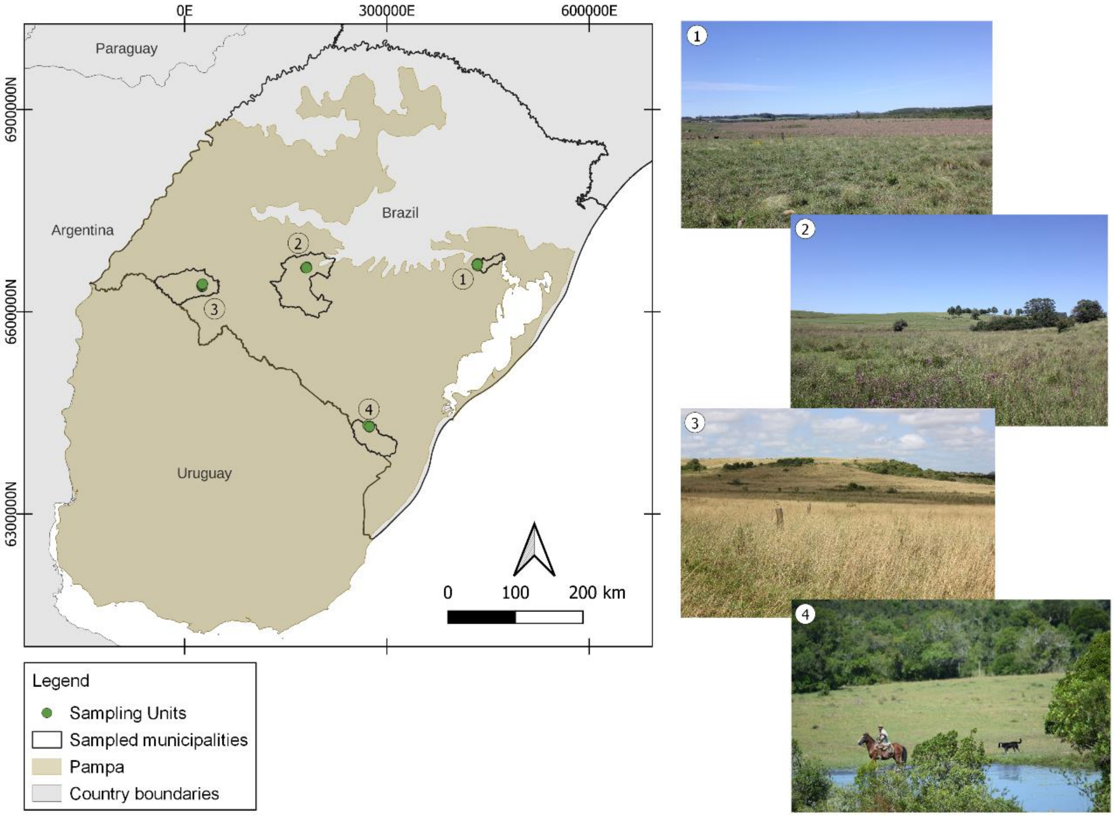

**Figure 4.** Map of the study localities in the Pampa biome, in the southern part of Rio Grande do Sul State, Brazil. The number of pictures corresponds to sample areas. 1. Eldorado do Sul (EEA), 2. São Gabriel (SG), 3. Quaraí (QR), 4. Jaguarão (JG).

### *Sampling procedure*

Fruit-feeding butterfly assemblages were sampled in two field expeditions, the first between December 2019 and February 2020 and the second between November 2020 and February 2021. The period between December and March corresponds to the summer season in the Southern Hemisphere, which is the best period of the year for sampling butterflies (Iserhard et al., 2017). Each local was sampled twice, except for EEA, and we alternated the month of sampling between the expeditions, resulting in seven independent sampling occasions. The sampling design was equal for all locals and consisted of seven sampling units in the native grasslands and seven sampling units in the riparian forests, except for SG, which had only six SU in the riparian forest at the first survey. We adopted standardized methods for sampling fruit-feeding butterflies in the Neotropical region (Freitas et al., 2014), which consisted in installing five bait traps (called Van Someren-Rydon) per sampling unit, which were baited with a mixture of mashed bananas and sugarcane juice (Freitas et al., 2014). The traps remained open for an average of 10 consecutive days, and we checked the traps daily to replace the bait and identify the individuals caught. For those individuals not easily identified during sampling, we collect the specimen for identification in the laboratory. In addition, we collected and fixed at least one specimen from each species for functional evaluation. All released individuals were marked with sequential numbering to avoid recounting. The total sampling effort was 48,500 trap hours (97 sampling units and 70 sampling days). We measure the temperature and humidity using a Hobo® datalogger, placed inside the traps in grasslands and forests. The dataloggers were programmed to take measures at an interval of every 30 minutes.

### *Data description*

Community data (matrix **Y**) – the community data is a matrix with the individual counts for each species in each sampling unit. We define a community as the set of species sampled from a municipality on each sampling occasion ($n = 1, ..., 7$). From Y, we derived the presence-absence data ($Y_{PA}$) for each community – the species have one if present and 0 if absent from the sampling unit. Further, we used the Hellinger transformation on abundance data to obtain a relative abundance matrix ($Y_{Normal}$), once the Hellinger transformation standardizes the abundances to sample totals and then square roots them (Legendre & Legendre, 2012).

Environmental data (matrix **X**) – the environmental data contains a categorical variable habitat with two levels: grassland or riparian forest, and the microclimatic variables temperature and humidity (as cited above), which were pooled across each sampling unit for each local and each survey. For each SU of each local, we extract the land use and land cover (LULC) as a proportion of Forest Formation, Natural Grassland, and Crop area in a buffer of 500m around the sampling unit centroid coordinate. These variables were chosen because they represent the natural cover of the Pampa (forest and grasslands), as well as anthropic land use for agriculture. The LULC data were acquired from the 6-collection of the MapBiomas project (https://mapbiomas.org/), using 2020 as the base year and a resolution of $30m^2$. Although the data on LULC have many classes, only 11 classes were included by the buffer (Forest Formation, Forest Plantation, Wetlands, Grasslands, Mosaic Agriculture and Pasture, Soybean, River, Lake and Ocean, Other Temporary Crops, Other non-Vegetated Areas, Rocky Outcrop, and Rice). We define the variable Crop as a sum of the proportion of the variables Forest Plantation, Mosaic Agriculture and Pasture, Soybean, Other Temporary Crops, and Rice. Wetlands, Rivers, lakes and oceans, Other non-Vegetated Areas, and Rocky Outcrop proportions were not used in the analysis.

Trait data (matrix **T**) – We gathered values of the wing-thorax ratio (WTR) for all 35 species sampled. To obtain trait data, we used digital images of species obtained from voucher specimens or specimens of museum collections (Unicamp, UFPel). The specimens were placed on a white background with a scale bar and photographed using a digital camera on a tripod. The linear measures were extracted from the images using the software TPSdig version 2.30 (Rohlf, 2015). We measure linear measures for forewing length (FWL) and the length and width of the thorax (TL, TW, respectively). We used these linear measures to calculate the WTR ($FWL^2/TV$, when TV is the thorax volume ($TL \times TW^2$) (Graça et al., 2017)). As we usually measured more than one specimen, the WTR represents an average value by specie.

Phylogenetic data (matrix **C**) – The phylogenetic data have the phylogenetic relationship among the 35 species of fruit-feeding butterflies sampled. We constructed our phylogenetic tree using as backbone a phylogenetic hypothesis proposed by Chazot et al. (2021). We used a species inclusion procedure based on the cladistic hierarchy. Following this procedure, species not represented in the phylogenetic tree were inserted as polytomies at the genus node. If there is no species representative of the genus in the backbone tree, we searched the literature to find which of the genera present in the phylogeny was closest to the species to be inserted and grafted the species as a sister group.

### *Data analysis*

We evaluated how the environmental variables affect the species' occurrences and relative abundances by fitting a joint species distribution model whit the Hierarchical Modelling of Species Communities framework (HMSC, Ovaskainen et al. 2017). We build two model structures, one to evaluate the occurrence (PA model) and one to evaluate the abundance of the species (Normal model). Despite this, we used the same variables to model fixed and random effects. The fixed component models the environmental filter, and we include matrix **X**, traits, and phylogenetic information (matrix **T** and **C**). As temperature, humidity, and habitat type as correlated, we have chosen to use only temperature as the microhabitat descriptor to avoid multicollinearity and hence the over-parametrizations of the model (Ovaskainen & Abrego, 2020). In this way, high temperatures indicate drier grassland areas, and low temperatures correspond to wetter riparian forest areas. For the random component, we opted to include spatially explicit random effects (XY), which describe the spatial coordinates for each sampling unity.

For each community (n=7) we fitted the model PA and Normal, resulting in 14 models. As occurrence was represented by presence-absence data, we employed a Bernoulli distribution with a probit-link function in PA model, while we used Gaussian distribution for relative abundance data in Normal model. We ran models with two Markov chain Monte Carlo (MCMC), a burn-in of 1000, a thin rate of 10, and 1000 posterior samples, using the default prior definitions by the HMSC framework (Tikhonov et al., 2020). We examine the MCMC convergence using the potential scale reduction factor (Gelman-Rubin convergence diagnosis (Gelman & Rubin, 1992). We assessed the explanatory power through $R^2$ for Normal and Tjur $R^2$ for PA models. We explored the parameters estimated in terms of variance partitioning from model components and examined how each variable affects the species niche and if traits and phylogeny mediate these effects ($\beta$, $\gamma$, and $\rho$). Also, we examine if there are residual species associations ($\Omega$) and spatial effects ($\alpha$). Regarding community-level structure, we predict the responses in terms of richness, relative abundance, and community-weighted trait mean (CWM) for each community in relation to environmental covariates included in models and accessed the strength of the relationships through correlation tests.

We performed a linear model to evaluate the relationship between the environmental filter and community size, using the model fit of the Normal model ($R^2$) as the response and community size (total abundance of each community) as the predictor variable. Additionally, we evaluated if the community size is related to the variance explained by each environmental variable and by the random part of the model.

The scripts and processed data to perform the analysis were stored on GitHub https://github.com/richterbine/MS_DiversityofPampaBfly.

## Results

We collected a total of 4,621 individuals distributed in 36 species and 28 genera belonging to the four subfamilies of fruit-feeding butterflies (Table S1). Satyrinae was the most abundant (95.8% of the total), being that 98% of the individuals belonged to Satyrini. Charaxinae was the second most abundant subfamily (2.2%), followed by Biblidinae with 1.9% and Nymphalinae with 0.2%. The riparian Forest plots were richer and more abundant than the grassland plots (Table S1; Fig. S1, Supplementary material). Regarding the communities, the EEA2 was the richest site (25 species) but the less abundant (159 individuals), while QR2 was the most abundant (1,326 individuals), and, together with JG2, had the lowest richness (12 species). SG1 and SG2 showed more equitability on abundance distribution (Fig. S1c, Supplementary material), and SG1 had higher diversity in terms of subfamily and tribe at grassland sites (Fig. S1 a-b, Supplementary Material).

For all communities and both model structures, the MCMC convergence was satisfactory (Fig. S2, and Table S2, Supplementary material), and the explanatory power was greater than 0.5 (Table 1), indicating a good fit of the PA and Normal models to the fruit-feeding butterflies' data.

**Table 3.** The mean explanatory power (EP), with their standard deviation (SD), the mean observed occurrence/relative abundance (Obs), and the mean predicted occurrence/relative abundance (Pred) for model PA, which model the occurrence of species, and model Normal, which model the relative abundance data for each community (EEA – Estação Experimental Agronômica, JG – Jaguarão, SG – São Gabriel; number 1 and 2 indicate the first and the second field expedition respectively).

| | PA | | | | Normal | | | |
|---|---|---|---|---|---|---|---|---|
| Local | EP (Tjur $R^2$) | SD | Obs | Pred | EP ($R^2$) | SD | Obs | Pred |
| EEA2 | 0.79 | 0.31 | 62 | 62.65 | 0.56 | 0.38 | 26.13 | 26.14 |
| JG1 | 0.78 | 0.26 | 66 | 66.01 | 0.62 | 0.31 | 24.14 | 24.13 |
| JG2 | 0.85 | 0.18 | 51 | 51.04 | 0.75 | 0.28 | 17.17 | 17.17 |
| QR1 | 0.52 | 0.31 | 66 | 65.63 | 0.58 | 0.26 | 21.90 | 21.89 |
| QR2 | 0.78 | 0.25 | 70 | 69.70 | 0.59 | 0.33 | 24.22 | 24.22 |
| SG1 | 0.64 | 0.28 | 60 | 60.57 | 0.53 | 0.31 | 22.72 | 22.72 |
| SG2 | 0.76 | 0.21 | 47 | 47.81 | 0.58 | 0.27 | 20.65 | 20.58 |

Although the explanatory variable that contributed most to the model predictions for presence-absence data differed among communities, landscape variables explained more than 50% of the variation in the data for five communities (EEA2, JG1, QR1, QR2, and SG1, Fig. 2), while in only two the microhabitat was more important (JG2 and SG2, Fig. 2). In addition, the random part of the PA model explained less than 5%, indicating that we included relevant environmental predictors in the fixed part of the model for all communities. On the other hand, for relative abundance data, in four communities, we had microhabitat conditions as the leading explanatory variable (SG1, SG2, JG1, and JG2, Fig. 3), while for two communities, the forest proportion was the most important in the variance partitioning (QR1 and QR2, Fig. 3). The random part of the model explained more than 10% for all communities, being that for EEA2, it explained more than 30% of the variation in the data (Fig. 3).

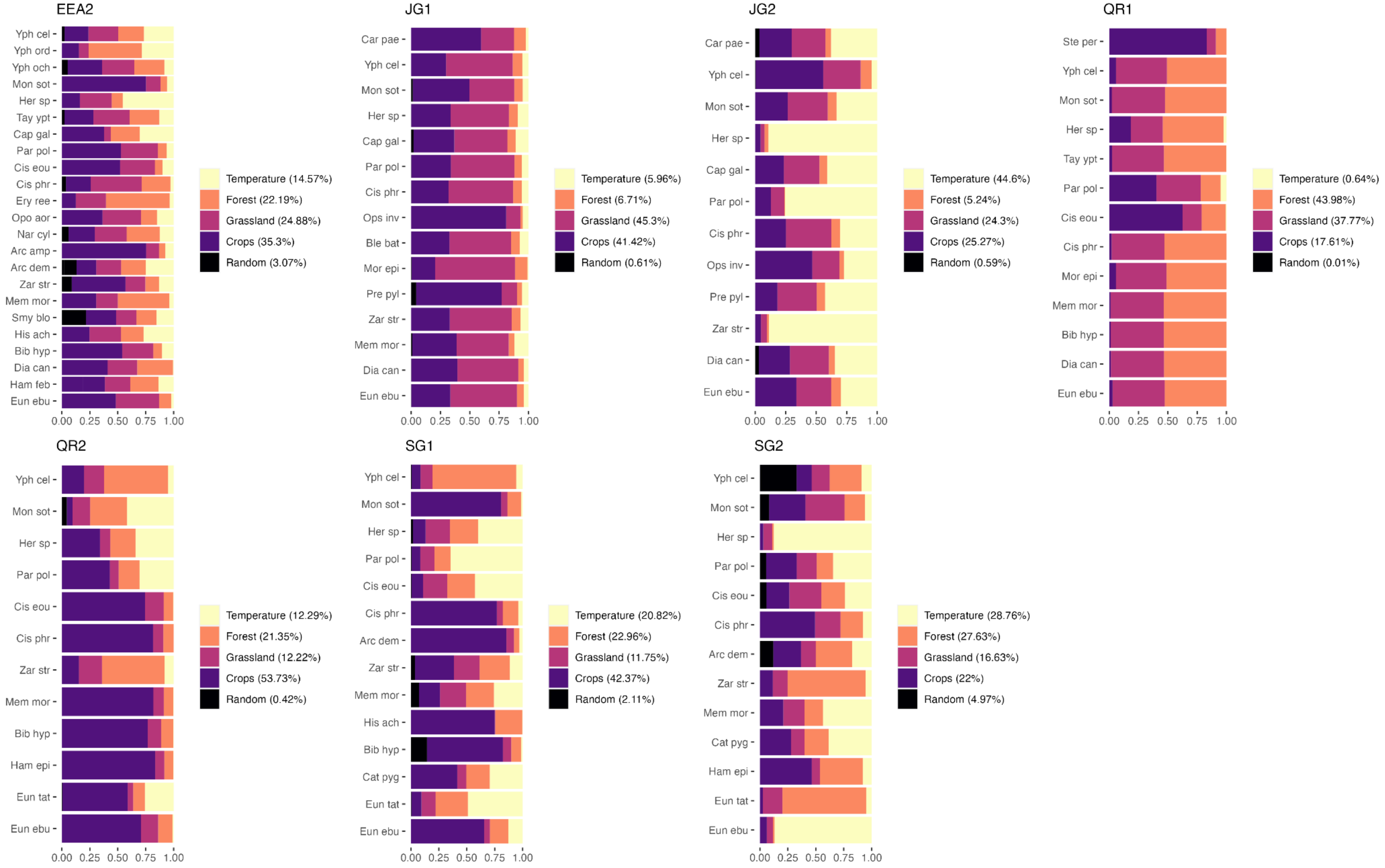

**Figure 5.** Variance partitioning for each community indicating the proportion of explanation from fixed (environmental variables) and random components of the model from presence-absence data (PA model).

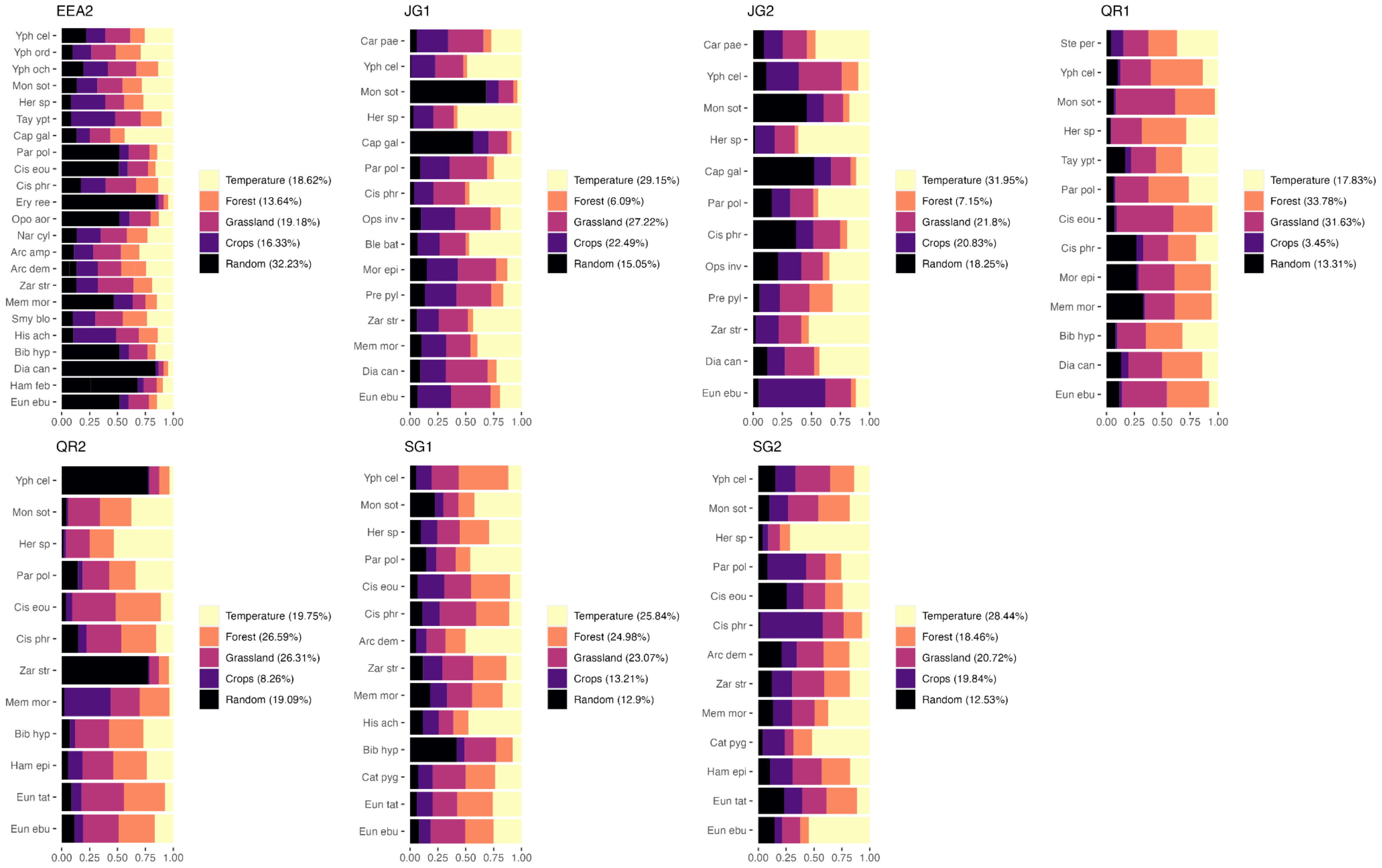

**Figure 6.** Variance partitioning for each community indicating the proportion of explanation from fixed (environmental variables) and random components of model from relative abundance data (Normal model).

Regarding species response, the WTR did not significantly affect species niche with at least 0.90 posterior probability for the model PA for any community. On the other hand, WTR was important in explaining more than 50% of the variation in species' occurrences in two communities (QR1 and SG1, Table S3, Supplementary material). For the relative abundance data, the opposite occurs. The trait was less correlated to abundance variation (Table S3, Supplementary material), but we captured a significative association between WTR and species niches for QR1 and SG1, although these associations differ between communities (Fig. S3, Supplementary material). We did not observe strong evidence for phylogenetic niche conservatism (PNC) in any community since the 95% credible interval includes zero (Table S4, Supplementary material), excepted by EEA2 ($\rho = 0.97$, CRI 95% = 0.78 - 1.00, Table S4) on Normal model.

We found a significant signal in the residual species association ($\Omega$) with a support level of 0.9 for both presence-absence and relative abundance data (Fig. S4 and S5, respectively, Supplementary material). Although the associations changed regarding data type (occurrence or abundance), we found mainly positive associations, indicating that species tended to co-occur more than expected by chance (Fig. S4 and S5, Supplementary material). Indeed, there was no support for spatial signal in the residual for any community or model structure, as the posterior distribution of the $\alpha$ overlapped with zero (Table S4, Supplementary material).

For the community-level response to environmental factors, species richness and relative abundance were negatively related to the microclimatic condition (Fig. 4a, Fig 5a). Regarding the landscape variables, the proportion of crops presented an overall positive relationship on richness but with less support (Fig. 4g), while for the other variables (forest and grassland proportion), we found few and mixed influences on richness with no clear patterns (Fig. 4c and e). The same lack of pattern in community responses to landscape variables was observed when abundance was evaluated, although the correlations were more robust than in PA models (Fig. 5c, e and g). Finally, considering the functional composition, communities located at higher temperatures tend to have a higher occurrence and abundance of species with high WTR (slower flights, Fig. 4b and Fig. 5b). The response of CWM using occurrence or abundance data to landscape variables was very similar, although we found fewer strong correlations using occurrence. An increase in the proportion of landscape variables (habitat homogenization) tends to increase the occurrence and abundance of species with low WTR (faster flights, Fig. 4 d, f, and h, Fig. 5 d, f, and h).

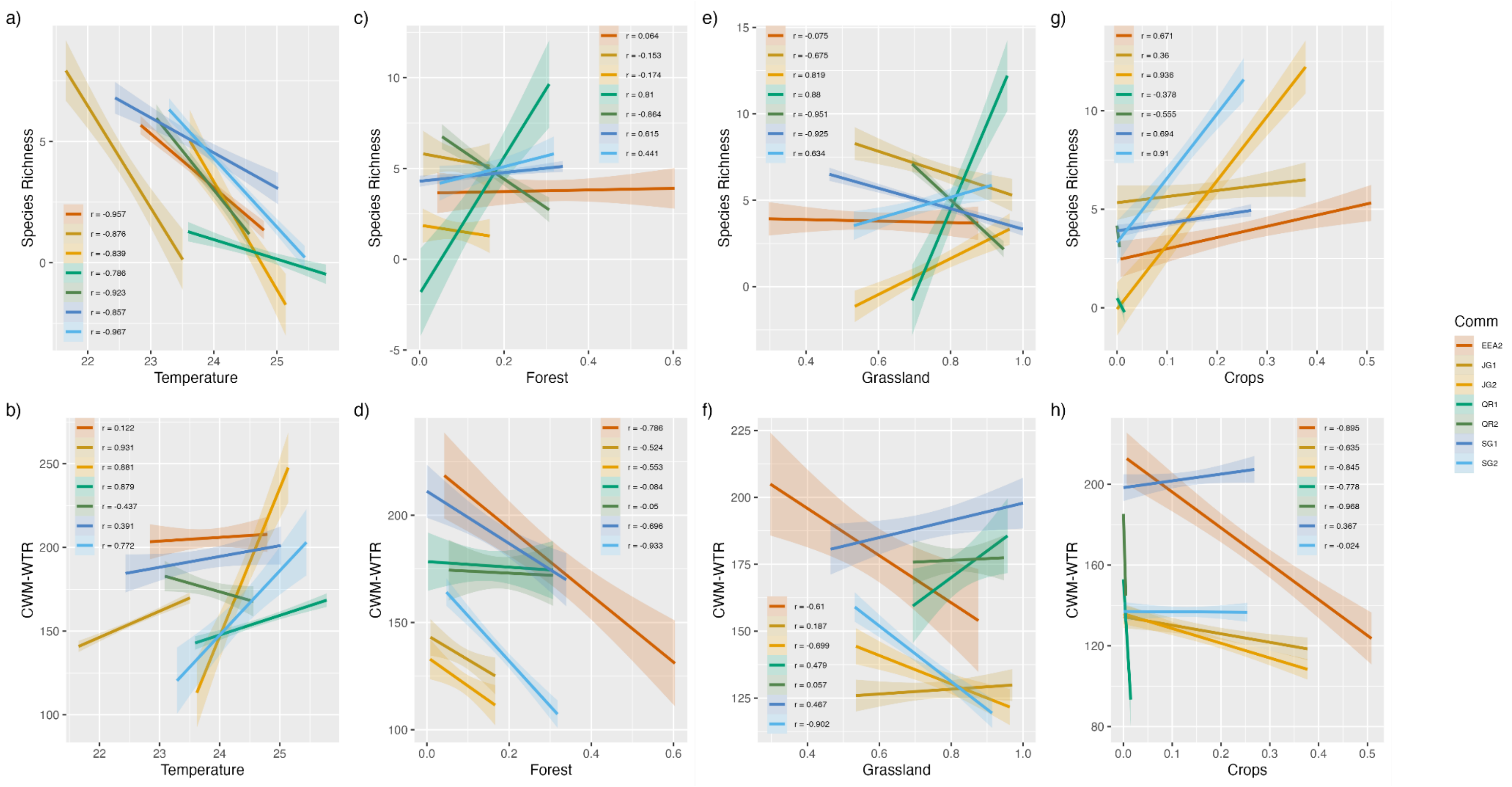

**Figure 7.** Predicted responses of species richness and CWM to environmental variables estimated by the PA model. Each colored line represents one community, and values within panels describe the strength in correlation (r) between community descriptor and environmental variables. The regression lines and confidence intervals (shaded area) were generated by performing a linear model.

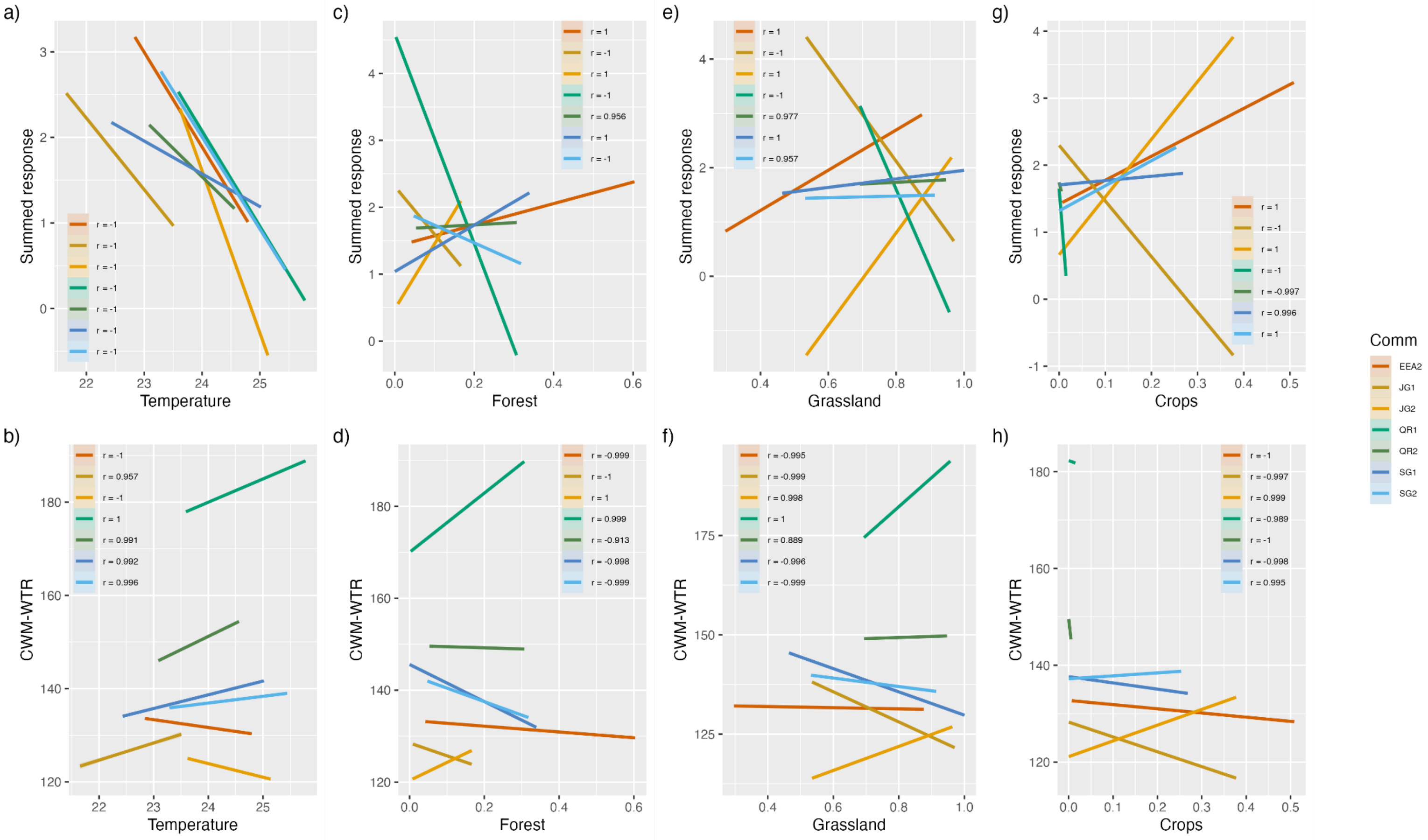

**Figure 8.** Predicted responses of relative species abundances and community-weighted trait mean to environmental variables estimated by the Normal model. Each colored line represents one community, and values within panels describe the strength in correlation (r) between community descriptor and environmental variables. The regression lines and confidence intervals (shaded area) were generated by performing a linear model.

We found that the size of communities positively affected the strength of the relationship between community composition and their responses to environmental factors (Fig. 6a). When decoupling the effects of each variable with community size, we found that native forest and grassland proportion were the main responsible variables in determining the strength of community responses and environmental variables (Fig. 6c and d), meaning that bigger communities are more affected by natural landscape variables than smaller communities. Further, we did not find a significant relationship between the random effect and community size.

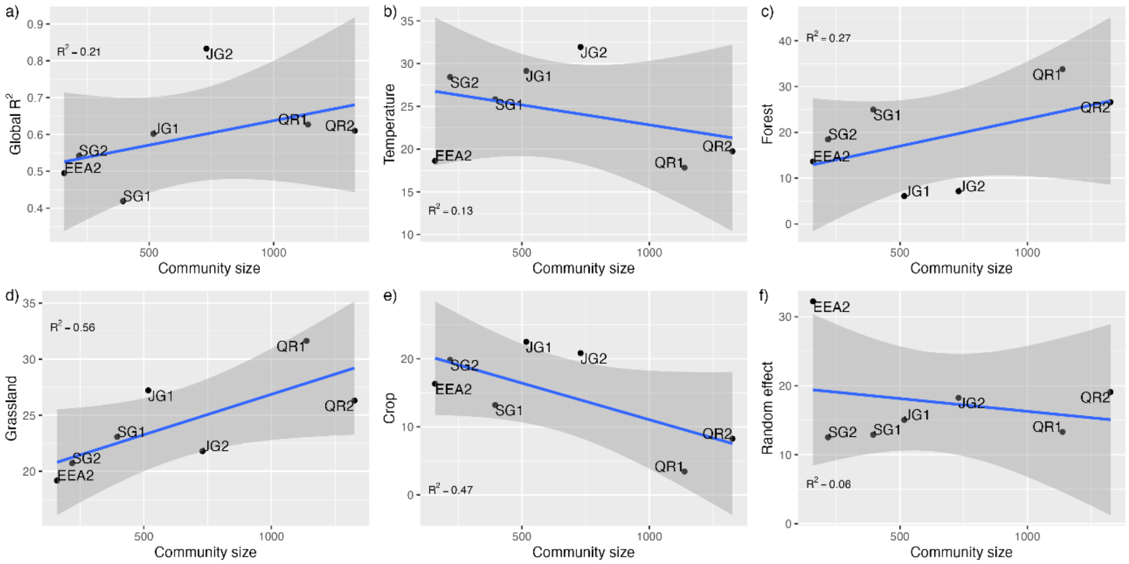

**Figure 9**. Relationship between community size (measured as the abundance sum of all species sampled for each community) and environmental predictors. a) effect of community size on the global model fit, measured as $R^2$. From b) to f) we evaluate the relationship between community size in the explained variation by environmental variables (b to e) and random effects (f). The $R^2$ value inside plots indicates the coefficient of determination for the linear model adjusted. The blue line indicates the predicted response, and the gray shadow indicates the confidence interval of 95%.

## Discussion

Due to the great difference in ecological processes that maintain biodiversity in grasslands and forested ecosystem (Luza et al., 2014; Pillar & Vélez, 2010), studies that assess how environmental filters shape species distribution in natural grasslands ecosystems are essential to guide actions of conservation and sustainable use (Lima et al., 2020; Pillar & Vélez, 2010). In this study, we investigate the role of environmental and ecological processes in shaping fruit-feeding butterflies' occurrences and abundance in the Brazilian Pampa biome. We highlight the different effects that the microhabitat and landscape variables have on the variation of occurrence and abundance of species and how landscape homogenization can change the functional composition, leading to communities with more generalist species in term of habitat requirements. Furthermore, we draw attention to the fragility of communities with few individuals. Since their relationships with the environmental filter (deterministic process) are weaker, our ability to predict how these communities will respond to climate and land use changes is reduced. Because small communities are more subjected to random loss of species and individuals than large communities, changes in their composition can lead to a loss in functional diversity and hence affecting ecosystem functioning (Gámez-Virués et al., 2015; Siqueira et al., 2020).

We would expect landscape variables to affect species occurrence since its permanence depends not only on microclimatic factors but also on habitat structure and the presence of host plants (Lawton, 1983). The landscape of Pampa is naturally heterogeneous, generally forming a grassland-forest mosaic maintained by low-intensity farm activities such as cattle raising and fire (Luza et al., 2014; Pillar & Vélez, 2010). These mosaics can enhance biodiversity because structurally complex landscapes can support more species than simple landscapes (Tscharntke et al., 2012). Our results corroborate the patterns found in temperate grasslands, in which semi-natural grasslands act as population sources exporting individuals to the surrounding habitats (Börschig et al., 2013; Ekroos et al., 2013; Öckinger & Smith, 2007). Further, landscape variables are important to determine the functional structure of communities, leading to an increase in the abundance of more-generalist species (higher flight mobility) according to an increase in landscape scale simplification (habitat homogenization, Gámez-Virués et al. 2015). Studies in temperate grasslands, and even at Pampa to other taxa, indicate that land use intensification leads to a biotic homogenization in insect communities (Börschig et al., 2013; Pires et al., 2022; Renner et al., 2018). For communities evaluated in this study, we observed an important role of landscape heterogeneity since the increment in the dominance of one land use class leads to an increase in the occurrence and abundance of mobile species. Although species that inhabit grasslands cannot directly depend on forest resources, they can gain from the natural barrier that forests create, increasing the humidity and reducing wind speed.

On the other hand, microclimate guides variation in species' relative abundances and strongly shapes patterns of richness and abundance of communities. Microclimate plays an important role in determining metapopulation dynamics both directly by promoting habitat suitability or indirectly by controlling food availability (Checa et al., 2014; Curtis et al., 2015), and traits are expected to mediate the relationship between species abundance and environmental gradients (Curtis et al., 2015). Individual and community-level responses indicate a negative association between species occurrences or relative abundance with temperature. High temperatures are associated with grassland habitat in our study, and consequently, we can conclude that grassland habitat has fewer species and individuals than plots located in the riparian forest. However, regarding functional composition, we observe that grassland habitats are characterized by individuals with low dispersal capacity (large WTR). This indicates that individuals inhabiting grasslands tend to be more specialized and less diverse than individuals living in riparian forest habitats. Xing et al. (2016), studying butterflies in a tropical woodland rainforest in Australia, reported the role of microhabitats in filtering darker and larger individuals to the forest interior, while smaller species preferred sunny habitats. Although all the riparian forests sampled showed some disturbance level caused by cattle trampling, such as reduced understory (Carlucci et al., 2016), they support lower temperatures and provide vertical structure, which could be related to other traits like body size and color that reflect distinct adaptive strategies (Shreeve et al., 2001; Xing et al., 2016). This result reinforces the importance of maintenance of the grassland-forest mosaic, which provides suitable habitat at the microhabitat scale and buffers the functional homogenization at the landscape scale (Andrade et al., 2015).

Deterministic processes (environmental filter) are notably relevant in shaping the butterfly communities in the Pampa, but we also observed that intrinsic characteristics of communities also influenced how communities respond to environmental gradients. Communities with more individuals show a larger influence of the environmental filter than the smaller communities. However, we did not find a huge effect of community size in determining the relationship between community composition and environment. By decoupling the proportion of explanation, we observe that forest and grassland proportions are the landscape features that most contribute to the positive association between community size and the environmental filter. This relationship is more evident in the Quaraí communities than in other locals since Quaraí sites remain in the highly conserved natural landscape (Andrade et al., 2015). Due to high stability, species can be more adapted to survive and persist at low disturbed sites, increasing the match between niche and environment. On the other hand, EEA2 has the smallest community size and has the highest proportion of variance explained by random components. Interestingly, it is in an anthropized matrix, either due to the presence of plantations or the proximity to large urban centers. These results may indicate that increasing land use change may reduce our ability to predict community response because we lose the correspondence between species and their niches, making small communities more fragile and subject to the effect of random factors (Siqueira et al., 2020).

## Concluding remarks

The occurrence of fruit-feeding butterflies, as well as of other taxa (Dala-Corte et al., 2016; Pires et al., 2022; Renner et al., 2018) in the Brazilian Pampa biome, is structured by both landscape and microclimate variables, influencing the species all aspects of community structures. In general, our results support that the homogenization of the landscape results in communities composed of more generalist species, and therefore, these findings show the importance of maintaining the mosaic of forested and grassland areas in the Pampa biome (Luza et al., 2014). Furthermore, in the Pampa biome, the new agricultural lands replace mainly native grasslands and have been growing in recent years (IBGE, 2020). This can either lead to a change in the cooccurrence of species due to change in land cover, as the extinction of grassland specialist species, modifying the community composition, ecological interactions, and even the ecosystem functioning (Börschig et al., 2013). We hope that these results will improve our understanding of how ecological, evolutionary, and stochastic processes act to shape communities, especially in highly threatened ecosystems such as the Pampa.

## References

Ackerly, D. D. (2003). Community assembly, niche conservativism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, *164*(S3), S165–S184. https://doi.org/10.1086/368401

Andrade, B. O., Koch, C., Boldrini, I. I., Vélez-Martin, E., Hasenack, H., Hermann, J. M., Kollmann, J., Pillar, V. D., & Overbeck, G. E. (2015). Grassland degradation and restoration: A conceptual framework of stages and thresholds illustrated by southern Brazilian grasslands. *Natureza e Conservacao*, *13*(2), 95–104. https://doi.org/10.1016/j.ncon.2015.08.002

Baeza, S., Baldassini, P., Bagnato, C., Pinto, P., & Paruelo, J. M. (2014). Caracterización del uso / cobertura del suelo en Uruguay a partir de series temporales de imágenes MODIS Land Use / Land Cover Classification in Uruguay Using Time Series of MODIS Images. *Agrociencia Uruguay*, *18*(2), 95–105. http://www.scielo.edu.uy/scielo.php?script=sci_arttext&pid=S2301-15482014000200011&nrm=iso

Baeza, S., & Paruelo, J. M. (2020). Land use/land cover change (2000-2014) in the rio de la plata grasslands: An analysis based on MODIS NDVI time series. *Remote Sensing*, *12*(3), 1–22. https://doi.org/10.3390/rs12030381

Börschig, C., Klein, A.-M., von Wehrden, H., & Krauss, J. (2013). Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic and Applied Ecology*, *14*(7), 547–554. https://doi.org/10.1016/j.baae.2013.09.002

Brown, K. S., & Freitas, A. V. L. (2000). Atlantic Forest Butterflies: Indicators for Landscape Conservation. *Biotropica*, *32*(4b), 934–956. https://doi.org/10.1111/j.1744-7429.2000.tb00631.x

Carlucci, M. B., Luza, A. L., Hartz, S. M., & Duarte, L. D. S. (2016). Forests, shrublands and grasslands in southern Brazil are neglected and have specific needs for their conservation. Reply to Overbeck et al. *Natureza e Conservacao*, *14*(2), 155–157. https://doi.org/10.1016/j.ncon.2016.08.001

Chai, P., & Srygley, R. B. (1990). Predation and the Flight, Morphology, and Temperature of Neotropical Rain-Forest Butterflies. *The American Naturalist*, *135*(6), 748–765.

Checa, M. F., Levy, E., Rodriguez, J., & Willmott, K. (2019). Rainfall as a significant contributing factor to butterfly seasonality along a climatic gradient in the Neotropics. *BioRxiv*, *May*. https://doi.org/10.1101/630947

Checa, M. F., Rodriguez, J., Willmott, K. R., & Liger, B. (2014). Microclimate Variability Significantly Affects the Composition, Abundance and Phenology of Butterfly Communities in a Highly Threatened Neotropical Dry Forest. *Florida Entomologist*, *97*(1), 1–13. https://doi.org/10.1653/024.097.0101

Curtis, R. J., Brereton, T. M., Dennis, R. L. H., Carbone, C., & Isaac, N. J. B. (2015). Butterfly abundance is determined by food availability and is mediated by species traits. *Journal of Applied Ecology*, *52*(6), 1676–1684. https://doi.org/10.1111/1365-2664.12523

Dala-Corte, R. B., Giam, X., Olden, J. D., Becker, F. G., Guimarães, T. de F., & Melo, A. S. (2016). Revealing the pathways by which agricultural land-use affects stream fish communities in South Brazilian grasslands. *Freshwater Biology*, *61*(11), 1921–1934. https://doi.org/10.1111/fwb.12825

DeVries, P. J. (1987). *The Butterflies of Costa Rica and Their Natural History: Papilionidae, Pieridae, Nymphalidae* (Issue v. 1). Princeton University Press. https://books.google.com.br/books?id=SqR1QgAACAAJ

Ekroos, J., Rundlöf, M., & Smith, H. G. (2013). Trait-dependent responses of flower-visiting insects to distance to semi-natural grasslands and landscape heterogeneity. *Landscape Ecology*, *28*(7), 1283–1292. https://doi.org/10.1007/s10980-013-9864-2

Filho, P. J. S. da S., Macedo, R. B., Vieira, M. de S., & das Neves, P. C. P. (2017). Floristic and structure of grassland vegetation in the Campos Arbustivos of São Gabriel (Arroio do Salso headwaters), Rio Grande do Sul, Brazil. *Iheringia - Serie Botanica*, *72*(3), 351–372. https://doi.org/10.21826/2446-8231201772305

Freire, G. B., Silva, T., Oliveira, H., Collier, C., Rodrigues, H. P., Dias, J. P., Santos, J. P., Marini-Filho, O. J., Freitas, A. V. L., Smilanich, A. M., Dyer, L. A., & Diniz, I. R. (2021). Good Things Come in Larger Packages: Size Matters for Adult Fruit-Feeding Butterfly Dispersal and Larval Diet Breadth. *Diversity*, *13*(12), 664. https://doi.org/10.3390/d13120664

Freitas, A. V. L., Iserhard, C. A., Santos, J. P. dos, Carreira, J. Y. O., Ribeiro, D. B., Melo, D. H. A., Rosa, A. H., Marini-Filho, O. J., Mattos Accacio, G., & Uehara-Prado, M. (2014). Studies with butterfly bait traps: An overview. *Revista Colombiana de Entomología*, *40*(2), 203–212. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-04882014000200013&lng=en&nrm=iso&tlng=en

Gámez-Virués, S., Perović, D. J., Gossner, M. M., Börschig, C., Blüthgen, N., de Jong, H., Simons, N. K., Klein, A. M., Krauss, J., Maier, G., Scherber, C., Steckel, J., Rothenwöhrer, C., Steffan-Dewenter, I., Weiner, C. N., Weisser, W., Werner, M., Tscharntke, T., & Westphal, C. (2015). Landscape simplification filters species traits and drives biotic homogenization. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms9568

Graça, M. B., Pequeno, P. A. C. L., Franklin, E., & Morais, J. W. (2017). Coevolution between flight morphology, vertical stratification and sexual dimorphism: what can we learn from tropical butterflies? *Journal of Evolutionary Biology, 30*(10), 1862–1871. https://doi.org/10.1111/jeb.13145

Hall, A. P. W., & Willmott, K. R. (2000). Patterns of feeding behaviour in adult male riodinid butterflies and their relationship to morphology and ecology. *Biological Journal of the Linnean Society*, *69*(June), 1–23. https://doi.org/10.1006/bijL

Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., & Roberts, C. (2004). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, *8*(1), 23–29. https://doi.org/10.1111/j.1461-0248.2004.00686.x

IBGE. (2020). *Contas de Ecossistemas: O uso da terra nos biomas brasileiros* (N. 73). https://biblioteca.ibge.gov.br/visualizacao/livros/liv101753.pdf

Iserhard, C. A., Romanowski, H. P., Richter, A., & Mendonça, M. de S. (2017). Monitoring Temporal Variation to Assess Changes in the Structure of Subtropical Atlantic Forest Butterfly Communities. *Environmental Entomology, 46*(4), 804–813. https://doi.org/10.1093/ee/nvx115

Kotiaho, J. S., Kaitala, V., Komonen, A., & Päivinen, J. (2005). Predicting the risk of extinction from shared ecological characteristics. *Proceedings of the National Academy of Sciences of the United States of America, 102*(6), 1963–1967. https://doi.org/10.1073/pnas.0406718102

Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, *29*(5), 592–599. https://doi.org/10.1111/1365-2435.12345

Lawton, J. H. (1983). Plant Architecture and the Diversity of Phytophagous Insects. *Annual Review of Entomology*, *28*, 23–39. https://doi.org/10.1146/annurev.en.28.010183.000323

Legendre, P., & Legendre, L. (2012). Numerical Ecology. In *Developments in Environmental Modelling* (3rd ed., Vol. 24, p. 1006). Elsevier. https://doi.org/10.1016/S0167-8892(12)70001-5

Lima, D. O. de, Crouzeilles, R., & Vieira, M. V. (2020). Integrating strict protection and sustainable use areas to preserve the Brazilian Pampa biome through conservation planning. *Land Use Policy*, *99*(January 2019), 104836. https://doi.org/10.1016/j.landusepol.2020.104836

Luza, A. L., Carlucci, M. B., Hartz, S. M., & Duarte, L. D. S. (2014). Moving from forest vs. grassland perspectives to an integrated view towards the conservation of forest-grassland mosaics. *Natureza e Conservacao*, *12*(2), 166–169. https://doi.org/10.1016/j.ncon.2014.09.005

McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, *21*(4), 178–185. https://doi.org/10.1016/j.tree.2006.02.002

Medan, D., Torretta, J. P., Hodara, K., de la Fuente, E. B., & Montaldo, N. H. (2011). Effects of agriculture expansion and intensification on the vertebrate and invertebrate diversity in the Pampas of Argentina. *Biodiversity and Conservation*, *20*(13), 3077–3100. https://doi.org/10.1007/s10531-011-0118-9

New, T. R. (1997). Are Lepidoptera an effective “umbrella group” for biodiversity conservation? *Journal of Insect Conservation*, *1*(1), 5–12. https://doi.org/10.1023/A:1018433406701

Öckinger, E., Schweiger, O., Crist, T. O., Debinski, D. M., Krauss, J., Kuussaari, M., Petersen, J. D., Pöyry, J., Settele, J., Summerville, K. S., & Bommarco, R. (2010). Life-history traits predict species responses to habitat area and isolation: a cross-continental synthesis. *Ecology Letters*, *13*(8), no-no. https://doi.org/10.1111/j.1461-0248.2010.01487.x

Öckinger, E., & Smith, H. G. (2007). Semi-natural grasslands as population sources for pollinating insects in agricultural landscapes. *Journal of Applied Ecology*, *44*(1), 50–59. https://doi.org/10.1111/j.1365-2664.2006.01250.x

Ovaskainen, O., & Abrego, N. (2020). Joint Distribution Modelling (with applications in R). In *Joint Species Distribution Modelling*.

Overbeck, G. E., Müller, S. C., Fidelis, A., Pfadenhauer, J., Pillar, V. D., Blanco, C. C., Boldrini, I. I., Both, R., & Forneck, E. D. (2007). Brazil’s neglected biome: The South Brazilian Campos. *Perspectives in Plant Ecology, Evolution and Systematics*, *9*(2), 101–116. https://doi.org/10.1016/j.ppees.2007.07.005

Pavoine, S., Baguette, M., Stevens, V. M., Leibold, M. A., Turlure, C., & Bonsall, M. B. (2014). Life history traits, but not phylogeny, drive compositional patterns in a butterfly metacommunity. *Ecology*, *95*(12), 3304–3313. https://doi.org/10.1890/13-2036.1.sm

Pillar, V. D. P., & Quadros, F. (1999). Grassland-forest boundaries in Southern Brazil. In *Conference on Recent Shifts in Vegetation Boundaries of Deciduous Forests, Especially Due to General Global Warming* (Vol. 12, Issues 2–3, pp. 301–316). Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8722-9_17

Pillar, V. D. P., & Vélez, E. (2010). Extinção dos Campos Sulinos em unidades de conservação: Um fenômeno natural ou um problema ético? *Natureza a Conservacao*, *8*(1), 84–86. https://doi.org/10.4322/natcon.00801014

Pires, M. M., Sahlén, G., & Périco, E. (2022). Agricultural land use affects the heterogeneity of Odonata communities in the Brazilian Pampa. *Journal of Insect Conservation*, *26*(3), 503–514. https://doi.org/10.1007/s10841-021-00349-0

Ramos, R. C., & Koester, E. (2014). Geologia da associação metamáfica-ultramáfica da região de Arroio Grande, sudeste do Escudo Sul-rio-grandense. *Pesquisas Em Geociências, 41*(1), 25. https://doi.org/10.22456/1807-9806.78030

Renner, S., Périco, E., Dalzochio, M. S., & Sahlén, G. (2018). Water body type and land cover shape the dragonfly communities (Odonata) in the Pampa biome, Rio Grande do Sul, Brazil. *Journal of Insect Conservation*, *22*(1), 113–125. https://doi.org/10.1007/s10841-017-0042-8

Roesch, L. F. W., Vieira, F. C. B., Pereira, V. A., Schünemann, A. L., Teixeira, I. F., Senna, A. J. T., & Stefenon, V. M. (2009). The Brazilian Pampa: A fragile biome. *Diversity, 1*(2), 182–198. https://doi.org/10.3390/d1020182

Rohlf, F. J. (2015). The tps series of software. *Hystrix, the Italian Journal of Mammalogy, 26*(1), 1–4. https://doi.org/10.4404/hystrix-26.1-11264

Santos, J. P., Sobral-Souza, T., Brown, K. S., Vancine, M. H., Ribeiro, M. C., & Freitas, A. V. L. (2020). Effects of landscape modification on species richness patterns of fruit-feeding butterflies in Brazilian Atlantic Forest. *Diversity and Distributions, 26*(2), 196–208. https://doi.org/10.1111/ddi.13007

Shreeve, T. G., Dennis, R. L. H., Roy, D. B., & Moss, D. (2001). An ecological classification of British butterflies: ecological attributes and biotope occupancy. *Journal of Insect Conservation*, *5*(Hodgson 1993), 145–161.

Silva, S. D. A., Luiz, J., & Carlos, L. (2004). *Arcabouço Geológico-Geomorfológico Da Bacia Hidrográfica Do Rio Quaraí - Rio Grande Do Sul/Brasil*. 1–15.

Siqueira, T., Saito, V. S., Bini, L. M., Melo, A. S., Petsch, D. K., Landeiro, V. L., Tolonen, K. T., Jyrkänkallio-Mikkola, J., Soininen, J., & Heino, J. (2020). Community size can affect the signals of ecological drift and niche selection on biodiversity. *Ecology*, *101*(6), 1–10. https://doi.org/10.1002/ecy.3014

Srygley, R. B., & Chai, P. (1990). Flight morphology of Neotropical butterflies: palatability and distribution of mass to the thorax and abdomen. *Oecologia*, *84*(4), 491–499. https://doi.org/10.1007/BF00328165

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution*, *11*(3), 442–447. https://doi.org/10.1111/2041-210X.13345

Tscharntke, T., Tylianakis, J. M., Rand, T. A., Didham, R. K., Fahrig, L., Batáry, P., Bengtsson, J., Clough, Y., Crist, T. O., Dormann, C. F., Ewers, R. M., Fründ, J., Holt, R. D., Holzschuh, A., Klein, A. M., Kleijn, D., Kremen, C., Landis, D. A., Laurance, W., … Westphal, C. (2012). Landscape moderation of biodiversity patterns and processes - eight hypotheses. *Biological Reviews*, *87*(3), 661–685. https://doi.org/10.1111/j.1469-185X.2011.00216.x

van Halder, I., Thierry, M., Villemey, A., Ouin, A., Archaux, F., Barbaro, L., Balent, G., & Benot, M. L. (2017). Trait-driven responses of grassland butterflies to habitat quality and matrix composition in mosaic agricultural landscapes. *Insect Conservation and Diversity*, *10*(1), 64–77. https://doi.org/10.1111/icad.12200

WallisDeVries, M. F., & van Swaay, C. A. M. (2006). Global warming and excess nitrogen may induce butterfly decline by microclimatic cooling. *Global Change Biology*, *12*(9), 1620–1626. https://doi.org/10.1111/j.1365-2486.2006.01202.x

Xing, S., Bonebrake, T. C., Tang, C. C., Pickett, E. J., Cheng, W., Greenspan, S. E., Williams, S. E., & Scheffers, B. R. (2016). Cool habitats support darker and bigger butterflies in Australian tropical forests. *Ecology and Evolution*, *6*(22), 8062–8074. https://doi.org/10.1002/ece3.2464

Zobel, M. (1997). The relative role of species pools in determining plant species richness: An alternative explanation of species coexistence? *Trends in Ecology and Evolution*, *12*(7), 266–269. https://doi.org/10.1016/S0169-5347(97)01096-3

## Supplementary material

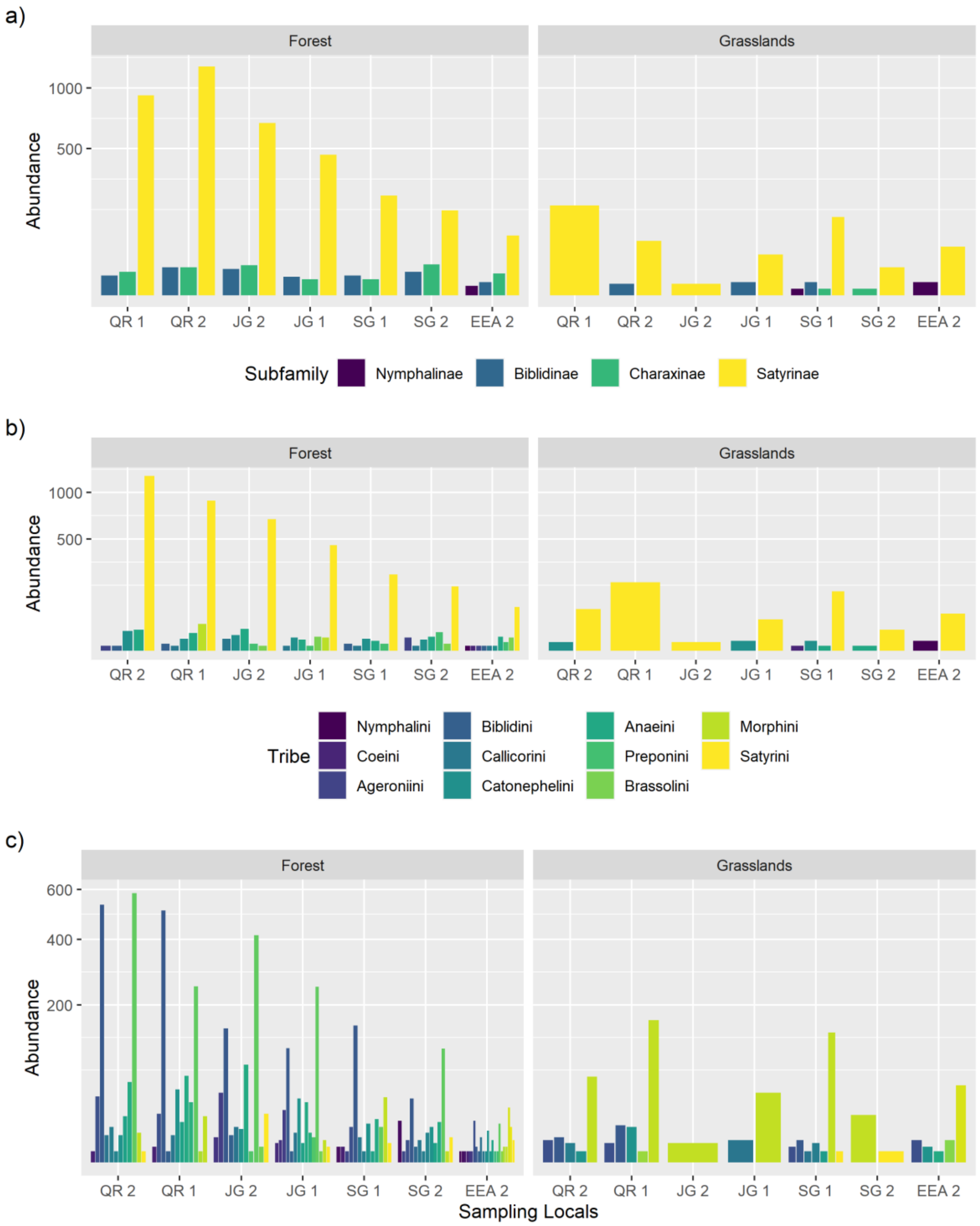

**Figure S1.** Distribution of fruit-feeding butterflies' abundances at the subfamily level (a), the Tribe level (b), and by species (c). The data was split between forest and grassland habits, and by communities – Estação Experimental Agronômica (EEA), Jaguarão (JG), Quaraí (QR), and São Gabriel (SG). The numbers that followed the community’s name indicate the survey season: 1 for summer 2020, and 2 for summer 2021. The abundance data (y-axis) was squared-rooted for better visualization of the data.

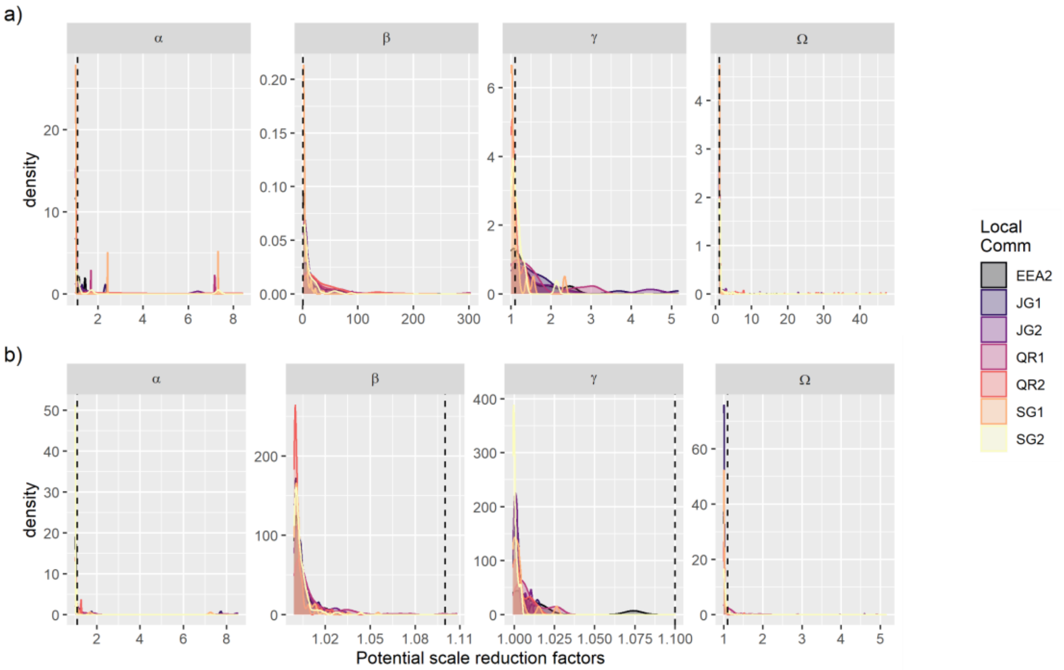

**Figure S2.** MCMC convergence diagnosis for the parameters estimated by the two model structures implemented: a) the presence-absence data with a probit distribution (PA), and b) the relative abundance data with a normal model (Normal), measured in terms of the potential scale reduction factor. The convergence of MCMC was evaluated for each community, represented by the colors. α – parameter related to the spatial structure of the model; β – parameter related to the species niche; γ – parameter related to the effects of traits on species niche; Ω - parameters related to the phylogenetic signal of the species' niche. The dashed line represents the expected threshold of 1.1, which indicates a good convergence of MCMC.

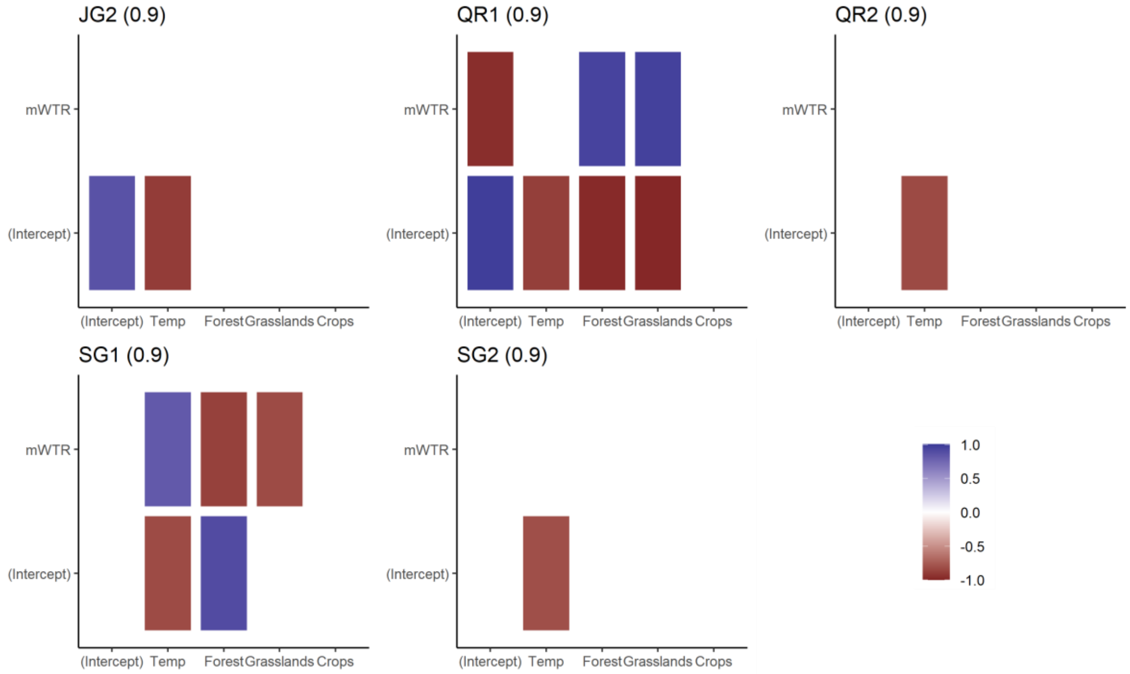

**Figure S3**. Heatmap of estimated γ parameters linking the mean wing-thorax ratio (mWTR) to species niches. Blue and red colors represent parameters that are estimated to be positive and negative, respectively, with at least 0.90 posterior probability. JG – Jaguarão; QR – Quaraí; SG – São Gabriel. The numbers that follow the community names indicate the sampling occasion.

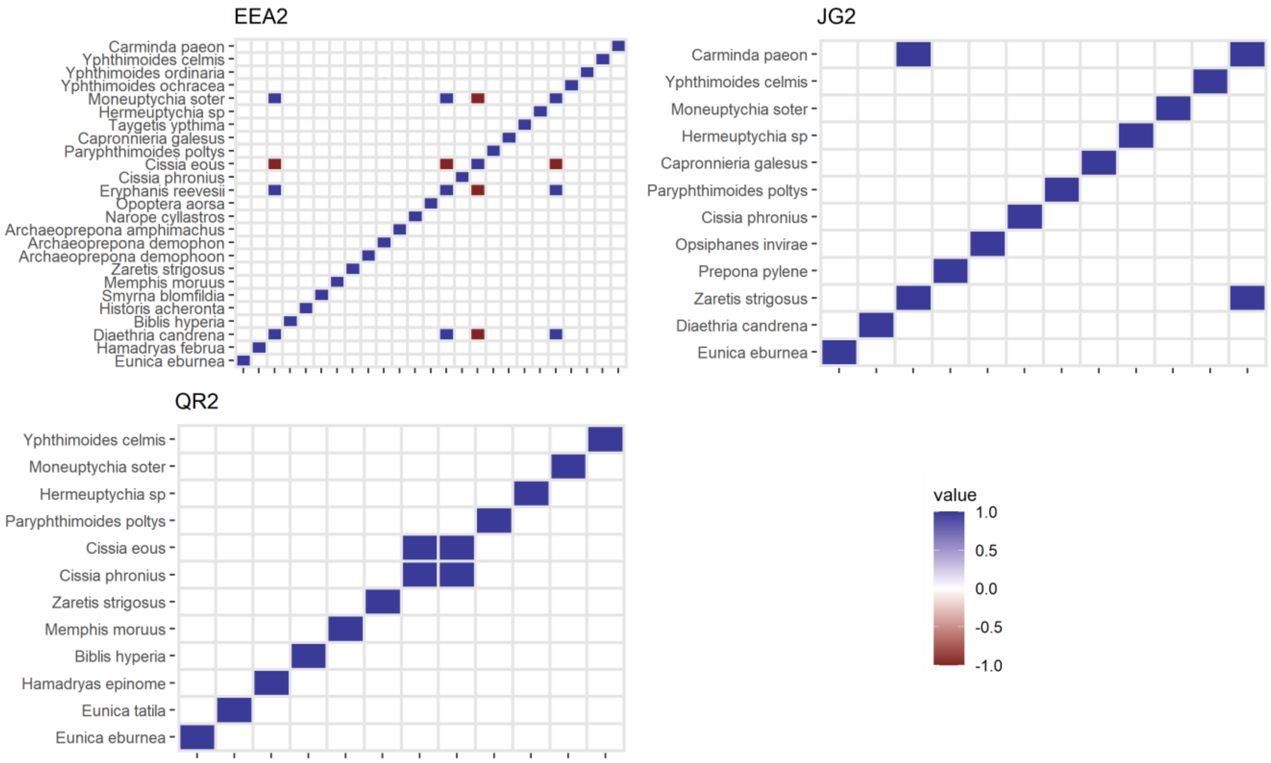

**Figure S4**. Residual species association (Ω) for presence-absence data (model PA) showing the associations that were positive or negative with at least 0.90 posterior probability.

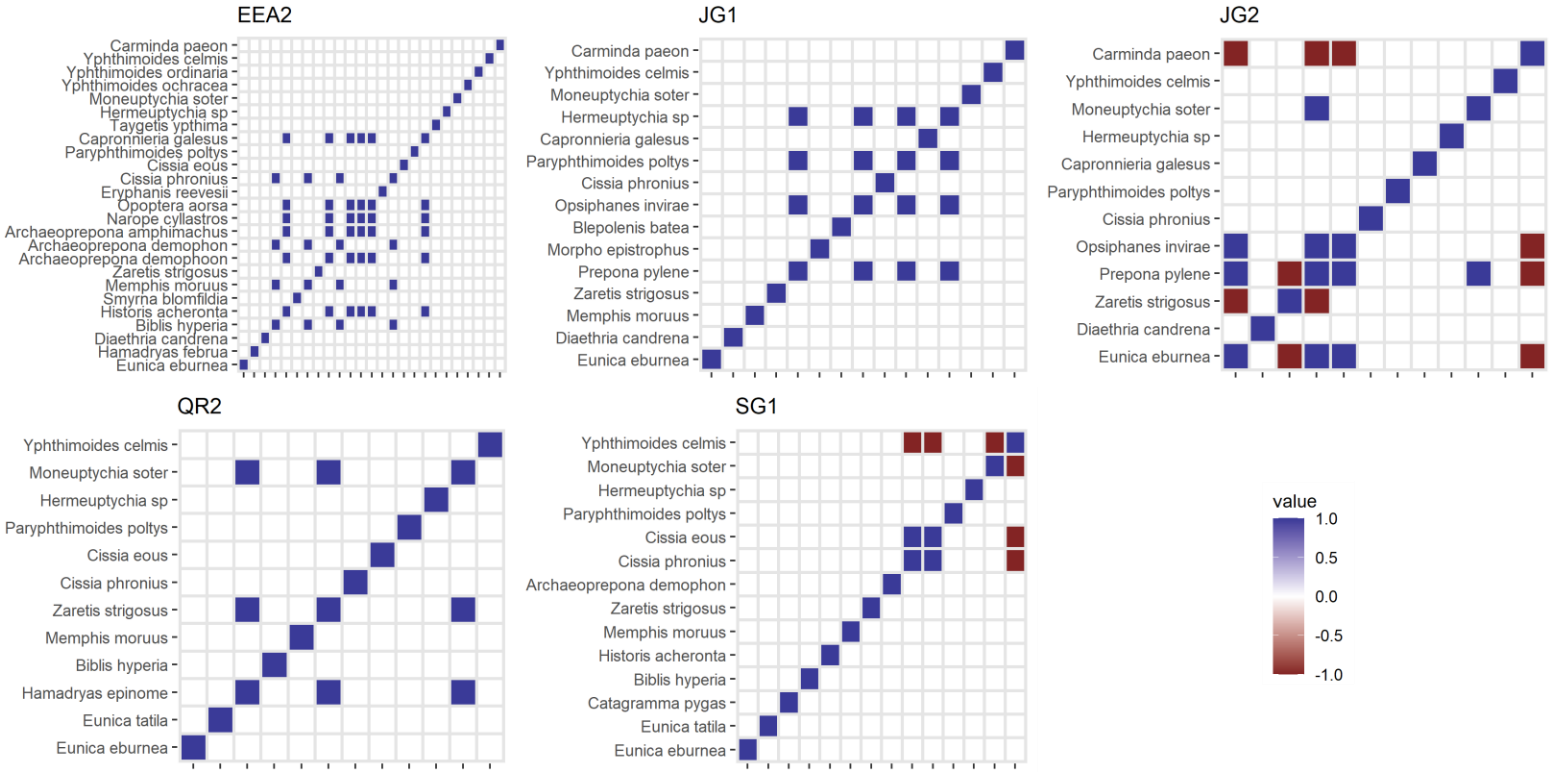

**Figure S5**. Residual species association (Ω) for relative abundance data (model Normal) showing the associations that were positive or negative with at least 0.90 posterior probability.

**Table S1**. List of fruit-feeding butterflies species sampled in four localities of the Brazilian Pampa – Estação Experimental Agronômica (EEA), Jaguarão (JG), Quaraí (QR), and São Gabriel (SG) – at native grasslands and riparian forest plots. The samples was collected in two summer seasons, one between December 2019 and February 2020 and another between November 2020 and March 2021.

| | Native Grasslands | | | | Riparian Forest | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| **Subfamily**/Tribe/*Species* | EEA | JG | QR | SG | EEA | JG | QR | SG | |
| **Biblidinae** | | **4** | **3** | **4** | **4** | **24** | **27** | **22** | **88** |
| Ageroniini | | | | | 1 | | 1 | 7 | 9 |
| *Hamadryas epinome* (C. Felder & R. Felder, 1867) | | | | | | | 1 | 7 | 8 |
| *Hamadryas februa* (Hübner, [1823]) | | | | | 1 | | | | 1 |
| Biblidini | | | | | 1 | | 3 | 2 | 6 |
| *Biblis hyperia* (Cramer, 1779) | | | | | 1 | | 3 | 2 | 6 |
| Callicorini | | | | | 1 | 7 | 1 | 2 | 11 |
| *Catagramma pygas* (Godart, [1824]) | | | | | | | | 2 | 2 |
| *Diaethria candrena* (Godart, [1824]) | | | | | 1 | 7 | 1 | | 9 |
| Catonephelini | | 4 | 3 | 4 | 1 | 17 | 22 | 11 | 62 |
| *Eunica eburnea* Fruhstorfer, 1907 | | 4 | | 1 | 1 | 17 | 12 | 9 | 44 |
| *Eunica tatila* (Herrich-Schäffer, [1855]) | | | 3 | 3 | | | 10 | 2 | 18 |
| **Charaxinae** | | | | **2** | **13** | **27** | **31** | **28** | **101** |
| Anaeini | | | | 2 | 8 | 24 | 31 | 12 | 77 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Memphis moruus* (Fabricius, 1775) | | | | | 4 | 3 | 30 | 4 | 41 |
| *Zaretis strigosus* (Gmelin, [1790]) | | | | 2 | 4 | 21 | 1 | 8 | 36 |
| Preponini | | | | | 5 | 3 | | 16 | 24 |
| *Archaeoprepona amphimachus* (Fabricius, 1775) | | | | | 1 | | | | 1 |
| *Archaeoprepona demophon* (Linnaeus, 1758) | | | | | 2 | | | 16 | 18 |
| *Archaeoprepona demophoon* (Hübner, [1814]) | | | | | 2 | | | | 2 |
| *Prepona pylene* Hewitson, [1854] | | | | | | 3 | | | 3 |
| **Nymphalinae** | **4** | | | **1** | **2** | | | | **7** |
| Coeini | | | | 1 | 1 | | | | 2 |
| *Historis achaeronta* (Fabricius, 1775) | | | | 1 | 1 | | | | 2 |
| Nymphalini | 4 | | | | 1 | | | | 5 |
| *Smyrna blomfildia* (Fabricius, 1781) | 4 | | | | 1 | | | | 5 |
| **Satyrinae** | **66** | **42** | **257** | **160** | **205** | **1150** | **2146** | **399** | **4425** |
| Brassolini | | | | | 7 | 9 | | 2 | 18 |
| *Blepolenis batea* (Hübner, [1821]) | | | | | | 3 | | | 3 |
| *Dinastor darius* (Fabricius, 1775) | | | | | | | | 2 | 2 |
| *Eryphanis reevesii* (E. Doubleday, [1849]) | | | | | 5 | | | | 5 |
| *Narope cyllastros* E. Doubleday, [1849] | | | | | 1 | | | | 1 |
| *Opoptera aorsa* (Godart, [1824]) | | | | | 1 | | | | 1 |
| *Opsiphanes invirae* (Hübner, [1808]) | | | | | | 6 | | | 6 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Morphini | | | | | | 7 | 29 | | 36 |
| *Morpho epistrophus* (Fabricius, 1796) | | | | | | 7 | 29 | | 36 |
| Satyrini | 66 | 42 | 257 | 160 | 198 | 1134 | 2117 | 397 | 4371 |
| *Capronnieria galesus* (Godart, [1824]) | | | | | 3 | 9 | | | 12 |
| *Carminda paeon* (Godart, 1824) | | | | | 5 | 61 | | | 66 |
| *Cissia eous* (Butler, 1867) | 4 | | 7 | 2 | 14 | | 54 | 14 | 95 |
| *Cissia phronius* (Godart, [1824]) | | | 16 | 4 | 28 | 249 | 1044 | 184 | 1525 |
| *Erichthodes narapa* (Schaus, 1902) | 9 | | | | | | | | 9 |
| *Hermeuptychia sp* Forster, 1964 | 2 | | 11 | | 11 | 42 | 49 | 22 | 137 |
| *Moneuptychia soter* (Butler, 1877) | 1 | | | | 24 | 106 | 112 | 28 | 271 |
| *Paryphthimoides poltys* (Prittwitz, 1865) | | | | | 74 | 663 | 833 | 114 | 1684 |
| *Stegosatyrus periphas* (Godart, [1824]) | 1 | | 1 | | | | | | 2 |
| *Taygetis ypthima* Hübner, [1821] | | | | | 3 | | 1 | | 4 |
| *Yphthimoides affinis* (A. Butler, 1867) | 1 | | | | | | | | 1 |
| *Yphthimoides celmis* (Godart, 1824]) | | 42 | 222 | 154 | 2 | 4 | 24 | 35 | 483 |
| *Yphthimoides ochracea* (Butler, 1867) | 48 | | | | 24 | | | | 72 |
| *Yphthimoides ordinaria* Freitas, Kaminski & Mielke, 2012 | | | | | 10 | | | | 10 |
| Total | 70 | 46 | 260 | 167 | 224 | 1201 | 2204 | 449 | 4621 |

**Table S2.** Potential scale reduction factor for the ρ parameters, considering two model structures: PA – presence-absence data with a probit distribution, and Normal – the Hellinger transformed abundance data with a normal distribution. Values in italic indicate values higher than the acceptable threshold of 1.1.

| Model | Local | Estimated | Upper CI |
|---|---|---:|---:|
| PA | EEA2 | 1.00 | 1.00 |
| | JG1 | 1.10 | *1.35* |
| | JG2 | 2.46 | *5.27* |
| | QR1 | 1.08 | *1.15* |
| | QR2 | 1.00 | 1.01 |
| | SG1 | 1.02 | 1.02 |
| | SG2 | 1.16 | *1.55* |
| Normal | EEA2 | 1.00 | 1.00 |
| | JG1 | 1.00 | 1.00 |
| | JG2 | 1.00 | 1.01 |
| | QR1 | 1.00 | 1.00 |
| | QR2 | 1.00 | 1.00 |
| | SG1 | 1.00 | 1.00 |
| | SG2 | 1.00 | 1.00 |

**Table S3**. Proportions of variance in species niches explained by the trait (mWTR) included in the model for presence-absence data (PA) and relative abundance data (Normal), as well as the proportion of variance that the mWTR explain out of species occurrence/relative abundances (Y).

| Model | Local | Intercept | Temp | Forest | Grassland | Crop | Y |
|---|---|---|---|---|---|---|---|
| PA | EEA2 | 16.37 | 14.66 | 18.63 | 12.75 | 11.48 | 4.39 |
| | JG1 | 16.59 | 16.95 | 15.63 | 16.22 | 18.18 | 12.32 |
| | JG2 | 10.24 | 11.10 | 3.07 | 2.66 | 1.60 | 7.40 |
| | QR1 | 56.98 | 5.21 | 61.22 | 60.12 | 82.97 | 72.86 |
| | QR2 | 1.74 | 2.66 | 1.56 | 3.79 | 63.56 | 28.82 |
| | SG1 | 36.21 | 28.49 | 21.23 | 48.27 | 65.19 | 60.70 |
| | SG2 | 11.90 | 7.56 | 59.56 | 37.99 | 48.38 | 32.26 |
| Normal | EEA2 | 11.57 | 7.83 | 33.34 | 34.33 | 33.28 | 10.66 |
| | JG1 | 9.25 | 1.58 | 39.61 | 52.89 | 51.62 | 4.06 |
| | JG2 | 14.92 | 7.08 | 33.01 | 49.88 | 48.49 | 11.65 |
| | QR1 | 51.16 | 16.19 | 79.28 | 83.06 | 15.39 | 28.65 |
| | QR2 | 22.11 | 12.36 | 56.87 | 57.65 | 16.96 | 14.16 |
| | SG1 | 10.13 | 12.31 | 57.31 | 52.89 | 29.77 | 22.91 |
| | SG2 | 8.38 | 5.20 | 37.37 | 40.48 | 25.42 | 18.67 |

**Table S4**. Estimated α and ρ parameters. Values in bold indicate a significative signal. We chose to show only the two leading latent factors (factor1 and factor2) because they explain most of the variation.

| | | | Probit | | | Lognormal Poisson | | | Normal | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2.5% | 50% | 97.5% | 2.5% | 50.0% | 97.5% | 2.5% | 50% | 97.5% |
| EEA2 | α | factor1 | 0.00 | 0.00 | 0.02 | 0.00 | 0.01 | 0.04 | 0.00 | 0.00 | 0.00 |
| | | factor2 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.03 |
| | ρ | | 0.00 | 0.12 | 0.39 | 0.00 | 0.15 | 0.64 | **0.83** | **0.97** | **1.00** |
| JG1 | α | factor1 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 |
| | | factor2 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| | ρ | | 0.00 | 0.00 | 0.41 | 0.00 | 0.24 | 0.88 | 0.00 | 0.81 | 1.00 |
| JG2 | α | factor1 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 |
| | | factor2 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| | ρ | | 0.00 | 0.25 | 0.78 | 0.00 | 0.00 | 0.62 | 0.00 | 0.82 | 1.00 |
| QR1 | α | factor1 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 |
| | | factor2 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| | ρ | | 0.00 | 0.00 | 0.41 | 0.00 | 0.30 | 0.90 | 0.00 | 0.76 | 1.00 |
| QR2 | α | factor1 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 |
| | | factor2 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.04 |
| | ρ | | 0.00 | 0.10 | 0.71 | 0.00 | 0.00 | 0.91 | 0.00 | 0.45 | 0.98 |
| SG1 | α | factor1 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 |
| | | factor2 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 |
| | ρ | | 0.00 | 0.13 | 0.69 | 0.00 | 0.42 | 0.99 | 0.00 | 0.82 | 1.00 |
| SG2 | α | factor1 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.02 |
| | | factor2 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.03 |
| | ρ | | 0.00 | 0.28 | 0.73 | 0.00 | 0.05 | 0.93 | 0.00 | 0.76 | 0.99 |

# Capítulo 3 – Processos que moldam a diversidade em macro escala

## THE HOTSPOTS INSIDE THE HOTSPOT: REVEALING IMPORTANT PROCESSES THAT SHAPE THE ATLANTIC FOREST DIVERSITY

1- Capítulo escrito na forma de artigo cientifico, em inglês e formatado de acordo com as normas da revista *Journal of Animal Ecology*.

2- Este capítulo também tem como coautores: Gabriel Nakamura, Cristiano Agra Iserhard e Leandro da Silva Duarte

## Abstract

1- The Atlantic Forest is among the world’s top five biodiversity hotspots and currently, only 40% of its total area is composed by natural coverage. Here, we evaluated the role of the landscape and climatic variables in defining species niches and how response traits and phylogenetic relationships mediate species responses. Also, we mapped the richness and functional composition of fruit-feeding butterflies along the Trinational Atlantic Forest, highlighting centers of diversity as well as the main filters that determine species distributions.

2- We used occurrence data of the fruit-feeding butterflies of 60 communities along the Atlantic Forest. We employed the Hierarchical Modelling of Species Communities framework to model the species niche and community-level responses to environmental filters (macroclimatic and landscape variables) and to predict the spatial distribution of richness and community-weighted mean traits.

3- Fruit-feeding butterfly communities are mainly composed of rare species (low occurrence), and species occurrence was determined primarily by macroclimatic conditions. The traits explained more than 50% of the variation in species niches, and we detected a phylogenetic niche conservatism in species responses. At the community level, richness and aspect ratio have similar responses to environmental variables, mainly positive. Community-mean body size respond negatively to open formations, temperature, and elevation.

4- Despite the high rates of landscape modification in the Atlantic Forest, the macroclimatic variables guided species’ responses to environmental filters. We observed joint responses of phylogenetic closed related species, indicating the strength of niche conservatism for fruit-feeding butterflies and corroborating the group-level response in this guild. On the macroecological scale, the dispersal trait is correlated with species richness, indicating that dispersion may be important to maintain the high levels of richness, mainly in the Araucarian transitional zone. On the other hand, the body size that responds to seasonality at small scales is constrained by elevation on larger scales, supporting the converse-Bergmann rule for this guild.

5- Our results enhance the importance of both landscape and climatic variables in structure species occurrence in the Atlantic Forest, emphasizing the role of traits and phylogeny in defining species niches. Identifying how macroecological patterns are linked to environmental predictors can improve our understanding of the key processes that make the Atlantic Forest a diversity hotspot.

**Keywords**: community assembly, environmental filters, fruit-feeding butterflies, Joint Species Distribution Models (JSDM), landscape, phylogenetic niche conservatism (PNC), species traits.

## Introduction

The Atlantic Forest is globally recognized for its biodiversity and is among the world's top five biodiversity hotspots (Marques & Grelle, 2021; Myers et al., 2000). However, since the arrival of the Portuguese colonizers, a long history of human interaction and transformation of this biome began, which profoundly altered its landscape, leaving only 12.6% of native vegetation cover in Brazil (IBGE 2020; Marques and Grelle 2021). Natural coverage (including forest and non-forest formation) represents approximately 40% of the total area of the Atlantic Forest, while around 57% of Atlantic Forest coverage is attributed to anthropic use (urbanization and agriculture/farming) (Projeto MapBiomas Mata Atlântica Trinacional, 2022). Changes in the landscape are expected to cause changes in local communities altering habitat structure, food resources, and species interactions, which are related to niches and processes that define species occurrences (Ovaskainen et al., 2017; Sobral-Souza et al., 2021). In addition to altering the taxonomic composition, landscape modifications can produce a homogenization of functional traits, shifting the communities from specialists to generalist species and hence modifying ecological interactions (Bagchi et al., 2018; Börschig et al., 2013; Gámez-Virués et al., 2015). In this way, determining to what extent species-specific traits are linked to environmental filtering and whether phylogenetically related species exhibit shared environmental responses could help us understand how climate and landscape changes affect species distributions in this threatened biome (Marques & Grelle, 2021; Poggiato et al., 2021; Tikhonov et al., 2020).

Community data obtained from non-manipulative observational approaches are shaped by the full complexity of the assembly processes, but since these processes can rarely be observed directly, we end up inferring them based on patterns (Ovaskainen et al., 2019; Pilowsky et al., 2022). However, the same pattern can be generated by distinct processes, and this ambiguity does not allow us to assess the importance of each process (ecological, evolutionary, or stochastic) in structuring communities (Pilowsky et al., 2022). Model-based approaches are a rising field within community ecology. They seek to incorporate the assembly processes that determine variation in species occurrence, identity, and traits across space and time (Tikhonov et al., 2020). At the macroecological scale, Species Distribution Models (SDM) are largely used to infer species-specific and community-level responses (stacked) to environmental variables, however this analytical tool models species independently, restricting our understanding of biotic filtering, evolutionary constraints, and distribution of rare species (Ovaskainen & Abrego, 2020). To address non-independence in species response, Joint Species Distribution Models (JSDM) explicitly accommodate the multivariate nature of biological communities by linking species responses to the environment (Poggiato et al. 2021). These joint responses of species could improve our ability to model biological communities, even those dominated by rare species (Ovaskainen & Abrego, 2020), opening new possibilities to answer questions about species interactions (Tobler et al., 2019) but also to test directly ecological hypotheses (Ellison, 1996).

Within modern theories of community assembly, the Assembly Rules Framework refers to any process that facilitates or restricts the occurrence of a species (Keddy, 1992). Processes can be viewed as filters that operate at different scales and ultimately define species composition in local communities. Within this framework of filters, processes related to large-scale historical patterns of speciation and migration are treated distinctly from ecological processes, which determine species composition due to small-scale dispersal events, abiotic variation (environmental filters), and biotic interactions (biotic filters) (Keddy, 1992; Ovaskainen & Abrego, 2020). Further, species traits and phylogenetic relationships can influence how species respond to environmental filters. Traits related to dispersal and competitive ability can indicate which species can reach and colonize certain areas and which species can secure adequate resources (McGill et al., 2006; Ovaskainen et al., 2017).

Including functional traits rather than focusing on the species per se allows for a better understanding of assembly processes (Pavoine & Bonsall, 2011) since we can link these processes with speciation and adaptation if we consider how these traits evolved (Ovaskainen & Abrego, 2020).

Macroecological variables can influence biodiversity patterns by modifying species richness directly through physiological constraints or indirectly by modulating energy availability in the system, i.e., host plants (Menéndez et al., 2007; Wright, 1983). In a recent study, Santos et al. (2020) found that climate and landscape predict high species richness for fruit-feeding butterflies in zones described as centers of endemism (*sensu* da Silva and Casteleti 2003) in the Atlantic Forest. Fruit-feeding butterflies are recognized as relevant study models because their response to environmental variations is rapid and robust, and they serve as a proxy for the response of other groups (Brown & Freitas, 2000). Despite the important results brought by the study of Santos et al. (2020) and the implications for conservation, the authors ignore biotic interactions and evolutionary processes when modeling species niches. Identifying and including ecological traits that influence species' response to environmental conditions can improve predictions of species occurrence and community-level responses to climate change (Ovaskainen & Abrego, 2020; Xing et al., 2018).

Furthermore, we can access ecogeographic patterns related to functional composition and relate this pattern to ecosystem functioning (biomass ratio hypothesis) (Grime, 1998) since the functional traits of dominant species are expected to be correlated to ecologically meaningful properties of communities (Duarte et al., 2018; Ricotta & Moretti, 2011). A recognized macroecological pattern for animals is the Bergmann rule, which postulates that body size is expected to increase with latitude/altitude (Shelomi, 2012). However, this relationship may be inverse for ectothermic organisms, such as butterflies, i.e., the organisms are smaller as latitude/altitude increases, resulting in the converse Bergmann rule. The converse Bergmann rule arises in insects due to restrictions in the "optimal time window", which decreases the time for foraging and growth, reducing body size (Lehnert et al., 2012; Shelomi, 2012).

Here we modeled the occurrence of fruit-feeding butterflies along the Trinational Atlantic Forest (TAF), employing a JSDM framework conceptually linked to the assembly rules framework. Specifically, we investigated: 1) how much variation in species occurrence is due to environmental filtering, co-occurrences, and random processes? 2) how do species traits and phylogenetic relationships correlate with ecological niches? 3) how do predicted richness and functional composition vary in response to environmental conditions and 4) how are these patterns distributed along TAF? We expected that 1) species will respond positively to an increase in the mean temperature and precipitation, as well as natural coverage (open or forest natural proportion), and negatively to anthropogenic disturbances (urbanization and agricultural proportion) and altitude, but due the high habitat heterogeneity of TAF (Marques & Grelle, 2021), landscape variables will be more critical than macroclimatic variables to explain variation in species occurrences; 2) Species with small body size and fast-flights (low aspect ratio) will increase their occurrence probabilities under human-modified landscapes. Large species need more resources for their development and are more vulnerable to habitat loss than smaller ones (Freire et al., 2021; Shahabuddin & Terborgh, 1999), while mobile organisms can cross barriers and exploit ephemeral and unpredictable resources (Denno et al., 1996; Shahabuddin & Terborgh, 1999). 3) At the community level, we expect an increase in species richness as forest cover, temperature, and precipitation increase, while community-weighted mean (CWM) will be associated with human-modified landscapes, in which we will observe a reduction in mean body size and increasing in dispersive capacity for communities under intensive land use; 4) at a macroecological scale, we expect to find hotspots matching known endemism centers in TAF (Silva and Casteletti, 2003) and a converse-Bergmann rule, where insects get smaller as latitude/altitude increases.

## Methods

### *Study area*

TAF is the second-largest forest in South America, being that 93% of its distribution is located in Brazilian territory, followed by Paraguay (5.3%) and Argentina (1.7%) (Fig. 1). The limits of TAF that we used in this study was defined based on the spatial file provided by MapBiomas Mata Atlantica Trinacional project (https://bosqueatlantico.mapbiomas.org/). Given its extent, the TAF has a wide range in both latitude and longitude (approximately 31° in latitude and 23° in longitude), covering everything from coastal environments to the interior of the continent, as well as experiencing distinct regimes of climatic conditions. The climate, according to the Köppen-Geiger climate classification, includes the major types A (tropical), B (arid), and C (temperate) (Peel et al., 2007). All these heterogeneities can be related to distance from the tropics or by association with high-altitude areas (Marques & Grelle, 2021). Unfortunately, the extent of the TAF does not represent the extent of the natural forest formation but rather a mosaic of forest fragments with anthropized areas. In Figure 2, we showed the variation in macroclimatic conditions (Fig 2A and B), topographic structure (Fig 2C), and land use and land cover (Fig. 2D) over TAF.

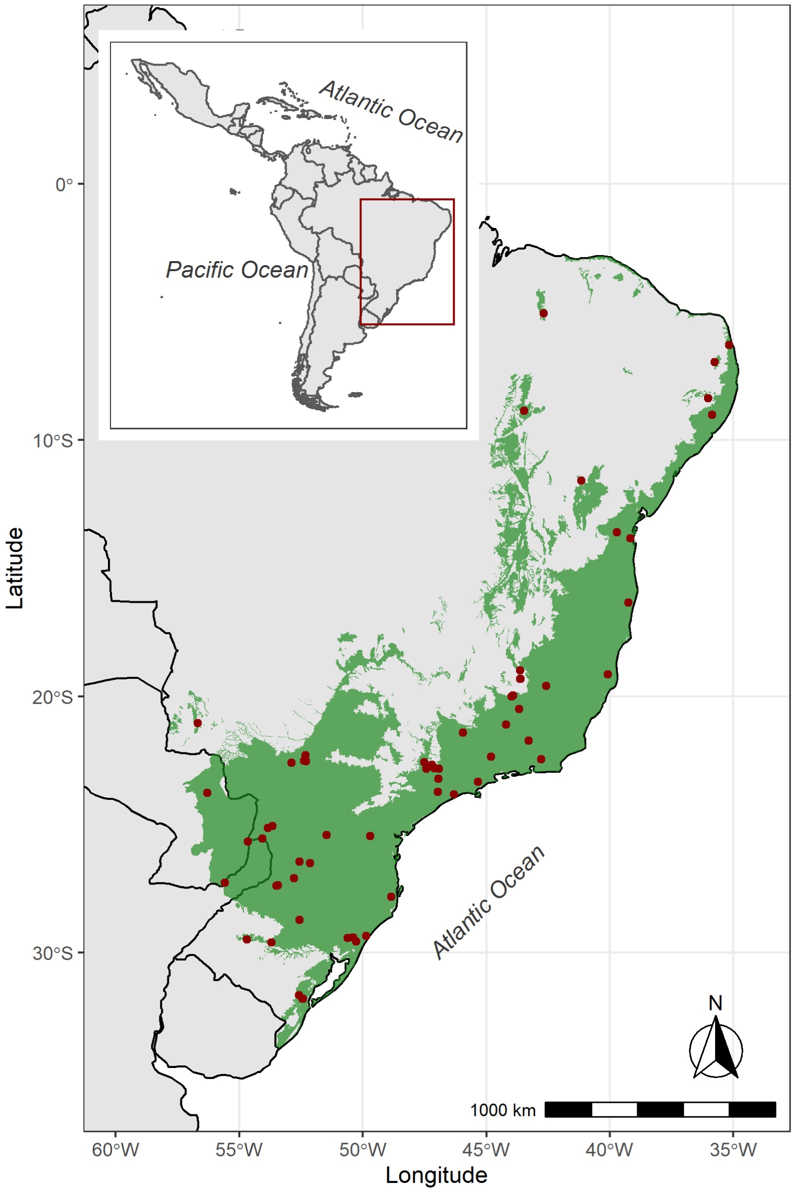

**Figure 10**. Map of the TAF domain (green polygon), which covers the coastal portion of Brazil, a small proportion of Paraguay, and Argentina. The red dots represent the coordinates of the fruit-feeding butterfly communities accessed in this study (n = 60). The black lines indicate country limits. The inset shows the position of the study site within the South American continent.

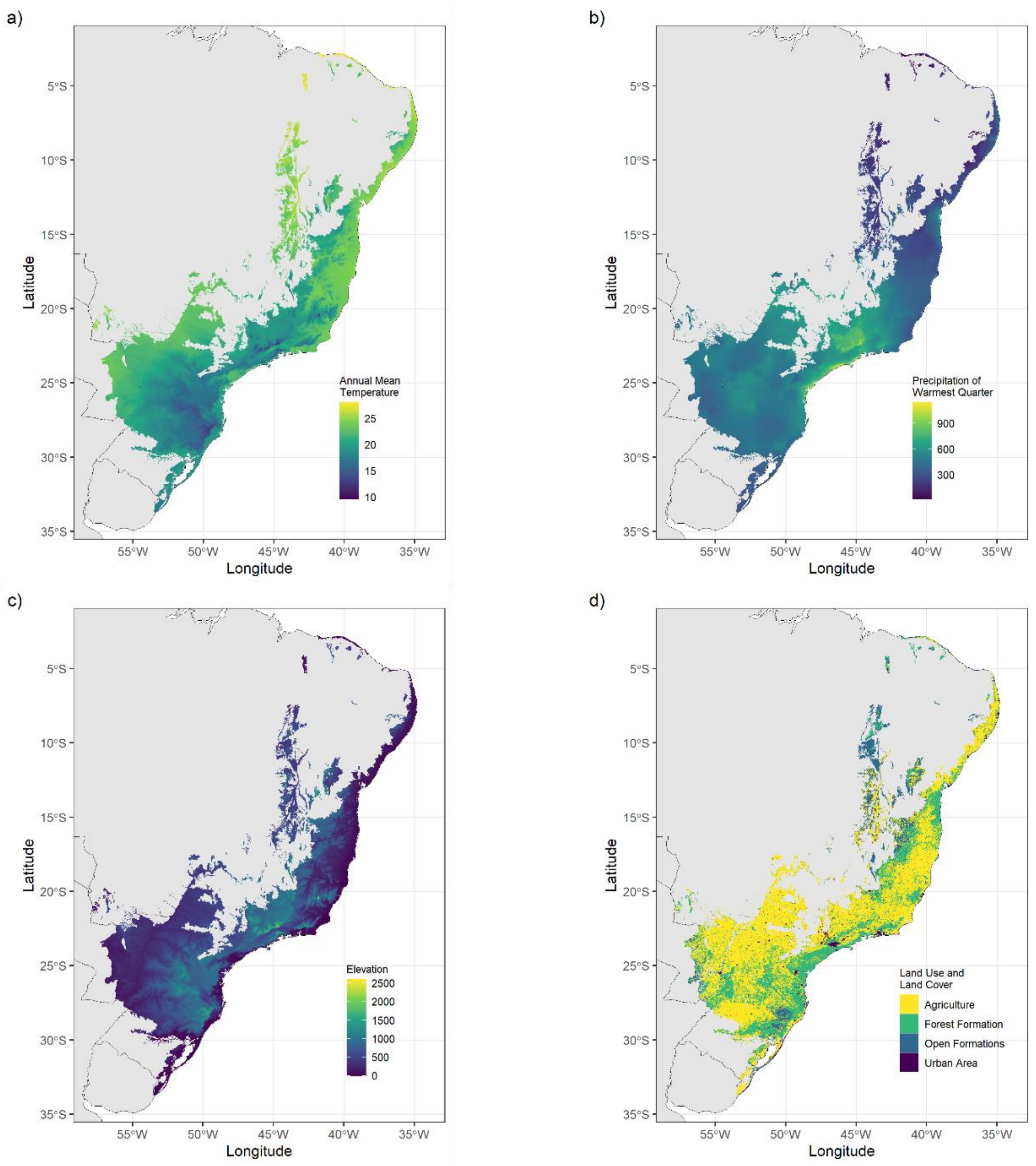

**Figure 11.** Patterns of distribution of macroclimatic variables, topographic variation, and classes of land use and land cover for tri-national Atlantic Forest. a) Annual mean temperature (°C); b) Precipitation of Warmest Quarter (mm); c) Elevation (m); d) Land Use and Land Cover classes. The distribution of variables was generated for a spatial resolution of 1 Km$^2$.

### *Fruit-feeding butterflies community data*

We started by reviewing the dataset from Santos et al. (2018), selecting only studies that sampled fruit-feeding butterflies with bait traps and standardized baits (Freitas et al., 2014) or a combination of active search with entomological nets and bait traps. Because this information was not explicit in the dataset, we reviewed each article cited by Santos et al. looking for the sampling method used and the total sampling effort of the study (trap hours). Studies that did not fit the methodological criteria cited above and those that we could not find online were removed from the dataset. Further, we complemented the dataset by performing a search of published and unpublished articles using the terms “fruit-feeding butterflies”, “bait-traps,” and “Atlantic Forest” on scientific search platforms such as Web of Science and Google Scholar. We added some studies published after 2018 and hence not included in the dataset by Santos et al. (2018), as well as studies carried out in the ecotones region between Atlantic Forest and Caatinga or Cerrado biomes.

We compiled 68 studies and took information from 75 sampling locations (hereafter called communities). However, we only used information about 60 communities since the remaining 15 communities were out of the TAF limit established in this study. For each community, we obtained information about the location as Country, State, Municipality, study area, and geographic coordinates when available. When the geographic coordinates were not provided, we took the coordinates at the centroid of the sampling sites from Google Earth. The species list comprises only species treated as true fruit-feeding butterflies that belong to four subfamilies of Nymphalidae (Charaxinae, Biblidinae (except *Dynamine*), Satyrinae, and the tribes of Nymphalini and Coeini belonging to Nymphalinae) (Freitas et al., 2014). Those subfamilies comprise a guild that feeds exclusively on rotting fruits, plant exudate, and decaying material, such as feces and carrion (Devries, 1988). The raw dataset encompasses a total of 280 species, 14 tribes, and 85 genera belonging to the four subfamilies of fruit-feeding butterflies. The species names were checked for taxonomic classification changes to meet this guild's most up-to-date nomenclature, especially for Satyrinae (Espeland et al., 2019). The metadata is available on GitHub https://github.com/richterbine/MS_Hotspots_inside_the_Hotspot/tree/main/data/raw/bfly.

### *Functional data*

We obtained information on functional traits for 258 of the 280 fruit-feeding butterfly species on TAF. We extract the functional data using digital images of the fruit-feeding butterflies registered in TAF. The images were obtained by taking photos of museum collections – Museu de Zoologia 'Adão José Cardoso', at Universidade Estadual de Campinas – or by searching in digital collections of museums (https://collections-zoology.fieldmuseum.org/list) and accessing images deposited in the Butterflies of America website (http://www.butterfliesofamerica.com/). We extracted nine linear measurements representing the width and length of the wings and body parts from images with a scale bar. We tried to obtain more than one measurement per species and select only males in order not to include variations related to sexual dimorphism. Therefore, trait values represent the mean value of traits for males of the species. In the way to select traits related to flight performance and competitive ability, we choose body size, represented by the mean forewing length (FWL), and the aspect ratio (AR), which is a derived measure, calculated as $4*FWL^2/FA*2$, where FA is the forewing area, calculated as the product of forewing length and forewing width. While body size variation can be associated with diet breadth and abundance of butterflies (Freire et al., 2021), AR represents the wing shape and is related to flight performance, being that species with long and narrow wings (large aspect ratio) are aerodynamically more efficient than short and wide wings (low aspect ratio) (Mena et al., 2020). We used the software TPSutil and TPSdig2 (http://www.sbmorphometrics.org/soft-dataacq.html) to rescale images and perform the linear measurement.

### *Phylogenetic information*

We used the phylogenetic tree of Nymphalidae proposed by (Chazot et al., 2021) to obtain the phylogenetic relationship among all species of fruit-feeding butterflies recorded within the TAF boundaries. Unfortunately, phylogenetic trees do not always carry all the information about the species relationships described for the taxon, so we adopted a species inclusion procedure based on the cladistic hierarchy. Following this procedure, species that occurred at TAF but were not represented in the Nymphalidae backbone tree were grafted as polytomies in the genus node if there were at least one species of the genus in the tree. In case of no representative of the genus, the species were placed nearest to the sister genus in the tree. Finally, we prune the Nymphalidae backbone tree to represent only the phylogenetic relationships of the 258 species of fruit-feeding butterflies, for which we also have functional information. We manipulate and prune the phylogenetic tree using the R environment (R Core Team et al., 2021), employing the packages *ape* v. 5.6-2 (Paradis & Schliep, 2019), *geiger* v 2.0.7 (Pennell et al., 2014), and *phytools* v. 1.0-1 (Revell, 2012).

### *Environmental and spatial data*

We use spatial data from two sources: the Wordclim 2.1 database (https://worldclim.org/data/worldclim21.html) and the MapBiomas Mata Atlantica Trinacional project (https://bosqueatlantico.mapbiomas.org/). From Worldclim, we extract the current climatic data from 19 bioclimatic variables (BIO) and the elevation at a resolution of 30s (each pixel has approximately 900m$^2$). We select two variables from 19 BIOs corresponding to the annual mean temperature (BIO 1) and the precipitation of the warmest quarter (BIO 18). These variables were chosen because they are related to butterfly activity and physiological requirements (Brown & Freitas, 2000). From MapBiomas, we obtained the land use and land cover (LULC) data, considering the first collection of the Trinational covering of the Atlantic Forest (Argentina, Brazil, and Paraguay) by the Google Earth Engine toolkit. The LULC is a raster composed of 12 classes of land use and land cover, mapped in a 30m scale (each pixel has 30m$^2$) considering the most recent year available, i.e., 2019. We started processing the data by resampling the pixel size of the LULC, BIO, and elevation to correspond to approximately 1 km$^2$. We reclassified the entire extension of the Atlantic Forest into six classes from MapBiomas: Forest Formation, Natural Open Formations (composed of Savanna Formation, Grassland, and Rocky outcrop), Agriculture (composed of Forest Plantation, Pasture, Agriculture, Mosaic of Agriculture and Pasture), Urbanization (Non-vegetated area), Non-Forest Natural Formations (composed by Wetland, Other Non-Forest Natural Formation), and Water (River, Lake, and Ocean). We used the new LULC classification to extract the percentage of each of the six classes in a 1km buffer around each community, therefore returning the proportion of coverage of each LULC class for each community. We manipulated the spatial data mainly using the package *raster* v. 3.5-15 (Hijmans, 2022) on the R environment (R Core Team et al., 2021).

### *Data analysis: HMSC Framework*

The Hierarchical Modeling of Species Community framework (HMSC) is a multivariate hierarchical generalized linear mixed model fitted with Bayesian inference and belongs to the JSDM approaches. Before performing analysis in the HMSC framework, we had to prepare the data in matrices. The community data (matrix **Y**) comprised the presence/absence of 258 species of fruit-feeding butterflies in each of the 60 communities distributed along the TAF (red dots, Fig. 1). The matrix **X** described the environmental covariates, that we called climatic variables (annual mean temperature and precipitation of the warmest quarter), a topographic variable (Elevation), and land use and land cover or landscape variables (proportion of natural forest formation, proportion of natural open formations, proportion of urban area, proportion of agricultural area). The matrix **T** described the mean values of the body size, measured as forewing length (FWL) and the aspect ratio (AR) for the 258 species. Also, we provide the phylogenetic relationship for the 258 species to consider the non-independent evolutionary history of species (matrix **C**). All these matrices were related to the fixed component of the HMSC. In the random component, we included an explicit spatial random effect at the sampling location level by providing the spatial coordinates of the communities (xy).

We fitted three concurrent models: mFULL, mENV, and mSPACE, all models include the matrix **T** and **C**, and we used probit regression to model species occurrence probabilities at each site. The first model, mFULL contains the environmental covariates (matrix **X**) as fixed effects and the spatial scale as random effects. The mENV only accounts for environmental covariates (matrix **X**), i.e., no random effects, whereas mSPACE only had random effects. We ran the HMSC framework with default prior definitions (Ovaskainen et al., 2017; Ovaskainen & Abrego, 2020) and standardized the X and T matrices to ensure reasonable performances of the priors. We ran the model for three MCMC chains, a thin rate of 50, and a sample of 100 values per chain. The first 25,000 values were discarded as transient. We used the package Hmsc v.3.0-11 (Tikhonov et al., 2020) in the R environment (R Core Team, 2021).

We evaluated the model's convergence based on the potential scale reduction factor (Gelman & Rubin, 1992) for parameters related to niche and residual species associations. We assessed the best-fitted model employing a Widely Applicable Information Criterion (WAIC) among the three models, which describes the goodness of fit concerning model complexity (Ovaskainen & Abrego, 2020). Using the best model selected (lower WAIC), we evaluated fixed and random effects on species occurrence and assessed the exploratory and predictive power via 4-k-fold cross-validation. We evaluated the niche parameters (β) and the effect of traits and phylogeny in the species niche (γ and ρ, respectively). Further, we used the parameters estimated to predict the community-level response considering each variable for richness, community-weighted mean body size (CWM – FWL), and community-weighted mean aspect ratio (CWM – AR). We measured the strength of the relationship between community structure and environmental covariate through a correlation test and considered a correlation coefficient higher than 0.7 as significant. Finally, we used the posterior predictive distribution of the parameters to predict the species' occupancy probability by cell grid (1 $km^2$) along with the TAF. From the expected occupancy matrix, we derived the mean occupancy probability by subfamily and tribe and the richness by summarizing information within communities. Further, we used the expected occupancy matrix to predict the spatial distribution of body size and AR along TAF. We used the package *ggplot2* v. 3.3.5 (Wickham, 2016) and its extensions to data visualization. All codes to perform the analysis were stored on GitHub (https://github.com/richterbine/MS_Hotspots_inside_the_Hotspot).

## Results

We extract information from 60 sampling sites regarding the presence or absence of 258 species of fruit-feeding butterflies. Most species had a prevalence lower than 0.3

(78.3%), and less than 4% had a prevalence higher than 0.7 (Fig. S1a, Supplementary material). Observed species richness varied from eight to 109 and was not correlated with sampling effort ($r = 0.17$, $t = 1.31$, $p = 0.19$). Most sites had a richness between 20 and 60 species (Fig. S1b, Supplementary Material). We considered that the convergence diagnostic of the MCMC for all models was satisfactory since, for most parameters, the potential scale reduction factor is close to the optimal value of one (Fig. S2 mFULL, Supplementary Material). Based on WAIC, the model mFULL, which included both environmental covariates and random spatial effects in its structure, fitted the data better than the model that only considered environmental covariates ($\mathrm{WAIC}_{\mathrm{mFULL}} = 77.04$, $\mathrm{WAIC}_{\mathrm{mENV}} = 91.94$) but hardly differed from the spatial model (which did not consider environmental covariates) ($\mathrm{WAIC}_{\mathrm{mSPACE}} = 77.57$). The model mFULL had an explanatory power of 0.30 (average Tjur $R^2$) and a predictive power of 0.10 (cross-validation average Tjur $R^2$), and the following analyses were done based on the parameters estimated by this model.

The variance partitioning indicated that the largest proportion of variation in species occurrences was explained by spatial random effects (38.56%) followed by climatic conditions (31.14%) (Fig. S3, Supplementary material). LULC and elevation explained approximately 30% of the variation in species occurrences. The traits explained 56% of the variation in species niche but only 5.28% of the variation in species occurrences probabilities. We found evidence for phylogenetic niche conservatism (PNC) since the estimates for parameter $\rho$ did not overlap zero (median = 0.48, 95% CRI = 0.33 – 0.63). Regarding spatial random effect, the leading factor is non-spatial (median = 1.03, 95% CRI = 0 – 2.75). Thus, the variation was independent among the sampling sites. However, we observed a spatial signal in the second latent factor (median = 6.18, 95% CRI = 4.12 – 9.45) that varies at the scale of ca. 5 km. The residual association between species pairs showed more positive than negative associations (Fig. S4, Supplementary material).

Generally, the species were rare (negative association with the intercept) (Fig. 3a), indicating an occurrence probability of less than 0.5. Regarding the effects of environmental variables, we observed more significant associations between species responses and climatic conditions than with landscape or topographic variables. Among the landscape variables, most of the significant responses were positive, and the proportion of forest formation had more associations with species than other variables (Fig. 3a, heatmap). In addition, we capture a positive association between this niche axis and the aspect ratio (Fig. 3b). This indicates that species with long and narrow forewings tend to occur more in forest habitats responding positively to this landscape feature. Few species responded to natural open formations (Fig. 3a, heatmap). Regarding climatic variables, annual mean temperature presented a positive association with most species of Biblidinae and Charaxinae and negative associations with Brassolini and Morphini, indicating that species of these clades tend to respond similarly (Fig. 3a, heatmap). Precipitation of the warmest quarter positively affects species' response, mainly for Brassolini and Morphini.

Further, we observed that precipitation was positively related to forewing length (Fig. 3b), indicating that larger butterflies occurred more in sites with high precipitation during the warmest months. For elevation, we observed a positive association with the genus *Yphthimoides* (Satyrini), while Brassolini, Morphini, and Charaxinae (genus *Archaeoprepona*) have negative associations. Elevation was negatively related to AR (Fig. 3b), indicating that species that occur at high elevations tend to exhibit low AR, i.e., fast and costly flights.

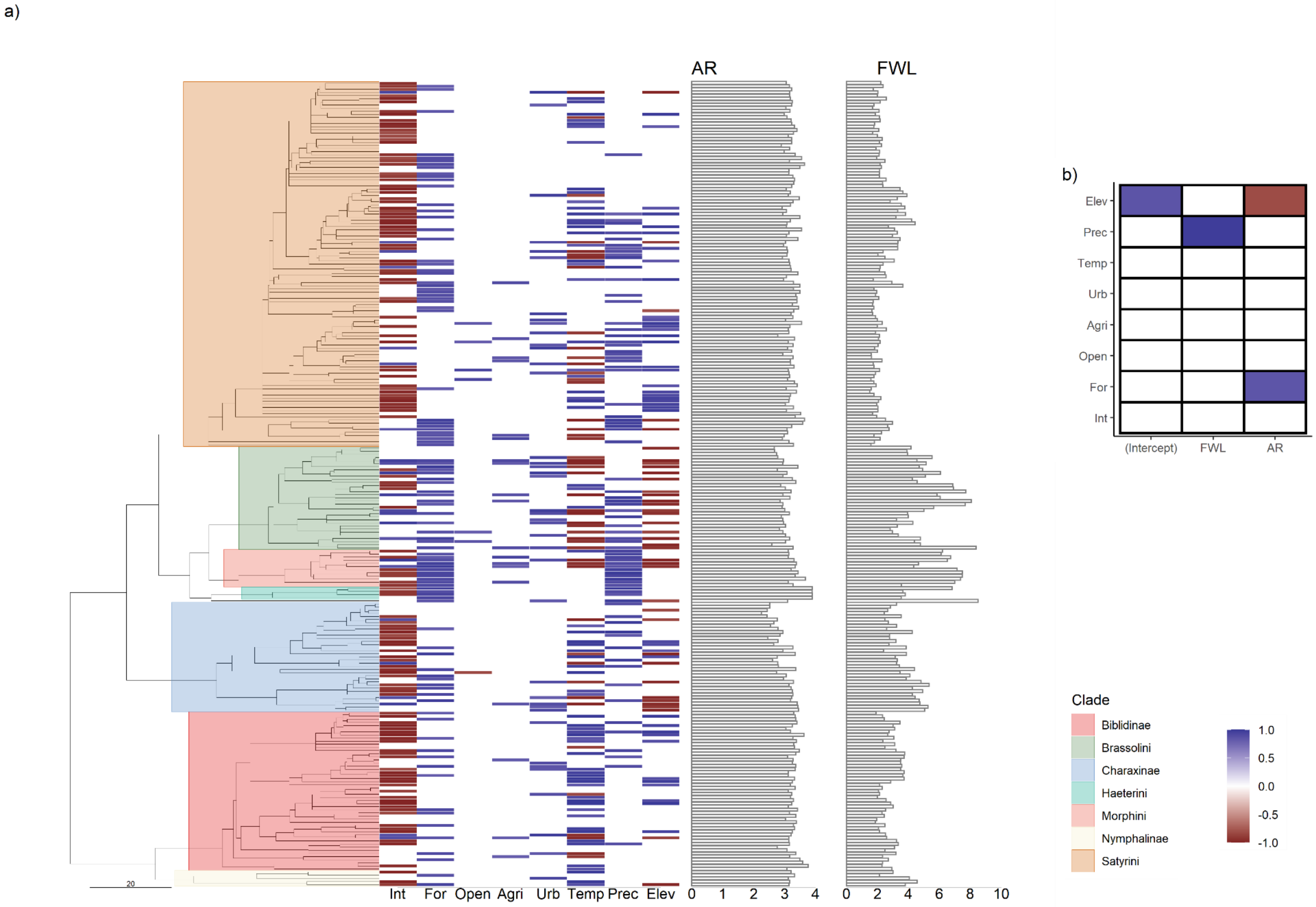
a)
AR
FWL
20
Int
For
Open
Agri
Urb
Temp
Prec
Elev
0
1
2
3
4
0
2
4
6
8
10
b)
Elev
Prec
Temp
Urb
Agri
Open
For
Int
(Intercept)
FWL
AR
Clade
Biblidinae
Brassolini
Charaxinae
Haeterini
Morphini
Nymphalinae
Satyrini
1.0
0.5
0.0
-0.5
-1.0

**Figure 12**. Estimated species niches responses (β parameters) and trait values for each fruit-feeding butterflies species in the TAF. Panel a) exhibits the phylogenetic relationship between species, highlighting the main clade division (Subfamilies and tribes), heatmap for species niche parameters (Int – intercept; For – Natural Forest Formation; Open – natural open formations; Agri – Agriculture; Urb – Urbanization; Temp – Annual mean temperature; Prec – Precipitation of the warmest quarter; Elev – Elevation), showing parameters that are estimated to be positive (blue) and negative (red), respectively, with at least 0.90 posterior probability, and the bar plots indicating the mean value for Aspect ratio (AR) and Forewing length (FWL in centimeters). Panel b) exhibits the heatmap of estimated γ parameters linking species traits to species niches indicating the positive (blue) and negative (red) associations with at least 0.90 posterior probability.

Regarding community-level response, species richness and CWM – AR generally had similar responses to environmental variables (Fig 4, first and third rows). LULC variables tended to increase species richness and increased the frequency of species with high AR (Fig. 4). The annual mean temperature did not affect species richness or CWM – AR, and the precipitation of the warmest quarter did not affect CWM – AR. Elevation had weaker correlations but was also positive for these communities' features (Fig. 4). Regarding the CWM – FWL, we observed that LULC variables have a positive correlation, except the natural open formations, which exhibited a negative correlation with CWM – FWL (Fig. 4, second row). Further, increasing temperature and elevation reduced the CWM – FWL, whereas precipitation increased it (Fig. 4).

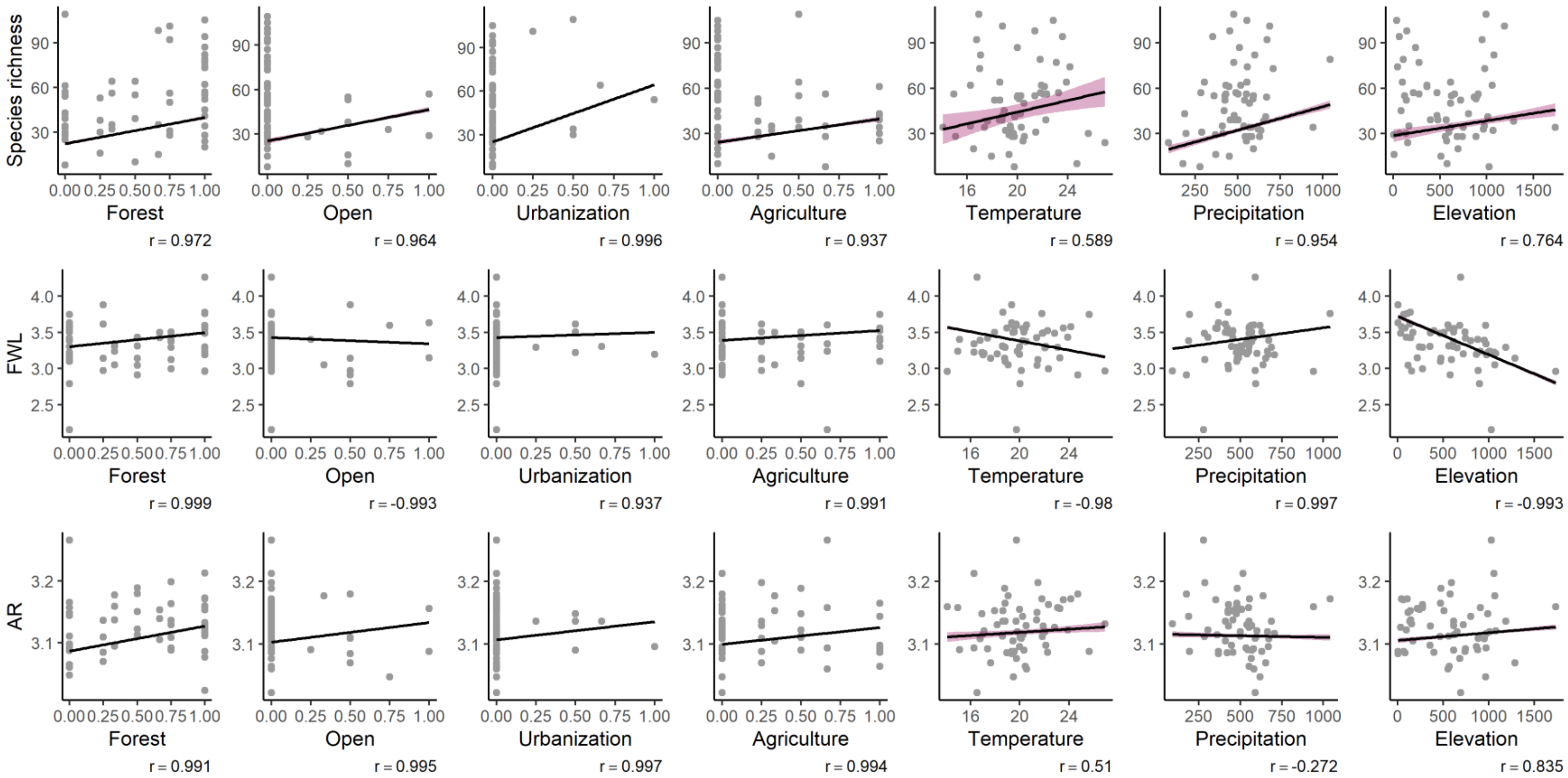

**Figure 13**. Predicted responses to variation in environmental conditions. The first row exhibits the relationship between species richness and focal environmental variables (controlling for the effect of other environmental covariates included in the model). The second and third rows exhibited the relationship between community weighted mean for the body size (FWL) and aspect ratio (AR) to environmental variables, respectively. The values above each panel indicate the correlation coefficient (r) assessed through a correlation test.

Employing the parameters estimated by mFULL model, we computed the posterior mean prediction for the occurrence of all species along all TAF in a grid of 1 km$^2$. The posterior mean richness for all TAF varies between 15 and 89 species (Fig. 5a), with a posterior standard deviation varying between 2 and 34 species, with the highest uncertainties coinciding with areas with high species richness estimates (Fig. S5, Supplementary material). We identified five regions of high species richness in the TAF, the first was related to the South of Bahia, northeastern Brazil, and despite the high estimated richness, it had low uncertainty associated (Fig. 5a; Fig. S5, Supplementary material). The second and third regions are associated with the metropolitan regions of São Paulo and Rio de Janeiro. The fourth region is related to the north of Santa Catarina and southeastern Paraná, and the fifth is associated with the north of the Misiones region and western Paraná. The distribution of the subfamilies/tribes of fruit-feeding butterflies were generally low but continuous along TAF (Fig. S6, Supplementary material). Regarding the functional composition, we observed a distinct pattern in the trait distribution along TAF. While high values of CWM–AR overlapping areas with high expected richness (Fig. 5c), the community-weighted mean for body size (Fig. 5b) indicated that communities near the coastal Atlantic Forest are composed of species with longer wing lengths.

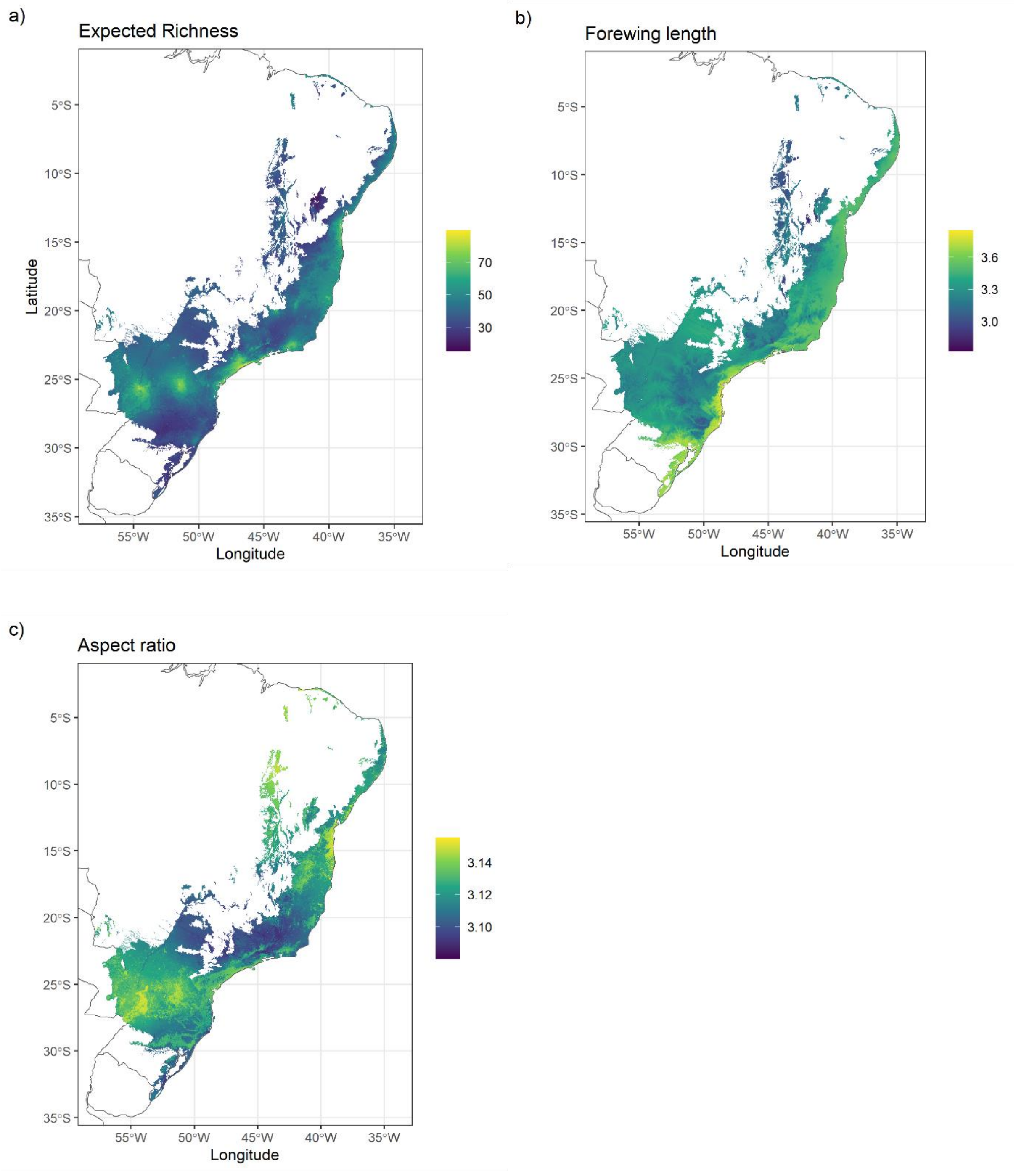

**Figure 14**. Predicted community features across Trinational Atlantic Forest. a) Expected richness (range from 15 to 89 species). b) The community-weighted mean (CWM) value considering forewing length (FWL) and c) for CWM considering the aspect ratio (AR).

## Discussion

Here we accessed the main filters that shaped the diversity of fruit-feeding butterflies along the tri-national Atlantic Forest. Within the environmental variables, climatic variables highly influenced the species niche, and the traits explained more than 50% of the variation in the niche. We found evidence for phylogenetic niche conservatism (PNC), indicating that response traits and phylogenetic relationships are essential for determining species niche. We highlight the importance of the random spatial effect in explaining the variation in species occurrence. Decoupling the role of processes that shape biological communities are crucial to predict how species will respond to climate and land use changes (Poggiato et al., 2021), and including traits and phylogenetic relationships, we assess can make inferences of ecosystem functioning (Ovaskainen et al., 2017; Perović et al., 2015).

At the community level, LULC variables mainly produce positive responses in all aspects of community structure. Further, since richness and CWM – AR hotspots have similar spatial patterns, we can infer that dispersal features can be important to maintain high diversity in TAF. On the other hand, for community body size, we found support for converse Bergmann’s rule since body size decreased with elevation, even though the lower temperatures caused body size to increase.

### *Niche and species occurrence*

In general, the species of fruit-feeding butterflies are rare, i.e., they have low occurrence probabilities along TAF. Joint species distribution models (JSDMs) can be advantageous when modeling community data since they assume that species share information, allowing the estimate of the response of rare species by borrowing information from common species (Ovaskainen & Abrego, 2020; Poggiato et al., 2021). Despite their rarity, these species are especially positively influenced by environmental covariates, whereas more common species can respond positively or even negatively to environmental variables (Fig. S7). Rare species are generally of conservation concern but are often disregarded in modeling frameworks because they lack sufficient data necessary for model fitting (Zipkin et al., 2010). Some rare butterflies occur at the same sites as other threatened taxa (like frogs and mammals), providing a valuable measure of habitat conservation status (Brown & Freitas, 2000; Uehara-Prado et al., 2007).

Considering the high habitat conversion in Atlantic Forest, we expected that landscape features instead macroclimatic conditions would be more critical to define variation in species occurrences. However, we observed that LULC variables are more related to the tribe or subfamily-level responses. This group-level response has already been reported for fruit-feeding butterflies (Brown & Freitas, 2000; Uehara-Prado et al., 2007) and was corroborated in this study as we detected a PNC. PNC may be related to the use of host plants since it is known that there is a relationship between host plant diversity and butterfly diversification (Chazot et al., 2021; Ferrer-Paris et al., 2013). Thus, the occurrence of species is limited by the presence of their host plants or their ability to adapt to new hosts (phylogenetic constraint) (Chazot et al., 2021), causing species to respond similarly to variations in the niche.

From a functional perspective, we observe that butterflies with longer and narrower wings (high AR) occur more in forest habitats but are restricted by elevation. Although flight performance depends on other traits (Betts & Wootton, 1988), it is recognized that high AR promotes power economy, allowing slow and endurance flights (Chai & Srygley, 1990; DeVries et al., 2010; Dudley & Srygley, 1994). Species of Morphini and Haeterini, known to employ this type of flight (DeVries et al., 2010; Stylman et al., 2019), showed a positive response to the forest formation while tending not to respond or decrease their occurrence at higher elevations. In addition, species in these clades showed a positive response to increased precipitation in the warmest quarter, and we observed that this niche axis is associated with body size (FWL). Seasonality is important in determining insect population dynamics (Wolda, 1988), while pluviosity-related variables are recognized as important factors affecting fruit-feeding butterflies' diversity (Checa et al., 2019; Santos et al., 2017). Larger species demand a greater amount of resources to complete their life cycle, and therefore they tend to occur in a narrow period when optimal resources are available (Ribeiro & Freitas, 2011). For Atlantic Forest, we can suggest that alteration in the time window due to climate changes can affect the occurrence of specific traits, such as body size, and may reflect alterations in the ecosystem functions.

### *Community-level responses*

The fruit-feeding butterfly richness is uneven for the Atlantic Forest, and it is not primarily driven by latitudinal gradient but is associated with centers of endemism and in the transition subregion of the Araucaria Forest. Similar results were found by Santos et al. (2020), indicating that the region of Serra do Mar, which concentrates the largest continuous forest remnants and more areas covered by protected areas (Ribeiro et al. 2009), had the potential to support more considerable species richness. However, we highlight two conspicuous richness hotspots in the Araucaria transition region that coincide with areas of forest remnants surrounded by an agricultural matrix. According to the habitat amount hypothesis, regardless of the spatial configuration, what matters is the total amount of habitat in an appropriate spatial extent of the local landscape (Fahrig, 2013; Rybicki et al., 2020). The *EcoLand* (which models the climatic and landscape filters independently) model proposed by Santos et al. (2020) points to these areas as of conservation importance in the face of climate change, indicating that in these locations, the appropriate landscape overrides the effects of climate to maintain species richness. We emphasize the importance of these regions, not only because of their high taxonomic diversity but also because these sites are composed of slow-flying and gliding species (high aspect ratio).

The overlap between areas of high richness and high AR may indicate an effect of dispersal on maintaining the number of species in the local community. For the Brazilian Atlantic Forest, previous studies suggest species richness is higher in forest fragments than in continuous areas (Filgueiras et al., 2016; Uehara-Prado et al., 2007). If some of these forest fragments are high-quality patches capable of maintaining large population sizes, then the richness of lower-quality patches is supported through mass effect (Brito et al., 2021; Leibold et al., 2004). Habitat alteration in these centers can lead to both a decline in species numbers and a change in functional composition, altering population and metacommunity dynamics and ecosystem functioning (Gámez-Virués et al., 2015; Santos et al., 2020). Communities of species that share similar traits are more likely to maintain ecosystem functioning if these species exhibit different responses to environmental filters (Gámez-Virués et al., 2015).

Regarding body size distribution, we observe that communities located at high elevations tend to be composed of species with smaller bodies, with sites near the southern coast showing the highest values. The insect body size and altitude relationship generally follow the converse-Bergmann cline or no clines (Shelomi, 2012). At high elevations, the growing season is short, which reduces the development time for ectotherms to a single season or less (Mousseau, 1997), whereas, at low elevations, the growing season is relatively longer, allowing more resources acquisition during the extended development time, resulting in larger butterflies. However, we observed that within the lowlands (coast), the southern portion of TAF concentrates larger species than the northern portion. This pattern can be due to a phylogenetic artifact (Ribeiro & Freitas, 2011) since the species with the largest bodies belong to Brassolini and Morphini (tribes of Satyrinae), and their distribution match areas of highly expected mean body size (FWL). On the other hand, the body size is related to many other life-history traits, such as diet breadth (Freire et al., 2021; Marini-Filho & Martins, 2010). The smaller species generally have a restricted feeding niche, being more specialized than the larger species (Freire et al., 2021; Perović et al., 2015). We found that increasing open natural formations increases the frequency of smaller species, i.e., acting as a filter preventing the persistence of generalist species. In this way, our result highlighted the importance of protecting the non-forested ecosystems, which is highly neglected in Brazilian conservation policies (Overbeck et al., 2015), to conserve both species and functions associated with these systems.

## Conclusion

In the face of climate change and intensifying land use, understanding the role of environmental and biotic processes in assembling communities is a central issue in ecological research. The use of JSDM can improve our capacity to estimate, predict and evaluate species' response to environmental gradients while considering the non-independence in species' responses (Ovaskainen & Abrego, 2020; Tobler et al., 2019). Our results bring new insights into processes that shape the diversity of fruit-feeding butterflies in the Atlantic Forest, showing that both traits and evolutionary history are essential to define species' responses to environmental filters. By including attributes and phylogeny in species modeling, we can test questions related to niche adaptations (Ovaskainen et al., 2017), but we can also assess how these characteristics are distributed in space. Identifying how macroecological patterns are linked to environmental predictors and other diversity patterns improves our understanding of the assembly processes that make the Atlantic Forest a diversity hotspot.

## References

Bagchi, R., Brown, L. M., Elphick, C. S., Wagner, D. L., & Singer, M. S. (2018). Anthropogenic fragmentation of landscapes: mechanisms for eroding the specificity of plant–herbivore interactions. *Oecologia*, *187*(2), 521–533. https://doi.org/10.1007/s00442-018-4115-5

Betts, C. R., & Wootton, R. J. (1988). Wing Shape and Flight Behaviour in Butterflies (Lepidoptera: Papilionoidea and Hesperioidea): A Preliminary Analysis. *Journal of Experimental Biology, 138*(1), 271–288. http://jeb.biologists.org/content/138/1/271.abstract

Börschig, C., Klein, A.-M., von Wehrden, H., & Krauss, J. (2013). Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. *Basic and Applied Ecology*, *14*(7), 547–554. https://doi.org/10.1016/j.baae.2013.09.002

Brito, M. R. M., Lion, M. B., Oliveira, I. F., & Cardoso, M. Z. (2021). Butterflies on the dry edge of the Atlantic Forest: water availability determines community structure at the Northern limit of Atlantic Forest. *Insect Conservation and Diversity*, *14*(4), 476–491. https://doi.org/10.1111/icad.12474

Brown, K. S., & Freitas, A. V. L. (2000). Atlantic Forest Butterflies: Indicators for Landscape Conservation. *Biotropica*, *32*(4b), 934–956. https://doi.org/10.1111/j.1744-7429.2000.tb00631.x

Chai, P., & Srygley, R. B. (1990). Predation and the Flight, Morphology, and Temperature of Neotropical Rain-Forest Butterflies. *The American Naturalist*, *135*(6), 748–765.

Chazot, N., Condamine, F. L., Dudas, G., Peña, C., Kodandaramaiah, U., Matos-Maraví, P., Aduse-Poku, K., Elias, M., Warren, A. D., Lohman, D. J., Penz, C. M., DeVries, P., Fric, Z. F., Nylin, S., Müller, C., Kawahara, A. Y., Silva-Brandão, K. L., Lamas, G., Kleckova, I., … Wahlberg, N. (2021). Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-25906-8

Checa, M. F., Levy, E., Rodriguez, J., & Willmott, K. (2019). Rainfall as a significant contributing factor to butterfly seasonality along a climatic gradient in the Neotropics. *BioRxiv*, *May*. https://doi.org/10.1101/630947

Denno, R. F., Roderick, G. K., Peterson, M. A., Huberty, A. F., Döbel, H. G., Eubanks, M. D., Losey, J. E., & Langelotto, G. A. (1996). Habitat persistance underlies intraspecific variation in the dispersal strategies of planthoppers. *Ecological Monographs*, *66*(4), 389–408. https://doi.org/10.2307/2963487

Devries, P. J. (1988). Stratification of fruit-feeding nymphalid butterflies in a Costa Rican rainforest. *Journal of Research on the Lepidoptera*, *26*(1–4), 98–108.

DeVries, P. J., Penz, C. M., & Hill, R. I. (2010). Vertical distribution, flight behaviour and evolution of wing morphology in Morpho butterflies. *Journal of Animal Ecology*, *79*(5), 1077–1085. https://doi.org/10.1111/j.1365-2656.2010.01710.x

Duarte, L. D. S., Debastiani, V. J., Carlucci, M. B., & Diniz-Filho, J. A. F. (2018). Analyzing community-weighted trait means across environmental gradients: should phylogeny stay or should it go? *Ecology*, *99*(2), 385–398. https://doi.org/10.1002/ecy.2081

Dudley, R., & Srygley, R. (1994). Flight Physiology of Neotropical Butterflies: Allometry of Airspeeds During Natural Free Flight. *The Journal of Experimental Biology*, *191*(1), 125–139. http://www.ncbi.nlm.nih.gov/pubmed/9317473

Ellison, A. M. (1996). An Introduction to Bayesian Inference for Ecological Research and Environmental Decision-Making. *Ecological Applications*, *6*(4), 1036–1046. https://doi.org/10.2307/2269588

Espeland, M., Breinholt, J. W., Barbosa, E. P., Casagrande, M. M., Huertas, B., Lamas, G., Marín, M. A., Mielke, O. H. H., Miller, J. Y., Nakahara, S., Tan, D., Warren, A. D., Zacca, T., Kawahara, A. Y., Freitas, A. V. L., & Willmott, K. R. (2019). Four hundred shades of brown: Higher level phylogeny of the problematic Euptychiina (Lepidoptera, Nymphalidae, Satyrinae) based on hybrid enrichment data. *Molecular Phylogenetics and Evolution*, *131*(August 2018), 116–124. https://doi.org/10.1016/j.ympev.2018.10.039

Fahrig, L. (2013). Rethinking patch size and isolation effects: The habitat amount hypothesis. *Journal of Biogeography*, *40*(9), 1649–1663. https://doi.org/10.1111/jbi.12130

Ferrer-Paris, J. R., Sánchez-Mercado, A., Viloria, Á. L., & Donaldson, J. (2013). Congruence and Diversity of Butterfly-Host Plant Associations at Higher Taxonomic Levels. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063570

Filgueiras, B. K. C., Melo, D. H. A., Leal, I. R., Tabarelli, M., Freitas, A. V. L., & Iannuzzi, L. (2016). Fruit-feeding butterflies in edge-dominated habitats: community structure, species persistence and cascade effect. *Journal of Insect Conservation*, *20*(3), 539–548. https://doi.org/10.1007/s10841-016-9888-4

Freire, G. B., Silva, T., Oliveira, H., Collier, C., Rodrigues, H. P., Dias, J. P., Santos, J. P., Marini-Filho, O. J., Freitas, A. V. L., Smilanich, A. M., Dyer, L. A., & Diniz, I. R. (2021). Good Things Come in Larger Packages: Size Matters for Adult Fruit-Feeding Butterfly Dispersal and Larval Diet Breadth. *Diversity*, *13*(12), 664. https://doi.org/10.3390/d13120664

Freitas, A. V. L., Iserhard, C. A., Santos, J. P. dos, Carreira, J. Y. O., Ribeiro, D. B., Melo, D. H. A., Rosa, A. H., Marini-Filho, O. J., Mattos Accacio, G., & Uehara-Prado, M. (2014). Studies with butterfly bait traps: An overview. *Revista Colombiana de Entomología*, *40*(2), 203–212. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-04882014000200013&lng=en&nrm=iso&tlng=en

Gámez-Virués, S., Perović, D. J., Gossner, M. M., Börschig, C., Blüthgen, N., de Jong, H., Simons, N. K., Klein, A. M., Krauss, J., Maier, G., Scherber, C., Steckel, J., Rothenwöhrer, C., Steffan-Dewenter, I., Weiner, C. N., Weisser, W., Werner, M., Tscharntke, T., & Westphal, C. (2015). Landscape simplification filters species traits and drives biotic homogenization. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms9568

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Grime, J. P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology*, *86*(6), 902–910. https://doi.org/https://doi.org/10.1046/j.1365-2745.1998.00306.x

Hijmans, R. J. (2022). *raster: Geographic Data Analysis and Modeling* (3.5-15). https://cran.r-project.org/package=raster

Keddy, P. A. (1992). Assembly and response rules: two goals for predictive community ecology. In *Journal of Vegetation Science* (Vol. 3).

Lehnert, M. S., Mark Scriber, J., Gerard, P. D., & Emmel, T. C. (2012). The “Converse to bergmann’s rule” in Tiger Swallowtail Butterflies: Boundaries of species and subspecies wing traits are independent of thermal and host-plant induction. *American Entomologist*, *58*(3), 156–165. https://doi.org/10.1093/ae/58.3.156

Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*(7), 601–613. https://doi.org/10.1111/j.1461-0248.2004.00608.x

Marini-Filho, O. J., & Martins, R. P. (2010). Nymphalid butterfly dispersal among forest fragments at Serra da Canastra National Park, Brazil. *Journal of Insect Conservation*, *14*(4), 401–411. https://doi.org/10.1007/s10841-010-9271-9

Marques, M. C. M., & Grelle, C. E. v. (2021). The Atlantic Forest: History, Biodiversity, Threats and Opportunities of the Mega-diverse Forest. In M. C. M. Marques & C. E. v. Grelle (Eds.), *The Atlantic Forest*. Springer International Publishing. https://doi.org/10.1007/978-3-030-55322-7

McGill, B. J., Enquist, B. J., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology and Evolution*, *21*(4), 178–185. https://doi.org/10.1016/j.tree.2006.02.002

Mena, S., Kozak, K. M., Cárdenas, R. E., & Checa, M. F. (2020). Forest stratification shapes allometry and flight morphology of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1937), 20201071. https://doi.org/10.1098/rspb.2020.1071

Menéndez, R., González-Megías, A., Collingham, Y., Fox, R., Roy, D. B., Ohlemüller, R., & Thomas, C. D. (2007). DIRECT AND INDIRECT EFFECTS OF CLIMATE AND HABITAT FACTORS ON BUTTERFLY DIVERSITY. *Ecology*, *88*(3), 605–611. https://doi.org/10.1890/06-0539

Mousseau, T. A. . (1997). Ectotherms Follow the Converse to Bergmann ' s Rule. *Evolution*, *51*(2), 630–632. https://doi.org/10.2307/2411138

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Ovaskainen, O., & Abrego, N. (2020). Joint Distribution Modelling (with applications in R). In *Joint Species Distribution Modelling*.

Ovaskainen, O., Rybicki, J., & Abrego, N. (2019). What can observational data reveal about metacommunity processes? *Ecography*, *42*(11), 1877–1886. https://doi.org/10.1111/ecog.04444

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576. https://doi.org/10.1111/ele.12757

Overbeck, G. E., Vélez-Martin, E., Scarano, F. R., Lewinsohn, T. M., Fonseca, C. R., Meyer, S. T., Müller, S. C., Ceotto, P., Dadalt, L., Durigan, G., Ganade, G., Gossner, M. M., Guadagnin, D. L., Lorenzen, K., Jacobi, C. M., Weisser, W. W., & Pillar, V. D. (2015). Conservation in Brazil needs to include non-forest ecosystems. *Diversity and Distributions*, *21*(12), 1455–1460. https://doi.org/10.1111/ddi.12380

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pavoine, S., & Bonsall, M. B. (2011). Measuring biodiversity to explain community assembly: A unified approach. *Biological Reviews*, *86*(4), 792–812. https://doi.org/10.1111/j.1469-185X.2010.00171.x

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., & Harmon, L. J. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, *30*(15), 2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Perović, D., Gámez-Virués, S., Börschig, C., Klein, A.-M., Krauss, J., Steckel, J., Rothenwöhrer, C., Erasmi, S., Tscharntke, T., & Westphal, C. (2015). Configurational landscape heterogeneity shapes functional community composition of grassland butterflies. *Journal of Applied Ecology*, *52*(2), 505–513. https://doi.org/10.1111/1365-2664.12394

Pilowsky, J. A., Colwell, R. K., Rahbek, C., & Fordham, D. A. (2022). Process-explicit models reveal the structure and dynamics of biodiversity patterns. *Science Advances*, *8*(31), 2271. https://doi.org/10.1126/sciadv.abj2271

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the Interpretations of Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, *36*(5), 391–401. https://doi.org/10.1016/j.tree.2021.01.002

Projeto MapBiomas Mata Atlântica Trinacional. (n.d.). *Coleção v1.0 da Série Anual de Mapas de Cobertura e Uso de Solo*. Fundação Brasileira Para o Desenvolvimento Sustentável. Retrieved September 14, 2022, from https://bosqueatlantico.mapbiomas.org/

R Core Team, Hsieh, T. C., Ma, K. H., & Chao, A. (2021). *R: A Language and Environment for Statistical Computing*. http://chao.stat.nthu.edu.tw/wordpress/software_download/

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Ribeiro, D. B., & Freitas, A. V. L. (2011). Large-sized insects show stronger seasonality than small-sized ones: A case study of fruit-feeding butterflies. *Biological Journal of the Linnean Society*, *104*(4), 820–827. https://doi.org/10.1111/j.1095-8312.2011.01771.x

Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J., & Hirota, M. M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, *142*(6), 1141–1153. https://doi.org/10.1016/j.biocon.2009.02.021

Ricotta, C., & Moretti, M. (2011). CWM and Rao's quadratic diversity: A unified framework for functional ecology. *Oecologia*, *167*(1), 181–188. https://doi.org/10.1007/s00442-011-1965-5

Rybicki, J., Abrego, N., & Ovaskainen, O. (2020). Habitat fragmentation and species diversity in competitive communities. *Ecology Letters*, *23*(3), 506–517. https://doi.org/10.1111/ele.13450

Santos, J. P. dos, Freitas, A. V. L., Brown, K. S., Carreira, J. Y. O., Gueratto, P. E., Rosa, A. H. B., Lourenço, G. M., Accacio, G. M., Uehara-Prado, M., Iserhard, C. A., Richter, A., Gawlinski, K., Romanowski, H. P., Mega, N. O., Teixeira, M. O., Moser, A., Ribeiro, D. B., Araujo, P. F., Filgueiras, B. K. C., … Ribeiro, M. C. (2018). Atlantic butterflies: a data set of fruit-feeding butterfly communities from the Atlantic forests. *Ecology*, *99*(12), 2875–2875. https://doi.org/10.1002/ecy.2507

Santos, J. P. dos, Iserhard, C. A., Carreira, J. Y. O., & Freitas, A. V. L. (2017). Monitoring fruit-feeding butterfly assemblages in two vertical strata in seasonal Atlantic Forest: Temporal species turnover is lower in the canopy. *Journal of Tropical Ecology*, *33*(5), 345–355. https://doi.org/10.1017/S0266467417000323

Santos, J. P., Sobral-Souza, T., Brown, K. S., Vancine, M. H., Ribeiro, M. C., & Freitas, A. V. L. (2020). Effects of landscape modification on species richness patterns of fruit-feeding butterflies in Brazilian Atlantic Forest. *Diversity and Distributions*, *26*(2), 196–208. https://doi.org/10.1111/ddi.13007

Shahabuddin, G., & Terborgh, J. W. (1999). Frugivorous butterflies in Venezuelan forest fragments : abundance , diversity and the effects of isolation. *Journal of Tropical Ecology*, *15*(September 2000), 703–722.

Shelomi, M. (2012). Where are we now? Bergmann's rule Sensu Lato in insects. In *American Naturalist* (Vol. 180, Issue 4, pp. 511–519). https://doi.org/10.1086/667595

Silva, J. M. C. da, & Casteleti, C. H. M. (2003). Status of the biodiversity of the Atlantic Forest of Brazil. In *The Atlantic Forest of South America: biodiversity status, threats, and outlook* (pp. 43–59).

Sobral-Souza, T., Santos, J. P., Maldaner, M. E., Lima-Ribeiro, M. S., & Ribeiro, M. C. (2021). EcoLand: A multiscale niche modelling framework to improve predictions on biodiversity and conservation. *Perspectives in Ecology and Conservation*, *19*(3), 362–368. https://doi.org/10.1016/j.pecon.2021.03.008

Stylman, M., Penz, C. M., & Devries, P. (2019). Large Hind Wings Enhance Gliding Performance in Ground Effect in a Neotropical Butterfly (Lepidoptera: Nymphalidae). *Annals of the Entomological Society of America*, *X*, 1–8. https://doi.org/10.1093/aesa/saz045

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution*, *11*(3), 442–447. https://doi.org/10.1111/2041-210X.13345

Tobler, M. W., Kéry, M., Hui, F. K. C., Guillera-Arroita, G., Knaus, P., & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, *100*(8), 1–14. https://doi.org/10.1002/ecy.2754

Uehara-Prado, M., Brown, K. S., & Freitas, A. V. L. (2007). Species richness, composition and abundance of fruit-feeding butterflies in the Brazilian Atlantic Forest: Comparison between a fragmented and a continuous landscape. *Global Ecology and Biogeography*, *16*(1), 43–54. https://doi.org/10.1111/j.1466-8238.2006.00267.x

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wolda, H. (1988). INSECT SEASONALITY: WHy? *Ann. Rev. Ecol. Sysi*, *19*, 1–18. http://www.annualreviews.org/doi/pdf/10.1146/annurev.es.19.110188.000245

Wright, D. H. (1983). Species-Energy Theory: An Extension of Species-Area Theory. *Oikos*, *41*(3), 496–506. http://www.jstor.org/stable/3544109

Xing, S., Cheng, W., Nakamura, A., Tang, C. C., Pickett, E. J., Huang, S., Odell, E., Goodale, E., Goodale, U. M., & Bonebrake, T. C. (2018). Elevational clines in morphological traits of subtropical and tropical butterfly assemblages. In *Biological Journal of the Linnean Society* (Vol. 123). http://www.ncbi.nlm.nih.gov/

Zipkin, E. F., Andrew Royle, J., Dawson, D. K., & Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation, 143*(2), 479–484. https://doi.org/10.1016/j.biocon.2009.11.016

## Supplementary material

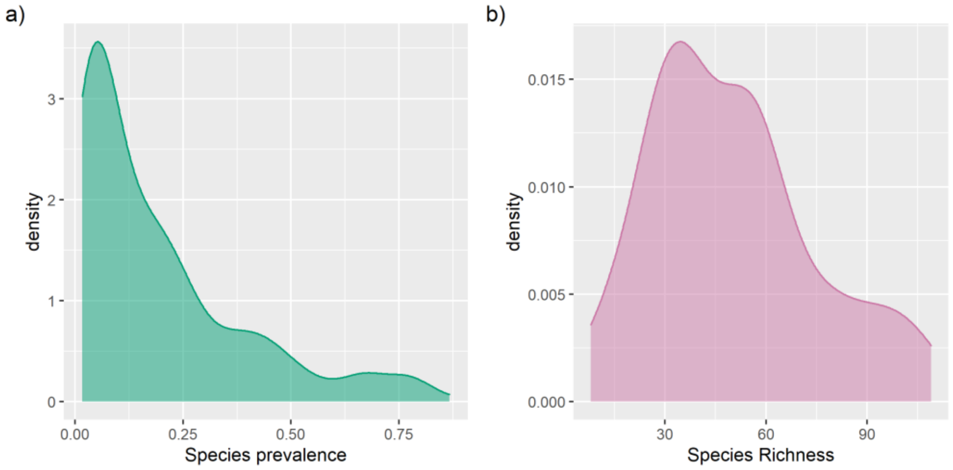

**Figure S6**. Basic summaries of the fruit-feeding butterflies recorded along the trinational Atlantic Forest. a) Species prevalence measured as the mean of occurrence of species over sampling units. b) Species richness measures as the sum of occurrence by sampling unit. y-axis (density) indicates number of species (a) and sampling units (b).

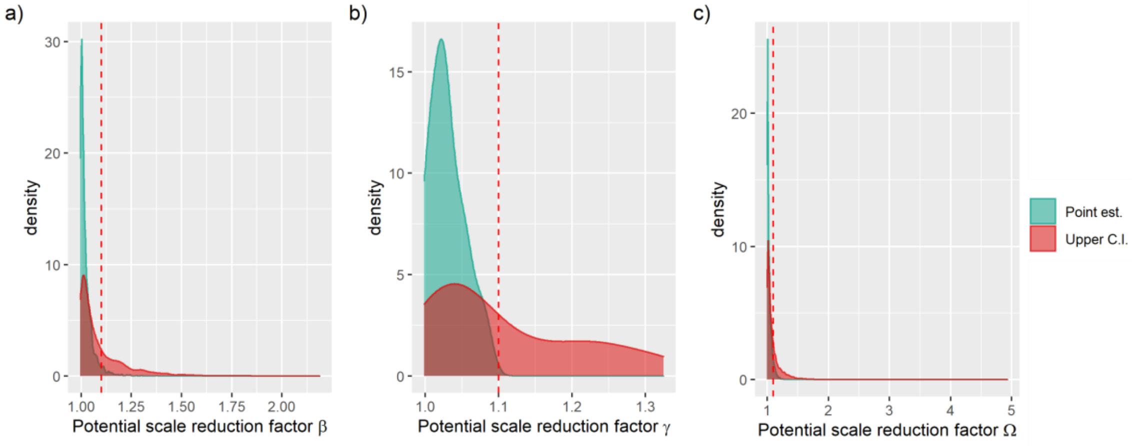

**Figure S7**. MCMC convergence diagnostics for Model FULL, which contains environmental covariates and spatial random effects. The panels show the distribution of the potential scale reduction factor for the β (a), gamma (b), and Ω (c) parameters. The red dashed line indicates the optimal value of 1.1.

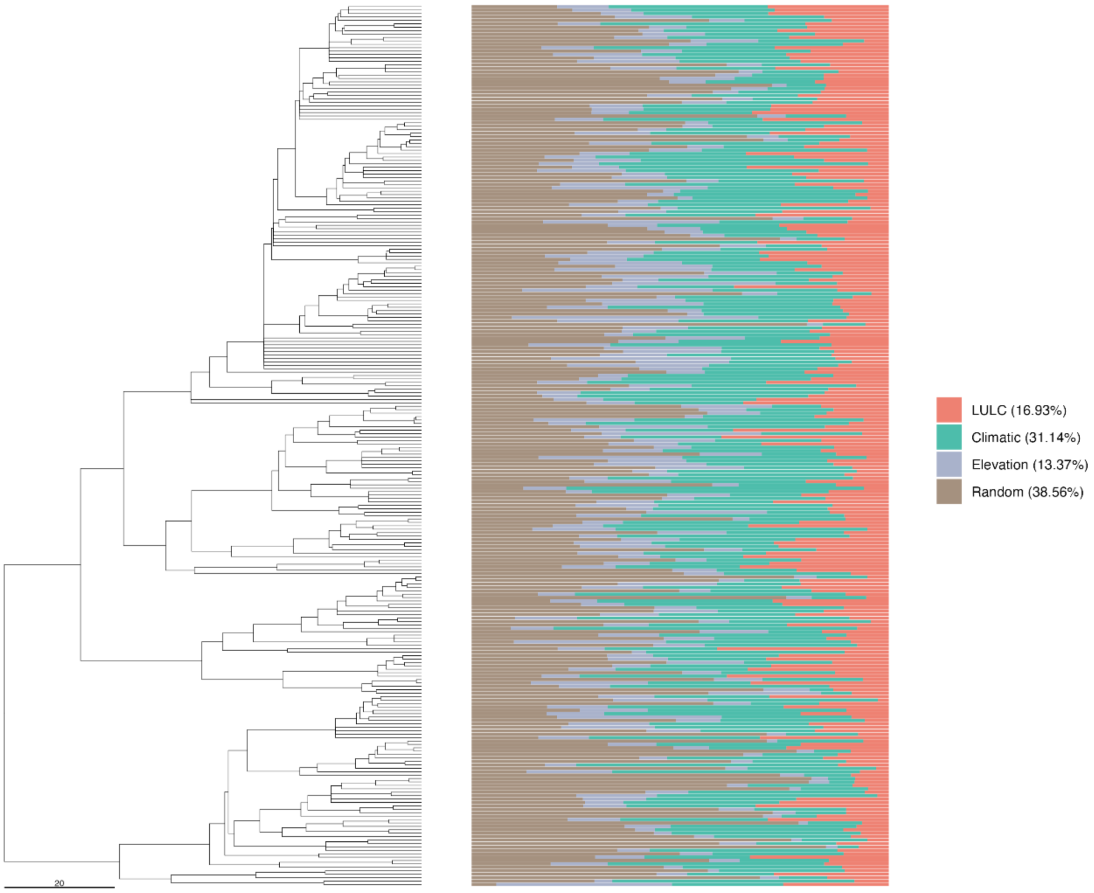

**Figure S8**. Variance partitioning among the fixed and random effects included in the mFull.

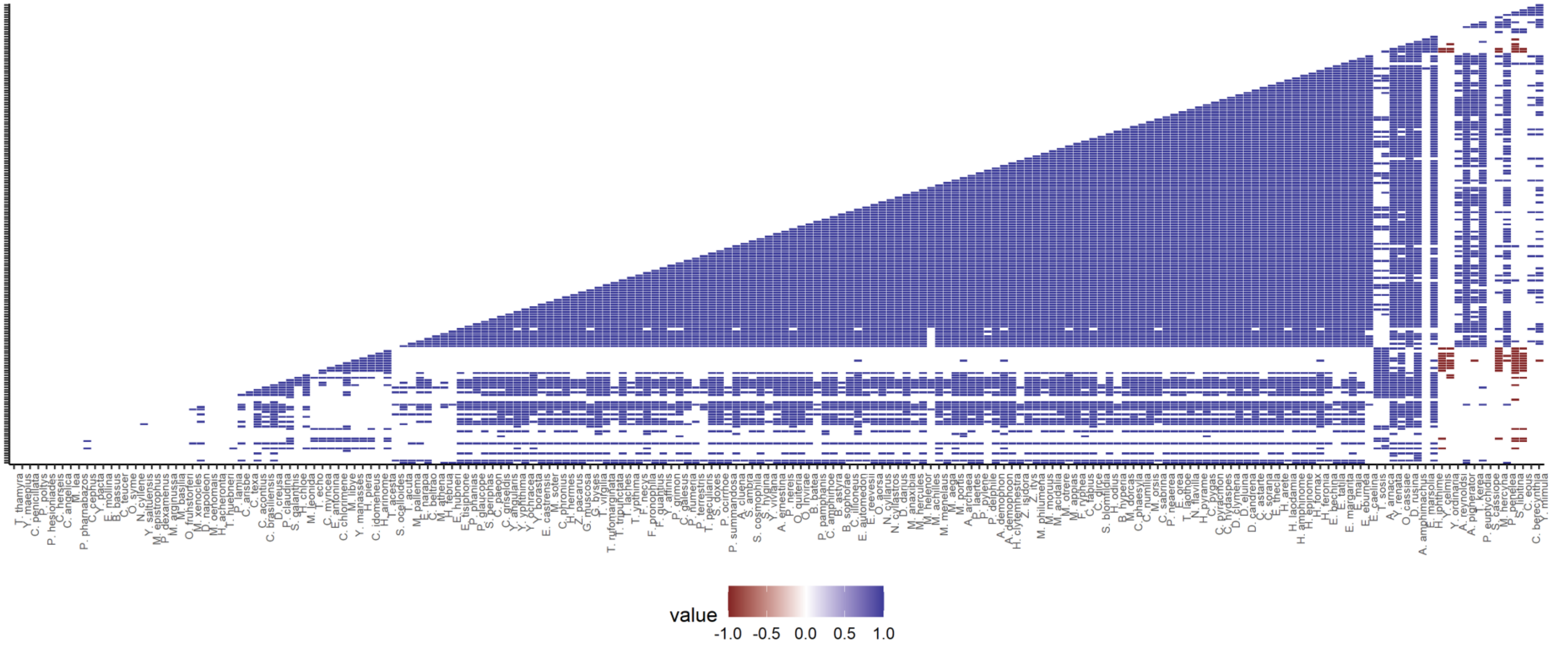

**Figure S9**. Residual species association. We removed species without any significant association to improve data visualization. The species was grouped using a hierarchical clustering method (“complete”) based on similarities.

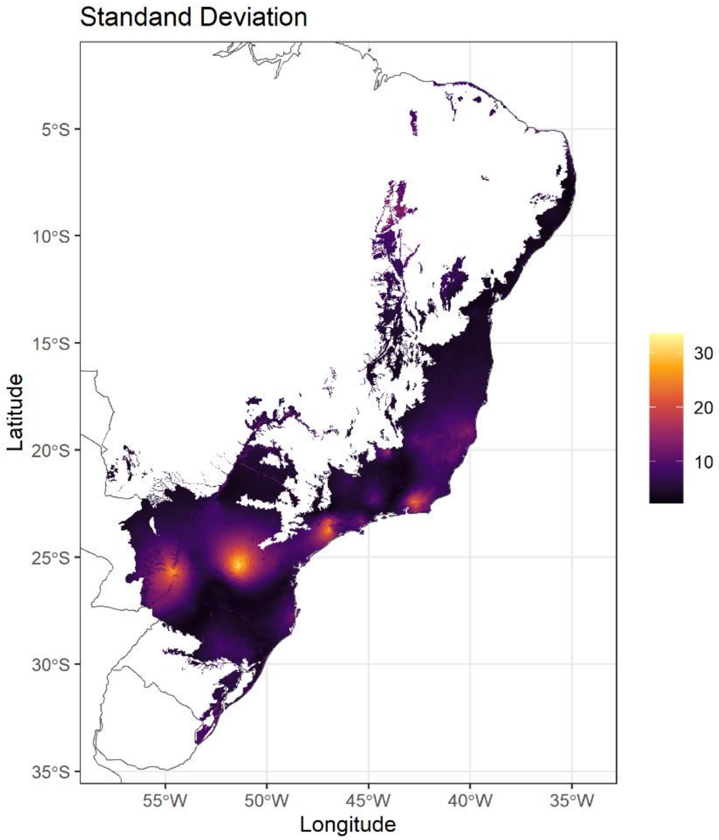

**Figure S10**. Illustration of uncertainty in model predictions. The panel shows the posterior standard deviation of species richness based on parameters estimated by mFull.

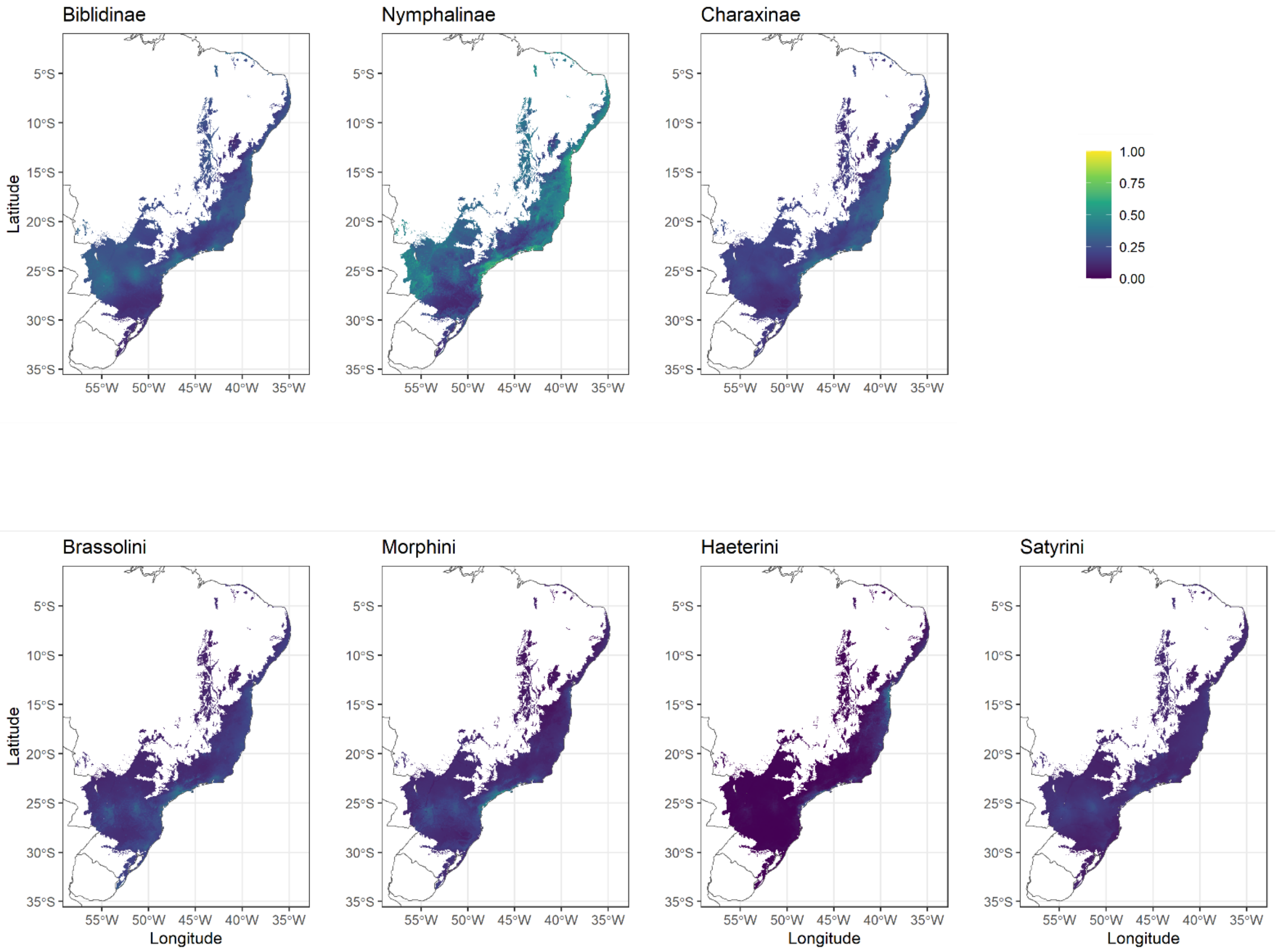
Biblidinae
Nymphalinae
Charaxinae
Brassolini
Morphini
Haeterini
Satyrini
Latitude
Longitude
5°S
10°S
15°S
20°S
25°S
30°S
35°S
55°W
50°W
45°W
40°W
35°W
1.00
0.75
0.50
0.25
0.00

**Figure S11**. The occurrence probability distribution at the subfamily and tribe level for fruit-feeding butterflies along the tri-national Atlantic Forest. Within Satyrinae, Haeterini had a high probability of occurrence on the Bahian coast, Morphini + Brassolini had hotspots in the Serra do Mar and along the southern coast. Biblidinae was more associated with transitional zones with Cerrado/Caatinga, and Nymphalinae, despite the lower species diversity, was highly occurrent along TAF, mainly at Bahia and Serra do Mar coast. The occurrence of Charaxinae tended to decrease with latitude and longitude and was more constrained to the coastal portion of TAF.

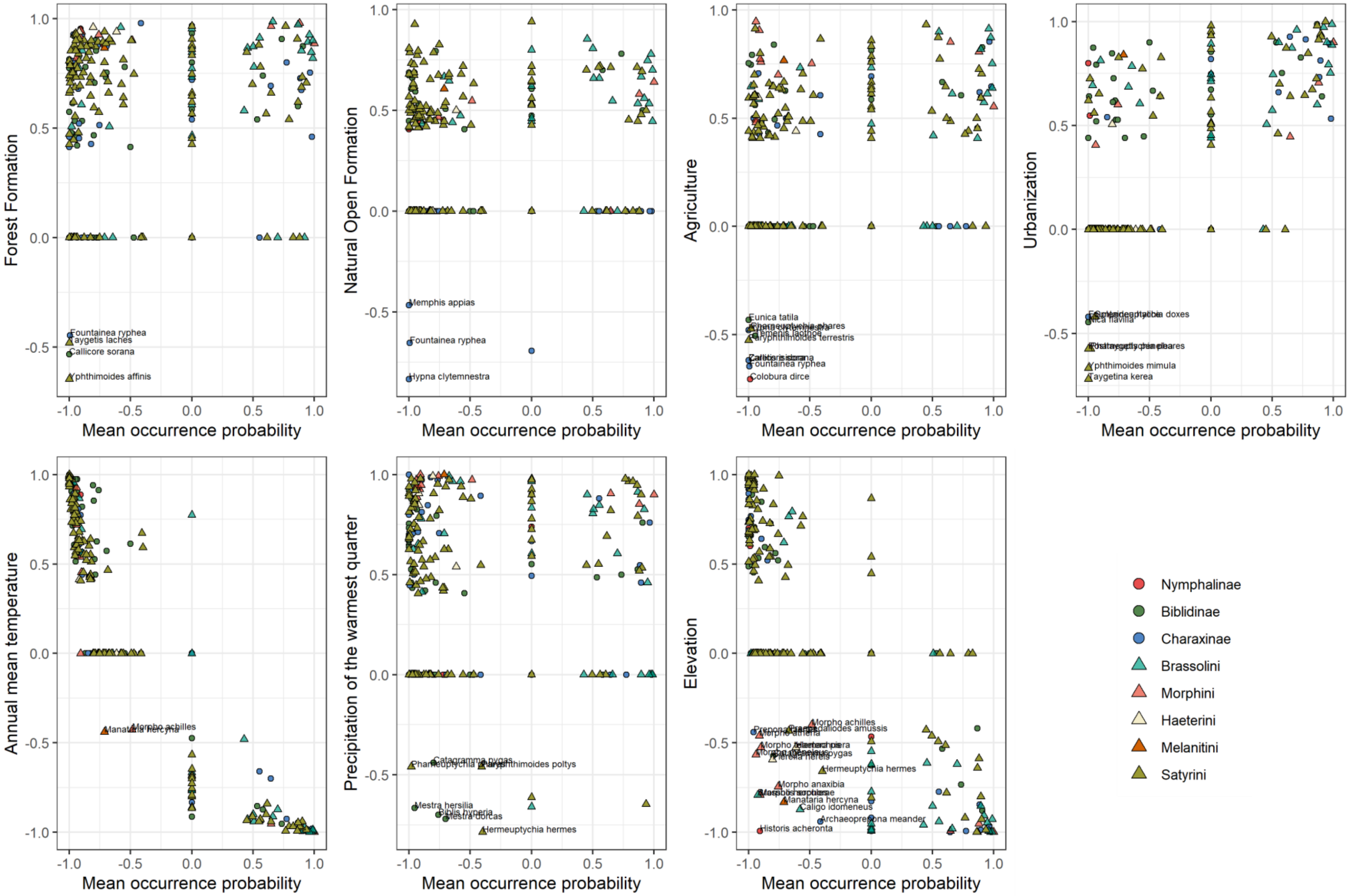
Forest Formation
Fountainea ryphea
Taygetis laches
Callicore sorana
Yphthimoides affinis
Natural Open Formation
Memphis appias
Fountainea ryphea
Hypna clytemnestra
Agriculture
Eunica tatila
Fountainea ryphea
Colobura dirce
Urbanization
Yphthimoides mimula
Taygetina kerea
Annual mean temperature
Morpho achilles
Manataria hercyna
Precipitation of the warmest quarter
Catagramma pygas
Yphthimoides poltys
Mestra hersilia
Biblis hyperia
Mestra dorcas
Hermeuptychia hermes
Elevation
Morpho achilles
Opsiphanes amussis
Morpho athena
Hermeuptychia hermes
Morpho anaxibia
Manataria hercyna
Caligo idomeneus
Archaeoprepona meander
Historis acheronta
Mean occurrence probability
1.0
0.5
0.0
-0.5
-1.0
Nymphalinae
Biblidinae
Charaxinae
Brassolini
Morphini
Haeterini
Melanitini
Satyrini

**Figure S12**. Illustration of variation in species niches. In panels, the *x*-axis represents the mean occurrence probability (intercept), while the *y*-axis represents the β for each environmental covariate. Negative and positive values in *x*-axis indicate, respectively, rare and common species. Values in the *y*-axis indicate the species' positive or negative response to the focal environmental covariates. Symbols represent the subfamily (circles) and tribe (triangles) to which the species belongs.

# Considerações finais

Concluí através desta tese que os fatores climáticos são os principais responsáveis por estruturar as comunidades de borboletas frugívoras, desde a escala local (microclima) até a escala biogeográfica (macro clima). Além disso, os padrões de diversidade também foram relacionados com variações na paisagem, ressaltando a importância da inclusão desse tipo de informações na modelagem da diversidade. Enquanto o aumento da proporção florestal leva a um aumento da riqueza de espécies na Mata Atlântica, a manutenção do mosaico campo-floresta aumenta a diversidade no Pampa. Esses resultados corroboram a ideia de que as respostas observadas são dependentes de escalas, mas são também dependentes do contexto regional de cada bioma, os quais são temporalmente dependentes. Vale também destacar a importância das características funcionais, principalmente aquelas ligadas a dispersão, bem como as relações filogenéticas sobre o nicho das espécies. Entender essa relação pode auxiliar a entender o funcionamento ecossistêmico, visto que mudanças na composição podem levar a homogeneização funcional ou filogenética. Espécies similares funcionalmente tem altas chances de sobreviver aos processos determinísticos e estocásticos se suas respostas as variáveis ambientais forem distintas (GÁMEZ-VIRUÉS et al., 2015).

Entender como as espécies respondem ao ambiente em diferentes sistemas é fundamental para explicar a diversidade, mas também para guiar práticas de conservação. Nesse sentido, a precisão e acurácia na caracterização dos padrões ecológicos provenientes de dados de diversidade são fundamentais. O uso de métodos de coleta passivo e padronizado, bem como o aumento do esforço amostral podem ajudar a reduzir o viés causado pela detecção imperfeita, porém a sua correção pode ser indispensável quando medimos a biodiversidade. Estudos de ecologia de comunidade geralmente empregam desenhos amostrais replicados no espaço ou tempo, o que permitiria o emprego de modelos de ocupação ou *N-mixture*. A baixa utilização destas ferramentas de modelagem em ecologia de comunidades pode ser devido à complexidade dos modelos, e da não popularização da inferência bayesiana nas análises de ecologia de comunidade. Nessa tese utilizei os modelos hierárquicos multi-espécies de diferentes formas, demonstrando a alta maleabilidade dessa abordagem para acessar questões importantes dentro da ecologia de comunidades, e que não seriam possíveis de serem incluídas em arcabouços analíticos frequentistas mais tradicionais (p.ex. detecção imperfeita e coocorrência entre espécies). Frente a isso, finalizo ressaltando os benefícios do uso desse tipo de modelagem, as principais limitações das abordagens utilizadas neste estudo e as perspectivas das análises multi-espécies para ecologia de comunidade.

### *Vantagens da abordagem hierárquica bayesiana*

A inferência bayesiana pode ser usada para estimar parâmetros ecologicamente significativos ao mesmo tempo que fornece uma medida explicita da quantidade de incerteza da estimativa do parâmetro (ELLISON, 1996). Além disso, esse tipo de inferência pode ser usado tanto para avaliar modelos ecológicos como para testar hipóteses, uma vez que ela fornece uma medida quantitativa da probabilidade da hipótese ser verdadeira dado os valores observados (ELLISON, 2004). Adicionando uma estrutura hierárquica, permitimos que as observações, ou seja, as espécies, compartilhem informações entre si. Embora os modelos de distribuição de espécies “stacked” (*stacked* SDM), juntem as informações específicas para compor as respostas a nível de comunidades, eles não consideram que as espécies compartilham informações quando modelam os processos que determinam as ocorrências. Assim, os modelos hierárquicos multi-espécies surgem como uma ferramenta robusta quando queremos predizer a distribuição das espécies no espaço, principalmente por permitirem a estimativa confiável de parâmetros para espécies raras. Comunidades ecológicas são naturalmente dominadas por muitas espécies raras, logo, descartar espécies menos frequentes pode modificar a observação de padrões ecológicos em comunidades. Além disso, espécies raras estão geralmente no foco de estudos de conservação pois tem alto risco de extinção, sendo fundamental estimar como elas respondem as variações no espaço e no tempo.

### *Principais limitações*

Infelizmente, ainda não dispomos de uma teoria unificadora na ecologia de comunidades (OVASKAINEN; ABREGO, 2020) e seria exagero pensar que teríamos uma única forma de analisar a biodiversidade. Com isso, cada tipo de ferramenta analítica tem seus benefícios e suas limitações, e não é diferente com os modelos hierárquicos multi-espécies. Devido a sua alta flexibilidade em acomodar parâmetros, os modelos hierárquicos são menos generalizáveis e muitas vezes altamente complexos, o que torna o seu uso mais limitado. Ainda, é preciso assegurar a robustez dos modelos, uma vez que na maioria das vezes usamos informações *a priori* vagas, para que as estimativas feitas não sejam decorrentes do modelo pressuposto (análises de sensibilidade). Em relação a capacidade computacional, o desenvolvimento de técnicas de MCMC (Markov chain Monte Carlo) gerou um grande avanço na inferência dos parâmetros, pois permitiu caracterizar as distribuições posteriores através de simulações. No entanto, as MCMC podem consumir muito tempo até convergirem, principalmente em modelos complexos e com grande número de parâmetros.

Considerando a aplicabilidade, decidir qual tipo de abordagem usar depende da natureza dos dados e do objetivo de cada estudo. Geralmente dados de comunidades são obtidos através de amostragem repetidas no tempo e no espaço. Porém, nem sempre esse tipo de informação é disponibilizado nos estudos, restringindo nossa capacidade de aplicar por exemplo modelos de ocupação ou *N-mixture*, os quais precisam das repetições de modo a distinguir o processo amostral do processo biológico. Dessa forma, modelos que consideram coocorrência devem ser empregados quando temos dados oriundos de um desenho amostral que garanta coletas de espécies que coocorrem em uma dada área. Isso implica que embora dados de pontos de ocorrência possam ser usados nesse framework, as interpretações resultantes podem não ter um sentido biológico.

### *Perspectivas*

Enquanto os estudos de monitoramento da vida selvagem e de biologia populacional tem incorporado a detecção imperfeita em suas estimativas a mais de 20 anos (MacKenzie et al. 2002; Pellet 2008), estudos em ecologia de comunidade geralmente ignoram as consequências da detecção imperfeita sobre os padrões observados (Jarzyna and Jetz 2016; Benoit et al. 2018). Infelizmente, a estimativa dos erros amostrais ainda não foi implementada no HMSC, embora os autores reconheçam a importância da inclusão desse processo (OVASKAINEN; ABREGO, 2020), e com isso nos capítulos 2 e 3 desconsideramos a detecção imperfeita nos dados. Recentemente, estudos tem tentado incorporar tanto a detecção imperfeita como a coocorrência das espécies na estimativa de ocorrência e abundância (DORAZIO; CONNOR; ASKINS, 2015; TOBLER et al., 2019). Esse tipo abordagem tem potencial de abrir novas possibilidades para responder uma gama de questões relacionadas a como as espécies respondem ao ambiente, mas também as interações ecológicas, enquanto leva em consideração possíveis erros que possam prejudicar a observação do padrão.

# Referências

BECCALONI, G. W. et al. **Catalogue of the hostplants of the Neotropical butterflies**. Zaragoza: m3m : Monografías Tercer Milenio, 2008. v. 8

BÖRSCHIG, C. et al. Traits of butterfly communities change from specialist to generalist characteristics with increasing land-use intensity. **Basic and Applied Ecology**, v. 14, n. 7, p. 547–554, nov. 2013.

BROWN, K. S. J. **Geologic, evolutionary, and ecological bases of the diversification of neotropical butterflies: implications for conservation**. **Tropical Rainforests: Past, Present, and Future**, 2005.

BROWN, K. S.; FREITAS, A. V. L. Atlantic Forest Butterflies: Indicators for Landscape Conservation. **Biotropica**, v. 32, n. 4b, p. 934–956, dez. 2000.

CHASE, J. M.; MYERS, J. A. Disentangling the importance of ecological niches from stochastic processes across scales. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 366, n. 1576, p. 2351–2363, 2011.

CHAZOT, N. et al. Conserved ancestral tropical niche but different continental histories explain the latitudinal diversity gradient in brush-footed butterflies. **Nature Communications**, v. 12, n. 1, 2021.

CHECA, M. F. et al. Microclimate Variability Significantly Affects the Composition, Abundance and Phenology of Butterfly Communities in a Highly Threatened Neotropical Dry Forest. **Florida Entomologist**, v. 97, n. 1, p. 1–13, mar. 2014.

CHECA, M. F. et al. Rainfall as a significant contributing factor to butterfly seasonality along a climatic gradient in the Neotropics. **bioRxiv**, n. May, 2019.

DEVRIES, P. J. Stratification of fruit-feeding nymphalid butterflies in a Costa Rican rainforest. **Journal of Research on the Lepidoptera**, v. 26, n. 1–4, p. 98–108, 1988.

DORAZIO, R. M.; CONNOR, E. F.; ASKINS, R. A. Estimating the Effects of Habitat and Biological Interactions in an Avian Community. **PLOS ONE**, v. 10, n. 8, p. e0135987, 19 ago. 2015.

ELLISON, A. M. An Introduction to Bayesian Inference for Ecological Research and Environmental Decision-Making. **Ecological Applications**, v. 6, n. 4, p. 1036–1046, nov. 1996.

ELLISON, A. M. Bayesian inference in ecology. **Ecology Letters**, v. 7, n. 6, p. 509–520, jun. 2004.

FREITAS, A. V. L. et al. Studies with butterfly bait traps: An overview. **Revista Colombiana de Entomología**, v. 40, n. 2, p. 203–212, 2014.

GÁMEZ-VIRUÉS, S. et al. Landscape simplification filters species traits and drives biotic homogenization. **Nature Communications**, v. 6, 2015.

JARZYNA, M. A.; JETZ, W. Detecting the Multiple Facets of Biodiversity. **Trends in Ecology & Evolution**, v. 31, n. 7, p. 527–538, jul. 2016.

KÉRY, M.; ROYLE, J. A. Applied Hierarchical Modeling in Ecology Analysis of distribution, abundance and species richness in R and BUGS. Em: [s.l.] Elsevier, 2016. V. 1p. i.

MACKENZIE, D. I. et al. Estimating site occupancy rates when detection probabilities are less than one. **Ecology**, v. 83, n. 8, p. 2248–2255, 1 ago. 2002.

MITTELBACH, G. G.; MCGILL, B. J. **Community ecology**. [s.l.] Oxford University Press, 2019.

OVASKAINEN, O. et al. How to make more out of community data? A conceptual framework and its implementation as models and software. **Ecology Letters**, v. 20, n. 5, p. 561–576, maio 2017.

OVASKAINEN, O.; ABREGO, N. **Joint Distribution Modelling (with applications in R)**. [s.l: s.n.].

POGGIATO, G. et al. On the Interpretations of Joint Modeling in Community Ecology. **Trends in Ecology and Evolution**, v. 36, n. 5, p. 391–401, 2021.

RAWAT, U. S.; AGARWAL, N. K. Biodiversity: Concept, threats and conservation. **Environment Conservation Journal**, v. 16, n. 3, p. 19–28, 2015.

SHIREY, V. et al. LepTraits 1.0 A globally comprehensive dataset of butterfly traits. **Scientific Data**, v. 9, n. 1, p. 382, 6 dez. 2022.

SIQUEIRA, T. et al. Community size can affect the signals of ecological drift and niche selection on biodiversity. **Ecology**, v. 101, n. 6, p. 1–10, 2020.

TOBLER, M. W. et al. Joint species distribution models with species correlations and imperfect detection. **Ecology**, v. 100, n. 8, p. 1–14, 30 ago. 2019.

TSCHARNTKE, T. et al. Landscape perspectives on agricultural intensification and biodiversity - Ecosystem service management. **Ecology Letters**, v. 8, n. 8, p. 857–874, 2005.

XING, S. et al. Cool habitats support darker and bigger butterflies in Australian tropical forests. **Ecology and Evolution**, v. 6, n. 22, p. 8062–8074, 2016.